An Effective Algorithm for Computing Global Sensitivity Indices (EASI)

Elmar Plischke[a]

[a] *Institut für Endlagerforschung, Technical University of Clausthal, 38678 Clausthal-Zellerfeld, Germany*

Abstract

We present an algorithm named EASI that estimates first order sensitivity indices from given data using Fast Fourier Transformations. Hence it can be used as a post-processing module for pre-computed model evaluations. Ideas for the estimation of higher order sensitivity indices are also discussed.

Key words: Global Sensitivity Analysis, Sobol' sensitivity indices, FAST Method, Correlation Ratio, Post-Processing Algorithm, Space Filling Curves

1. Introduction

Global sensitivity analysis investigates the relationship between uncertainty in the inputs of a computational model and the uncertainty in the output. So-called variance-based techniques are based on a decomposition of the variance in the model output into components each depending on just one input variable, components each depending on two variables and so forth. Correspondingly, the output variance can be decomposed into contributions each coming from only one input variable ("first order effects"), from just two variables ("second order"), etc. A major drawback of most of the available algorithms for the estimation of this variance decomposition like (extended) Fourier amplitude sensitivity test/(E)FAST[1], random balance design/RBD[2], Ishigami-Saltelli-Homma-method/IHS[3] or the Sobol´ algorithm[4] (see also [5, 6]) is the requirement of special sampling schemes or additional model evaluations so that available data from previous model runs (e.g., from an uncertainty analysis based on SRS or LHS schemes) cannot be reused.

The EASI algorithm is a Fourier-based technique for performing variance-based methods of global sensitivity analysis for the computation of first order effects (a.k.a. Sobol´ indices, main effects, correlation ratios), hence belonging into the same class of algorithms as FAST and RBD. Algorithms of this type are using a frequency-based approach, i.e., signals of known frequencies are assigned to the input factors, and a frequency analysis is carried out on the output that

Email address: elmar.plischke@tu-clausthal.de (Elmar Plischke)

computes the influence of each input factor on the output, see Figure 5 for a demonstration of a frequency response.

EASI is a computationally cheap method for which existing data can be used. Unlike the FAST and RBD methods which use a specially generated sample set that contains suitable frequency data for the input factors, in EASI these frequencies are introduced by sorting and shuffling the available input samples. The sorted input will be a nearly symmetric, periodic signal of frequency 1 (irrespectively of the input distribution). The output data are sorted accordingly matching the resorted input samples, hence avoiding re-evaluation of the model. These sorted data can then be analysed using the power spectrum of the output. This latter analysis forms the standard back-end procedure of the Fourier-based techniques.

For higher-order effects, the sorting algorithm is implemented via a multi-dimensional search curve. The sorted and shuffled input obtained with this method will be a perturbed signal with a certain frequency spectrum. It is hoped that the non-periodic perturbations are distributed over the whole spectrum and therefore are of little influence for the spectral analysis.

2. VARIANCE-BASED SENSITIVITY ANALYSIS

We consider a computational model $y = f(x_1, \ldots, x_k)$ with k (scalar) input parameters x_j and a (scalar) output y. The values of the input parameters are not exactly known. We assume that this uncertainty can be handled by using random variables X_j, $j = 1, \ldots, k$ of known distributions. Then the model output is also a random variable $Y = f(X_1, \ldots, X_k)$.

The first order effect for the input factor j is the fraction of the variance of the output Y which can be attributed to the input X_j. It is defined by

$$S_j = \frac{\mathrm{V}[\mathrm{E}[Y|X_j]]}{\mathrm{V}[Y]} \tag{1}$$

where $\mathrm{E}[Y|X_j]$ is the conditional expectation of Y given X_j and $V[\cdot]$ denotes the variance of a random variable. To estimate the value of S_j we require realisations of the input distributions and the associated model evaluations. We will denote one realisation of a parameter set with $(x_1, \ldots, x_k)$, multiple realisations are shown in matrix notation $\mathbf{X} = (x_{ij})_{i=1,\ldots,n,j=1,\ldots,k}$. The associated output is then denoted by $\mathbf{Y} = f(\mathbf{X})$.

3. EFFECTIVE ALGORITHM FOR SENSITIVITY INDICES

The Extended Fourier Amplitude Sensitivity Test (EFAST) and the Random Balance Design (RBD) algorithms depend on the map

$$G_\omega(s) = {}^1\!/\!_\pi \arccos(\cos(2\pi\omega s)), \quad [0,1] \to [0,1] \tag{2}$$

(or a variant thereof) to produce frequency dependent input data. For $\omega = 1$, this map is a triangle-shaped zig-zag line. For FAST, different values of ω are chosen for different input factors, and with the help of a power spectrum the output is analysed for resonances with respect to these different input frequencies to yield main effects (first order sensitivity indices). The relationship between this frequency approach and the variance-based formulation (1) is established by Parseval's theorem.

RBD uses only $\omega = 1$, but randomly permutes the zig-zag map (2) for different input factors. Undoing the appropriate permutation on the output, main effects can be computed for different factors, again using power spectral methods.

Now, the EASI method can be thought of an "inverse" RBD approach, as a permutation is constructed from given data. In particular, the map $G_\omega(s)$ with $\omega = 1$ can be approximated from existing random data via a straight-forward sorting-and-shuffling procedure as follows. Let us assume that x is a real vector of length n which is the realisation of some distribution X_j on $\mathbb{R}$. In order to keep the notation short the dependency on j is dropped. However, keep in mind that x is a column of the data matrix $\mathbf{X}$.

We order $x = (x_i)$ increasingly to obtain an ordered vector $(x_{(i)})$ with $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. Now, taking all odd indices from $(x_{(i)})$ in increasing order followed by all even indices in decreasing order gives us a vector $(x_{[i]})$ with

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

for which the elements satisfy the following zig-zag relation

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

We therefore call this type of vector triangular-shaped. Figure 1 demonstrates the reshaping process for 50 uniformly distributed realisations.

As the permutation associated with the sorting-and-shuffling is invariant under monotonic transformations of x, other input distributions pose no problems for EASI.

To compute an estimate of the first order sensitivity index S_j for the k-parametric model $y = f(x_1, x_2, \ldots, x_j, \ldots, x_k)$ is now straight-forward. If (x_i) is the n dimensional realisation of the random variable X_j and if $\pi((x_i)) = (x_{[i]})$ denotes the permutation which transforms (x_i) to the triangular-shaped vector $(x_{[i]})$ then we look out for resonances of period $\omega = 1$ and its higher harmonics in the power spectrum of the permuted output $\pi(y)$ using the standard procedure implemented in all Fourier-based sensitivity methods. In particular, if $c_m = \sum_{\kappa=1}^{n} (\pi(y))_\kappa \zeta_n^{(\kappa-1)m}$, $\zeta_n = e^{-\frac{2\pi i}{n}}$, $m = 0, \pm 1, \pm 2, \ldots, \pm [n/2]$ are the complex coefficients of the discrete Fourier transform of $\pi(y)$ then an estimate of the first order sensitivity index is given by

$$\hat{S}_j = \frac{\sum_{m=1}^{M} |c_m|^2 + |c_{-m}|^2}{\sum_{m \neq 0} |c_m|^2} = 2 \frac{\sum_{m=1}^{M} |c_m|^2}{\sum_{m \neq 0} |c_m|^2} \tag{3}$$

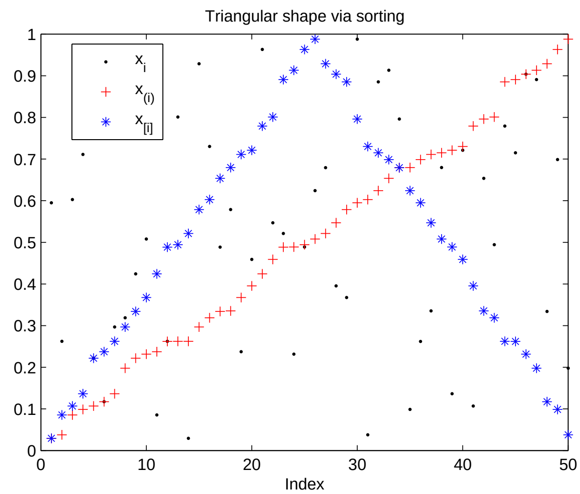


Figure 1: Sorting produces a triangular shape

where the maximum harmonic M is usually 4 or 6. But if the output depends non-continuously on the input parameters then the quadratic convergence properties of the series in (3) are lost and higher values of M are required.

As EASI is a post-processing algorithm, statistical methods of bootstrapping or jackknifing can be applied to test the robustness of the indicators.

4. Partitioning of Hypercubes via Space Filling Curves

If a sorting procedure as presented above is available in higher dimensions then sensitivity indices for index sets may be computed. We will now derive such a higher-dimensional sorting procedure. With the help of a search curve we can assign an address to each sample. Sorting these addresses allows to use an Fourier analysis based algorithm like the one described above in (3). This approach of computing higher-order sensitivity indices is a mixture of a graphical method (partitioning of the data) combined with the standard Fourier Amplitude method.

The search curve is constructed as a finite-length approximation of a space-filling curve. Moreover, changes of direction of this search curve should be detectable via a frequency analysis (such an idea is mentioned in [7]). These two conditions are mirrored in the following address assignment. Let us assume that the index group of interest, I, has dimension $\#I = \ell \leq k$. If we partition each dimension into P intervals then the ℓ-dimensional hypercube has P^ℓ sub-hypercubes which can be enumerated from 0 to $P^\ell - 1$, hence assigning an address to any sample located inside this sub-hypercube. In order to be able to work with arbitrary input distributions, we use the ranks of the input factors for address assignment. As experiment shows, the use of ranks also leads to a uniformly populated hypercube.

We use a "plough-track" curve which continuously connects the sub-hypercubes, see Figure 2 for a demonstration using $P = 5$ partitions and $\ell = 3$ dimensions. Note that with such a curve, the mapping from an ℓ-dimensional coordinate into the natural numbers follows a Gray-code rule [8, 9] (the coordinates of adjacent indices differ only by one in one of the coordinates) and changes of direction occur for multiples of powers of P which can be detected by using a Fourier transformation.

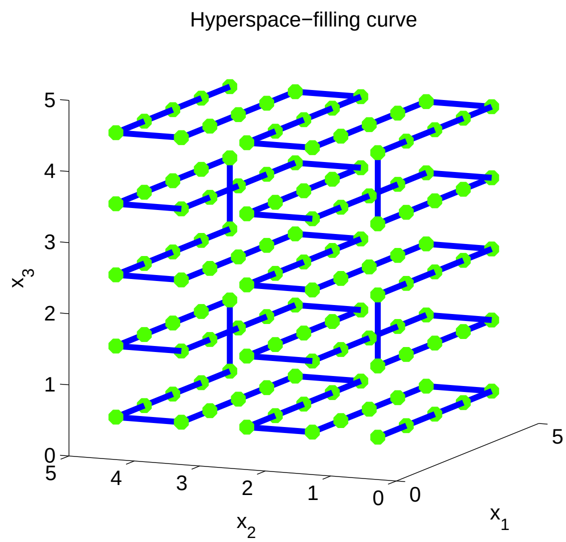


Figure 2: Plough-track curve for a three-dimensional index assignment.

Before starting the index assignment, we replace the data in each dimension by their appropriate ranks (this works for arbitrary distributions and resolves possible problems introduced by clustering). We then scale and round the ranks to fit into $\mathcal{P} = \{0, 1, \ldots, P-1\}$ via the function scale : $r \mapsto \tilde{r} = \lfloor P\frac{2r-1}{2n} \rfloor$. The plough-track curve is implemented by either retaining this value $\tilde{r}$ (forward direction) or changing it to $P - \tilde{r} - 1$ (backward direction) depending on an even-odd rule for the higher dimensions,

$$\text{eo} : \mathcal{P}^\ell \to \mathcal{P}^\ell, p = (p_i) \mapsto q = (q_i) \text{ where } q_i = \begin{cases} p_i, & \prod_{j>i}(-1)^{p_j} = 1, \\ P - p_i - 1, & \text{otherwise.} \end{cases}$$

This even-odd rule is an inverse-Gray-code map. The final hyperindex is then computed by weighting different columns of the data with powers of P (including $P^0 = 1$) and then summing over the rows,

$$\text{padic} : q \mapsto \sum_{i=1}^{\ell} q_i P^{i-1}.$$

Combining the above, a vector-valued coordinate-to-index map is given by

$$\text{hyperindex}: \qquad \mathbb{R}^{n\times\ell} \to \mathbb{R}^n, \quad X \mapsto \text{padic} \circ \text{eo} \circ \text{scale} \circ \text{ranks}(X).$$

As an improvement, the leading dimension with weight $P^0 = 1$ can use positions after the decimal place as this dimension does not appear in the even-odd calculation. Then a sorting algorithm has access to information from within the sub-hypercube.

Figure 3 shows some tests with clustered data: on the left, an index assignment without ranked data, in the middle a curve using ranked data (note that each plough-track uses roughly the same amount of points), and on the right a curve using ranked data, but no sub-hypercube information, i.e., using only data rounded to integer values which leads to some “backwards leaps”. All these index assignments start in the lower left corner and are based upon a 10×10 grid using $P = 10$. Here, the index assignment in the middle figure produces

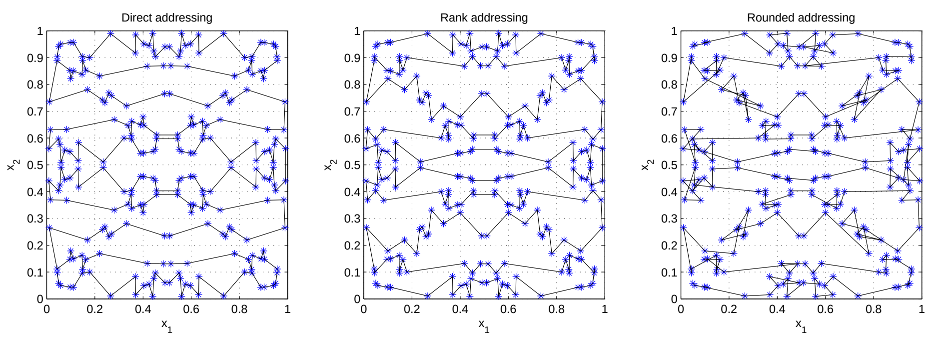


Figure 3: Indexing a two-dimensional data set

the clearest Fourier power spectrum.

5. Higher-Order Effects

With the index-assignment derived above we can estimate the sensitivity index S_I of an index group I for the model $y = f(x^I, x^{\hat{I}})$. Here $x^I \in \mathbb{R}^\ell$ is the realisation of an ℓ dimensional random vector $(X_{j_1}, \ldots, X_{j_\ell})$ with $j_i \in I$ and $x^{\hat{I}} \in \mathbb{R}^{(k-\ell)}$ is the realisation of an $k - \ell$ dimensional random vector collecting the remaining variables X_j, $j \notin I$.

Hence, the multiple realisation matrix $\mathbf{X}$ is split into $\mathbf{X}' = (x_{ij})_{i=1,\ldots,n,\, j\in I} \in \mathbb{R}^{n\times\ell}$ and $\bar{\mathbf{X}} = (x_{ij})_{i=1,\ldots,n,\, j\notin I} \in \mathbb{R}^{n\times(k-\ell)}$. Again, we drop the dependency on I for the ease of notation.

As a first step we assign an index to each ℓ-dimensional row entry of $\mathbf{X}'$ via the plough-track curve introduced in the last section.

When the rows of the data matrix $\mathbf{X}'$ are sorted with respect to this index we obtain the matrix $(\mathbf{X}'_{(i)})$. A triangular reshape produces $(\mathbf{X}'_{[i]})$ again by taking odd rows of $(\mathbf{X}'_{(i)})$ in increasing order followed by even rows in decreasing order. Next, if $\pi(\mathbf{X}') = (\mathbf{X}'_{[i]})$ denotes the associated permutation of the rows then we again look for resonances in the power spectrum of the resorted output $\pi(\mathbf{Y})$. Due to the nature of the plough-track curve, if $I = \{j_1, \ldots, j_\ell\}$ then the first order effects of the input factor j_ℓ are located at the frequency $\omega_\ell = 1$ and its higher harmonics, first order effects of the input factor $j_{\ell-1}$ are located at the frequency $\omega_{\ell-1} = P$ and its higher harmonics etc., so that first order effects of the factor j_1 are located at the frequency $\omega_1 = P^{\ell-1}$ and its higher harmonics. For higher order effects corresponding to the index group I, the superposition principle for trigonometric function shows that the affected frequencies are $\pm\omega_1 \pm\omega_2 \pm\cdots\pm\omega_\ell$. When respecting the higher harmonics of the individual frequencies, the following list contains all the frequencies contributing to the sensitivity index S_I,

$$\Omega_I = \{\pm m_1\omega_1 \pm m_2\omega_2 \pm \cdots \pm m_\ell\omega_\ell, m_i \in \{1, 2, \ldots, M\}\}\,. \tag{4}$$

Finally, we can compute an estimate for S_I analogously to the first order case using the complex coefficients c_m of a discrete Fourier transform of the resorted output vector $\pi(\mathbf{Y})$,

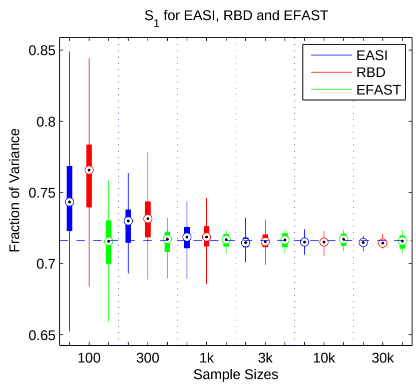
$$\hat{S}_I = \frac{\sum_{m\in\Omega_I} |c_m|^2}{\sum_{m\neq 0} |c_m|^2}, \tag{5}$$

see also (3). The composition of the set Ω_I also shows that the basic frequency P must satisfy $P > 2M$ to avoid overlaps in the higher harmonics. Moreover, the maximum frequency in use, $M(1 + P + \cdots + P^{\ell-1}) = M\frac{P^\ell-1}{P-1}$, must not exceed the Nyquist frequency $[\frac{n}{2}]$ of the Fourier transform. This yields $2M < P < \sqrt[\ell]{n+1}$ as a conservative estimate for the relation between the sample size n, the maximum harmonic M and the basic frequency P. For total effects (here: the accumulated effect of all indexed parameters), the frequency set Ω_I is augmented by frequency components from the subsets of I, $\Omega_{TI} = \bigcup_{J\subset I} \Omega_J$. We then have

$$S_{TI} = \sum_{i\in I} S_i + \sum_{i,j\in I, i<j} S_{ij} + \sum_{i,j,k\in I, i<j<k} S_{ijk} + \ldots \tag{6}$$

If the basic frequency satisfies $P = 2M + 1$ then the set Ω_{TI} contains all frequencies from 1 up to $M(P^{\ell-1}+\cdots+P+1)$ (and their negative terms). Hence, calculation of the total effects requires only the summation over the first few Fourier coefficients of $\pi(y)$. If one is only interested in total effects, and not in individual contributions of higher terms or different factors, then the frequencies may overlap, i.e., $P - M \not> M$.

In contrast to the one-dimensional case, where a monotonic structure is enforced onto the input signal, the multi-dimensional resorting process produces "dirty" step functions so that the overall signal quality deteriorates. A further analysis of the error introduced via resorting is under investigation.



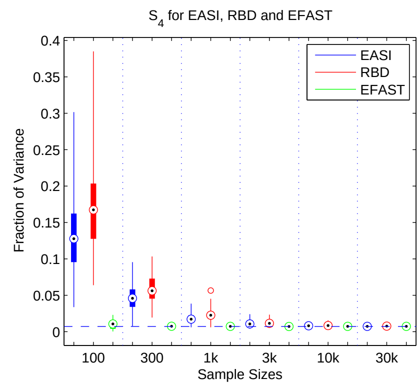


Figure 4: Box plots of First and Fourth Sensitivity indices for the Sobol´ g-function computed with EASI, RBD and EFAST.

6. Examples

We apply the one-dimensional algorithm on a widely used test function. This benchmark function is non-monotonic and has been proposed by Sobol´, see [10] and [5, §2.9.3]. It is given by

$$Y = \prod_{i=1}^{k} \frac{|4X_i - 2| + a_i}{1 + a_i}$$

where $k = 8$, $(a_i) = (0, 1, 4.5, 9, 99, 99, 99, 99)$ and $X_i \sim U(0, 1)$. The analytical values for the first order effects are given by

$$V_i = V(E(Y|X_i)) = \frac{1}{3(1 + a_i)^2}, \quad V = V(Y) = \prod_{i=1}^{8}(V_i{+}1){-}1, \text{ whence } S_i = \frac{V_i}{V}.$$

We tested the EASI implementation against a RBD and an EFAST implementation. We performed 150 runs per sample size $(100, 300, 1000, 3000, 10000, 30000)$. The results are reported in Figure 4. EASI was used with input data from a Latin hypercube sampling algorithm while RBD used random permutations of equidistant samples from the zig-zag function, EFAST uses a sample-size dependent frequency selection scheme. We conclude that all of the methods perform equally well. Moreover, in this example EASI and RBD exhibit the same flaws: For small sample sizes the true value is over-estimated, while for large sample sizes a slight under-estimation occurs. This might be remedied by using bias-correction techniques which are under investigation for EASI, see (7). In contrast, EFAST is able to estimate small sensitivity indices with only a few realisations. The sharp bend in the model at $x_1 = 0.5$ is not resolved by harmonic functions. One has to consider higher harmonics to achieve unbiased results.

We continue with a test of the multi-dimensional algorithm using the so-called Ishigami function from [5, §2.9.3], given by

$$Y = \sin X_1 + 7.0 \sin^2 X_2 + 0.1 X_3^4 \sin X_1$$

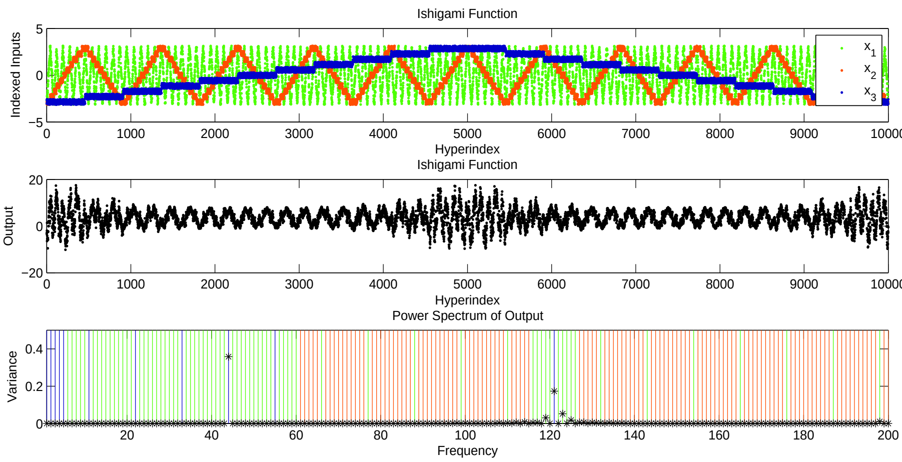


Figure 5: Analysis of Ishigami function.

where $X_i \sim U(-\pi, \pi)$. A short calculation reveals that the expectation conditioned on X_3 satisfies $E(Y|X_3) = E(Y)$, whence the first order sensitivity index of X_3 is $S_3 = V(E(Y|X_3))/V(Y) = E(E(Y|X_3) - E(Y))^2/V(Y) = 0$. Moreover, the first order sensitivity indices S_1 and S_2 account for 76% of the variance. Hence 24% of the variance must be due to higher order effects. We have simulated 10,000 runs using Latin hypercube sampling, and then carried out the multi-dimensional EASI algorithm for $\ell = 3$ dimensions. Figure 5 shows sorted inputs in the upper part. We see that X_3 (blue/black) is associated with frequency $\omega_3 = 1$, while X_2 (red/medium grey) is associated with $\omega_2 = 11$ and X_1 (green/light grey) with $\omega_1 = 11^2 = 121$. Note also the dirty step functions which are produced by the multi-dimensional sorting algorithm. The sorted output is presented in middle part of the plot. It also shows periodic behaviour, however, the frequencies are not directly linked to those in the input. With the reordering of the data an artificial time-scale has been introduced so that a frequency analysis may now be carried out. This is shown in the lower part by providing the power spectrum of the sorted output.

In this frequency plot, we see that X_3 has no main effect (blue/dark background lines) which would have been found in the frequencies $\omega_3 = 1, \ldots, M\omega_1 = 5$, while X_1 ($\omega_1 = 121$) and X_2 show main effects, however for X_2 the effect is not found in the fundamental frequency, but in the fourth harmonic $\omega = 44 = 4\omega_2$, hence the influence of this factor is not visible when using sensitivity measures based on linear regression. We spot second order effects (green/light grey background lines) at frequencies 119 and 123 corresponding to $\omega_1 \pm 2\omega_3$, hence there is an interaction between X_1 and X_3 (or rather X_3^2), but there are no third order effects visible (red/medium grey background lines).

With all $k = 3$ factors listed in the index set $I = \{1, 2, 3\}$ all effects in the output should be ascribed to a specific input factor or an interaction of input factors. However, the total effect of I is estimated by $\hat{S}_{TI} = 0.91$ in

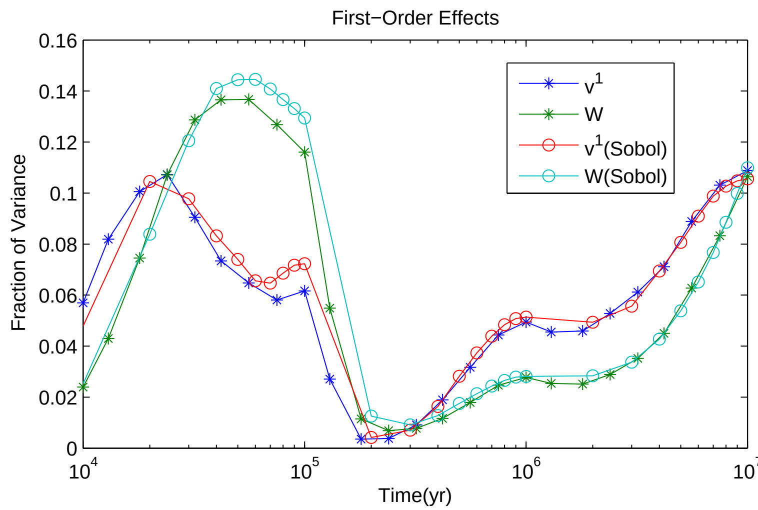


Figure 6: Time-dependent sensitivity indices for Level E example.

contrast to theoretically $S_{TI} = 1.0$. Thus, the absolute values derived from the frequency decomposition of $\pi(y)$ are not reliable. Moreover, it also shows that the sample size can also be too large: From the 10,000 samples available, only $M(1+P+P^2) = 665$ Fourier coefficients are used for spectral analysis, all others contribute to sporadic errors which only form part of the overall variance, but not part of input-factor related effects. Here, also the need for bias-correction techniques is apparent.

Let us now apply the EASI method to a more realistic test case. In various publications (see [2], and [11] for a review), the PSACOIN Level E code [12] was used both as a benchmark of Monte Carlo simulations and as a benchmark for sensitivity analysis methods. This computational model predicts the radiological dose 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 uncertainties are either uniformly or log-uniformly distributed. The parameters of the distributions have been selected on the basis of expert judgement.

A sample containing $2^{15} = 32768$ realisations[1] of the $k = 12$ dimensional input distribution has been generated using Latin hypercube sampling. The time-dependent output is analysed using the EASI method for first-order effects. Results for the most influential factors v^1, the water velocity in the first geosphere layer, and W, the stream flow rate, can be found in Figure 6. These values are plotted against the asymptotic results taken from [2] which were ob-

[1]Evaluation of all the model runs took less then three hours.

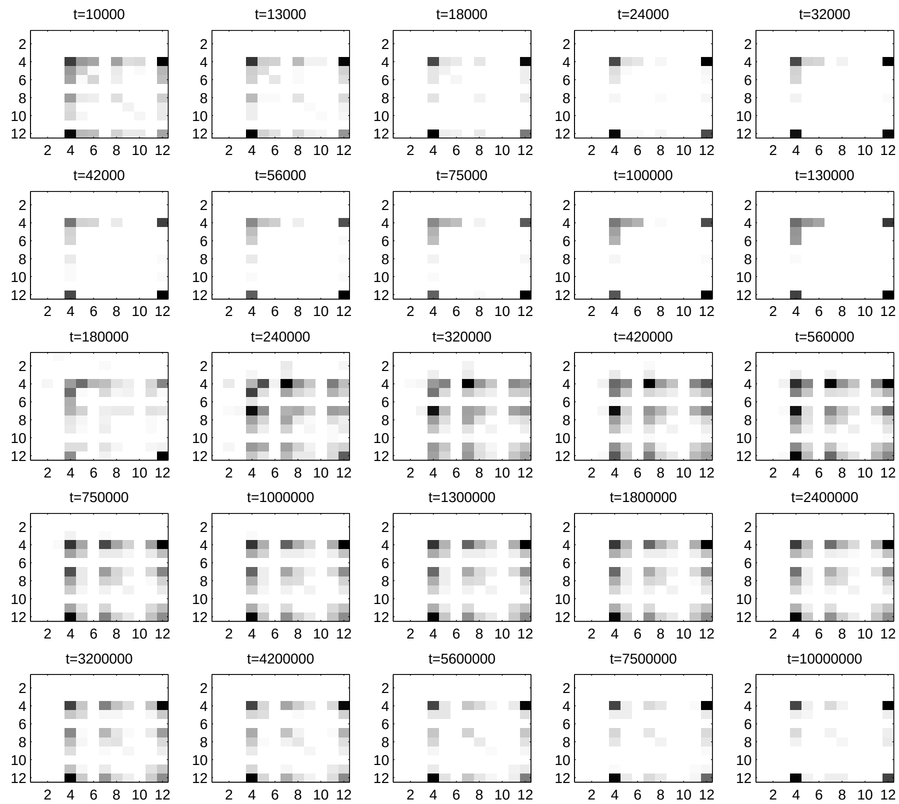


Figure 7: Normalised first and second order effects for Level E example.

tained from a Sobol' algorithm. As a rule of thumb the estimation error is of order $\frac{1}{\sqrt{n}} \approx 0.6\%$. Hence the results show a good agreement despite the use of different time discretisations.

The analysis of first and second order effects is shown in Figure 7. For each timestep, a matrix containing first order effects as diagonal entries and second order effects in the off-diagonal entries is shown. This matrix is normalised so that the maximal value is plotted in black. We see that parameters 4 (v^1) and 12 (W) are of importance, here v^1 interacts with several other parameters. For some time steps, the interaction between v^1 and W is the dominating influence on the dose output. Moreover, for early times parameters 7 and 11 (Neptunium chain retardation factors) have no noticeable influence while for later times parameters 6 and 10 (Iodine retardation factors) are of no influence. Note that the Iodine decay dominates the dose at early times, while decay products from the Neptunium chain are responsible for the dose at later times which perfectly agrees with these results. The source terms in parameters 1 to 3 are of very little influence.

For further examples on the application of variance-based sensitivity indices methods including EASI see [13].

7. Conclusions and Outlook

The proposed algorithm connects variance-based sensitivity analysis ideas with Monte-Carlo-based ones. The examples suggest that the performance of EASI is on-par with established methods like RBD. Moreover, the accuracy achieved by analysing given data is enough for practical purposes. The algorithm for first order effects is easily implemented using a `sort()` and an `fft()` algorithm. Unfortunately, the estimation of total effects for models with large numbers of parameters is still out of reach for the EASI method.

As EASI is a post-processing algorithm, there are connections to the statistical tool of the correlation ratio introduced in [14] which estimates (1) by using piecewise constant approximations of $\mathrm{E}[Y|X_j]$. For correlation ratios many theoretical results are available which may also be used for improvements of the EASI method. As an example, one quickly notices that the estimates produced by EASI are biased with respect to the maximal number of harmonics M. This problem has also been spotted in the theory of correlation ratios. In [15] an unbiased estimator was derived which is easily adapted to our set-up,

$$\hat{S}_j' = \tfrac{1}{n-2M}\left(n\hat{S}_j - 2M\right), \tag{7}$$

as we are using $2M$ out of n Fourier coefficients. For higher order effects, the calculation of the degrees of freedom is also straight-forward.

The idea of sorting and shuffling data might be of use for signal processing applications which require periodic data, e.g., wavelet decompositions using periodic boundaries or cyclic moving averages.

8. Acknowledgements

This work is one of the results of the PAMINA project (Performance Assessment Methodologies in Application to Guide the Development of the Safety Case) which is part of the 6th framework programme of the European Commission. The author thanks all members of Workpackage 2.1.D of the PAMINA programme for fruitful discussions and Stefano Tarantola, JRC STATINT, for providing Level-E data.

References

[1] A. Saltelli, S. Tarantola, K. Chan, A quantitative, model independent method for global sensitivity analysis of model output, Technometrics 41 (1999) 39–56.

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

[3] A. Saltelli, Making best use of model evaluations to compute sensitivity indices, Computer Physics Communication 145 (2002) 280–297.

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

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

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

[7] M. R. Schroeder, Number Theory in Science and Communication. With Applications in Cryptography, Physics, Digital Information, Computing, and Self-Similarity, 5th Edition, Springer, Berlin, 2009.

[8] D. E. Knuth, The Art of Computer Programming. Vol. 4, Fasc. 2: Generating All Tuples and Permutations, Addison-Wesley, Upper Saddle River, NJ, 2005.

[9] D. E. Knuth, The Art of Computer Programming. Vol. 4, Fasc. 3: Generating All Combinations and Partitions, Addison-Wesley, Upper Saddle River, NJ, 2005.

[10] D. Gatelli, S. Kucherenko, M. Ratto, S. Tarantola, Calculating first-order sensitivity measures: A benchmark of some recent methodologies, Reliability Engineering&System Safety 94 (2009) 1212–1219.

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

[12] Nuclear Energy Agency, PSACOIN level E intercomparison, Tech. rep., OECD, Paris (1989).

[13] K.-J. Röhlig, E. Plischke, R. Bolado Lavín, D.-A. Becker, P.-A. Ekström, S. Hotzel, Lessons learnt from studies on sensitivity analysis techniques in the EU project PAMINA: A benchmark study, in: Reliability, Risk and Safety. Theory and Applications. Proceedings of the ESREL 2009 Annual Conference, Vol. 3, CRC Press, Boca Raton, 2009, 1769–1775.

[14] K. Pearson, On the General Theory of Skew Correlation and Non-linear Regression, Vol. XIV of Mathematical Contributions to the Theory of Evolution, Drapers' Company Research Memoirs, Dulau & Co., London, 1905, reprinted in: Early Statistical Papers, Cambridge University Press, Cambridge, UK, 1948.

[15] T. L. Kelley, An unbiased correlation ratio measure, Proc. Natl. Acad. Sci. USA 21 (9) (1935) 554–559.