

Sampling Strategies in Density-Based Sensitivity Analysis

W. Castaings^{a,b,*}, E. Borgonovo^c, M.D. Morris^d, S.Tarantola^e

^a*Lab. EDYTEM (CNRS, Université de Savoie), Campus scientifique, F-73376 Le Bourget du Lac, France*

^b*Université de Toulouse, INPT, UPS, IMFT (Institut de Mécanique des Fluides de Toulouse), F-31400 Toulouse, France*

^c*Dpt. of Decision Sciences and ELEUSI, Bocconi University, Via Roentgen 1, 20135, Milano, Italy*

^d*Dpt. of Statistics, Iowa State University, Snedecor Hall, Ames, IA, 50011*

^e*Joint Research Center of the European Commission, Via Enrico Fermi 1, 21027, Ispra (VA), Italy*

Abstract

Decision and policy-makers benefit from the utilization of computer codes in an increasing number of areas and applications. Several authorities and agencies recommend the utilization of proper sensitivity analysis methods in order to confidently entrust model results. In this respect, density-based techniques have recently attracted interest among academicians and practitioners, for their property to characterize uncertainty in terms of the entire distribution of an output variable. However, their estimation is a challenging task and, without a proper methodical approach, errors in the estimates can lead to misleading conclusions. In this work, we propose sampling plans for reducing the computational burden of sensitivity estimates while improving and controlling the accuracy in the estimation. We compare designs based on column substitutions and designs based on permutations. We investigate their behavior in terms of type I and type II errors. We apply the methods to the Level E model, a

*Corresponding author

Email addresses: william.castaings@univ-savoie.fr (W. Castaings),
emanuele.borgonovo@unibocconi.it (E. Borgonovo), mmorris@iastate.edu (M.D. Morris),
stefano.tarantola@jrc.it (S.Tarantola)

computational tool developed by the Nuclear Energy Agency of the OECD for the assessment of nuclear waste disposal sites. Results show that application of the proposed sampling plans allows one to obtain confidence in the sensitivity estimates at a number of model runs several orders of magnitude lower than a brute-force approach. This assessment, based upon the entire distribution of the model output, provides us with ways to effectively reduce uncertainty in the model output, either by prioritizing the model factors that need to be better known or by prioritizing the areas where additional modelling efforts are needed. *Keywords:* uncertainty, global sensitivity analysis, importance measures, density-based, moment independent, computer experiments, sampling design.

1. Introduction

Computer models play a central role in studying physical phenomena and in supporting the decision-making process in several disciplines. Especially in the environmental, climate and physical sciences, scientific models are complex machines (Craig et al. 2001, Drignei and Morris 2006). The intricacy of the phenomena under investigation, their space and time scales and the variety of features that models aim at capturing, make it impossible to obtain a straightforward (let alone analytical) understanding of the input-output relationship. Lack of transparency limits analyst's ability in defending inferences from criticism and can lead to model rejection by stakeholders (see Stokstad (2008) for an example in the food industry). Several regulatory bodies and institutions [the United States Environmental Protection Agency, (US EPA 2009), the European Commission (European Commission 2009; p. 24), the White House office of Management and Budget (Saltelli 2009, White House 2006)] explicitly recommend the utilization of uncertainty and global sensitivity analysis methods as part of model audit and validation. By probabilistic sensitivity, in facts,

analysts are helped in *pinpointing which assumptions are appropriate candidates for additional data collection to narrow the degree of uncertainty in the results* (White House 2006). In the environmental literature, sensitivity analysis methods are the subject of increased attention by both academicians and practitioners (Campolongo and Braddock 1999, He et al. 2000, Newham et al. 2003, Pappenberger et al. 2006, Campolongo et al. 2007, Norton 2008, Ziehn and Tomlin 2009, Saltelli and Annoni 2010, Ravalico et al. 2010, Confalonieri et al. 2010, Nossent et al. 2011).

Scientists have shown interest in the analysis of datasets that are obtained from computer codes since the early 1990s. The works of Sacks et al. (1989a;b), Welch et al. (1992) have pioneered future developments in the solutions of issues as: *i*) approximating the input/output mapping provided by the computer model with simplified constructs (surrogate models) (Sacks et al. 1989b, Friedman and Stuetzle 1981, Friedman 1991, Santner et al. 2003, Sudret 2008, Bayarri et al. 2009, Marrel et al. 2009, Ratto and Pagano 2010) ; *ii*) characterizing the uncertainty of predictions given the uncertainty in factors (uncertainty analysis) (see Helton et al. (2006) for a review); and *iii*) identifying the factors that mostly affect the output uncertainty (global sensitivity analysis, see Saltelli et al. (2000), Oakley and O'Hagan (2004)). In this work, we focus on item *iii*), namely the identification of key-uncertainty drivers by means of global sensitivity analysis methods. Our numerical experiments are performed through the Level E computer code (OECD 1989). LevelE is a case study designed by the OECD Nuclear Energy Agency for the performance assessment of nuclear waste disposal sites. By its frequent use in sensitivity studies, LevelE has become the benchmark for sensitivity analysis and , practically , all global sensitivity methods have been tested through LevelE [non-parametric methods in Saltelli and

Marivoet (1990), variance-based methods in Saltelli and Tarantola (2002), and a new type of emulator associated with variance-based methods in Ratto et al. (2007)]. The first benchmark study was launched specifically on this model by OECD (OECD 1993), with the purpose of ranking uncertain factors in order of importance. The results of that benchmark were difficult to interpret because of lack of a rigorous definition of importance. Working on Level E, Saltelli and Tarantola (2002) addressed the problem in a rigorous fashion, identifying the so-called variance-based sensitivity analysis settings.¹

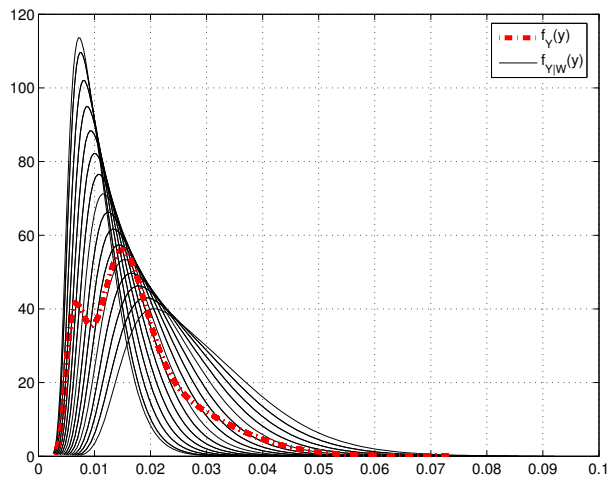


Figure 1: Unconditional and conditional densities of the maximum radiological dose simulated by the Level E model. The conditional densities are obtained for given values of the streamflow rate W . Due to the large positive skewness of the model output probability density function, a monotonic transformation ($y^{1/4}$) is displayed.

Figure 1 displays probability density functions of the main output of LevelE, namely the maximum radiological dose released to humans over a given time-

¹SA settings are specific questions that the sensitivity analysis is supposed to answer. Two examples are: *Which is the factor that should be fixed to achieve the greatest reduction of the output uncertainty?*; and *What is the minimal subset of factors that one should fix in order to achieve a prescribed reduction in the output uncertainty?*. These questions have been called later on *factor prioritization* and *variance cutting* (Saltelli et al. 2004).

frame. The conditional densities are obtained by fixing one of the uncertain factors (stream flow rate W , in the specific case). The distributions in Figure 1 are skewed and can be multimodal. Under these circumstances, several works have underlined that variance is not a sufficient descriptor of uncertainty (Soofi 1994, Oakley 2009). Alternative (to variance) approaches for measuring the effect of a random factor on the model output based on high-order conditional moments are presented in Ratto et al. (2009). Approaches that do not rely on any specific moment but, instead, measure the separation between the conditional and unconditional densities have been introduced by several authors (Park and Ahn 1994, Krykacz-Hausmann 2001, Auder and Iooss 2008, Chun et al. 2000, Borgonovo 2007, Liu and Homma 2009, Borgonovo et al. 2011b;a). Some authors rely on a distance-based metric (Chun et al. 2000, Borgonovo 2007, Liu and Homma 2009) others on an entropy-based one (Park and Ahn 1994, Krykacz-Hausmann 2001, Auder and Iooss 2008). However, independently of the selected metric, the estimation of density-based sensitivity statistics requires the assessment of the unconditional and conditional densities of the model output. A brute force estimation is associated with a computational cost M of the order of $k * N^2$ model runs (where k is the number of uncertain factors and N is the sample size). In this work, we present and compare alternative designs for the estimation of density-based sensitivity statistics, with the purposes of reducing computational burden and improving estimation accuracy. The first design rests on the combination of column substitution and quadrature (Davis and Rabinowitz 1984). This design achieves an overall computational cost of $M = k * N_{ext} * N$, with $N_{ext} \ll N$.² The second method rests on column per-

²This design is thoroughly discussed for the first time in this work. However, it is employed in Borgonovo et al. (2011a), where the estimation of density-based sensitivity measures through emulators is studied.

mutations (Morris et al. 2006; 2008) for the generation of conditional samples. A permutation-based scheme eliminates the dependence of the numerical cost on the number of model factors k . M is, then, reduced to $(r + 1) * N$, where r is the number of replicates. Two permuted-column sampling plans are proposed, the first relying on random permutations [replicated Latin Hypercube sampling (McKay 1995)], the second on deterministic sampling based on orthogonal arrays (Morris et al. 2008). Robustness of the sampling plans is tested first on analytical case studies. Convergence is obtained for all the proposed sampling plans as M increases. Nevertheless, spurious estimates of the sensitivity measures of the least influential factors are registered in the rLHS design, especially for low sample sizes. To reduce the bias, we study a sampling plan in which orthogonal arrays replace random permutations. Results show an evident reduction of the low-sample-size bias. We then challenge the proposed designs through of computationally intensive model, namely, LevelE. In this case, the determination of the degree of confidence in the numerical estimates is obtained by bootstrap (Efron (1979); see Davison et al. (2003) for a review). In sensitivity analysis, assessing as important a non important factor corresponds to a Type I error and assessing as non important an important factor is a Type II error. Both the analytical test cases and the LevelE results reveal the following. *i*) A notable reduction in computational burden is achieved by using the proposed approaches in respect to current practice; *ii*) substitution-based sampling plans identify the least influential factors already at a small sample size with good accuracy, buy they are prone to type I errors; conversely, permutation methods identify the most important factors at a reduced sample size, but they are prone to type II errors. We discuss that, by combining the advantages of the two approaches, one obtains the identification of the key-uncertainty-drivers at a relatively small sample size, albeit in the presence of an intensive model as

LevelE.

The remainder of the paper is organized as follows. The sensitivity measures of interest are described in section 2. The sampling plans are proposed and discussed in section 3. In section 4, we assess numerically the performance of the sampling strategies on several test cases. In section 5, the most efficient strategies are compared in the context of the LevelE model. Finally, section 6 offers conclusions.

2. Density-based sensitivity analysis

In this section, we briefly review the principles of global sensitivity analysis.

Some notation is offered first. Let $\Omega_X \subseteq \mathbb{R}^k$ denote the model input space, \mathbf{X} the random model factor vector and \mathbf{x} one of its realizations. μ_X is the probability measure that reflects the decision-maker's degree of belief on \mathbf{X} .

The joint probability density of \mathbf{X} has to be provided by the analyst. Using the scientific literature, expert judgment or physical experiments, it is usually provided as marginal distributions (noted f_{X_i}) supplemented by a correlation matrix when factors are correlated.

$y = g(\mathbf{x})$, $g : \Omega_X \subseteq \mathbb{R}^k \rightarrow \Omega_Y \subseteq \mathbb{R}$, is the relationship that links the model factors to the model output. It can be a simple expression or a complex computer code. Ω_Y , the image of g , coincides with the model output (y) support. Uncertainty in \mathbf{x} makes y a random variable, which we denote by Y . As in Gelfand and Smith (1990), we assume that densities for both the model factor and output and for all marginal and conditional distributions exist.

Global sensitivity methods can be categorized as follows. Screening methods have the purpose of identifying the non-influential model factors through a limited number of model runs (Morris 1991, Campolongo et al. 2007). For the identification of the most relevant factors, Morris method can provide results

comparable to those obtained with more computationally intensive techniques (Confalonieri et al. 2010). However, this method can produce inaccurate measures for non-monotonic functions when their characteristic length of variation is much smaller than the step size used for building the trajectories exploring the input space (Sobol and Kucherenko 2009). Non-parametric methods make use of correlations and rank-correlations to identify the main uncertainty drivers (Saltelli and Marivoet 1990, Helton 1993, Storlie et al. 2009). Variance based methods are based on the functional ANOVA decomposition of the model output (Efron and Stein 1981, Takemura 1983, Sobol’ 1993, Rabitz and Alis 1999, Owen 2003, Sobol’ 2003) and are among the most widely utilized ones (Saltelli et al. 2000, Saltelli and Tarantola 2002, Oakley and O’Hagan 2004). The founding SA setting of variance-based methods is established in Saltelli and Tarantola (2002): “*We are asked to bet on the factor that, if determined (i.e., fixed to its true value), would lead to the greatest reduction in the variance of Y* [Saltelli and Tarantola (2002), p. 705].” The associated sensitivity measures are defined as (Iman and Hora 1990, Homma and Saltelli 1996, Saltelli and Tarantola 2002)

$$S_i = \frac{V[Y] - \mathbb{E}_{X_i}[V\{Y|X_i\}]}{V[Y]} = \frac{V_{X_i}[\mathbb{E}\{Y|X_i\}]}{V[Y]} \quad (1)$$

where $V[Y]$ is the model output variance. $S_i > S_j$ implies that fixing X_i leads on average to a greater reduction in $V[Y]$ than fixing X_j .

Under the assumption $\mu_{\mathbf{X}} = \prod_{i=1}^k \mu_{X_i}$ (factor independence) the sensitivity measures in eq. (1) are generalized to global sensitivity indices of order m , defined as

$$S_{i_1, i_2, \dots, i_m} := \frac{V_{i_1, i_2, \dots, i_m}}{V[Y]} \quad (2)$$

where

$$V_{i_1, i_2, \dots, i_m} = \int g_{i_1, i_2, \dots, i_m}^2 d\mu_{i_1} d\mu_{i_2} \dots d\mu_{i_m} \quad (3)$$

is a partial variance in the functional ANOVA expansion of g (Hoeffding 1948)

which is defined as

$$g = g_0 + \sum_{m=1}^k \sum_{i_1 < i_2, \dots, < i_m} g_{i_1, i_2, \dots, i_m}(x_{i_1}, x_{i_2}, \dots, x_{i_m}) \quad (4)$$

In Equation 4, $g_0 = \mathbb{E}[g]$ and the functions $g_{i_1, i_2, \dots, i_m}(x_{i_1}, x_{i_2}, \dots, x_{i_m})$ are obtained by conditional expectations and nested subtractions following a Gram-Schmidt orthogonalization procedure (Sobol' 1969, Efron and Stein 1981, Takemura 1983).

Besides first and higher order sensitivity indices, of particular relevance are the total order indices (S_i^T). They are formally defined as

$$S_i^T := \frac{\mathbb{E}[V(Y|\mathbf{X}_{\sim i})]}{V[Y]} = 1 - \frac{V[\mathbb{E}(Y|\mathbf{X}_{\sim i})]}{V[Y]} \quad (5)$$

where $\mathbf{X}_{\sim i}$ denotes the set of all model factors excluding X_i . S_i^T is the expected variance that would be left if all factors but X_i could be fixed

Being eqs. (2) and (5) the expectation of conditional moments, the numerical estimation of all sensitivity indexes (up to order k) involves a double loop computation to be repeated $2^k - 1$ times, for a total of $k \cdot N^2 \cdot (2^k - 1)$, where N is the Monte Carlo sample size. Such cost would make their estimation infeasible for most realistic models. Several works have then focussed on their estimation and led to a notable reduction in computational burden (Jansen et al. 1994, Homma and Saltelli 1996, Saltelli et al. 1999; 2000, Saltelli 2002, Saltelli et al. 2010). Saltelli et al. (2010) provide a recent review and comparison of existing and new practices. Theorem 1 in Saltelli (2002) states that, under the independence assumption, it is possible to compute S_i , S_i^T and each of the $\binom{k}{2}$ second order indices, V_{i_t, i_s} , at the cost of $N(k+2)$ model evaluations. Variance-based sensitivity measures have been extensively studied both from the theoretical and numerical viewpoints (Cukier et al. 1978, Efron and Stein 1981, Bedford 1998, Rabitz and Alis 1999, Saltelli et al. 2000, Saltelli and

Tarantola 2002, Owen 2003, Oakley and O’Hagan 2004) . Similar considerations cannot be stated for the estimation of density-based sensitivity methods. The growing attention and utilization of these methods by both academicians and practitioners makes research on this aspect necessary. However, when the factors are dependent, Bedford (1998) shows that the value of V_{i_1, i_2, \dots, i_m} depends on the order with which factors are considered. Consequently, sensitivity indices of order 2 or higher are not univocally defined, making the creation of a corresponding sensitivity setting cumbersome³. Also, variance is sufficient to characterize uncertainty under either *a*) the assumption of a quadratic utility function of the decision-maker (Oakley (2009)); or *b*) the assumption of normal distributions (Huang and Litzenberger (1998); p. 61). However, contradictory results can be obtained if either of these assumptions fail. In Borgonovo (2006) , a test case is presented where a decision-maker representing uncertainty by means of variance would refuse perfect information about an uncertain factor. In Soofi (1994)[p. 1244], it is shown that *variance cannot be used for comparing uncertainties associated with two Pareto distributions when $a < 2$ for one or both distributions*. If $a < 2$, in fact, the variance of a Pareto random variable is infinite. Nonetheless, comparison of uncertainties associated with any two Pareto variables is always possible using density separation (Soofi 1994). In fact, Figure 1, shows that the effect of fixing W is a modification in the entire density of the model output. The rationale of density-based sensitivity measures is, then, to quantify the change in shape between the conditional and unconditional model output densities, without reference on a particular moment (*e.g.*, variance). In this way, one can also capture oscillations in higher order moments as skewness, kurtosis, etc. The corresponding setting is, then: *We are asked to bet on the*

³Research for coming to a unified definition of variance-based sensitivity measures is ongoing (Rabitz 2010, Mara and Tarantola 2011)

model factor that, if determined, would lead to the greatest expected modification in the distribution of Y (Borgonovo and Tarantola 2008). To the authors’s knowledge, the first work introducing a global sensitivity measure based on density separation is Park and Ahn (1994), where the Kullback-Leibler divergence is employed. This intuition is next generalized by Chun et al. (2000), where the Minkowski-distance of order 2 is used and in Borgonovo (2006) and Borgonovo (2007) — an historical overview in the development of distributional sensitivity methods is offered in Borgonovo et al. (2011b). The sensitivity measure of interest in this work is:

Definition 1.

$$\delta_i := \frac{1}{2} \mathbb{E}_{X_i} [s_i(X_i)] = \frac{1}{2} \int f_{X_i}(x_i) s_i(x_i) dx_i \quad (6)$$

where

$$s_i(x_i) := \int |f_Y(y) - f_{Y|X_i=x_i}(y)| dy \quad (7)$$

δ_i [eq. (6)] has the following interpretation. $s_i(x_i)$ quantifies the shift between the unconditional and the conditional output distribution given that X_i is fixed at x_i . As proven in Glick (1975), the operation $\int (|f_Y(y) - f_{Y|X_i}(y)|) dy$ is a separation measurement with respect to the L^1 norm, for the set of all probability densities. More generally, one can use the Minkowski-distance of order p : $\int (|f_Y(y) - f_{Y|X_i}(y)|^p)^{1/p} dy$. For instance, $p = 2$ is used in Chun et al. (2000) (for a thorough overview of alternative metrics in distributional sensitivity analysis, we refer to Borgonovo et al. (2011b)). The choice $p = 1$, however, grants one with the advantage that the sensitivity measure has a direct geometrical interpretation and possesses attractive normalization and scale invariance properties. Geometrically, $s_i(x_i)$ represents the area between $f_Y(y)$ and $f_{Y|X_i}(y)$ (Fig. 2).

δ_i possesses the following properties (Borgonovo 2007, Borgonovo et al. 2011b):

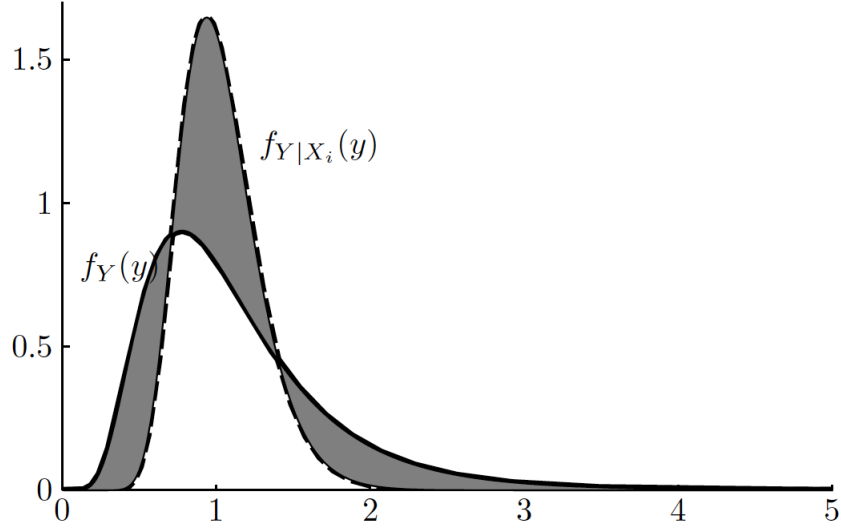


Figure 2: Geometric interpretation of $s_i(X_i)$

1. individual normalization: $0 \leq \delta_i \leq 1$, with $\delta_i = 0$ if and only if Y is independent of X_i ;
2. joint normalization ;
3. scale invariance.

The first property insures that the importance of an uncertain factor is normalized between zero and unity. If Y is unaffected by X_i (whatever the value), $\delta_i = 0$. Property 2 states that the distance between the unconditional distribution of Y and the distribution corresponding to the situation in which all model factors are known is unity. This is explained as follows. The $\delta_{1,2,\dots,k}$ for getting to know all factors is given by:

$$\delta_{1,2,\dots,k} = \frac{1}{2} \mathbb{E}_X \left[\int |f_Y(y) - f_{Y|X_1=x_1, X_2=x_2, \dots, X_n=x_k}(y)| dy \right] \quad (8)$$

where $f_{Y|X_1=x_1, X_2=x_2, \dots, X_k=x_k}(y)$ is the density of Y given that all model factors are fixed at \mathbf{x} . This density is, then, Dirac δ -density centered at $y = g(\mathbf{x})$.

One can then show that the distance between any density and the Dirac- δ -density is equal to 2, independently of \mathbf{x} (unity after normalization). Property 3 follows by the invariance for monotonic transformation of the L^1 -norm. Scale invariance is particularly useful in practical applications when the numerical model output spans several orders of magnitude. A technique often adopted by analysts to improve numerical processing is re-scaling (typically one uses log-transformations). For instance, Iman and Hora (1990) underline a lack of robustness in the estimation of variance-based sensitivity measures, which are highly influenced by outliers in the presence of fat-tailed input/output distributions. In order to circumvent this problem, Iman and Hora (1990) employ a log transformation. However, as emphasized by Saltelli and Sobol' (1995), and investigated in detail for rank transformation, the outcomes of a variance-based sensitivity analysis on re-scaled data cannot be easily transferred back to the original model. By the scale invariance property, δ does not suffer from this limitation. In fact, it assumes the same values if computed on y or on data obtained through any monotonic transformation of y (*e.g.*, $\log y$). The transformed data can then be fully exploited in applications (see section 5).

Furthermore, consider all situations in which the model output acts as a decision-support criterion, and is used by policy-makers to make environmental decisions. Then, most generally, the decision-maker preferences are represented by assessing a von Neuman-Morgenstern utility function over y , that we denote by $u(y)$. In many situations, however, it is difficult to assess the precise functional form of $u(y)$. However, monotonicity of $u(y)$ is a standard requirement. In fact, if $u(y)$ and $t(y)$ are two monotonic functions, then they are equivalent representations of a decision-maker preferences. Then, if a sensitivity measure is scale invariant, it produces the same ranking for $u(y)$ and $t(y)$. That is, the ranking obtained by computing the sensitivity on y is unaffected by imprecisions in the

assessment of $u(y)$. These properties are connected with the scale-invariance of the L^1 -norm and do not hold if $p \neq 1$.

Finally, we emphasize that δ_i and S_i/S_T are complementary sensitivity measures. δ_i is a measure of statistical dependence and allows one to detect whether a factor is influential on the model output, no matter what moment of the distribution she is considering. In this respect, δ_i allows one to avoid type I errors. Conversely, S_i/S_T by their link to the functional ANOVA decomposition allow one to obtain insights on model structure (e.g., interactions).

As mentioned in the introduction, the estimation of density-based sensitivity measures can be challenging in the presence of model output generated by large computer codes. In the next section, we address the principles of numerical estimation of δ_i and propose ways for reducing the computational burden.

3. Description of alternative sampling plans

In this section, the computational and numerical aspects underlying the estimation of density-based importance measures are discussed. It is readily noted that a key-role is played by the estimation of the internal-loop statistics, $s_i(x_i)$, in eq. (7). In the estimation of δ_i , an essential aspect is the generation of conditional and unconditional model output vectors that allow an efficient approximation of the conditional ($f_{Y|X_i=x_i}(y)$) and unconditional model output densities ($f_Y(y)$). In fact, $f_{Y|X_i=x_i}(y)$ and $f_Y(y)$ are essential to insure that the internal statistics ($s_i(x_i)$) is properly computed. In a numerical estimation, $f_{Y|X_i=x_i}(y)$ and $f_Y(y)$ are obtained by processing the conditional and unconditional model output vectors using kernel density estimation (Parzen 1962, Silverman 1986). However, the problem is generating samples of the uncertain factors at which to evaluate the model to obtain the conditional and unconditional output vectors efficiently. In the remainder, therefore, we focus on the

description of the sampling strategies. The terminology used by Morris et al. (2006) is adopted in order to classify the different sampling plans evaluated in this paper.

3.1. Substituted column sampling plans

The family of plans we discuss in this subsection is based on substituted column (*substituted columns*) sampling. We describe a basic substituted column design first.

Let A_0 denote the unconditional sample of size N (A_0 an $N \times k$ matrix) drawn from $f_{\mathbf{X}}(x)$ using any (quasi) random sampling scheme. Evaluating the model in correspondence of A_0 , one obtains the unconditional model output vector b_0 . b_0 can then be used for the estimation of $f_Y(y)$ through kernel density.

In order to obtain the conditional model output vectors required for the estimation of $f_{Y|X_i=x_i}(y)$ (for different values x_i), one needs to generate a conditional input sample from $f_{X_i|x_i}(x)$. Assuming that $x_i^{(j)}$ is the value of X_i sampled for the j^{th} Monte Carlo run, let us denote by $A_i^{(j)}$ the corresponding conditional input sample and by $b_i^{(j)}$ the calculated model output vector. To obtain the conditional input sample $A_i^{(j)}$, one needs to distinguish the cases of dependent and independent model factors. In the latter, one can simply fix all elements of the i^{th} column of A_0 at $x_i^{(j)}$ to obtain $A_i^{(j)}$. In the former, one has to re-sample from the conditional factors distribution given that $X_i = x_i^{(j)}$. Repeating this procedure for $j = 1, 2, \dots, N$, one obtains N conditional sample matrices $A_i^{(j)}$ of size $N \times k$. The operation needs then also to be repeated for the k model factors. Correspondingly, the computational cost associated with the substituted columns design is $M = N(1 + kN)$ model runs (Table 1). This cost makes the estimation prohibitive for most models as N or k grow, unless the computational time is negligible (this could be done by substituting the model

Sampling plans	Tot. numb. of eval. (M)
Substituted Columns	$M = N(1 + kN)$
Improved Substituted Columns	$M = N(1 + kN_{ext})$ $N_{ext} \ll N$
Permuted columns	$M = N(1 + r)$ with $r \simeq N$

Table 1: Summary of the computational cost for the various sampling plans with k number of model factors, N unconditional sample size, N_{ext} number of explored conditional values and r number of replicates for rLHS

through a surrogate model; see Borgonovo et al. (2011a)). However, a first way for lowering M is to use a quadrature method (Davis and Rabinowitz 1984) to estimate the one-dimensional integral given by eq. 6

The number of conditional input points $x_i^{(j)}$ to be explored for an accurate estimation of δ_i is then limited to $N_{ext} \ll N$. The total number of model evaluations associated with this *improved substituted columns* sampling plan is $N(1 + kN_{ext})$ (Table 1). In our numerical experiments we shall adopt a Gauss-Legendre quadrature proceeding as follows. The N_{ext} abscissas used for quadrature are first given by the roots of Legendre polynomials and then transformed using the inverse of the marginal cumulative distribution function of X_i .

We observe that both substituted columns and improved-substituted column sampling plans are independent of the random number generation method. Therefore, one can adopt stratification procedures (McKay et al. 1979) or low discrepancy sequences (Sobol' 1976) to improve the numerical accuracy of δ_i . In this work, both techniques are used and compared.

Let us denote by X_{-i} all factors but X_i . An interesting feature of substituted column sampling plans is that when the model factors are independent the same sampling points for X_{-i} are used for the calculation of the outputs b_0 and $b_i^{(j)}$ (*i.e.* the base case *vs.* the conditional case). The resulting advantage is that any change between $f_Y(y)$ and $f_{Y|X_i=x_i^{(j)}}(y)$ is due to $X_i = x_i^{(j)}$ and not spuriously

provoked by sampling variability⁴ for the factors X_{-i} .

This prevents the occurrence of bias. In fact, given that the L^1 -norm is used for the difference between the unconditional and conditional densities, if X_i is a dummy factor which does not play any role in the calculation of Y , the statistic δ_i cannot be strictly null if different sampling points are used for the factors X_{-i} (see Sections 3.2 and 4).

A limitation, common to both substituted columns and improved-substituted column sampling plans, is that the conditional sample matrices $A_i^{(j)}, \forall j = 1, 2, \dots, N$ are used only once for estimating δ_i . Consequently, the total computational cost depends on the number of factors k . A way for reducing the computational burden is to make use of permuted column sampling.

3.2. Permuted columns sampling plans

McKay (1995) introduces a permuted column (PC) sampling plan for the numerical estimation of first-order variance-based importance measures, S_i . The proposed design, named replicated Latin Hypercube sampling, is denoted rLHS in the remainder of the paper. It relies on a base Latin Hypercube sample A_0 of size N and on r additional matrices (*i.e.* $\{A_1, \dots, A_r\}$ $N \times k$ matrices) obtained by r independent permutations of all columns of A_0 . Each replicate of N sample points is therefore constructed by random combination of the set of sampled values available of each factor X_i in the base sample A_0 . We now show that rLHS can be used for the estimation of density-based sensitivity measures. In fact, by construction, each of the arrays A_j ($j = 1, \dots, r$) contains the same N values of X_i that are matched with different values of the other factors. By sorting the rows of the whole sample matrix according to the values of X_i , the entire sample can be seen as made of N groups of r points sharing the same

⁴Sampling variability refers to how much the estimate varies from sample to sample.

value $x_i^{(j)}$. This is valid for all factors X_i .

With rLHS, the number of required replicates r is equal to the number of points considered over a conditional value (*i.e.* the sample size used for approximating $f_{X|X_i}(x)$ in Eq. 7). Therefore, r has to be set not too far from N (see section 4).

The total number of sample points required for the computation of δ is independent of the number of model factor k . The computational cost is, therefore, reduced to $N(1 + r)$. As emphasized by Morris et al. (2006), an attractive feature of this sampling strategy is that all model runs (*i.e.* the whole sample of size $N(r + 1)$) contribute to the estimation of each sensitivity index S_i .

Because the number of model evaluations characterizing rLHS is independent of the number of model factor k , this method can be cheaper than improved-substituted column for high values of k . The reader is referred to Table 1 for a comparison of the numerical costs.

Conversely to both substituted columns and improved-substituted column, the limitation of rLHS is that permutations lead to spurious changes between $f_Y(y)$ and $f_{Y|X_i=x_i^{(j)}}(y)$, which are not necessarily provoked by conditioning on a model factor. As we are to see, a dummy variable can be associated with a non-null (albeit small) value of the global sensitivity measures (both variance-based and density-based).

As noted, the motivation for the randomized rLHS is that random recombination of N distinct values for each factor yields a collection of r runs over which any one factor can be fixed, while the others vary randomly. However, because N is finite, distinct combinations of pairs (or even triples, et cetera) of factors can appear by chance in these arrays. The result is that estimates of sensitivity indices are biased under these plans, and this bias is relatively larger for small N and/or large k . Morris et al. (2008) showed how this bias could be eliminated

by using orthogonal arrays of strength 2, rather than random permutation, as the basis for recombining factor values in these plans. Briefly, an orthogonal array of strength 2 is, in our context, an array or matrix in which each column contains N distinct values, and for which, for every pair of columns, every combination of N^2 pairs of values appear together in the same number of rows. As a result, the total number of rows in the orthogonal array must be a multiple of N^2 . The approach calls for constructing an orthogonal array of strength 2 in $k + 1$ columns. For each of the first k columns (independently), the N coding values are replaced by the mid-points of equal-probability intervals derived from the distribution specified for the corresponding factor. The last column is used to split the orthogonal array into N sub-arrays, r of which are retained and used as $\{A_1, \dots, A_r\}$ as described above. This retains the general structure of the rLHS, while assuring that distinct pairs of values are not repeated for any two factors across the collection of arrays, and eliminating the estimation bias associated with rLHSs in which the arrays are constructed through unconstrained random permutation within columns. There are a number of algebraic techniques available for constructing orthogonal arrays; an excellent general reference on this topic is the book by Hedayat et al. (1999). (For readers interested in more detail, we note that the particular construction technique we have used in the examples is described as Construction 2 by these authors, following Theorem 3.20 on page 50, but will not repeat that description here because it is well beyond the technical scope of this paper, and because any other construction technique would be equally valid for this purpose.)

The performance of the different sampling plans is discussed in the next section.

4. Comparative analysis of the sampling plans

In this section, we investigate the efficiency and accuracy of the sampling plans by means of analytical test cases.

The first test case consists of the additive model:

$$y = \sum_{i=1}^{10} X_i \quad (9)$$

with X_i distributed as follows: $X_1 \sim N(0, \sqrt{16})$, $X_2 \sim N(0, \sqrt{14})$, $X_3 \sim N(0, \sqrt{8})$, $X_4 \sim N(0, \sqrt{7})$, $X_5 \sim N(0, \sqrt{6})$, $X_6 \sim N(0, \sqrt{3})$, $X_7 \sim N(0, \sqrt{2})$, $X_8 \sim N(0, \sqrt{1})$, $X_9 \sim N(0, \sqrt{0.5})$, $X_{10} \sim N(0, \sqrt{0.5})$. This configuration of model factors and output allows us to obtain δ_i analytically (Proposition 4 in Borgonovo et al. (2011b)). In fact, letting

$$y = \sum_{i=1}^n a_i x_i \quad (10)$$

with $\mathbf{X} \sim N(\mathbf{x}, \mathbf{m}, \Sigma)$ with mean values $m_i = \mathbb{E}[X_i]$ and non degenerate covariance matrix Σ ($\det \Sigma \neq 0$), then

$$\delta_i = \mathbb{E}_{X_i} [N(y_1; m_Y, V_Y) + N(y_2; m_{Y|X_i}, V_{Y|X_i=x_i}) - N(y_2; m_Y, V_Y) - N(y_1; m_{Y|X_i}, V_{Y|X_i=x_i})] \quad (11)$$

where

$$\begin{aligned} V_Y &= \mathbf{a} \Sigma \mathbf{a}^T \\ V_{Y|X_i} &= \mathbf{a} \Sigma_{Y|X_i} \mathbf{a}^T \\ m_Y &= \sum_{s=1}^n a_s m_s \\ m_{Y|X_i} &= \sum_{s=1, s \neq i}^n a_s \left[m_s + (x_i - m_i) \frac{\sigma_{s,i}}{\sigma_i} \right], \quad i = 1, 2, \dots, n. \\ y_{1,2} &= \frac{1}{V_Y - V_{Y|X_i}} \left(V_Y m_{Y|X_i} - V_{Y|X_i} m_Y \pm \sqrt{V_Y V_{Y|X_i} \left[(a_i x_i)^2 + (V_Y - V_{Y|X_i}) \ln \left(\frac{V_Y}{V_{Y|X_i}} \right) \right]} \right) \end{aligned}$$

For our particular choice of the parameters one obtains

$$\{\delta_1, \dots, \delta_{10}\} = \{0.196, 0.180, 0.128, 0.119, 0.109, 0.074, 0.060, 0.042, 0.029, 0.029\}$$

This result shows that the model factors can be divided in groups of high (δ_1, δ_2) , moderate $(\delta_3, \delta_4, \delta_5)$, low $(\delta_6, \delta_7, \delta_8)$ and very low (δ_9, δ_{10}) importance. As we are to see, this allows one to compare the performance of the proposed sampling plans with respect to their ability in identifying the most (factor prioritization) or least (factor fixing) important model factors.

We start by analyzing substituted columns sampling plans (*i.e.* substituted columns and improved-substituted column). As we mentioned in Section 3, these designs could be obtained using any random generation scheme, which can be a crude Monte Carlo, a quasi-Monte Carlo or some form of stratified sampling.

We then compare the performance of substituted columns and improved-substituted column schemes using both standard Latin Hypercube Sampling (LHS) (McKay et al. 1979) and quasi-Monte Carlo (quasi-Monte Carlo) generation. The latter is obtained by using Sobol' sequences (Sobol' 1976). In order to establish the accuracy of the estimates a root mean square error (RSME) is computed. For each sample size, RMSE is calculated repeating the exercise 100 times. Figure 3 shows convergence results for $\delta_1, \delta_3, \delta_6, \delta_9$ (a representative of each category is shown for space reasons).

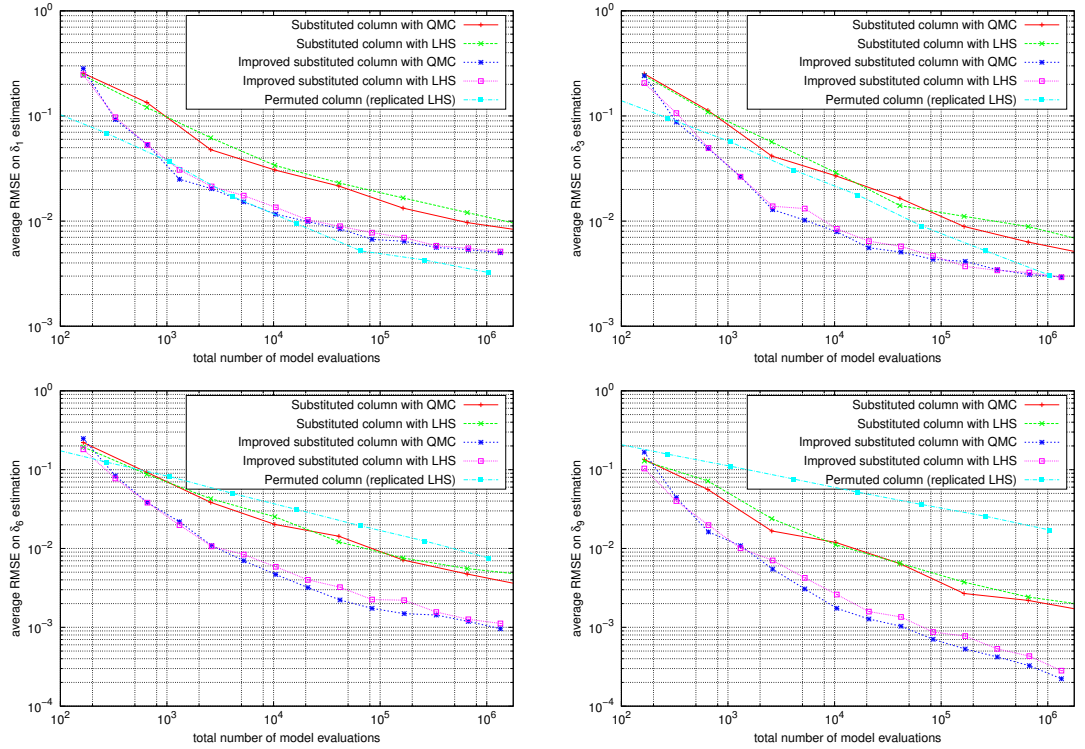


Figure 3: RMSE (average across 100 replicates) for the different sampling plans, for model factors δ_1 , δ_3, δ_6 , δ_9 .

The first four lines in Figure 3 represent the RMSE as a function of N for substituted columns and improved-substituted column with quasi-Monte Carlo and LHS, respectively. One notes that the RMSE tends to zero as N increases, with improved-substituted column outperforming substituted columns both when an LHS and a quasi-Monte Carlo scheme are used. The choice of the sampling method (LHS or quasi-Monte Carlo) does not affect the improved-substituted column performance significantly. Figure 3 also reports the results obtained with rLHS. In these experiments, we have set $r = N$ (number of replicates equal to the unconditional sample size) for the reasons discussed earlier and that we now investigate further.

Figure 3 shows that rLHS outperforms both improved-substituted column

and substituted columns for the most important model factor δ_1 . However, performances gradually deteriorate for factors featuring moderate to very low importance ($\delta_3, \delta_6, \delta_9$). Using rLHS the estimation of the change between unconditional and conditional densities is also influenced by sampling variability (*i.e.* not strictly due to conditioning on a given value of X_i) and a positive bias for the estimates of δ_i occurs.

To investigate this feature, we perform additional numerical experiments using a well-known sensitivity analysis case study, namely the Ishigami function (Ishigami and Homma 1990). The model output is given analytically by: $Y(X_1, X_2, X_3, X_4) = \sin X_1 + 7 \sin^2 X_2 + 0.1 X_3^4 \sin X_1$.

The four factors used for the calculation are X_i ($i = 1, \dots, 4$) $\sim U[-\pi - \pi]$. X_4 is a dummy factor. Because Y is independent of X_4 the value of δ_4 is known to be null. Figure 4 reports the results of the computational experiments.

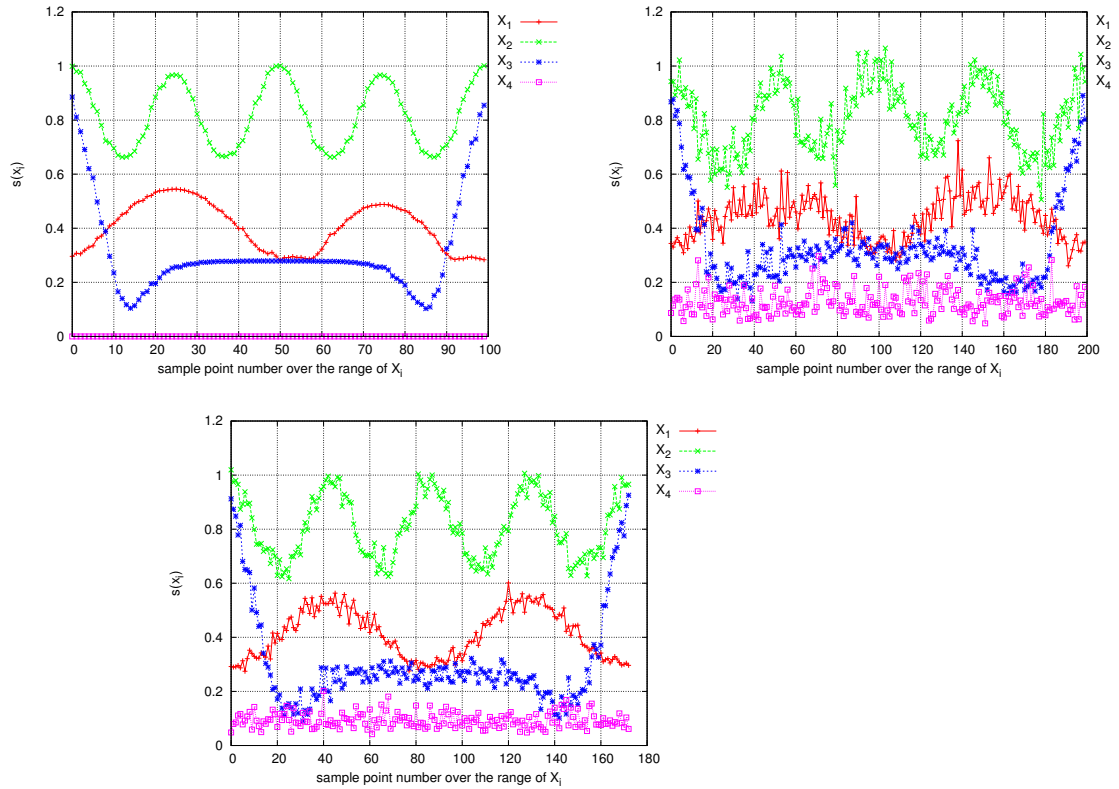


Figure 4: $s_i(x_i)$ across the model factors range for the Ishigami function, using substituted columns (top left figure where $s(x_4)$ is strictly null), rLHS (with $r = N$, top right figure) and Morris et al. (2008) design (with $r = N - 1$, bottom figure). In order to have a comparable total number of model evaluations, we set $N = 100$ for substituted columns, $N = 200$ for rLHS and $N = 173$ for Morris et al. (2008) design (see table 1).

Figure 4 displays the separation $s_i(x_i)$ (i.e. across the range) for the four model factors obtained with substituted columns, rLHS and Morris et al. (2008) design. For a comparable number of model runs M (i.e. $M = 100(1 + 4 \times 100) = 40100$, $M = 200(1 + 200) = 40200$ and $M = 173(1 + 172) = 29929$), the shape of the curves is much smoother using the substituted columns sampling plan. The noise observed in the curves obtained with rLHS is another consequence of the positive bias on δ_i which appears visible for less important factors. This

noise is less pronounced with Morris et al. (2008) design for a smaller total number of model runs. As a result, one notes that $s_4(x_4)$ is always null when using substituted columns, while it is non-null using permuted column sampling.

For the same case study, let us compare in more detail rLHS with the orthogonal array replication scheme of Morris et al. (2008), with the construction described in Section 3.2. Figure 5 reports the results.

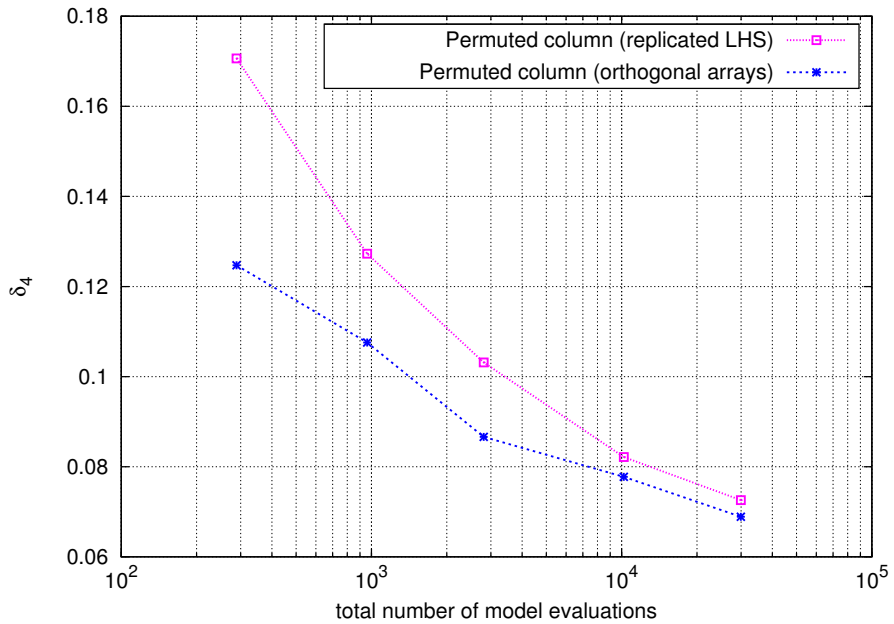


Figure 5: Results for the dummy variable of the Ishigmi model; comparison of the Morris et al. (2008) design with rLHS

Figure 5 reports the numerical estimates of δ_4 at 300, 1000, 3000, 10000, and 30000 model runs obtained using rLHS and Morris et al. (2008) designs. Because δ_4 is a dummy, its importance is expected to be null; this is indeed the case with both rLHS and the orthogonal array designs for increasing M . However, one notes the bias reduction obtained by orthogonal arrays, especially at a small sample sizes (low M).

Finally, we explore how different combinations of r and N in the rLHS sampling plan influence the estimation of δ considering a dummy factor X_4 in the Ishigami function.

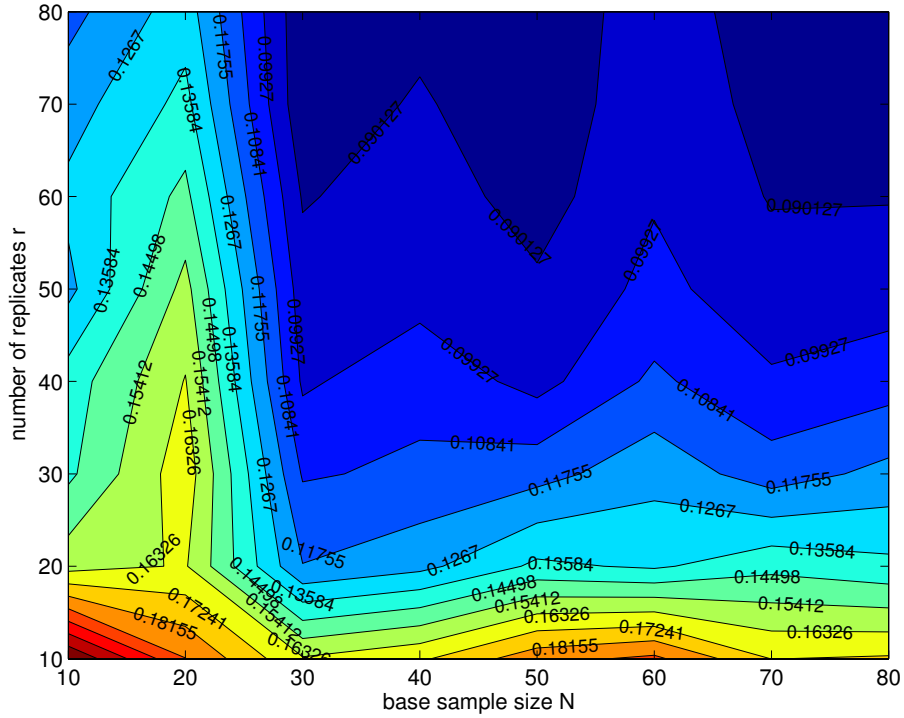


Figure 6: Error of density-based sensitivity measure for a dummy factor introduced for the Ishigami function.

Figure 6 shows the estimation error on δ_4 revealing that for a fixed value of N the approximation error significantly decreases when r grows towards N . Given the range of values explored for this specific experiment (N ranging from 10 to 80) and the convergence rate reported by figure 5, the reduction of the error is not really significant after $N = 30$.

In the remainder of the work, we shall maintain $r = N$. [We investigated the effects of the combination of N and r for other analytical test cases (not

reported here for space reasons) and the results confirm the observations stated above.]

In the next section, the performance of the proposed sampling plans is investigated by means of the benchmark model for sensitivity analysis, namely, LevelE.

5. Application to safety assessment for nuclear waste disposal

The Level E code⁵ is a case study designed to forecast the radiological dose to humans due to the underground migration of radionuclides from a nuclear waste disposal site over geological timescales (from 0 to 10^7 years). The radionuclides are iodine (^{129}I) and the neptunium, uranium and thorium chain ($^{237}Np \rightarrow ^{233}U \rightarrow ^{229}Th$). The repository is represented as a point source and the migration happens in two geosphere layers characterized by different hydro-geological properties. The processes being considered in the model are radioactive decay, dispersion, advection and chemical reaction between the migrating nuclides and the porous medium. The resulting model is then a set of coupled partial differential equations, whose analytical expression can be found in Saltelli and Tarantola (2002). Twelve uncertain model factors are present, whose distributions have been assigned on the basis of expert judgment (OECD 1989; 1993) and are listed in Table 2 together with a set of parameters which are assumed constant.

In Saltelli and Tarantola (2002), variance-based sensitivity analysis techniques were applied on the same model with the same factor distributions and the effect of dependence among factors were analysed. The results obtained through density-based sensitivity measures estimated via the proposed sampling

⁵The computer code is available from
<http://sensitivity-analysis.jrc.ec.europa.eu/software/index.htm>

plans are presented in the following paragraphs.

Notation	Definition	Distribution	Range	Units
T	containment time	uniform	[100, 1000]	y
k_I	leach rate for iodine	log-uniform	[10^{-3} , 10^{-2}]	mol/y
k_C	leach rate for Np chain nuclides	log-uniform	[10^{-6} , 10^{-5}]	mol/y
$V^{(1)}$	water velocity in geosphere's 1st layer	log-uniform	[10^{-3} , 10^{-1}]	m/y
$L^{(1)}$	length of geosphere's 1st layer	uniform	[100, 500]	m
$R_I^{(1)}$	retention factor for I (1st layer)	uniform	[1, 5]	-
$R_C^{(1)}$	factor to compute retention coefficients for Np chain nuclides (1st layer)	uniform	[3, 30]	-
$V^{(2)}$	water velocity in geosphere's 2nd layer	log-uniform	[10^{-2} , 10^{-1}]	m/y
$L^{(2)}$	length of geosphere's 2nd layer	uniform	[50, 200]	m
$R_I^{(2)}$	retention factor for I (2nd layer)	uniform	[1, 5]	-
$R_C^{(2)}$	factor to compute retention coefficients for Np chain nuclides (2nd layer)	uniform	[3, 30]	-
W	stream flow rate	log-uniform	[10^5 , 10^7]	m^3/y

Table 2: List of model inputs for LevelE.

The output of interest here is the maximum radiological dose to humans due to the joint effect of the four radionuclides. We perform uncertainty analysis first, using a Sobol' quasi random sample (Sobol' 1967) of size 2^{15} (*i.e.* 32768), which allows us to obtain the unconditional model output distribution. Model output values range across several orders of magnitude. The unconditional distribution has skewness $\gamma_1 = 7.7841$ and kurtosis $\gamma_2 = 106.8809$. We then apply a log transformation to the model output. This leads to $\gamma_1 = -0.0339$ and $\gamma_2 = 2.5011$. We process both the raw and the log-transformed data to appreciate the impact of the log-transformation. The registered effect is an improvement in the estimation accuracy for all sampling strategies when the log-transformed data are employed. We recall that the scale invariance property of δ_i allows us to use directly results for the log-transformed data [see Section 2].

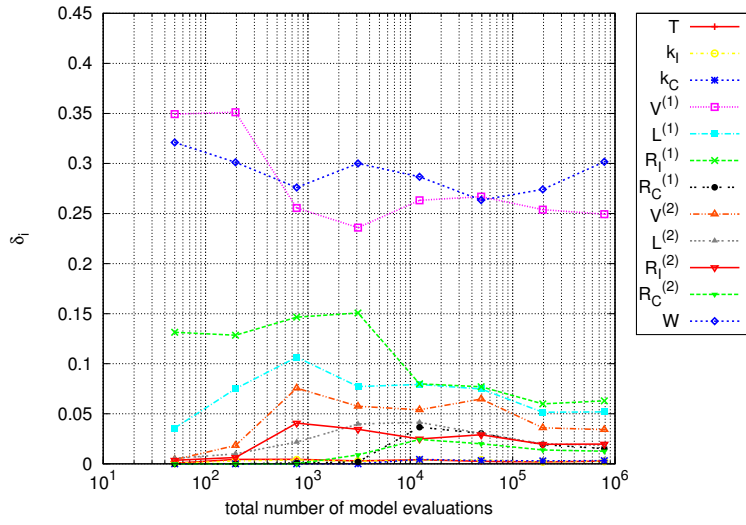


Figure 7: Moment independent sensitivity analysis for the LevelE maximum dose, convergence of δ_i estimates using substituted column sampling with Sobol's quasi-random sequences (point estimates)

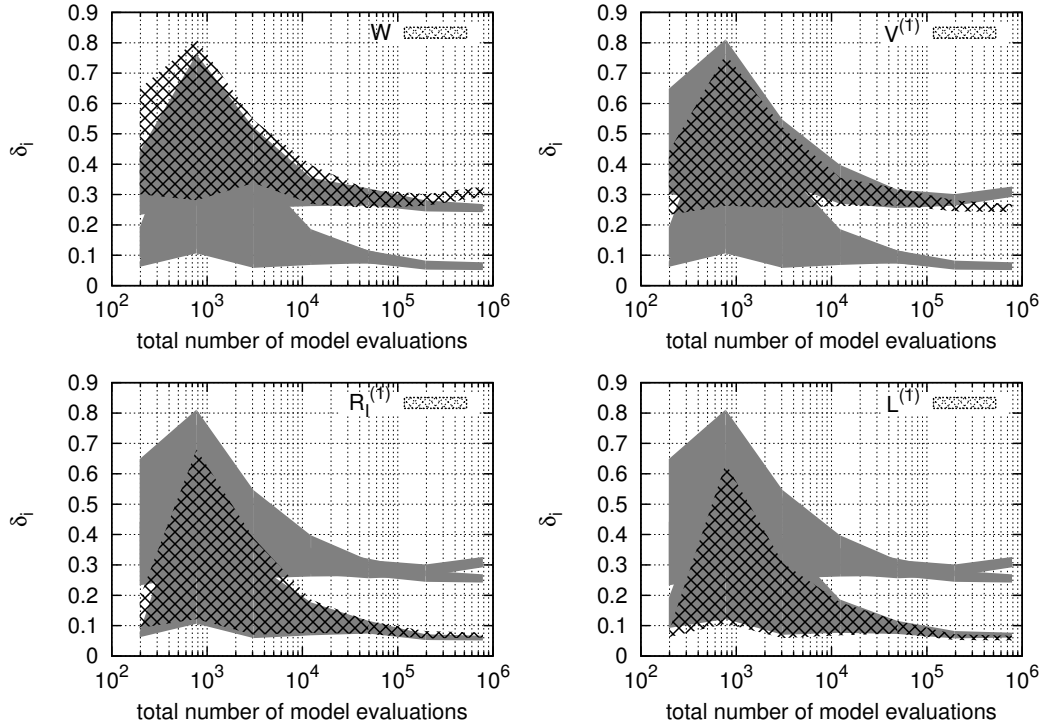


Figure 8: Moment independent sensitivity analysis for the LevelE maximum dose, convergence of δ_i estimates using substituted column sampling with Sobol's quasi-random sequences, confidence intervals obtained by bootstrapping for the 4 most important model factors (i.e. W , $V^{(1)}$, $R_I^{(1)}$ and $L^{(1)}$). The factor of interest is evidenced using a grid pattern and the others are displayed in background.

Figure 7 displays convergence results obtained using the substituted columns with Sobol' quasi-Monte Carlo for the 12 uncertain model factors. On the horizontal axis is the total number of model evaluations (M). M ranges from 50 to almost 8×10^5 . On the vertical axes, the point estimates of δ_i are displayed as M varies. Overall, Figure 7 shows that W and $V^{(1)}$ are identified as most important factors, followed by $R_I^{(1)}$ and $L^{(1)}$, which have a moderate importance, while the remaining factors have a minor influence. The analysis of Figure 7 reveals that the most important model factors (i.e. W and $V^{(1)}$) are identified

by point estimates even at very low values of M . However, as M increases, we observe instability of the point estimates, that lead to a ranking reversal between W and $V^{(1)}$ at $M \simeq 10^3$ model evaluations. Conversely, Figure 7 reveals a rapid convergence in the importance measures of the moderate and least important model factors. The bootstrap analysis (Figure 8) confirms these observations. In fact, there is a notable overlapping of the confidence intervals of W and $V^{(1)}$, which does not allow us to draw a definitive conclusion about their ranking until $M \simeq 10^5$. Furthermore, only starting with $M \geq 10^4$ we register a neat separation between the values of $(\delta_W, \delta_{V^{(1)}})$ and the sensitivity measures of the remaining 10 factors. The results in Figure 7 and 8 indicate that using a simple substituted column approach a large number of model runs is required to obtain convergence. This finding is a consequence of the fact that this sampling strategy is a plain brute force approach. Let compare the results Figure 7 and 8 to the results obtained using an improved-substituted column plan with quasi-Monte Carlo generation (Figure 9 and 10) .

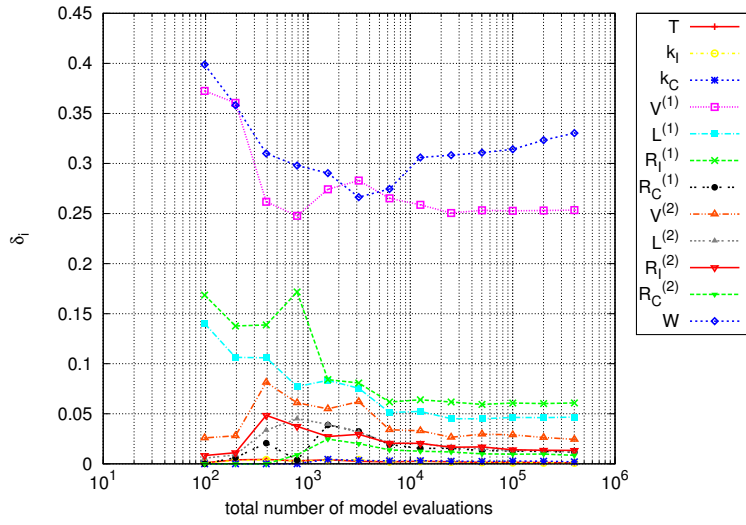


Figure 9: Moment independent sensitivity analysis for the LevelE maximum dose, convergence of δ_i estimates using improved-substituted column sampling plan (Gauss-Legendre quadrature with 4 points for the external loop) with Sobol's quasi-random sequences (point estimates)

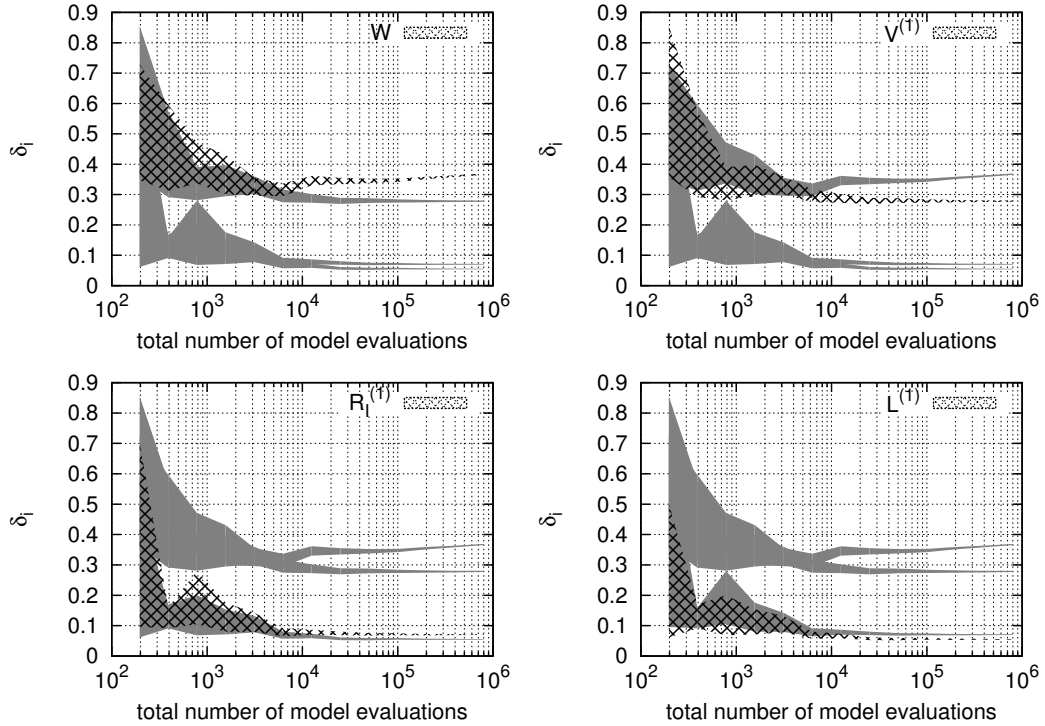


Figure 10: Moment independent sensitivity analysis for the LevelE maximum dose, convergence of δ_i estimates using improved-substituted column sampling plan (Gauss-Legendre quadrature with 4 points for the external loop) with Sobol's quasi-random sequences, confidence intervals obtained by bootstrapping for the 4 most important model factors (i.e. W , $V^{(1)}$, $R_I^{(1)}$ and $L^{(1)}$). The factor of interest is evidenced using a grid pattern and the others are displayed in background.

Figure 9 shows that, using an improved-substituted column sampling plan, the identification of the two most important factors occurs at a much lower M . Compare, for instance, the confidence intervals at $M = 10^3$ in Figure 10 against the ones at $M = 10^4$ in Figure 8. In fact, the confidence intervals of W and $V^{(1)}$ are sharply separated from the remaining factors already at $M \simeq 10^3$. That is, at a value of M at least one order of magnitude lower than with basic substituted column sampling. Furthermore, the asymptotic separation of the confidence bounds occurs around 10^4 model evaluations, *i.e.*, again for M one

order of magnitude lower than for the previous sampling plan (Figure 9). Both Figure 9 and 10 show that for the moderate and non-influential factors one obtains a very sharp distinction of the sensitivity measures by this sampling scheme.

Let us now compare these results with those obtained using permuted columns sampling schemes (Section 3.2). Results for rLHS are reported in Figure 11 and 12.

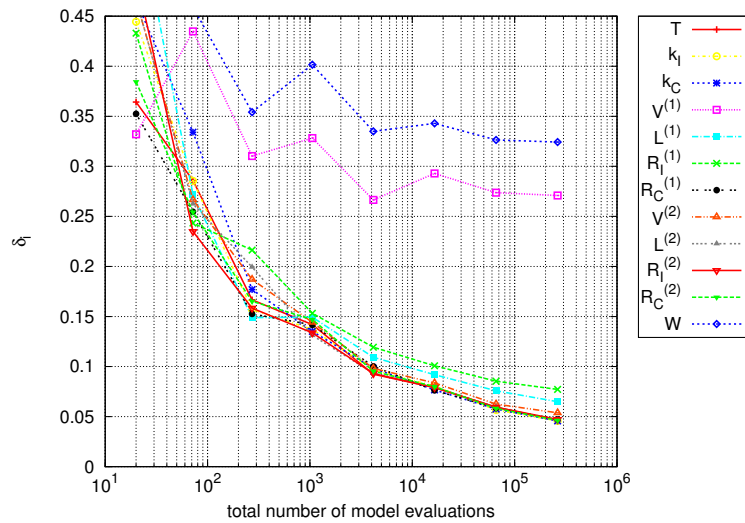


Figure 11: Moment independent sensitivity analysis for the LevelE maximum dose, convergence of δ_i estimates using a permuted column sampling plan based on replicated Latin Hypercube with $r = N$ (point estimates)

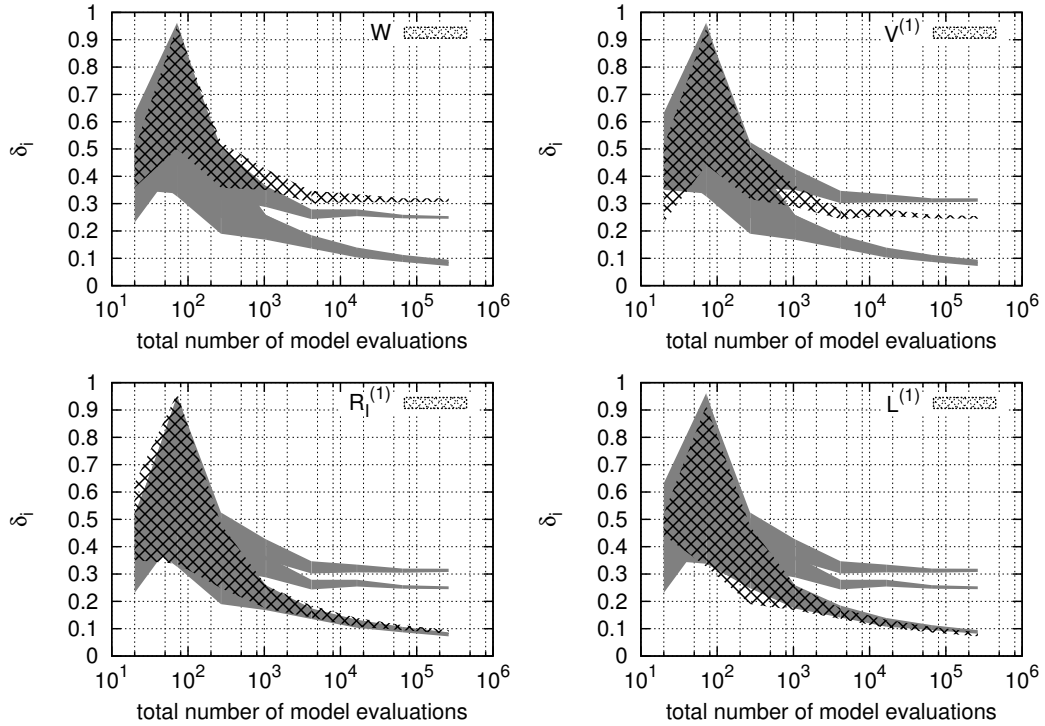


Figure 12: Moment independent sensitivity analysis for the LevelE maximum dose, convergence of δ_i estimates using a permuted column sampling plan based on replicated Latin Hypercube with $r = N$, confidence intervals obtained by bootstrapping for the 4 most important model factors (i.e. W , $V^{(1)}$, $R_I^{(1)}$ and $L^{(1)}$). The factor of interest is evidenced using a grid pattern and the others are displayed in background.

Figure 11 shows that rLHS provides a stable ranking of W and $V^{(1)}$ already at few hundreds model runs and ranking reversal never occurs. However, the sensitivity estimates for the factors featuring moderate ($R_I^{(1)}$, $L^{(1)}$ and $V^{(2)}$) to low importance are overestimated at low sample size, with the bias that progressively reduces as the number of model evaluations increases. Another consequence of this bias is that the less important model factors cannot be robustly distinguished (Figure 12)

As we discussed in Section 3.2 and evidenced in Section 4, the biasing at

low sample sizes can be reduced by using orthogonal arrays instead of random permutations (Morris et al. 2008). Let us then analyse results obtained by the last sampling scheme proposed in this work, namely, permuted column sampling with orthogonal arrays (see Figure 13).

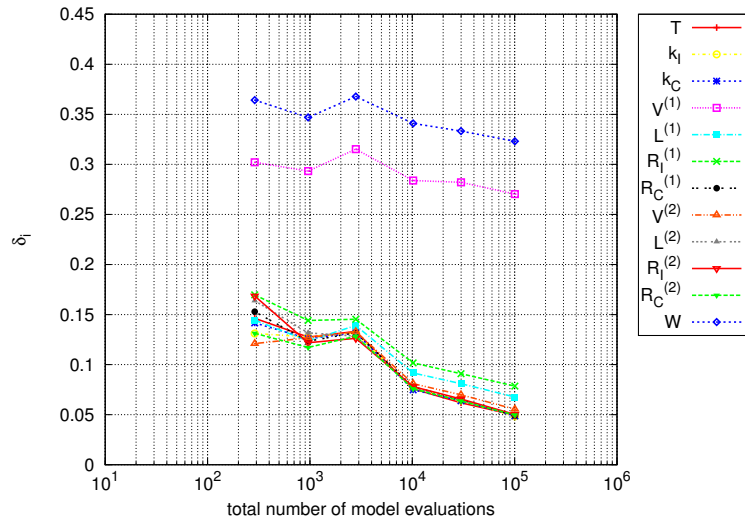


Figure 13: Moment independent sensitivity analysis for the LevelE maximum dose, convergence of δ_i estimates using a permuted column sampling plan with orthogonal arrays (point estimates)

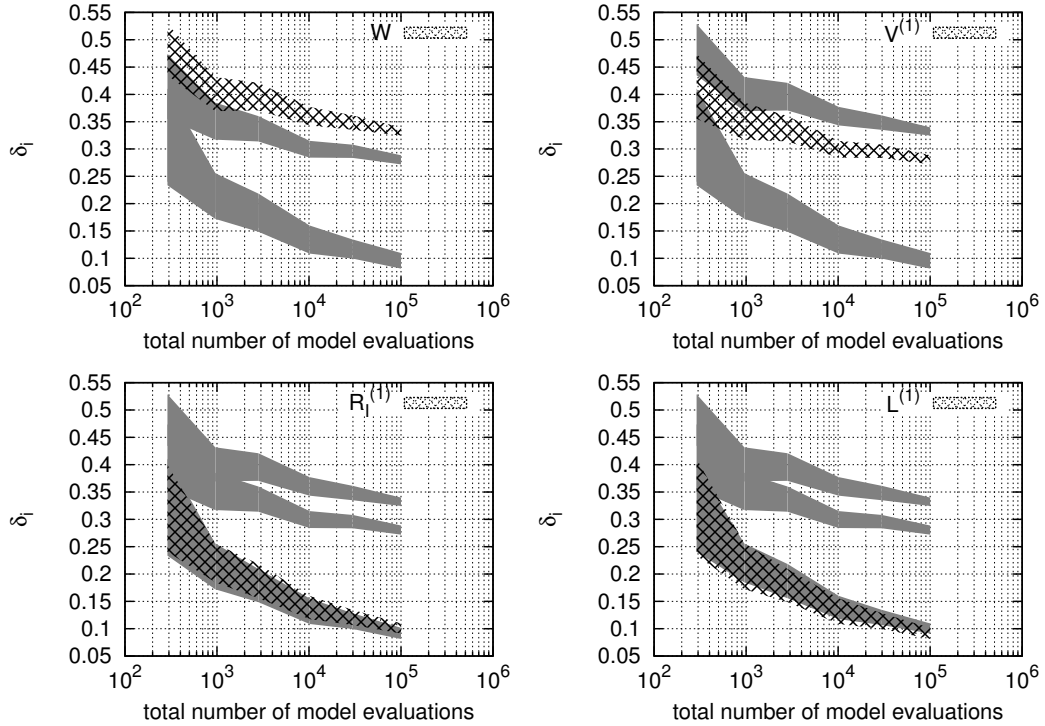


Figure 14: Moment independent sensitivity analysis for the LevelE maximum dose, convergence of δ_i estimates using a permuted column sampling plan with orthogonal arrays, confidence intervals obtained by bootstrapping for the 4 most important model factors (i.e. W , $V^{(1)}$, $R_I^{(1)}$ and $L^{(1)}$). The factor of interest is evidenced using a grid pattern and the others are displayed in background.

Figure 13 confirms the efficacy of orthogonal arrays in reducing the bias at low sample sizes. For instance, by comparing results for $10^2 \leq M \leq 10^3$ in Figures 12 and 14, one observes the sharper confidence bounds obtained by orthogonal arrays. Also, Figure 14 shows that W and $V^{(1)}$ are neatly identified as the most important factors at all sample sizes. However, the precision of the estimates for the less important model factors is not improved.

Let us now offer an overall comparison of results obtained with a substituted column sampling scheme to a permuted column scheme (see Figures 7 to 14).

We note that:

- both the improved-substituted column and the permuted column designs represent a definitive improvement over the basic substituted column plan;
- the improved-substituted columns sampling plan is more prone to Type I errors than permuted column sampling plan, but more robust to type II error.

This complementarity suggests the following way for identifying the minimum M at which the ranking obtained can be confidently entrusted. First, one utilizes permuted column sampling and considers the bootstrap confidence intervals as M increases. Then, one stops at the value M_1 , at which one is confident about the most important factors. Then, one can decide between continuing with permuted column with larger sample sizes or switch to the improved-substituted column sampling. By this sampling scheme, in fact, one obtains a neater assessment of the importance of the low relevance factors. For the present case study, the application of this strategy allows one to obtain confidence in the ranking of the model factors using density-based sensitivity measure at a number of model runs of around $M = 8000$. [We note that to obtain such confidence using a brute force approach one would have to set M at around 1 million model runs.]

Finally, the insights obtained are as follows. W and $V^{(1)}$ are the most influential variables. Variables of moderate influence are: $L^{(1)}$, $R_I^{(1)}$, $V^{(2)}$, $L^{(2)}$, and $R_I^{(2)}$. The bootstrap confirms that the variables T , k_I , k_C , $R_C^{(1)}$ and $R_C^{(2)}$ have negligible effect on the distribution of the maximum radiological dose.

Because this assessment is performed in a moment-independent fashion, this information is then providing us indications on what factors to focus efforts in data collection and further modelling for most effectively managing uncertainty in the model predictions.

6. Conclusions

In this work, we have presented sampling plans for density-based sensitivity measures utilized in the context of global sensitivity analysis of model output. This work adds to the contribution of Saltelli and Tarantola (2002) in which a variance decomposition approach is used to identify key-uncertainty drivers and to Morris et al. (2008), where orthogonal arrays are used for improving the estimation of variance-based sensitivity measures, and to Borgonovo et al. (2011a), where density-based sensitivity measures are estimated via an emulator. The estimation of density-based importance measures is, in fact, a challenging task, especially in the presence of numerically intensive models. Our work proposes novel sampling plans that reduce the computational cost while improving accuracy of the estimation process. Building on a quasi-random generated sample, we test column substitution designs (basic and improved via Gauss-Legendre quadrature formulas) and column permutation designs (based on random permutations and orthogonal arrays) on analytical tests to check convergence of sensitivity estimates. We also test these designs on a model-based dataset for the performance assessment of nuclear waste disposal sites.

Concerning sampling plans performance, results show that the improved substituted column and the permuted column plans achieve a notable reduction in computational burden. Also, using orthogonal arrays instead of random permutations improves the estimation efficacy of permuted column sampling plans. Convergence analysis performed using the root mean square error for analytical test cases and bootstrapping for LevelE, shows that column substitution sampling plans are more robust to Type II errors, while column permutation sampling plans are more robust to Type I errors.

Concerning environmental insights, results show that analysts are equipped with effective tools for determining key-uncertainty drivers, thus identifying areas

where further modelling and data collection are needed for reducing variability in model predictions caused by parametric uncertainty.

Acknowledgements

The authors wish to thank the anonymous referees for the very constructive comments. The French National Research Agency (ANR, FLASH project) and the thematic network for advanced research in Sciences and Techniques of Aeronautics and Space (RTRA STAE, CYMENT project) are gratefully acknowledged by W. Castaings for funding this research. Financial support from the ELEUSI research center is gratefully acknowledged by E. Borgonovo.

References

- B. Auder and B. Iooss. Global sensitivity analysis based on entropy. In *ESREL 2008 Conference, ESRA, Valencia, Spain, September 2008.*, 2008.
- M. J. Bayarri, Berger J., and D.M. Steinberg. Special issue on computer modeling. *Technometrics*, 51(4):353–353, 2009.
- T. Bedford. Sensitivity indices for (tree)- dependent variables. In *Proceedings of the Second International Symposium on Sensitivity Analysis of Model Output, Venice (Italy)*, pages 17–20, 1998.
- E. Borgonovo. Measuring uncertainty importance: Investigation and comparison of alternative approaches. *Risk Analysis*, 26(5):1349–1361, 2006.
- E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784, 2007.
- E. Borgonovo and S. Tarantola. Moment independent and variance-based sensitivity analysis with correlations: An application to the stability of a chemical

- reactor. *International Journal of Chemical Kinetics*, 40(11):687–698, 2008. ISSN 1097-4601.
- E. Borgonovo, W. Castaings, and S. Tarantola. Model emulation and moment independent sensitivity analysis: An application to environmental modelling. *Environmental Modelling & Software*, (in press), 2011a.
- E. Borgonovo, W. Castaings, and S. Tarantola. Moment independent importance measures: New results and analytical test cases. *Risk Analysis*, 31: 404–428, 2011b.
- F. Campolongo and R. Braddock. Sensitivity analysis of the image greenhouse model. *Environmental Modelling and Software*, 14(4):275–282, 1999.
- F. Campolongo, J. Cariboni, and A. Saltelli. An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22 (10):1509–1518, October 2007. doi: doi:10.1016/j.envsoft.2006.10.004.
- M-H. Chun, S-J. Han, and N.-I.L. Tak. An uncertainty importance measure using a distance metric for the change in a cumulative distribution function. *Reliability Engineering & System Safety*, 70(3):313–321, 2000.
- R. Confalonieri, G. Bellocchi, S. Tarantola, M. Acutis, M. Donatelli, and G. Genovese. Sensitivity analysis of the rice model warm in europe: Exploring the effects of different locations, climates and methods of analysis on model sensitivity to crop parameters. *Environmental Modelling and Software*, 25(4):479–488, 2010.
- P.S. Craig, M. Goldstein, J.C. Rougle, and A.H. Seheult. Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association*, 96 (454):717–729, 2001.

- R.I. Cukier, H.B. Levine, and K.E. Shuler. Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, 26:1–42, 1978.
- P.J. Davis and P. Rabinowitz. *Methods of Numerical Integration*. Academic Press, 1984. ISBN-0122063600.
- A.C. Davison, D.V. Hinkley, and G.A. Young. Recent developments in bootstrap methodology. *Statistical Science*, 18(2):141–157, 2003. ISSN 0883-4237.
- D. Drignei and M.D. Morris. Empirical bayesian analysis for computer experiments involving finite-difference codes. *Journal of the American Statistical Association*, 101 (476):1527–1536, 2006.
- B. Efron. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7(1):126, 1979.
- B. Efron and C. Stein. The jackknife estimate of variance. *Annals of Statistics*, 9(3):586–596, 1981.
- European Commission. Impact assessment guidelines. *SEC(2009)*, 2009.
- J.H. Friedman. Multivariate adaptive regression splines (with discussion). *Annals of Statistics*, 19(1), 1991.
- J.H. Friedman and W. Stuetzle. Projection pursuit regression. *Journal of the American Statistical Association*, 376(76):817–23, 1981.
- A.E. Gelfand and A.F.M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 485 (410): 398–409, 1990.
- N. Glick. Measurements of separation among probability densities or random variables. *Canadian Journal of Statistics*, 3(2):267–276, 1975.

- S. He, G.R. Carmichael, A. Sandu, B. Hotchkiss, and V. Damian-Iordache. Application of adifor for air pollution model sensitivity studies. *Environmental Modelling and Software*, 15(6-7 SPEC. ISS):549–557, 2000.
- A.S. Hedayat, N.J.A. Sloane, and J. Stufken. *Orthogonal Arrays: Theory and Applications*. Springer Series in Statistics, Springer Verlag, New York, 1999.
- J.C. Helton. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety*, 42(2-3):327–367, 1993.
- J.C. Helton, J.D. Johnson, C.J. Sallaberry, and C.B. Storlie. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering & System Safety*, 91(10-11):1175–1209, 2006.
- W. Hoeffding. A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325, 1948. ISSN 00034851.
- T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, 52:1–17, 1996.
- C.-F. Huang and R.H. Litzenberger. *Foundations for Financial Economics*. Prentice Hall, Upper Saddle River, NJ, USA, 1998. 365 pages, ISBN 0-13-500653-8.
- R.L. Iman and S.C. Hora. A robust measure of uncertainty importance for use in fault tree system analysis. *Risk Analysis*, 10(3):401–406, 1990. ISSN 1539-6924.
- T. Ishigami and T. Homma. An importance quantification technique in uncertainty analysis for computer models. In *Proceedings of the ISUMA90, First International Symposium on Uncertainty Modelling and Analysis*, pages 398–403, 1990.

- M.J.W. Jansen, W.A.H. Rossing, and R.A. Daamen. *Predictability and Non-linear Modelling in Natural Sciences and Economics*, chapter Monte Carlo estimation of uncertainty contributions from several independent multivariate sources, page 334-343. Kluwer Academic Publishers, Dordrecht, 1994.
- B. Krykacz-Hausmann. Epistemic sensitivity analysis based on the concept of entropy. In *Proceedings of SAMO2001, Madrid*, pages 31–35, 2001.
- Q. Liu and T. Homma. A new computational method of a moment-independent uncertainty importance measure. *Reliability Engineering & System Safety*, 94(7):1205–1211, 2009. ISSN 0951-8320.
- T.A. Mara and S. Tarantola. Variance-based sensitivity indices for models with dependent inputs. In *proceedings of the 6th international Conference on Sensitivity Analysis of Model Output (SAMO 2010)*, 2011.
- A. Marrel, B. Iooss, B. Laurent, and O. Roustant. Calculations of Sobol indices for the Gaussian process metamodel. *Reliability Engineering and System Safety*, 94:742–751, March 2009. doi: 10.1016/j.ress.2008.07.008.
- M.D. McKay. Evaluating prediction uncertainty. Technical Report NUREG-CR-6311, LA-12915-MS, Los Alamos Laboratories, 1995.
- M.D. McKay, R.J. Beckman, and W.J. Conover. 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, 1979.
- M.D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33:161–174, 1991.
- M.D. Morris, L.M. Moore, and M.D. McKay. Sampling plans based on balanced incomplete block designs for evaluating the importance of computer model inputs. *Journal of Statistical Planning and Inference*, 136(9):3203–3220, 2006.

- M.D. Morris, L.M. Moore, and M.D. McKay. Using orthogonal arrays in the sensitivity analysis of computer models. *Technometrics*, 50(2):205–215, 2008.
- L.T.H. Newham, J.P. Norton, I.P. Prosser, B.F.W. Croke, and A.J. Jakeman. Sensitivity analysis for assessing the behaviour of a landscape-based sediment source and transport model. *Environmental Modelling and Software*, 18(8-9): 741–751, 2003.
- J.P. Norton. Algebraic sensitivity analysis of environmental models. *Environmental Modelling and Software*, 23(8):963–972, 2008.
- J. Nossent, P. Elsen, and W. Bauwens. Sobol’ sensitivity analysis of a complex environmental model. *Environmental Modelling & Software*, 26(12):1515–1525, 2011.
- J.E. Oakley. Decision-theoretic sensitivity analysis for complex computer models. *Technometrics*, 51(2):121–129, 2009. ISSN 0040-1706.
- J.E. Oakley and A. O’Hagan. Probabilistic sensitivity analysis of complex models: a Bayesian approach. *J. Royal Stat. Soc. B*, 66:751–769, 2004.
- OECD. OECD/NEA PSAC User group, PSACOIN Level E Intercomparison. An International Code Intercomparison Exercise on a Hypothetical Safety Assessment Case Study for Radioactive Waste Disposal Systems. Technical report, OECD - NEA publication, 1989.
- OECD. OECD/NEA PSAG User group, PSACOIN Level S Intercomparison. An International Code Intercomparison Exercise on a Hypothetical Safety Assessment Case Study for Radioactive Waste Disposal Systems. Technical report, OECD - NEA publication, 1993.
- A.B. Owen. The dimension distribution and quadrature test functions. *Statistica Sinica*, 13(1):1–17, 2003.

- F. Pappenberger, I. Iorgulescu, and K.J. Beven. Sensitivity analysis based on regional splits and regression trees (SARS-RT). *Environmental Modelling and Software*, 21(7):976–990, 2006.
- C.K. Park and K. Ahn. A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment. *Reliability Engineering & System Safety*, 46(3):253–261, 1994.
- E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.
- H. Rabitz. Global sensitivity analysis for systems with independent and/or correlated inputs. *Procedia - Social and Behavioral Sciences*, 2(6):7587–7589, 2010. ISSN 1877-0428. doi: DOI: 10.1016/j.sbspro.2010.05.131. Sixth International Conference on Sensitivity Analysis of Model Output.
- H. Rabitz and A. Alis. General foundations of high-dimensional model representations. *Journal of Mathematical Chemistry*, 25:197–233, 1999. ISSN 0259-9791. 10.1023/A:1019188517934.
- M. Ratto and A. Pagano. Recursive algorithms for efficient identification of smoothing spline anova models. *Advances in Statistical Analysis*, 94:367–388, 2010.
- M. Ratto, A. Pagano, and P.C. Young. State dependent parameter metamodelling and sensitivity analysis. *Computer Physics Communications*, 177(11): 863–876, 2007.
- M. Ratto, A. Pagano, and P.C. Young. Non-parametric estimation of conditional moments for sensitivity analysis. *Reliability Engineering & System Safety*, 94(2):237–243, February 2009. ISSN 0951-8320.

- J.K. Ravalico, G.C. Dandy, and H.R. Maier. Management option rank equivalence (more) - a new method of sensitivity analysis for decision-making. *Environmental Modelling and Software*, 25(2):171–181, 2010.
- J. Sacks, S.B. Schiller, and W. J. Welch. Designs for computer experiments. *Technometrics*, 31(1):4147, 1989a. ISSN 0040-1706.
- J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn. Design and analysis of computer experiments. *Statistical science*, 4(4):409–423, 1989b. ISSN 0883-4237.
- A. Saltelli. Making best use of model valuations to compute sensitivity indices. *Computer Physics Communications*, 145:280–297, 2002.
- A. Saltelli. Editorial. *Reliability Engineering and System Safety*, 94(7):1133–1134, 2009.
- A. Saltelli and P. Annoni. How to avoid a perfunctory sensitivity analysis. *Environmental Modelling and Software*, 25(12):1508–1517, 2010.
- A. Saltelli and J. Marivoet. Non-parametric statistics in sensitivity analysis for model output: a comparison of selected techniques. *Reliability Engineering & System Safety*, 28(2):229–253, 1990.
- A. Saltelli and I.M. Sobol'. About the use of rank transformation in sensitivity analysis of model output. *Reliability Engineering & System Safety*, 50(3):225–239, 1995. ISSN 0951-8320.
- A. Saltelli and S. Tarantola. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. *Journal of the American statistical association*, 97(459):702–709, 2002.

- A. Saltelli, S. Tarantola, and K. P. . Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1): 39–56, 1999. URL www.scopus.com. Cited By (since 1996): 292.
- A. Saltelli, S. Tarantola, and F. Campolongo. Sensitivity analysis as an ingredient of modelling. *Statistical Science*, 19 (4):377–395, 2000.
- A Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley and Sons, 2004.
- A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2): 259–270, February 2010. ISSN 0010-4655.
- T. Santner, B. Williams, and W. Notz. *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. Springer Verlag, New York, 2003.
- B. W. Silverman. *Density estimation for statistics and data analysis / B.W. Silverman*. Chapman and Hall, London ; New York :, 1986. ISBN 0412246201 0412246201.
- I.M. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7:86–112, 1967.
- I.M. Sobol'. Multidimensional quadrature formulas and haar functions. Nauka, Moscow, Russia, 1969. (in Russian).
- I.M. Sobol'. Uniformly distributed sequences with additional uniformity properties. *USSR Comput. Math. Math. Phys.*, 16(5):236–242, 1976.

- I.M. Sobol'. Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, 1:407–414, 1993. English translation of Russian original paper.
- I.M. Sobol'. Theorems and examples on high dimensional model representation. *Reliability Engineering & System Safety*, 79(2):187–193, 2003. ISSN 0951-8320.
- I.M. Sobol and S. Kucherenko. Derivative based global sensitivity measures and their links with global sensitivity indices. *Mathematics and Computers in Simulation*, 79:30093017, 2009.
- E. Soofi. Capturing the intangible concept of information. *Journal of the American Statistical Association*, 89(428):1243–1254, 1994.
- E. Stokstad. Dueling visions for a hungry world. *Science*, 319:1474–1476, 2008.
- C.B. Storlie, L.P. Swiler, J.C. Helton, and C.J. Sallaberry. Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering & System Safety*, 94(11):1735–1763, 2009. ISSN 0951-8320.
- B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979, July 2008. ISSN 0951-8320.
- A. Takemura. Tensor analysis of anova decomposition. *Journal of the American Statistical Association*, 78(384):pp. 894–900, 1983. ISSN 01621459.
- US EPA. Guidance on the development, evaluation, and application of environmental models. Technical Report EPA/100/K-09/003, US Environmental Protection Agency, 2009. URL www.epa.gov/crem.

W.J. Welch, J. Sacks, H.P. Wynn, T.J. Mitchell, and M.D. Morris. Screening, predicting, and computer experiments. *Technometrics*, 34(1):15–25, 1992. ISSN 0040-1706.

White House. Proposed risk assessment bulletin. *The White House and Office of Management and Budget*, January, 2006.

T. Ziehn and A.S. Tomlin. GUI-HDMR - a software tool for global sensitivity analysis of complex models. *Environmental Modelling and Software*, 24(7): 775–785, 2009.