EUROPEAN COMMISSION

Community research

PAMINA
Performance Assessment Methodologies in Application to Guide the Development of the Safety Case

(Contract Number: FP6-036404)

SENSITIVITY ANALYSIS BENCHMARK BASED ON THE USE OF ANALYTIC AND SYNTHETIC PA CASES MILESTONE (N°: M2.1.D.11)

Author(s): **Elmar Plischke (TU Clausthal, Germany),**
Klaus-Jürgen Röhlig (TU Clausthal, Germany),
Anca Badea (JRC Petten, Netherlands),
Ricardo Bolado Lavín (JRC Petten, Netherlands),
Per-Anders Ekström (Facilia SA, Sweden),
Stephan Hotzel (GRS Cologne, Germany)

Date of issue of this report : **08/06/2009**

Start date of project : **01/10/2006** Duration : 36 Months

Project co-funded by the European Commission under the Euratom Research and Training Programme on Nuclear Energy within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
RE	Restricted to a group specified by the partners of the [PAMINA] project	
CO	Confidential, only for partners of the [PAMINA] project	

Foreword

The work presented in this report was developed within the Integrated Project PAMINA: **P**erformance **A**ssessment **M**ethodologies **IN A**pplication to Guide the Development of the Safety Case. This project is part of the Sixth Framework Programme of the European Commission. It brings together 25 organisations from ten European countries and one EC Joint Research Centre in order to improve and harmonise methodologies and tools for demonstrating the safety of deep geological disposal of long-lived radioactive waste for different waste types, repository designs and geological environments. The results will be of interest to national waste management organisations, regulators and lay stakeholders.

The work is organised in four Research and Technology Development Components (RTDCs) and one additional component dealing with knowledge management and dissemination of knowledge:

- In RTDC 1 the aim is to evaluate the state of the art of methodologies and approaches needed for assessing the safety of deep geological disposal, on the basis of comprehensive review of international practice. This work includes the identification of any deficiencies in methods and tools.

- In RTDC 2 the aim is to establish a framework and methodology for the treatment of uncertainty during PA and safety case development. Guidance on, and examples of, good practice will be provided on the communication and treatment of different types of uncertainty, spatial variability, the development of probabilistic safety assessment tools, and techniques for sensitivity and uncertainty analysis.

- In RTDC 3 the aim is to develop methodologies and tools for integrated PA for various geological disposal concepts. This work includes the development of PA scenarios, of the PA approach to gas migration processes, of the PA approach to radionuclide source term modelling, and of safety and performance indicators.

- In RTDC 4 the aim is to conduct several benchmark exercises on specific processes, in which quantitative comparisons are made between approaches that rely on simplifying assumptions and models, and those that rely on complex models that take into account a more complete process conceptualization in space and time.

The work presented in this report was performed in the scope of RTDC 2.

All PAMINA reports can be downloaded from http://www.ip-pamina.eu.

PAMINA Milestone M2.1.D.11: Sensitivity Analysis Benchmark Based on the Use of Analytic and Synthetic PA Cases (Topic Report)

Elmar Plischke,[*] Klaus-Jürgen Röhlig,[†] Anca Badea,[‡]
Ricardo Bolado Lavín,[§] Per-Anders Ekström,[¶] Stephan Hotzel[‖]

June 8, 2009

Contents

[*] elmar.plischke@tu-clausthal.de
[†] klaus.roehlig@tu-clausthal.de
[‡] anca.costescu-badea@ec.europa.eu
[§] ricardo.bolado-lavin@jrc.nl
[¶] peranders.ekstrom@facilia.se
[‖] stephan.hotzel@grs.de

1 Introduction

The PAMINA task 2.1.D–Techniques for Sensitivity and Uncertainty Analysis–compares the relative advantages and disadvantages of different methods for applying Sensitivity Analysis (SA) to performance assessment calculations. This report is part of a benchmark study aimed at testing a wide range of Sensitivity Analysis techniques on test cases. In a previous Milestone Report [11], we issued the plan for such a benchmark study. In this Milestone Report, we report on the results on this benchmarks study by analysing nonlinear techniques for SA. In order to gain experience with the tools and techniques of SA we first apply them to analytical test models, the results of which are reported in Section 5. As a more application-oriented model, we study a geosphere transport model in Section 6.
Uncertainty Analysis (UA) and Sensitivity Analysis (SA) form an integral part of gaining knowledge of computational models used for the analysis and prediction in many parts of engineering and applied sciences. UA is handled by computing standard statistical indicators (mean, variance) based on a given sample of corresponding input data and output data. Asides from these standard indicators, Sensitivity Analysis indicators should provide hints to questions like:

- Which uncertain parameters mostly contribute to the output uncertainty?
- Are there any parameters whose uncertainties have negligible effects on the output?
- Is there a set of uncertain parameters which has a combined effect on the output variability while the individual influences are not noticeable?

In most practical cases, SA is performed using indicators based on linear regression techniques, including rank-based techniques. However, for complex models where no linear or monotonous dependency is apparent, such a SA is not very powerful. Especially, the dependency on higher order terms cannot be fully explained when using only linear regression techniques.
When the model under inspection is a linear one, variance-based Sensitivity Analysis indicators yield similar results in terms of information about parameter importance as Pearson Correlation Coefficients, Partial Correlation Coefficients or Standard Regression Coefficients. Additionally,

for nonlinear, non-monotonous models those indicators provide information not retrievable from indicators based on global linear regression techniques.
For our purposes, we distinguish two types of sensitivity indicators, named *first order effects* (also called: main effects, Sobol' indices, correlation ratios, importance measures) which can be used in a Factor Prioritisation (FP) setting by detecting direct influences from the input parameters to the output parameter, and *total effects* which can be used in a Factor Fixing (FF) setting by detecting indirect influences. In a FP setting most influential parameters are identified for which further research (for reducing their lack-of-knowledge uncertainty) improves the model significantly, in a FF setting unessential parameters are identified which may be fixed to a constant value without changing the overall model behaviour. For further discussion see [14, Section 1.2].

2 Problem formulation

We consider a computational model $y = f(x_1, \ldots, x_k)$ with k (scalar) input parameters x_i and a (scalar) output y. However, the values of the input parameters are not exactly known. We assume that this uncertainty can be handled by using random variables X_i of known distributions. Then the model output is also a random variable $Y = f(X_1, \ldots, X_k)$. To determine its properties one uses the tools of Uncertainty Analysis and Sensitivity Analysis. These tools require realisations of the input distributions and model evaluations, e.g., to compute the mean as an estimate for the expected value of Y. We will denote one realisation of a parameter set with $(x_1, \ldots, x_k)$, multiple realisations are shown in matrix notation $\mathbf{X} = (x_{ji})_{j=1,\ldots,n, i=1,\ldots,k}$.

3 Sensitivity Indices

For a short introduction to Sensitivity Analysis, see [1]. We will draw most of our attention to methods described therein in Section 5.2 (Monte Carlo based methods) which contains a discussion of regression techniques and in Section 5.3 (Variance decomposition based methods) where the Sobol' indices are introduced and methods for their estimation are presented.
For convenience, some details for variance decomposition are mentioned here (cf. [19]). The variance of Y can be expressed in terms of the conditional variance

$$\mathrm{V}[Y] = \mathrm{E}[\mathrm{V}[Y|X_I]] + \mathrm{V}[\mathrm{E}[Y|X_I]] \tag{1}$$

where $X_I = (X_i)_{i \in I}$ is a random vector, and $I \subset \{1, \ldots, k\}$ is an index set of "interesting" factors. $\mathrm{E}[Y|X_I]$ is the conditional expectation of Y given X_I and $\mathrm{V}[Y|X_I] = \mathrm{E}[(Y - \mathrm{E}[Y|X_I])^2|X_I]$ is the conditional variance, respectively. These two are random variables of X_I.
The *Sensitivity Index (SI)* with respect to the index set I is then given by

$$S_I = \frac{\mathrm{V}[\mathrm{E}[Y|X_I]]}{\mathrm{V}[Y]} \qquad \text{or} \qquad S_I = 1 - \frac{\mathrm{E}[\mathrm{V}[Y|X_I]]}{\mathrm{V}[Y]}.$$

If the index set contains just one element $i \in I$, the SI is called *first order effect* or main effect of i. If I contains all but one index $i \notin I$ then $S_{Ti} = 1 - S_I$ is the *total effect* of i. Analogously, the total

effect of an index set I is defined as $S_{TI} = 1 - S_{I^C}$ where $I^C = \{1,\dots,k\} \setminus I$ is the complement set of I. The name "total effect" should not distract from the fact that in case of dependent random vectors the total effect does not include the part of the variance of Y which escapes through the dependency of X_{I^C} on X_I.
The Sensitivity Index has the following properties:

- $S_I \in [0,1]$.
- If $S_I = 1$ then Y is a function of X_I.
- If X_I and Y are independent then $S_I = 0$.

For the first order and total effects of single factors i the following results hold true:

- If $X_1,\dots,X_k$ are independent then $\sum_{i=1}^k S_i \leq 1$ and $\sum_{i=1}^k S_{Ti} \geq 1$.
- If $f(x_1,\dots,x_k)$ is an additive function in all of the parameters x_i (in particular, if f is linear) then $S_i = S_{Ti}$ and $\sum_{i=1}^k S_i = 1$.

For the derivation of some of the methods presented below, the variance of the conditional expectation of Y satisfies the following limit,

$$\mathrm{V}[\mathrm{E}[Y|X_I]] = \lim_{\mathrm{E}[\varphi(X_I)-\mathrm{E}[Y|X_I]]\to 0} \mathrm{E}\left[(\mathrm{E}Y - \varphi(X_I))^2\right] \tag{2}$$

for a sequence of (square-integrable) functions $\varphi : \mathbb{R}^\ell \to \mathbb{R}$ of X_I where $\ell = \operatorname{card} I \leq k$ denotes the number of factors in I (which is the dimension of the random vector X_I). Setting $\varphi(x) = \mathrm{E}[Y|X_I = x]$ yields equality in (2). This special choice of the function φ is a non-parametric regression curve. With (2) in mind the first order effect S_i is the fraction of the variance of Y that is explained by a functional dependency on X_i alone, while the total effect S_{Ti} is the fraction of the variance of Y that is *not* explained by a functional dependency on all parameters but X_i.

4 Overview of the Tested Algorithms

Some of the methods which we present in the next subsections have not caught much attention in the current literature. These are "cheap methods" in the sense that the Sensitivity Indices can be estimated from a given sample of realisations of the input variables and its associated set of model outputs which were, e.g, already used for Uncertainty Analysis. Hence these methods are of special value for practitioners when keeping efficiency in mind.
The formula (2) implies that an estimator for the Sensitivity Index with respect to the index set I is given by

$$\hat{S}_I = \frac{\sum_{j=1}^n (\hat{y}(x_{Ij}) - \bar{y})^2}{\sum_{j=1}^n (y_j - \bar{y})^2}, \qquad \bar{y} = \frac{1}{n}\sum_{j=1}^n y_j, \tag{3}$$

where $\hat{y}$ is a model prediction based on the input data from the parameter group I with $\mathrm{E}[(\hat{y} - \mathrm{E}[Y|X_I])^2]$ small. Clearly, one cannot consider all possible functions for $\hat{y}$. Instead, we use different model classes which are "rich" enough to provide a good estimate of $\mathrm{E}[Y|X_I]$. If $\hat{y}$ is constructed from a *linear* regression model then we have already noted that the result of (3) (which is, in this case, the standard regression coefficient) is not necessarily good in case of non-linear, non-monotonic output data. We may therefore try a class of locally constant models (Correlation Ratios), models performing higher-order polynomial model fitting or models using locally linear regression techniques. Other regression techniques, like kernel density estimators, have not been used in this benchmark.
A different approach to computation of SA requires special sampling schemes. We will briefly discuss the used methods based upon repetitive model evaluations (Sobol') or upon a frequency response setting (FAST). The use of other sampling schemes like winding stairs sampling or alternative orthogonal transformations (Walsh-Hadamard) was not within the scope of the benchmark.

4.1 Correlation Ratios: Graphical method extended

One straight-forward method of estimating $\mathrm{V}[\mathrm{E}[Y|X_i]]$ is to consider $\mathrm{E}[Y|X_i \in \mathcal{I}_m]$ where $\{\mathcal{I}_m, m = 1,\ldots,\ell\}$ is a partition of the whole input parameter range for X_i into ℓ subsamples. Then from an estimate of the conditional variance for $y_j = f(x_{j1}, x_{j2}, \ldots, x_{ji}, \ldots, x_{jk})$, $j = 1,\ldots,n$ we obtain

$$\bar{y}_m = \frac{1}{n_m} \sum_{x_{ji} \in \mathcal{I}_m} y_j, \qquad n_m = \mathrm{card}\{x_{ji} \in \mathcal{I}_m\},$$

$$\hat{S}_i = \frac{\sum_{m=1}^{\ell} n_m (\bar{y}_m - \bar{y})^2}{\sum_{j=1}^{n} (y_j - \bar{y})^2} \qquad \text{(CR-VCE)}$$

where $\mathrm{card}\,\mathcal{A}$ denotes the number of elements in the set $\mathcal{A}$. Figure 1 demonstrates this process of taking local means.
Alternatively, based on decomposition (1) one can also compute the mean of the conditional variances instead of estimating the variance of the conditional means,

$$s_m^2 = \frac{1}{n_m - 1} \sum_{x_{ji} \in \mathcal{I}_m} (y_j - \bar{y}_m)^2,$$

$$\hat{S}_i = 1 - \frac{(n-1) \sum_{m=1}^{\ell} s_m^2}{(n-\ell) \sum_{j=1}^{n} (y_j - \bar{y})^2}. \qquad \text{(CR-ECV)}$$

For the number of subsamples in a partition, [6] suggests $\ell = \lfloor \sqrt{n} \rfloor$. The TU Clausthal implementations use the upper integer bound, $\ell = \lceil \sqrt{n} \rceil$ (which we will call "rule of thumb" in the following text). Hence we can expect ℓ realisations in each of the ℓ subsamples. However, the tests performed by R. Bolado and A. Badea, JRC-Petten, for this benchmark exercise suggest that this choice can be sub-optimal.
Using a higher-dimensional partition, correlation ratio methods are also able to compute higher-order effects. Unfortunately, for computing the interaction between multiple factors as for total

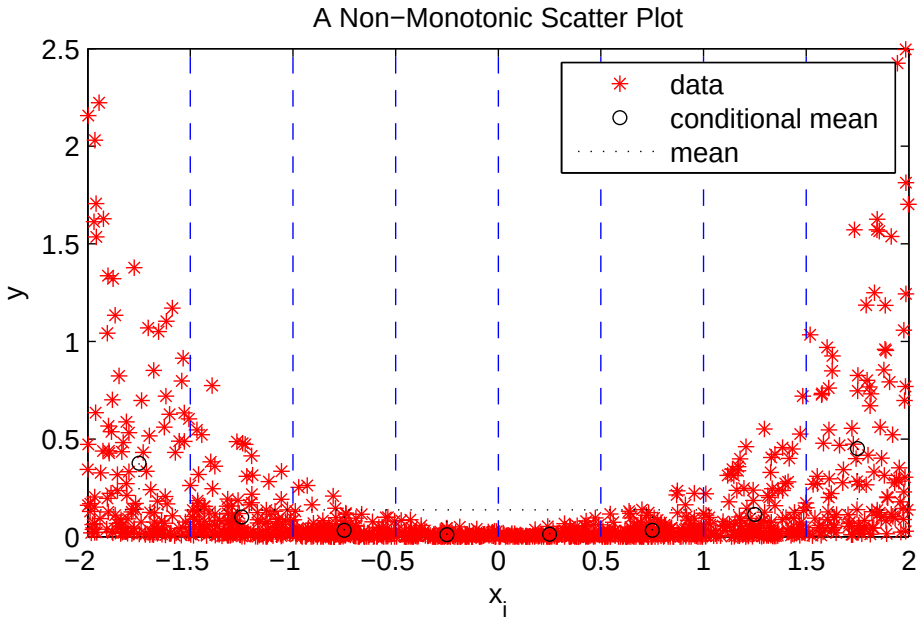

Figure 1: Correlation Ratios: Computing the variance of conditional means.

effects this method suffers from the curse of dimensionality: The number of subsamples in a partition grows with the power of the length of the index set. Moreover, in this situation it is unclear if there are enough realisations available in each subsample of a high-dimensional space. With respect to (3), in (CR-VCE) we are using the step function $\hat{y} : x \mapsto E[Y|\mathcal{I}_m]$, $x \in \mathcal{I}_m$ which can be rewritten by utilising the characteristic function of $\mathcal{I}_m$,

$$\hat{y}(x) = \sum_{m=1}^{\ell} \mathbf{1}_m(x) E[Y|\mathcal{I}_m], \qquad \mathbf{1}_m(x) = \begin{cases} 1 & \text{if } x \in \mathcal{I}_m, \\ 0 & \text{if } x \notin \mathcal{I}_m \end{cases}$$

For the estimation of S_i via Correlation Ratio this yields

$$\sum_{j=1}^{n} (\hat{y}(x_{ij}) - \bar{y})^2 = \sum_{j=1}^{n} \left(\sum_{m=1}^{\ell} \mathbf{1}_m(x_{ij}) E[Y|\mathcal{I}_m] - \bar{y} \right)^2 = \sum_{m=1}^{\ell} n_m (E[Y|\mathcal{I}_m] - \bar{y})^2.$$

In the original publication [9] the Correlation Ratio ${}_y\eta_{x_i}$ is defined as the square root of the Sensitivity Index.

4.2 Polynomial Fit: If a linear model is not enough

Another cheap approach consists of fitting a polynomial model of the given input data to the given output data (with hidden error term),

$$Y = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \ldots \beta_M X_i^M$$

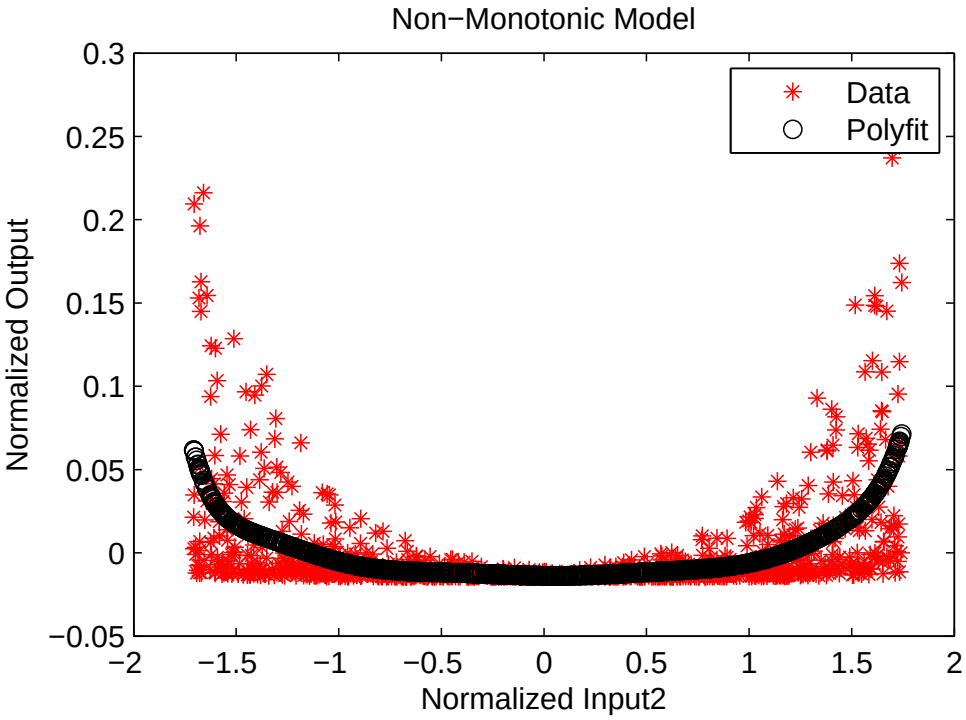

Figure 2: Polynomial Fit: Computing the variance for the model-predicted output.

Then we compute the goodness-of-fit

$$\hat{S}_i = \frac{\sum_{j=1}^{n} (\hat{y}|_{p_M(X_i)}(x_{ji}) - \bar{y})^2}{\sum_{j=1}^{n} (y_j - \bar{y})^2} \qquad \text{(FIT)}$$

where $\hat{y}|_{p_M(X_i)}(\cdot)$ is the model predicted output from a polynomial model in X_i with maximal power M. This maximal power should be chosen large enough to capture sudden changes of the output. However, one has to consider the problem of over-fitting the data. The value in (FIT) already gives an estimate for the first order Sensitivity Index. Further details can be found in [7]. In Figure 2 the polynomial fit with $M = 10$ is applied to the same set of data as in Figure 1.
With respect to (3), the used model $\hat{y}$ is a polynomial $\hat{y}|_{p_M(X_i)}(x)$. This global polynomial fit clearly is of limited use when there are discontinuities in the output.
When computing higher-order effects using polynomial regression we have to consider products of powers of the input factors. A feasible way to handle this mass of monomials is to restrict the sum of the powers of the individual factors by M (i.e, to prescribe a maximal length of the associated multi-indices). As an example, the computation of S_{T1} for a $k = 3$ parameter model with $M = 2$ uses a polynomial fit of the form

$$Y = \beta_{00} + \beta_{10}X_2 + \beta_{01}X_3 + \beta_{20}X_2^2 + \beta_{11}X_2X_3 + \beta_{02}X_3^2 = \sum_{|\alpha| \le M} \beta_\alpha \vec{X}^\alpha, \qquad \vec{X} = [X_2 X_3].$$

As the design matrices for the regression obtained by this method get very large we need a least squares algorithm that can cope with intermediate results which have close-to-singular precision.

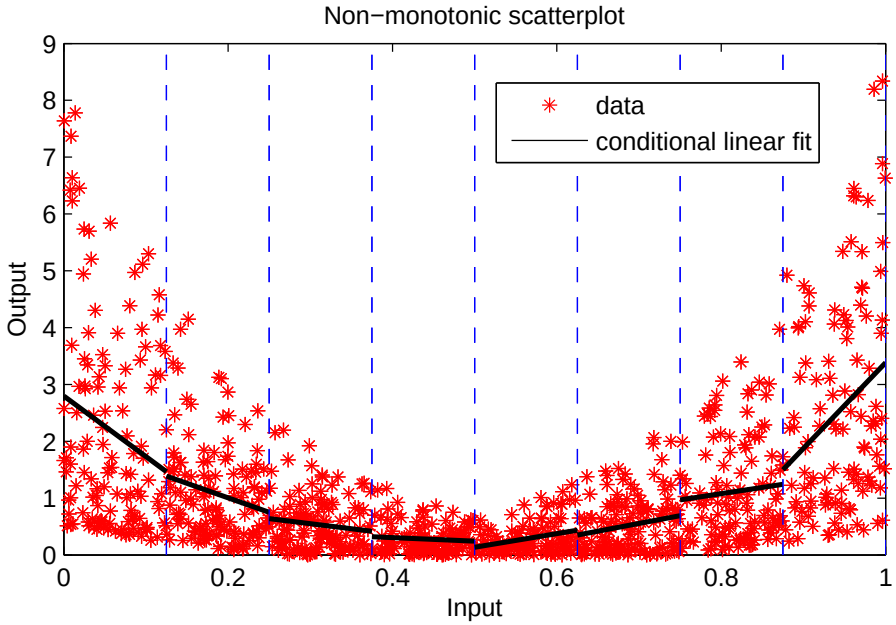

Figure 3: Conditional Linear Fit: Piecewise linear model-predicted output.

4.3 Conditional linear fit

Instead of fitting a polynomial to the data of the whole input parameter space as in Subsection 4.2, we can perform the model fitting conditioned on some suitable partition as in Subsection 4.1. For local interpolation purposes, low order polynomials should suffice. In fact, we use linear models, see Figure 3. Compared to the previous figures, a slightly different function has been used to create the data. As the local approach allows for jump discontinuities, it also helps in handling non-continuous output. However, the computational effort of higher-order effects (in particular, total effects) is of polynomial order in the length of the factor group I.
The TU Clausthal implementation of the conditional linear fit uses a fixed partition size of $\ell = 5$ for first order effects and of $\ell = 5^{k-1}$ for total effects unless otherwise noted. The subsamples are determined by partitioning the ranked data into equally sized intervals.

4.4 Methods using special input sampling

Asides from the direct computation of the Sensitivity Indices a large amount of algorithms have been developed that need special sampling schemes. Overviews of the available algorithms can be found in [13, 16, 14]. These methods offer better estimates compared to the cheap methods. However, there are some drawbacks and pitfalls which one should be aware of when using these algorithms. We will report on this in Section 5.

4.4.1 Ishigami-Homma-Saltelli/Sobol´

These methods use two input samples, the basic sample $\mathbf{X} = (x_{ji})_{j=1,\ldots,n,\, i=1,\ldots,k}$ and the alternative sample $\mathbf{X}' = (x'_{ji})_{j=1,\ldots,n,\, i=1,\ldots,k}$ with the associated output from the model evaluations $\mathbf{Y} = f(\mathbf{X}) = (f(x_{j1},\ldots,x_{jk}))_{j=1,\ldots,n}$ and $\mathbf{Y}' = f(\mathbf{X}') = (f(x'_{j1},\ldots,x'_{jk}))_{j=1,\ldots,n}$.
For each input factor i, a new sample $\mathbf{X}_i$ is created by replacing the ith column of $\mathbf{X}'$ with that of $\mathbf{X}$. The first order Sensitivity Indices are computed by determining the correlation coefficient between the model output $\mathbf{Y}_i = (f(x'_{j1},\ldots,x'_{j(i-1)},x_{ji},x'_{j(i+1)},\ldots,x'_{jk}))_{j=1,\ldots,n}$ associated with the input sample $\mathbf{X}_i$ and the basic model output $\mathbf{Y}$, in matrix notation given by

$$\hat{S}_i = \rho(\mathbf{Y}_i,\mathbf{Y}) = \frac{(\mathbf{Y}_i - \bar{\mathbf{Y}}_i)^{\mathrm{T}}(\mathbf{Y} - \bar{\mathbf{Y}})}{\sqrt{(\mathbf{Y}_i - \bar{\mathbf{Y}}_i)^{\mathrm{T}}(\mathbf{Y}_i - \bar{\mathbf{Y}}_i)}\sqrt{(\mathbf{Y} - \bar{\mathbf{Y}})^{\mathrm{T}}(\mathbf{Y} - \bar{\mathbf{Y}})}}. \tag{IHS}$$

There are many variants of this formula in use which exploit that the output variables Y, Y' and Y_i have the same expectation. To obtain total effects, we compute the correlation coefficient between the modified model output $\mathbf{Y}_i$ and the output $\mathbf{Y}'$ from the alternative sample, $\hat{S}_{Ti} = 1 - \rho(\mathbf{Y}_i,\mathbf{Y}')$. The total number of model evaluations is given by $n(k+2)$ and provides us with first order effects and total effects. If there are dependencies between the input data then the row insertion has to take marginal probabilities into account.
For the Sobol´ method, a special quasi-random sampling scheme named LP$_\tau$ is used as input sample generator which has favourite convergence properties for Monte-Carlo integration compared to simple random sampling. There are further constraints[1] on the basic sample size and on the maximal number of parameters when using LP$_\tau$.
These methods have the drawback that they may produce negative values or, for total effects, values larger than 1. These are meaningless values as fractions of variance are estimated.

4.4.2 Fourier Amplitude Sensitivity Test/Random Balance Design

For Fourier Amplitude Sensitivity Tests (FAST) the parameter realisations are chosen along a search curve with a special frequency behaviour. This introduces an artificial time-scale via the placement of the realisations in the sample. A common choice for this sample is

$$\begin{aligned} x_{ji} &= G_{\omega_i}(s_j + r_i), \qquad r_i \in [0,1], \quad s_j = {}^{j}\!/_{n}, \quad i = 1,\ldots,k, \quad j = 1,\ldots,n, \\ G_\omega :&\quad \mathbb{R} \to [0,1], \quad s \mapsto {}^{1}\!/_{\pi}\arccos(\cos(2\pi\omega s)) \end{aligned} \tag{4}$$

where r_i is a random shift parameter and $\omega_i \in \mathbb{N}$ is an integer frequency assigned to the ith input factor. The frequency selection is handled by a special algorithm [3, 17, 4] to avoid frequency interference. As a vital constraint all the frequencies including higher harmonics[2] up to a given maximum M have to stay below the Nyquist frequency, $M\sum\omega_i < \lfloor \frac{n}{2} \rfloor$. For small sets of parameters the choice $\omega_i = \omega_0^{i-1}$ also works well. In this case, the basic frequency ω_0 controls the precision of the algorithm. Moreover, this frequency selection scheme also allows for the computation of higher order effects and total effects, see below.

[1] Due to number-theoretical reasons the performance is best for power-of-two sample sizes.
[2] The maximal harmonic M is also called interference factor.

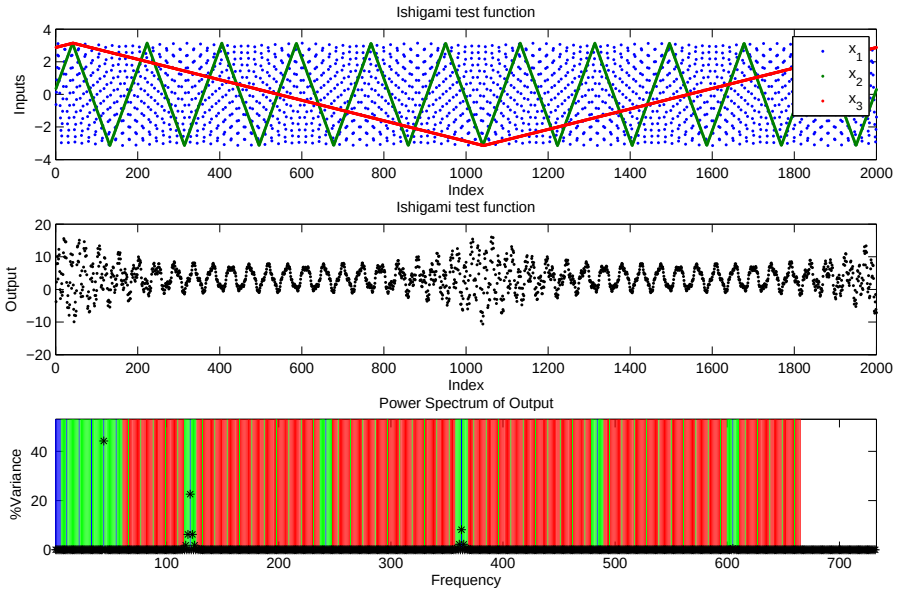

Figure 4: FAST: Prescribed frequencies in the inputs, resonances in the output.

Note that with (4) the generated sample is quasi-uniformly $[0,1]$ distributed. If other distributions are needed then the sample has to be modified by suitable transformations, e.g., via inverse cumulative distribution functions. To keep this presentation short, we only consider uniform $[0,1]$ input distributions and therefore need no parameter transformations to other distribution types.
After the frequencies are assigned to all input factors, the output is analysed for resonances using a Fast Fourier Transformation, see Figure 4 for an illustration and also Section 5.1 for further details on this particular example. If the complex discrete Fourier coefficients of $\mathbf{Y} = (y_j)_{j=1,\dots,n}$ are given by

$$c_m = \sum_{j=1}^{n} y_j \zeta_n^{(j-1)m}, \qquad \zeta_n = e^{-\frac{2\pi i}{n}}, \qquad m = 0,\dots,n-1, \tag{5}$$

then the part of the output attributed to the frequency ω_i (and hence to the input factor i) is found in the set $\{c_{m\omega_i} | m = 1,\dots,M\}$ where the maximal higher harmonic M is usually 4 or 6. The fraction of the variance attributed to the frequency ω_i (and hence to the input factor i) is given by

$$\hat{S}_i = 2\frac{\sum_{m=1}^{M} |c_{m\omega_i}|^2}{\sum_{m\neq 0} |c_m|^2}. \tag{FAST}$$

However, if the output depends non-continuously on input parameters then the quadratic convergence properties of the series in (FAST) are lost and higher values of M are required.
Analogously to the cheap methods the formula can be derived from (3) by using a model prediction

based on the frequency ω_i giving a regression depending on the sampling sequence (s_j),

$$\hat{S}_i = \frac{\sum_{j=1}^n (\hat{y}|_{\omega_i}(s_j) - \bar{y})^2}{\sum_{j=1}^n (y_j - \bar{y})^2}, \qquad \hat{y}|_\omega(s) = \frac{1}{n}\left(c_0 + 2\sum_{m=1}^M \mathrm{Re}\left(c_{m\omega} e^{2\pi i m\omega s}\right)\right), \qquad \bar{y} = \tfrac{1}{n}c_0.$$

The application of Parseval's Theorem then yields (FAST).
To compute total effects with FAST we use the frequency scheme $\omega_i = \omega_0^{i-1}$ where the basic frequency $\omega_0 = 2M+1$ is given by the maximal harmonic. Each frequency $\omega \le M\sum_{\ell=0}^{k-1}\omega_0^\ell = \frac{1}{2}(\omega_0^k - 1)$ in the Fourier spectrum of the output $\mathbf{Y}$ can be uniquely decomposed into

$$\omega = \sum_{i=1}^k \alpha_i(\omega)\omega_0^{i-1}, \qquad \alpha_i(\omega) \in \{-M, 1-M, \ldots, -1, 0, 1, \ldots, M-1, M\}.$$

If $\alpha_i(\omega) \neq 0$ then ω contributes to the total effect of input factor i. Hence to compute this value we can use two different approaches

$$\hat{S}_{Ti} = 2\frac{\sum_{\alpha_i(\omega)\neq 0}|c_\omega|^2}{\sum_{m\neq 0}|c_m|^2}, \qquad \hat{S}_{Ti} = 1 - 2\frac{\sum_{\alpha_i(\omega)=0}|c_\omega|^2}{\sum_{m\neq 0}|c_m|^2}, \quad \omega \in \left\{1, \ldots, \frac{1}{2}(\omega_0^k - 1)\right\}.$$

For computing higher order effects the combined zero patterns of the $\alpha_i(\omega)$ can be exploited.
In a different frequency selection scheme named "Extended FAST" (EFAST) a factor i of interest is assigned to a relative large frequency $\omega_i \gg 1$ and all others are assigned to low frequencies (say, $\omega_{j\neq i} = 1$). EFAST can be used to compute total effects as all frequencies below $\omega_0 = \omega_i - M\sum_{j\neq i}\omega_j$ do not contribute to the variance from factor i up to the Mth order. Hence

$$\hat{S}_{Ti} = 1 - 2\frac{\sum_{m=1}^{\omega_0 - 1}|c_{m\omega_i}|^2}{\sum_{m\neq 0}|c_m|^2} \qquad \text{(EFAST)}$$

Clearly, a new sample is needed for each of the factors. But this set-up also allows for the computation of the first order effect S_i via (FAST). Moreover, is has a smaller memory footprint compared to the full resolution FAST described above.
The Random Balance Design (RBD) [20] uses only the frequency $\omega = 1$ and generates a one-dimensional sample $\mathbf{U}$ with realisations $u_j = G_1(s_j)$ where $s = (s_j)_{j=1,\ldots,n}$ is equidistantly spaced in $[0,1]$. Then k random permutations $\pi_i : \{1,\ldots,n\} \to \{1,\ldots,n\}$ are generated, and $\mathbf{X}_i = \pi_i(\mathbf{U})$ is a sample for the ith input parameter. To find the first order Sensitivity Indices for the ith factor, the output $\mathbf{Y} = f(\mathbf{X}) = f(\mathbf{X}_1, \ldots, \mathbf{X}_k)$ is sorted with respect to the inverse of the ith permutation, $\pi_i^{-1}(\mathbf{Y})$, and then analysed via a Fourier transformation using (5) with y_j replaced by $\pi_i^{-1}(\mathbf{Y})_j$ and (FAST). This method can only estimate first order effects. Higher order effects can be estimated by introducing groups of different frequencies.

4.5 Effective Algorithm: Combining given data with Fourier amplitude

We can modify the idea behind RBD so that it can be applied to given data: Instead of generating random permutations, we construct permutations π_i from the columns $\mathbf{X}_i$ of the given data matrix

$\mathbf{X}$ such that each $\pi_i(\mathbf{X}_i)$ has a zig-zag-like shape and therefore has a power spectrum where the frequency $\omega = 1$ is predominant. These permutations are obtained by sorting and shuffling the input data. In particular, let $x = (x_j)_{j=1,\ldots,n}$ be a vector of realisations of the random variable X_i. To keep the notation short, we drop the dependency on i. We order $x = (x_j)$ increasingly and obtain an ordered vector $(x_{(j)})$ with $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. Now, taking all odd indices from $(x_{(j)})$ in increasing order followed by all even indices in decreasing order gives us a vector $(x_{[j]})$ with

$$x_{[j]} = \begin{cases} x_{(2j-1)}, & j \leq \frac{n+1}{2}, \\ x_{(2(n+1-j))}, & j > \frac{n+1}{2}, \end{cases} \qquad j = 1, 2, \ldots, n$$

that satisfies

$$x_{[j]} \leq x_{[j+1]} \quad \text{if } j \leq \tfrac{n+1}{2}, \qquad x_{[j]} \geq x_{[j+1]} \quad \text{if } j > \tfrac{n+1}{2}.$$

This shows that the entries in the vector $(x_{[j]})$ are in zig-zag order. There exists a permutation π_i with $\pi_i((x_j)) = (x_{[j]})$. As in RBD, this permutation is also applied to the output $\pi_i(\mathbf{Y})$. The Fourier transformation of the permuted output is analysed for frequency responses. Further details can be found in [10].
This approach is called “Effective Algorithm” for the computation of SI (EASI). It has been developed in the course of the PAMINA project. This benchmarking exercise is also used to test its performance.

5 The analytical benchmark cases

A first round of PA benchmark studies were performed by the members of the PAMINA 2.1.D work group, see Appendix B. In order to unify the results and to draw more attention to the nonlinear SA indicators we asked in a second simulation round for selected benchmarks cases with a prescribed setting. This setting consists of 25 runs at sample sizes of 100, 300, 1000, 3000, and 10000, computing mean, variance, R^2, R^{2*}, and first and total order effects (where available). The choice of the SI algorithms was left to the participants of this second round. Contributions were received from Facilia (Sweden), GRS Cologne (Germany), JRC Petten (The Netherlands), and TU Clausthal (Germany). In the following we sometimes mark the contributions of these participants with the abbreviations FCL, GRS, JRC, and TUC, respectively.
Most of the following graphics are shown in form of box plots derived from the available 25 runs per sample size. The box plots show the lower and the upper quartile, the median value is marked with a dot. The whiskers in the plots are lines illustrating the data range. Outliers are detected using three-times the inter-quartile range.

5.1 Ishigami function: A model with three input parameters and higher order effects

The Ishigami test function is a three-parameter model. It is in so far interesting as the second and third input factors have a Pearson Correlation Coefficient of zero. A variance-based SA

retrieves a 44% first order effect for the second input factor, but the third input factor shows no first order effect. Only when estimating total effects, the third factor is attributed to 24% of the output variance.
The Ishigami function is given by

$$Y = \sin X_1 + 7\sin^2 X_2 + 0.1X_3^4 \sin X_1$$

where $X_i \sim U(-\pi,\pi)$ are uniformly distributed in $[-\pi,\pi]$. The values of $R^2 \approx R^{2*} \approx 20\%$ imply that the results from a standard or rank-transformed linear regression are not very powerful. Hence we have to look into nonlinear SA methods.
Figure 4 displays inputs and outputs from this model prepared for the FAST method. From the upper graphics showing the input we see that $\omega_3 = 1$, $\omega_2 = 11$ (by counting the peaks), and $\omega_1 = 11^2 = 121$. Note that the scatter-plot of the ω_1 input data shows Moiré-patterns which indicates that the sampling size, $n = 2000$ (Nyquist frequency 1000), is small compared to the maximal frequency in use $5(1+11+121) = 665$. Considering the model output in the center of Figure 4, we find a periodic behaviour which is not directly related to the input frequencies. The power spectrum of the output in the lower part of the figure shows more details. First order effects are coloured blue, second order effects green and third order effects are coloured red. We find noticeable first order effects for frequencies 44 (x_2^4), 121 (x_1), and 363 (x_1^3). Second order effects group around the first order effects of x_1, for frequencies 119 and 123 ($x_1x_3^2$), 117 and 125 ($x_1x_3^4$), 361 and 365 ($x_1^3x_3^2$). A well-equipped eye might also spot frequencies 115 and 127 ($x_1x_3^6$), but as the maximal harmonic is $M = 5$ this part of the variance is mis-classified as third order effect.
As x_2 enters as fourth power into the model and x_3 enters the model only indirectly in combination with x_1, their influences are not detected by linear regression techniques.
Let us now consider the performance of different algorithms for this example. Figures 5, 6 show the results of the first order effects for x_1 and x_2, respectively. Nine different algorithms were used. The TUC implementation of the FAST method errs on too few realisations for sizes 100 and 300, EFAST only for sample size 100. For a FAST analysis of a $k = 3$ parameter model with a maximal harmonic $M = 3$ the TUC implementation needs at least $2M(1+(2M+1)+\cdots+(2M+1)^{k-1}) = 6\cdot(1+7+49) = 342$ realisations, for EFAST $2kM(2M+1) = 126$ realisations.
The first five algorithms for each sample size are cheap methods working on the same data set generated with simple random sampling. Their performance is nearly the same. The indicators generated via the correlation ratio method which uses the mean of conditional variances (ECV) differ from those generated by calculating the variance of the conditional mean (VCE).
The Ishigami-Homma-Saltelli (IHS) algorithm seems to be the only algorithm which produces unbiased estimates. But the sample size n is only the basic sample size for use in the IHS method so that a total of $(3+2)n = 5n$ model evaluations are needed. However, there is no explanation for the wide variance compared to the cheap methods when estimating S_1. An overview of the performance of the different IHS implementations is found in Fig. 7. The TUC version uses a formula [18] that reduces the error introduced via cancellation, hence the true values should be better approximated. However, this theoretical result does not become apparent in the figure.
The behaviour of the IHS algorithm changes drastically when using Sobol's LP_τ sequence as pseudo random number source, see Figure 11. Then even for small sample sizes good estimates are computed. Here, the basic sample size is rounded to the next power of 2.

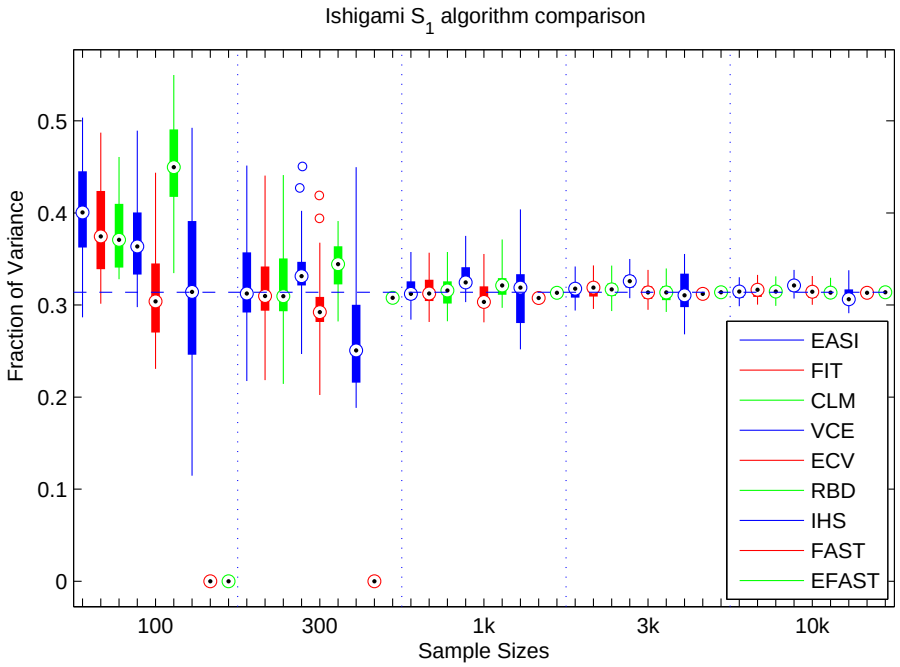

Figure 5: TUC results – Box plots for S_1.

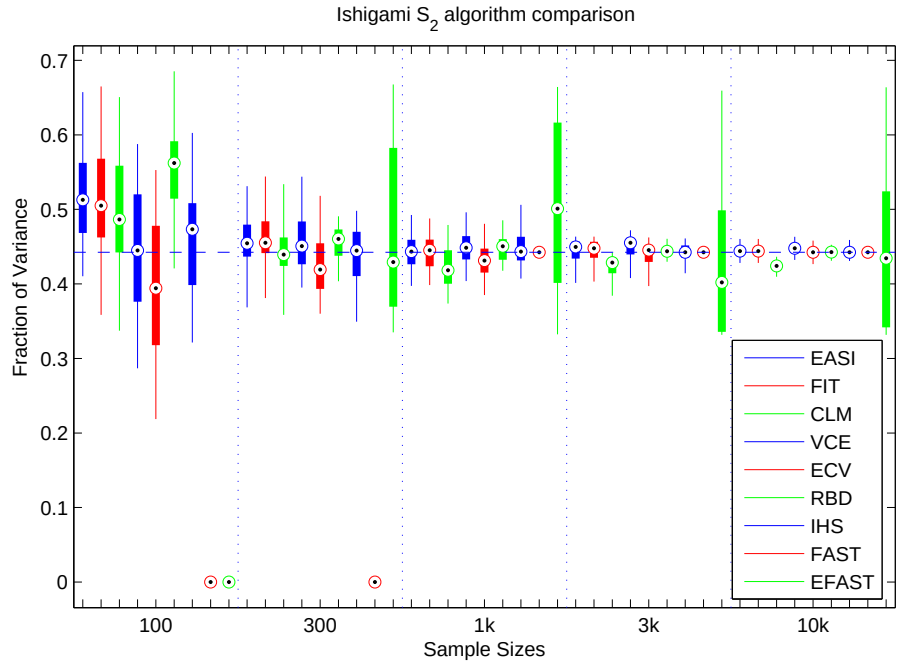

Figure 6: TUC results – Box plots for S_2.

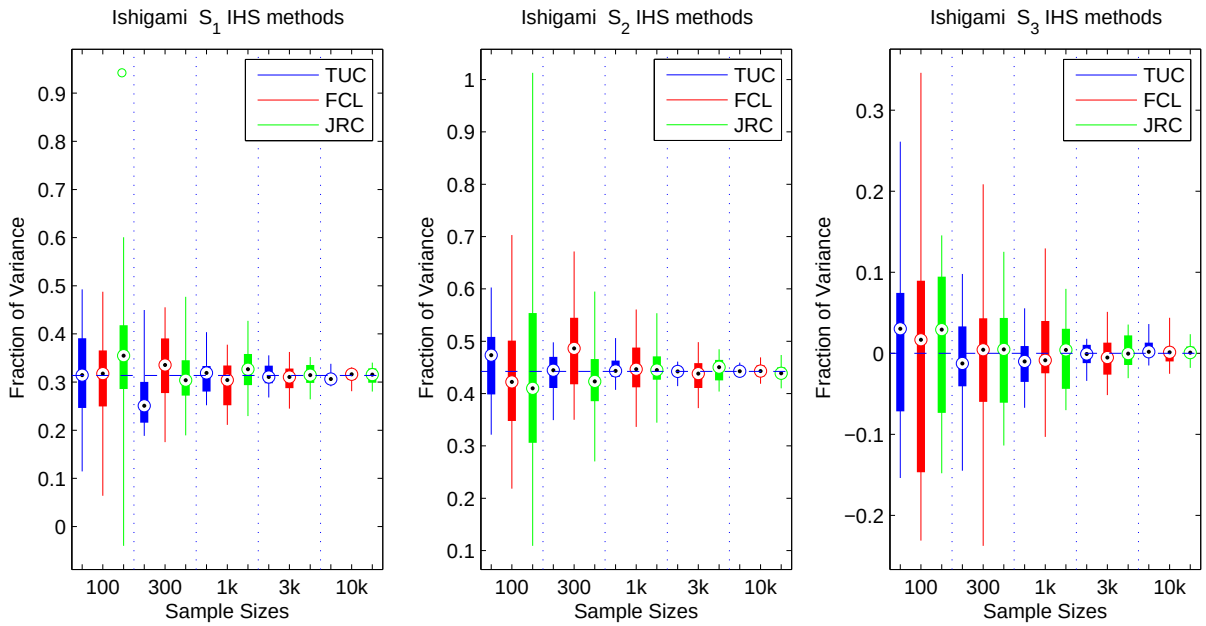

Figure 7: Ishigami-Homma-Saltelli algorithm comparison.

The RBD method –although using an artificially generated sample– seems to have no better properties than the cheap methods, see also Figure 9 where all Fourier-based methods are compared. A comparison of correlation ratio implementations from TUC, GRS-Cologne and JRC-Petten for all parameters is found in Figure 8. Note that the different correlation ratio implementations use simple random sampling (ECV,VCE,SRS,CR2P,CR,CP5S), Latin hypercube sampling (LHS), and Latin hypercube sampling with the selection of the conditional mean in each subinterval (LHS-M). However, the use of different input sampling schemes does not produce significantly different results. The CR2P, CR and CP5S methods study different subsample sizes: CR2P uses a two-interval partition, while CR5S requests a partition which is constructed in such way that every subsample contains five realisations, and CR uses a subsample size resembling the rule-of-thumb $\ell = \lceil \sqrt{n} \rceil$. Here, the subsample-size-five setting overestimates the true values while the two-intervals setting produces an underestimation, all other estimators produce consistent results. For S_3, the estimation of true zero values via CR methods is also difficult, only ECV and CR2P produce unbiased results.

If FAST methods are available then they produce exact estimates for moderately-sized samples. For the computation of S_2 via EFAST(TUC) strange things happen: The range of the computed estimates is not reduced by increasing the sample size. Maybe there are some resonances internal to the model that react to joint input frequencies $\omega_1 = \omega_3 = 1$. Figure 9 shows the comparison of Fourier-based methods from TUC and Facilia. Facilia's implementation of EFAST uses a simple frequency selection for sample size 100, hence now the same-sized box plot appears as for the TUC implementation with sample sizes larger or equal to 300. For larger sample sizes, the Facilia version of EFAST uses a different frequency selection scheme utilising different small frequencies and produces much better results which are on par with FAST.

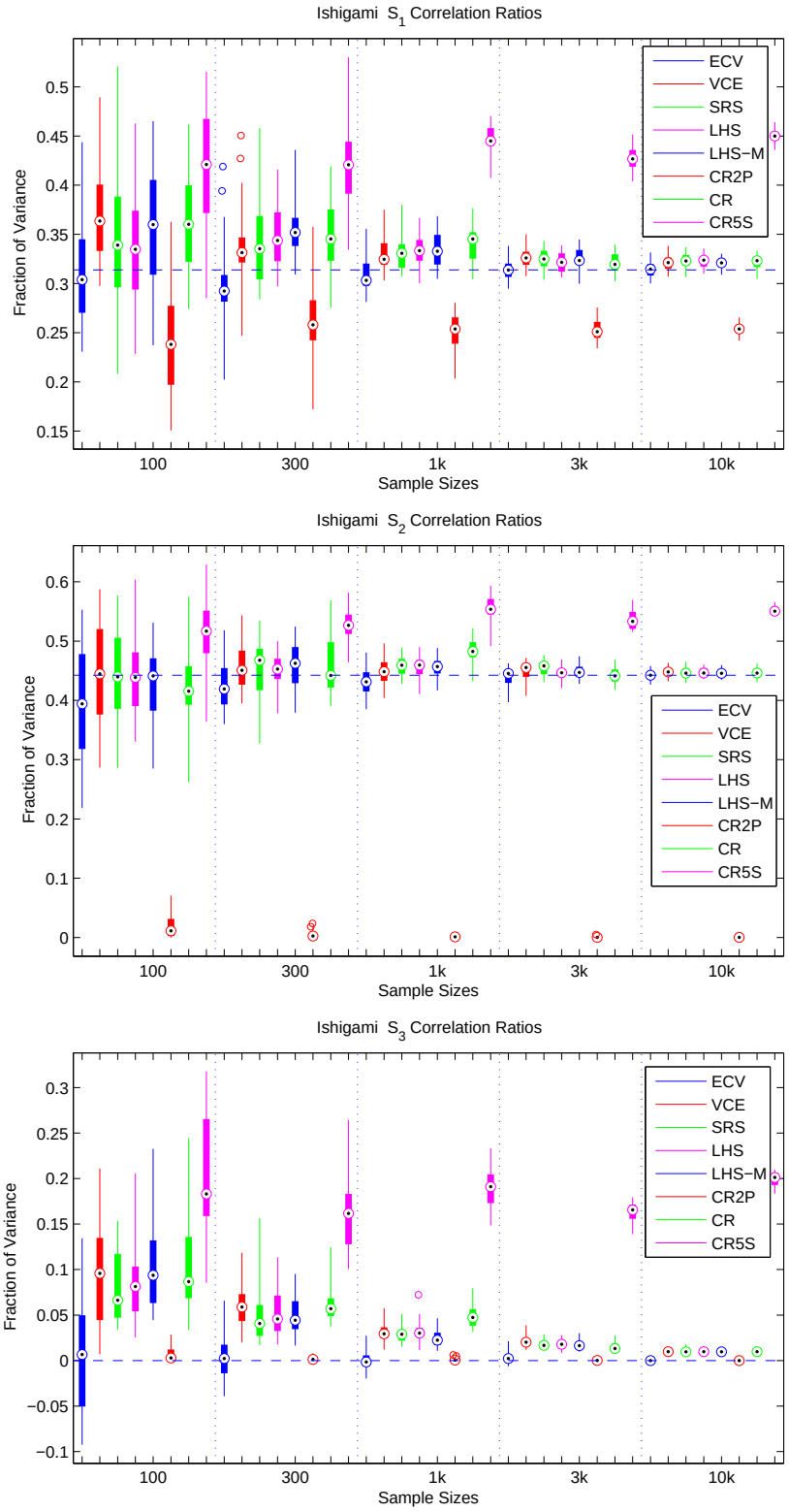

Figure 8: Correlation Ratio methods for the Ishigami function.

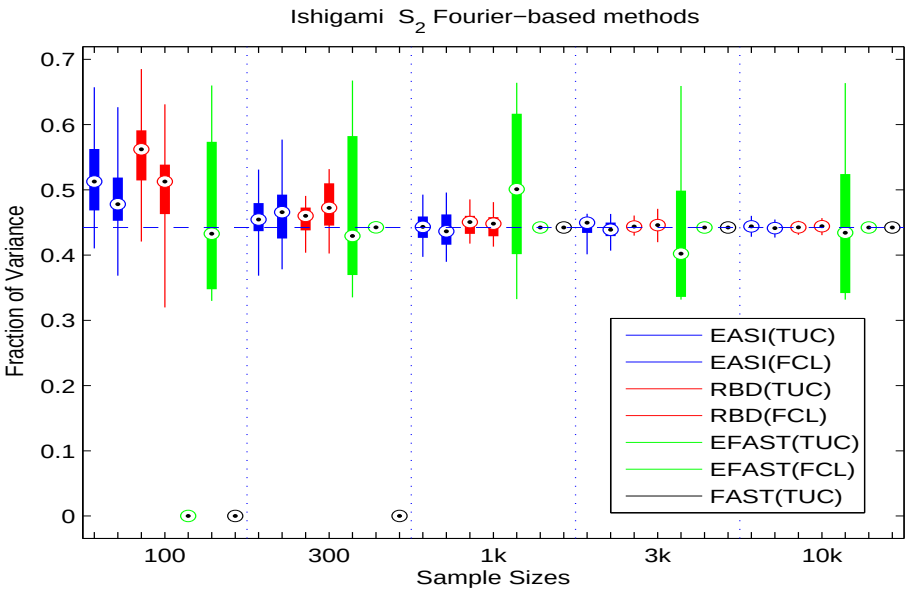

Figure 9: Fourier-based methods for the Ishigami function.

The results for the total effects of x_3 are presented in Figure 10. The number of available algorithms for the computation of total effects is smaller than the number of algorithms for the computation of main effects. The cheap estimators seem to be biased but consistent, IHS is unbiased but has large variations: even negative estimates are generated for sample size 100 (which are clipped out from the graphics). The FAST methods produce estimates that are nearly uninfluenced by the random frequency shift, there are only little differences in the performance of the different implementations.
For the Sobol' method TUC used the next power of 2 for the sample size which produces good estimates, while Facilia used the given value as basic sample size which produces less accurate results for small sample sizes, see Figure 11. As the Sobol' method uses a special sampling scheme there is only one estimate available per sampling size. The figure therefore shows all main and total effects in one graphics. Note that $S_2 = S_{T2}$ so that two values are plotted on the same spot.

5.2 A discontinuous switch example

If a computational model has input parameters that drastically change the output behaviour then these discontinuities may impose numerical problems for the used SA algorithms. Hence we analyse the following test function,

$$Y = \begin{cases} X_2, & \text{if } X_1 > \frac{1}{2} \\ -X_2, & \text{if } X_1 \leq \frac{1}{2} \end{cases}, \qquad X_1, X_2 \sim U(0,1).$$

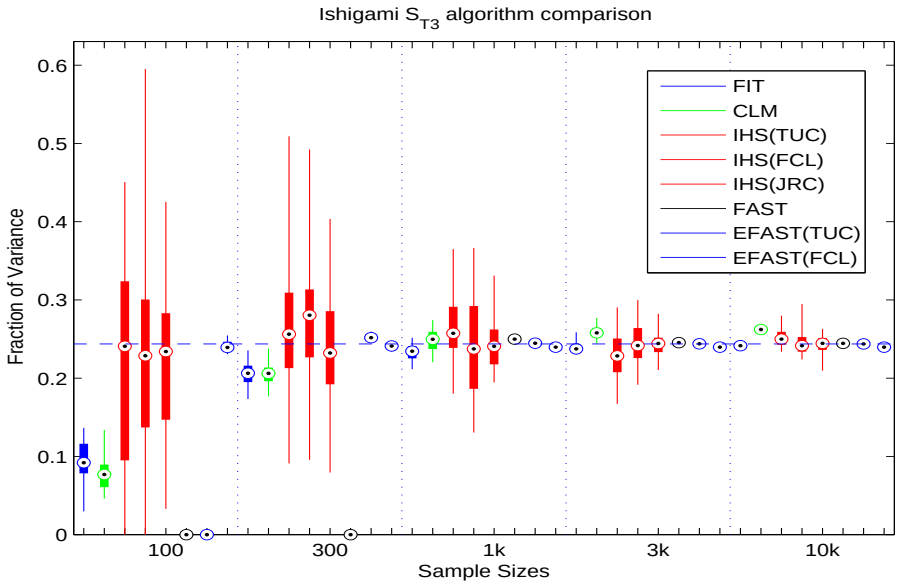

Figure 10: Box plots for S_{T3}.

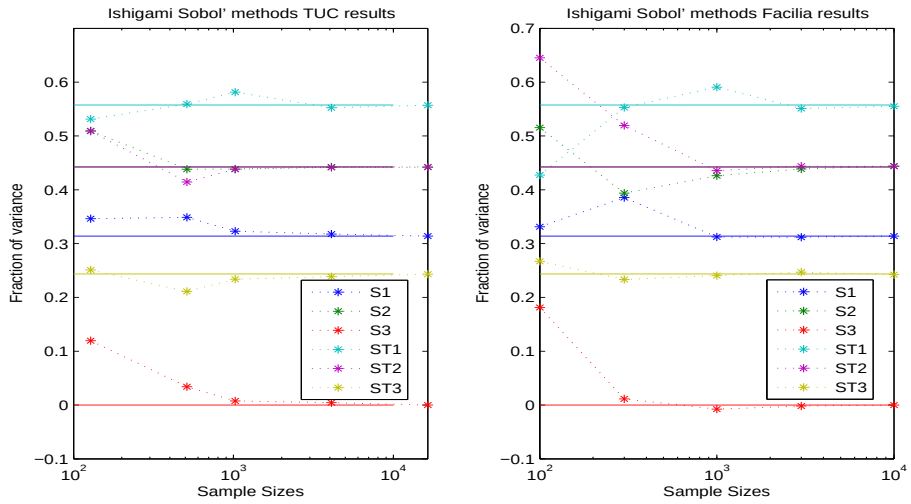

Figure 11: First and total effects using Sobol's method for the Ishigami function.

The expected values are $S_1 = \frac{3}{4}$, $S_2 = 0$, $S_{T1} = 1$, $S_{T2} = \frac{1}{4}$. As $R^2 \approx R^{2*} \approx 56\%$ the results from a standard or rank-transformed linear regression are only of limited use.

Figure 12 shows the results of the estimation of S_1 using the TUC computations. We conclude that in this case only the conditional linear model (CLM), the correlation ratio methods (VCE, ECV), the IHS algorithm, and EFAST can cope with the non-linearity, all other algorithms systematically produce too low estimates. Using the correlation ratio methods, for all but the 3000 sample size the number of subsamples in the partition given by the rule of thumb is even so that the

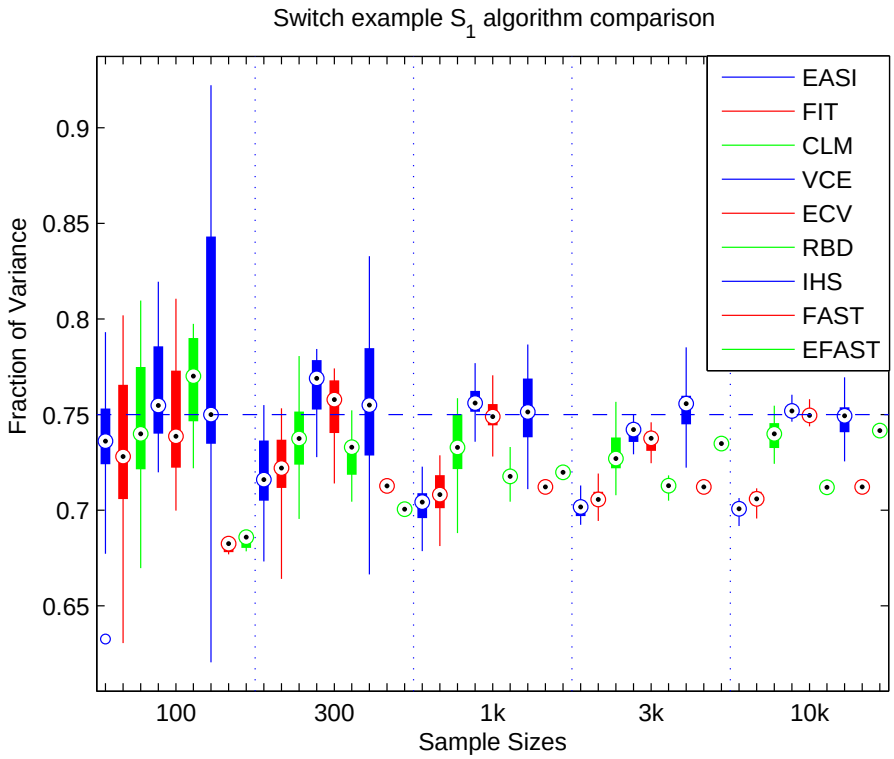

Figure 12: TUC results – Box plots for S_1.

discontinuity at $x_1 = 0.5$ is resolved by the associated partition. One sees that the performance of the CR methods is slightly different for the 3000 sample size: While the estimates for the other sampling sizes are nearly unbiased, the estimate for the size-3000 sample underestimates the true value. Figure 16 shows the results of other CR methods. For the computation of S_1 via CR the use of Latin Hypercube sampling schemes is advantageous, but there are no differences in the results of S_2 for varying sampling schemes (SRS,LHS,LHS-M). As already noted in the previous example the CR estimates based on a subsample size of five realisations or on two intervals produce inferior results.

Returning to Figure 12 the conditional linear model (CLM) uses a partition with an even number of intervals, $M = 6$. Hence this method benefits from the same effect as the CR methods.

The results for the TUC implementations of FAST and EFAST differ considerably. While EFAST converges slowly FAST remains stubbornly at a low level. This is due to the fact that for the sample of size 10,000 EFAST uses the maximal harmonic $M = 35$ and frequency $\omega = 71$ while FAST is restricted to maximal harmonic $M = 8$ and frequencies $\omega_1 = 17, \omega_2 = 1$. Other implementations of Fourier-based methods suffer also from the fixed maximal harmonic M, see Figure 15 which shows that the algorithms with the same value of M produce equivalent estimates.

Figure 14 shows first and total effects estimated via IHS methods. The variations are large when compared to other methods. However, unbiased estimates are produced. Compared to these results from the IHS methods, the Sobol' algorithm gives almost the correct estimates even for small sample sizes, cf. Fig. 17. Even a sampling size which is not a power-of-2 produces no

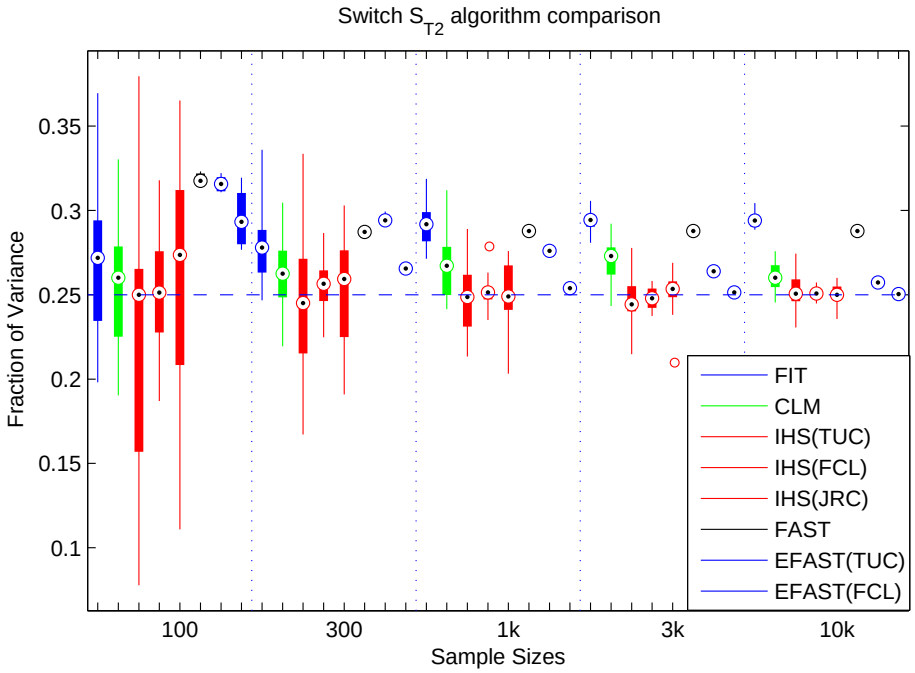

Figure 13: TUC results – Box plots for S_{T2}.

eye-catching effects.
For a two-parameter model the total effects are given by $S_{Ti} = 1 - S_{3-i}$, $i = 1,2$. Therefore the explicit calculation of total effects is not necessary. Nevertheless, Figure 13 shows computed total effects. Again, the polynomial fit and the FAST algorithm with bounded maximal frequency fail to catch the exact value. Figure 17 shows that the total effects computed by the Sobol' method are in good agreement with the theoretical values.

5.3 A linear model with dependent input data

In theory, independent input parameters are required for performing variance-based SA. It is not clear what happens with the SA algorithms in the presence of dependencies between the input parameters or to what extend the results can be interpreted. This example highlights some of the problems encountered when processing dependent data. The function under inspection is given by the linear model $Y = X_1 + X_2$ where the input parameters have a joint probability density function given by

$$p(X_1, X_2) = \begin{cases} 2 & \text{if } 0 \le X_1, X_2 \le 0.5 \text{ or } 0.5 \le X_1, X_2 \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

The expected values are $S_i = \frac{\mathrm{V}[\mathrm{E}[Y|X_i]]}{\mathrm{V}[Y]} = \frac{13}{14} = 0.9285714\ldots$ hence $S_{Ti} = 1 - S_{3-i} = \frac{1}{14} = 0.0714285\ldots$, $i = 1,2$. In a linear model with independent parameters we would expect

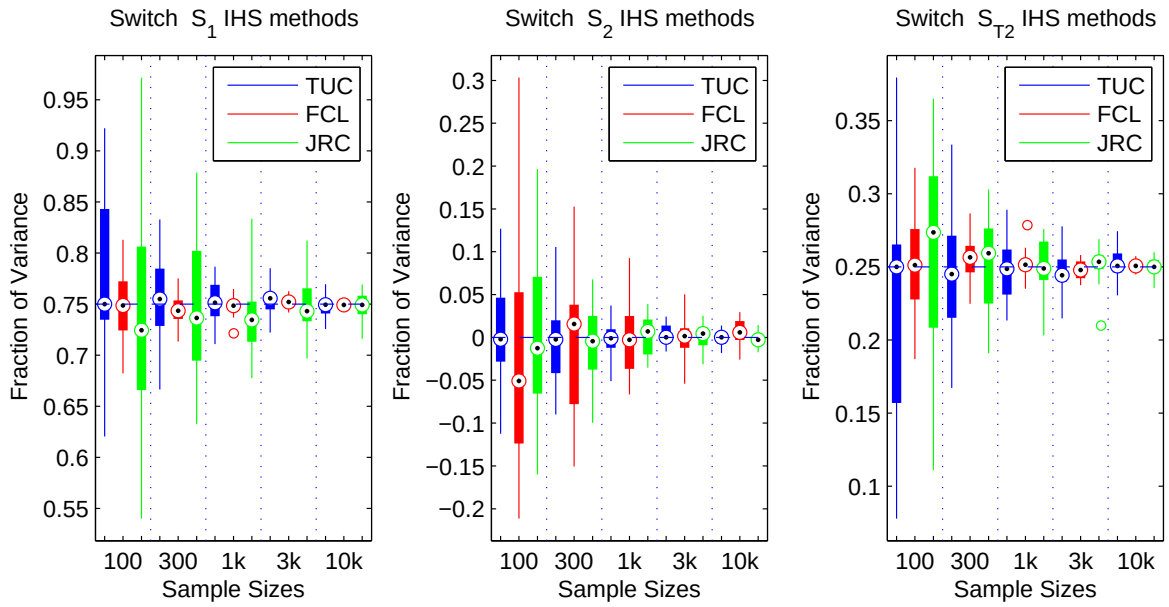

Figure 14: Ishigami-Homma-Saltelli algorithm comparison.

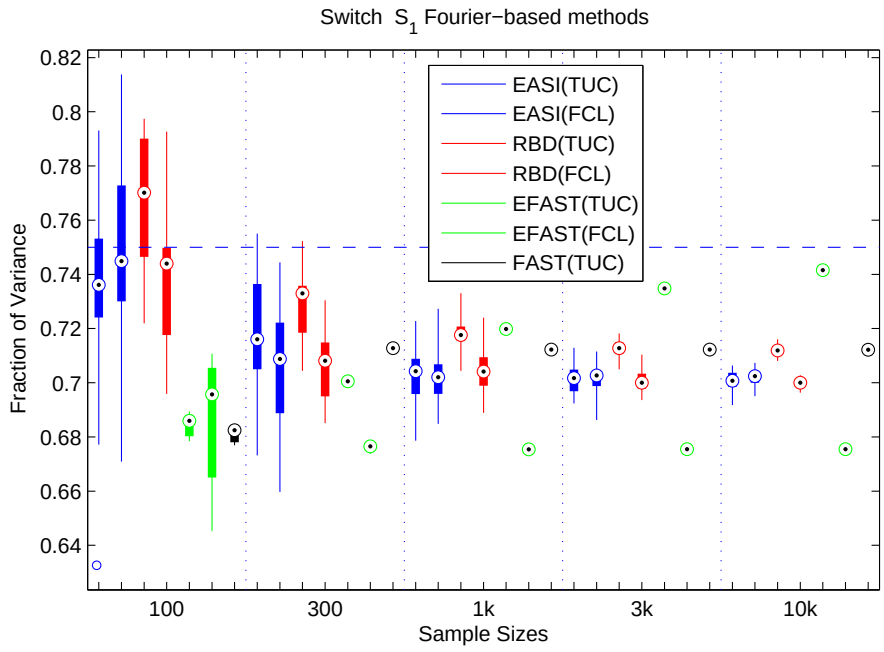

Figure 15: Fourier-based methods for the Switch example.

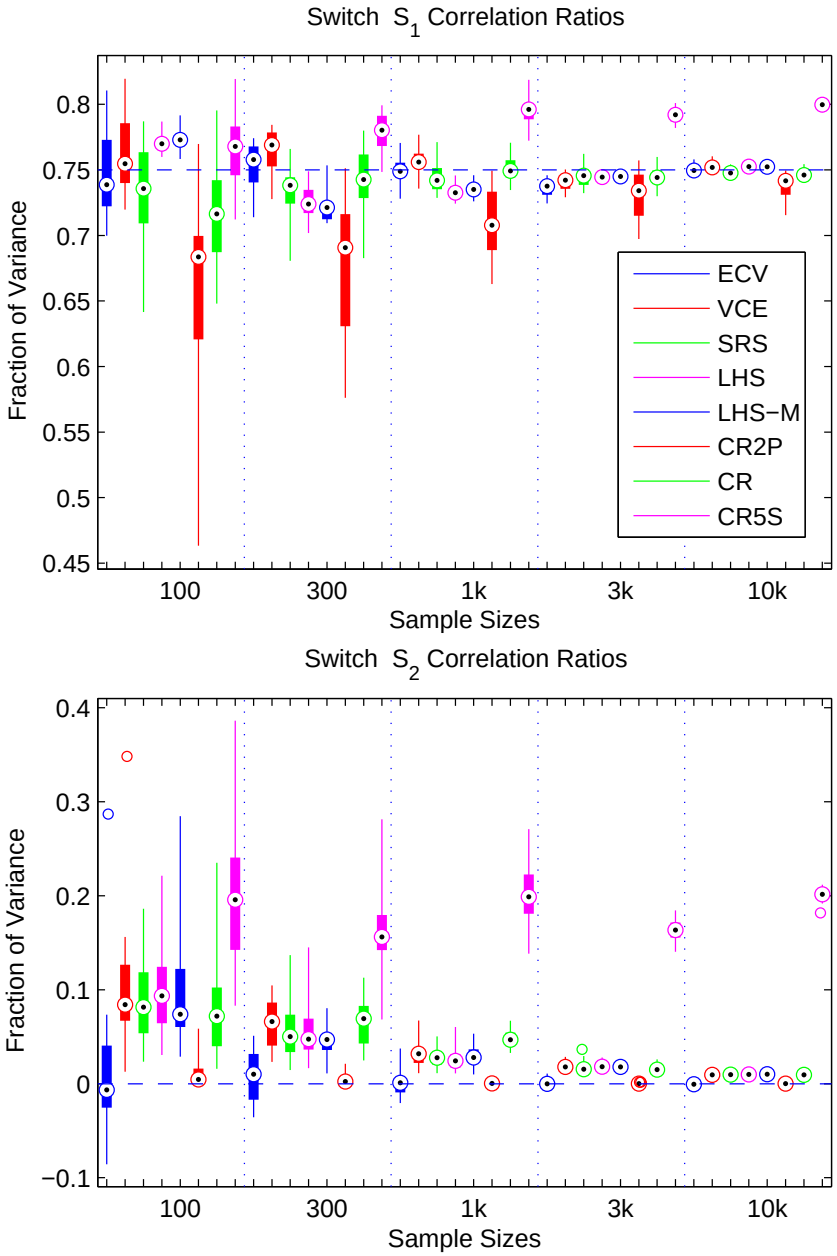

Figure 16: Correlation Ratio methods for the Switch example.

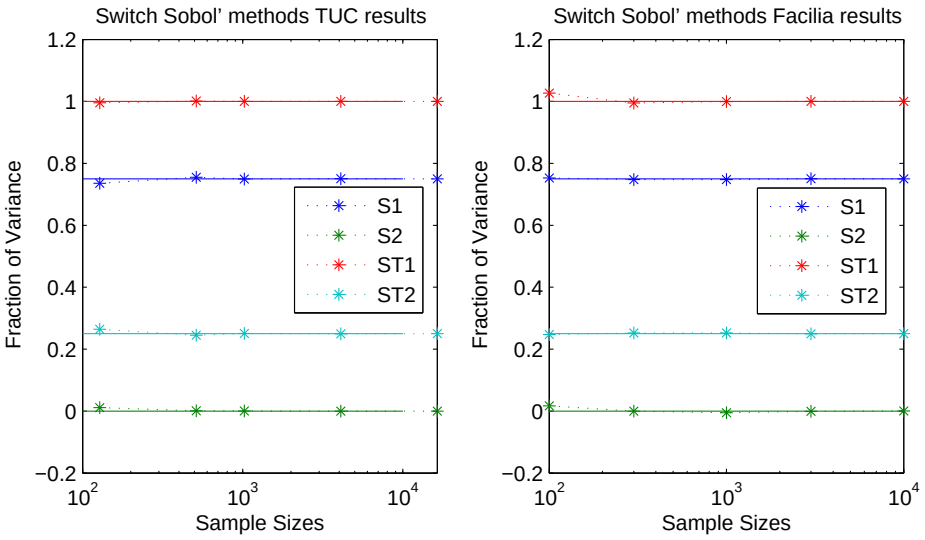

Figure 17: Sobol's method for the Switch example.

(a) the sum of all main effects to be 1, and

(b) the total effects to be larger or equal to the main effects.

However, we are not dealing with an independent input parameter setting. Figures 18 and 19 show the results for the Sensitivity Indices S_1 and S_2. Since the model is symmetric in x_1 and x_2 the results should also be the same. This is the case where the methods for calculation of the Sensitivity Index require no additional information for parameter transformations, i.e., for the cheap methods. For these, the results show a fair agreement with the expected results, and since the model is linear the results are already valid for small sample sizes. The problems arise in case of Sobol', IHS, RBD, FAST or EFAST methods as the sample generation not only has to satisfy the needs of a special sampling scheme but also to realise the joint probability. Depending on the input parameter transformation which may use the marginal distribution $p(X_1|X_2)$ or $p(X_2|X_1)$ we see different results, in this case the results of S_2 are definitely wrong. However, with the careful choice of a parameter transformation one obtains the correct results, as the Facilia implementation of the IHS method and the GRS-Cologne implementations of the special sampling scheme CR methods (SRS, LHS, LHS-M) show.

Moreover, the CR methods, IHS and EFAST seem to be consistent, for the rest of the algorithms a small bias seems to be present since an increase in the sample size does not lead to better results. Figure 20 shows a selection of CR methods.

The IHS algorithm shows the largest variation. Again, when using the Sobol' sequence for the sample generation the quality improves drastically (not shown). For total effects, the cheap methods produce good estimates. For other methods which already gave bad estimates the "wrong" parameter transformation is now in effect for S_{T1} while S_{T2} is estimated correctly (not shown).

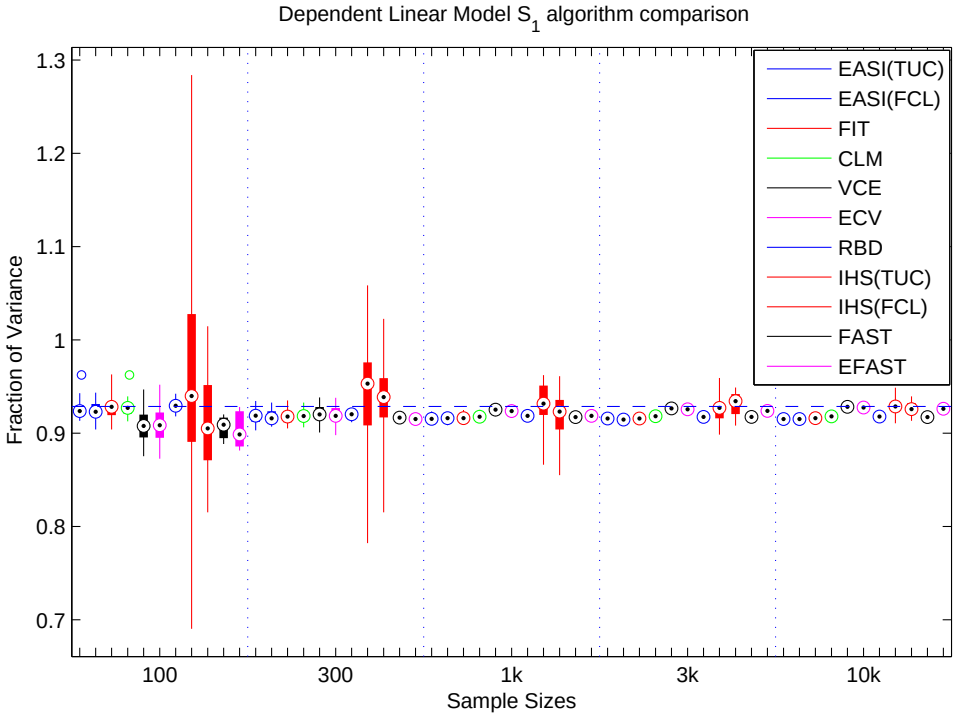

Figure 18: TUC and Facilia results – Box plots for S_1.

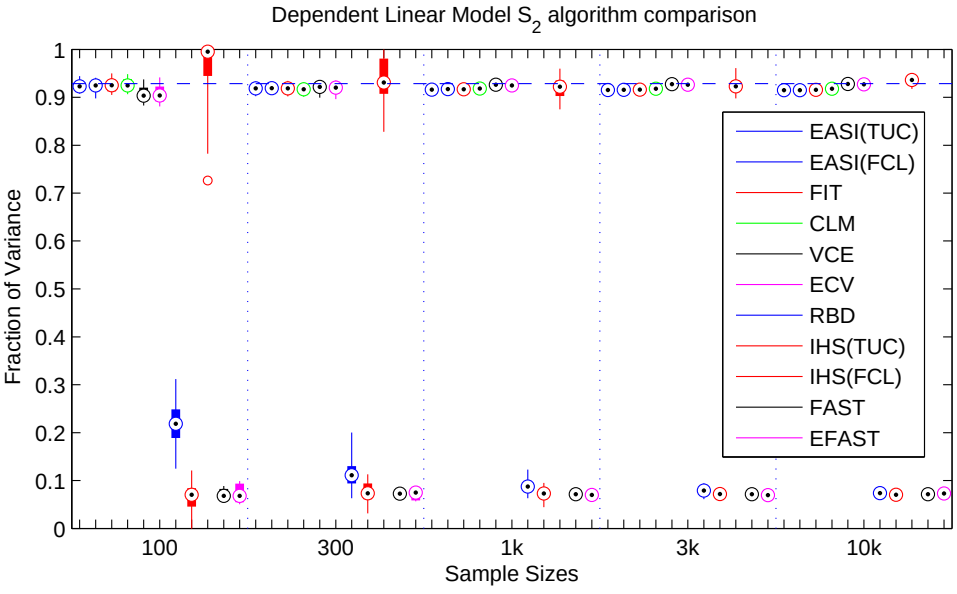

Figure 19: TUC and Facilia results – Box plots for S_2.

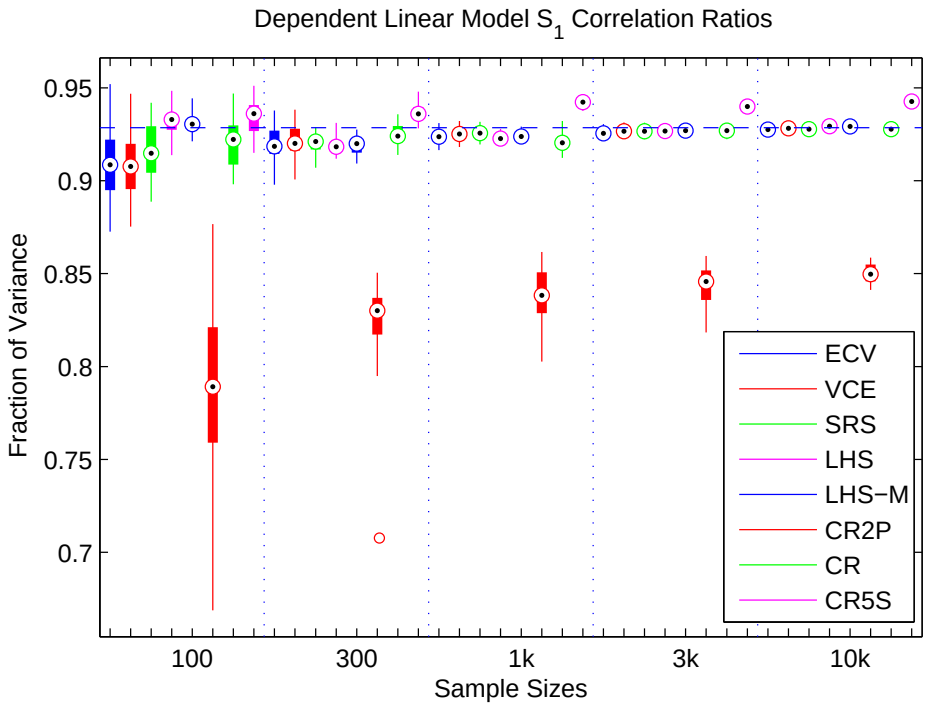

Figure 20: Correlation Ratio methods for the dependent linear model.

5.4 Sobol' g function: A model with eight input parameters

Real-world models have many input parameters. Hence a test case where many input parameters are considered shows if an algorithm is robust enough. A well-studied test function is the non-monotonic Sobol' g-function which is given by

$$Y = \prod_{i=1}^{k} \frac{|4X_i - 2| + a_i}{1 + a_i}, \qquad X_i \sim U(0,1) \tag{6}$$

where $k = 8$, $(a_i) = (0, 1, 4.5, 9, 99, 99, 99, 99)$. The first parameter is most influential, the influence decreases through the rest of the parameters until parameters five to eight become equally uninfluential. Due to the symmetry in the formula, we have $R^2 = R^{2*} = 0$. Hence the results based on linear regression are of no value for the Sensitivity Analysis.
The results for S_1 are reported in Figure 21. A full resolution FAST with $k = 12$ parameters needs more than 10,000 realisations so that there are no results available for this particular method. Instead, we feature a guest appearance of Jansen's Winding Stairs algorithm. Its results should be comparable to the IHS method (as both require the same number of model evaluations).
Here, the performance of the Winding Stairs algorithm for S_1 is slightly better than the results from IHS.
The TUC EFAST implementation needs at least $2kM(2M+1) \geq 366$, $M \geq 3$, realisations to work. Hence the first two sample sizes allow no EFAST(TUC) analysis. For the other sample sizes, the

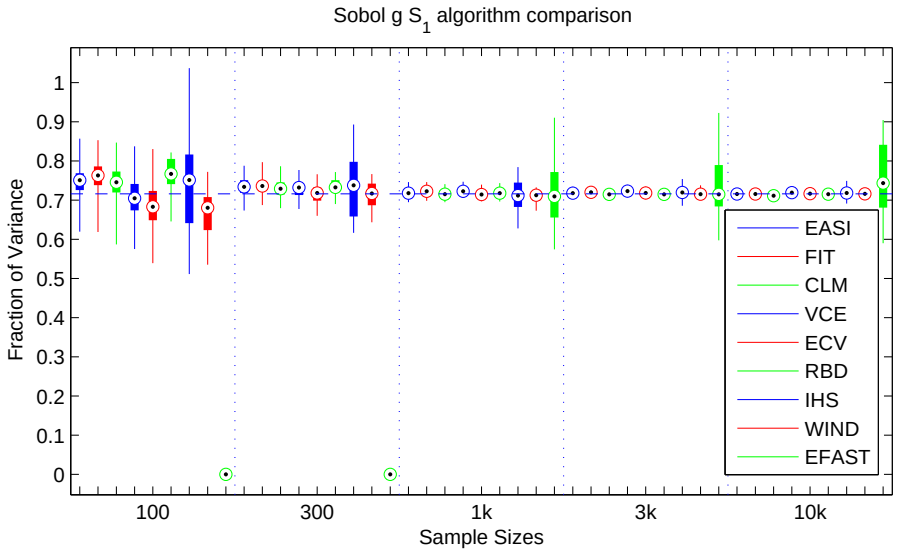

Figure 21: TUC results – Box plots for S_1.

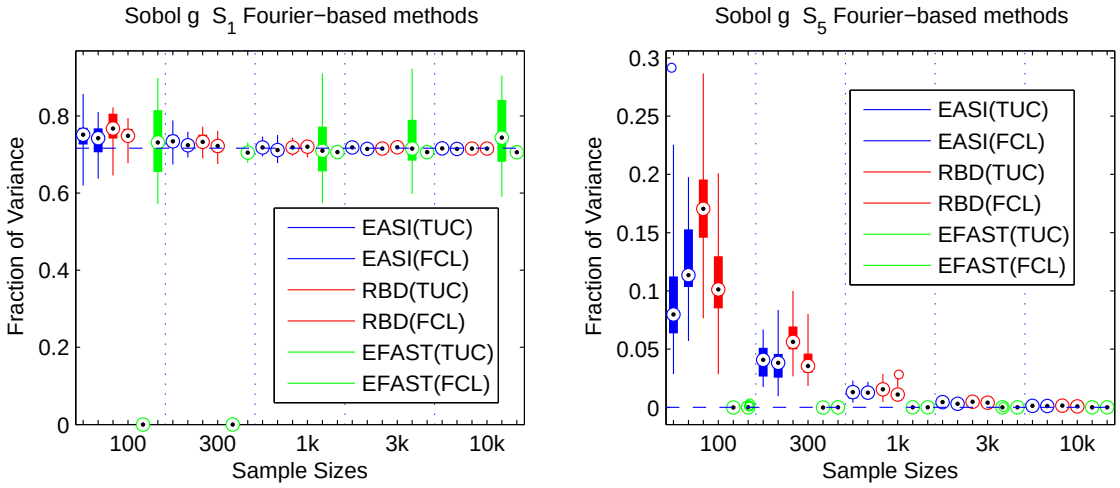

Figure 22: Fourier-based methods for Sobol' g function.

simple frequency scheme of EFAST(TUC) is not well suited: Again, there is no convergence to the real value. The performance of other Fourier-based implementations is reported in Figure 22. The results for S_5 are reported in Figure 24. As the fifth parameter is un-influential its Sensitivity Index is close to zero. Note that for the first two sample sizes, EFAST produces an exact zero as there are not enough realisations available. The cheap methods (asides ECV) and RBD overestimate the exact value.

Compared to the IHS method that produces good estimates even for small sample sizes, the Winding Stairs implementation gives the worst estimates of S_5 of all tested algorithms.

Nearly all Correlation Ratio methods have problems identifying a close-to-zero Sensitivity Index, see Figure 25 where S_5 is estimated.

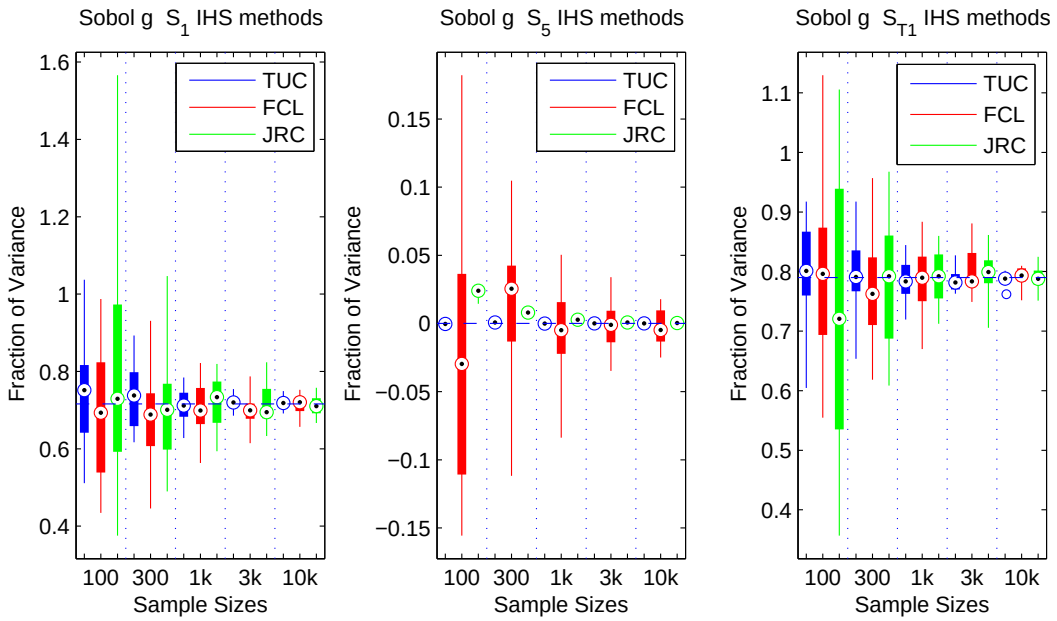

Figure 23: Ishigami-Homma-Saltelli algorithm comparison for Sobol' g function.

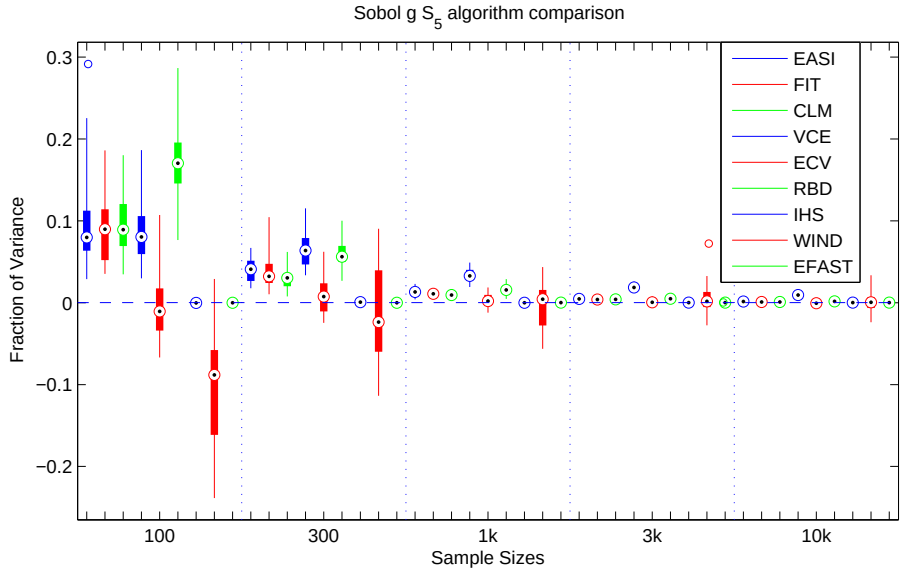

Figure 24: TUC results – Box plots for S_5.

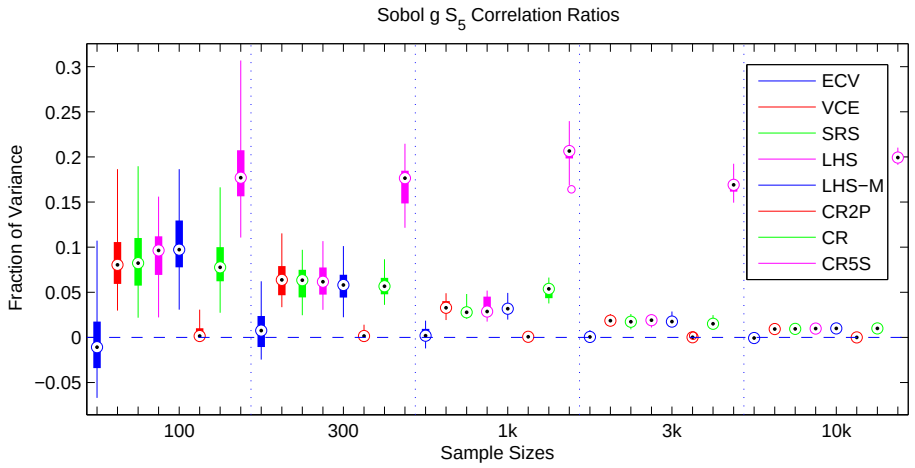

Figure 25: Correlation Ratio methods for Sobol' g function.

There are noticeable differences in the performance of the different implementations of the IHS method, see Figure 23. For the Sobol' method no abnormalities can be spotted (not shown). The estimates of the Sensitivity Indices for the uninfluential factors 5 to 8 are below 0.2% even for the basic sample size of 128.

6 The PA case example: The GTM Level-E model

After discussing the analytical in the previous section we now draw our attention to a complex geosphere transport model (GTM). In various publications (see [20], and [15] for a review), the PSACOIN Level E code [8] was used both as a benchmark of Monte Carlo simulations and as a benchmark for Sensitivity Analysis methods. This computational model calculates the radiological dose rate to humans over geological time-scales due to the underground migration of radionuclides from a hypothetical nuclear waste disposal site through a system of idealised natural and engineered barriers. The model has a total of 33 parameters, 12 of which are taken as independent uncertain parameters. The distributions of these uncertainties are either uniform or log-uniform distributions, the parameters of which have been selected on the basis of expert judgement. For a description of these uncertain input parameters see Table 1. These values are taken from [12] where further information including a mathematical description of the GTM Level-E model is available. The supplied binary model outputs the dose rate in $\frac{\mathrm{Sv}}{a}$ which stem from the radioactive Iodine-129 nuclide and the dose rate from the Neptunium-237 decay chain, moreover the maximum of the dose rates up to a given point (for I and Np) and the total dose rate per time-step. The change from the Iodine decay to the Neptunium decay chain introduces non-linearities into the model which are major obstacles for the Sensitivity Analysis. The Level-E model is also discussed in [1, Annex 1]. Here only the influence of Iodine is studied, the Neptunium decay chain is not considered.

The issue of time-dependent results deserves some further attention. The Level-E benchmark therefore provides sensitivity measures for the following entities,

Table 1: Uncertain input parameters for GTM Level-E.

Parameter		Description	Distribution	Range	Unit
1	T	Containment time (source)	uniform	$10^2 \ldots 10^3$	a
2	k_I	Leach rate for Iodine (source)	log-uniform	$10^{-3} \ldots 10^{-2}$	$1/a$
3	k_C	Leach rate for Np decay chain (source)	log-uniform	$10^{-6} \ldots 10^{-5}$	$1/a$
4	v^1	Water velocity (1^{st} layer)	log-uniform	$10^{-3} \ldots 10^{-1}$	m/a
5	l^1	Length (1^{st} layer)	uniform	$100 \ldots 500$	m
6	R_I^1	Iodine retardation (1^{st} layer)	uniform	$1 \ldots 5$	$-$
7	γ_C^1	Np chain retardation multiplier (1^{st} layer)	uniform	$3 \ldots 30$	$-$
8	v^2	Water velocity (2^{nd} layer)	log-uniform	$10^{-2} \ldots 10^{-1}$	m/a
9	l^2	Length (2^{nd} layer)	uniform	$50 \ldots 200$	m
10	R_I^2	Iodine retardation (2^{nd} layer)	uniform	$1 \ldots 5$	$-$
11	γ_C^2	Np chain retardation multiplier (2^{nd} layer)	uniform	$3 \ldots 30$	$-$
12	W	Stream flow rate (biosphere)	log-uniform	$10^5 \ldots 10^7$	m^3/a

- Peak of total dose rate,
- Time of occurrence of this peak, and
- Total dose rate (time-dependent, 25 time-steps equally distributed over a logarithmic scale from 10^4 to 10^6 years).

From the experience gained in analysing the analytical test cases, TUC decided to approach the SA problem by the following two paths. On the one hand, a simple random input sample of size $75,000 \times 12$ was generated and the associated model output was analysed using cheap methods, allowing for an analysis of 25 samples of size 3000 each. On the other hand, a basic and an alternative input sample of size 4096×12 were generated using Sobol's LP_τ sequence and the samples $\mathbf{X}_i$ were added to this input set, yielding an overall input sample size of $57,344 \times 12$. Both methods allow for estimates using smaller sample sizes by picking suitable submatrices.
Facilia computed first order and total effects using IHS and EFAST methods, and first order effects using RBD and EASI methods. For each method, 25 runs of sample sizes 100, 300 and 1000 were computed.

6.1 Peak of total dose rate

The peak of the total dose rate is not directly available as model output. Two simple approaches can be used, either by taking the maximum of the total dose (ignoring effects in-between time-steps) or by taking the maximum of the two peak doses "up to" the latest available time-step (ignoring effects which occur when Iodine as well as the decay chain both significantly contribute to the dose). There are cases where both values differ by a factor of over 5000 which suggests numerical problems with the model. For our analysis we have chosen the data from the first approach which seems numerically more stable.

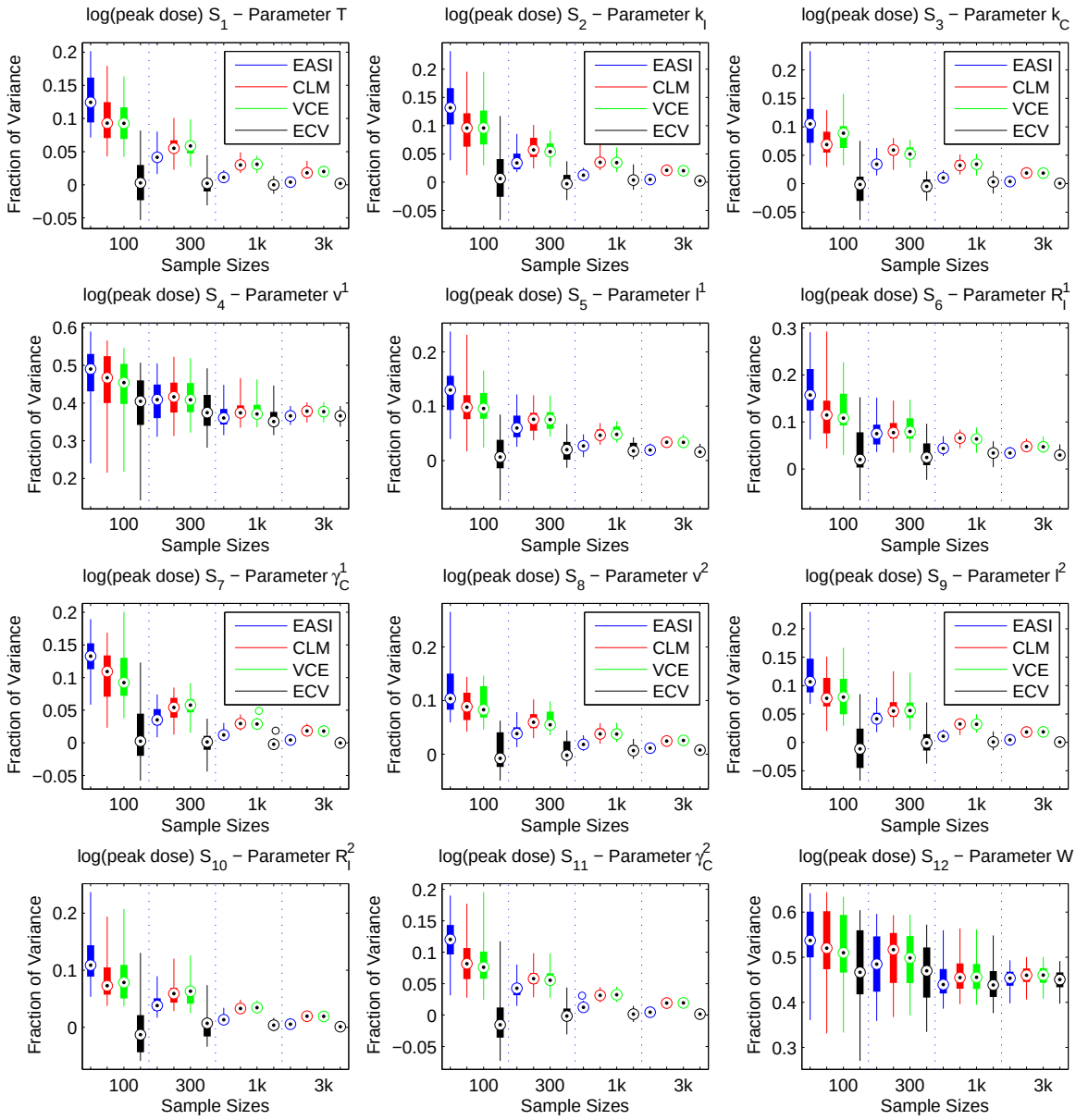

Figure 26: Cheap Sensitivity Analysis of $\log(\text{peak dose rate})$.

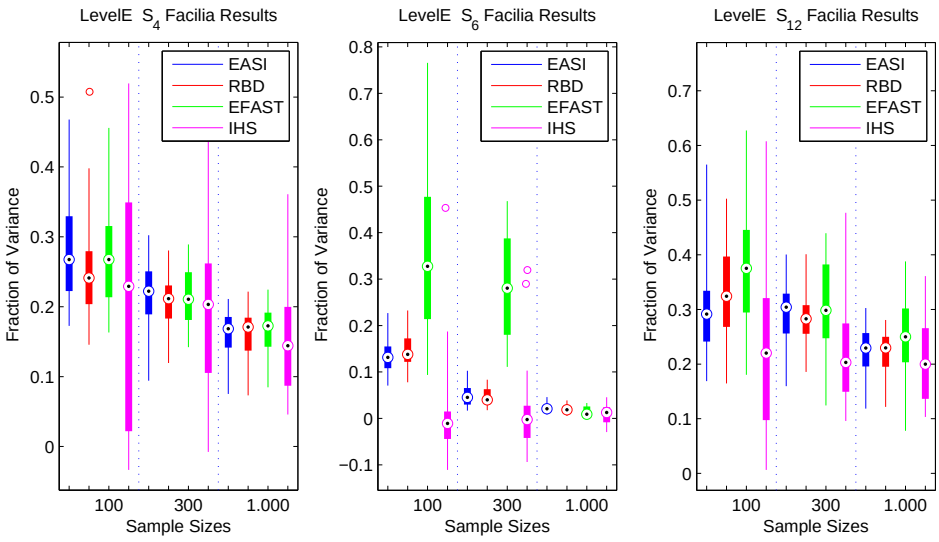

Figure 27: Sensitivity Analysis of the peak dose rate, parameters v_1, R_I^1, and W.

Figure 26 shows the results obtained with cheap methods when analysing the logarithm of the peak dose rate. The ECV method immediately catches one's eye as its bias seems to be minimal compared to the other methods. The added value of a linear fit for the CLM method cannot stand out against the VCE method.

If the data were not log-transformed then the Sensitivity Analysis would attribute 15% of the variance to v^1 and 21% of variance to W (opposed to 37% and 46% shown in Fig. 26). The untransformed data were analysed by Facilia, see Figure 27 for an illustration. Here we see that the methods using special sampling schemes offer no advantage when compared to the cheap EASI method. Sometimes their performance is even worse. Moreover, 1000 realisations are not enough to capture all the effects of the non-linearity in the model. There were small effects visible for the parameter R_I^1 in the log-transformed output data, for the untransformed output data this influence on the output has completely vanished.

Let us now discuss the results obtained from the Sobol' method. The LP_τ sequence was taken from the GNU Scientific Library, http://www.gnu.org/software/gsl/. Figure 28 shows the results for different basic sample sizes ranging from 100 to 4096. The linear connection between the points is deceiving, there are nonlinear effects between the shown sample sizes. The most-influential parameters W and v^1 are identified for small sample sizes. However, to fix a percentage value the maximal basic sample size, 4096, is still too small.

For total effects see Figure 29, the influence of W and v^1 is also detected with a few 100 realisations. In this example the parameter v_2 produces large negative values.

The total effects as computed by Facilia for the parameters v_1, R_I^1, v_2 and W can be found in Figure 30. As for the Sobol' method, we encounter problems with the IHS method. The total effects from the parameter v_1 show a large negative outlier for sample size 300, and from the analysis of the parameters v_2 and W we encounter large negative values. The results of EFAST look more promising: Their variation is small compared to the IHS methods and they seem to

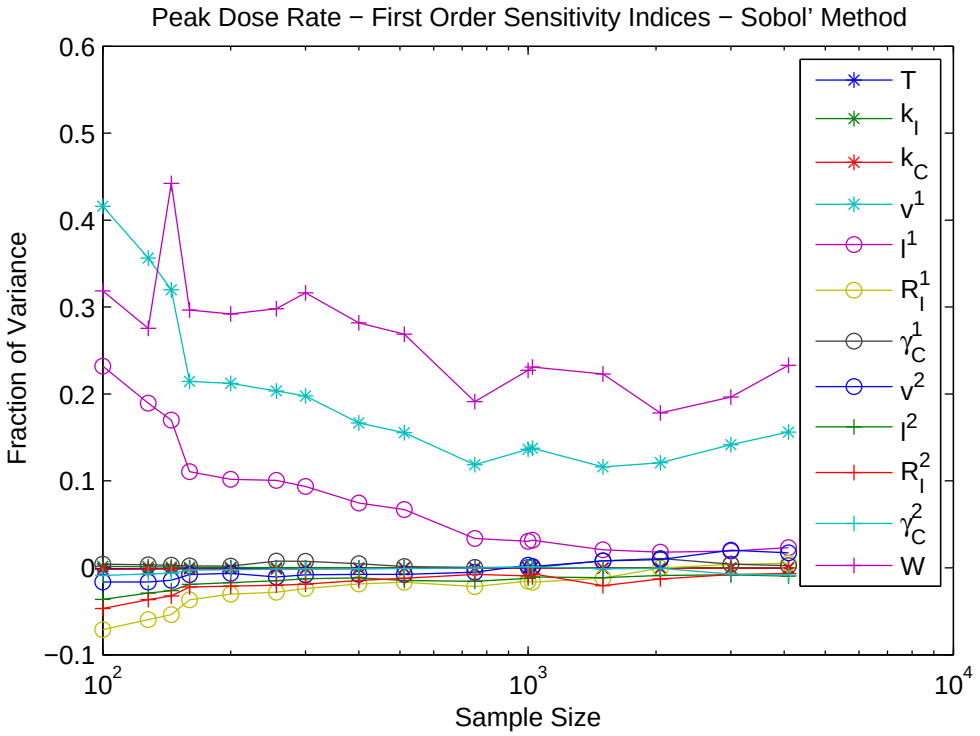

Figure 28: Sobol' method for different sampling sizes, main effects.

converge for v_1,v_2 and W, while the results for R_I^1 show sudden changes between sample sizes 300 and 1000.

6.2 Time of occurrence of the peak

For the determination of the time of occurrence of the maximal dose rate we are confronted with the same problem as above, the data are not directly available in the output. For an analysis we have chosen the time step where the maximal total dose rate is attained. Essentially, this makes the peak time a discrete random variable which picks one out of the 25 specified time-steps.
Figure 31 shows the Sobol' indices of the peak rate depending on the sample size. They look utterly uninformative. Maybe there is a cluster of slightly important variables v^1, γ_C^1, γ_C^2 which makes some sense as the velocity and the retardation multipliers should influence the occurrence of the peak. The results of the cheap methods highlight this impression. Figure 32 shows the results for the above-mentioned parameters, indeed showing a subtle, but noticeable influence. Facilia's results are displayed in Figure 33. It can be seen that S_7 (i.e., the sensitivity of γ_C^1) is positive. The influence of the other two parameters is not so clearly visible for the maximal sample size 1000 as the IHS method has a large variation compared to the other methods which hinders the decision if the Sensitivity Index is non-zero.
The peak time covers orders of magnitude, so that one can suggest a logarithmic transformation of the time scale. The results obtained in this way differ substantially from the untransformed

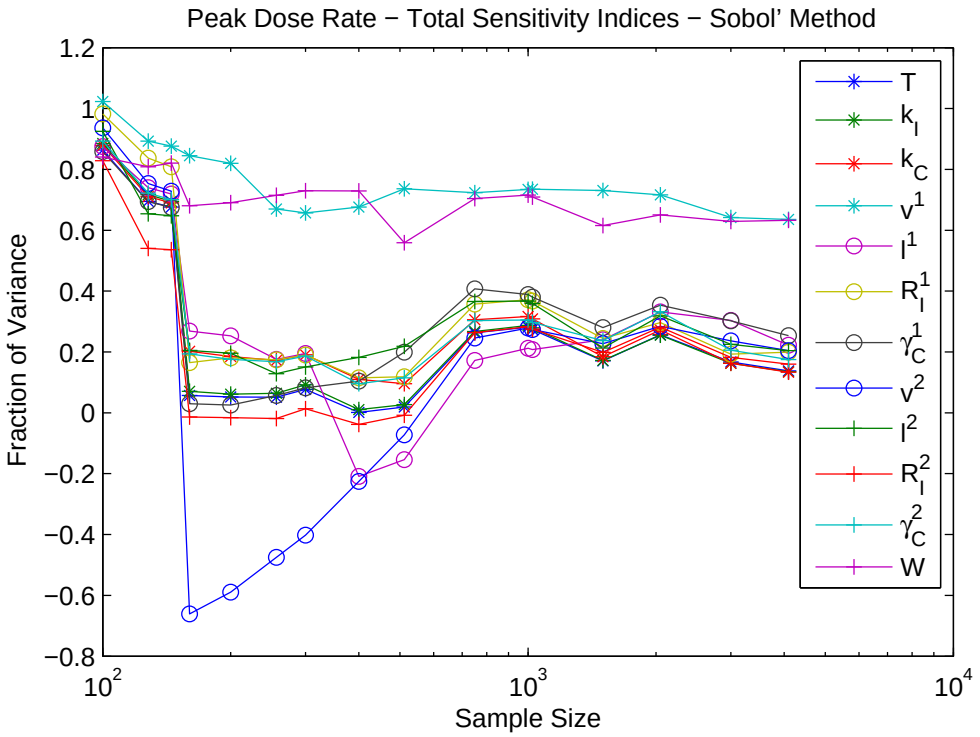

Figure 29: Sobol' method for different sampling sizes, total effects.

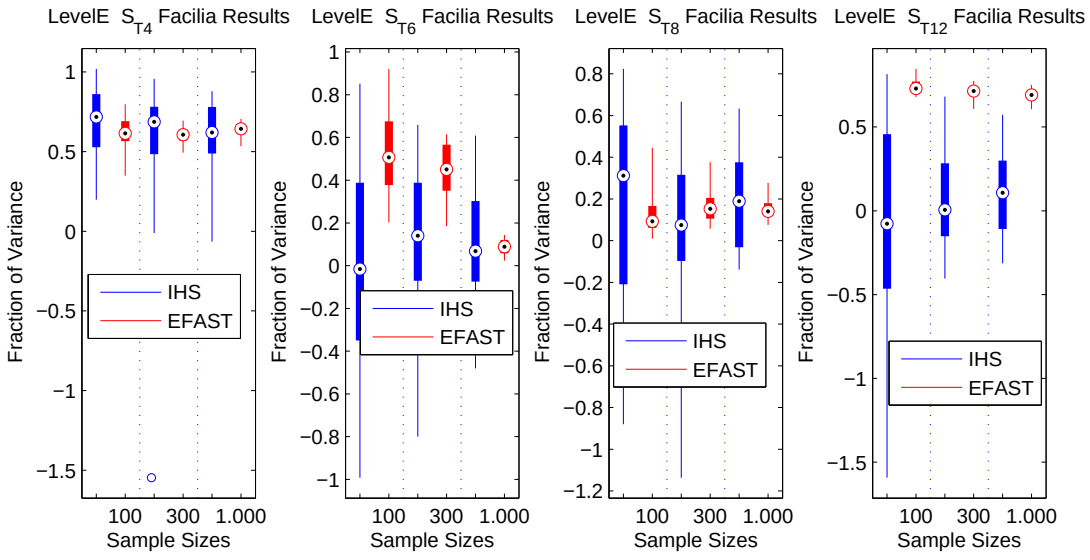

Figure 30: Total effects for the peak dose rate, Facilia's results.

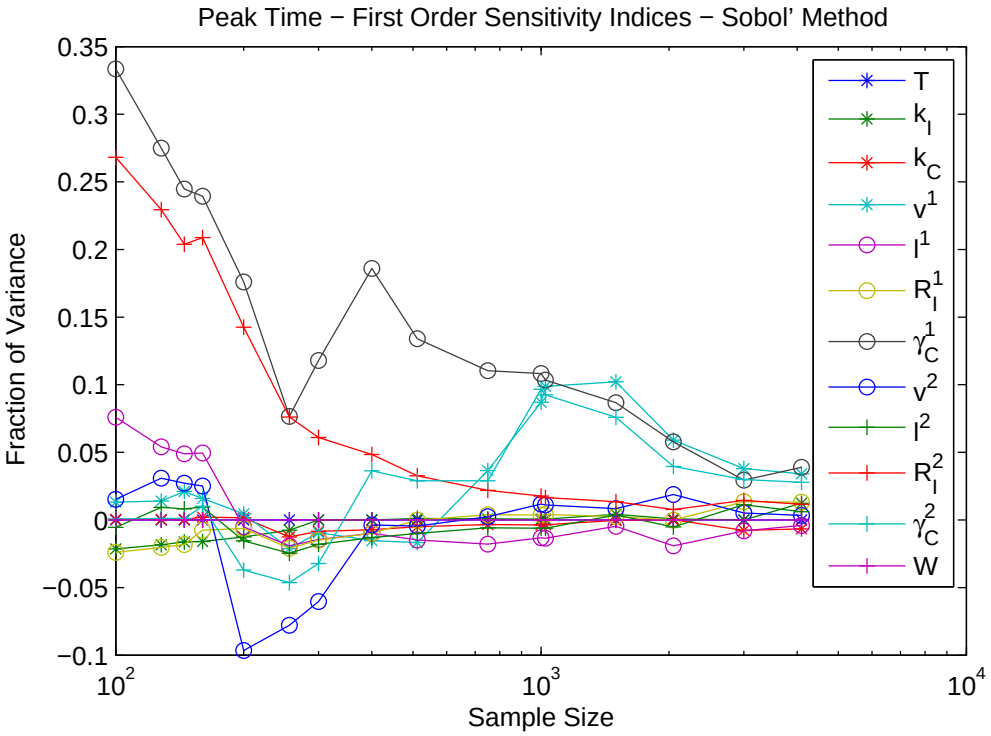

Figure 31: Sobol' method for different sampling sizes, main effects.

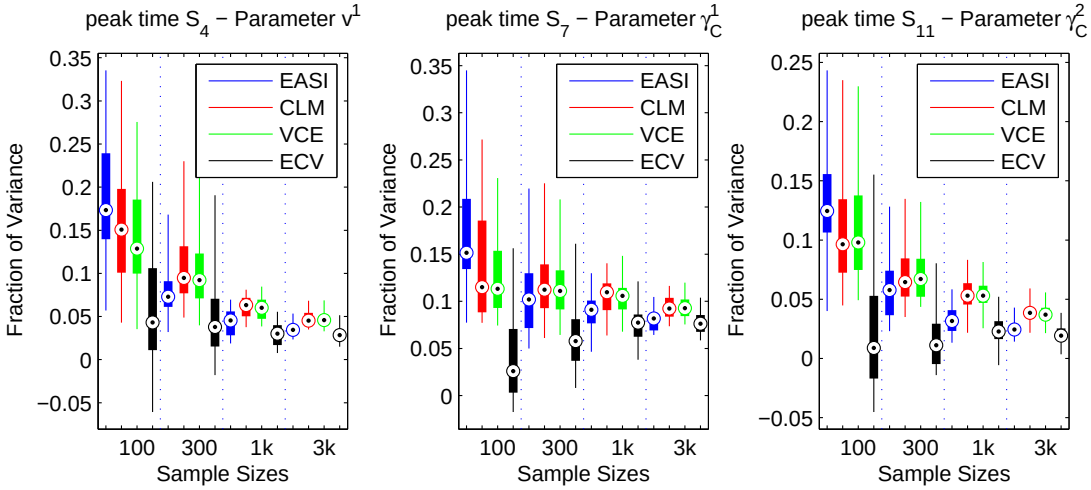

Figure 32: Cheap Sensitivity Analysis of the time of the peak dose rate.

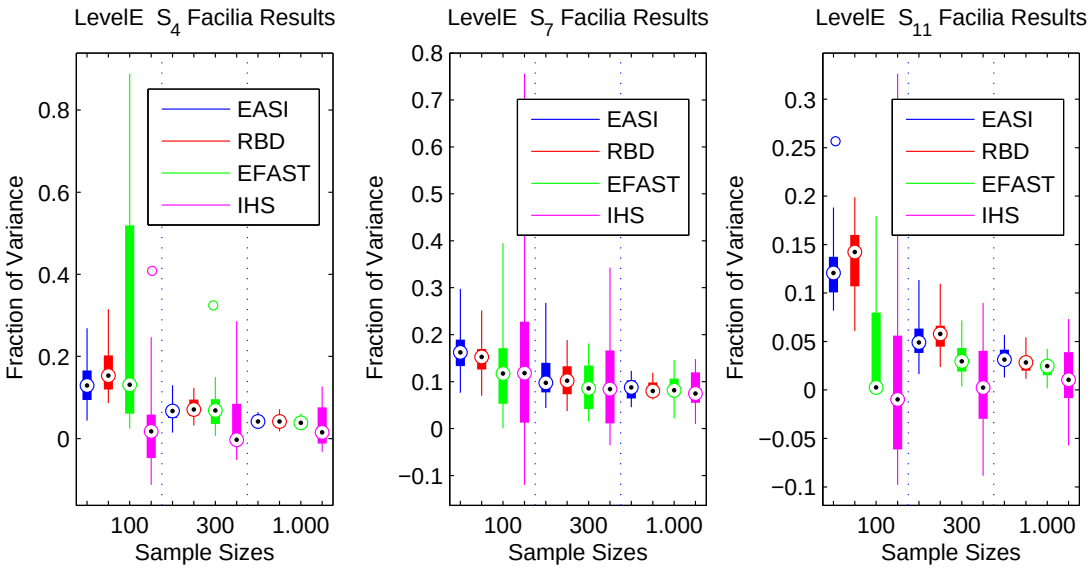

Figure 33: Sensitivity Analysis of the time of the peak dose rate, parameters v_1, γ_C^1, and γ_C^2.

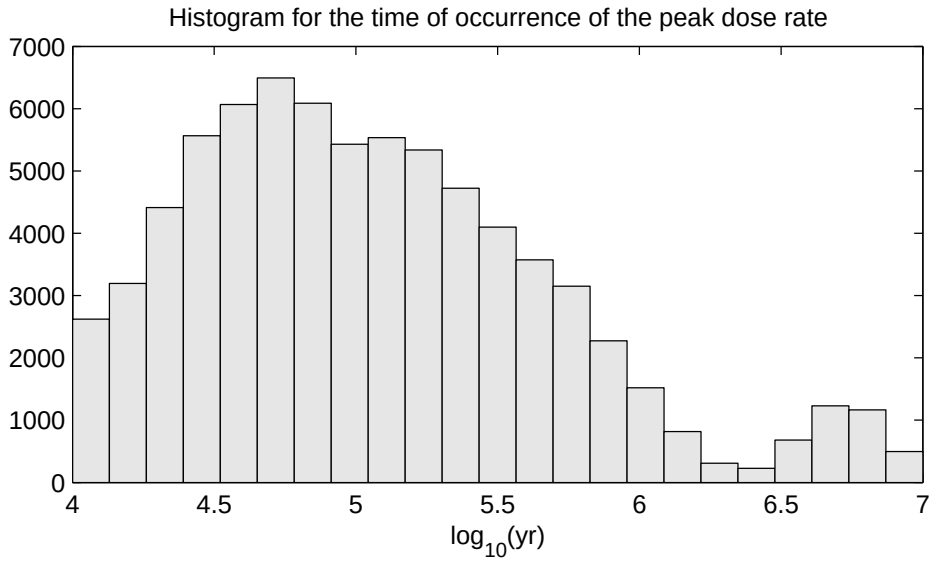

Figure 34: Histogram for the time of the peak dose rate.

results. For this, let us first comment on the histogram for the peak time, see Figure 34. We see two local maxima, one at 50,000 years and the other one at 500,000 years. While the first one is due to the Iodine decay, the second one is due to the Neptunium decay chain.
Without a logarithmic transformation of the time data only the “late” outliers feel the strength of the Sensitivity Analysis, hence the SA qualifies the Neptunium decay chain retardation multipliers which lead to late maxima as influential. With a logarithmic transformation the late maxima move much more closely to the other maxima and more strength in the SA is given to the “early” max-

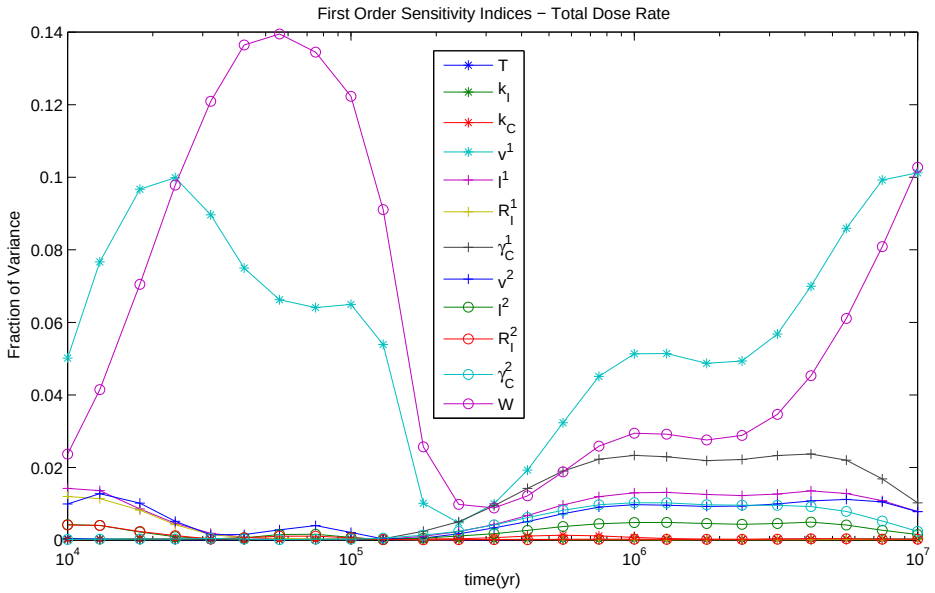

Figure 35: Time-dependent SA of the total dose rate based upon 75,000 realisations.

ima. This is also visible in the results with respect to log-transformed time-data as the following sensitivities are reported for the layer-1-velocity $v^1 \approx 31\%$, the Iodine retardation $R_I^1 \approx 7\%$, and the Np retardation multipliers $\gamma_C^1 \approx 5\%$ and $\gamma_C^2 \approx 2\%$ (without illustration). Hence the analysis puts more emphasis on the "early" Iodine maxima.

6.3 Time-dependent total dose rate

For the peak dose rate we already identified v^1 and W as the most influential parameters. Let us now analyse how these influences change over time.

Figure 35 shows a SA performed with EASI over all of the available 75,000 realisations. For the rest of the section we will concentrate on the four parameters W, v^1, v^2 and γ_C^1 which are most influential. Note that sum of the Sensitivity Indices is smaller than 1, hence there exist parameter interactions which are not be captured by first order effects.

The results from the Sobol' algorithm with a basic sample size of 4096 can be found in Figure 36. Although the total amount of model runs is $4096 \cdot (k+2) = 57,344$, the results still show some negative values, hence the precision of the sensitivity estimates is much worse than for those obtained via the EASI analysis of Figure 35 which uses a just 30% larger sample.

Let us now consider an analysis based upon the 3000 realisations for a cheap method. Figure 37 reports the time-dependent results, showing $\min$ and $\max$ (dotted lines), median (dashed lines) and mean (solid lines) from the 25 available runs for the four parameters of interest using a ECV correlation ratio method with 55 subsamples per partition. The means and the medians are nearly indistinguishable, and the whole analysis looks sound.

Last, but not least we have a look at the results from Facilia. Results for sample sizes of up to

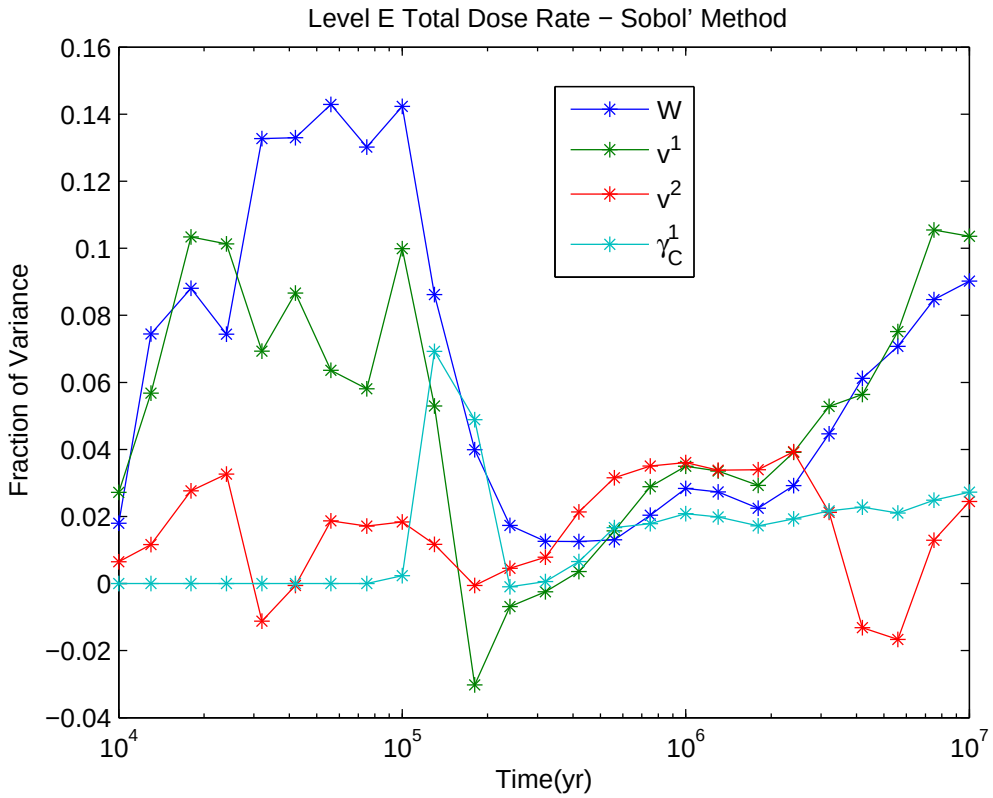

Figure 36: Time-dependent main effects of the total dose rate with the Sobol' method.

1000 realisations are available. We only show the statistics for the Sensitivity Index of parameter v^1 based upon the 25 available runs of samples of 1000 realisations. Figures 38 and 39 show minimum, maximum, mean and median of the estimates obtained with four different methods. The Fourier-based methods more or less deliver the same results and parameter ranges are comparable with those reported in Fig. 37, only IHS performs much worse (remember that IHS uses a basic sample size of 1000, hence the estimate is based upon 14,000 model evaluations). For the total dose rate, we do not discuss the results of the estimation of total effects.
As this benchmark is mainly a test for the Sensitivity Analysis benchmark we do not try to interpret the obtained Sensitivity Indices, and to enlighten the roles of the parameters involved in the model, which would be the next step in a real world analysis. Such an interpretation would allow us to find answers to questions of the type we mentioned in the introduction.

7 Conclusions

A lot of insight into the internals of variance-based Sensitivity Analysis has been gained during the course of this benchmark exercise. We collect and present the lessons learnt in a condensed form.

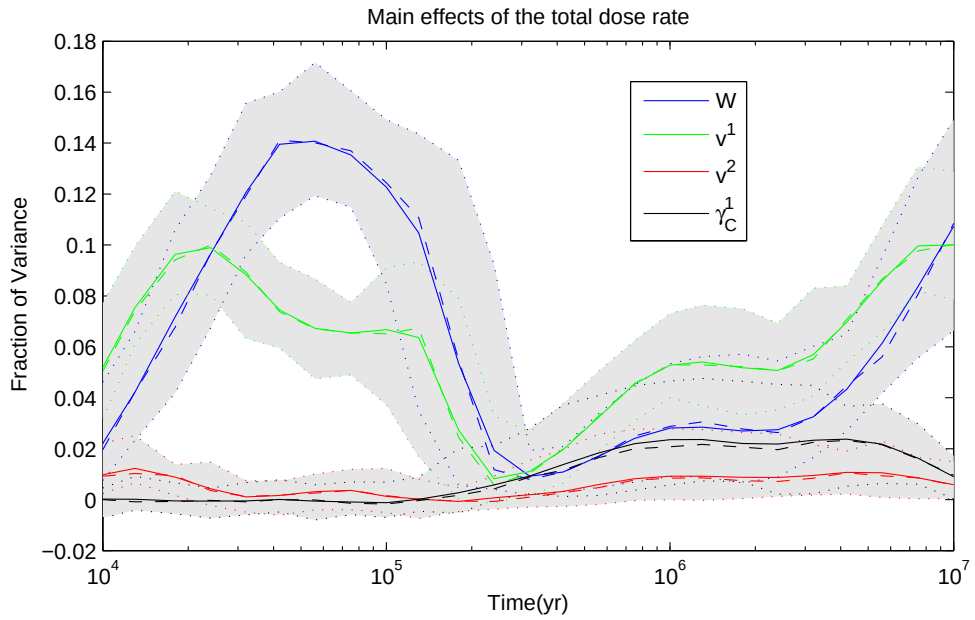

Figure 37: Min, Max, Mean and Median of the main effects for the total dose rate.

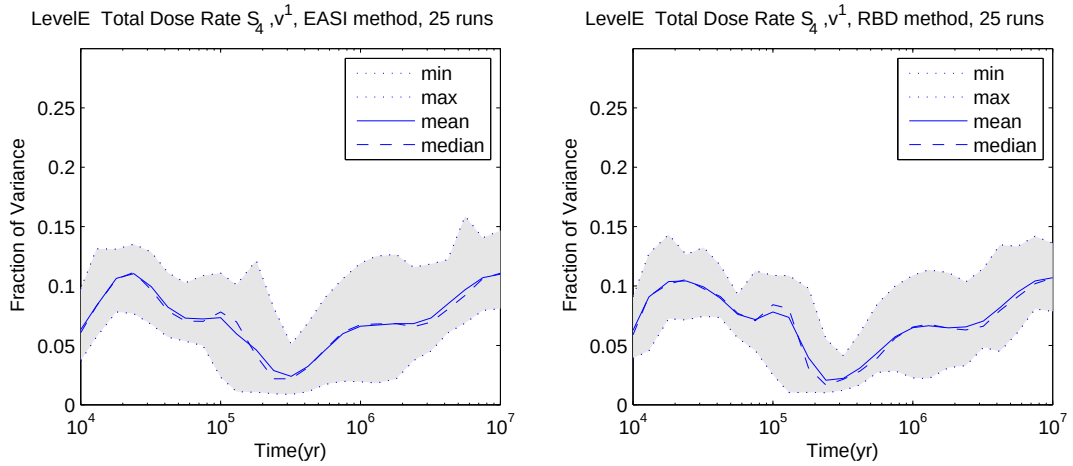

Figure 38: Statistics of the main effects for the total dose rate, EASI and RBD methods.

First of all, we noted that for the standard algorithms the different implementations seem to be very stable and produce results with only subtle differences. Moreover, results obtained with cheap methods are very much comparable to those obtained with more sophisticated methods. However there are some pitfalls which should be kept in mind when performing a variance-based SA.

- Sobol'/IHS without special Monte-Carlo-integration sequence performs worse than a cheap method.
- Sobol' LP_τ without a sample size which is a power of 2 is sub-optimal for small sample

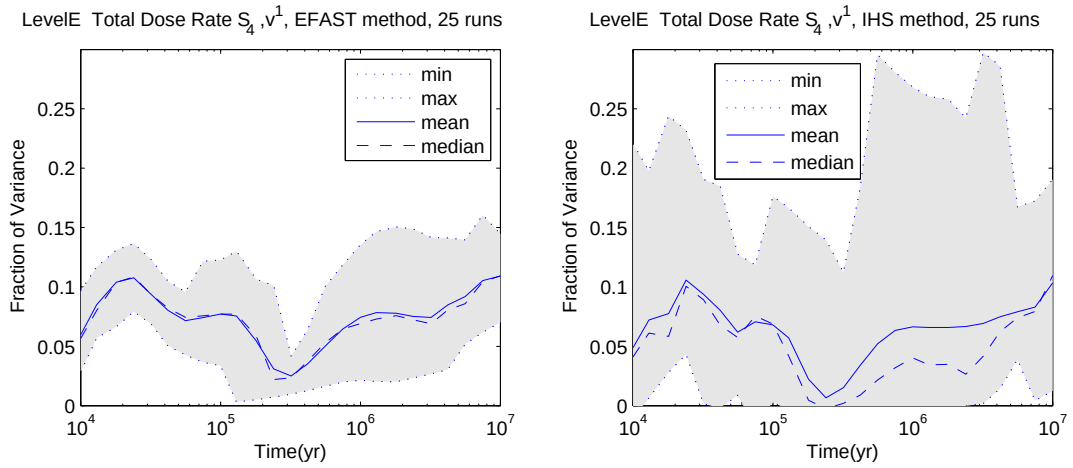

Figure 39: Statistics of the main effects for the total dose rate, EFAST and IHS methods.

sizes.

- For large number of parameters, Sobol' LP_τ needs a large number of realisations.
- Algorithms with fixed maximal harmonic/numbers of subsamples do not capture discontinuities.
- Fourier-based methods and models with periodic output may have unwanted resonances in the frequencies which render results useless. This may happen for EFAST and small sample sizes, i.e., if a simple frequency selection scheme is in use.
- For CR methods, if jump discontinuities are not resolved by the choice of the partition then the results are sub-optimal. Moreover, the influence of the subsample size is not neglectable.
- Random Balance Design shows no advantages when compared with cheap methods.
- For small Sensitivity Indices nearly all methods show bad convergence properties.
- For EFAST, one has the added value of computing total effects. But if a simulation run is already available then a cheap method will provide first order effects with no additional simulation costs.

There are still open problems related to SA and this benchmark exercise.

- Cheap methods can also deal with the estimation of total effects. However, one has to keep the curse of dimensionality in mind when choosing subsample sizes.
- Cheap methods provide consistent results in situations with dependent input data. It is unclear how to interpret these results.

- The good performance of the ECV correlation ratio method (in combination with a rank-based partition) is currently not well understood.
- The effect of $\log$-transforming the output data on the Sensitivity Indices is not studied in detail. It is clear that when taking the logarithm of a product there are parts of the variance which are transferred from higher order effects to main effects.
- These empirically distilled advices are currently not always backed up by theoretical results.

Acknowledgements

The authors thank all the participants of the PAMINA task 2.1.D for providing such a stimulating working environment, for their willingness to run the benchmark tests and to provide the data, and for their ideas and constructive remarks.

References

[1] A. Badea and R. Bolado. Milestone M.2.1.D.4: Review of sensitivity analysis methods and experience. Technical report, PAMINA Project, Sixth Framework Programme, European Commission, 2008. http://www.ip-pamina.eu/downloads/pamina.m2.1.d.4.pdf.

[2] R. Bolado and A. Badea. JRC's contribution to the benchmark based on synthetic PA cases. Technical report, PAMINA Project, Sixth Framework Programme, European Commission, 2008.

[3] R. Cukier, C. Fortuin, K. Shuler, A. Petschek, and J. Schaibly. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. Theory. *J. Chem. Phys.*, 59:3873–3878, 1973.

[4] R. Cukier, J. Schaibly, and K. Shuler. Study of the sensitivity of coupled reaction systems to uncertainties in rate cofficients. III. Analysis of the approximations. *J. Chem. Phys.*, 63:1140–1149, 1975.

[5] P.-A. Ekström. *Eikos – A Simulation Toolbox for Sensitivity Analysis*. http://www.luthagen.org/ekstrom/docs/Eikos_A_Simulation_toolbox_for_Sensitivity_Analysis.pdf, 2005.

[6] B. Krzykacz. *SAMOS: A Computer Program for the Derivation of Empirical Sensitivity Measures of Results from Large Computer Models*. Garching, Germany, 1990. Report GRS-A-1700, Contract No. 73 708, 31 050.

[7] D. Levandowski, R. M. Cooke, and R. J. Duintjer Tebbens. Sample-based estimation of correlation ratio with polynomial approximation. *ACM Transactions on Modeling and Computer Simulation*, 18(1):3:1–3:16, 2007.

[8] Nuclear Energy Agency. PSACOIN level E intercomparison. Technical report, OECD, Paris, 1989.

[9] K. Pearson. *On the General Theory of Skew Correlation and Non-linear Regession*, volume XIV of *Mathematical Contributions to the Theory of Evolution*. Drapers' Company Research Memoirs, Cambridge University Press, Cambridge, UK, 1905.

[10] E. Plischke. An effective algorithm for computing global sensitivity indices (EASI). *Reliability Engineering&System Safety*, 2009. Submitted Manuscript.

[11] E. Plischke and K.-J. Röhlig. Milestone M.2.1.D.3: Plan for benchmark, including specification of synthetic PA cases. Technical report, PAMINA Project, Sixth Framework Programme, European Commission, 2008.

[12] P. Prado-Herrero. *SimLab and GTM1. An External Model Example – The PSACOIN Level E*. http://sensitivity-analysis.jrc.it/tutorial/SimLab%2520and%2520GTM1-TT1.pdf, 2005.

[13] A. Saltelli, K. Chan, and E. Scott. *Sensitivity Analysis*. John Wiley&Sons, Chichester, 2000.

[14] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis – The Primer*. John Wiley&Sons, Chichester, 2008.

[15] A. Saltelli and S. Tarantola. On the relative importance of input factors in mathematical models: Safety assessment for nuclear waste disposal. *J. Am. Stat. Assoc.*, 97(459):702–709, 2002.

[16] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto. *Sensitivity Analysis in Practise – A Guide to Assessing Scientific Models*. John Wiley&Sons, Chichester, 2004.

[17] J. Schaibly and K. Shuler. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. II. Applications. *J. Chem. Phys.*, 59:3879–3888, 1973.

[18] I. Sobol´, S. Tarantola, D. Gatelli, S. Kucherenko, and W. Mauntz. Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering&System Safety*, 92:957–960, 2007.

[19] A. Stuart, K. Ord, and S. Arnold. *Kendall's advanced theory of statistics: Classical inference and the linear model*, volume 2A. John Wiley&Sons, Hoboken, NJ, 2009.

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

A A short UA/SA implementation

In the course of the benchmark exercise, TU Clausthal developed a set of MATLAB scripts for performing UA/SA. This section gives an overview of the available implementations of SA methods. Some design decisions were made to keep the code simple, e.g., Fourier coefficients are computed via `fft()` and not by a direct calculation, partitions are accessed using a `find()`, and code snippets were re-used between the different methods.

Name	Syntax	Description
SUSI	`Si=susi(x,y)`	Main effects from given data (Correlation Ratio)
EASI	`Si=easi(x,y)`	Main effects from given data (Fourier-based)
FITSI	`Si=fitsi(x,y)`	Main effects from given data (polynomial fit)
XFITSI	`STi=xfitsi(x,y)`	Total effects from given data (polynomial fit)
CLMSI	`[Si,STi]=clmsi(x,y)`	Main and total effects from given data (local fit)
IHSSI	`[Si,STi]=ihssi(k,n,m,t)`	Main and total effects (Ishigami-Homma-Saltelli)
SOBOL	`[Si,STi]=sobol(k,n,m,t)`	Main and total effects (Sobol' LP_τ)
JANSEN	`[Si,STi]=jansen(k,n,m,t)`	Main and total effects (Jansen Winding Stairs)
XFAST	`[Si,STi]=xfast(k,n,m,t)`	Main and total effects (FAST)
EFAST	`[Si,STi]=xfast(k,n,m,t)`	Main and total effects (EFAST)
RBD	`Si=rbd(k,n,m,t)`	Main effects (RBD)

For methods using given data the standard syntax is `Si=method(x,y)` where `x` is a matrix of inputs and `y` is an output vector. For methods using a special sampling-scheme the standard syntax is `[Si,STi]=method(k,n,model,trafo)` where `model` is the function under inspection, `trafo` is the transformation from the unit cube to the input distribution, `k` is the number of the model parameters and `n` is the number of the requested basic sample size. The internal program flow is given by `u=quasirand(n,k); x=trafo(u); y=model(x);`. The transformation offers no additional parameters to model different marginal distributions in case of dependent data.
Some of the scripts offer further options which are documented in the online help. The archive containing these scripts is available upon request.

B More benchmark results

In a first series of PA benchmarking we asked the participants of PAMINA 2.1.D for their results on a number of benchmarking examples, see the Milestone report [11]. A diverse range of available algorithms was in use, starting from linear regression over variance-based global Sensitivity Analysis to screening methods and statistical tests for performing Monte-Carlo Filtering.
This appendix gathers the individual contributions of the partners. The following contributions have been received.

- ANDRA (France): L. Loth, G. Pepin

  The results of the analytical and threshold cases have been performed by Andra with the Alliances computing platform. Sensitivity Analysis indicators based on linear regression were calculated.

Table 2: Computational coverage of the first round of benchmark examples

No.	Name	k	Reference	ANDRA	FACILIA	JRC	TUC
1	Linear model	3	[13, §2.9.1: 1]	X	X	X	X
2	Linear model with interactions	2	[13, §2.9.1: 2]		X		X
3	Linear Sobol' function	22	[13, §2.9.1: 3]		X	X	X
4a	Monotonic model	2	[13, §2.9.2: 4(a)]	X	X	X	X
4b	Monotonic model	2	[13, §2.9.2: 4(b)]		X	X	
4c	Monotonic model	2	[13, §2.9.2: 4(c)]	X	X	X	X
5a	Exponential Sobol' function	6	[13, §2.9.2: 5(a)]		X	X	X
5b	Exponential Sobol' function	20	[13, §2.9.2: 5(b)]		X	X	X
6a	Quotient model	2	[13, §2.9.2: 6(a)]	X	X	X	X
6b	Quotient model	2	[13, §2.9.2: 6(b)]	X	X	X	X
7	Sobol' g function	8	[13, §2.9.3: 7]	X	X	X	X
8	** missing **						
9	Ishigami function	3	[13, §2.9.3: 9]		X	X	X
10	Morris function	20	[13, §2.9.3: 10]		X	X	X
11	Bungee jumping man	3	[16, §3.1]	X	X	X	X
12a	Distance of two spheres	6	[16, §3.5]		X	X	X
12b	Distance of two spheres	6	[16, §3.5]		X	X	X
13a	Smooth switch	2	[11]		X	X	X
13b	Smooth switch	2	[11]		X	X	X

- Facilia (Sweden): P.-A. Ekström

  All computational work has been performed with Eikos[5], a simulation toolbox for Sensitivity Analysis written in MATLAB.

- JRC Petten (The Netherlands): A. Badea

  The software in use was R (see http://cran.r-project.org/), a free software environment for statistical computing and graphics. The functions needed for SA where provided by the additional package "sensitivity".

- TUC (Germany): E. Plischke

  Algorithms for UA/SA were developed using MATLAB. As an alternative option, the SimLab 3.0 software (http://sensitivity-analysis.jrc.it/) was to be tested. Unfortunately, major problems were encountered which prevented its use for the benchmarking.

Table 2 lists the examples and their coverage by the participants. The analysis of some of the models has been marked as optional for the participants, hence not necessarily all examples are covered. Table 3 shows the applied UA/SA methods per participant. Some of the methods are only applied to certain models. Furthermore, note that in this first round the cheap methods are not covered.
The benchmarks were used to gain knowledge of operating the UA/SA frameworks and to build confidence in the obtained results.

Table 3: UA/SA methods used by the participants for the first round

Method	ANDRA	FACILIA	JRC	TUC
Mean	X	X	X	X
Variance	X	X	X	X
Skewness, Kurtosis		X		
Quartiles, Min/Max		X		
R^2	X	X	X	X
R^{2*}	X	X	X	
Pearson Correlation Coeff.	X	X	X	X
Spearman Correlation Coeff.	X	X	X	
Partial Correlation Coeff.	X	X	X	X
Partial Rank Correlation Coeff.	X	X	X	
Standard Regression Coeff.		X	X	X
Standard Rank Regression Coeff.		X	X	
Smirnov		X		
Sensitivity Indices (first order)				
FAST		X	X	
IHS		X	X	
EASI				X
Sensitivity Indices (total order)				
FAST/EFAST		X	X	
IHS		X	X	
Morris OAT		X	X	

The descriptions of the individual benchmark results are available in electronic form upon request.