



EUROPEAN
COMMISSION

Community research

PAMINA

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

(Contract Number: **FP6-036404**)



EVALUATION OF APPROACHES TO SENSITIVITY ANALYSIS DELIVERABLE (D-N°:2.1.D.1)

Author(s):

**D.-A. Becker, S. Spießl (GRS)
K.-J. Röhlig, E. Plischke (TU Clausthal)
R. Bolado-Lavin, A. Badea (JRC)
J. L. Cormenzana (Enresa)
T. J. Schröder, J. Hart (NRG)
R. Avila, P. A. Ekström, R. Broed (Facilia)**

Date of issue of this report : **11/14/09**

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 **I**N **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>.

Executive Summary

Probabilistic sensitivity analysis is the general name for a number of mathematical methods that have been developed for analysing the sensitivity of models against their different input parameters under consideration of the possible interactions by means of statistical procedures. A higher number of model evaluations with individually distributed input parameter values is performed and the results are evaluated using appropriate methods. In the field of long-term safety assessment for final repositories, probabilistic sensitivity analysis is a valuable tool for identifying those parameters that have a high influence on the model results and therefore have to be measured or determined with special care or can be used for optimising the system.

The methods for probabilistic sensitivity analysis can be grouped according to their mathematical nature. The simplest methods are based on correlation or regression analysis and have been applied for many years, but their applicability to nonlinear systems is limited. More sophisticated methods, e.g. based on variance, are available, but their suitability for models with the specific behaviour of repository systems has not been investigated in detail. This is the goal of the report at hand.

In its first part, the report gives an overview of the mathematical basics of known sensitivity analysis methods. The second part describes a benchmark exercise, which was undertaken with a number of analytical models in order to investigate the behaviour of different sensitivity analysis methods under known mathematical conditions. The third part describes the findings of several application exercises with different methods applied to realistic repository systems. All types of host formations were taken into account.

Table of Contents

1. Introduction.....	7
2. Sensitivity Analysis methods review	9
2.1 Screening methods	9
2.2 Global methods	11
2.2.1 Graphic methods	12
2.2.2 Monte Carlo based methods.....	17
2.2.3 Variance based methods	23
3. Sensitivity Analysis benchmark, analytical cases.....	31
3.1 The analytical benchmark cases	31
3.1.1 Ishigami function: A model with three input parameters and higher order effects.....	31
3.1.2 A discontinuous switch example	39
3.1.3 A linear model with dependent input data	44
3.1.4 Sobol' g function: A model with eight input parameters.....	47
3.2 The PA case example: The GTM Level-E model.....	50
3.2.1 Peak of total dose rate.....	51
3.2.2 Time of occurrence of the peak	55
3.2.3 Time-dependent total dose rate	58
3.3 Conclusions	61
4. Application of Sensitivity Analysis methods to realistic repository PA models	64
4.1 NRG application case: rock salt repository	65
4.1.1 Description of the system and the model	65
4.1.2 Objectives and peculiarities of the study	68
4.1.3 Varied input parameters and analysed output parameters	68
4.1.4 SA Methods applied and main results.....	70
4.1.5 Conclusions	77
4.2 NRG application case: repository design in argillaceous rock.....	78
4.2.1 Description of the system and the model	78
4.2.2 Considered scenario.....	79
4.2.3 Modelled geometry	79
4.2.4 Objectives of the study/variability of input parameters	80
4.2.5 Output parameters analysed	82
4.2.6 Statistical methods	82
4.2.7 Results and discussions	83
4.2.8 Conclusions.....	88
4.3 Enresa application case	89
4.3.1 Description of the system	89

4.3.2	System evolution and modelling	90
4.3.3	Sensitivity analysis performed	93
4.3.4	Main results	94
4.3.5	Conclusions and recommendations	102
4.4	GRS application case: generic rock salt repository.....	105
4.4.1	Background	105
4.4.2	Test case.....	106
4.4.3	Sensitivity Analysis	108
4.4.4	Summary	116
4.5	Facilia application case: Biosphere Assessment	116
4.5.1	Description of the studied system	117
4.5.2	Sensitivity analyses performed	117
4.5.3	Inputs and outputs analysed	118
4.5.4	Main results	119
4.5.5	Summary	126
4.6	JRC application case: French clay site.....	126
4.6.1	Conceptual model.....	126
4.6.2	Input uncertain parameters and outputs analysed	128
4.6.3	Study objectives	130
4.6.4	Main results	130
4.6.5	Conclusions of the JRC application case (French clay site).....	136
4.7	Lessons learnt.....	137
5.	Summary	140
6.	References	143



1. Introduction

For the long-term safety assessment of geological repositories for high-level radioactive wastes numerical calculations have to be performed, which are subject to a variety of uncertainties. Proper handling of uncertainties is a very important issue in this field. While, on one hand, underestimation of uncertainties might lead to misleading or wrong conclusions about the long-term safety of the repository, on the other hand an overestimation could reduce the significance of the safety assessment or, in extreme cases, make it completely worthless. Therefore, PAMINA RTDC 2 was dedicated to the management of uncertainties.

Uncertainties can be due to random-dependent intrinsic properties of the system, such as the lifetime of containers. If the system to be assessed is once defined, such uncertainties cannot be reduced, or their reduction lies outside the responsibility of the safety assessor. They have to be accepted and taken into account in the assessment. The more important type of uncertainties, however, is due to lack of knowledge about the system. These uncertainties are called epistemic. They can be reduced by putting effort in measuring or research programmes, but such programmes can be expensive and time-consuming. Uncertainties can also be avoided or mitigated by appropriate design measures. Therefore, there is a big interest – not only on the side of safety analysts, but also on the side of repository developers and implementers, decision makers, and other stakeholders in identifying the uncertainties that are really relevant for the system under consideration.

The work described in the report at hand addresses the subject of sensitivity analysis. In this context this means exclusively the analysis of the sensitivity of the output of numerical performance assessment (PA) models against parameter uncertainties. Other types of uncertainties, such as scenario or model uncertainties, which are also important for PA, can be handled by mapping them to parameter uncertainties.

One has to distinguish between local and global sensitivity analysis (SA). In a local SA study the majority of parameters are kept constant while one or a few ones are varied in order to analyse the reaction of the model to a specific influence. In contrast, in a global SA all uncertain parameters are varied together and the model response is analysed by adequate methods, separating the influences of the individual parameters. This allows analysing the interactions between the parameters.

Furthermore, one distinguishes between deterministic and probabilistic SA. In a deterministic study some selected values of each parameter to be analysed are fed into the model. Such an analysis is mostly performed as a local SA and typically requires only a low number of runs. For a probabilistic SA, however, a big number of model runs is done with a set of parameter values that is statistically drawn according to their given probability distribution functions. Normally, this is performed as a global SA.

While local and deterministic SA exercises are very helpful for the modeller to investigate the behaviour of the model and to improve understanding of the system, probabilistic sensitivity

analysis is increasingly considered as an essential part of the safety case, since it allows identifying the most sensitive parameters and important interactions and thus contributing to model and/or system understanding and focussing further research and development on reducing or avoiding uncertainties. Therefore, the interest of performance assessors in probabilistic sensitivity methods and tools has been rising for several years. The work described in this report exclusively deals with probabilistic SA.

A variety of different methods for probabilistic sensitivity analysis are available. Some of them are well established and have been applied for years in the field of repository PA. In view of the increasing importance of probabilistic SA, however, it became questionable whether they are really adequate for all PA situations and yield reliable results. The PAMINA work package reported here was launched in order to improve the understanding of the functioning of different types of methods, especially when applied to typical PA systems, and to create a basis for future research and application.

The goals of the work package were:

- a review of available probabilistic SA methods,
- testing of different methods with simple mathematical systems,
- application of a variety of methods to realistic repository PA systems.

A compilation of different SA methods with their specific properties is given in chapter 2.

For testing the different methods a benchmark exercise was set up with a number of simple and well-defined mathematical models that are amenable to exact analysis of their sensitivity to parameter variations. By applying probabilistic SA methods to such systems and comparing the results to the theoretical ones one can assess their accuracy, robustness and reliability. As a more realistic approach to the subject of final repository PA a simple standardised repository model, the Level-E system, was included in the benchmark test, though it cannot be analytically analysed. The benchmark exercise is described in chapter 3.

As the last step different probabilistic SA methods were applied to realistic repository models, based on the different concepts of the participating organisations. Such models sometimes show specific behavioural patterns (interaction of inputs, nonlinearity, non-monotony, threshold behaviour) that are not covered by the simple mathematical models and can lead to unexpected results. Specific advantages and disadvantages of the different methods in conjunction with performance assessment of typical repository systems were identified. This work is described in chapter 4.

2. Sensitivity Analysis methods review

A PA model typically involves several hundred input parameters, an important fraction of which are uncertain, thus a joint probability density function is needed to characterise their uncertainty and the possible dependence structure among them. If all inputs are independent, the individual (marginal) probability density functions (pdfs) are enough to characterise such uncertainty. The use of the Monte Carlo method allows mapping the input space onto the output variable space and estimate consequences.

In addition to characterising as accurately as possible the consequences associated to a repository, which is the target of an uncertainty analysis, identifying the most relevant input parameters, whatever this means, is a key task in a PA. A real problem arises when we ask for 'relevant' or 'important' input parameters: the interpretation of these words; what means 'relevant', 'important' and similar words? An input parameter can be considered important with respect to a given output variable if a strong correlation exists between both (linear relation), but it could also be considered so if the output takes remarkably high values when the input takes values in a given region, or if that input contributes a large fraction of the output variance (considered as a measure of uncertainty).

Another issue that arises in the SA area is the study of interactions. Two input parameters interact when the joint effect of both is different from the addition of their individual effects (interactions of order 2). This concept is naturally extended to any number of input factors. In general, main effects (individual effects of each input parameter) are more important than second order interactions, second order interactions are more important than third order interactions and so on, though this is not always true. Interactions deserve to be studied in order to know the true structure of the system model under study. Not all SA techniques are able to study interactions and in some cases, though they are able, the study could be impractical due to different reasons (extreme computational cost, too large diversity of possibilities, etc).

The existence of different interpretations of importance has triggered the development, over the last twenty-five to thirty years, of a variety of SA methods designed to study the model from different points of view, each one developed according to each given interpretation. Nowadays a large corpus is available to the SA practitioner, who may choose appropriate methods to perform a specific type of SA attending to his/her interests and needs. PAMINA pays attention to screening methods and, in particular, to global methods, which are further described in the next sections. Further insights about all methods described in this section may be found in two reports generated within PAMINA; see [Badea and Bolado 2008] and [Plischke et al. 2009].

2.1 Screening methods

The target of screening methods is to identify at a low computational cost (hopefully) non-relevant input parameters. This allows model simplification and all the benefits that this pro-

vides (conceptually simpler models, less computational cost, less input parameter characterisation efforts, etc.). Screening methods focus on the functional relation between inputs and outputs disregarding input parameter distributions, but paying attention to the ranges of each input.

The origin of screening methods for computer codes is in the area of Design of Experiments, which is the branch of Statistics dedicated to study the structure of systems (natural, industrial, commercial, etc.) whose performance depend on a set of controlled factors and another set of not known factors. The latter introduce variability (noise) in the response; see [Box and Draper 1987]. The oldest and most simple method consists in fixing one point (typically a middle point, though not necessarily) and modifying one by one each input, recording the results. This method is quite inefficient when the number of important factors is small and is completely useless to detect interactions, which needs additionally the simultaneous variation of the involved input parameters, unless something else is done, as for example in Morris' method. This method is shortly described in this section.

Typically, several levels may be defined for each input, those levels are combined in all possible ways and the output is computed for each possible crossing (factorial design). This strategy is completely useless for models with many input parameters due to the huge sample size needed. A possibility is to reduce the number of levels to two (low and high level respectively) per input parameter. This strategy (full two-level factorial design) is too expensive (2^k runs needed) even when the number (k) of input parameters is moderate (1048576 runs needed to analyse a 20-input parameter model). A strategy to reduce the computational cost, paying the price of not being able to estimate some interactions or even main effects, is to take only a convenient fraction of the full design (fractional factorial design or 2^{k-p} design). Saturated designs allow to study a number of input parameters with number of runs = number of parameters+1, but they do not exist for any number of input parameters. Supersaturated designs need fewer runs than the number of parameters. Among them the most used one is the Sequential Bifurcation method, which is a group-screening method that will be shortly described in this section.

Morris Factorial Sampling method, see [Morris 1991], is becoming more and more popular among SA practitioners as a screening tool. This is a One-At-a-Time design (OAT). It consists in taking a number of levels per input factor and a number of trajectories randomly generated. Trajectories start at different points chosen at random and are built by successively selecting at random one of the inputs and moving it to one of its possible next levels. These trajectories are used to estimate the mean value and the standard deviation of each main effect. A high estimated mean main effect indicates that the input parameter is important; a high estimated standard deviation indicates important interactions of that input parameter. Figure 1 shows the results of applying Morris' screening method to a model with 20 input parameters. Input parameters 8, 9 and 10 have an important impact on the output but do not participate in important interactions, while parameters 1 to 7 are important by themselves and also because of their interactions. The rest of the parameters are irrelevant.

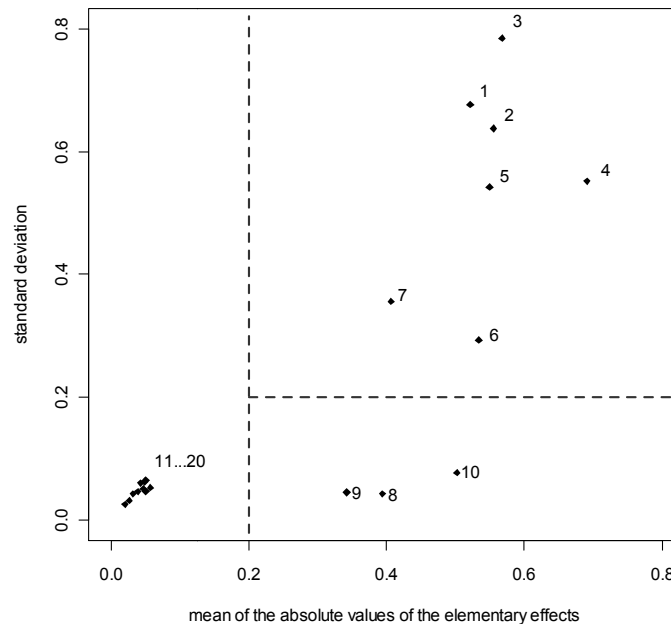


Figure 1 Result of applying Morris' method to a model with 20 input parameters.

Sequential bifurcation has been developed in [Bettonvil and Kleijnen 1996]. It allows screening a large number of factors. One of the main advantages is that the number of simulations to be performed is smaller than the number of factors to be screened. The drawback is that the relationship between each input factor and the output has to be monotonic and that the type of monotony has to be known a priori. Whenever this condition is fulfilled, the method is very effective.

2.2 Global methods

Global methods pay attention to how the whole input space maps onto the output space, taking into account the input distributions. We may classify global methods as Monte Carlo based methods, variance based methods and graphical methods. Monte Carlo based methods do not need a specific sampling plan to be applied, a normal sample used to propagate uncertainties in a PA obtained via Simple Random Sampling (SRS) or Latin Hypercube Sampling (LHS) may be used to compute the corresponding sensitivity indices. Variance based methods study the contribution of each input parameter and its related interactions to the output variance. Variance based methods usually need a specific sampling strategy, with the exception of the simplest method (correlation ratios - CR) and the recently developed methods Random Balanced Design (RBD) and EASI. Graphical methods provide complementary visual information that helps understanding the meaning of numerical sensitivity indices and the global structure of the system model.

2.2.1 Graphic methods

Three types of graphic methods are of paramount importance in a PA: scatter-plots, cobweb plots and contribution to the sample mean plots (CSM plots). These methods are described in the next sections.

2.2.1.1 Scatter-plots

Let us call $\mathbf{X}=(X_1, X_2, \dots, X_k)$ the vector of input parameters and Y to a given scalar output variable. For a given input X_i , the scatter-plot is the projection of the sample points (\mathbf{X}, Y) on the (X_i, Y) plane. This representation allows the examination of the dependence between Y and X_i . Scatter-plots are very helpful to identify linear relations, monotonic relations and the existence of thresholds among other potential trends. The use of transformations may also provide a lot of information about input/output relations. They may be used as supporting material to explain the results obtained by means of numeric sensitivity techniques, but also to prevent the use of inadequate techniques. The most frequently used transformations are: 1) Logarithmic transformation (base 10), and 2) Rank transformation. Both transformations are useful, though they also have shortcomings. The logarithmic transformation is more intuitive, but may not be used when a fraction of the values (usually output values) take value 0, unless the software at hand has been designed to cope with such problem and it avoids null values before applying the transformation. The transformation into the ranks is not affected by such problem. In this transformation the smallest sample value is transformed into 1, the second smallest into 2 and so on until the largest value, which is transformed into n (sample size). So, the new scale is between 1 and n . Nevertheless, this transformation is not so intuitive (two values separated by several orders of magnitude can get contiguous ranks, say h and $h+1$, the same as two other very close values). Figure 2 shows the relation between an output variable and an input parameter. It is difficult to guess the actual nature of the relation between both from this plot. The logarithmic transformation in both axes uncovers the linear relation between the logarithms of both (Figure 3). Figure 4 shows the relation between the ranks, which is also linear.

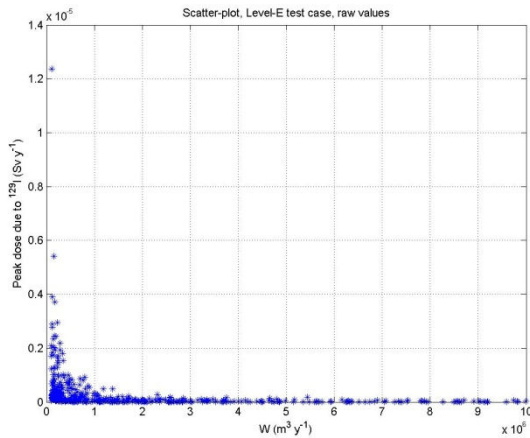


Figure 2 Output variable vs. Input parameter (linear scale in both axes).

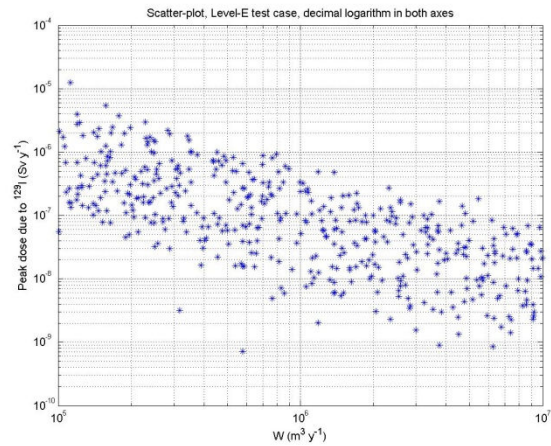


Figure 3 Output variable vs. Input parameter (log. scale in both axes).

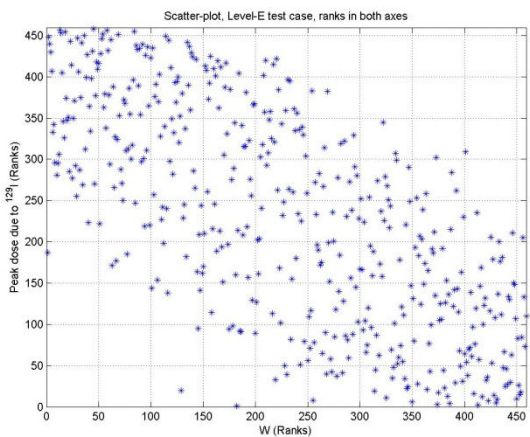


Figure 4 Output variable vs. input parameter (rank scale in both axes).

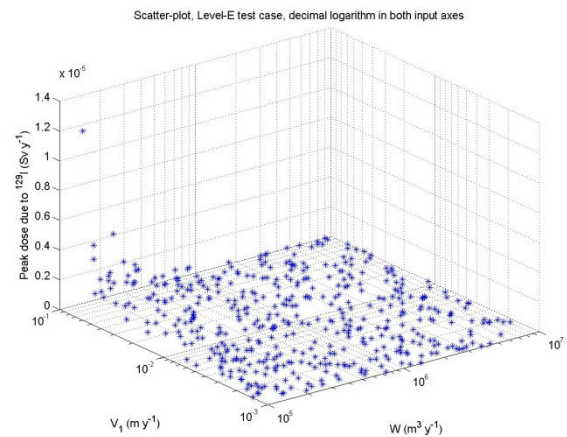


Figure 5 3-D scatter-plot that allows identifying an interaction.

Three-dimensional (3-D) scatter-plots or XYZ plots show the projection of the sample points (X,Y) on the (X_i,X_j,Y) space. The information they are able to provide is also valuable. The extraction of such information is limited, though challenging, due to obvious interpretation problems when a 3-D figure is shown on a 2-D display. Software packages that allow changing the angle of the view may enhance and broaden their applicability. As in the case of normal 2-D scatter-plots, the use of convenient scale transformations may help identifying interesting model features. Figure 5 shows the relation between an output variable and two input parameters. The convenient use of a logarithmic scale in the two input axes and of a linear scale in the output helps identifying the strong interaction between both input parameters with respect to that output (pay special attention to the left corner).

2.2.1.2 Cobweb plots

Cobweb plots have been designed to show multidimensional samples in a two-dimensional graph, see [Cooke and Van Noortwijk 1999]. Vertical parallel lines separated by equal distances are used to represent the sampled values of a given number of inputs/outputs, usually not more than ten or twelve, in order to keep the plot sufficiently clear. Each vertical line is used for a different input/output and either the raw values or the ranks may be represented (either raw values or ranks in all lines, never mixed). Sampled values are marked in each vertical line and jagged lines connect the values corresponding to the same run. Coloured lines can be used to display the different regions of any input parameter or output variable. Moreover, flexible conditioning capabilities enable an extensive insight into particular regions of the mapping. The cobweb plots are usually provided with 'cross densities' showing the density of line crossings midway between the vertical axes. Therefore, an informed and careful analysis of cobweb plots enables the characterisation of dependence and conditional dependence.

Figure 6 and Figure 7 are cobweb plots that show the relation between 9 input parameters and an output variable. All these cobweb plots are based on the ranks; no one is based on the raw values. Figure 6 is an unconditional cobweb plot, where coloured lines have been used to distinguish between runs related to the lower (yellow), lower-medium (green), upper-medium (blue) and upper (black) quartile regions of the output variable. The output variable is represented in the last column; the rest of the columns are dedicated to each input parameter according to the legend behind them. Paying attention to what colours predominate in each region of each input parameter, the reader may get an idea about the complex relation between inputs and output. Nevertheless, the use of conditional cobweb plots is usually more convenient. In this type of plots only a fraction of the runs obtained are represented. Figure 7 shows the runs that produced the 10% highest output values. These doses are always obtained when *W* takes values below its own median and *V1* takes values above its own median (with only one exception). For the rest of the input parameters, such a clear relation cannot be observed, in fact, almost any region of the other input parameters may produce large values of the output.

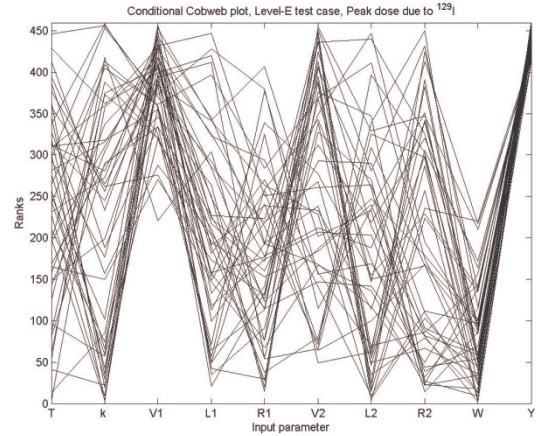
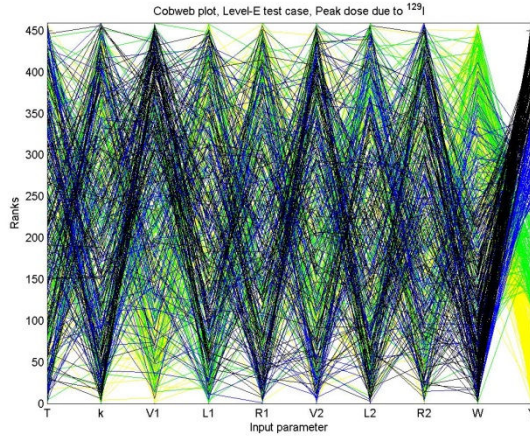


Figure 6 Unconditional cobweb plot based on the ranks. Figure 7 Conditional cobweb plot based on the ranks.

2.2.1.3 Contribution to the sample mean plots (CSM plots)

In order to create a CSM plot, a random sample \mathbf{S} of size n of the input factors and the corresponding sample of the output variable considered must be available. The random sample must be obtained via a sampling technique that introduces no bias. Suitable sampling schemes may be, for example, simple random sampling, LHS and proportional stratified sampling, while non-proportional stratified sampling (i.e. optimal stratified sampling) or importance sampling would not be acceptable schemes. To build the CSM plot for a given input variable, let us say X_i , and the response Y , the following procedure is applied:

- the realisations of X_i are sorted generating the series of values $\{x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)}\}$,
- the corresponding series of values $\{y^{[1]}, y^{[2]}, \dots, y^{[n]}\}$ is created,
- the ancillary variable M_i is defined, whose sampled values are obtained from the sampled values of X_i and Y as

$$m_i^q = \frac{1}{n} \sum_{j=1}^q y^{[j]}, \quad q = 1, 2, \dots, n, \quad (1)$$

- M_i is normalised dividing the values m_i^q by the sample mean of Y ,
- The sampled values of M_i are plotted versus the cumulative relative frequency of X_i .

Using the approach described previously, the estimates of the following quantity are represented on the y-axis

$$\frac{\int_{\Omega_{\mathbf{x}(-i)}} d\mathbf{x}_{(-i)} \int_{x_i \leq x_i^q} f_{\mathbf{x}}(\mathbf{x}) y(\mathbf{x}) dx_i}{\int_{\Omega_{\mathbf{x}}} f_{\mathbf{x}}(\mathbf{x}) y(\mathbf{x}) d\mathbf{x}} \quad (2)$$

which is the fraction of the output mean corresponding to values of X_i smaller or equal to its quantile of order q . The subindex $(-i)$ indicates the exclusion of the input parameter X_i . In expression (2) $\Omega_{\mathbf{x}}$, $\Omega_{\mathbf{x}(-i)}$ and $x_i \leq x_i^q$ mean that the respective integrals are computed on the whole input space, on the whole input space excluding input parameter X_i and on X_i up to its quantile of order q . Given the definition of the plot in step 5, each point in a CSM plot represents the fraction of the output mean due to any given fraction of smallest values of an input parameter.

Figure 8 shows the CSM plot for one output variable and four input parameters of a simulation model based on a simple random sample of 100 realisations. In principle, an input parameter could be considered as non-important if knowing its value doesn't provide much information about the value of the output, which means that the value of the output depends more on the values of other input parameters. In this case, a completely non-important variable, in relation with a given output variable would be characterised by the fact that the values of both would be randomly associated. High, low and intermediate output values would be equally obtained in any region of a non-important variable, which means that a behaviour similar to the ones of X_1 and X_4 in Figure 8 could be expected (lines close to the diagonal).

Input parameters could be considered as important if knowing its value provides relevant information about the value of the output. This means that concentrations of low, intermediate and high values of the output could be expected to be associated to different regions of the input parameter. This is the type of behaviour observed in input parameters X_2 and X_3 in Figure 8, where clear departures from the diagonal can be observed.

Roughly speaking, each region of equal probability of X_1 and X_4 are responsible for the same percentage of the output sample mean (the lines are always close to the diagonal). Regarding X_2 , its 60% smallest values are responsible for more than 99% of the output sample mean. Only values of X_3 between its median and its quantile of order 0.85 produce output values significantly different from 0. This region is responsible for more than 99% of the output sample mean. Other similar statements could be made, as for example: values between the quantiles of order 0.70 and 0.80 of X_3 produce more than 55% out the output sample mean.

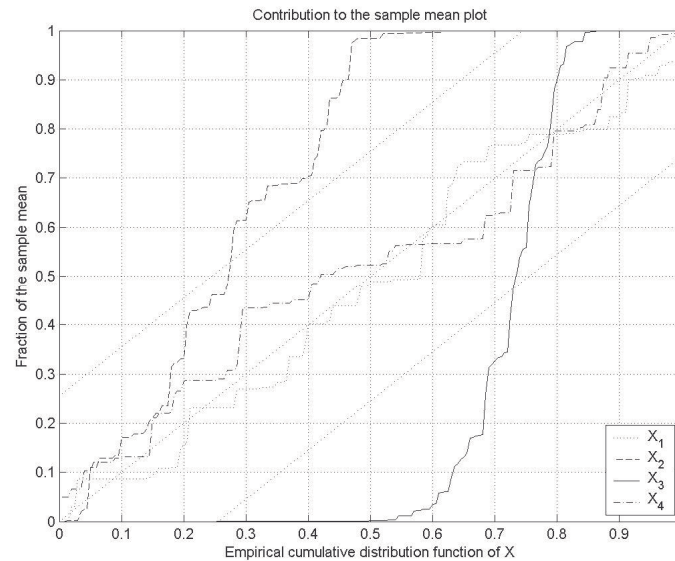


Figure 8 CSM plot for one output variable and four input parameters

[Bolado et al. 2009] have proposed a test to study when the deviations from the diagonal observed in a CSM plot are statistically significant and when they may be obtained due to randomness. The band represented in Figure 8 is the implementation of the test: Lines that keep all their points within the band correspond to non-important parameters, while lines that have at least one point outside the band correspond to important parameters.

A CSM plot represents indirectly a contribution to the variance. Indeed, if for a given quantile range all realisations of Y are very close to the mean, this implies that there is a very low contribution to the variance. Any significantly low and significantly high contribution to the mean represents locally an important contribution to the variance.

2.2.2 Monte Carlo based methods

The Monte Carlo method consists in sampling at random the vector of input parameters, running the system model computer code for each sample of that vector and getting a sample of the vector of output variables. Later on, the input and output samples may be used to perform SA. The main techniques based on Monte Carlo samples are correlation/regression and Monte Carlo Filtering (MCF) based techniques.

2.2.2.1 Correlation based statistics

The most commonly used analysis is the one based on the computation of the correlation coefficient. The Pearson correlation coefficient, or just correlation coefficient, provides a measure of the strength of the linear relationship between any input factor X_j and the output Y . It is defined as:

$$CC(x_j, y) = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(y_i - \bar{y})}{\left[\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \right]^{1/2} \left[\sum_{i=1}^n (y_i - \bar{y})^2 \right]^{1/2}} \quad (3)$$

where $\bar{x}_j = \sum_{i=1}^n x_{ij} / n$ and $\bar{y} = \sum_{i=1}^n y_i / n$ are the sample means. The values of this coefficient are between -1 and +1. A positive value means that both X_j and Y are increasing or decreasing together while a negative value means that Y increases when X_j decreases and vice versa. An absolute value of $CC(x_j, y)$ close to one corresponds to a linear relationship, while an absolute value close to zero corresponds to no linear relationship between X_j and Y .

Whenever a nonlinear but monotonic relationship between X_j and Y exists, a rank transformation can be used to get a linear relationship. Spearman correlation coefficient (denoted by RCC - rank correlation coefficient) is computed using the same expression (3) as the Pearson correlation coefficient except that the ranks of each variable are used instead of the raw values.

2.2.2.2 Regression based statistics

In the class of linear models, the simplest ones are the first order polynomials:

$$y(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon \quad (4)$$

The coefficients $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$ are computed using the sample, such that the error ε is minimized in the least squares sense. With the matrix notations the model (4) becomes:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \quad (5)$$

where

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{bmatrix}, \quad \mathbf{y} = (y_1, \dots, y_n)^T \text{ is the sampled output and } \varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T \text{ is the}$$

vector of random perturbations.

The least squares estimator of β is

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (6)$$

and the fitted regression model is

$$\hat{y} = X\hat{\beta}. \quad (7)$$

An important feature of the linear regression is the decomposition of the total sum of squares $SS_T = \sum_{i=1}^n (y_i - \bar{y})^2$ into a sum of squares due to the regression $SS_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ and a sum of squares due to the error $SS_E = \sum_{i=1}^n (\hat{y}_i - y_i)^2$:

$$SS_T = SS_R + SS_E. \quad (8)$$

The ratio

$$R^2 = \frac{SS_R}{SS_T} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

measures the extent to which the regression model matches the data and is called the coefficient of multiple determination. A value of R^2 close to 1 indicates that the regression model is accounting for most of the uncertainty in Y , while a value of R^2 close to 0 indicates that the regression model does not explain the uncertainty in Y .

When the input factors are independent, the value of R^2 can be decomposed as $R^2 = R_1^2 + R_2^2 + \dots + R_k^2$, where each individual R_j^2 represents the contribution of x_j to the R^2 value. R_j^2 is computed as the R^2 value for the model where the only input factor is x_j .

The equation (7) describes a hyper plane in the k -dimensional space of the input variables. The parameter $\hat{\beta}_0$ represents the intercept of the hyper plane, while each of the parameters $\hat{\beta}_j$ represents the expected change in the output Y per unit change in X_j when all the remaining independent variables are constant. With this interpretation, the parameters $\hat{\beta}_j$ are candidates for sensitivity indicators except that each one is influenced by the units in which the corresponding x_j is expressed, and that it does not give any information on the distribution of X_j . To fix this problem, the regression model (4) is replaced by a model where all the variables have been standardized:

$$y \leftarrow \frac{y - \bar{y}}{s}, \quad x_j \leftarrow \frac{x_j - \bar{x}_j}{s_j}, \quad j = 1, \dots, k, \quad (10)$$

s, s_j being the sample standard deviations for y and x_j .

The regression coefficients of this new model are called Standardized Regression Coefficients (SRC). They provide a sensitivity measure, or more precisely a measure of variables importance:

- an input parameter X_j is more important than another input parameter X_i if the absolute values of their standardized regression coefficients satisfy the condition $|SRC_j| > |SRC_i|$;
- the sign of SRC_j indicates whether X_j and Y are either positively or negatively correlated.

Moreover if the input factors X_j are independent, the inclusion or exclusion of one factor has no effect on the estimation of the $SRCs$ of the other factors in the model. If the input factors are not independent, the information provided by the $SRCs$ is not reliable (for the variable importance).

The relationship between the standardized regression coefficients SRC_j and the regression coefficients $\hat{\beta}_j$ is

$$SRC_j = \frac{s_j}{s} \hat{\beta}_j \quad (11)$$

Sometimes the correlation between X_j and Y may be due to a third variable. In order to avoid that, the Partial Correlation Coefficient (PCC) is defined as follows: Two regression models have to be constructed

$$x_j = \alpha_0 + \sum_{i=1, i \neq j}^d \alpha_i x_i + \delta \quad \text{and} \quad y = \beta_0 + \sum_{i=1, i \neq j}^d \beta_i x_i + \varepsilon, \quad (12)$$

with the corresponding fitted regression models

$$\hat{x}_j = \hat{\alpha}_0 + \sum_{i=1, i \neq j}^d \hat{\alpha}_i x_i \quad \text{and} \quad \hat{y} = \hat{\beta}_0 + \sum_{i=1, i \neq j}^d \hat{\beta}_i x_i. \quad (13)$$

The PCC is the CC of the variables $x_j - \hat{x}_j$ and $y - \hat{y}$.

Both $PCCs$ and $SRCs$ provide measures of importance. PCC_j characterizes the linear relationship between X_j and Y after excluding the linear effects on Y of the other input factors. When the input factors are independent, $PCCs$ and $SRCs$ give the same ranking of variable importance, i.e. an ordering based on the absolute value of $PCCs$ is the same as an ordering based on the absolute value of $SRCs$. If the input factors are correlated, the results based on $SRCs$ or on $PCCs$ may be misleading, as some authors remark, see [Helton et al. 2006].

Data transformations

The linear regression model proposed as a tool to perform sensitivity analysis has some evident limitations to analyse the dependence between the input factors and the output variables, i.e. not being able to tackle appropriately non-linear input-output relations.

A simple and quite effective method is the transformation of either a part of the inputs or the output, or both into their logarithms. The coefficient of determination is the key indicator to judge if the transformation was useful or not. If, after applying the transformation(s), R^2 increases, the variability of the output may be explained in a more appropriate way in the new scales. Then, the analysis based on the raw values should be left in favour of the one based on the transformed values.

A widely used transformation is the rank transformation. Whenever a nonlinear but monotonic relationship between X_j and Y exists, a rank transformation leads to a linear relationship between the ranks. After rank transformation, standardized rank regression coefficients (SRRCs) and Partial Rank Correlation Coefficients (PRCCs) can be computed in a similar way as the SRCs and PCCs described in this section.

2.2.2.3 *Monte Carlo Filtering*

Sensitivity can be interpreted in different ways that have nothing to do with either linear or monotonic relations. An input parameter can be considered important with respect to a given output variable if a clear link exists between specific regions of both. This link may be impossible to uncover by regression based techniques. A method to uncover such relations is *Monte Carlo filtering*. Monte Carlo Filtering (MCF) is based on dividing the output sample in two or more subsets according to some criterion (achievement of a given condition, exceeding a threshold, etc.) and testing if the inputs associated to those subsets are different or not. As an example, we could divide the output sample in two parts, the one that exceeds a safety limit and the rest. It could be tested by some tools if runs in both subsamples are related to different regions of a given input. If this is the case, a further analysis of the range of values of that input parameter that leads to the exceeding of the safety limit would be important in order to predict under which conditions the safety limit will be exceeded. The tools used in the present study to provide adequate answers to this type of questions are a two non-parametric statistics and their associated tests:

- The two-sample *Smirnov* test and
- the *Mann-Whitney* (or *Wilcoxon*) two-sample test.

The two-sample *Smirnov* test is based on comparing empirical cumulative distribution functions, while the *Mann-Whitney* test is based on average ranks. Normally, when any of these tests is applied in other areas, two samples are taken via random sampling from two or more populations. When the test is used for the purpose of sensibility analysis, the samples used are in fact subsamples of an input parameter sample. These two samples are obtained in the following way:

- Consider the samples obtained via simulation for a given input parameter X and an output variable Y : (x_1, x_2, \dots, x_n) and (y_1, y_2, \dots, y_n)
- Then, sort (y_1, y_2, \dots, y_n) , obtaining $(y^{(1)}, y^{(2)}, \dots, y^{(n)})$, and re-order (x_1, x_2, \dots, x_n) according to the order of the sample of Y , obtaining $(x^{[1]}, x^{[2]}, \dots, x^{[n]})$.

- Set a threshold on $(y^{(1)}, y^{(2)}, \dots, y^{(n)})$ based on a given criterion and divide it in the corresponding two subsamples $(y^{(1)}, y^{(2)}, \dots, y^{(h)})$ and $(y^{(h+1)}, y^{(h+2)}, \dots, y^{(n)})$. Let us call 'C' the criterion.
- Divide accordingly the sample $(x^{[1]}, x^{[2]}, \dots, x^{[n]})$ in the two subsamples $(x^{[1]}, x^{[2]}, \dots, x^{[h]})$ and $(x^{[h+1]}, x^{[h+2]}, \dots, x^{[n]})$. These two subsamples will be the samples used to perform the test.

The two-sample Smirnov test

The procedure to perform this test has the following steps:

- Assumptions: a) Both samples are random and mutually independent and b) the variable considered should be continuous in order to get an exact test (having a discrete variable does not preclude the use of the test unless it takes only a few different values).
- Hypotheses:
- $H_0: F_{x|C1}(x|C1) = F_{x|C2}(x|C2) \quad \forall x \in IR$
- $H_1: \exists x' \in IR / F_{x|C1}(x'|C1) \neq F_{x|C2}(x'|C2)$, where stands for the cumulative distribution function, C1 indicates fulfilling the criterion and C2 means not fulfilling the criterion
- Discrepancy measure (test statistic): $T_{Smirnov} = \sup_x |S_{C1}(x) - S_{C2}(x)|$, which is the maximum vertical distance between the empirical distribution functions ($S_{C1}(x)$) of $(x^{[h+1]}, x^{[h+2]}, \dots, x^{[n]})$, the subsample of the input parameter linked to the subsample of the output variable that fulfils the criterion, and the empirical distribution functions ($S_{C2}(x)$) of $(x^{[1]}, x^{[2]}, \dots, x^{[h]})$.
- Decision rule: As a generally accepted rule, the null hypothesis is accepted if the p-value of the test is above 0.05, otherwise it is rejected. The distribution of the test statistic under the null hypothesis may be found in [Conover 1980]. As an alternative to this general rule, [Saltelli et al. 2004] propose the following decision rule
- P-value < 0.01 ==> the input parameter is critical.
- $0.01 \leq \text{p-value} \leq 0.1$ ==> the input parameter is important
- p-value ≥ 0.1 ==> the input parameter is not important

The Mann-Whitney two-sample test

The implementation of this test is the same as for the Smirnov two-sample test. The only difference is the test statistic used as a measure of discrepancy between the null hypothesis and the alternative hypothesis, which in this case is

$$T_{Mann-Whitney} = \sum_{i=h+1}^n R(x^{[i]}), \quad (14)$$

where $R(x^{[i]})$ is the rank associated to observations $x^{[i]}$ of the input parameter corresponding to the output values that satisfy condition C1.

Figure 9 and Figure 10 represent respectively the implementation of the Smirnov and the Mann-Whitney tests for one output variable and one input variable (the same in both cases). The output variable sample, whose size was 459, was split in the 10% largest samples and the rest 90% samples. The values of the Smirnov statistic is 0.6733, which corresponds to a p-value=0. The value of the Mann-Whitney statistic is 17534, which corresponds to a p-value=0. The result of both tests is conclusive, certainly the input values that produce the largest 10% values of the output variable are different from the values that produce the rest of the values of the same output variable.

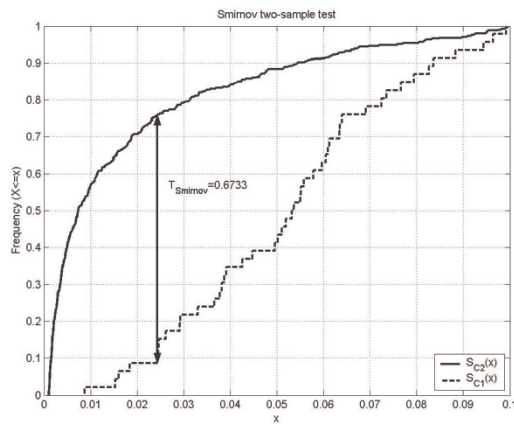


Figure 9 Graphic representation of a two-sample Smirnov test.

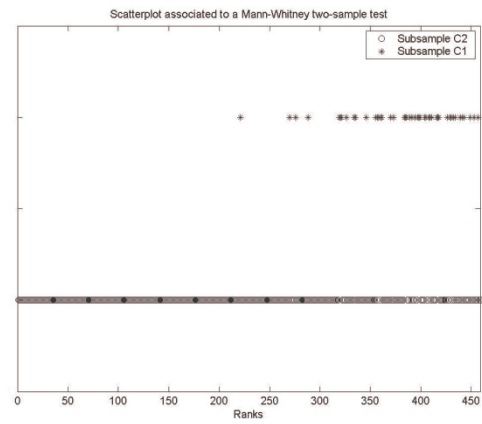


Figure 10 Graphic representation of a two-sample Mann-Whitney test.

2.2.3 Variance based methods

The variance, or equivalently the standard deviation, and the entropy, are the main measures of uncertainty in the theory of Probability. The larger the variance of a random variable is, the less accurate our knowledge about it is. Decreasing the variance of a given output variable is quite an attractive target, that may be achieved sometimes by decreasing the variance of input parameters (this is not always true, remember the possibility of risk dilution). This is what makes so attractive methods that try to find out what fraction of the output uncertainty (variance) may be attributed to the uncertainty (variance) in each input.

Variance based methods find their theoretical support in Sobol's decomposition of any integrable function $Y = f(X_1, \dots, X_d)$ in the unit reference hypercube into 2^k orthogonal summands of different dimension:

$$Y = f(X_1, \dots, X_k) = f_0 + \sum_i f_i(X_i) + \sum_{i < j} f_{ij}(X_i, X_j) + \dots + f_{1\dots k}(X_1, \dots, X_k) \quad (15)$$

This decomposition exists and is unique. The term f_0 is the mean of Y (a constant). The first order function $f_i(X_i)$ of the decomposition represents the effect of the input factor X_i acting independently (and generally in a nonlinear way) upon the output Y . The second order function $f_{ij}(X_i, X_j)$ represents the joint effect of the input factors X_i and X_j upon the output Y . Higher order functions represent the joint effect of the corresponding input factors upon the output. The last term of the decomposition gives the residual influence of all the input factors together.

The component functions of the HDMR (15) are a set of projectors that are mutually orthogonal (the integral of the product of any pair of them over the whole input space is null); their expressions are:

$$\begin{aligned} f_0 &= E[Y] \\ f_i(X_i) &= E[Y | X_i] - f_0 \\ f_{ij}(X_i, X_j) &= E[Y | X_i, X_j] - f_i(X_i) - f_j(X_j) - f_0 \\ f_{ijk}(X_i, X_j, X_k) &= E[Y | X_i, X_j, X_k] - f_{ij}(X_i, X_j) - f_{ik}(X_i, X_k) - f_{kj}(X_k, X_j) \\ &\quad - f_i(X_i) - f_j(X_j) - f_k(X_k) - f_0 \\ &\dots \end{aligned} \quad (16)$$

Where $E[Y | X_i]$ is the expectation of Y conditional on X_i , $E[Y | X_i, X_j]$ is the expectation of Y conditional on X_i and X_j and so on. The details for obtaining these formulas may be found in [Sobol 1993].

2.2.3.1 The decomposition of variance

For *independent* input factors and using the fact that any two different components of the HDMR decomposition are orthogonal, the variance of the model $Y = f(X_1, \dots, X_d)$ can be written as:

$$V = \text{Var}(Y) = \sum_i V_i + \sum_{i < j} V_{ij} + \dots + V_{1\dots d}, \quad (17)$$

where

$$\begin{aligned} V_i &= \text{Var}(E[Y | X_i]) \\ V_{ij} &= \text{Var}(E[Y | X_i, X_j]) - V_i - V_j \\ V_{ijk} &= \text{Var}(E[Y | X_i, X_j, X_k]) - V_{ij} - V_{jk} - V_{ki} - V_i - V_j - V_k \\ &\dots \end{aligned} \quad (18)$$

[Sobol 1993] shows that the individual terms in expression (16) are the variances of the functions of the corresponding indices in the HDMR decomposition (15), which means that:

$$\begin{aligned} V_i &= \text{Var}(f_i(X_i)) \\ V_{ij} &= \text{Var}(f_{ij}(X_i, X_j)) \\ V_{ijk} &= \text{Var}(f_{ijk}(X_i, X_j, X_k)) \\ &\dots \end{aligned} \tag{19}$$

It is important to remark that Sobol's decomposition is equivalent to the classical Analysis of Variance (ANOVA) decomposition of variability used in Statistics.

2.2.3.2 Sobol sensitivity indices

Using the variance decomposition (17), Sobol sensitivity indices are defined as:

- first order sensitivity indices $S_i = \frac{V_i}{V}$
- second order sensitivity indices, which gives the sensitivity of the variance of the output to the interaction between the input factors X_i, X_j : $S_{ij} = \frac{V_{ij}}{V}$,
- and so on until order k .

We can divide both sides of (17) and obtain a very useful relation:

$$1 = \sum_i S_i + \sum_{i < j} S_{ij} + \dots + S_{1\dots k} \tag{20}$$

The interpretation of the sensitivity indices is straightforward, as the sum of all indices is 1 and as they are all positive: the larger the sensitivity index (close to 1), the more influent the corresponding input (or group of inputs). However, the number of all these indices is $2^d - 1$ and when the number of input factors is too large their computation and interpretation becomes impossible. This is why [Homma and Saltelli 1996] introduced the total sensitivity indices which assess the sensitivity of the variance of the output with respect to the standalone and every interaction of the considered input factor, by:

$$S_{T_i} = \sum_{\substack{k \in \text{all the sets of} \\ \text{indices containing} \\ \text{the index } i}} S_k \tag{21}$$

For a model with three input factors, we have $S_{T_1} = S_1 + S_{12} + S_{13} + S_{123}$, and similar expressions for the two other total indices. There is a simpler definition of the total indices:

$$S_{T_i} = 1 - \frac{V_{-i}}{V}, \quad (22)$$

where $V_{-i} = \text{Var}(E[Y | X_{-i}])$ represents the variance explained by all the factors except X_i . This definition makes the total indices as easy to estimate as the first order indices.

2.2.3.3 Methods to estimate sensitivity indices

Sobol/Saltelli estimation method

This method, due to [Sobol 1993] and addressed by [Archer et al. 1997] and [Saltelli 2002], allows the computation of all the terms in the variance decomposition. Let us consider a sample of size n of the input vector, $(x_{k1}, \dots, x_{kd})_{k=1, \dots, n}$. The expectation f_0 and the variance V of Y are estimated as usual. To estimate the first order sensitivity indices we need to estimate $V_i = \text{Var}(E[Y | X_i])$. This variance can be written as:

$$V_i = \text{Var}(E[Y | X_i]) = \underbrace{E[E[Y | X_i]^2]}_{U_i} - E[E[Y | X_i]]^2 = U_i - E[Y]^2, \quad (23)$$

using the identity $E[E[Y | X_i]] = E[Y]$. [Sobol 1993] estimates U_i by means of

$$\hat{U}_i = \frac{1}{n} \sum_{k=1}^n f(x_{k1}, \dots, x_{k(i-1)}, x_{ki}, x_{k(i+1)}, \dots, x_{kd}) f(x'_{k1}, \dots, x'_{k(i-1)}, x_{ki}, x'_{k(i+1)}, \dots, x'_{kd}) \quad (24)$$

where $(x'_{k1}, \dots, x'_{kd})_{k=1, \dots, n}$ represents a second sample of size n of the input vector and the conditioning is taken into account by “resampling” all the variables except x_{ki} . The first order sensitivity indices are then estimated by means of

$$\hat{S}_i = \frac{\hat{U}_i - \hat{f}_0^2}{\hat{V}} \quad (25)$$

For the second order sensitivity indices, the conditional variance is written as

$$\text{Var}(E[Y | X_i, X_j]) = \underbrace{E[E[Y | X_i, X_j]^2]}_{U_{ij}} - E[E[Y | X_i, X_j]]^2 = U_{ij} - E[Y]^2, \quad (26)$$

and the quantity U_{ij} is estimated by means of

$$\begin{aligned} \hat{U}_{ij} = \frac{1}{n} \sum_{k=1}^n f(x_{k1}, \dots, x_{k(i-1)}, x_{ki}, x_{k(i+1)}, \dots, x_{k(j-1)}, x_{kj}, x_{k(j+1)}, \dots, x_{kd}) \\ \times f(x'_{k1}, \dots, x'_{k(i-1)}, x_{ki}, x'_{k(i+1)}, \dots, x'_{k(j-1)}, x_{kj}, x'_{k(j+1)}, \dots, x'_{kd}) \end{aligned} \quad (27)$$

The second order sensitivity indices are then estimated as

$$\hat{S}_{ij} = \frac{\hat{U}_{ij} - \hat{f}_0^2 - \hat{V}_i - \hat{V}_j}{\hat{V}} \quad (28)$$

with $\hat{V}_i = \hat{U}_i - \hat{f}_0^2$. The procedure continues for the estimation of all the indices. The total sensitivity indices are estimated as

$$\hat{S}_{T_i} = 1 - \frac{\hat{U}_{\sim i} - \hat{f}_0^2}{\hat{V}} \quad (29)$$

with

$$\hat{U}_{\sim i} = \frac{1}{n} \sum_{k=1}^n f(x_{k1}, \dots, x_{k(i-1)}, x_{ki}, x_{k(i+1)}, \dots, x_{kd}) f(x_{k1}, \dots, x_{k(i-1)}, x'_{ki}, x_{k(i+1)}, \dots, x_{kd}) \quad (30)$$

Fourier Amplitude Sensitivity Test (FAST)

This method is due to [Cukier et al. 1973, 1975, 1978] and [Schaiby and Shuler 1975]. It is based on the fact that it is possible to convert the k -dimensional integrals defining the variances in the variance decomposition into one-dimensional integrals, by using the transformation functions $G_i, i=1, \dots, k$ defined as: $x_i = G_i(\sin(\omega_i s))$, where s is a scalar variable $s \in]-\pi, \pi[$ and $\{\omega_i\}$ is a set of integer angular frequencies. Cukier et al. showed (using the properties of Fourier series) that for properly chosen ω_i and G_i , the expectation and the variance of Y can be approximated respectively by:

$$f_0 = E[Y] = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) ds \quad (31)$$

$$\begin{aligned} Var(Y) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(s) ds - f_0^2 \approx \sum_{j=-\infty}^{\infty} (A_j^2 + B_j^2) - (A_0^2 + B_0^2) \\ &= 2 \sum_{j=1}^{\infty} (A_j^2 + B_j^2) \end{aligned} \quad (32)$$

with $f(s) = f(G_1(\sin(\omega_1 s)), \dots, G_k(\sin(\omega_k s)))$ and A_j, B_j the Fourier coefficients of the cosine and sine series.

The part of the variance due to one input factor X_i is the sum of the squares of the Fourier coefficients due to the angular frequency ω_i and its harmonics:

$$V_i = 2 \sum_{k=1}^{\infty} (A_{k\omega_i}^2 + B_{k\omega_i}^2) \quad (33)$$

The sensitivity index S_i is estimated as:

$$S_i = \frac{\sum_{k=1}^{\infty} (A_{k\omega_i}^2 + B_{k\omega_i}^2)}{\sum_{j=1}^{\infty} (A_j^2 + B_j^2)} \quad (34)$$

However, in order to use the formula (34), several conditions have to be fulfilled:

- a maximum number of harmonics has to be fixed in expression 2.2.3.3.12; it is denoted by M [Cukier et al. 1975] have empirically shown that a good compromise between a good quality of the indices and their estimation cost is $M = 4$ or $M = 6$)
- the functions G_i have to be chosen; the initial choices were given in [Cukier et al. 1973] and in [Koda et al. 1979]
- $x_i = G_i(\sin(\omega_i s)) = x_i^0 \exp(\nu_i \sin(\omega_i s))$
- $x_i = G_i(\sin(\omega_i s)) = x_i^0 (1 + \nu_i \sin(\omega_i s))$
- the choice of the angular frequencies ω_i such that are free of inferences up to a certain order (usually up to order 4). These frequencies change with the number of input parameters. Sets of such frequencies are given in [Schaibly and Shuler 1973].
- the sampling of the variable $s \in]-\pi, \pi[$, for the computation of the Fourier coefficients A_j, B_j has to respect the Nyquist-Shannon sampling theorem (i.e. the sampling frequency should be at least twice the maximal frequency of the function to be sampled); if we denote by N_s the size of this sample, then its minimal value should be $N_s > 2M \max(\omega_i)$.
- In addition to the computational cost, the FAST method to estimate Sobol sensitivity indices is limited to estimating first order sensitivity indices.

Extended Fourier Amplitude Sensitivity Test (EFAST)

In order to be able to estimate total sensitivity indices, [Saltelli et al. 1999] introduced the *extended FAST* (EFAST) method, where the part of variance due to all the variables except X_i is the sum of the squares of the Fourier coefficients due to all the angular frequency $\omega_{\sim i}$ different from ω_i , leading to:

$$S_{T_i} = 1 - \frac{\sum_{k=1}^{\infty} (A_{k\omega_i}^2 + B_{k\omega_i}^2)}{\sum_{j=1}^{\infty} (A_j^2 + B_j^2)} \quad (35)$$

Another difference from the initial FAST method is that the functions G_i have been chosen such that they are “filling” in a more suitable way the input space:

- $x_i = G_i(\sin(\omega_i s)) = \frac{1}{2} + \frac{1}{\pi} \arcsin(\sin(\omega_i s))$
- $x_i = G_i(\sin(\omega_i s)) = \frac{1}{2} + \frac{1}{\pi} \arcsin(\sin(\omega_i s + \varphi_i))$

where φ_i is a random phase chosen uniformly in $[0, 2\pi]$. Moreover, the choice of the angular frequencies ω_i is no longer constrained by the “no interferences up to a certain order” rule. For a given variable X_i , the pair of indices (S_i, S_{π}) is computed simultaneously; the only constraint is that the angular frequency ω_i (corresponding to x_i) should be large compared to the ones corresponding to the rest of the variables (for instance $\omega_i = 124$, $\omega_{\sim i} = 1$).

Estimation of sensitivity indices by Random Balanced Design (RBD)

The RBD method, see [Tarantola et al. 2006] or [Saltelli et al. 2008], is similar to FAST or EFAST. It uses the parametrisation $x_i = G_i(\sin(\omega_i s))$, $s \in]-\pi, \pi[$ is the scalar variable and $\{\omega_i\}$ is the set of integer angular frequencies. The scalar variable is then discretised. The main difference with FAST – EFAST is that the input space is explored using the same frequency ω , to avoid the use of a complicated algorithm for choosing the different frequencies. Doing so, the set of discretized input parameters x_i (i.e. the design) is no longer space-filling. Therefore, random permutations of the coordinates of those points are taken in order to generate a set of scrambled points that cover the input space. To compute the first order sensitivity index S_i , the output y is re-ordered in increasing order of the values taken by the factor X_i . The Fourier spectrum of the output is calculated for the frequencies $\{\omega, 2\omega, \dots, M\omega\}$ to estimate the sensitivity index S_i . The procedure is repeated for each input factors to obtain all the first order sensitivity indices.

Its main advantages with respect to (E)FAST is that there is no minimum sample size to avoid interference problems and no special algorithm for providing the free interference set of frequencies. Also, the number of harmonics in the spectrum (i.e. M) may be increased without affecting the size of the sample. The main disadvantage is that only first order sensitivity indices may be estimated.

Effective Algorithm for computing Sensitivity Indices (EASI)

The idea behind RBD can be modified to be applied to given data: Instead of generating random permutations, permutations π_i are done on the columns X_i of the given data matrix X

such that each $\pi_i(X_i)$ 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 $x = (x_j)_{j=1,\dots,n}$ 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 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. Now, taking all odd indices from $(x_{(j)})$ in increasing order followed by all even indices in decreasing order gives us a vector $(x_{[j]})$ that satisfies $x_{[j]} \leq x_{[j+1]}$ if $j \leq (n+1)/2$, and $x_{[j]} \geq x_{[j+1]}$ if $j > (n+1)/2$.

There exists a permutation π_i with $\pi_i((x_j)) = (x_{[j]})$. As in RBD, this permutation is also applied to the output $\pi_i(Y)$. The Fourier transformation of the permuted output is analysed for frequency responses.

Correlation Ratios (CR)

One straightforward manner to estimate first order sensitivity indices is to estimate the numerator, $V[E[Y | X_i]]$, by creating a partition of a random sample according to the values of X_i . The values of X_i are sorted and then divided in several sets (the smallest values of X_i in the first set, the second smallest values of X_i in the second set and so on). The values of the output are classified according to this partition of the values of X_i . Then the average of Y is computed in each set and the variance of the averages is used as an estimator of $V[E[Y | X_i]]$. This method (CR-VCE, Correlation Ratios – Variance of Conditional Expectations) may be extended to compute higher order sensitivity indices, but then it suffers the curse of dimensionality.

Extensions of this method, such as CR-ECV (CR-Expectation of Conditional Variances), CR-FIT (CR-Polynomial Fit) and CR-CLM (CR-Conditional Linear Model) may be found in [Plischke et al. (2009)].

3. Sensitivity Analysis benchmark, analytical cases

For testing the different SA methods a benchmark exercise was set up with a number of simple and well-defined mathematical models that are amenable to exact analysis of their sensitivity to parameter variations. By applying probabilistic SA methods to such systems and comparing the results to the theoretical ones one can assess their accuracy, robustness and reliability. The models considered (except for the one described in section 3.1.2 below) were taken from [Saltelli et al., 1997] and [Saltelli et al., 2000]. A first round of PA benchmark studies were performed by the members of the PAMINA 2.1.D work group. Contributions were received from ANDRA (France), Facilia (Sweden), JRC Petten (The Netherlands) and TU Clausthal (Germany). 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.

Nine different methods have been tested in the benchmark. Some methods allow the estimation of sensitivity indices using the sample of realisations already used for the probabilistic assessment. These methods are called “cheap methods” and five of them are tested: EASI, FIT, CLM, VCE and ECV. By the contrary, other methods require special sampling schemes, although usually offer better estimates compared to cheap methods. Four “expensive methods” have been tested: RBD, HIS, FAST and EFAST.

3.1 The analytical benchmark cases

In the following we discuss these results from the second round. Contributions were received from Facilia (Sweden), GRS Cologne (Germany), JRC Petten (The Netherlands), and TU Clausthal (Germany). 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.

3.1.1 *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

$$Y = \sin X_1 + 7 \sin^2 X_2 + 0.1 X_3^4 \sin X_1, X_i \sim U(-\pi, \pi)$$

where the random variables X_i are uniformly distributed in $U(-\pi, \pi)$. 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.

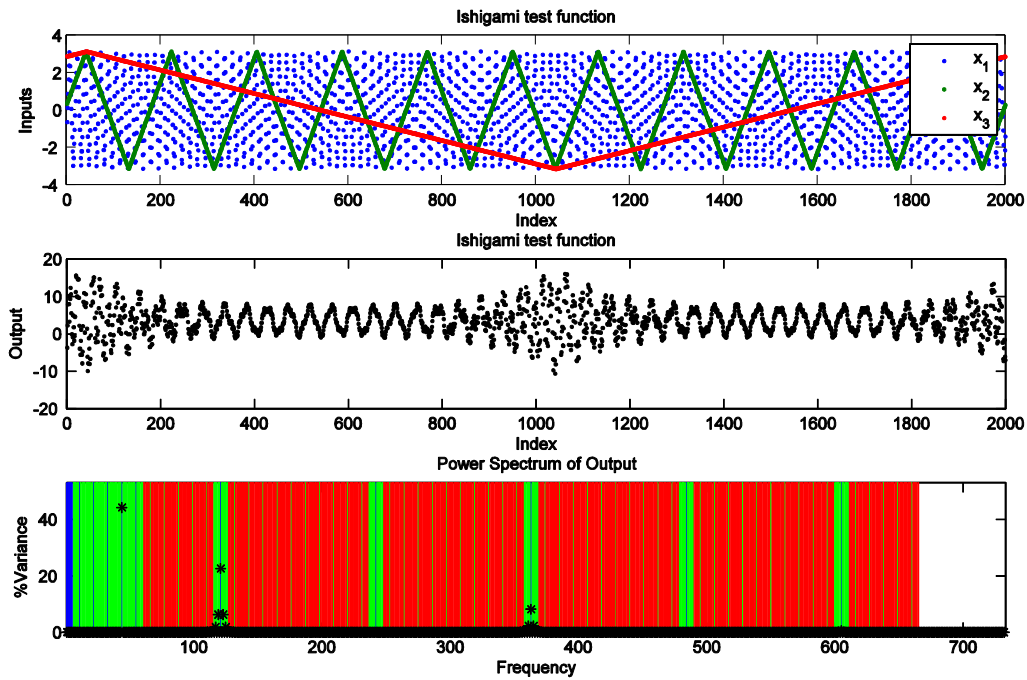


Figure 11 FAST analysis of the Ishigami function

Figure 11 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 11 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 ($x_1 x_3^2$), 117 and 125 ($x_1 x_3^4$), 361 and 365 ($x_1^3 x_3^4$). A well-equipped eye might also spot frequencies 115 and 127 ($x_1 x_3^6$), 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. Figure 12 and Figure 13 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 $2M(1 + (2M + 1) + \dots + (2M + 1)^{k-1}) = 6(1 + 7 + 49) = 342$ realisations, for EFAST $2kM(2M + 1) = 126$ realisations.

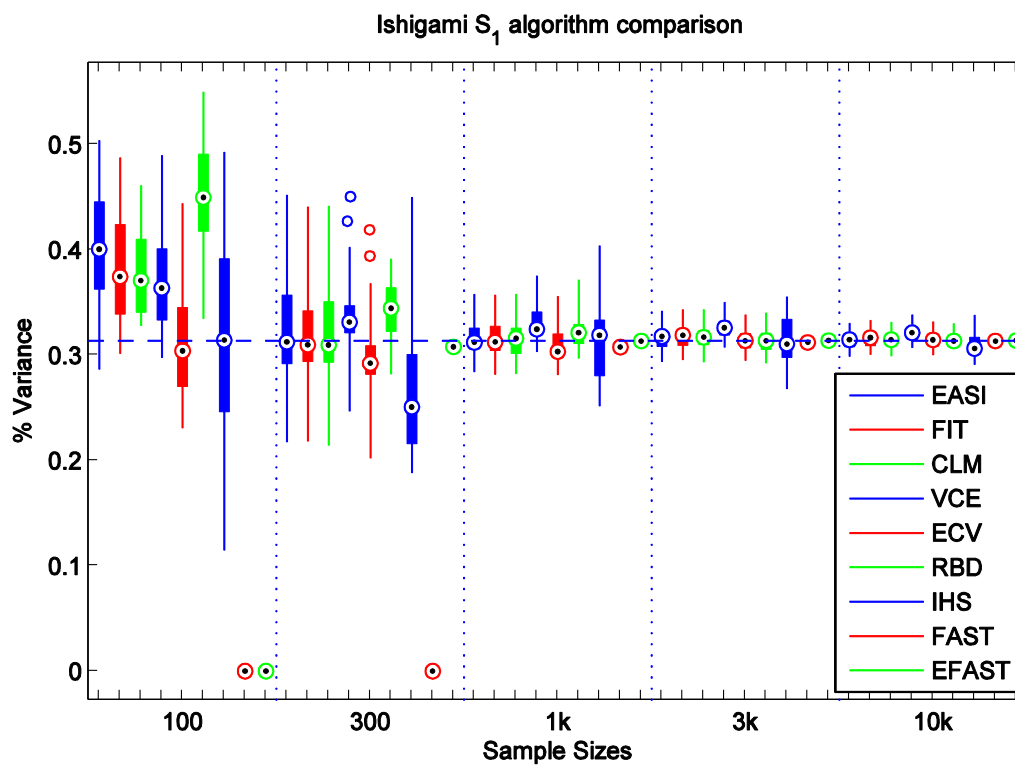


Figure 12 TUC results - Box plots for S_1

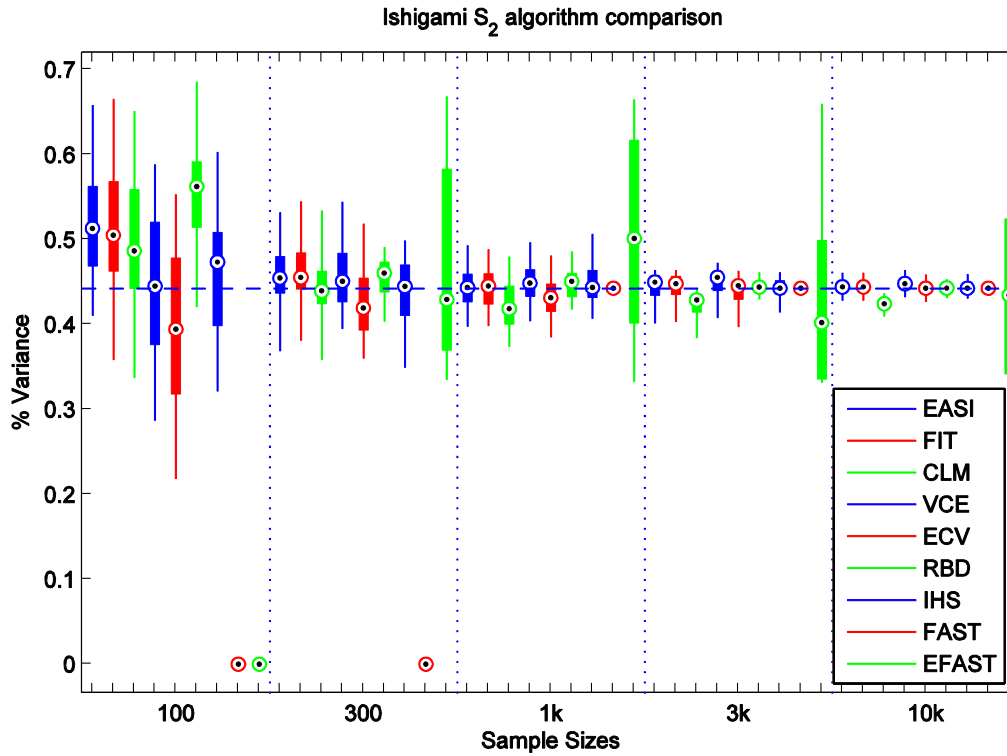


Figure 13 TUC results - Box plots for S_2

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 Figure 14. The TUC version uses the following formulas [Sobol' et al. 2007] that reduce the error introduced via cancellation,

$$\hat{V}_i = \frac{1}{n} \sum_{k=1}^n f(x_{k1}, \dots, x_{kd}) (f(x'_{k1}, \dots, x_{ki}, \dots, x'_{kd}) - f(x'_{k1}, \dots, x'_{kd}))$$

$$\hat{V}_{\sim i} = \frac{1}{n} \sum_{k=1}^n f(x'_{k1}, \dots, x'_{kd}) (f(x'_{k1}, \dots, x_{ki}, \dots, x'_{kd}) - f(x_{k1}, \dots, x_{kd}))$$

hence the true values should be better approximated. However, this theoretical result does not become apparent in the figure.

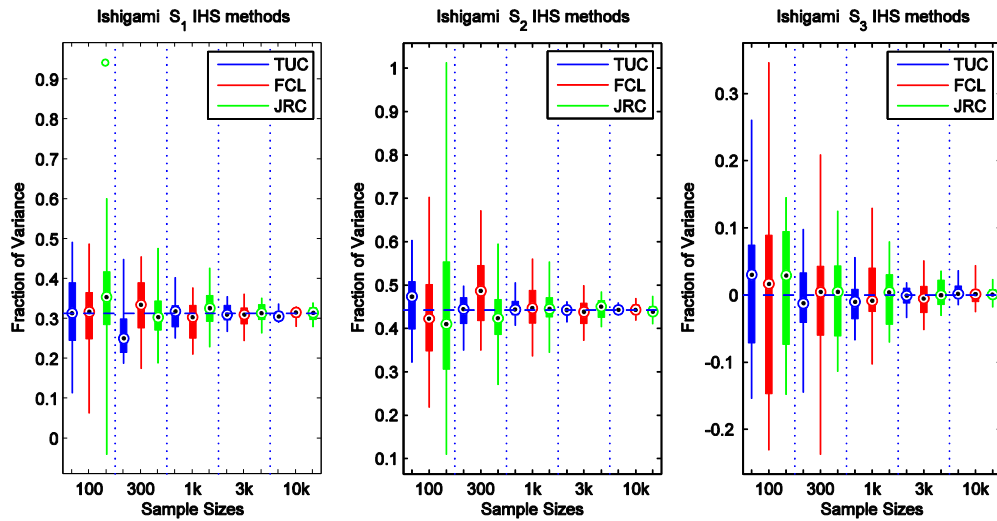


Figure 14 Ishigami-Homma-Saltelli algorithm comparison

The behaviour of the IHS algorithm changes drastically when using Sobol's LP_τ sequence as pseudo random number source, see Figure 18. Then even for small sample sizes good estimates are computed. Here, the basic sample size is rounded to the next power of 2.

The RBD method-although using an artificially generated sample-seems to have no better properties than the cheap methods, see also Figure 16 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 15. 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 sub-sample 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 sub-sample size resembling the rule-of-thumb $\ell = \lfloor \sqrt{n} \rfloor$. 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.

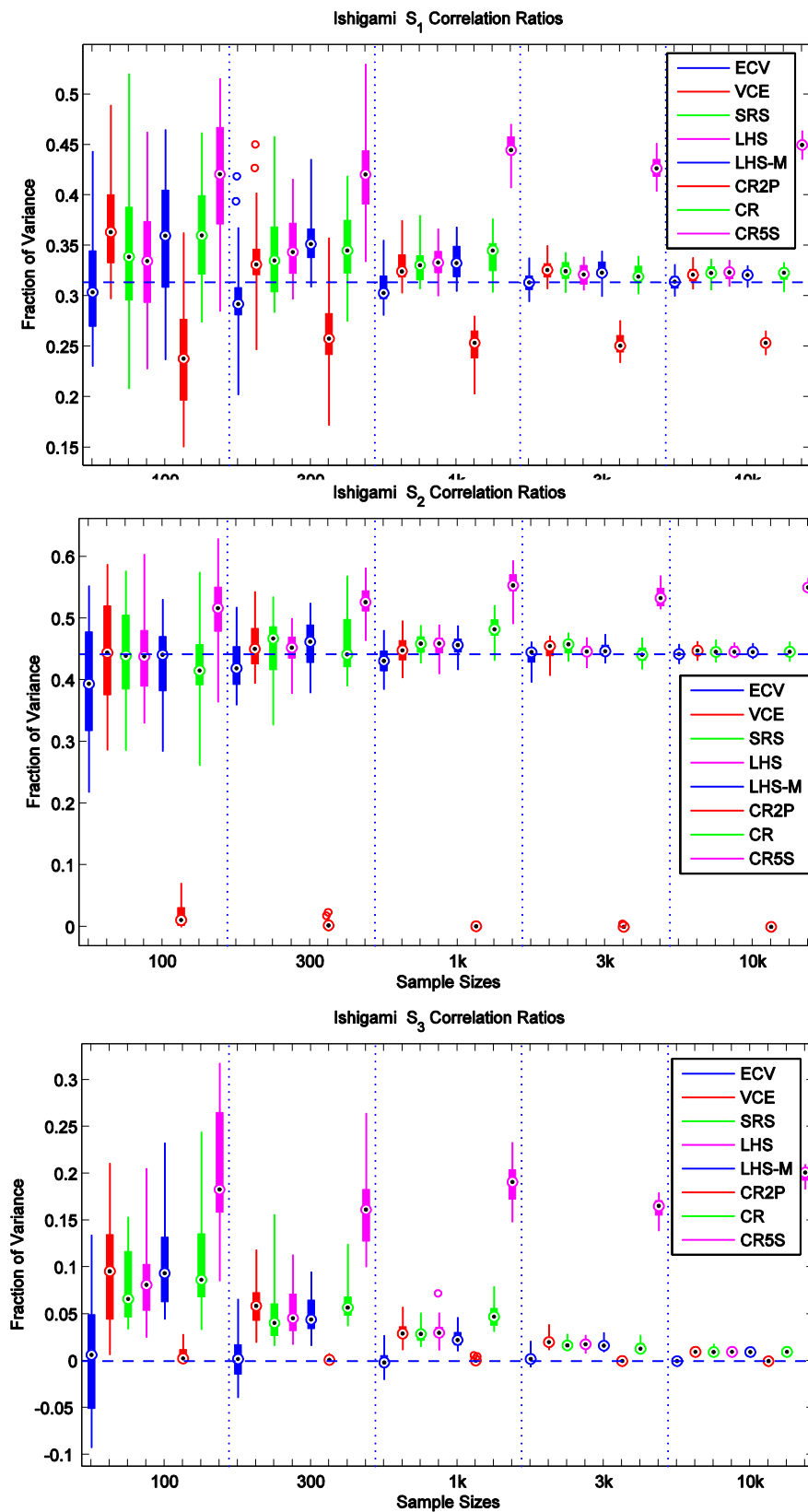


Figure 15 Correlation Ratio methods for the Ishigami function.

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 16 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.

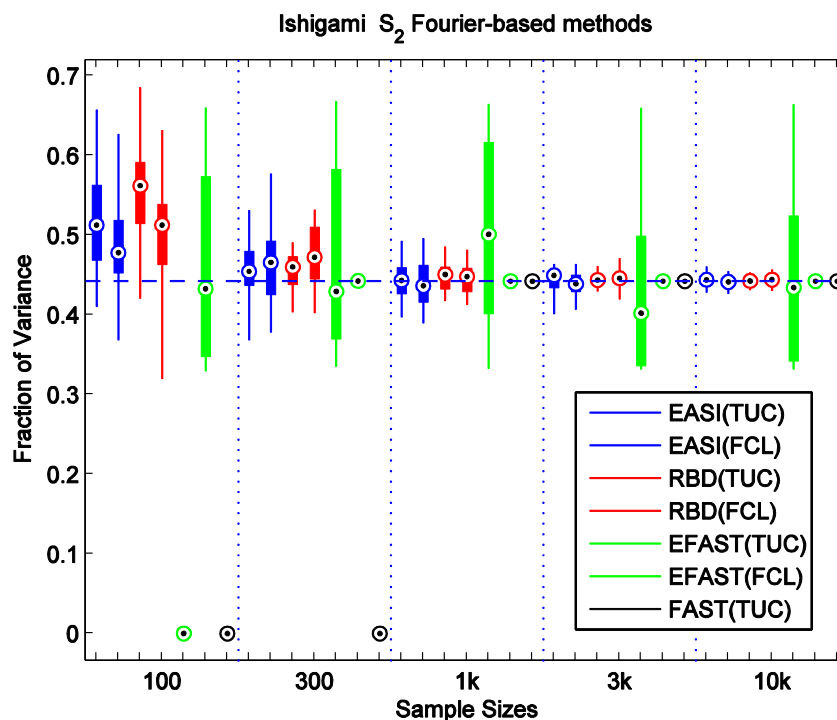


Figure 16 Fourier-based methods for the Ishigami function.

The results for the total effects of x_3 are presented in Figure 17. 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.

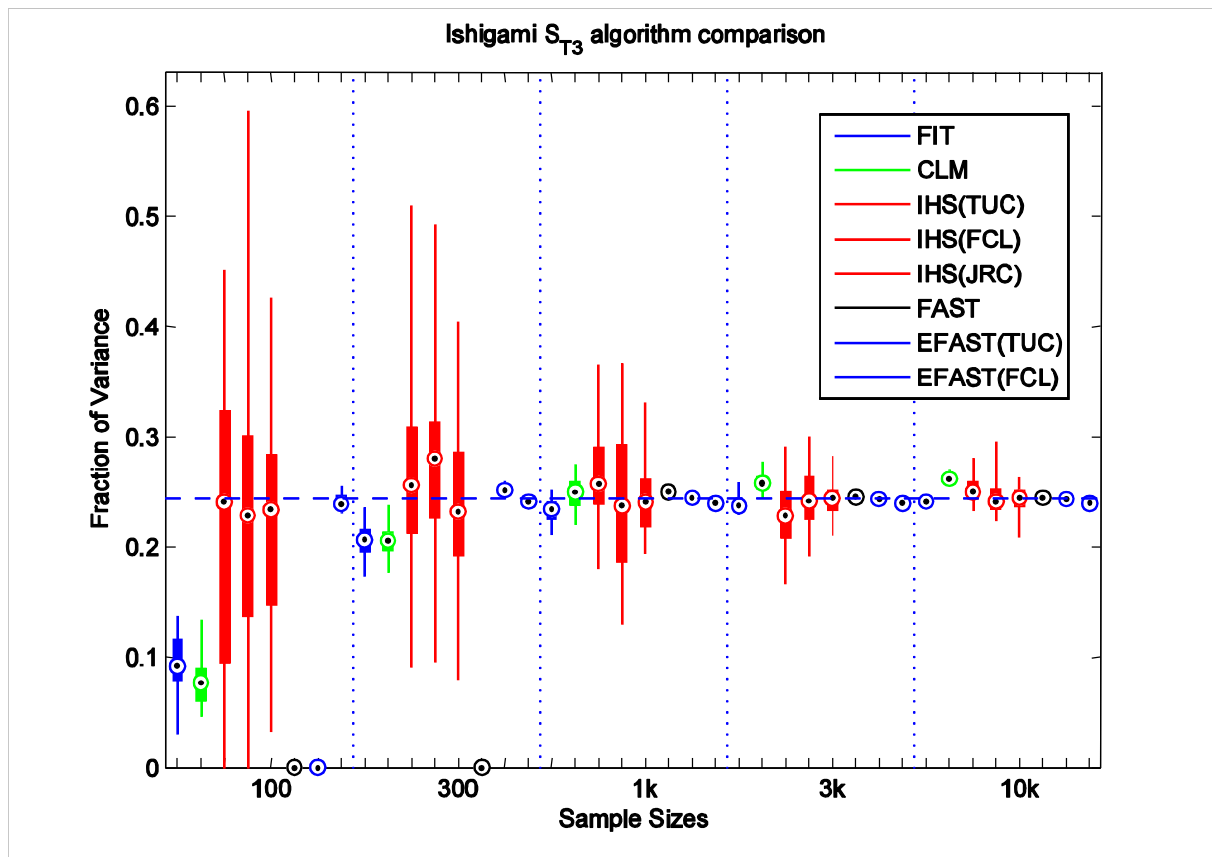


Figure 17 Box plots for S_{T3}

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 18. 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.

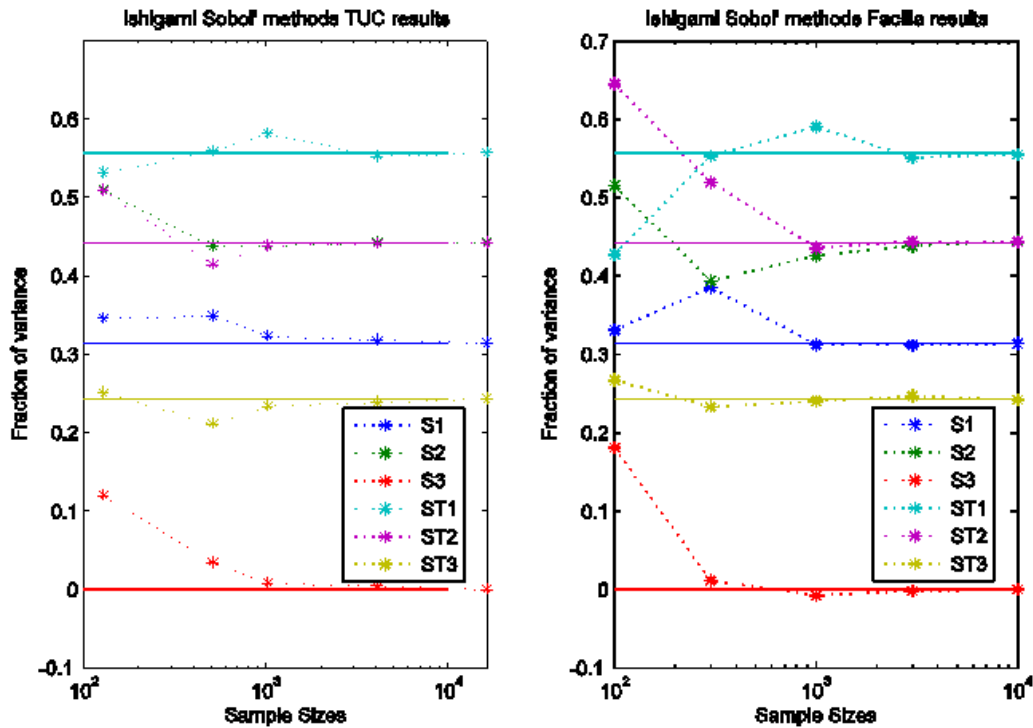


Figure 18 First and total effects using Sobol's method for the Ishigami function.

3.1.2 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,

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

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.

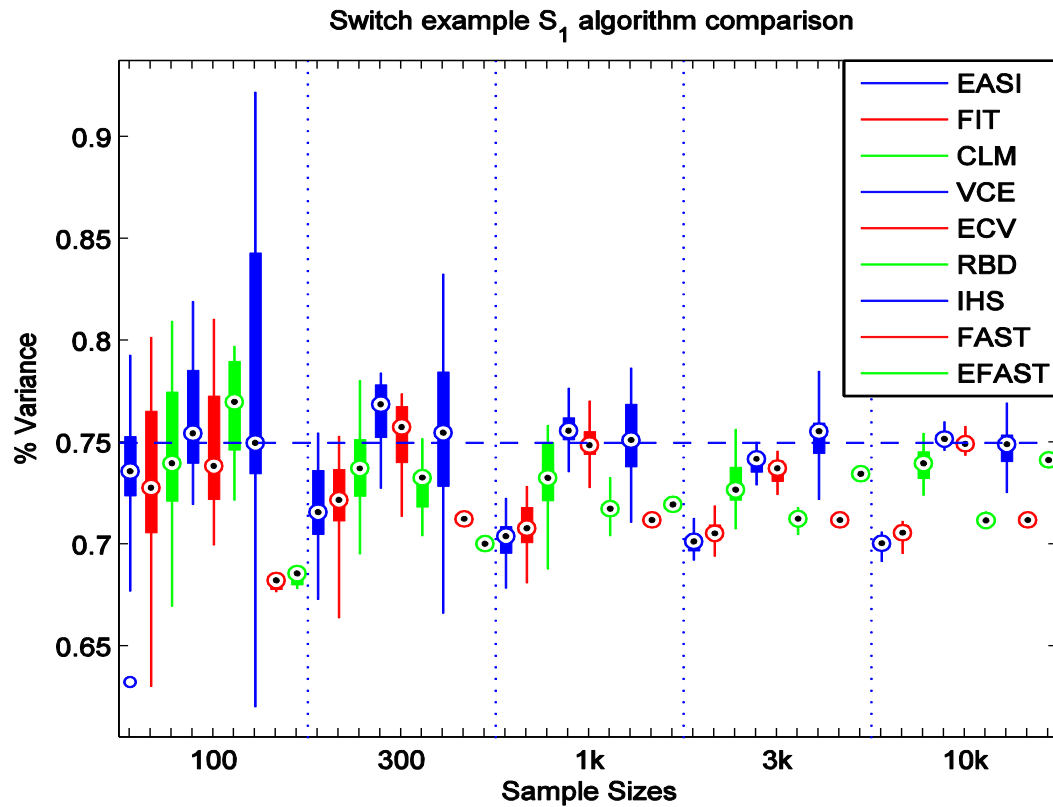


Figure 19 TUC results - Box plots for S_1

Figure 19 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 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 23 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 19 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 22, which shows that the algorithms with the same value of M produce equivalent estimates. Figure 21 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. Figure 24. Even a sampling size which is not a power-of-2 produces no 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 20 shows computed total effects. Again, the polynomial fit and the FAST algorithm with bounded maximal frequency fail to catch the exact value.

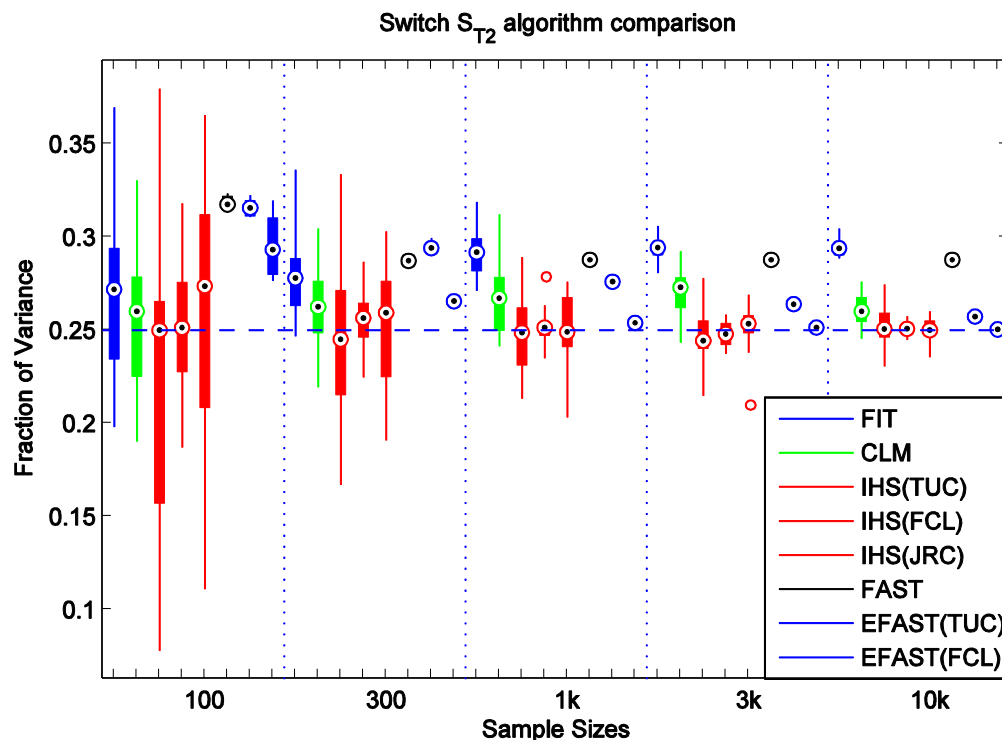


Figure 20 TUC results - Box plots for S_{T2}

Figure 24 shows that the total effects computed by the Sobol' method are in good agreement with the theoretical values.

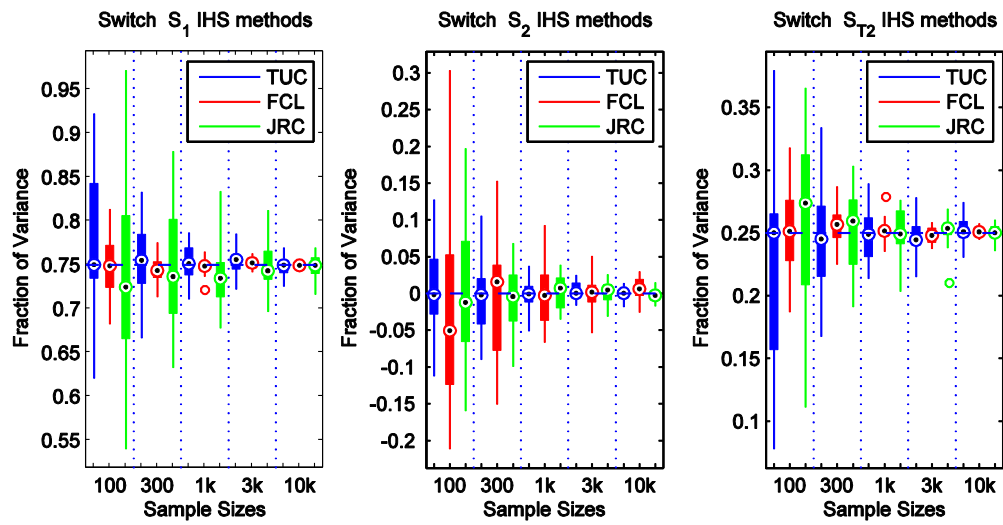


Figure 21 Ishigami-Homma-Saltelli algorithm comparison.

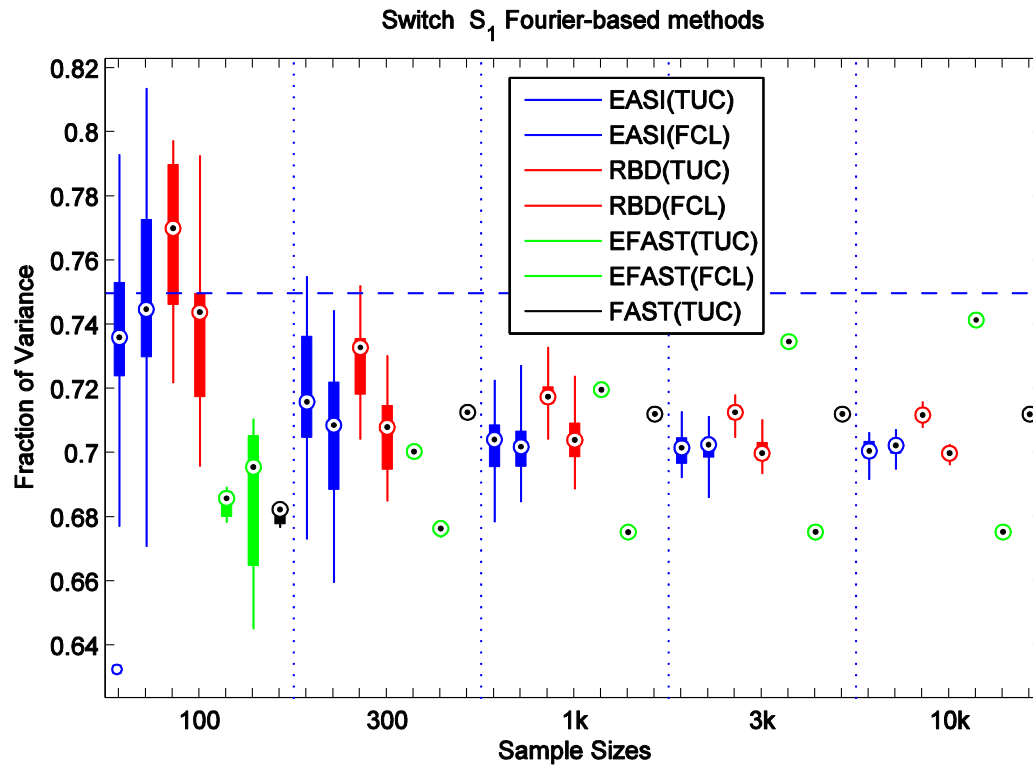


Figure 22 Fourier-based methods for the Switch example.

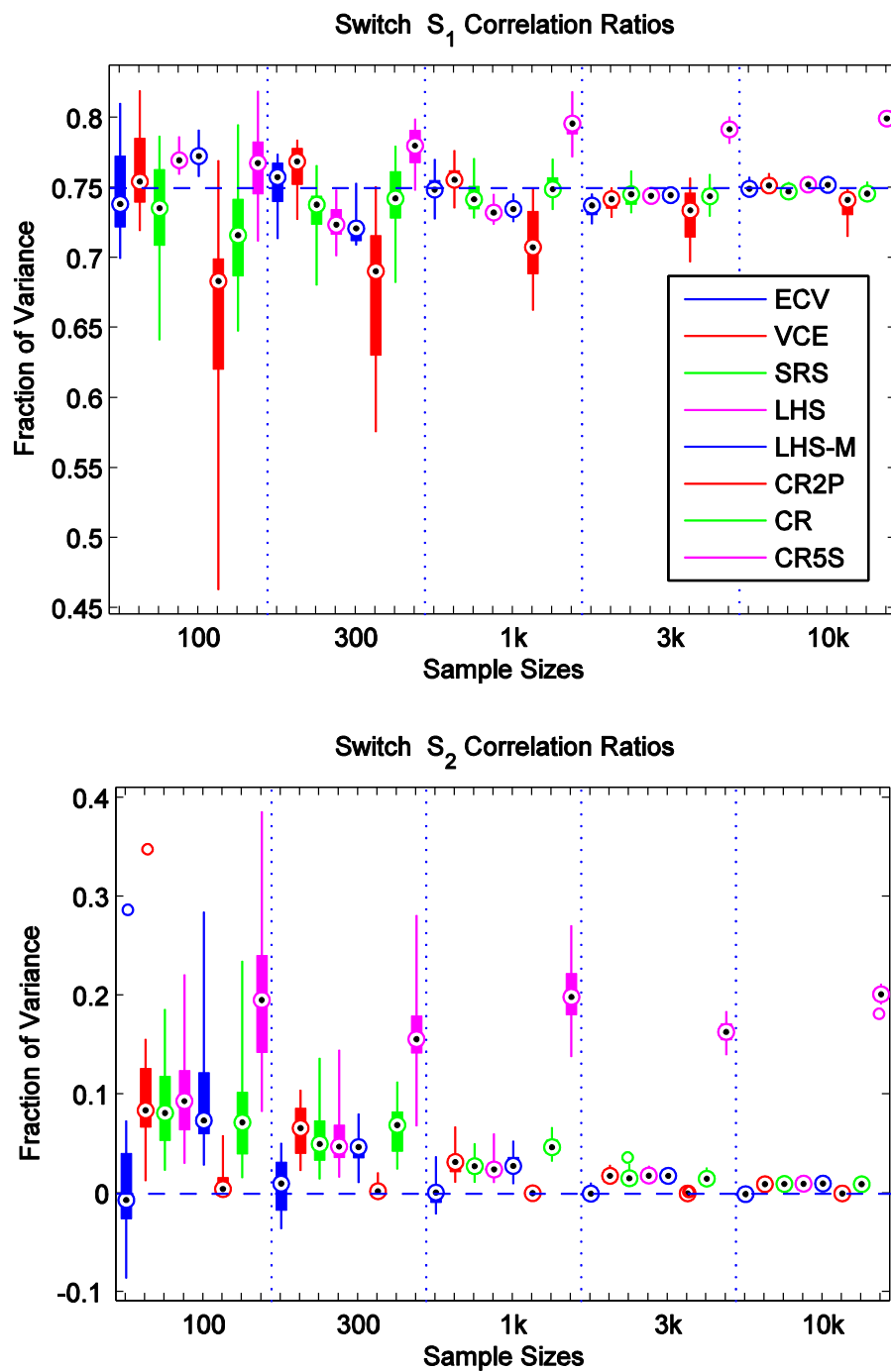


Figure 23 Correlation Ratio methods for the Switch example.

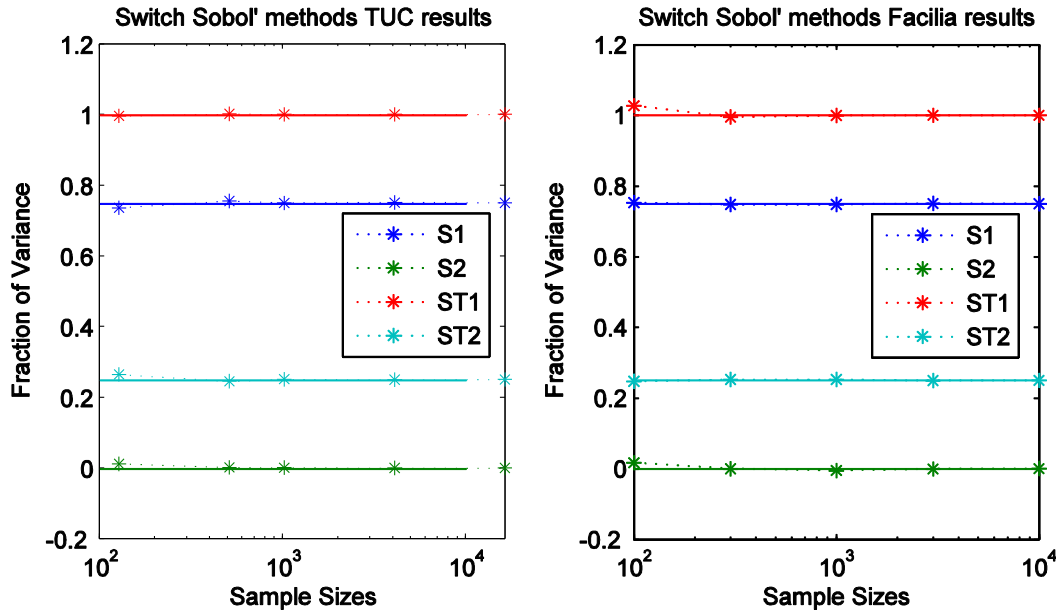


Figure 24 Sobol' method for the Switch example.

3.1.3 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 extent 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

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

The expected values are $S_i = \frac{\text{Var}[E[Y|X_i]]}{\text{Var}[Y]} = \frac{13}{14} = 0.9286$ hence $S_{Ti} = \frac{1}{14} = 0.0714$. In a linear model with independent parameters we would expect

- the sum of all main effects to be 1, and
- the total effects to be larger or equal to the main effects.

However, we are not dealing with an independent input parameter setting. Figure 25 and Figure 26 show the results for the Sensitivity Indices S_1 and S_2 . Since the model is symmetric in x_1 and x_2 , these results should be the same for both values. 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 $p(X_2 | X_1)$ or $p(X_1 | X_2)$ 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 27 shows a selection of CR methods.

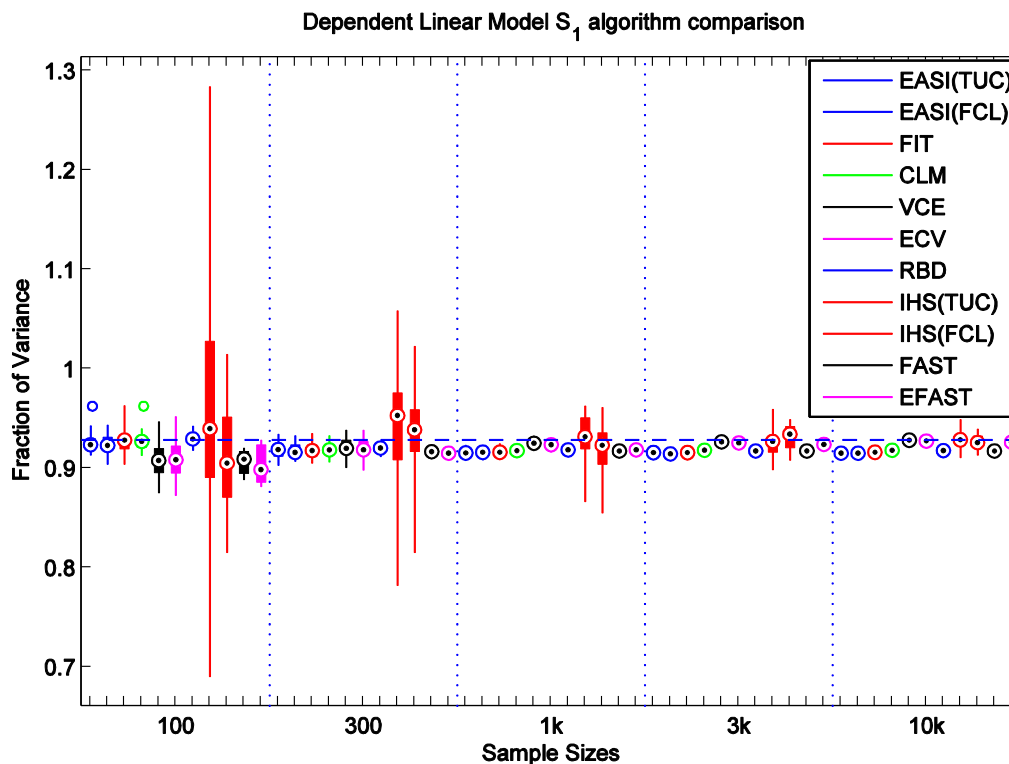


Figure 25 TUC and Facilia results - Box plots for S_1

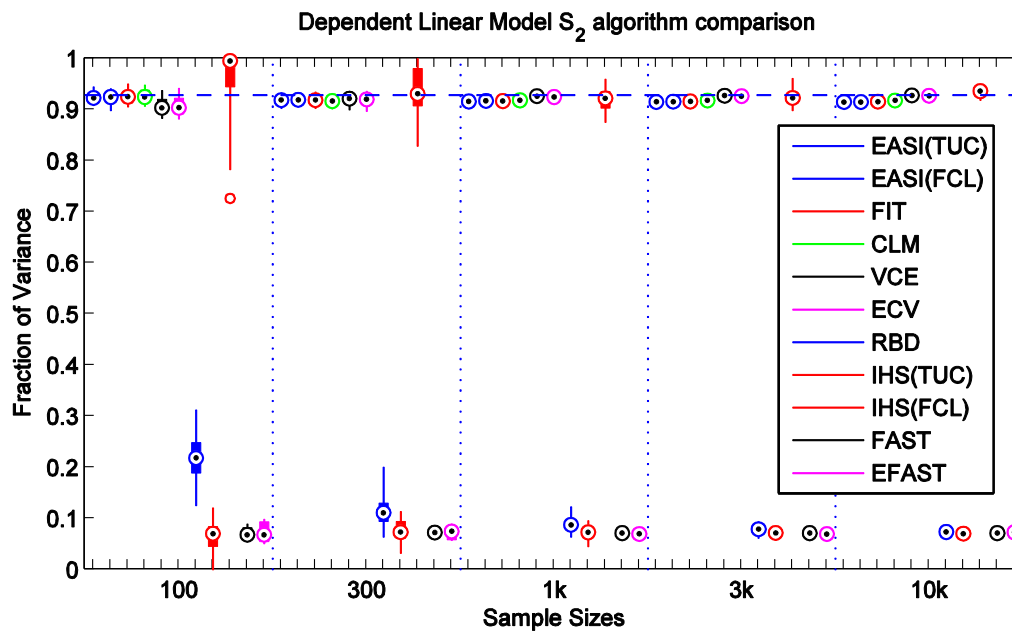


Figure 26 TUC and Facilia results - Box plots for S_2

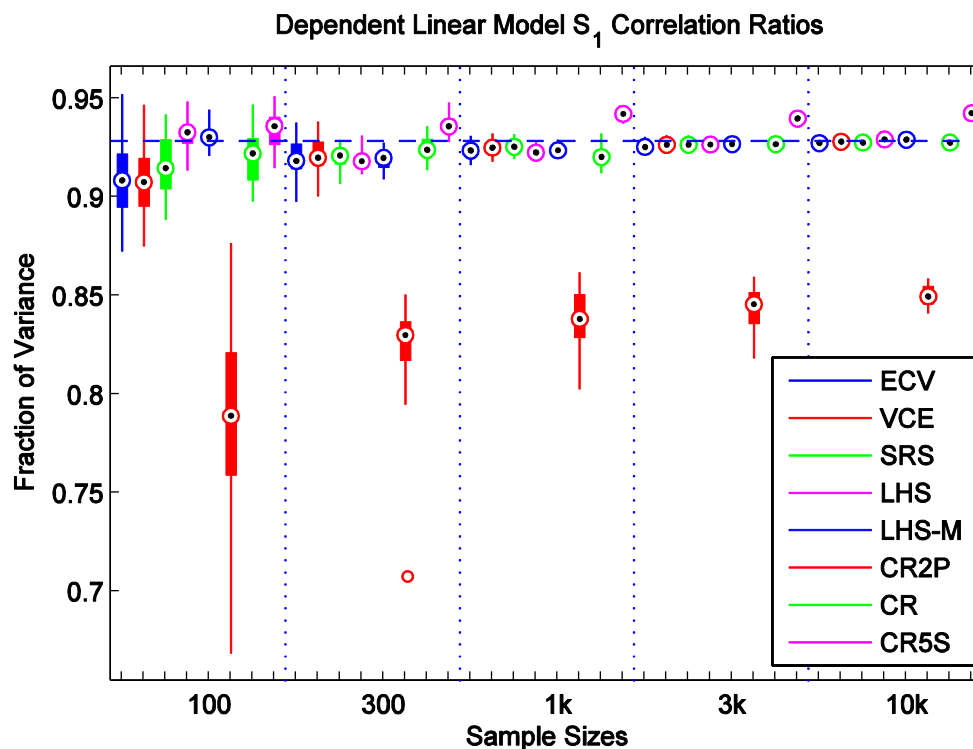


Figure 27 Correlation Ratio methods for the dependent linear model.

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 esti-

mates the “wrong” parameter transformation is now in effect for S_{T1} while S_{T2} is estimated correctly (not shown).

3.1.4 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

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

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.

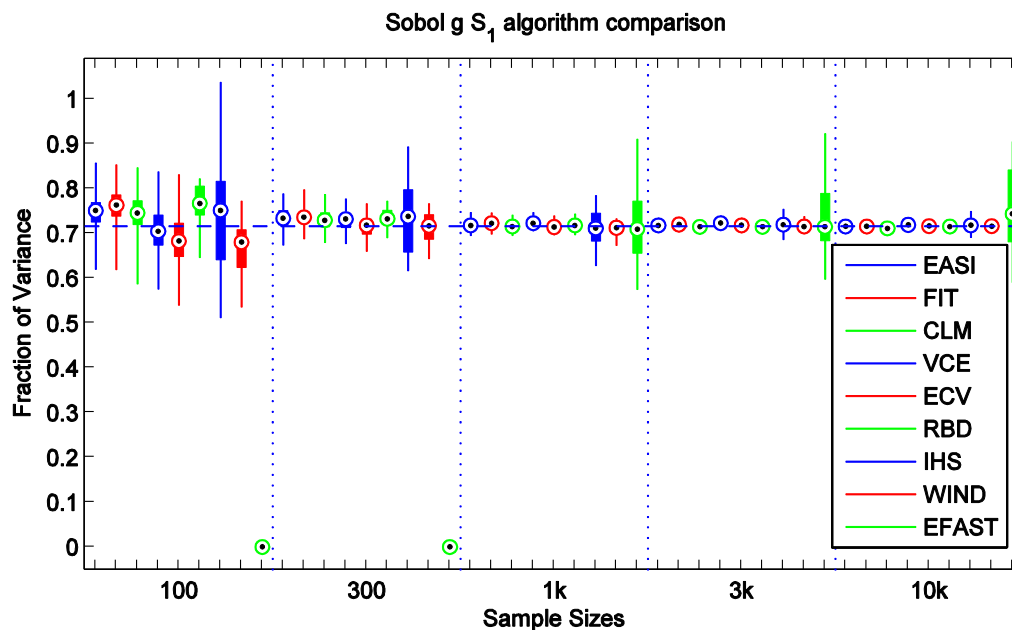


Figure 28 TUC results - Box plots for S_1

The results for S_1 are reported in Figure 28. 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 $2kM(2M + 1) \geq 366$ realisations (for $M \geq 3$) to work. Hence the first two sample sizes allow no EFAST(TUC) analysis. For the other sample sizes, the 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 29.

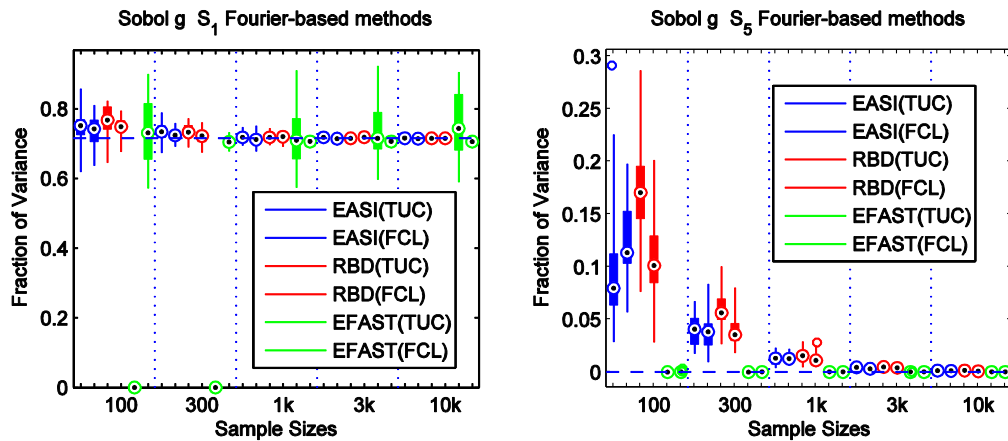


Figure 29 Fourier-based methods for Sobol' g function

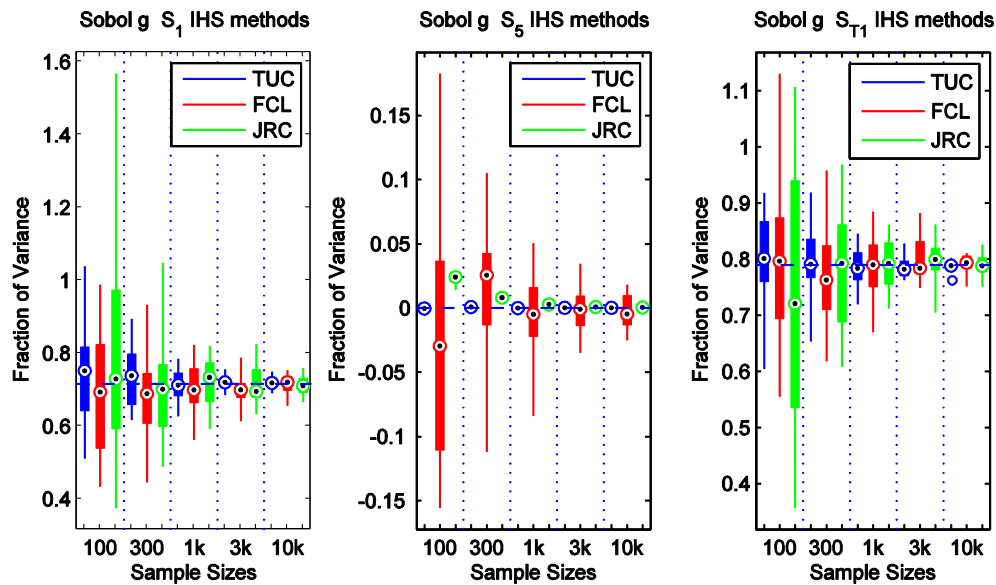


Figure 30 Ishigami-Homma-Saltelli algorithm comparison for Sobol' g function

The results for S_5 are reported in Figure 31. 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.

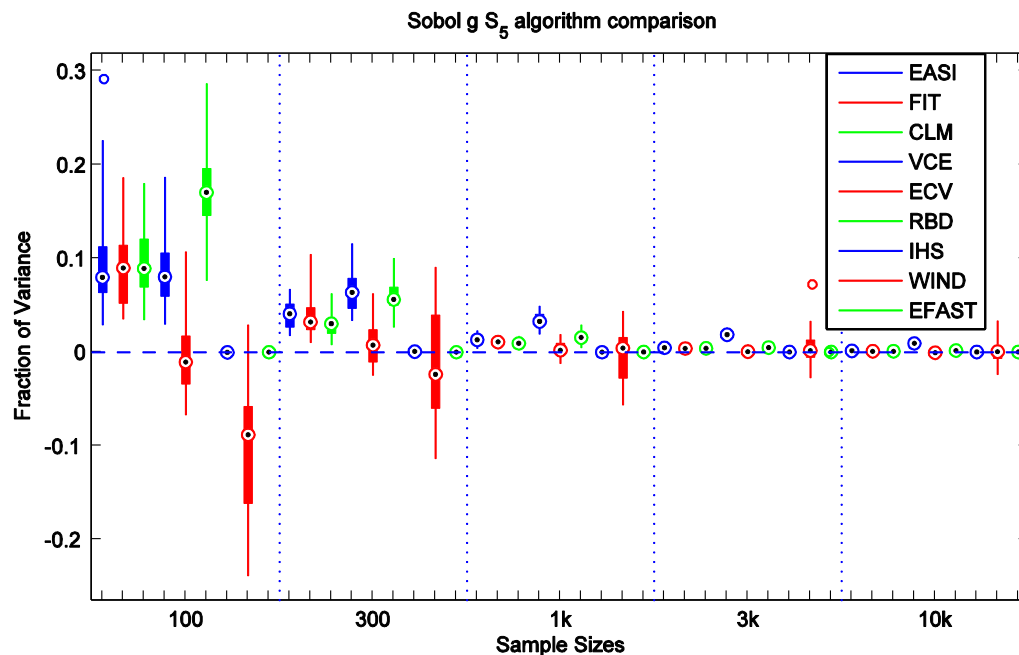


Figure 31 TUC results - Box plots for S_5

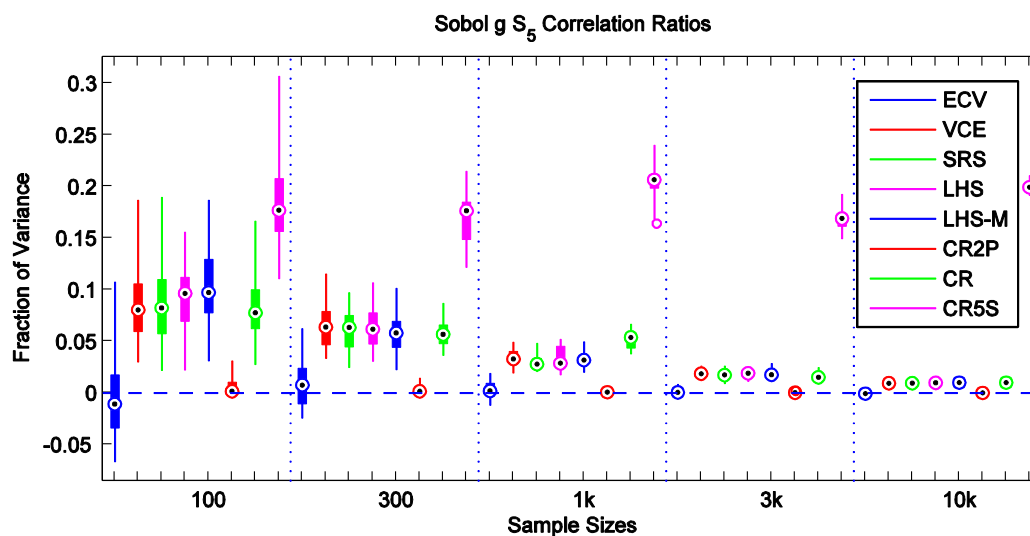


Figure 32 Correlation Ratio methods for Sobol' g function

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 32 where S_5 is estimated.

There are noticeable differences in the performance of the different implementations of the IHS method, see Figure 30. 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.

3.2 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 [Tarantola et al. 2006] and [Saltelli & Tarantola 2002] for a review), the PSACOIN Level E code (PSAC 1989) 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 [Prado-Herrero 2005], where further information including a mathematical description of the GTM Level-E model is available. The supplied binary model outputs the dose rate in Sv/a due to Iodine-129 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 [Badea & Bolado 2008]. 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,

- 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).

Table 1 Uncertain input parameters for GTM Level-E.

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

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 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.

3.2.1 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.

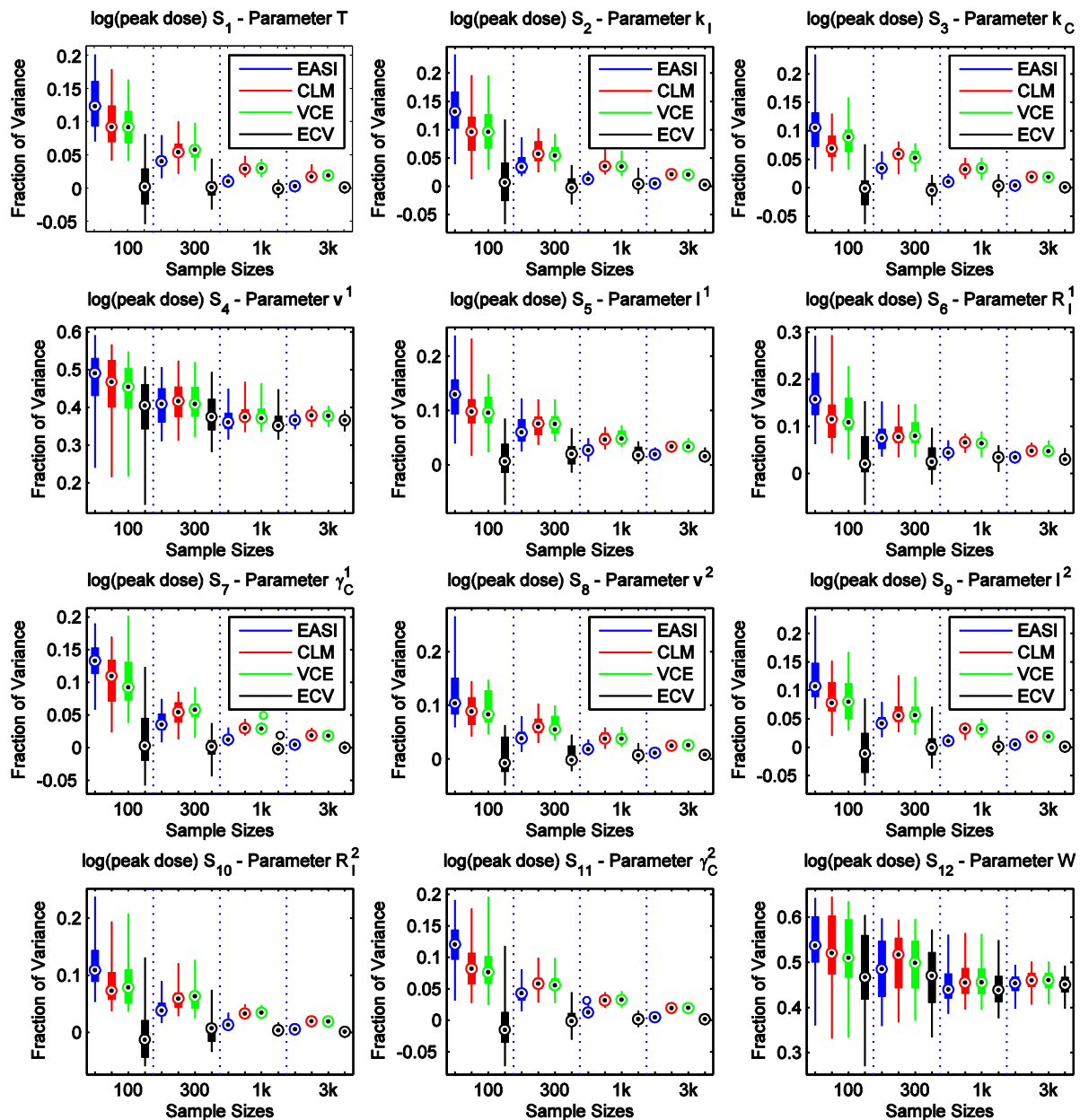


Figure 33 Cheap Sensitivity Analysis of log (peak dose rate).

Figure 33 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 Figure 33).

The untransformed data were analysed by Facilia, see Figure 34 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.

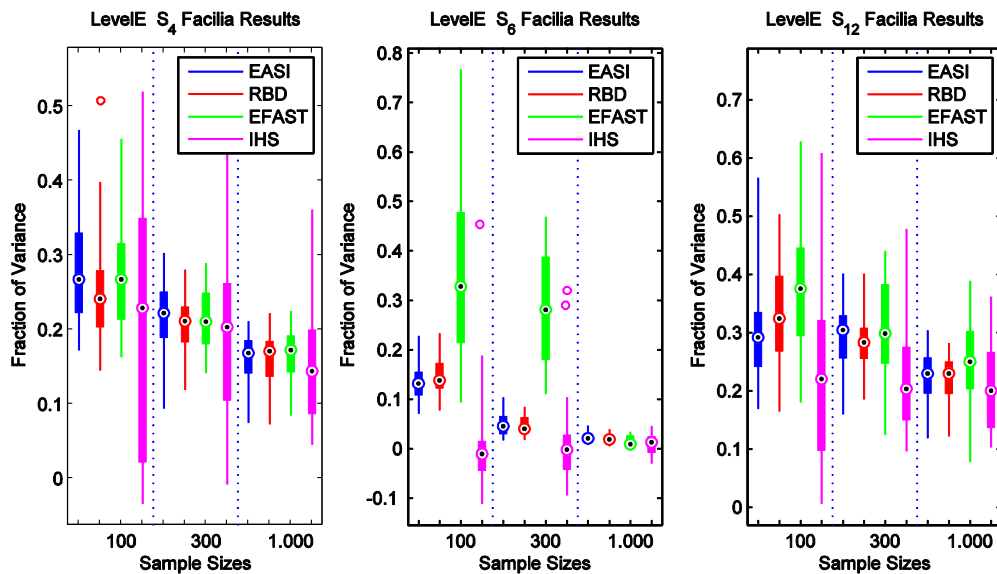


Figure 34 Sensitivity Analysis of the peak dose rate, parameters v^1 , R_I^1 and W

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 35 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 36, 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.

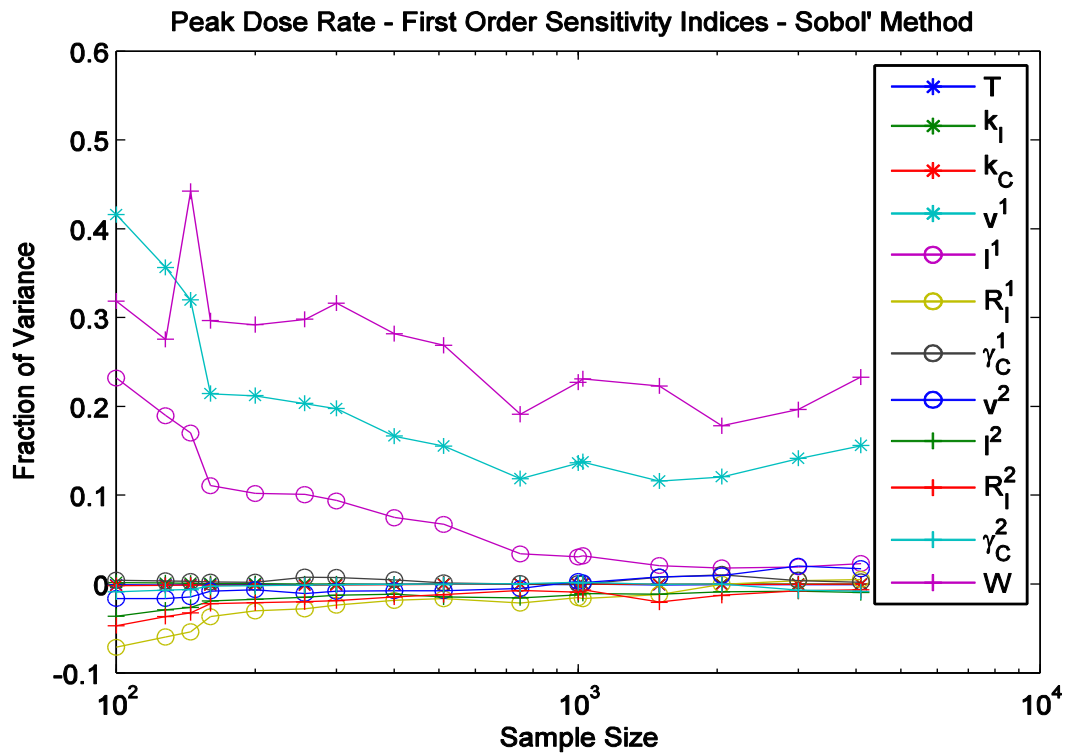


Figure 35 Sobol' method for different sampling sizes, main effects.

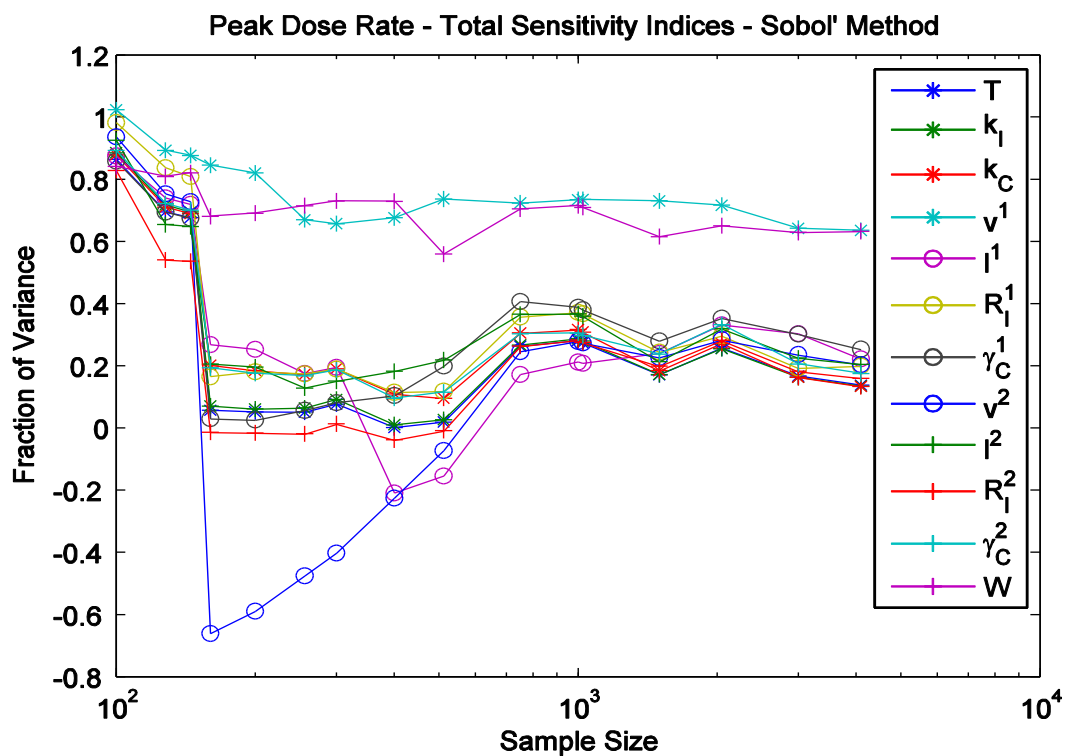


Figure 36 Sobol' method for different sampling sizes, total effects.

The total effects as computed by Facilia for the parameters v^1 , R_l^1 , v^2 and W can be found in Figure 37. As for the Sobol' method, we encounter problems with the IHS method. The

total effects from the parameter ν^1 show a large negative outlier for sample size 300, and from the analysis of the parameters ν^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 converge for ν^1 , ν^2 and W , while the results for R_l^1 show sudden changes between sample sizes 300 and 1000.

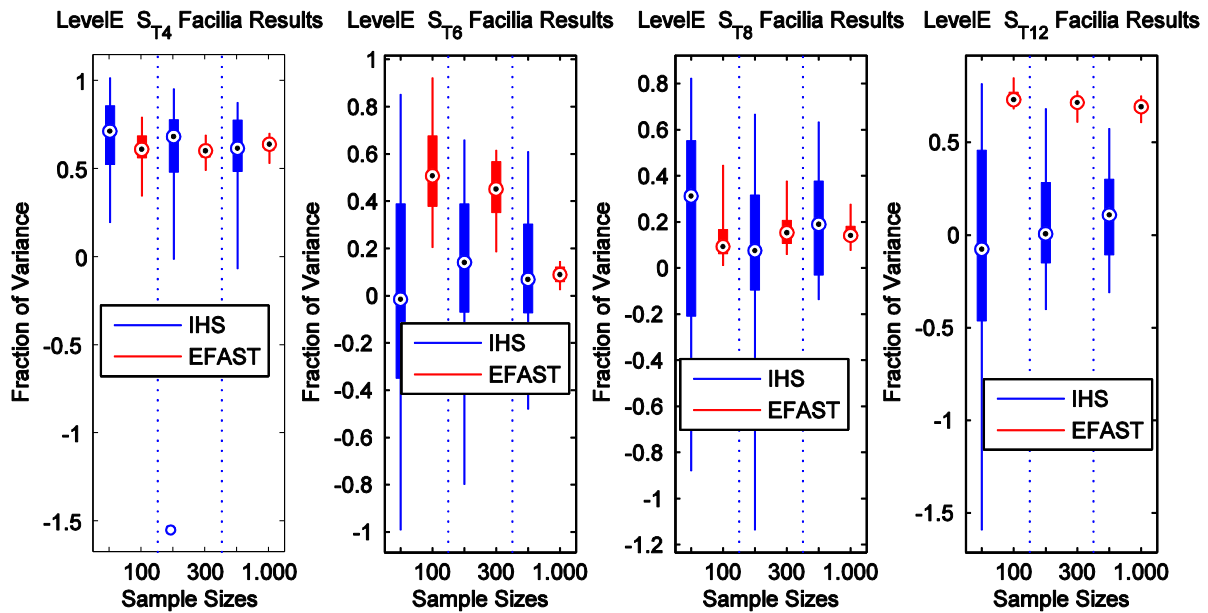


Figure 37 Total effects for the peak dose rate, Facilia's results.

3.2.2 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 38 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 ν^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 39 shows the results for the above-mentioned parameters, indeed showing a subtle, but noticeable influence. Facilia's results are displayed in Figure 40. It can be seen that S_γ (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.

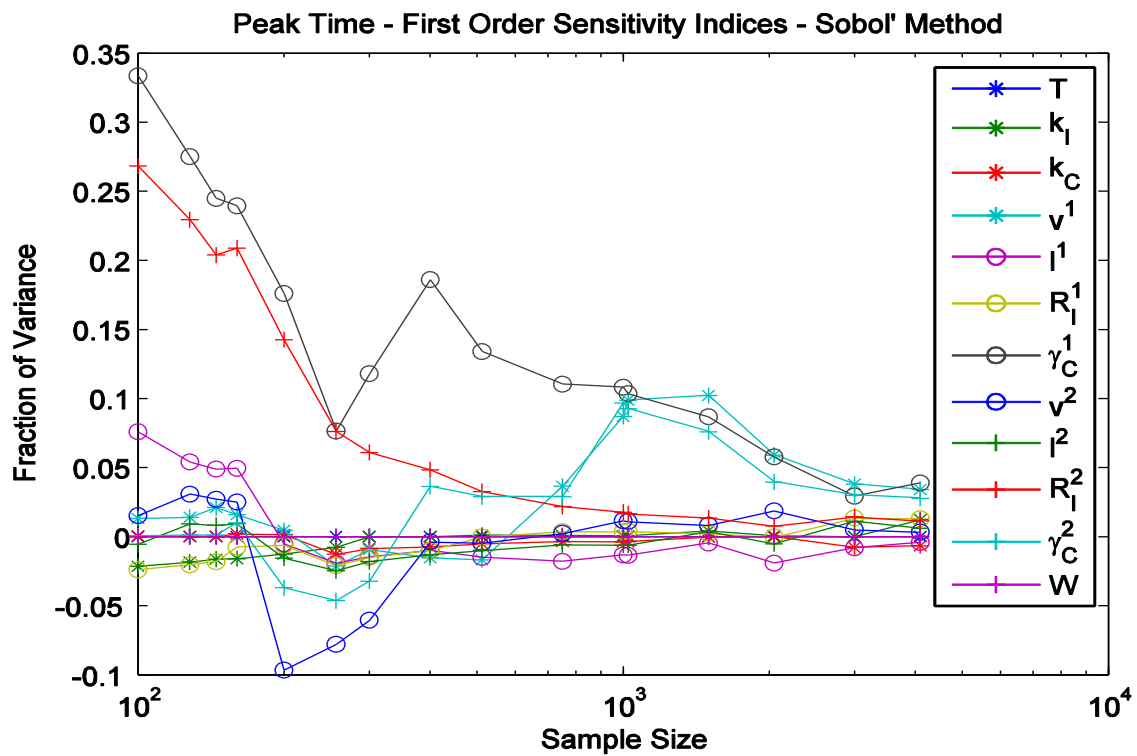


Figure 38 Sobol' method for different sampling sizes, main effects.

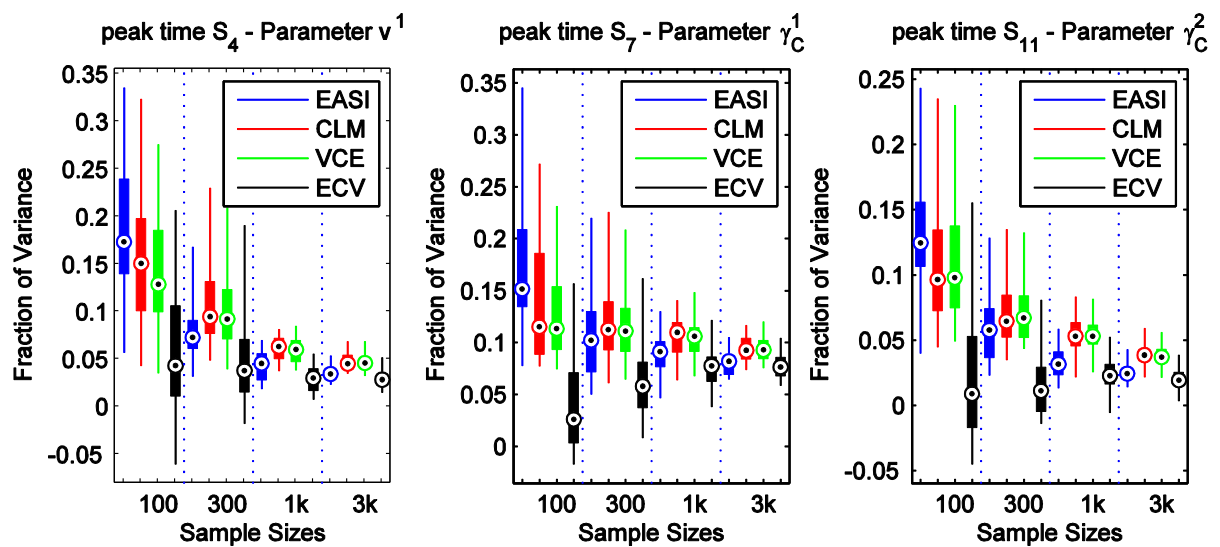


Figure 39 Cheap Sensitivity Analysis of the time of the peak dose rate.

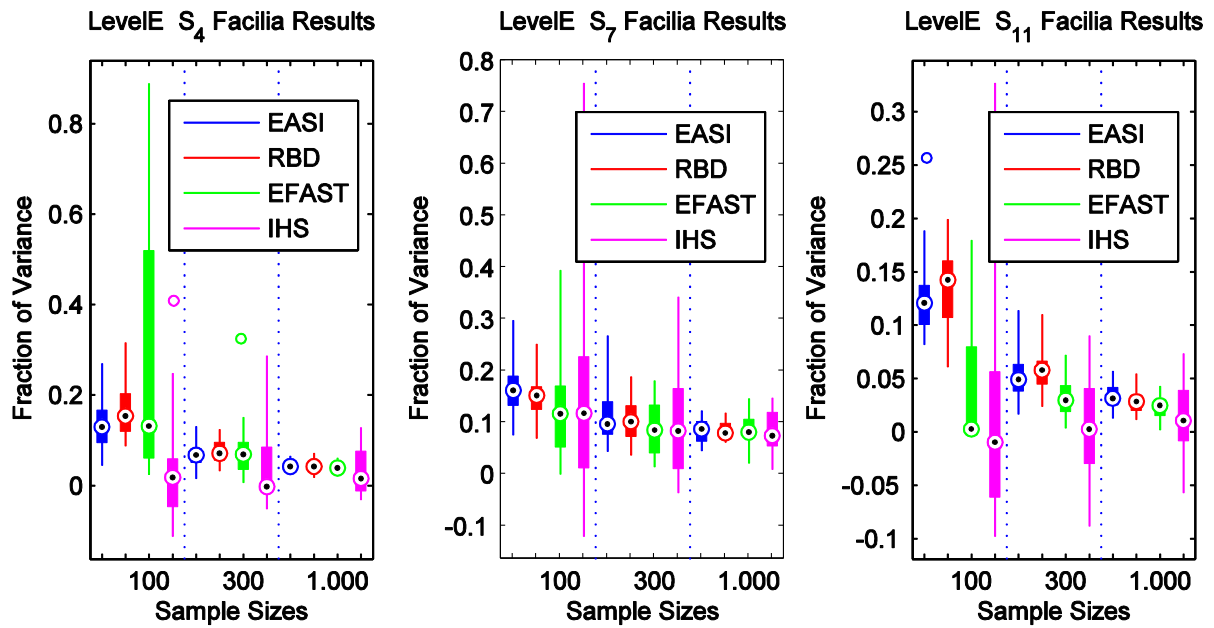


Figure 40 Sensitivity Analysis of the time of the peak dose rate, parameters ν^1 , γ_c^1 and γ_c^2

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 results. For this, let us first comment on the histogram for the peak time, see Figure 41. 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.

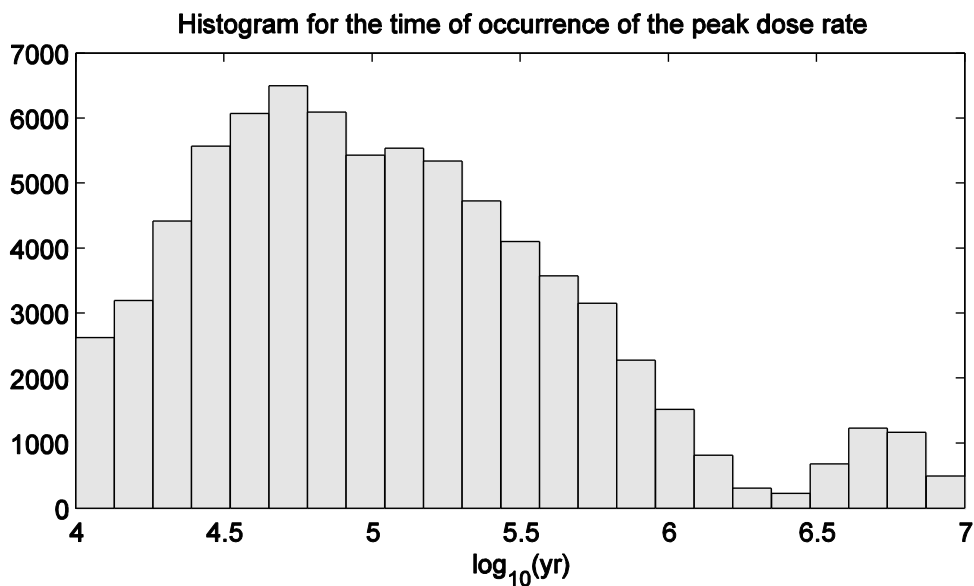


Figure 41 Histogram for the time of the peak dose rate

Fig-

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” maxima. 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 $R_I^1 \approx 7\%$ $R_I^1 \approx 7\%$, 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.

3.2.3 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 42 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 much 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 43. Although the total amount of model runs is $4096(k + 2) = 57,344$, 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 42 which uses a just 30% larger sample.

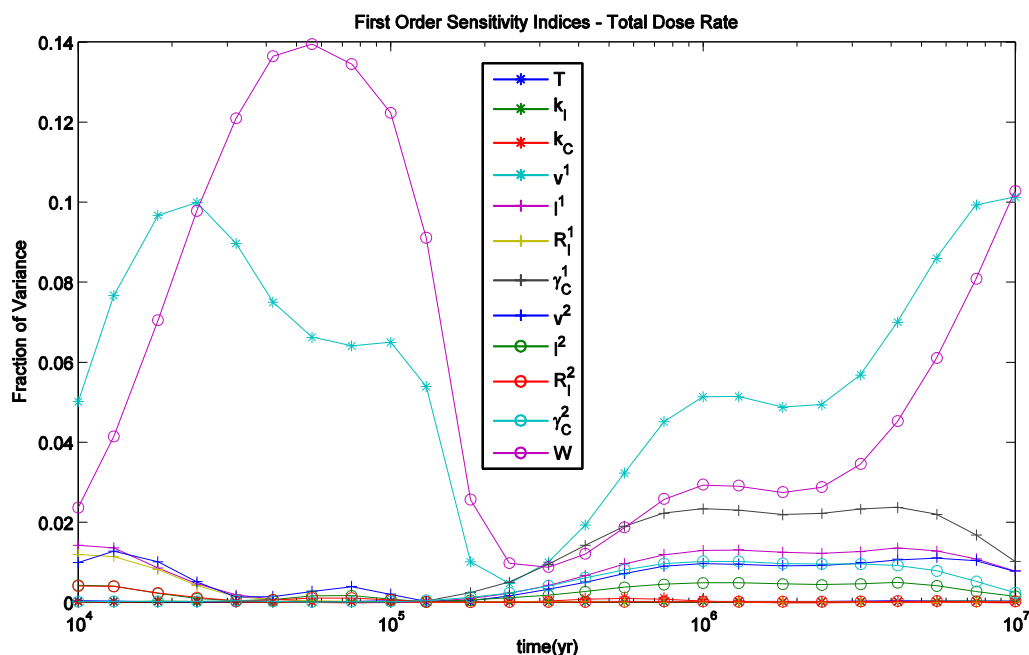


Figure 42 Time-dependent SA of the total dose rate based upon 75,000 realisations.

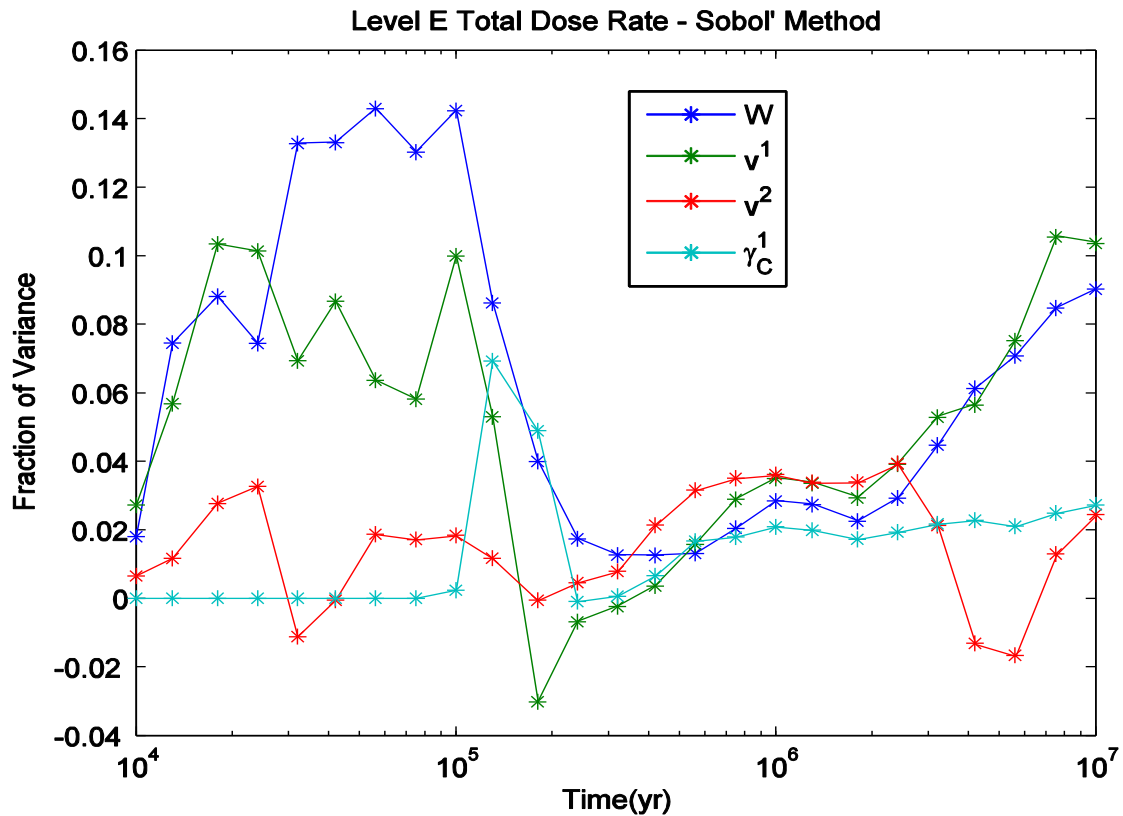


Figure 43 Time-dependent main effects of the total dose rate with the Sobol' method.

Let us now consider an analysis based upon the 3000 realisations for a cheap method.

Figure 44 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 an ECV correlation ratio method with 55 subsamples per partition. The means and the medians are nearly indistinguishable, and the whole analysis looks sound.

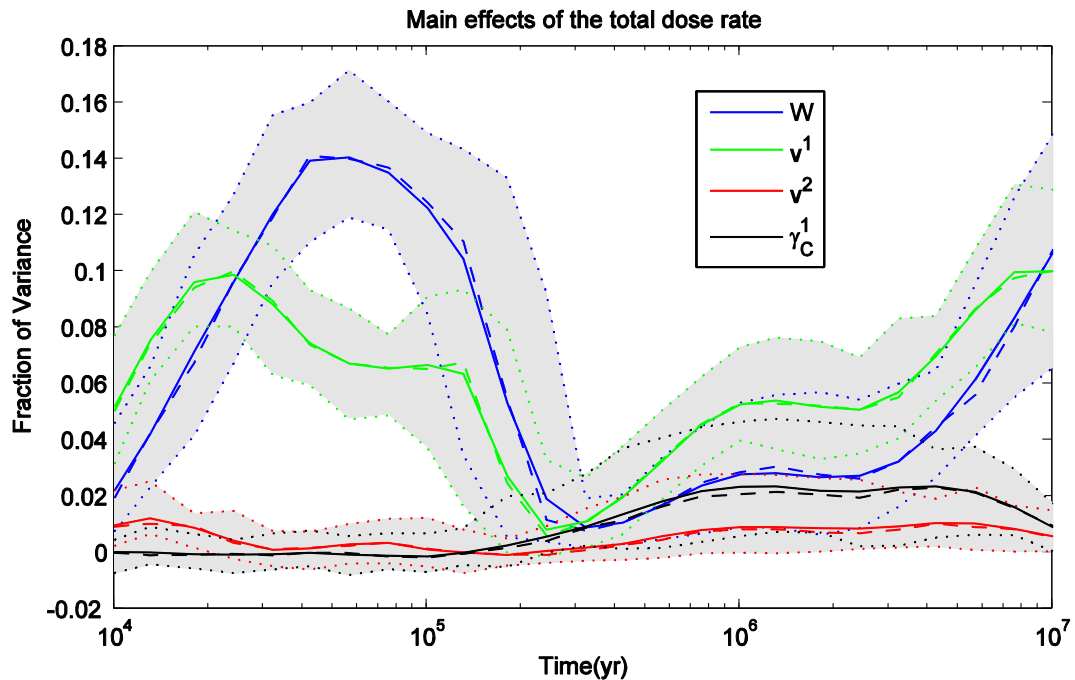


Figure 44 Min, Max, Mean and Median of the main effects for the total dose rate.

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

Figure 45 and Figure 46 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 Figure 44, 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).

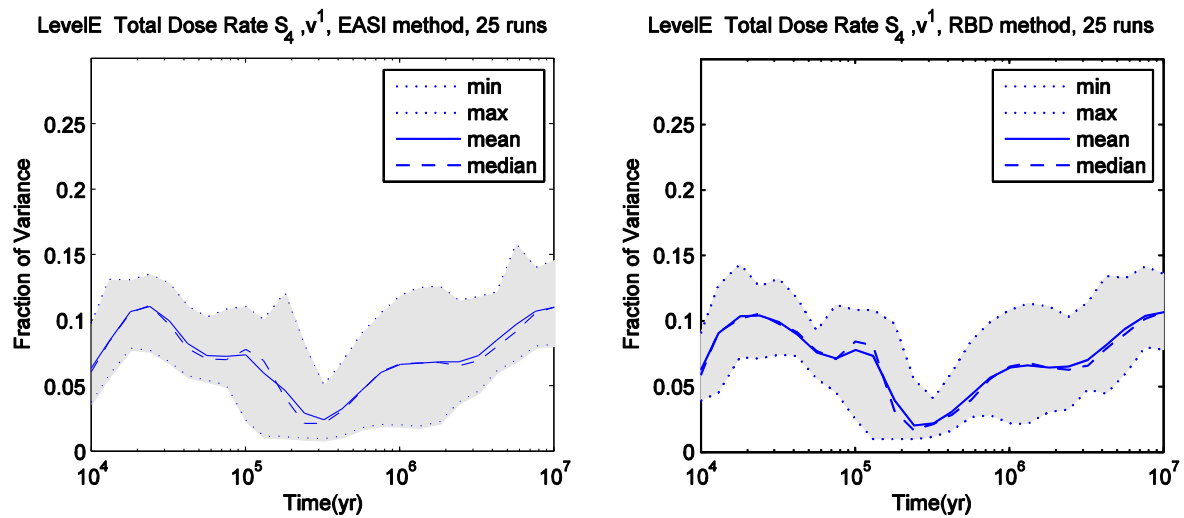


Figure 45 Statistics of the main effects for the total dose rate, EASI and RBD methods.

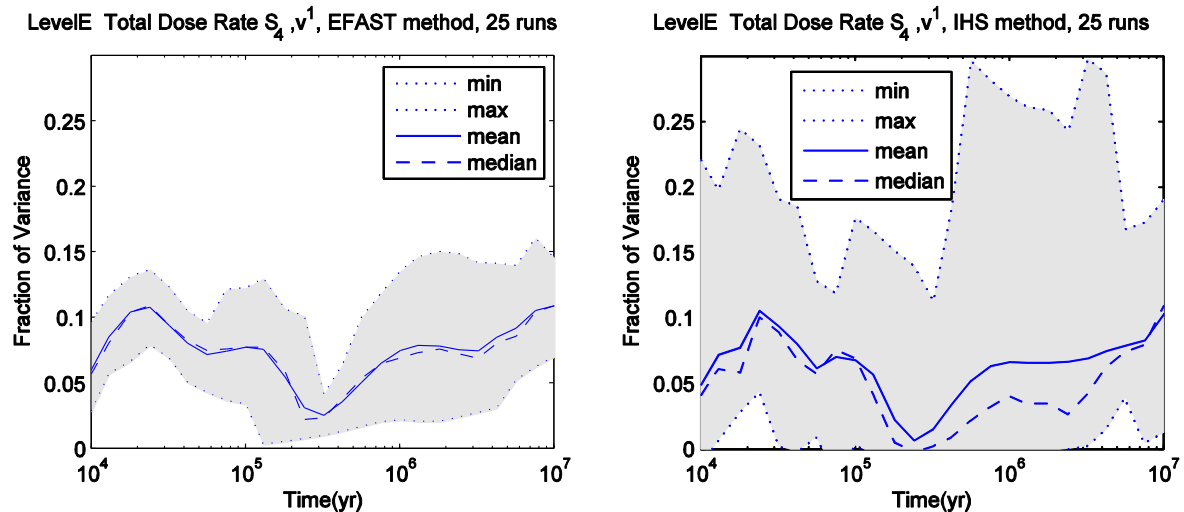


Figure 46 Statistics of the main effects for the total dose rate, EFAST and IHS methods.

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.

3.3 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.

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, the 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 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 negligible.
- 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 the analytic 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.

4. Application of Sensitivity Analysis methods to realistic repository PA models

This chapter is about application of different SA methods to realistic repository model systems. Such systems usually show a much more complex behaviour than the analytic models considered in chapter 3, and even the Level E test case. Consequently, it is harder to interpret the results correctly. In contrast to the benchmark exercise with analytical models, which showed how the different methods react to specific known model properties and how they reflect the sensitivity of systems with a known behaviour to parameter variations, the investigations presented in this chapter focus to systems whose behaviour is not known a priori, at least not in detail. This can lead to results that are sometimes surprising and hard to interpret.

In total, six systems were analysed: two salt systems (NRG, GRS), two clay systems (NRG, JRC/ANDRA) and two granite systems (Enresa, Facilia). Different SA approaches were applied by the different organisations.

For the Dutch salt concept NRG calculated Pearson's correlation coefficients (PEAR), Spearman's rank correlation coefficients (SPEA), standardised regression coefficients (SRC), standardised rank regression coefficients (SRRC), partial correlation coefficients (PCC) and partial rank correlation coefficients (PRCC) for six model parameters. The clay concept model was analysed for three parameters using SRC and SRRC and with graphical methods (CSM plots, cobweb plots).

ENRESA investigated the sensitivity of the Spanish granite concept model with 135 statistically distributed input parameters, calculating regression coefficients for the values and for the ranks (PEAR and SPEA). Moreover, graphical methods (CSM plots, cobweb plots) were applied.

GRS performed a variance-based sensitivity analysis for a generic German salt concept with six statistically varied parameters. This model has the specific property of yielding zero output for 85 % of all runs, which makes it somewhat hard to analyse. Additionally, correlation coefficients (PEAR, SPEA) and the Smirnov test were calculated and graphical methods (CSM plots, scatterplots, cobweb plots) were applied.

Facilia made a sensitivity analysis study on a biosphere model using a landscape model developed for the Olkiluoto Island. A number of biosphere-related parameters were considered. The Morris screening method was applied to identify potentially important parameters. Two Monte Carlo simulations with 10000 iterations each were carried out and analysed using different methods. The most important parameters were additionally investigated with the EFAST method.

JRC, in co-operation with ANDRA, analysed a model for release of radionuclides from an ILW disposal cell in a generic French clay site. 24 uncertain parameters were taken into account. The sensitivity analysis was done by calculating SRC and SSRC.

4.1 NRG application case: rock salt repository

The underlying study assessed the performance of the rock salt plug that seals the waste disposal cells in a rock salt based repository and focused on the uncertainties associated with the compaction rate of salt as a result of the back pressure of the overlying host rock and the relation between the porosity and the permeability of compacted salt grit. Two cases are looked at, a “dry” scenario and a “brine intrusion” scenario. Under dry conditions no brine will be present in the repository and therefore no radionuclides are able to leave the repository. However, in the event of an intrusion of brine into the disposal cell, brine may get contaminated by radionuclides. If the waste canisters inside the disposal cell are damaged due to compaction or corrosion, small fractions of vitrified waste can dissolve in the brine. As long as the sealing plug contains open flow paths, the convergence of the host rock may drive contaminated brine out of the disposal cell and consecutively also out of the repository. Although this risk is assumed to be extremely small in the case of salt-based repositories, it can not be excluded. The potential effects of this event are calculated in the “brine intrusion” scenario, assuming that the repository is instantaneously flooded.

4.1.1 Description of the system and the model

For performing the probabilistic uncertainty analysis for PAMINA, a generic repository design similar to that of the EVEREST [Becker et al. 1997] concept is adopted, whereas the geometry of the disposal cell is based on the Torad-B design [Poley 2000]). In order to calculate the compaction behaviour of the disposal cell plugs and the backfill, as well as the dissolution, migration and human exposure by radionuclides, the computer program EMOS has been applied.

The conceptual EVEREST repository design [Becker et al. 1997] for high-level waste (HLW) and medium-level waste (MLW) has a horizontal extension of 14 x 4 km² and is located in a salt dome at a depth of 840 m. The central field consists of shafts, underground working places and warehouses, and galleries which connects the vertical and horizontal disposal cells. Upon the termination of the disposal operations, the central field will be backfilled. The drifts which lead from the central field to the shafts will be sealed with dams.

The Torad-B borehole concept (Figure 48) has been developed as a blueprint of a deep underground facility for the disposal of large quantities of heat-generating high-level radioactive waste in deep lined vertical boreholes in rock salt serving as disposal cells. This concept accounts for retrievability of the waste by adding a liner to the borehole gives an overview of the Torad-B borehole concept. After emplacement of the waste canisters the boreholes are closed with a compacted salt grit plug.

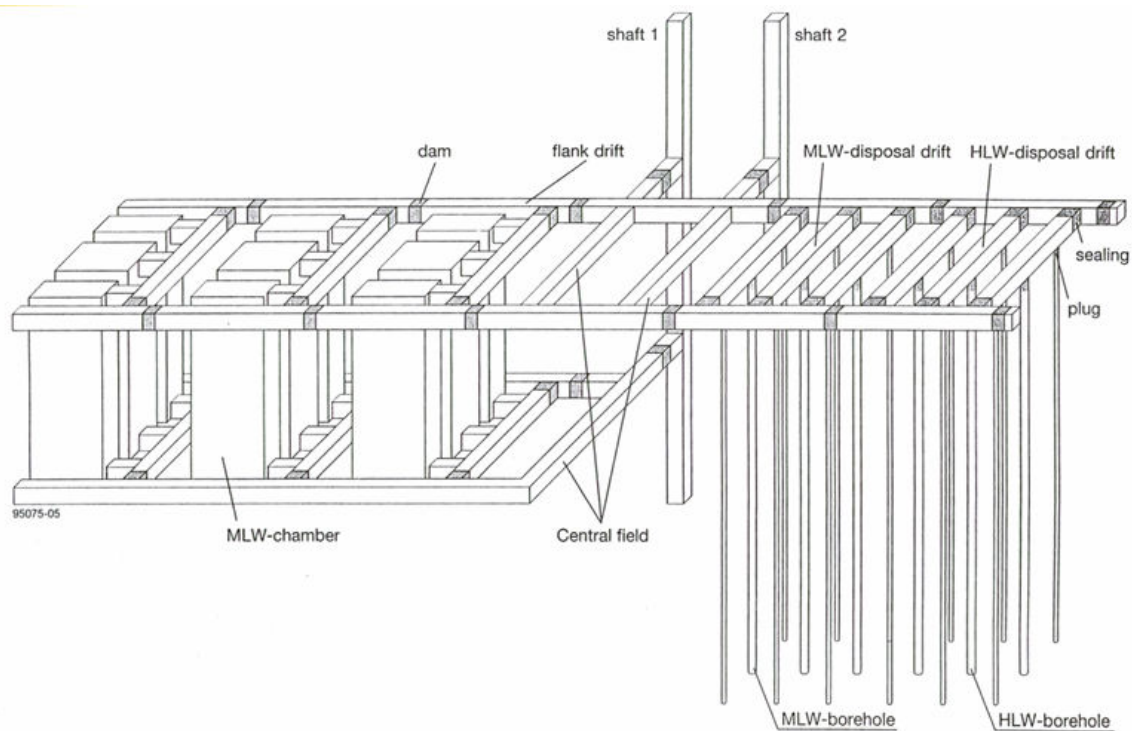


Figure 47 Model of the EVEREST repository design.

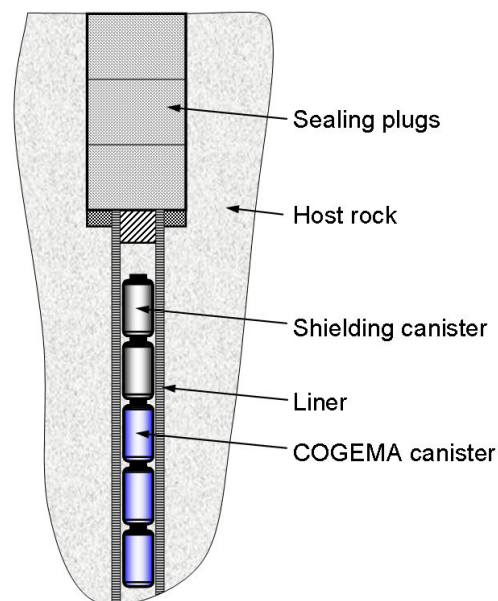


Figure 48 Overview of the Torad-B borehole concept.

Due to the pressure of the surrounding host rock and the heat developed in the waste, the host rock will start to creep and the borehole plugs will be pressurized and compacted in time. This leads to a reduction of the porosity and permeability of the salt plugs as time progresses.

The compaction behaviour of salt grit and the relation between the porosity and the permeability are key-processes for the safety assessment of a salt-based repository and has been the main subject of the underlying study. Experiments have shown that when the porosity of rock salt has decreased to a value of about 5%, the intragranular pore network in compacted salt will be disconnected and the permeability will decrease significantly. The threshold value, below which the compacted salt grit may be considered as impermeable, is also referred to as percolation threshold.

The computer program *EMOS (Endlagerbezogene Modellierung von Szenarien)* [Storck et al. 1990] has been applied to assess the impact of the borehole plug compaction on the repository safety. EMOS has originally been developed by GSF, Germany, for the assessment of radionuclide dissolution in a nuclear waste repository in rock salt, the subsequent nuclide transport from the waste containers through the repository into the geosphere, and the distribution of the radionuclides in the biosphere.

In the present study an EMOS model of a simplified repository design based on the Torad-B concept has been developed (Figure 49), consisting of three segments:

- the waste canisters
- the disposal cells
- the central field

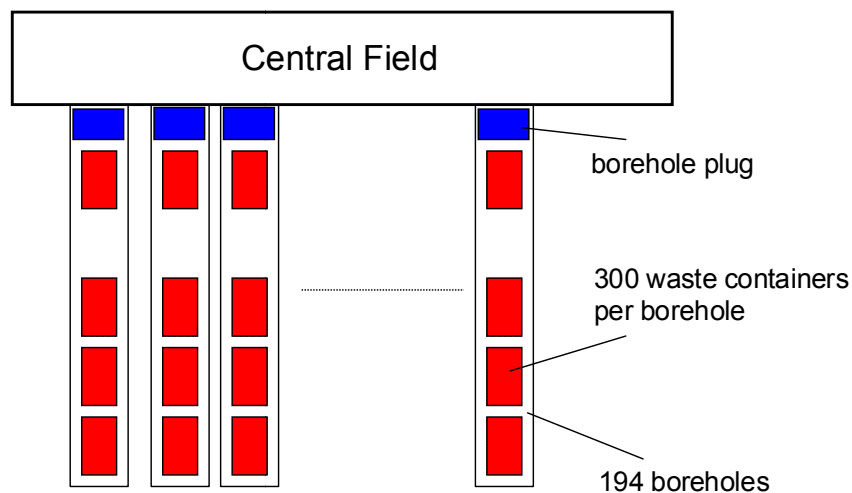


Figure 49 Compartment representation of the Torad-B facility design

From the 194 boreholes of the Torad-B concept, for the simplicity only one borehole is modelled in this study, assuming the case of a brine intrusion in one borehole only and a consecutive instant failure of all 300 containers. For the canisters containing vitrified high-level waste (HLW), a COGEMA type has been assumed.

In the original source code of *EMOS*, several modifications have been implemented by NRG/ECN, mainly with respect to the modelling of the transient convergence and compaction behaviour of crushed salt [Heijdra et al. 1995]. The most recent modification has been implemented in the framework of the PAMINA project [Schröder 2008b].

4.1.2 Objectives and peculiarities of the study

The objective of the study was to assess the impact of the uncertainties of the compaction behaviour of the compacted salt grit borehole plug on the safety of a salt-based geological repository. This was accomplished by variation of the input parameters of the *EMOS* modules that describe the compaction behaviour as well as the relation between the porosity and permeability of the salt grit. Output parameters relating to the safety of the model repository were analysed with different statistical methods in order to establish the relative importance of the model input parameters on calculated safety indicators.

A peculiarity of the present study was that the *EMOS* simulations were not necessarily terminated at the same end-points. Because the simulations of the disposal cell were terminated when the porosity of the sealing plug reached the threshold value of 0.0003 (i.e. the percolation threshold), different simulation times occurred for the set of calculations. From the time of termination of a particular simulation, no values of output parameters were available for that simulation to perform the statistical analysis. As a consequence the number of time steps and output parameters available for the sensitivity and uncertainty analysis decreased progressively for the sets of 1000 simulations.

For the sensitivity analysis, three options for the replacement of the missing output values were studied: 1.) no replacement of missing values, 2.) missing-value replacement by a threshold value, and 3.) missing-value replacement by the last available value of a data series.

4.1.3 Varied input parameters and analysed output parameters

When an external pressure is exerted on a rock salt core, the salt core will deform as a result of the viscoplastic properties of rock salt. This phenomenon is referred to as compaction creep. By compaction creep, empty spaces between the salt grains will reduce in size, resulting in a compaction of the core and a decrease of the porosity. The compaction creep phenomenon was one of the main subjects of the NF-PRO project, e. g. [Zhang and Grupa 2006], in which it was established that compaction creep depends on two mechanisms, namely:

- (dislocation coupled) recrystallization creep
- pressure solution creep

These two mechanisms are independent creep mechanisms that work simultaneous.

During the *NF-PRO* project the so-called ‘*Coupled Creep Model*’ (*CCM*) was developed describing the compaction rate of salt grit [Grupa and Houkema 2000]. The *CCM* is based on mechanistic considerations but also needed to be parameterized by the measurement of several model parameters during compaction experiments. Another important parameter in this respect is the permeability of the compacted salt which appears to be a function of its porosity.

The parameterization of the models that describes the creep rate of salt grit and the relation between its porosity and permeability is subject to some degree of uncertainty. For example, Figure 50 shows experimental data from various sources including data determined within the EC Framework *NF-PRO* project [Spiers et al. 2005]. Usually, a “*best fit*” is used to parameterize a model like the porosity-permeability relation (solid line in Figure 50). However, the variability of the measured data leads to an uncertainty in the parameterization of the relation. Since the compaction creep of a borehole plug and the porosity-permeability relation may influence a potential release of radionuclides from a repository in many ways, the uncertainty of these parameters affects the potential risks to man and environment in a complex and *a priori* unpredictable manner.

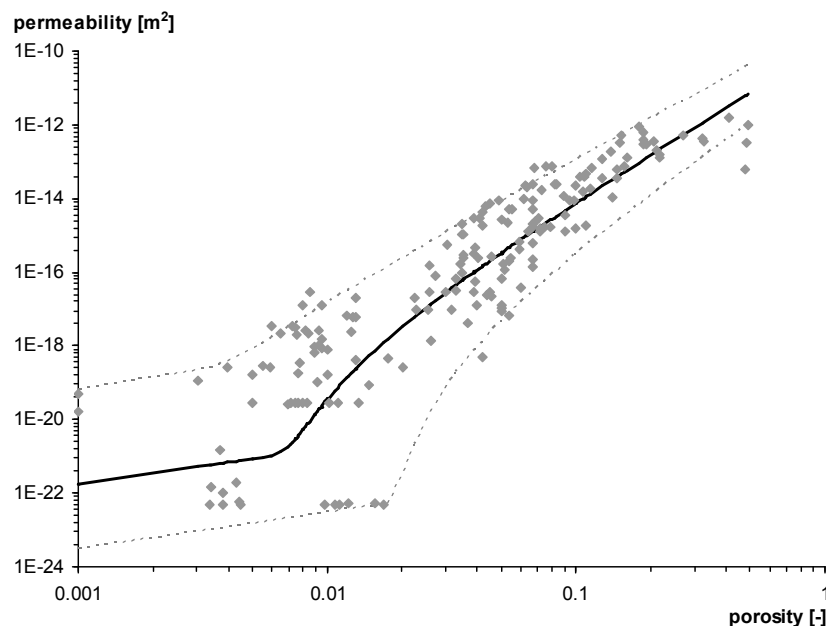


Figure 50 Permeability as a function of porosity. Solid curve: Best estimate fit model. Dashed curves represent the 95%-confidence interval

Within the underlying study the uncertainties of the creep behaviour and the permeability development of compacted salt were investigated together with the effect of these uncertainties on the safety assessment of a repository in salt rock. In this study, the following six model parameters, all related to the performance of the sealing rock salt plug, have been selected to perform the uncertainty and sensitivity analyses [Schröder 2008a]:

- B_{recr} recrystallization creep constant
- A_{ps} pressure creep constant
- F_{09} permeability constant of salt grit
- n constant, power factor in the porosity- permeability relation
- F_{46} threshold porosity below which a seal is considered “closed”
- F_{47} threshold permeability below which a seal is considered “closed”

Together, these model parameters determine the closure behaviour of the compacted salt plug and thus the migration behaviour of contaminated brine that may be pressed out through the plug of the disposal cell.

For six input parameters, relevant for the creep behaviour of the plug and the porosity-permeability relation, average values and their distribution were derived from experimental data. A sample with 1000 realizations was generated for all input parameters and used for performance assessment calculations using *EMOS* of the normal evolution scenario (the “dry case”) and the brine intrusion scenario (the “wet case”). As output, five point values and nine time-dependent output parameters were selected, all related to the safety and performance of the repository. Statistical analysis covered uncertainty analysis (mean, standard deviation, skewness, kurtosis and percentiles) as well as sensitivity analysis (R^2 , Pearson, Spearman, *SRC*, *PCC*, *SRRC* and *PRCC*).

4.1.4 SA Methods applied and main results

An uncertainty and sensitivity analysis has been performed for the indicators computed by *EMOS*). For every indicator 1000 realizations with different values of the input parameter have been computed. A first analysis showed that the distribution of most of the output parameters was strongly skewed. Therefore it was decided to perform the uncertainty and sensitivity analysis on the *log*-transformed data, too. Table 2 and Table 3 give an overview of the statistical analyses performed.

Table 2 Uncertainty indicators

		Quantitative indicators (Tables)	Qualitative indicators (Figures)
Point values		(geometric) mean (geometric) standard deviation quantiles: $q_{1\%}$, $q_{5\%}$, $q_{25\%}$, $q_{50\%}$ (median), $q_{75\%}$, $q_{95\%}$, $q_{99\%}$ skewness and kurtosis	estimation of the probability density function (pdf) empirical cumulative distribu- tion function (ecdf)
Time- dependent values	all data		plots of all realizations evolution of mean, standard deviation, quantiles, skew- ness and kurtosis
	predefined (fixed) mo- ments	(geometric) mean (geometric) standard deviation quantiles: $q_{1\%}$, $q_{5\%}$, $q_{25\%}$, $q_{50\%}$ (median), $q_{75\%}$, $q_{95\%}$, $q_{99\%}$ skewness and kurtosis	estimation of the probability density function (pdf) empirical cumulative distribu- tion function (ecdf)
	maximum values and time to the maximum values	(geometric) standard deviation quantiles: $q_{1\%}$, $q_{5\%}$, $q_{25\%}$, $q_{50\%}$ (median), $q_{75\%}$, $q_{95\%}$, $q_{99\%}$ skewness and kurtosis	estimation of the probability density function (pdf) empirical cumulative distribu- tion function (ecdf)

The following quantities were used in this study as statistical indicators for the sensitivity analysis: *Pearson correlation coefficient*, *Spearman correlation coefficient*, *SRC*, *PCC*, *SRRC* and *PRCC*. The first two coefficients are based on the computation of a correlation. A correlation provides a measure of the strength of the linear relationship between any input factor (denoted in this section by X_j) and the output (denoted by Y). For the calculation of these measures we supposed the input factors to be statistically independent.

Table 3 Sensitivity indicators

		Quantitative indicators (Tables)	Qualitative indicators (Figures)
Point values		Pearson correlation coefficient Spearman correlation coefficient SRC, PCC SRRC, PRCC R^2 , R^2_{adj}	scatter plots (output as a function of each input)
Time-dependent values	all data		evolution of Pearson correlation coefficient, Spearman correlation coefficient, SRC, PCC, SRRC, PRCC evolution of R^2 , R^2_{adj}
	maximum values and time to the maximum values	Pearson correlation coefficient Spearman correlation coefficient SRC, PCC SRRC, PRCC R^2 , R^2_{adj}	scatter plots (output as function of each input)

With *EMOS*, several output variables have been calculated for all 1000 sets of input data. Examples of the *EMOS* output are depicted in Figure 51 and Figure 52, showing the realizations of the evolution of respectively the porosity and permeability of the borehole plug. The variety of the calculated dose rate in the biosphere, which is the main safety-related output variable has been depicted in Figure 53.

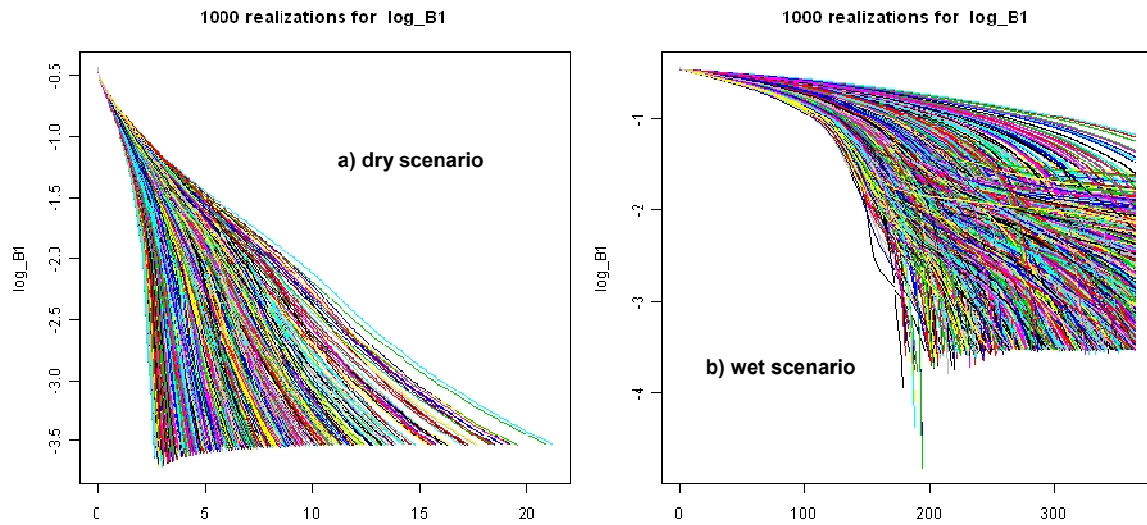


Figure 51 Evolution of the porosity of the plug [-] over time for a) the dry scenario and b) the wet scenario.

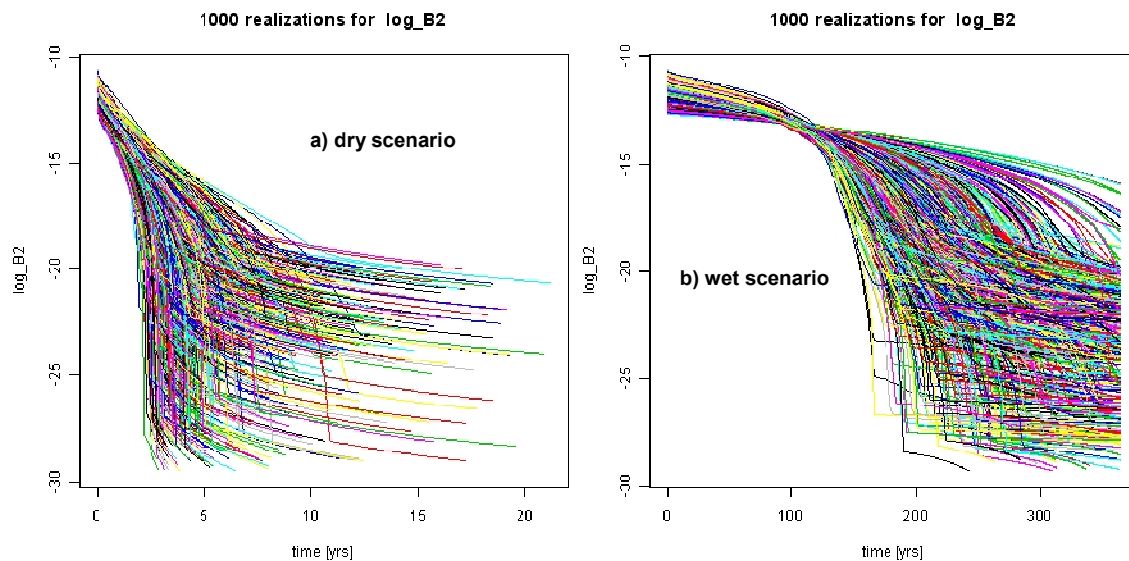


Figure 52 Evolution of the permeability of the plug [m²] over time for a) the dry scenario and b) the wet scenario.

As a result of the threshold criterion for the borehole plug to become impermeable (i.e. its porosity decreases below 0.0003), more and more simulations are terminated as time progresses (cf. Figure 51 to Figure 52). The consequences of these “missing values” for the sensitivity analysis are discussed below. Despite the relatively large variation of the plug porosity and permeability, the distribution of the maximum dose rate in the biosphere has a small dispersion (the ratio between the 99% percentile and the 1% percentile is less than 2) and it is spread around $6.8 \cdot 10^{-13}$ Sv/year, which is orders of magnitude below the reference

value of 0.3 mSv/year. The reason for the rather small variability of the dose rate is the very slow transport of contaminated brine that enters the deep groundwater layer surrounding the host rock to the biosphere. The closure time of the plug is reached in 95% of the realizations in less than 1000 years in the case of the wet scenario, while the maximum dose rate is found after approximately 100'000 years.

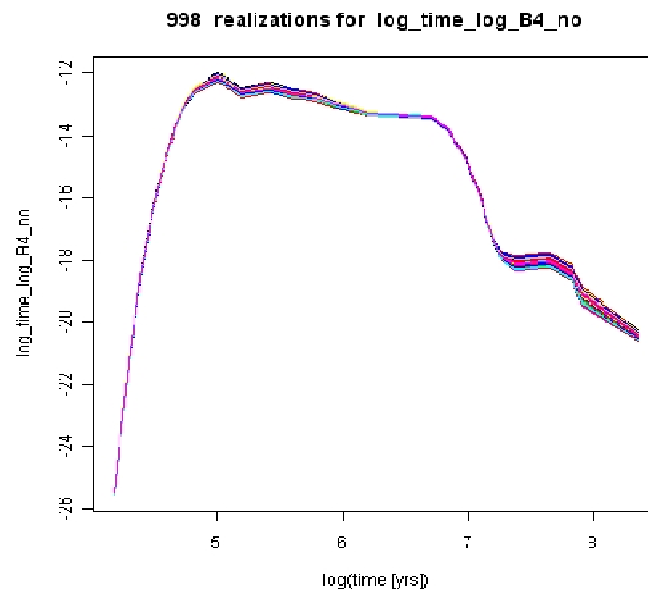


Figure 53 Evolution over time of the dose rate [Sv/year] in the biosphere (wet scenario)

Two regression models have been used for the sensitivity analysis of the maximum dose rate in the biosphere, the first based on the logarithm of the raw values, the second one based on the ranks (Table 4). When moving from the first regression model to the second one the coefficient of determination increases from 0.36 to 0.96. Taking into account the difference in the output variability explained by the regression models, the model based on the ranks must be chosen. The rank-based model accounts for 95.5% of the output variability, what makes the results of the analysis very reliable, and indicates a strong monotonic relation between the maximum dose rate and some of the input parameters. Table 5 shows that the three most important input parameters are F_{09} (negative correlation with the output) followed by B_{recr} and n (negative and positive correlation respectively with the output). As can be derived from the $SRRC$ -values in Table 5, F_{09} is clearly the most important parameter, while n and B_{recr} are less relevant, despite the high $PRCC$ -values.

Table 4 Coefficient of determination for the maximum dose rate in the biosphere

	linear regression	rank regression
R^2	0.362	0.955

Table 5 Regression based sensitivity indices for the maximum dose rate in the biosphere (A.4)

	Pearson	Spearman	SRC	PCC	SRRC	PRCC
B_{recr}	-0.164	-0.295	-0.130	-0.160	-0.239	-0.748
A_{ps}	-0.073	-0.060	-0.066	-0.082	-0.054	-0.245
F_{09}	-0.566	-0.921	-0.554	-0.569	-0.906	-0.974
n	0.140	0.220	0.132	0.163	0.203	0.692
F_{46}	0.062	-0.041	0.070	0.087	-0.056	-0.254
F_{47}	-0.024	0.008	-0.009	-0.012	0.003	0.013

Table 6 and Figure 54 show the statistical results of the uncertainty analysis of the dose rate distribution in the biosphere at selected time steps for the “wet” scenario (brine intrusion). In general the distribution is rather symmetric and the kurtosis is close to 3 at all time steps. A careful study of the quantiles and the skewness coefficient (Table 6) indicates that the distributions are slightly skewed towards the small values at early times and towards the large at late times (after 112'000 years).

Table 6 Results of the uncertainty analysis at selected times for the dose rate in the biosphere

time [year]	geomet- ric mean [Sv/year]	geometric standard deviation [-]	quantiles [Sv/year]					skewness [-]	kurtosis [-]
			1%	5%	50%	95%	99%		
30'000	2.91E-17	1.05	2.58E-17	2.69E-17	2.92E-17	3.13E-17	3.21E-17	-0.316	3.223
60'000	1.73E-13	1.08	1.45E-13	1.53E-13	1.73E-13	1.94E-13	2.02E-13	-0.120	2.749
112'000	5.81E-13	1.14	4.56E-13	4.84E-13	5.75E-13	7.33E-13	8.02E-13	0.386	2.823
250'000	3.33E-13	1.15	2.54E-13	2.71E-13	3.30E-13	4.30E-13	4.78E-13	0.396	2.833
490'000	1.96E-13	1.13	1.54E-13	1.64E-13	1.95E-13	2.43E-13	2.67E-13	0.390	2.859
1'000'000	8.69E-14	1.08	7.33E-14	7.69E-14	8.65E-14	9.98E-14	1.07E-13	0.224	3.008

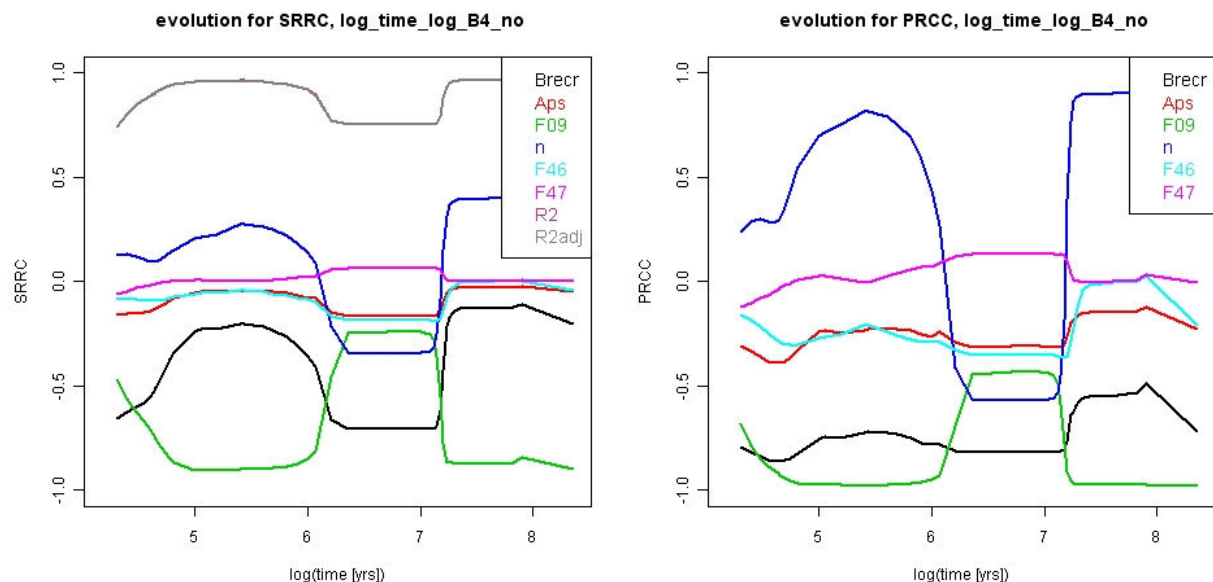


Figure 54 Regression based sensitivity indices for the dose rate in the biosphere – wet scenario

Figure 54 shows that the coefficient of determination R^2 remains above 0.75 all the time and is close to 1 except for early times and times between 10^6 years and 10^7 years. The results obtained from this sensitivity analysis are thus quite reliable. The most important input parameter is F_{09} , the permeability of salt grit. In the interval with lower absolute values of the $SRRC$ and $PRCC$ (between 10^6 and 10^7 years), R^2 decreases noticeably, too. B_{recr} and n are the next most important input parameters; between 10^6 and 10^7 years they replace F_{09} in the ranking. The shifting $SRRC$ and $PRCC$ values in the three intervals before 10^6 years, between 10^6 to 10^7 years, and after 10^7 years are related to the transport properties of the different radionuclides emitted to the biosphere.

Analysis of the EMOS results reveals that in the first interval up to 10^6 years, the mobile nuclides ^{79}Se and ^{129}I dominate the dose rate (data not shown). In the second interval (10^6 to 10^7 years), the decay product of the ^{238}U chain, ^{226}Ra and her daughter nuclides, is dominating, and in the third interval, after 10^7 years, ^{229}Th is the most relevant radionuclide contributing to the dose rate. The interpretation of the factors influencing the differences in release rates of the above mentioned radionuclides is complex due to several possible interactions between the used models.

4.1.5 Conclusions

A peculiarity of the *EMOS* simulations was that the simulations were not necessarily terminated at fixed end-points in all compartments. As a consequence the number of time steps and output parameters available for the sensitivity and uncertainty analysis decreased progressively for the sets of 1000 simulations.

Three strategies have been adopted in this study to deal with missing values after termination of the calculation and they provided in some cases important differences.

The first strategy was not to replace missing values. In this case the analysis was done with less and less realisations until either the number of remaining realizations was less than 20 or the end of the output sample file was reached. With this strategy, the results obtained refer at each time step only to the realizations indeed present in the sample. The second strategy was to replace missing-values by a threshold ($3 \cdot 10^{-4}$ in the porosity case and $3 \cdot 10^{-30} \text{ m}^2$ in the permeability case), whereas the third strategy was to replace missing values by the last available value of a realization.

At early times, when either no run or just a few have been stopped, all three strategies deliver the same results for the statistical analysis. At later times, depending on the output variable, results may differ a lot. Without data replacement, at late time steps large fluctuations of the sensitivity indices can appear due to the reduction of data available for the estimation. In general, the first imputation strategy (replace missing-values by a threshold) results in smaller R^2 values at late time steps. This strategy does not account for the variability present in the last values and makes the used regression techniques too insensitive at late times. Better results were obtained without imputation and with the second imputation strategy (replace missing-values by the last available value). For the given dataset, the second imputation strategy preserves well the differences between the different runs. The advantage of the no-imputation strategy is that it provides information related (only) to the available data at each time step.

Most of the distributions of the output variables were found to be highly skewed, resulting in a large decrease of R^2 values. The logarithmic transformation showed appropriate, transforming them into either symmetric or almost symmetric variables. The *log*-transformation was also adequate to perform the sensitivity analysis, though regression analysis based on the

rank transformed values has proved the best option. R^2 values were large enough in most cases to reach well-supported conclusions.

SRRC's and *PRCC*'s have been used simultaneously in the present study. With independent input parameters both indicators provide always the same ranking. *SRRC*'s are more useful from the point of view of interpretation (slopes), while *PRCC*'s have larger absolute values, which makes them more convenient for ranking input parameters, particularly when time-dependent outputs are analysed.

From the point of view of repository safety, the variability of the parameterization of the plug model has only a limited influence on the repository safety in terms of potential dose rate for the wet scenario. The lack of experimental data on the permeability behaviour of salt grit at very low porosities (< 0.005) do not influence the safety of the repository: for the given set-up, any monotonic assumption on the permeability evolution at porosities below 0.005 is suitable to show that the expected dose rates are far below any legal limit.

4.2 NRG application case: repository design in argillaceous rock

This section describes the probabilistic uncertainty and sensitivity analysis for a repository design in an argillaceous host rock - the Boom Clay - undertaken by JRC-IE and NRG.

4.2.1 Description of the system and the model

In the *TRUCK-II* design [Barnichon et al. 2000], the reference repository is located at a depth of 500 m in a 100 m thick clay layer of the Boom Clay formation and includes shafts, galleries, disposal cells and other open volumes (e.g. workplaces). In the original *TRUCK-II* layout, galleries are surrounded by small horizontal disposal cells each enclosing a single *COGEMA* container of vitrified HLW. For this study, the geometry of the disposal cell is modified to store one "*Supercontainer*" (Figure 55) following a recent Belgian concept [De Bock et al. 2004], where a Supercontainer consists of two *COGEMA* containers enclosed by a large concrete buffer (1.9 m diameter). This results in a disposal cell of 6.5 m depth (including a sealing plug of 3 m) and 2 m diameter. When all containers are stored, the access galleries and other open volumes will be backfilled and the shafts will be sealed.

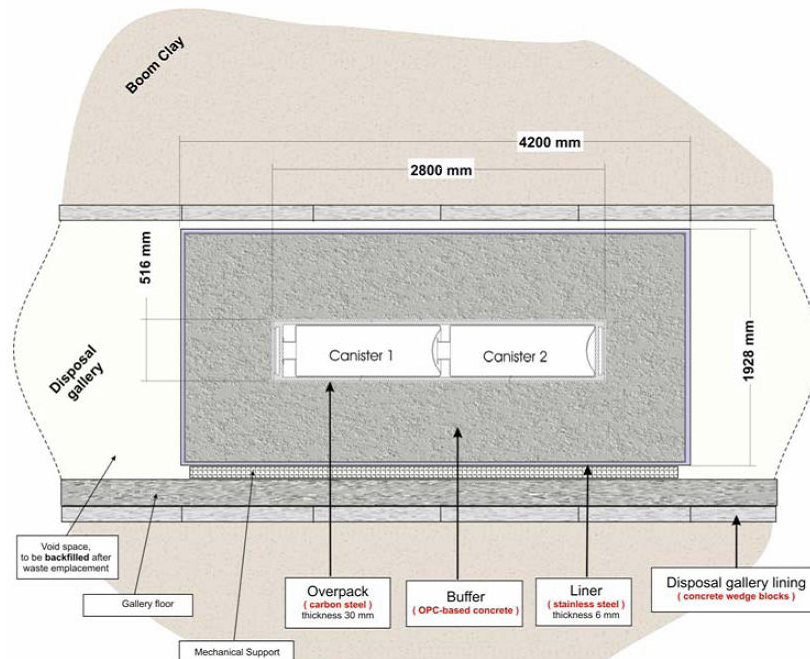


Figure 55 Radial cross-section of the reference Supercontainer.

4.2.2 Considered scenario

In the present study, the so-called “abandonment” scenario is analysed. The geological disposal of radioactive waste is a long-term operation that will be realized in subsequent stages, leading to extended time periods in which the facility is open and accessible. Political, societal or technical problems may lead to an abandonment of the facility. When shafts and galleries have not or only inaccurately been sealed, an “abandonment” situation may appear and the repository may get flooded. In case of it is assumed that

- waste canisters have been emplaced in the disposal boreholes
- the disposal galleries have been sealed-off with a sealing plug
- the shafts and access galleries have not yet been backfilled and sealed
- the access galleries are filled with water
- the main shaft that is connected to shallow groundwater layers is filled with water

4.2.3 Modelled geometry

To analyse the repository performance in an abandonment situation, only one single disposal cell has been modelled, containing the inventory of one Supercontainer. Only diffusive transport of radionuclides is modelled, using sorption parameters from the BENIPA project [Sillen 2002]. Precipitation of radionuclides is not taken into account. Migration of radionuclides from

the upper border of the clay layer to the geosphere is neglected. The barriers formed by the stainless steel container, the steel overpack and the outer liner of the galleries are not taken into account, too.

Figure 56 shows the conceptual model of the repository as been implemented in the *PORFLOW* computer code [PORFLOW 1999]. To represent a situation of abandonment, the disposal cell is connected to an open gallery, flooded and without backfilling material. The gallery is assumed to be directly connected to overlying aquifers. The model is established in two-dimensional cylindrical coordinates.

The grid size of the computational domain of the gallery in x -direction is taken equidistant with lengths of 1 m (0.35 m in the disposal cell, 0.1 m in the plug). The cell size in the gallery in r -direction is 0.125 m. To reduce the calculation efforts, the cell sizes in the clay layer in r -direction increases from 0.0125 m to 1 m at the top of the domain. The model discretization is tested for numerical artefacts.

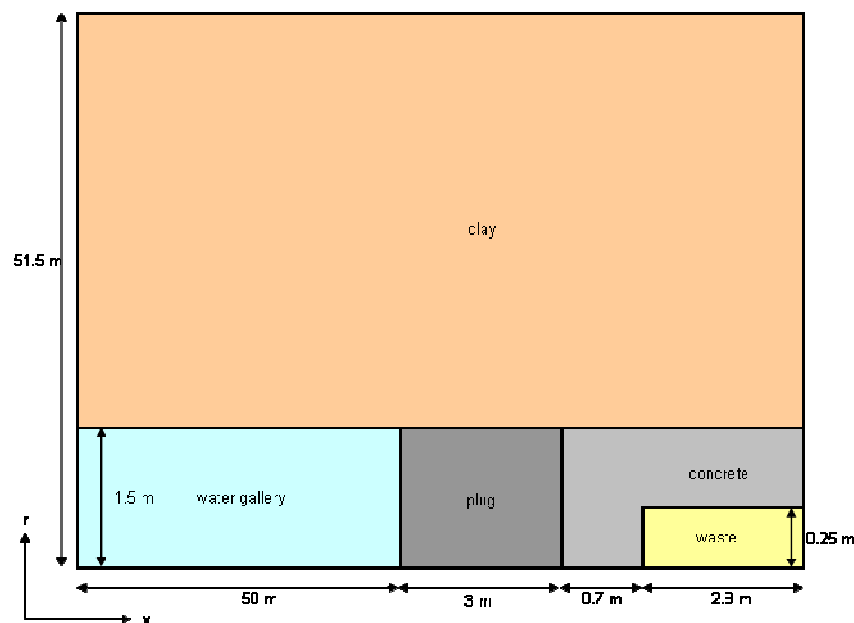


Figure 56 Structural model of the abandonment scenario (not on scale)

4.2.4 Objectives of the study/variability of input parameters

This study focuses on the variability of three input parameters: the composition of the sealing plug, the apparent diffusion coefficient in clay and the dissolution time of the vitrified waste.

Plug material

Several constructions of sealing plugs are conceivable, but at this moment, for the Supercontainer-concept no selection of material for the plug has been made [De Bock et al. 2004]. In

the present study it is assumed that the plug is a mixture of an inert material and a varying fraction of clay, ranging from 0 to 100%.

Diffusion

The apparent diffusion coefficient of a soluble material in clay strongly depends on the porosity and charge of clay material. The properties of Boom Clay reported in [Sillen 2002] are typical for the Mol underground laboratory, but may vary dependent on the exact location of a future repository. Insufficient data on the properties and variability of the Boom Clay sediments on the horizontal scale are available, but since most parts of the Boom Clay layer in The Netherlands are situated at greater depth (± 500 m), the diffusion rate may be lower than in the Belgian underground laboratory. The “best guess” value is based on values as were found for the Boom Clay layer of the Belgian underground laboratory in Mol, and for the minimum- and maximum-parameterization we varied the apparent diffusion coefficient by a factor of 10 (Table 7).

Glass dissolution rate

In [Grupa and Houkema 2000] it has been assumed that the vitrified contents of the waste canisters will dissolve completely with a constant rate in a time frame of 75'000 years. This assumption was based on the fact that the presence of silicates in groundwater, both from natural background and from glass dissolution, strongly controls the glass dissolution rate [Curti, 2003]. Recalibration of the rate constant by data of the FP6 *NF-PRO* project [Godon et al. 2008] delivered a range of dissolution rate equivalent to a time frame of about 20'000 to 200'000 years for the original *TRUCK-II* repository design. However, the leachate from the concrete buffer of the Supercontainer may lead to pH values of up to 13.5, increasing the dissolution rate by a factor of 400 [Grambow, et. al. 2007]; according to [Minet and Lartigue 2008] the “ $V_0 \rightarrow V_\infty$ ”-model is not valid for that high pH values. To cover the large uncertainty around the dissolution rate of glass, we used a simplified dissolution model assuming the glass is dissolved linearly in only one time interval. The parameterization in Table 7 is conservative and uses rather short intervals to reflect the large uncertainty on the glass dissolution rates in contact with large concrete buffers at this moment.

Table 7 Model parameter and distribution

parameter	minimum	maximum	distribution
clay content in plug [%]	0	100	uniform
$D_{a \text{ clay}}$ [m^2/a]	$1.26 \cdot 10^{-3}$	$1.26 \cdot 10^{-2}$	uniform
glass dissolution time [a]	100	10'000	uniform

4.2.5 Output parameters analysed

For the statistical analysis, a selection of output parameters has been made that are relevant for the safety and performance of the given repository design or that are assumed to be helpful in understanding the repository behaviour. These indicators are summarized in Table 8. The indicators are analysed as sum of all relevant nuclides or as individual values for 4 relevant nuclides: ^{79}Se , ^{126}Sn , ^{229}Th and ^{241}Am .

Table 8 Output parameter analysed

	safety and performance indicators	unit
A.1 - A.5	Dose rate in biosphere: sum of all nuclides, ^{79}Se , ^{126}Sn , ^{229}Th and ^{241}Am	[Sv/a]
B.1	Radiotoxicity concentration at the end of gallery: sum of all nuclides	[Sv/m ³]
C.1 - C.5	Radiotoxicity concentration in the disposal cell: sum of all nuclides, ^{79}Se , ^{126}Sn , ^{229}Th and ^{241}Am	[Sv/m ³]
D.1 - D.5	Radiotoxicity flow from plug to gallery: sum of all nuclides, ^{79}Se , ^{126}Sn , ^{229}Th and ^{241}Am	[Sv/a]
E.1 - E.5	Radiotoxicity flow from waste to disposal cell: sum of all nuclides, ^{79}Se , ^{126}Sn , ^{229}Th and ^{241}Am	[Sv/a]

4.2.6 Statistical methods

An uncertainty and sensitivity analysis has been performed for the indicators computed by *PORFLOW*. For every indicator 500 realizations with different values of the input parameters have been computed. A first analysis showed that the distribution of most of the output parameters was strongly skewed. Therefore it was decided to perform the uncertainty and sensitivity analysis on the *log*-transformed data, too. Table 9 and Table 10 give an overview over the statistical analyses performed.

Table 9 Uncertainty indicators

	Quantitative indicators (Tables)	Qualitative indicators (Figures)
all data		plots of all realizations evolution of mean, standard deviation and quantiles
predefined (fixed) times	mean standard deviation quantiles: $q_{1\%}$, $q_{5\%}$, $q_{25\%}$, $q_{50\%}$ (median), $q_{75\%}$, $q_{99\%}$ skewness and kurtosis	empirical cumulative distribution function (ecdf)
maximum values and time to the maximum values	standard deviation quantiles: $q_{1\%}$, $q_{5\%}$, $q_{25\%}$, $q_{50\%}$ (median), $q_{75\%}$, $q_{99\%}$ skewness and kurtosis	empirical cumulative distribution function (ecdf)

Table 10 Sensitivity indicators

	Quantitative indicators (Tables)	Qualitative indicators (Figures)
all data		evolution of SRC / SRRC evolution of R^2
maximum values and time to the maximum values	SRC / SRRC R^2 Smirnov test Mann-Whitney test	CSM plot Conditional cobweb plot

4.2.7 Results and discussions

From the 500 realizations calculated, 499 realizations are used for the statistical analysis.¹ From all indicators calculated, two aspects will be highlighted: the evolution of the dose rate in the biosphere and the migration of ^{241}Am in through the repository. All regression analyses are performed on the raw data, the *log*-transformed data and on the rank transformed data. The later gives consistently the highest R^2 for the regression based methods; all results discussed in the next sections are therefore based on the rank transformed data.

Dose rate in biosphere

Figure 57a shows the evolution of the dose rate in the biosphere in terms of main quantiles, mean and standard deviation. The maximum dose rate is found around 12'000 years (mean

¹ For one realization, PORFLOW used different time steps without apparent reasons. Analysis of this realization shows neither extreme values in the sample nor in the output.

$1.26 \cdot 10^4$ years, median $1.10 \cdot 10^4$ years). The peak dose rate is mainly caused by ^{79}Se , a non-sorbing anion that leaves the repository unretarded.

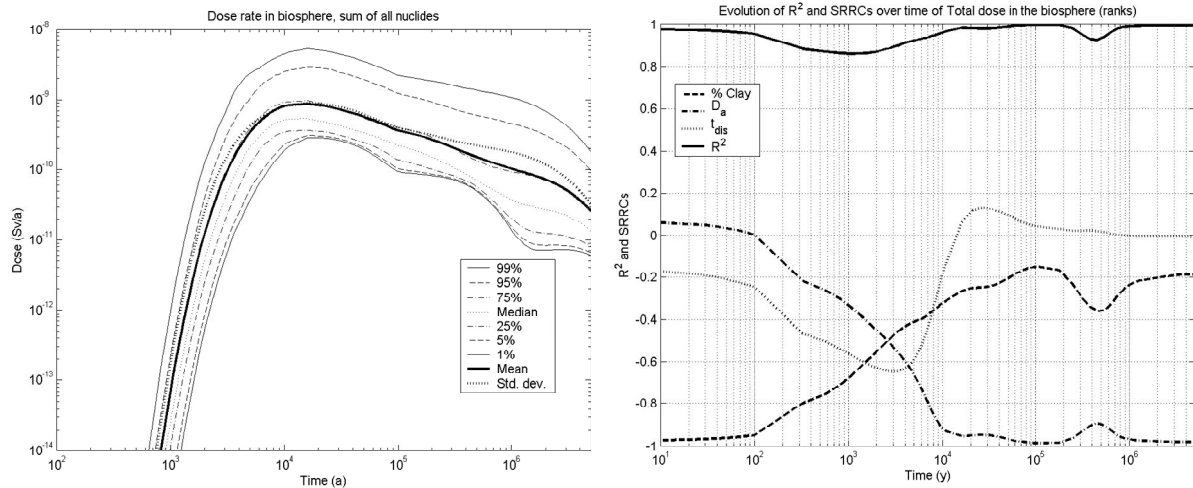


Figure 57 Dose rate in the biosphere (sum of all nuclides). a) evolution of the dose rate in time; b) evolution of R^2 and SRRCs for the rank transformed data

Figure 57b shows the evolution of the R^2 and SRRCs for the three input parameter for the dose rate in the biosphere. The R^2 is above 0.85 for the whole interval calculated, indicating very reliable statistical results. In the first 2000 years, %clay is the most important parameter influencing the dose rate. After that period, D_a is the most important input parameter, gaining SRRC values of almost one. The dissolution time t_{dis} only affects the dose rate in the interval of about 100 to 10'000 years.

Table 11 shows the results of the sensitivity analysis of the maximum dose rate. The value of the maximum dose rate is mostly influenced by the apparent diffusion coefficient D_a and to a lesser extent by the fraction of clay in the sealing plug (%clay). The influence of the dissolution time of the waste matrix t_{dis} is found to be insignificant, both for the regression based indicators and for the Monte Carlo filtering based indicators (data not shown). The time to the maximum dose rate ("time to peak") does behave differently: for that parameter, t_{dis} has the largest influence. However, as depicted in Figure 58, due to the discrete time steps of the numerical calculations, the cumulative distribution function shows a step-like pattern. This also affects the R^2 value that is lower than for the maximum dose rate (0.72 vs. 0.98, respectively). Interpretation of these results should therefore be done with care.

Table 11 SRRCs and R² values for the peak dose rate and time to the peak dose rate in the biosphere

	SRRC			R ²
	%clay	D _a	t _{dis}	
peak value	-0.30	-0.95	0.01	0.98
time to peak	0.35	0.14	0.77	0.72

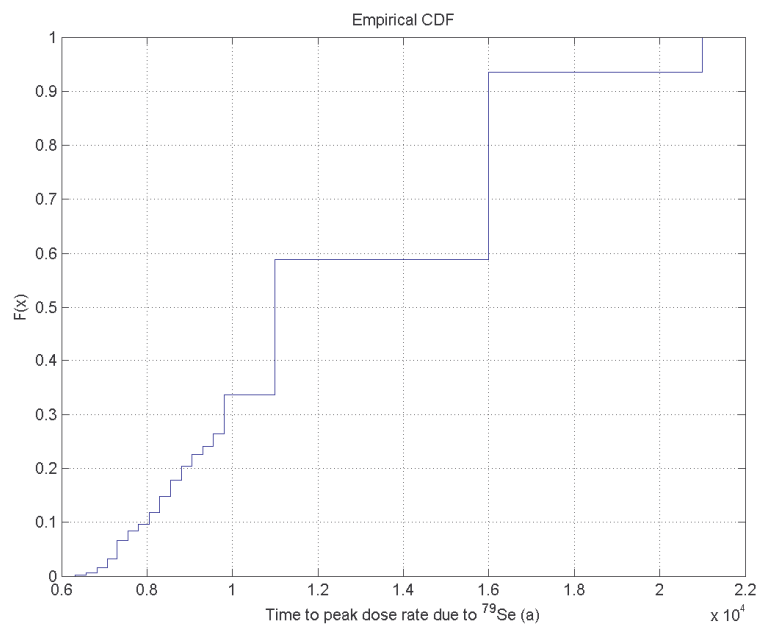


Figure 58 Empirical cumulative distribution function (ecdf) of the time to the peak dose

4.2.7.1 Migration of ²⁴¹Am through the repository

Figure 59a shows the evolution of the dose rate of ²⁴¹Am in the biosphere. The mean curve shows two peaks, a feature that is most evident for ²⁴¹Am but can also be found for other radionuclides. Figure 59b shows that unlike in case of the total dose rate, %clay has a rather long lasting and also stronger influence on the dose rate of ²⁴¹Am. Generally, two possible pathways from the waste to the gallery are existing: through the plug or through the clay that encloses the plug. In case of ²⁴¹Am, the second (more delayed) route is less relevant because of the relatively short half life of 432 years, explaining why these two peaks are more evident than for the other nuclides. The regression analysis in Figure 60b shows that for the first pathway through the plug, the %clay is clearly the most relevant input parameter over the whole interval of the calculations. The apparent diffusion coefficient D_a has only a minor role that also depends on the % clay, because D_a influences both the diffusion in the host rock and the diffusion in the fraction of clay in the plug.

In Figure 60a a noticeable peak at 100 years is visible that is one order of magnitude higher than the other peaks. For this early peak, the dissolution time of the waste t_{dis} is expected to

have a relevant influence. However, this is neither supported by the regression based indicators in Figure 60b nor in other regression based analysis performed in the present study: the *SRRC* is very close to zero over almost the whole time interval and never exceeds a value of 0.1 (Figure 60a and Figure 60b). The strongest relation with t_{dis} comes from the analysis of the time to the peak (Table 12), but this relation is very weak. However, the weak relation may also be the result of the comparable small number of discrete time steps chosen for the output of data.

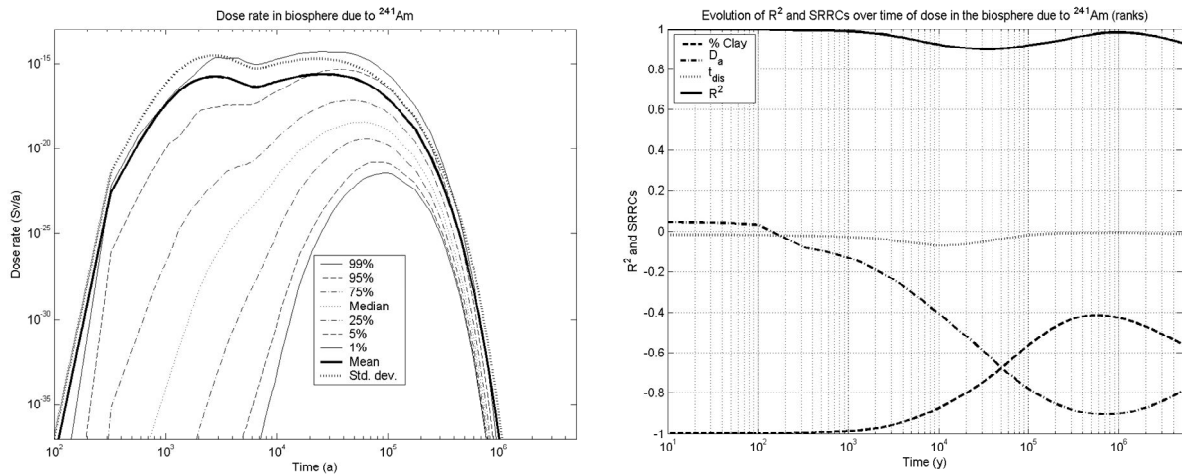


Figure 59 Dose rate in the biosphere due to ^{241}Am . a) evolution of the dose rate in time; b) evolution of R^2 and SRRCs for the rank transformed data.

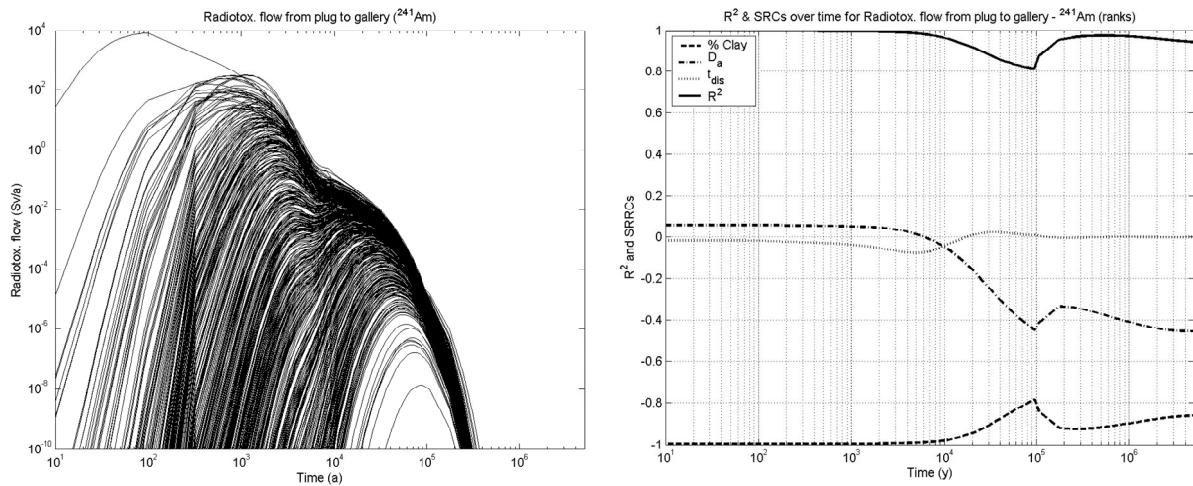


Figure 60 Radiotoxicity flow from the plug to the gallery due to ^{241}Am . a) evolution of the radiotoxicity in time; b) evolution of R^2 and SRRCs for the rank transformed data.

Table 12 SRRCs and R² values for the peak radiotoxicity flow and time to the radiotoxicity flow from the plug to the gallery due to ²⁴¹Am.

	SRRC			R ²
	%clay	D _a	t _{dis}	
peak value	-0.995	0.001	-0.059	0.992
time to peak	0.985	-0.075	0.074	0.980

Some further insights about the behaviour of the peak flow provided by Monte Carlo filtering techniques are given by the cobweb plot (Figure 61a) and the CSM plot (Figure 61b). Figure 61b shows that %clay is the only input parameter with a statistically significant departure from the diagonal; its 5% smallest values contribute roughly 95% of the output mean.

Figure 61a indicates a clear relation between the largest five values of the peak flow and small values of %clay and t_{dis}. This plot highlights that the largest peak flow is obtained when the smallest %clay and the smallest t_{dis} values are combined. This is statistically also confirmed by *Smirnov* and *Mann-Whitney* tests for which the results are summarized in Table 13. Thus indeed a strong interaction exists between %clay and t_{dis}, with very low values of both resulting in very large output values. This effect may also be visible by a 3D plot representing the peak radiotoxicity flow vs. %clay and t_{dis} (not shown).

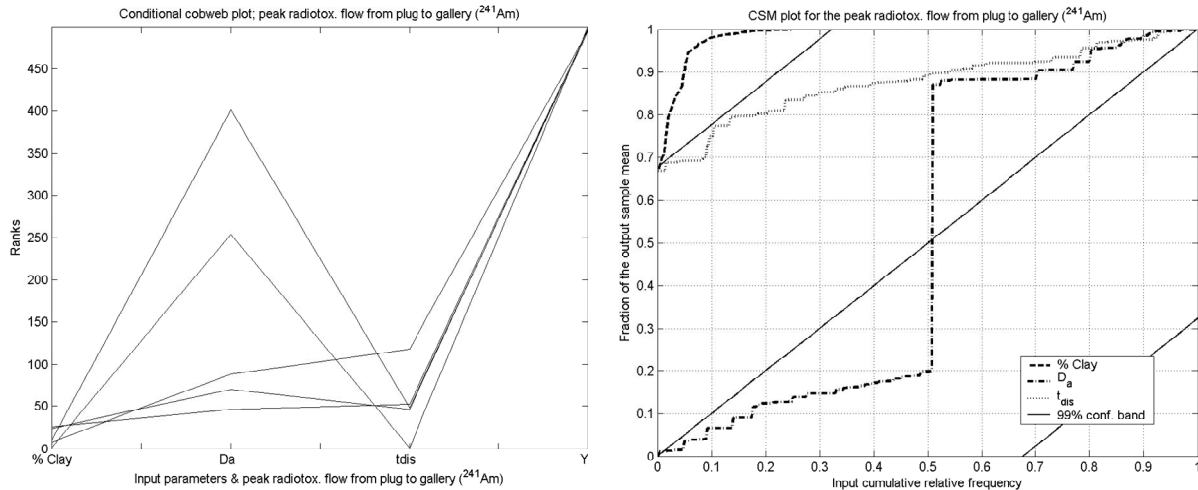


Figure 61 Radiotoxicity flow from the plug to the gallery due to ²⁴¹Am. a) Conditional cobweb plot (1%/99% sample division); b) CSM plot

Table 13 Results of Smirnov and Mann-Whitney test for the peak radiotoxicity flow from plug to gallery due to ^{241}Am (1%/99% sample division)

p-value	%clay	D_a	t_{dis}
Smirnov	5.4e-05	0.24	2.2e-03
Mann-Whitney	2.3e-04	0.23	2.2e-03

4.2.8 Conclusions

General findings

Most of the findings presented in the previous section describe general patterns found for almost all indicators:

- When analysing time dependent output variables, the quality of the regressions (in terms of fraction of the output variability explained – R^2) improved when both output and inputs were *log*-transformed. Further improvement was consistently obtained when rank based regressions were used (R^2 values of at least 0.8 for all indicators at any time step). Regression analysis based on raw values performed particularly poorly when the output is spread over (many) orders of magnitude.
- The same systematic improvement was observed for the peak values of the output variables.
- When analysing the time to the peak of the output variables rank based regression produced the best results, followed by the regressions based on the raw values. Unlike the other statistical analyses, in case of the time to the peak, the regressions based on *log*-transformed outputs and inputs delivered the poorest results. The analysis of the distributions of this type of output variables supports the finding that the linear scale is more appropriate than the logarithmic scale.
- A remarkable degree of consistency has been observed when comparing the results provided by CSM plots, cobweb plots, Monte Carlo filtering techniques and regression analysis based on ranks. The order of input parameters according to their importance was almost always the same.
- The use of results provided by uncertainty analysis tools (runs vs. time, quantiles vs. time) has been identified as an important ancillary tool to help interpreting correctly results provided by SA techniques (regression).
- No tool used in this study has been designed to study interaction between input parameters. Nevertheless, an adequate combination of them (cobweb plots, CSM plots) - together with uncertainty analysis tools - have been shown to be useful in uncovering relevant interactions. The *ad hoc*, not systematic, use of scatter plots was also found to be helpful.

Suitability of the procedure for sensitivity analyses

The calculation of diffusion processes in a two-dimensional system for the purpose of sensitivity analysis is found to have certain limitations. The time necessary to calculate 500 realizations for the system of this study is very long, but due to the use of increasing time step sizes calculations could be performed within a few hundred hours with only limited numerical artefacts. The artefacts were most apparent in the waste container, but at the times these artefacts became relevant, the influence of the waste compartment on the repository safety has become negligible.

Implication to repository safety

For the repository safety, the most relevant safety feature of the system studied is the adsorption of radionuclides at the gallery/clay interface; the clay efficiently removes radionuclides on their way from the plug to the geosphere. The statistical analysis shows that a large diffusion coefficient of radionuclides in clay improves the repository safety in case of an abandonment situation, despite the larger flux through the sealing plug. However, one should keep in mind that this study only analyses a simplified setup with one Supercontainer and also not accounts for the radionuclide outflow at the upper surface of the clay layer. Extrapolation of the results should therefore be done with care.

4.3 Enresa application case

4.3.1 *Description of the system*

ENRESA reference concept in granite is based on the disposal of spent fuel in carbon steel canisters in 500m long horizontal drifts excavated by means of tunnel boring machines. Canisters are surrounded by bentonite. Access is accomplished by means of "main drifts" which run perpendicular to the disposal drifts. The main drifts meet at a central area, which includes the required underground infrastructure. Communication between the surface and the central underground area is accomplished by means of 3 access shafts and a ramp.

The carbon steel canister measures 4.54 m in length and 0.90 m in diameter, has a thickness of 10cm, and contains 4 PWR or 12 BWR fuel elements in a subcritical configuration. After being unloaded from the reactor, the fuel elements are temporarily stored during 50 years to allow the thermal power to decay to a level at which they may be disposed of, with a thermal power of 1,200 W per canister. A total of 3,600 canisters will be required for the total inventory of spent fuel estimated for the Spanish nuclear programme.

DISPOSAL CONCEPT

- **Deep disposal**
- **Crystalline rock**
- **Spent fuel**
- **Carbon steel canister**
- **Horizontal emplacement**
- **Bentonite buffer**

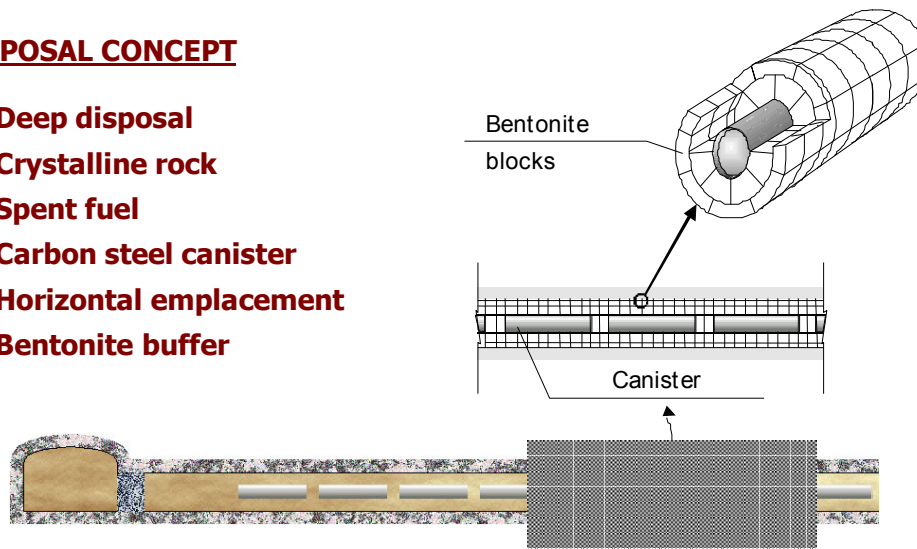


Figure 62 Longitudinal section of a disposal drift

Canisters are emplaced in cylindrical disposal cells, built with blocks of pre-compacted bentonite. The disposal drifts of 500 m in length and 2.4 m in diameter (see Figure 62) are located at a depth of 500 m in the host formation. The separation between canisters is determined mainly by thermal constraints. Separations of 1.4 m between canisters and 35 m between disposal drifts have been established, in order not to exceed a temperature of 100° C in the bentonite. Actual separation is a function of the properties of the host rock, and thermal calculations have been made for a reference generic site.

Once a disposal drift is filled with canisters and bentonite blocks, it is sealed with a 6 m long seal made of bentonite blocks and closed with a concrete plug at its entry. After completion of all the disposal drifts, the main drifts, ramp, shafts and other remaining rock cavities will be backfilled with a mixture of bentonite and natural sand or an appropriate crushed material. The backfilling material will consist of 10 % bentonite (increasing up to 20 % at the top of the drifts) and suitably graded sand.

4.3.2 System evolution and modelling

Due to high suction pressures in bentonite, it is expected that bentonite will be fully saturated a few decades after repository closure. When water contacts the carbon steel canister a process of anaerobic generalized corrosion starts. In the calculations performed canisters have a minimum duration of 20,000 years and fail sequentially up to 100,000 years approximately. In addition, a few canisters (10 as a maximum) are assumed to have an undetected defect and fail around one hundred years after disposal.

After canister failure, and since no credit is given to the cladding as a barrier, there is an instantaneous release of some volatile radionuclides (such as I129, Cl36 and Cs135) and the gradual release of the radionuclides in the UO₂ matrix starts.

The radionuclides released from the waste dissolve or precipitate in the water in the canister cavity, depending on their solubility limits. Dissolved radionuclides are transported by diffusion through the bentonite buffer and pass to the groundwater flowing around the disposal drifts. Transport calculations in the near field are done using a 1D axisymmetric model of the bentonite and imposing a “mixing tank” boundary condition in the bentonite outer surface, where radionuclides leave the bentonite and pass to the groundwater flowing in the granite.

In the transport calculations the granite is modelled as a single one-dimensional planar fracture (or stream tube) with a distribution of travel times provided by the hydro-geological model. Longitudinal dispersion and matrix diffusion into the adjacent rock are modelled, including sorption onto the granite. Neither solubility limits nor sorption on fracture surfaces or infill are considered.

The radionuclides released from the granite formation are discharged to a river with a flow rate of $10^6 \text{ m}^3/\text{year}$. Radionuclides concentrations in river water are calculated dividing the release rate from the geosphere by the river flow rate. River water is used by the critical group to produce most of the aliments they consume.

4.3.2.1 Variables used in the calculations.

In the repository model all the uncertain parameters are represented by probability distribution functions (pdf's), with the exception of Biosphere parameters, which are constants.

The model uses 135 random input parameters, each one being represented by an independent pdf. Many of the parameters are element-specific (such as the distribution coefficients in the bentonite), and a given radionuclide can be affected by 22 parameters at most. Random input parameters are presented in Table 14, classified on the basis of the component of the disposal systems affected by the parameter. Element-specific parameters are shown in grey cells in Table 14.

Table 14 List of random input parameters for chemical element X.

Waste(3)	
R1000	UO ₂ matrix alteration rate after 1000 years of cooling
METAL LIFETIME	Duration of the structural components of the fuel elements
GAPX	Fraction of the inventory of element X released immediately when the canister fails – Instant Release Fraction (IRF)
Canisters (3)	
FAIL	Number of canisters with initial defects that fail in a century approximately
MEAN	Mean duration of the canisters due to generalised corrosion
SLOPE	Parameter for the Weibull distribution of canister failures
Near field (6)	
VC1	Volume of water in the canister cavity
POR ANIONS BUFFER	Fraction of bentonite porosity accessible to anions
WATER NF	Groundwater flow in the granite of the near field
SOLX	Solubility of element X in the water of the canister cavity
KDBEX	Distribution coefficient bentonite-water for element X
DIFFX	Diffusion coefficient in bentonite porewater (Dp) for element X
Water flow (4)	
WATER TT0	Water travel time for a kinematic porosity of 10 ⁻⁴
KINEM POR	Kinematic porosity of the granite
PECLET	Peclet number (longitudinal dispersion)
PATH LENGTH	Length of the geosphere pathway, from repository to biosphere
Granite matrix (6)	
FWA	Flow wetted area
THICKM	Thickness of the granite matrix
PORMAT	Porosity of the granite matrix
POR ANIONS GRANITE	Fraction of granite matrix porosity accessible to anions
MATCOR	Correction factor for applying K _d values measured in batch to intact granite
KDRX	Distribution coefficient granite-water for element X (in batch)

The experience gained in previous Safety Assessments allows identifying some key parameters that can control the doses caused by the repository and are a combination of the random input parameters. In this document such parameters are called “derived parameters” and some of them are presented in Table 15.

Table 15 List of “derived parameters”

TT NF	Travel time through the near field.
TT FF	Travel time through the far field
$\text{DecayT}=\exp(-0.693 \cdot \text{TT FF}/T_{1/2})$	Decay of the radionuclide during transport
MF NF	Mass flow from the near field, that is defined as the product SOLX·WATER NF(for solubility controlled species)

4.3.3 Sensitivity analysis performed

Figure 63 shows the mean doses obtained in a stochastic calculation with 5000 runs for the Spanish disposal concept in granite. Figure 63 presents all the radionuclides that cause mean doses greater than 10^{-11} Sv/yr.

The sensitivity analysis performed was limited to the doses caused by the repository system, and only for the radionuclides presented in Figure 63 (plus total dose and radioactive series). Nevertheless, the methodology presented in the next sections can be used to study other results of the model, such as the radionuclide fluxes leaving the near field of the repository.

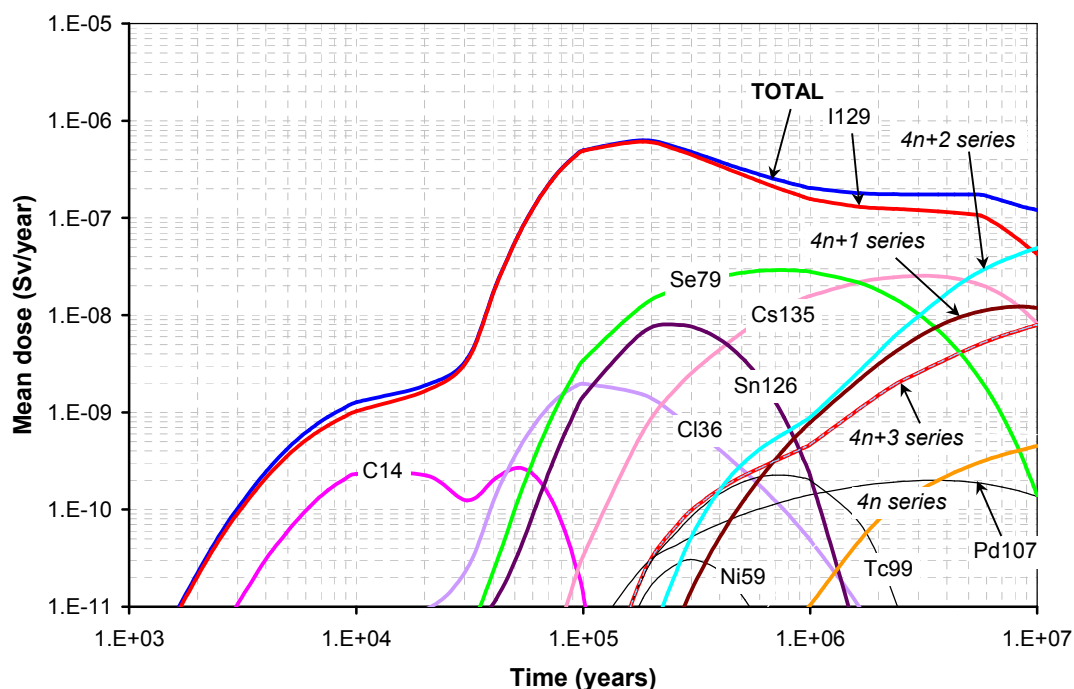


Figure 63 Mean doses (total and per radionuclide).

The input parameters considered in the analysis are both the random input parameters and the “derived parameters” identified as potentially relevant. In the tables in section 4.3.4 random input parameters are presented in white cells, while “derived parameters” are shown in grey cells.

The sensitivity analysis methods have been applied to a 25,000 run calculation. Other calculations with 200, 1000 and 5000 runs have been done to study the effects of the sample size on the results. The sensitivity analysis has been done using only graphic (2.2.1) and Monte Carlo methods (2.2.2).

The main results obtained with the different methods are presented in sections 4.3.4.1 to 4.3.4.6.

4.3.4 *Main results*

4.3.4.1 *Correlation and regression analysis for the peak doses*

Due to the relevance of the peak doses in the Safety Assessment, the correlation coefficients of all the input parameters (both random and derived) with the total peak dose and peak dose due to a single radionuclide or a radioactive series have been calculated. Correlation coefficients have been calculated both for values (CCs) and for ranks (RCCs).

Table 16 presents the results of the correlation analyses for peak total dose and the peak doses due to I-129, Cl-36 and Cs-135. It can be seen that “derived parameters” are much more correlated with the peak doses than the original random input parameters, confirming that the “derived parameters” are those which really control the behaviour of the radionuclides in the repository and consequently, their doses.

As seen in Table 16 the transformation into ranks leads to a significant increase in the values of the correlation coefficients. This confirms the monotonic relationship between input variables and the peak dose.

Table 16 Correlation and regression analyses for total, I129, Cl36 and Cs135 peak doses.

	Total (*)		I129		Cl36		Cs135	
	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC	CC SRC	RCC SRRC
Waste(3)								
R1000	0.107	0.152	0.131	0.153	0.108	0.112	0.092	0.049
GAPX	0.307 (I)	0.384 (I)	0.403	0.394	0.427	0.403	0.030	
Canisters (3)								
MEAN	-0.056	-0.047	-0.071	-0.046	-0.108	-0.080		
SLOPE	0.026		0.037		0.056	0.036		
Near field (6)								
POR ANIONS BUFFER		-0.022		-0.023	-0.050	-0.050		
WATER NF	0.352	0.565	0.400	0.558	0.084	0.112	0.189	0.102
SOLX	0.084(U)							
KDBEX	-0.161 (I)	-0.245 (I)	-0.221	-0.252			-0.070	-0.032
TT NF	-0.399	-0.618 (I)	-0.495	-0.615	-0.166	-0.119	-0.151	-0.107
Water flow (4)								
WATER TT0	-0.330	-0.340	-0.380	-0.334	-0.610	-0.590	-0.295	-0.438
KINEM POR	-0.223	-0.332	-0.297	-0.340	-0.421	-0.542		
PECLET	0.087	0.126	0.122	0.131	0.102	0.102		-0.189
Granite matrix (6)								
FWA	-0.066	-0.061	-0.057	-0.046	-0.075	-0.076	-0.178	-0.526
THICKM	-0.046	-0.042	-0.033	-0.032	-0.051	-0.057	-0.161	-0.365
PORMAT						-0.023		
MATCOR	-0.029						-0.170	-0.378
KDRX							-0.119	-0.270
TT FF	-0.379	-0.504 (I)	-0.465	-0.502	-0.687	-0.862	-0.125	-0.958
DecayT	0.378	0.504 (I)	0.466	0.502	0.753	0.862	0.685	0.958
1/(TT NF+TT FF)	0.579	0.863 (I)	0.667	0.862	0.623	0.878	0.928	0.955
GAP/(TT NF+TT FF)	0.662	0.948 (I)	0.780	0.952	0.720	0.964	0.921	0.919
DecayT/(TT NF+TT FF)	0.579	0.862 (I)	0.667	0.862	0.613	0.877	0.906	0.965

(*) For the total peak dose the chemical element X is shown in brackets

Regression analyses have been done for sets of independent (orthogonal) parameters that can be random inputs or derived parameters. For independent parameters the regression coefficients are equal to the equivalent correlation coefficients: Standardized Regression Coefficients (SRCs) are equal to the Correlation Coefficients (CCs) and the Standardized Rank Regression Coefficients (SRRCs) are equal to the Rank Correlation Coefficients (RCCs). The coefficient of determination (R^2) for independent parameters is calculated simply as the sum of the squares of the correlation coefficients: $R^2 = CC_1^2 + CC_2^2 + \dots$

Results of the regression analyses for the peak dose due to Se79 are presented in Table 17. Each "regression analysis" is presented in three columns: a column of wedges that identify the parameters included in the analysis, the SRCs and the SRRCs. Table 17 shows that the variance of the peak dose due to Se79 can be explained by the variance of a few (or even just one) derived parameters, specially when the transformation into ranks is done.

It is noteworthy that the variance of the peak dose due to Se79 is totally controlled by a single "derived parameter" MF NF*DecayT, that is a function of the Solubility of the Selenium, the water flow leaving the near field (WATER NF), the travel time of Selenium in the far field (TT FF(Se)) and the half-life of Se79.

Table 17 Regression analyses for different sets of random input and derived parameters and the peak dose due to Se79

Se79		SRC	SRRC		SRC	SRRC		SRC	SRRC		SRC	SRRC
SOLSE	►	0.492	0.627	►	0.492	0.627						
WATER NF	►	0.437	0.582	►	0.437	0.582						
TT FF	►	-0.120	-0.395									
DecayT				►	0.220	0.395	►	0.220	0.395			
MF NF							►	0.837	0.863			
MF NF*DecayT										►	0.974	0.995
R ²		0.447	0.888		0.481	0.898		0.749	0.901		0.949	0.990
R ² with all the random input parameters											0.475	0.887

4.3.4.2 Time evolution of correlation and regression coefficients

Correlation and regression analyses (both in values and ranks) have been done for the time dependent total dose in the stochastic calculation with 25,000 runs. The results of the analyses in ranks are shown in Figure 64, which presents all the random input parameters whose correlation coefficient is greater than 0.1 (in absolute value) at least in one instant.

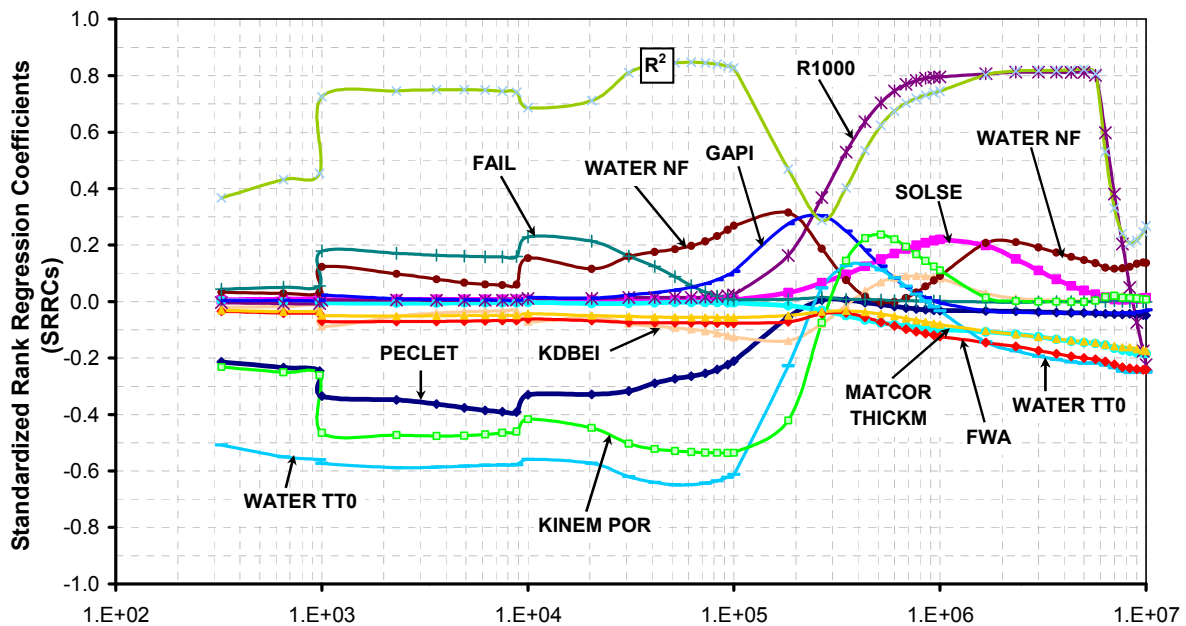


Figure 64 Time evolution of the Standardized Rank Regression Coefficients (SRRCs) and the coefficient of determination (R^2) for the total dose.

Around 1,000 and 10,000 years the curves in Figure 64 have strange jumps that are related to the code used in the calculations (at those instants the dose becomes non-zero in a great number of runs). In any case, doses are very small up to 10,000 years.

In the calculations the minimum duration of the canisters is 20,000 years. Only a few canisters (FAIL) are assumed to present a defect that causes their failure in about a hundred years. As a consequence, FAIL is relevant at the beginning of the calculations but becomes irrelevant when many more canisters have failed.

This class of figures show the parameters that control the dose at each time, but must be used with care. The UO_2 matrix alteration rate (R1000) seems to be far more important than the Instant Release Fraction of iodine (GAPI), but peak total dose is controlled by GAPI. When the I129 inventory in the gap has been released from the formation, ulterior I129 doses are roughly proportional to R1000.

Figure 64 presents a significant decrease in the value of the coefficient of determination (R^2) between 100,000 and 400,000 thousand years. The reason for such drop was traced back to the change of sign of the regression coefficients of several important parameters for the dose due to the I129 in the Instant Release Fraction (IRF) of the inventory. Since total doses between 100,000 and 400,000 thousand years are controlled by the I129 in the IRF, the effect was transmitted to the total dose.

4.3.4.3 Monte Carlo filtering (Smirnov, Mann-Whitney and *t* tests)

Monte Carlo filtering has been tested with the peak doses of different radionuclides. The output variable (peak dose) sample is divided in two subsets and the different statistics test whether the inputs associated to those subsets are different or not.

Three different criteria are used to generate the 2 subsets: 10% of runs producing highest doses vs. the remainder 90% (10%-90%), 10% of runs producing lowest doses vs. the remainder 90% (90%-10%) and runs that cause a peak dose equal to zero vs. runs that produce a non-zero peak dose (zero vs. non zero).

- Monte Carlo filtering was done using Smirnov, Mann-Whitney and *t* two-sample statistics and associated tests. Mann-Whitney statistic was modified to make it independent of the number of runs and to take values between -1 and 1 (M-W*). *t* statistic was modified to make it independent of the number of runs (*t**).
- Table 18 presents some of the result obtained using Monte Carlo filtering. Since peak doses are controlled by I129, very similar results are obtained for the peak total dose and the peak dose due to I129. For I129 the parameters with high absolute values for the 3 statistics in the 10%-90% case and the 90%-10% case are roughly the same. For all the parameters the Mann-Whitney* statistic has different sign in the 10%-90% and the 90%-10% cases: if the run with highest doses are related with high values of a given parameter, then the runs with lowest doses are related with the low values of the same parameter, and vice versa. This is a logical result, consequence of the monotony of the model.

Table 18 Monte Carlo filtering for peak total dose (10% of highest doses) and I129 peak doses (10% of highest doses and 10% of lowest doses). Values of the statistics.

	Total (10%-90%)			I129 (10%-90%)			I129 (90%-10%)		
	Smir	M-W*	t*	Smir	M-W*	t*	Smir	M-W*	t*
Waste(3)									
R1000	0.081	0.101	0.055	0.084	0.104	0.057	0.256	-0.324	0.173
GAPI	0.406	0.516	0.278	0.429	0.548	0.297	0.334	-0.441	0.236
Canisters (3)									
MEAN	0.104	-0.126	0.066	0.104	-0.126	0.066			
SLOPE	0.059	0.064	0.033	0.062	0.068	0.035			
Near field (6)									
WATER NF	0.438	0.541	0.270	0.428	0.530	0.261	0.578	-0.677	0.208
SOLU			0.028						
KDBEI	0.162	-0.201	0.105	0.177	-0.225	0.118	0.294	0.381	0.202
TT NF	0.505	-0.593	0.208	0.504	-0.594	0.208	0.633	0.737	0.713
Water flow (4)									
WATER TT0	0.451	-0.554	0.317	0.470	-0.542	0.308	0.161	0.212	0.108
KINEM POR	0.233	-0.310	0.136	0.251	-0.335	0.149	0.216	0.283	0.187
PECLET	0.086	0.108	0.059	0.098	0.124	0.067	0.112	-0.135	0.066
Granite matrix (6)									
FWA	0.073	-0.095	0.048	0.056	-0.063	0.038		0.035	0.020
THICKM		-0.050	0.026						
TT FF	0.574	-0.689	0.278	0.577	-0.690	0.280	0.272	0.333	0.259
1/(TT NF+TT FF)	0.759	0.877	0.644	0.763	0.881	0.635	0.733	-0.882	0.200
GAPI/(TT NF+TT FF)	0.850	0.942	0.895	0.865	0.955	0.904	0.880	-0.950	0.200

4.3.4.4 Cobweb plots and other graphical results

Conditional cobweb plots, where only a fraction of the runs obtained are represented, were tested, with analyses focused on the runs that produce the 10% highest doses.

Due to the great number of lines represented (500 for the 10% of highest doses in a 5,000 run calculation) cobweb plots can only present a few parameters in the X axis, the graphic becomes blurred and only strong correlations can be identified. Finally, cobweb plots can produce visual artefacts leading to the identification of non-relevant parameters as having an effect on the peak values of the output variable.

An alternative graphical representation has been developed using only the mean ranks of the parameters in the sub-sample selected. The mean ranks for all the parameters relevant for a radionuclide (about 20 input parameters plus 3 to 5 derived parameters) for several conditions on the output variable are represented in a single figure. In our case the results for the 10% or runs with highest doses (Upper 10%) and the lowest doses (Lower 10%) are represented together. Figure 65 is an example of this class of graphic for Se79.

Parameters that have no effect on the highest/lowest doses have a mean rank around 12500.5. Mean ranks close to the 1250.5 or 23750.5 indicate a strong relation of the low/high values of parameter with the smallest or the highest peak doses. In general, for a given parameter the points for the upper and lower 10% of runs present some symmetry (using the line Mean rank=12500.5 as symmetry axis): if the high values of the parameter are related with the higher/lower doses, then the low values of the parameter are related with the lower/higher doses. Nevertheless, the importance of a given parameter can be different for higher and lower doses: for instance, the relevance of FWA for the highest peak doses due to Se79 is much smaller than for the lowest peak doses due to Se79.

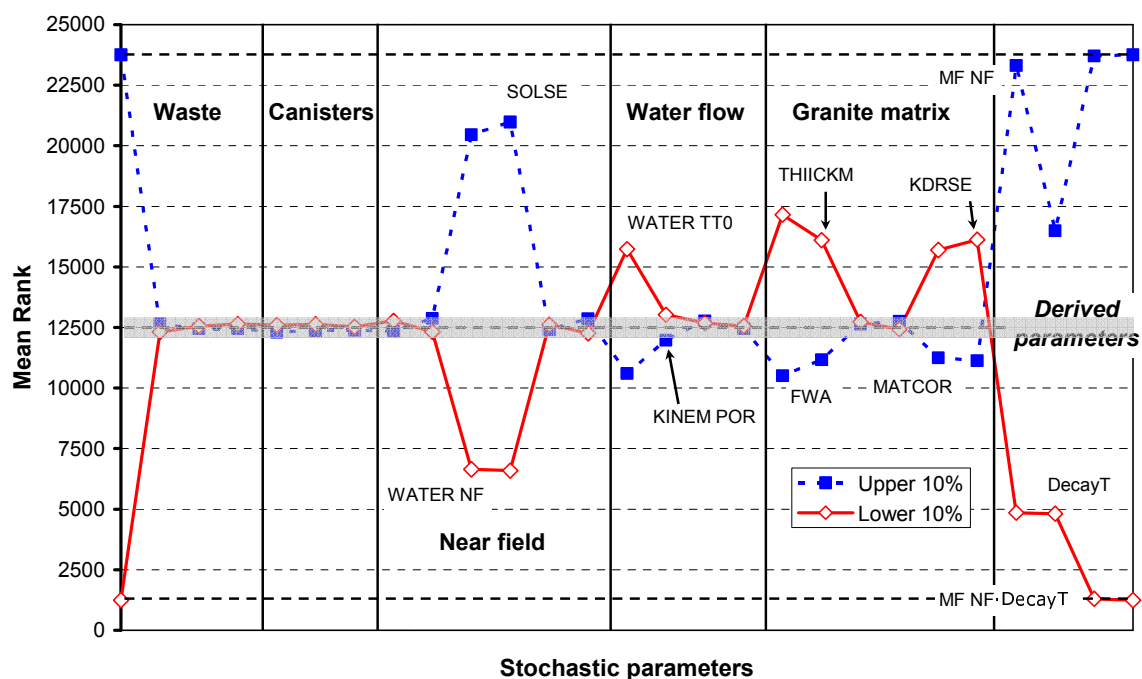


Figure 65 Mean ranks for the independent and derived parameters for Selenium in the 10% of runs with the highest and lowest peak dose due to Se79 (Upper and Lower 10%).

4.3.4.5 Contribution to the sample mean (CSM) plots.

In the CSM plots the y-axis represents the fraction of the output variable (in our calculations, peak dose due to a radionuclide) mean corresponding to values of an input parameter smaller or equal to a given quantile while the x-axis represents the empirical cumulative dis-

tribution function of the parameters. The test statistic adopted is the “maximum vertical distance to the diagonal” (D_m).

Low values of the test statistic mean that the whole range of the input parameter has a similar contribution to the mean of the peak dose. By the contrary, high values of the statistic mean that the contribution of the parameter to the mean is controlled by a small range of values.

For I129 the test statistic takes small values for all the input parameters. By the contrary, for Ni59 the mean of the peak dose is controlled by the 5% of runs with lowest values of the travel time in the far field (TT FF), as can be seen in Figure 66.

The importance of TT FF for Ni59 was already identified in the correlation analysis and the Monte Carlo filtering analysis but the CSM plot allows to identify exactly the range of values of TT FF that really control the doses. This class of information can be useful to guide R&D efforts: if the lowest 5% of values of TT FF for Nickel can be ruled out (by additional investigations) the doses due to Ni59 will decrease several orders of magnitude.

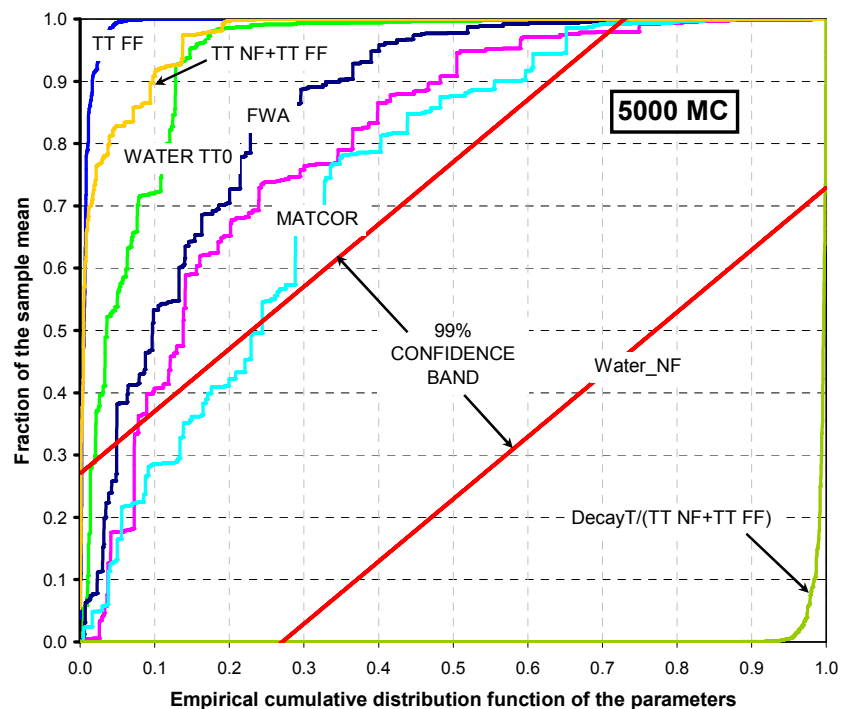


Figure 66 CSM plot for relevant parameters and the peak dose due to Ni59 (output variable).

4.3.4.6 Ranking of parameters

Comparison of the values of the statistics generated in the previous sections shows that for a given radionuclide the parameters identified as relevant for the peak dose are nearly the same for the different statistics. Although each statistic provides a particular information on

the parameters, the same small group of parameters controls the doses due to a given radionuclide in all the cases.

The absolute values of the statistic can be used to produce an importance ranking of the parameters for each statistic, as shown in Table 19 for the peak doses due to Se79. The rankings obtained for each radionuclide are quite similar for the 6 statistics, and the most relevant parameters always are “derived parameters”.

Table 19 Importance ranking of the main parameters affecting Se79 peak dose. (Importance(1)>Importance(2)>....)

	Values				Ranks	
	CC SRC	Smirnov 10%-90%	t* 10%-90%	CSM	RCC SRRC	M-W* 10%-90%
Near field (6)						
POR ANIONS BUFFER	12			12		
WATER NF	4	4	4	4	4	4
SOLSE	3	3	3	3	3	3
MF NF	2	2	2	2	2	2
Water flow (4)						
WATER TT0	6	6	6	6	7	6
KINEM POR	11	11	11	11	11	11
Granite matrix (6)						
FWA	7	7	7	7	6	9
THICKM	8	9	9	9	8	8
MATCOR	10	10	10	10	10	10
KDRSE	9	8	8	8	9	7
DecayT	5	5	5	5	5	5
MF NF*DecayT	1	1	1	1	1	1

4.3.5 Conclusions and recommendations

The different statistics used in the sensitivity analysis identify the same set of relevant parameters with a similar ranking of importance. This agreement confirms the robustness of the results obtained in the sensitivity analysis. The different sensitivity analysis methods are complementary, and each one provides specific information on the parameters.

The capability to identify small effects in the stochastic model increases with the number of runs, as expected. The threshold value above which an statistic is considered statistically significant decreases with the number of runs n following the “rule of thumb”:

$$\text{threshold} \propto \frac{1}{\sqrt{n}}$$

The sampling method used (Simple Random Sampling - SRS or Latin Hypercube Sampling: LHS) makes no difference in the correlation and regression analysis. No formal comparison of both sampling schemes has been done for the rest of statistics. SRS is preferred from a practical perspective because it allows performing several parallel stochastic calculations using different computers (or cores of a multi-core PC) and then append the runs to produce a calculation with many runs (25,000 in our case). With LHS this approach is not possible and the 25,000 runs must be calculated in the same batch.

In the time dependent regression analysis for the total dose (Figure 64) a significant decrease in the value of the coefficient of determination (R^2) between 100,000 and 400,000 thousand years was observed. This is a real effect that was traced back to the change of sign of the regression coefficients of several important parameters with the dose due to the I129 released from the gap. When a change of sign of important parameters and the resulting (and unavoidable) drop in the value of R^2 is observed in a time dependent regression analysis, the reasons of the drop should be identified.

Use of “derived parameters” in the sensitivity analysis

Using the experience gained in the Safety Assessment it is possible to identify some parameters that have an important effect on repository behaviour and are a combination of some random input parameters. These parameters have been called “derived parameters” in this document. The “travel time in the far field for a radionuclide” in the case of a repository in granite is an example of a “derived parameter”.

“Derived parameters” should be identified at the beginning of the sensitivity analysis and treated in the same way than the random input variables through the analysis. It must be kept in mind that these “derived parameters” are correlated with some input parameters.

In the sensitivity analysis performed, “derived parameters” always got the highest values of the different statistics, showing that they have the greatest influence on the peak doses. The rankings of importance of the parameters (section 4.3.4.6) for different statistics calculated are always led by “derived parameters”.

Use of non-relevant parameters to derive threshold values for the statistics

The random input parameters that are known for sure that can not have any effect on the output variable can be used to derive empirical values for the minimum value of a statistic that is statistically significant (empirical thresholds). These “empirical thresholds” have been

calculated for correlation coefficients, Smirnov, Mann-Whitney and t-tests and CSM plots, and are in good agreement with the “predicted thresholds” based on statistics theory.

In the model used any output variable can be affected by 22 random input parameters at most and there are about 110 non-relevant random input parameters, because all the input parameters were independent. In other models input parameters can be correlated and non-relevant parameters can be scarce. In this case “dummy” random parameters could be included in the model with pdf’s similar to the real input parameters. These dummy parameters would be sampled in the stochastic calculation but would play no role in the calculations, and could serve to derive “empirical thresholds” for statistics.

Monte Carlo filtering

It has been found that is more useful to present the results of the Monte Carlo filtering in terms of the statistics instead of the p-values. The statistics tend to the “real” value when the number of runs increases while the p-values tend to zero. The statistics are more robust parameters that allow making a ranking of the relative importance of the input parameters.

Mann-Whitney and t-test statistics were modified to make them independent of the number of runs, and in the case of Mann-Whitney statistic to have a range of variation between -1 and 1. The resulting M-W* and t* statistic were found more useful than the original M-W and t statistics to present the results of the Monte Carlo filtering analysis.

The t* statistic in general produces results consistent with the other statistics (the same parameters are identified as significant with a similar ranking), but some anomalies with the t* statistic have been observed. Since the t-test is based on the sampling distribution of the mean of normal variables, and none of the input random or derived parameters in the model have normal distributions, the strange results could be a consequence of the not applicability of the t-test to the model and parameters under study. As a consequence, it is recommended to use only Smirnov and Mann-Whitney tests for Monte Carlo filtering in values and ranks respectively, and avoid the use of the t statistic.

Graphic representations (cobweb plots)

Due to the great number of lines represented (one per run), cobweb plots can only represent a few parameters in X axis, they can become blurred and only strong correlations can be identified visually. In addition, cobweb plots are prone to visual artefacts.

An alternative graphic representation (Figure 65) has been developed where only the mean ranks of the parameters in several subsets of the sample are represented. The figure presents the mean ranks for all the input parameters relevant for the dose due to a radionuclide in the 10% of runs with highest peak doses and in the 10% of runs with lowest peak doses. Input parameters are grouped on the basis of the part of the system that affect, and “derived parameters” are included too.

This type of plot provides much information in a condensed way: which input parameters do not affect at all the output variable, whether high values of the output variable are related to low or high values of each input (or derived) parameter and whether low values of the output variable are related to high or low values of the same input (or derived) parameters.

Contribution to the sample mean (CSM) plots

CSM plots are visual tools useful to identify if the mean value of the output variable is controlled by a particular region of values of an input parameter. In addition an statistic is calculated. Low values of the test statistic mean that the whole range of the input parameter has a similar contribution to the mean of the peak dose. By the contrary, high values of the statistic mean that the contribution of the parameter to the mean is controlled by a small range of values.

The CSM statistic allows selecting the parameters whose contribution to the mean is controlled by a small region of values, and the CSM plots are used to identify exactly such region of values.

Effects of a stochastic treatment of the Biosphere

Most of the results obtained in the sensitivity analysis of a calculation done using a constant biosphere remain valid for a case with stochastic biosphere: important non-biosphere parameters and their ranking of importance remain the same. The numerical values of the statistics will decrease with the introduction of the stochastic biosphere.

The effects of an stochastic biosphere in the sensitivity analysis of Safety Assessment calculations can be broadly predicted on the basis of the additional variance introduced by the stochastic biosphere.

4.4 GRS application case: generic rock salt repository

4.4.1 Background

In Germany a number of sensitivity analysis studies for final repository systems were performed using Spearman's rank correlation coefficients (SPEA), partial rank correlation coefficients (PRCC), standardised rang regression coefficients (SRRC) and the Smirnov test. It was found, however, that these methods are not always adequate to reveal the actual sensitivities of the model under consideration. If, for example, there is a high probability of zero output, as it is typical for repositories in rock salt, the mentioned methods perform rather poorly. Therefore, more sophisticated methods of sensitivity analysis, based on the variance of the model output, have been tested in the investigation described here.

One candidate host formation for a deep geological repository in Germany is rock salt, as there are large salt structures in the north of Germany and salt has got some beneficial prop-

erties. Repository systems in rock salt typically show specific behavioural patterns, which cannot be observed in other host rock formations. For this reason, the investigations performed in the work described here were confined to a rock salt system.

All openings in rock salt formations are subject to convergence. Due to the plastic creep properties of rock salt open voids decrease and are finally closed under the rock pressure. For final disposal of hazardous substances this effect can be favourable or unfavourable. On one hand the convergence leads to permanent watertight enclosure of the waste containers if it can act long enough without being disturbed. On the other hand, if by some reason the substances get in contact with brine before the openings are closed, the convergence will drive the contaminated fluid through and out of the repository structure and accelerate the release to the biosphere. It is most probable that the wastes are tightly enclosed by salt before they can get in contact with brine and the probability of contaminant release is considered rather low, but in case such a release takes place it could cause relatively high doses.

The computer code LOPOS is the rock salt module of the EMOS package for long-term safety analysis calculations for final repository systems. It yields a zero output if the convergence leads to a tight enclosure of the wastes before they get in contact with brine. For this reason there are a high number of zero-runs in a typical probabilistic set of calculations, which is a situation that can cause strange behaviour of sensitivity analysis methods.

For the investigations described here, a "realistic" generic HLW/SF rock salt repository model was selected as a test case for the different sensitivity methods, since it shows typical important properties of a PA system. Firstly, the distribution of radiation exposure is highly skewed and heavily-tailed which typically spans over several orders of magnitude. Secondly and thirdly, the system of the test case shows non-linear and non-monotonic behaviour. A complete repository setup in rock salt formations is considered in the test case, including the near field, the far field and the biosphere. As model output the annual effective dose to an adult human individual is calculated with the software package EMOS.

4.4.2 *Test case*

A generic performance assessment study for a hypothetical SF/HLW repository in a German salt formation has recently been made. On the basis of this study a test system was set up, which was used for testing different methods and techniques for sensitivity analysis. This system shows the typical properties of repositories in rock salt. In the normal evolution scenario there is no release of contaminants. Therefore, a disturbed evolution scenario with shaft seal failure and dam failure was chosen for the investigation.

The generic SF/HLW repository model represents a hypothetical repository for High-Level-Waste (HLW) and Spent Fuel (SF) to be set up in a German rock salt formation. It consists of eight emplacement fields for SF, one for HLW and one for intermediate-level waste (ILW). From the inner transfer drifts boreholes are drilled to a depth of 300 m, including 10 m for a plug for closing the boreholes. Each emplacement drift contains three boreholes. The dis-

tance between the central field and the waste sections of the repository is about 450 m. The inner drifts are replaced by one model drift per field. It is assumed that the total amounts of high-level and heat-generating waste that are expected to accumulate in Germany until 2080 are disposed of. The main engineered barriers are the shaft seal, the drift seals, the borehole plugs and the backfill material. A graphical representation of the model structure is given in Figure 67.

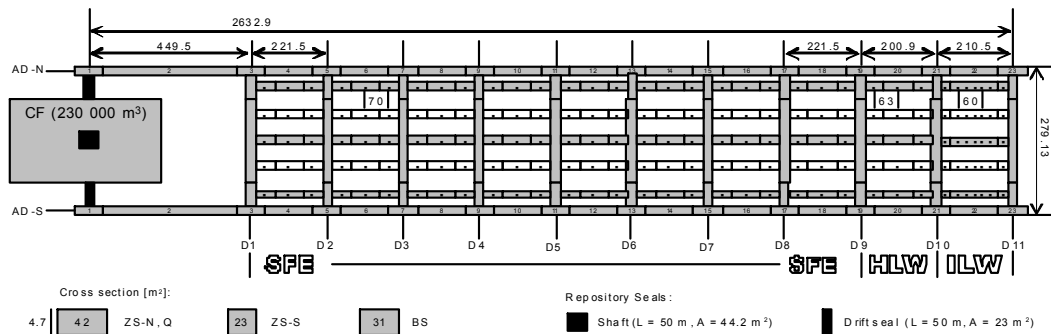


Figure 67 Model Structure for the Generic Repository

Although it is rather improbable, a shaft failure scenario was selected for investigation, in order to demonstrate that the consequences remain low even in such a situation. Six uncertain parameters (4 near field and 2 far field) were considered for the probabilistic investigations. Table 20 lists the parameters along with their distribution types and ranges.

Table 20 Parameter distributions and ranges of the parameters of the test case

Number	Parameter	Distribution	Minimum	Maximum
1	Shaft permeability after failure [m²]	Log uniform	1×10^{-17}	1×10^{-14}
2	Dam permeability [m²]	Log uniform	1×10^{-17}	1×10^{-14}
3	Shaft failure time [yr]	Uniform	25	75
4	Reference convergence rate [1/yr]	Log uniform	1×10^{-3}	1×10^{-1}
5	Relative sorption coefficients in the far field	Log uniform	0.1	10
6	Diluting water flow in the aquifer [m³/yr]	Uniform	1×10^4	8.6×10^4

The sorption coefficients (K_d -values) in the far field are element-specific. Their values are more or less independent of each other, but for the present investigation they are varied together. That means that there is only one random variable, representing a common factor, by which the reference values are multiplied. This factor is varied between 0.1 and 10, where 10 means high sorption and 0.1 low sorption of all elements.

4.4.3 Sensitivity Analysis

The sensitivity analysis was performed using a number of different methods. Moreover, scatterplots, contributions to the sample mean plots (CSM plots) and cobweb plots were utilised as graphical methods to find important parameters.

Four different samples were generated and used for the analysis. EFAST requires a specific sampling scheme, which was applied for two samples with sizes of 3030 and 6054. The other two samples were produced using a random sampling scheme. For optimal comparability, the sample sizes were also chosen as 3030 and 6054.

4.4.3.1 Analysis of the maximum dose rate

The investigations described in this chapter only consider the peak values of the individual runs, regardless of their times of occurrence. By such an analysis the sensitivity of the maximum model output against variations of the input parameters is investigated.

Input and Output Analysis

Figure 68 presents the time-evolution of the calculated dose rate for all simulations using the random samples with 3030 and 6054 model runs. The majority of simulations predict zero or very low dose rates, while a few ones lead to rather high values. The result is a highly skewed and heavily-tailed distribution of the calculated maximum values, which spans several orders of magnitude. This is illustrated with a frequency plot of the peak annual doses in Figure 69. The zero runs are not included in the figure, but it has to be kept in mind that there are 5136 or 2579 zero results, respectively.

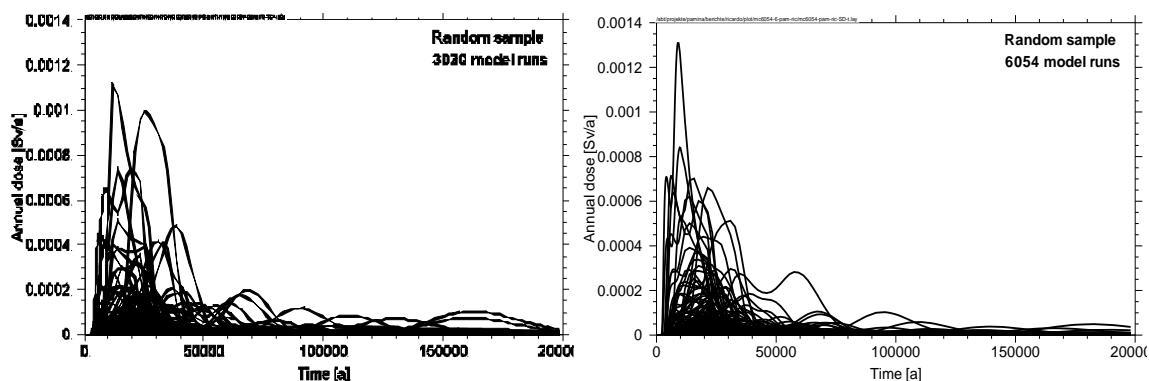


Figure 68 Time evolution of the annual dose rate for all simulations (random samples)

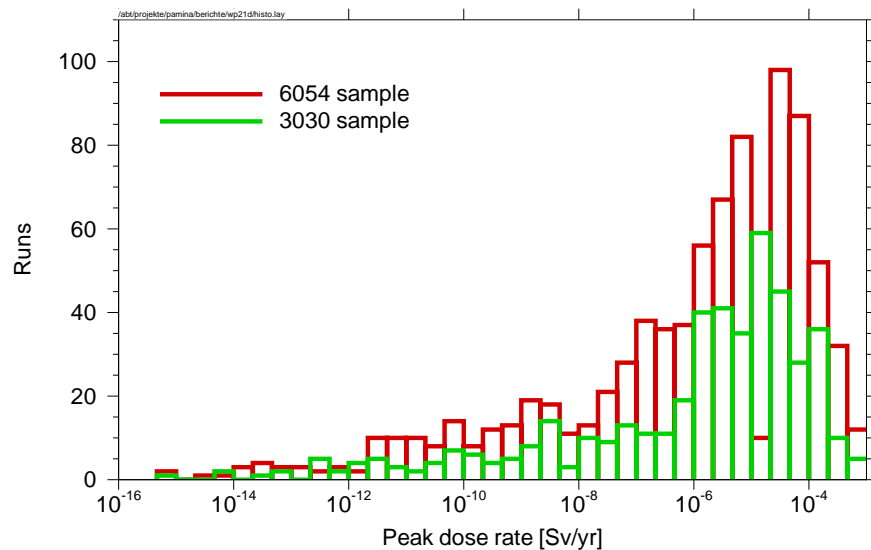


Figure 69 Frequency histogram of the peak dose rates for both EFAST samples

Scatterplots

Figure 70 shows scatterplots for the random-based set of 6054 model runs. For each parameter the calculated maximum dose rates are depicted versus the corresponding parameter values. Since uniform or log-uniform distributions were used for all parameters, there is a unique, but more or less arbitrarily chosen interval for each of them.

Because of the high number of zero-runs only 918 data points can be seen in the figures. The first parameter (shaft permeability after failure) shows a strong accumulation of peak dose rates at the upper end of the range while the values in the lowest decade produce either very low or zero output. The second parameter (dam permeability) shows a similar behaviour, but the data points are distributed over the whole parameter interval. This difference is due to the uncertainty intervals of the two parameters. This is a good example of how the selection of parameter distributions can influence the results of the sensitivity analysis.

The fourth parameter (reference convergence rate) indicates quite the contrary: accumulation of peak dose rates at the lower end of the range and none at all at the upper end. For the rest of the parameters the peak dose rates are more or less homogeneously distributed all over the range.

Just from the appearance of the parameters in the scatterplots, it can be inferred that the shaft permeability, the reference convergence rate and the dam permeability are important, while the rest are less important since they do not have great impact upon the model results.

In the case of the factor for the sorption coefficients at least the clear tendency can be seen that higher sorption leads to lower maxima, which agrees with the expectation.

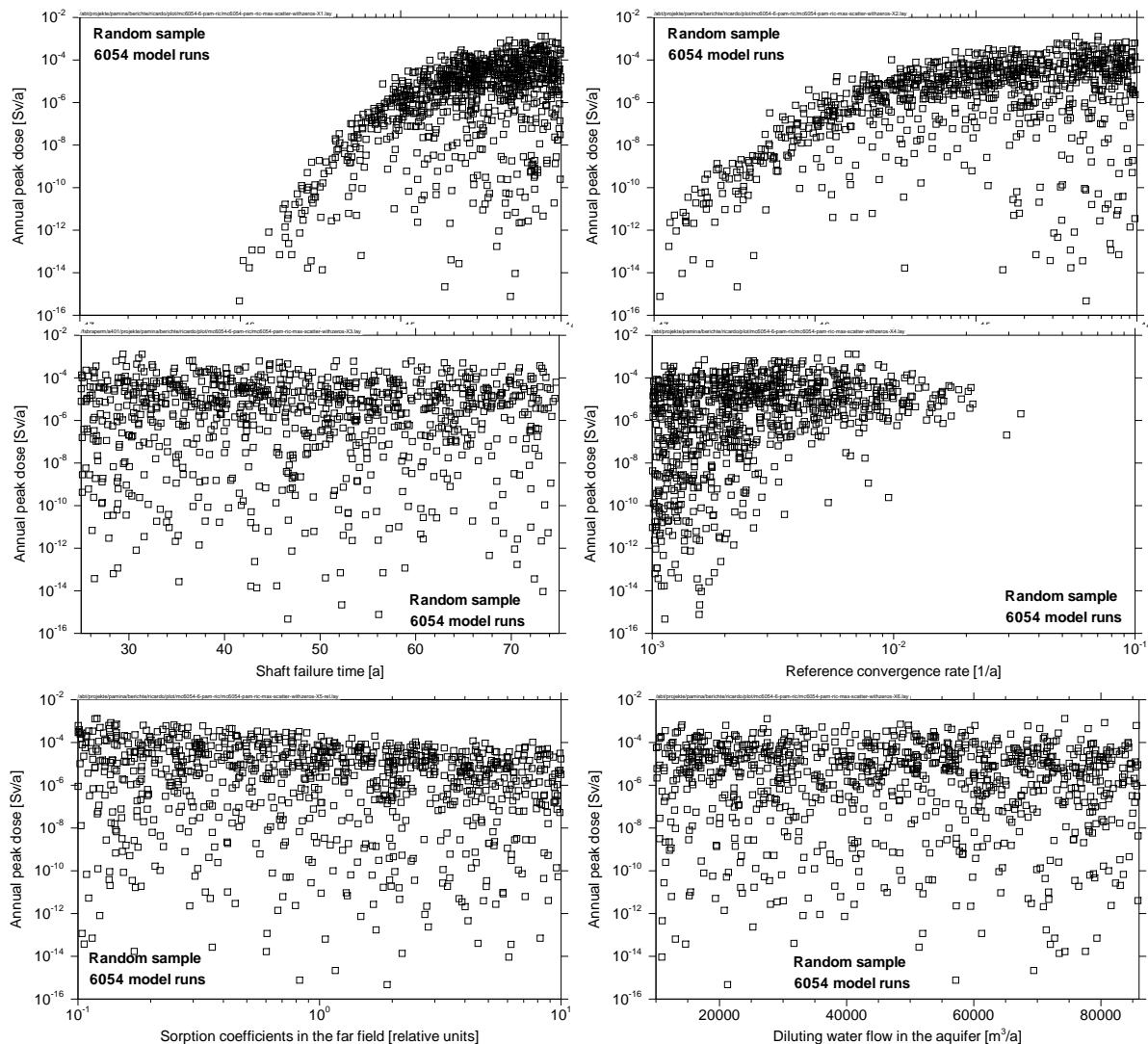


Figure 70 Scatterplots for the 6054 random sample.

CSM plots

Figure 71 presents CSM (Contribution to the Sample Mean) plots for the 6054 runs samples. For this kind of plots the N model runs are ordered according to the values of the input parameter under consideration. Then the fraction of the sample mean of the peak values is plotted versus the fraction of the sample size. As a rule of the thumb, important parameters create curves that significantly deviate from confidence bands around the diagonal, while non-important ones generate curves close to the bands of the diagonal.

A significance band of 25 % is considered in Figure 71. For the four samples in this study, the parameter with the most significant deviation from the diagonal is the shaft permeability, followed by the dam permeability and the reference convergence rate. For the random-based sample, also the factor for the sorption coefficients in the far field is identified as an important parameter. This is in line with what could be seen from the scatterplots.

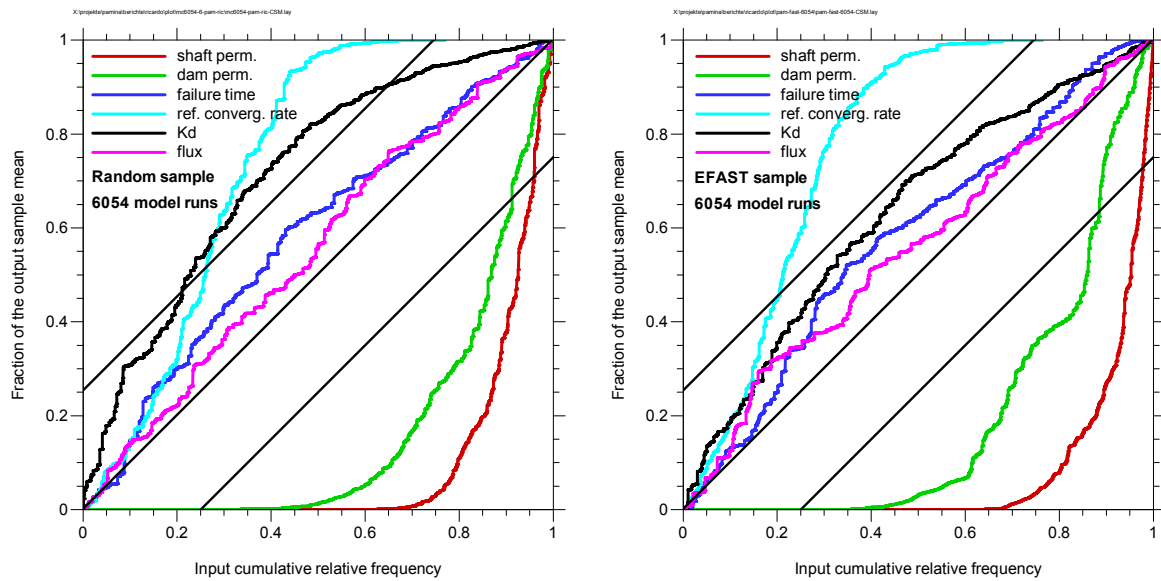


Figure 71 CSM plots for the 6054 runs samples.

For all samples about 70 % of the runs, according to the lowest shaft permeabilities, are responsible for less than 1 % of the mean, since they generate either very low or even zero output. This means that if due to further experimental work, the 30% upper range of this input parameter could be discarded as a possible range of values, the peak dose rate would decrease down to approximately 1% of the estimate obtained in this study.

Cobweb plots and Smirnov test

The two random samples were analysed using cobweb plots, which are a graphical means for showing the importance of the different parameters in one figure, see Figure 72. For each run a line is drawn that connects the ranks of all input parameters and the output value (the peak dose rate). Since the zero-runs are not taken into account, only the highest 15 % of the output ranks appear in the plots. They are predominantly connected to the lower ranks of the reference convergence rate and the higher ranks of the shaft permeability, which shows that these are the most relevant parameters.

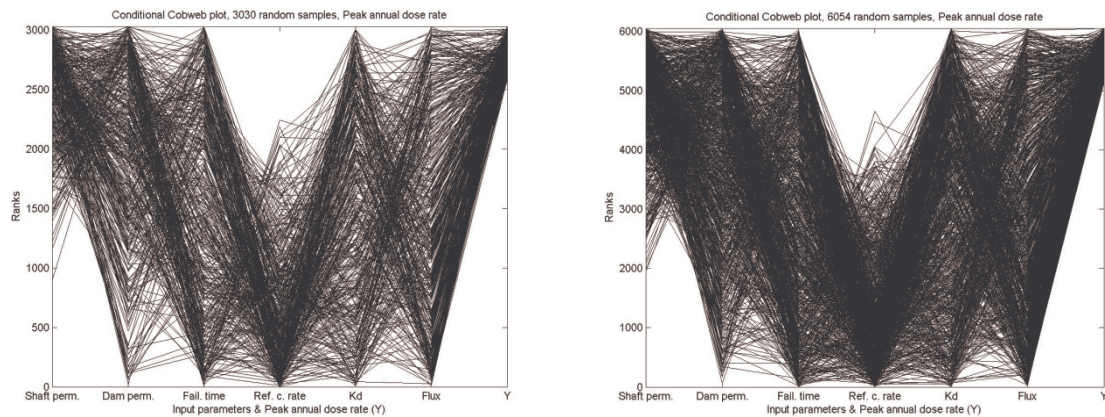


Figure 72 Cobweb plots for the random samples.

For a more quantitative ranking the Smirnov test was performed on both sets of calculations. The results are shown in Table 21. The order of importance calculated by the Smirnov test does not change when passing from 3030 to 6054, except for the two parameters identified as least important (sorption factor and flux). It is, however, surprising that the sorption factor is identified as the least or second least important parameter, while it is on rank 4 and clearly significant according to the CSM plots and, unlike the failure time, can also be identified as relevant from the cobweb plots as well as from the scatterplots. This shows that the results of the Smirnov test are not always reliable.

Table 21 Results of the Smirnov test for the 3030 and the 6054 run random samples.

Input parameter	3030 run random sample p-value / importance order	6054 run random sample p-value / importance order
Shaft permeability	1.01E-107 (1)	3.33E-243 (1)
Dam permeability	4.04E-029 (3)	6.03E-047 (3)
Failure time	2.37E-001 (4)	3.62E-003 (4)
Ref. convergence rate	9.27E-088 (2)	1.29E-203 (2)
Sorption factor	4.03E-001 (5)	3.73E-001 (6)
Flux	5.72E-001 (6)	1.37E-001 (5)

4.4.3.2 Time-dependent analysis

In the following the model is analysed for a number of separate points in time, so that the time developments of the different sensitivity measures can be derived. The evaluations were done using the tool SIMLAB 3.

Correlation-based methods and Smirnov test

Due to technical reasons the random samples could not be used for a time-dependent evaluation. Although EFAST samples are not optimal for evaluation with random-based correlation methods, such an evaluation can be done. The results presented in Figure 73 refer to the EFAST samples exclusively. It shows the time curves of PEAR and SPEA coefficients as well as the time-dependent evaluation of the Smirnov test. The rank-based evaluation identifies the reference convergence rate as the second most important parameter, while the value-based method yields higher absolute coefficients for the dam permeability. The Smirnov tests shows for both samples a clear dominance of the shaft permeability and assesses the dam permeability and the reference convergence rate as nearly equally important. The remaining three parameters are identified as unimportant. Although the sorption factor definitely has an influence on the model, this is obviously not identified by the Smirnov test, which was already found out during the sensitivity analysis of the maxima.

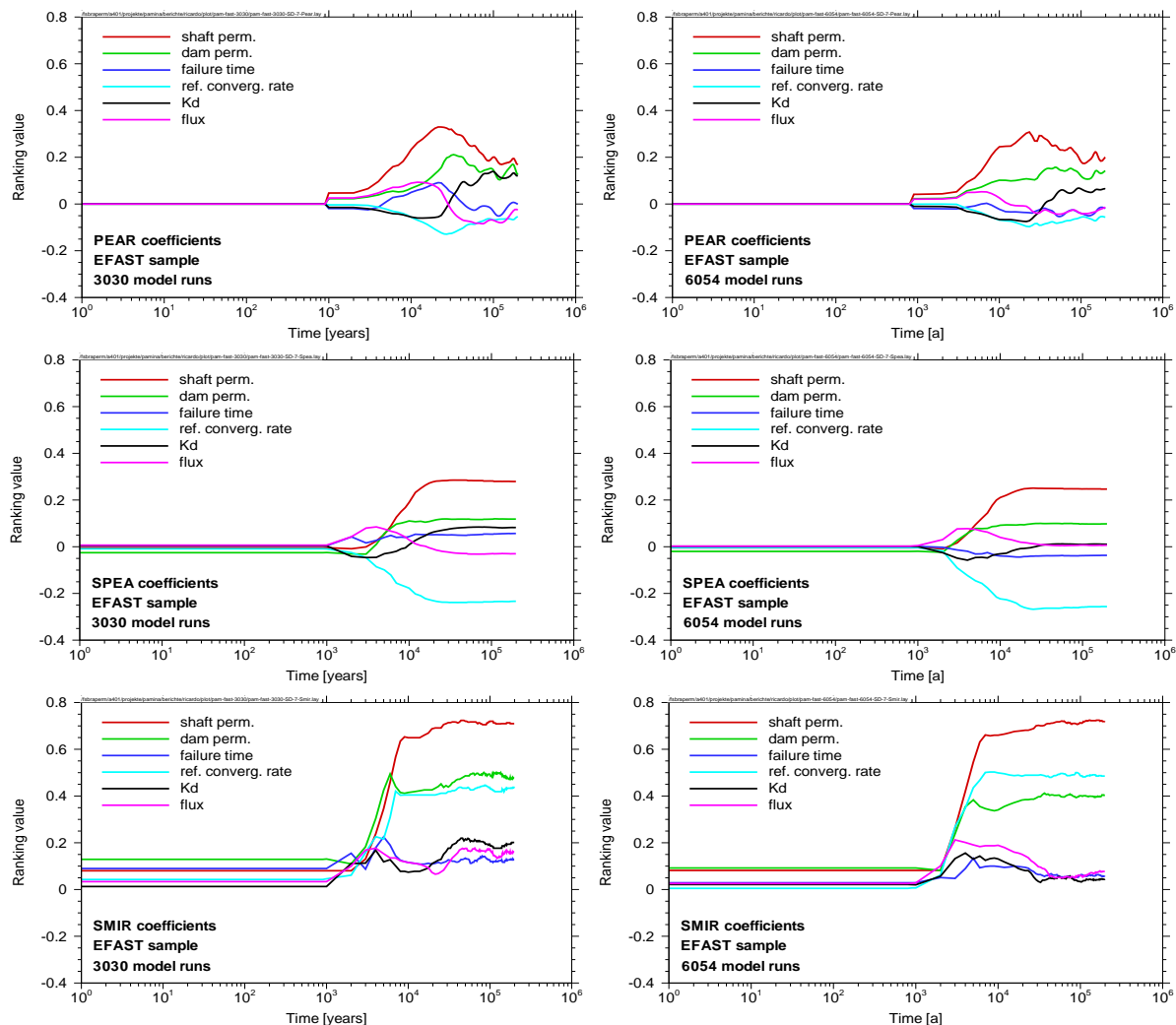


Figure 73 Time-dependent evaluation of random-based SA methods.

Variance-based evaluation

The regression- or correlation-based sensitivity analysis techniques are best suitable for models with a close-to-linear behaviour, which is not a typical property of repository models. Variance-based sensitivity analysis is a means to get around this problem as it does not require linearity or monotonicity.

The principle of FAST is to calculate the first-order sensitivity indices

$$\frac{\text{Var}_{X_j}[E(Y|X_j)]}{\text{Var}(y)}$$

by scanning the parameter space with periodic functions using interference-free frequencies and performing a Fourier analysis on the output. The Extended FAST method (EFAST) [Saltelli et al., 1997] additionally yields the total order sensitivity indices, which describe the influence of a parameter to the output in interaction with all others.

The EFAST method was applied to the model under consideration as a time-dependent evaluation using the two EFAST samples with 3030 and 6054 runs, respectively. The results are unsatisfactory as shown in Figure 74. It is apparent that the shaft permeability is the most important parameter, though it does not reach a sensitivity index of more than 0.15. The other parameters have very low sensitivity indices and show no unique behaviour. The agreement between the two evaluations is poor.

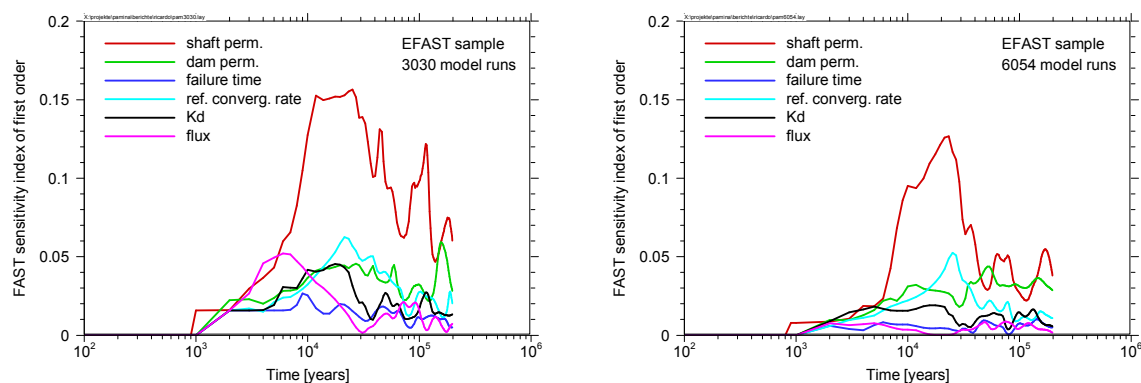


Figure 74 First order EFAST sensitivity index for the two sets.

Only 430 out of 3030 or 885 out of 6054 runs, yield a non-zero output, which means a 15 % probability. Therefore, only these 15 % of the runs can contribute to the calculation of sensitivity indices. Moreover, the periodicity of the model output and with it the Fourier analysis is disturbed by the zero-runs.

The variance of a set of data values can be decomposed in the individual contributions of subsets. In the set of 6054 runs, the 85 % zero-runs are responsible for only a very small fraction of less than 2 % of the variance. On the other hand, the two highest maxima alone

make nearly 20 % and the 12 highest maxima more than 50 % of the variance. It becomes obvious from this investigation that the variance of the model output depends on the randomness of the sample and a variance-based analysis will not be very robust if performed with sample sizes of a few thousand. So far it cannot be said which number of simulations would be necessary for the method to produce stable results with the considered model.

To improve the robustness of the method the following transformation of the model output was done [Becker and Spießl, 2008]:

$$y^* = \log_2(y/a + 1)$$

with some adequate value a . In contrast to simply taking the logarithm this transformation does not overvalue extremely small values and allows even zero. It maps 0 to 0, a to 1 and big values practically to their logarithm. The value a indicates the transition from “low” to “high” values. In the investigation described here it was calculated individually for each point in time in such a way that the expectation of the transformed distribution gets equal to 1. This gives equal statistical weights to “low” and “high” values. With this transformation, applied to the set of the peak output values of the 6054 runs sample, the contribution of the highest value to the total variance reduces from 9.9 % to 0.7 % and that of the zero-output runs increases from 1.9 % to 5 %. Figure 75 shows the variance decomposition of the transformed output before and after the transformation.

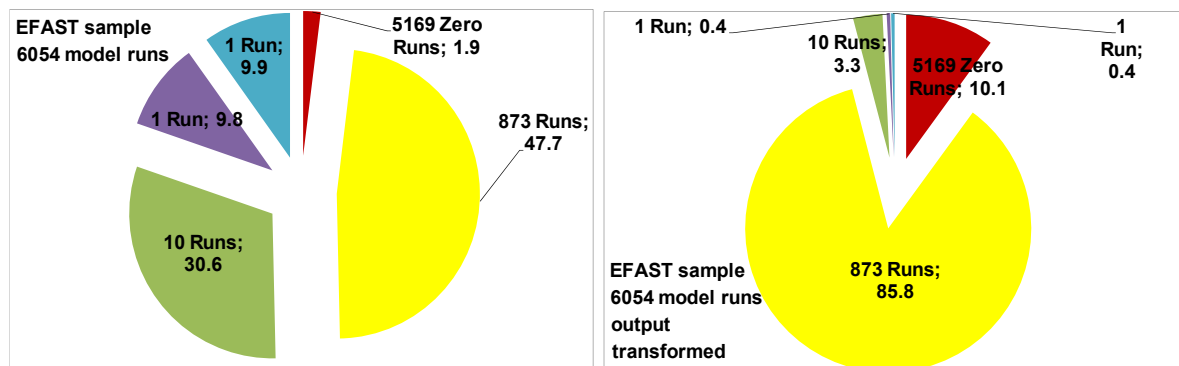


Figure 75 Decomposition of the variance of the untransformed and transformed model output (peak values).

It is obvious that the weighting becomes much more balanced through the transformation, since the contributions of the highest values decrease and those of the zero-runs increase.

In Figure 76 the time-development of the first order sensitivity indices of the transformed model output, calculated using EFAST, is depicted. In comparison with Figure 74 the curves look much smoother and allow a more unique ranking of the input parameters. Although there are still differences between the curves of the two evaluations, they are much more similar and the robustness of the method has obviously increased.

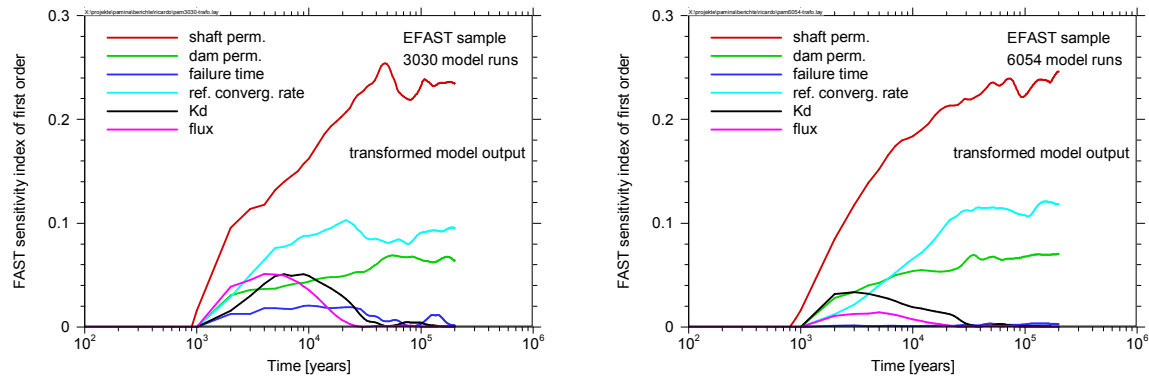


Figure 76 First order EFAST sensitivity index for the two sets (model output transformed).

4.4.4 Summary

All applied methods agree that the most important parameter for the model system is the shaft permeability, followed, in varying orders, by the dam permeability and the reference convergence rate. The remaining parameters seem less important.

The graphical methods (scatterplots, CSM plots and cobweb plots) provide a visual impression of the sensitivity of the model against parameter variations. They allow a quick examination of the parameters which ones are important and which ones are less important.

The correlation- and regression-based methods are easy to apply, but yield relatively unclear results concerning the less important parameters.

For the variance based method EFAST, it was shown that an adequate output transformation can improve both the robustness and the significance of the evaluation. In view of the high proportion of zero-runs the FAST evaluation of the transformed output is surprisingly good. Generally, variance-based sensitivity analysis, if applied and interpreted carefully, seems promising to support the safety case better than regression or correlation methods.

4.5 Facilia application case: Biosphere Assessment

This section presents results from a sensitivity analysis study of a biosphere model, which has been used in the safety assessment of a geological repository for spent fuel. A particularity of the studied biosphere model is that it uses a larger number of uncertain parameters. As a consequence, a large number of model simulations are required in order to perform a full range of sensitivity analyses.

The objective of the sensitivity study presented here was to test the applicability of a graded approach to the analysis, where first screening methods are applied in order to select the most important parameters, which are then studied with more advance, variance based methods. The intention with the graded approach is to reduce the required simulation time

and to provide a clearer picture of the contribution of the different parameters to the uncertainties of the predictions for selected endpoints.

4.5.1 Description of the studied system

The sensitivity analysis study was carried out using a landscape model, which was developed for the Olkiluoto Island on the western shore of Finland (northern Baltic Sea). This island has been selected as the site for a spent nuclear fuel repository in Finland. Since the site is on an area with high post-glacial uplift rate, terrain forecasts and their analysis in the context of the dose assessment are an essential part of the safety case. The model used in the present study is one of four biosphere models representing different assumptions for the landscape in this area, described in [Ikonen *et al.* 2008].

The selected landscape model (Figure 77) consisted of the following interconnected biosphere objects: 5 forests, 4 lakes, 5 rivers, 4 agricultural lands and 2 sea objects. For agricultural lands irrigation with water from adjacent lakes and rivers was also considered. A release scenario with high release rates to the biosphere, as result of gas expulsion from the waste canisters, was selected. The releases were directed to the first forest object (F1 in Figure 77).

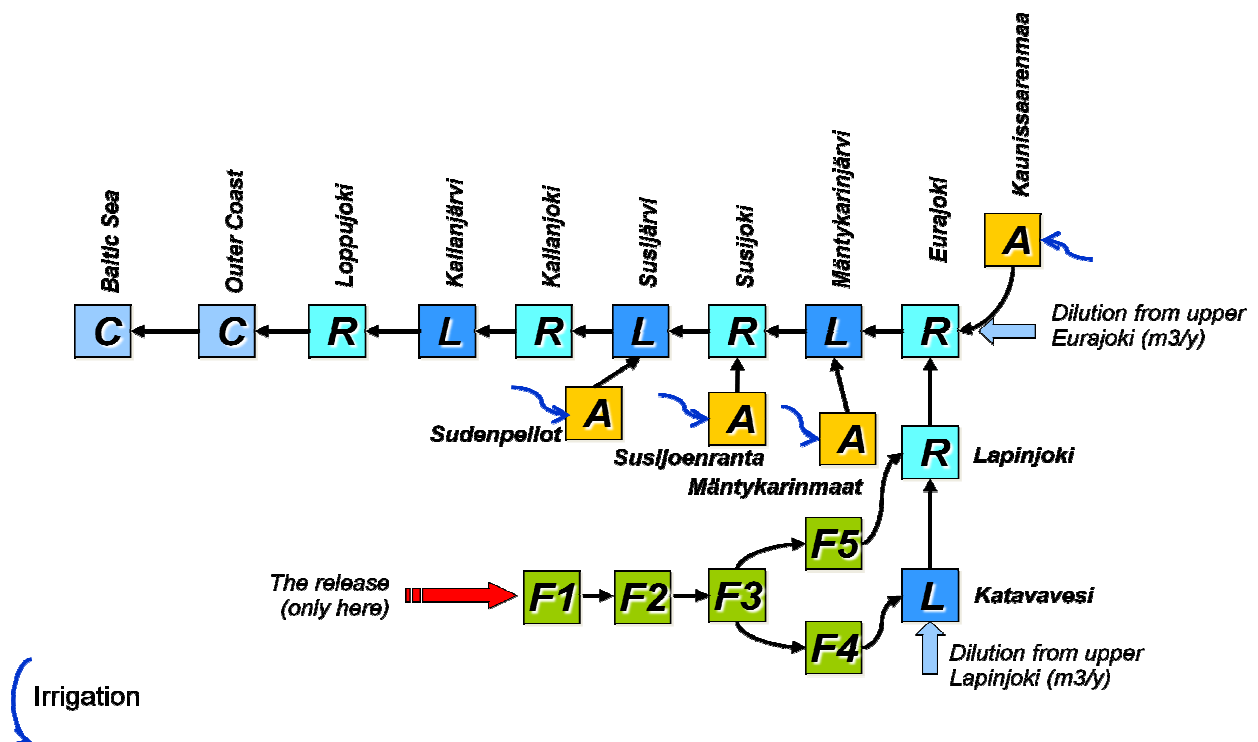


Figure 77: Schematic view of the landscape model used in this study.

4.5.2 Sensitivity analyses performed

The following sensitivity analysis methods were used in the study:

- The Morris Factorial Sampling Method (Morris, 1991) as screening method.
- Monte Carlo based methods. Correlation and regression based statistics were obtained from a probabilistic simulation with 10000 iterations: Pearson Correlation Coefficients (CC), Spearman Correlation Coefficients (RCC), Partial Correlation Coefficients (PCC), Partial Rank Correlation Coefficients (PRCC), Standardized Regression Coefficients (SRC) and Standardized Rank Regression Coefficients (SRRC).
- The variance based method Fourier Amplitude Sensitivity Test (EFAST) described in [Saltelli *et al.* 1999].

4.5.2.1 Graded approach to the sensitivity study

The sensitivity study was carried out in three steps:

1. The Morris Factorial Sampling Method was applied to identify potentially important parameters. In this study potentially important parameters are defined as those with more than 1 % of a predefined ranking criteria obtained as follows: firstly the square root of the sum of the squared values of the Mean and Standard Deviation of each elementary effect is calculated, and secondly the obtained values for each elementary effect are normalized by the sum across all elementary effects.
2. Two Monte Carlo simulations, with 10000 iterations in each, were carried out. In one simulation all uncertain parameters were varied according to the assigned PDFs. In the other simulation PDFs were used only for parameters that were identified as potentially important from the screening study and for other parameters the same value was used in all iterations.
3. For the parameters identified as potentially important from the screening study the EFAST method was applied

4.5.3 Inputs and outputs analysed

The two endpoints considered in the sensitivity analyses were the average and the maximum value of the doses across all landscape objects for the time point when the peak doses are observed within the simulation period (10000 years). The time of the peak dose was found from a deterministic simulation, where best estimate values were assigned to all model parameters. The radionuclides considered in this study were: ¹³⁷Cs and ¹³¹I.

The model has a very large number of parameters, including parameters that define sizes of the landscape objects. In this specific study only parameters which are considered uncertain for a given simulation were included (see Table 22). Other parameters, mainly those describing the size of the objects, are kept constant during the simulations. For such parameters, the contribution to the uncertainty is studied by considering different landscape configurations (scenarios). The type of distribution assigned to the uncertain parameters is indicated in the second column of Table 22. The deterministic values and the values of the distribution parameters were taken from [Broed, 2007].

Table 22 Parameters considered in the sensitivity analyses

Parameter	Type of distribution	Description
LP	Normal	Production of leaves
CR_L	Lognormal	Concentration ratio from soil to leaves
density_soil	Normal	Soil density
h	Normal	Soil thickness
WP	Normal	Production of wood
CR_W	Lognormal	Concentration ratio from soil to wood
TC_WtoLi	Lognormal	Transfer rate coefficient from wood to litter
TC_UtoLi	Lognormal	Transfer rate coefficient from understory to litter
TC_LitoS	Lognormal	Transfer rate coefficient from litter to soil
TC_LtoLi	Lognormal	Transfer rate coefficient from leaves to litter
UP	Normal	Production of understory plants
CR_U	Lognormal	Concentration ratio from soil to understory
Runoff	Lognormal	Sub-surface runoff
theta	Normal	Water content in soil
Kd_soil	Lognormal	Distribution coefficient in forest soils
Kd_lake	Lognormal	Distribution coefficient for suspended matter in lakes
Kd_riversediment	Lognormal	Distribution coefficient in river sediments
Kd_river	Lognormal	Distribution coefficient for suspended matter in rivers
Kd_sea	Lognormal	Distribution coefficient for suspended matter in sea
Kd_agriland	Lognormal	Distribution coefficient in agricultural soils
CRforest	Lognormal	Concentration ratio from soil to forest products
Forest1.Multiple1	Uniform	Catchment area/area of the forest 1
Forest2.Multiple2	Uniform	Catchment area/area of the forest 2
Forest3.Multiple3	Uniform	Catchment area/area of the forest 3
Forest4.Multiple4	Uniform	Catchment area/area of the forest 4
Forest5.Multiple5	Uniform	Catchment area/area of the forest 5

Note: the Kd and CR are nuclide specific - different parameters are used for different radionuclides.

4.5.4 Main results

The results obtained from the sensitivity analysis are presented below.

4.5.4.1 Screening with the Morris method

Figure 78 and Table 23 present the results obtained with the Morris screening method for the total doses from CI-36. It can be observed that 8 parameters are important by themselves and also because of their interactions. Other parameters have marginal impact on the uncertainty of the predictions. The results obtained for the average doses were very similar.

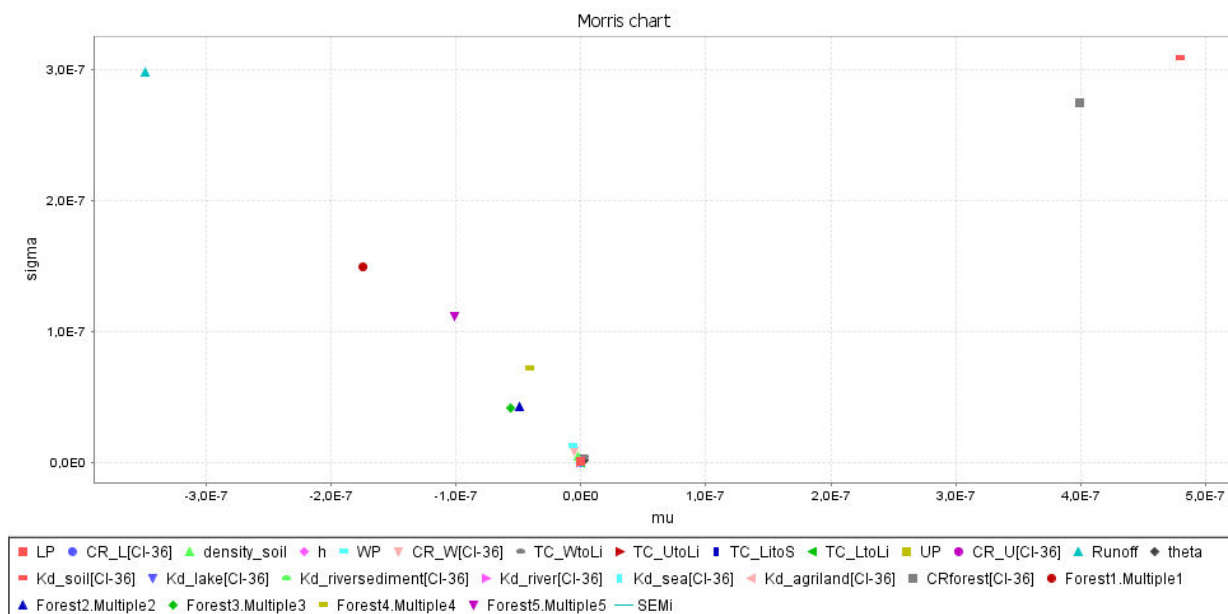


Figure 78 Scatter plot of the mean (μ) and standard deviation (σ) of the elementary effects obtained with the Morris method for the total dose from CI-36.

Table 23 Ranking of parameters, obtained with the Morris method, by their importance for the Total Doses from CI-36

Parameter	Ranking measure ($(\text{Mean}^2 + \text{SD}^2)^{1/2}$)
Kd_soil	26,5%
CRforest	22,5%
Runoff	21,3%
Forest1.Multiple1	10,6%
Forest5.Multiple5	7,0%
Forest4.Multiple4	3,8%
Forest3.Multiple3	3,2%
Forest2.Multiple2	3,0%
Sum for other parameters	2,0%

The most important parameters are the distribution coefficient for forest soils (Kd_soil), the concentration ratio for forest products (CRforest) and the runoff. Interactions between these

parameters are to be expected. For example, for high values of the K_d _soil the runoff should play a less important role in the predictions, as the same radionuclide amount will be retained in the soil independently of the runoff value. The K_d _soil and the CRforest are known to be correlated, as high retention in the soil leads to lower bioavailability of the radionuclides for uptake by plants and consequently lower concentration ratios.

Figure 79 and Table 24 present the results obtained with the Morris screening method for the total doses from I-129. It can be observed that 10 parameters are important by themselves and also because of their interactions. Other parameters have marginal impact on the uncertainty of the predictions. The results obtained for the average doses were very similar. In general the results for I-129 are very similar to the results for Cl-36.

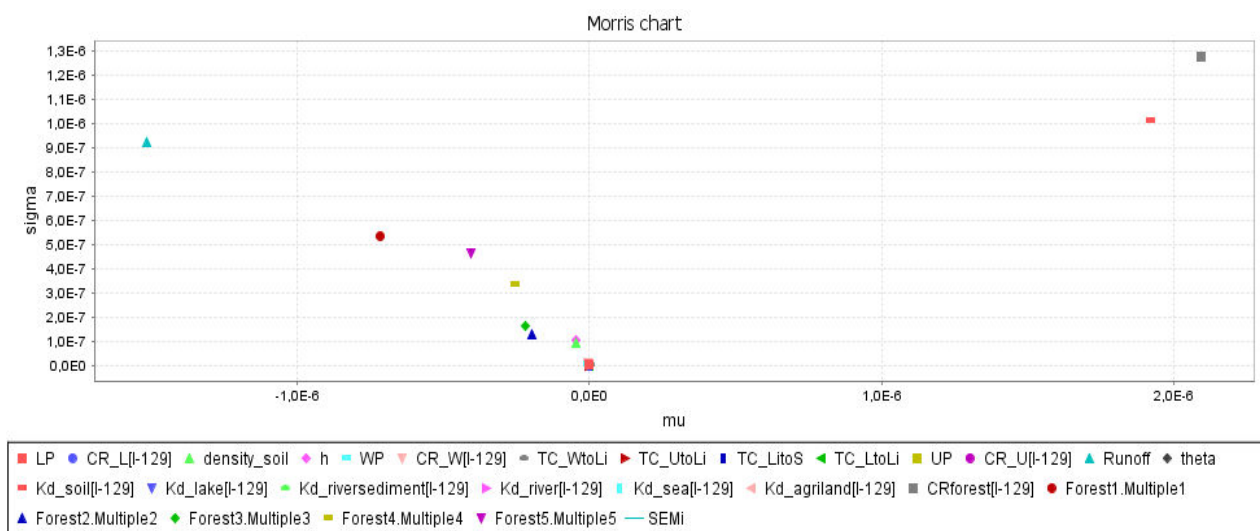


Figure 79 Scatter plot of the mean (μ) and standard deviation (σ) of the elementary effects obtained with the Morris method for the total dose from I-129.

Table 24 Ranking of parameters, obtained with the Morris method, by their importance for the Total Doses from I-129

Parameter	Ranking measure ($\text{Mean}^2 + \text{SD}^2$) ^{1/2}
CRforest	27,0%
Kd_soil	23,9%
Runoff	19,5%
Forest1.Multiple1	9,8%
Forest5.Multiple5	6,8%
Forest4.Multiple4	4,7%
Forest3.Multiple3	3,0%
Forest2.Multiple2	2,6%
h	1,3%
density_soil	1,1%
Sum for other parameters	0,3%

4.5.4.2 Uncertainty analysis

The results of the uncertainty analyses for the two considered endpoints are presented in Table 25 for CI-36 and in Table 26 for I-129. The statistics are presented for two cases. In one case all uncertain parameters were varied in the simulations and in the other only important parameters, identified from the screening study, were varied and for others the same value was used in all iterations. It is clear from these tables that the results in both cases are practically identical. The standard deviation in both cases was practically the same, which indicates that the parameters identified as potentially important from the screening study fully explain the variation observed from the probabilistic simulations.

Table 25 Results of the uncertainty analysis for the Total and Average doses from CI-36. The results are presented for two cases, one in which all parameters are varied in the simulations and other in which only important parameters are varied.

Statistics	Average Dose, Sv/y		Total Dose, Sv/y	
	All parameters	Important parameters	All parameters	Important parameters
Mean	2,0E-10	2,1E-10	6,7E-07	7,0E-07
Std dev	3,1E-10	3,2E-10	1,1E-06	1,1E-06
Median	1,0E-10	1,1E-10	3,2E-07	3,4E-07
5.0%	2,1E-11	2,1E-11	3,9E-08	3,8E-08
95.0%	7,2E-10	7,3E-10	2,5E-06	2,5E-06

Table 26 Results of the uncertainty analysis for the Total and Average doses from I-129. The results are presented for two cases, one in which all parameters are varied in the simulations and other in which only important parameters are varied.

Statistics	Average Dose, Sv/y		Total Dose, Sv/y	
	All parameters	Important parameters	All parameters	Important parameters
Mean	1,0E-09	1,0E-09	2,7E-06	2,7E-06
Std dev	1,2E-09	1,2E-09	4,3E-06	4,1E-06
Median	6,3E-10	6,4E-10	1,3E-06	1,3E-06
5.0%	3,0E-10	3,0E-10	1,5E-07	1,4E-07
95.0%	2,9E-09	2,9E-09	9,5E-06	9,5E-06

4.5.4.3 Correlation and regression based statistics

Table 27 and Table 28 present correlation and regression based statistics obtained for the total doses from CI-36 and I-129, respectively. The results for the average doses were very similar. All considered correlation and regression statistic show results that are consistent with each other and with the results obtained from the screening study. However, the coefficient of determination (R^2) around 0,5-0,6 which indicates to the presence of non-linearity in the dependencies and/or interactions between the parameters.

Table 27 Correlation and regression statistics for the total doses from CI-36

Parameter	Pearson	Spearman	SRC	SRRC	PCC	PRCC
LP	0,004	-0,007	0,003	0,002	0,005	0,007
CR_L[CI-36]	0,004	-0,014	0,007	0,001	0,011	0,003
density_soil	-0,003	0,002	-0,004	0,004	-0,006	0,016
h	-0,001	0,009	-0,008	-0,002	-0,012	-0,007
WP	-0,018	-0,017	-0,021	-0,014	-0,033	-0,055
CR_W[CI-36]	-0,021	-0,015	-0,015	-0,016	-0,024	-0,063
TC_WtoLi	0,018	0,004	0,016	0,007	0,026	0,029
TC_UtoLi	-0,004	-0,005	-0,001	0,000	-0,001	-0,001
TC_LitoS	-0,014	-0,015	0,000	0,000	0,000	-0,001
TC_LtoLi	0,013	0,013	0,005	0,002	0,009	0,010
UP	-0,002	0,001	0,000	0,002	0,000	0,008
CR_U[CI-36]	-0,002	-0,005	-0,006	-0,002	-0,009	-0,006
Runoff	-0,294	-0,448	-0,291	-0,450	-0,421	-0,868
Theta	0,001	-0,004	0,006	0,003	0,009	0,010
Kd_soil[CI-36]	0,343	0,531	0,341	0,533	0,477	0,901
Kd_lake[CI-36]	0,001	0,003	0,006	0,000	0,009	-0,001
Kd_riversediment[CI-36]	-0,003	-0,006	0,005	-0,002	0,008	-0,007
Kd_river[CI-36]	0,004	0,004	0,003	0,000	0,005	0,001
Kd_sea[CI-36]	-0,018	-0,013	-0,002	-0,007	-0,003	-0,026
Kd_agriland[CI-36]	-0,005	-0,004	-0,001	0,001	-0,002	0,005
CRforest[CI-36]	0,611	0,633	0,613	0,642	0,698	0,928
Forest1.Multiple1	-0,131	-0,162	-0,129	-0,164	-0,200	-0,537
Forest2.Multiple2	-0,043	-0,042	-0,044	-0,048	-0,070	-0,183
Forest3.Multiple3	-0,042	-0,054	-0,041	-0,051	-0,065	-0,194
Forest4.Multiple4	-0,027	-0,049	-0,039	-0,054	-0,062	-0,206
Forest5.Multiple5	-0,080	-0,085	-0,080	-0,084	-0,126	-0,309

Table 28 Correlation and regression statistics for the total doses from I-129

Parameter	Pearson	Spearman	SRC	SRRC	PCC	PRCC
LP	0,012	0,012	0,005	0,004	0,009	0,014
CR_L[CI-36]	-0,016	-0,003	-0,013	0,003	-0,020	0,011
density_soil	-0,018	0,000	-0,020	-0,009	-0,031	-0,036
H	-0,029	-0,013	-0,019	-0,009	-0,030	-0,034
WP	-0,010	-0,016	-0,003	0,001	-0,005	0,004
CR_W[CI-36]	-0,003	-0,002	-0,010	-0,010	-0,015	-0,038
TC_WtoLi	0,000	-0,008	-0,003	0,001	-0,005	0,004
TC_UtoLi	0,001	0,006	0,001	-0,002	0,002	-0,007
TC_LitoS	-0,006	-0,009	0,004	0,001	0,007	0,003
TC_LtoLi	0,002	0,004	-0,006	-0,003	-0,009	-0,012
UP	-0,003	-0,017	0,002	-0,004	0,002	-0,014
CR_U[CI-36]	0,000	0,000	0,002	-0,004	0,003	-0,017
Runoff	-0,298	-0,445	-0,291	-0,436	-0,421	-0,859
Theta	0,002	0,004	-0,003	0,004	-0,005	0,014
Kd_soil[CI-36]	0,342	0,535	0,345	0,532	0,482	0,899
Kd_lake[CI-36]	0,000	0,002	-0,004	0,001	-0,006	0,003
Kd_riversediment[CI-36]	0,008	0,004	0,002	-0,004	0,003	-0,014
Kd_river[CI-36]	-0,008	-0,009	-0,001	0,004	-0,002	0,017
Kd_sea[CI-36]	0,006	0,000	0,005	0,000	0,009	0,000
Kd_agriland[CI-36]	-0,011	0,011	-0,012	0,000	-0,020	0,002
CRforest[CI-36]	0,615	0,644	0,619	0,642	0,702	0,927
Forest1.Multiple1	-0,111	-0,157	-0,123	-0,164	-0,192	-0,534
Forest2.Multiple2	-0,010	-0,031	-0,026	-0,046	-0,042	-0,173
Forest3.Multiple3	-0,034	-0,038	-0,040	-0,048	-0,063	-0,182
Forest4.Multiple4	-0,026	-0,047	-0,032	-0,059	-0,051	-0,221
Forest5.Multiple5	-0,067	-0,094	-0,070	-0,092	-0,110	-0,335

4.5.4.4 Variance based methods

The results obtained with the EFAST method for the total doses from CI-36 are presented in Table 29. The results for the average doses were very similar. Around 61 % percent of the variance can be explained by first order interactions and 53 % are attributable to three parameters. Around 39 % of the variance is due to higher order interactions between the pa-

rameters, which confirm the findings from the screening study and the correlation/regression analyses.

Table 29 Sensitivity indexes obtained with the EFAST method for the total doses from CI-36

Parameter	EFAST	
	First Order SI	Total Order SI
CRforest	0,30	0,63
Runoff	0,12	0,43
Kd_soil	0,11	0,29
Forest1.Multiple1	0,04	0,16
Forest2.Multiple2	0,01	0,09
Forest5.Multiple5	0,01	0,05
Forest4.Multiple4	0,00	0,05
Forest3.Multiple3	0,00	0,02
CR_W	0,00	0,02
unexplained	0,39	

4.5.5 Summary

A graded approach to the sensitivity analysis was applied to a biosphere model with a large number of uncertain parameters. This graded approach started with the application of the Morris screening method with the aim of identifying potentially important parameters. The screening was followed by uncertainty analyses, which were made for two cases: one with all uncertain parameters being varied probabilistically and other with only important parameters, identified in the screening study, being varied between iterations. The results for the two cases were practically the same, confirming the conclusion about important radionuclides made from the screening study. From the Monte Carlo simulations correlation and regression statistics were obtained which also pointed to the same important parameters and indicated that interactions between parameters can have an important contribution to the observed variation of the simulation results. This was corroborated by the analyses with the EFAST method, which showed that around 39 % of the variance is related with higher order interactions between the parameters.

4.6 JRC application case: French clay site

4.6.1 Conceptual model

In this study we analyse the release of radionuclides from wastes of an Intermediate Level Waste (ILW) disposal cell embedded in a porous geological formation for a generic French

clay site. The disposal cell, which only contains non-organic waste forms, does not release hydrogen; it includes both activated waste and compacted hulls and end-caps that are supposed to be homogeneously filled.

Figure 80 is a simplified representation of the repository. A 2D vertical slice along the symmetry axis of the disposal cell is shown in the upper left corner. The complete geometrical model includes the thickness (130 meters) of the undisturbed host rock (Callovo-Oxfordien layer) and the disposal cell located in the middle of the clay layer. The waste packages domain is considered as a rectangle located in the centre of the circular disposal cell. The materials are also indicated in Figure 80.

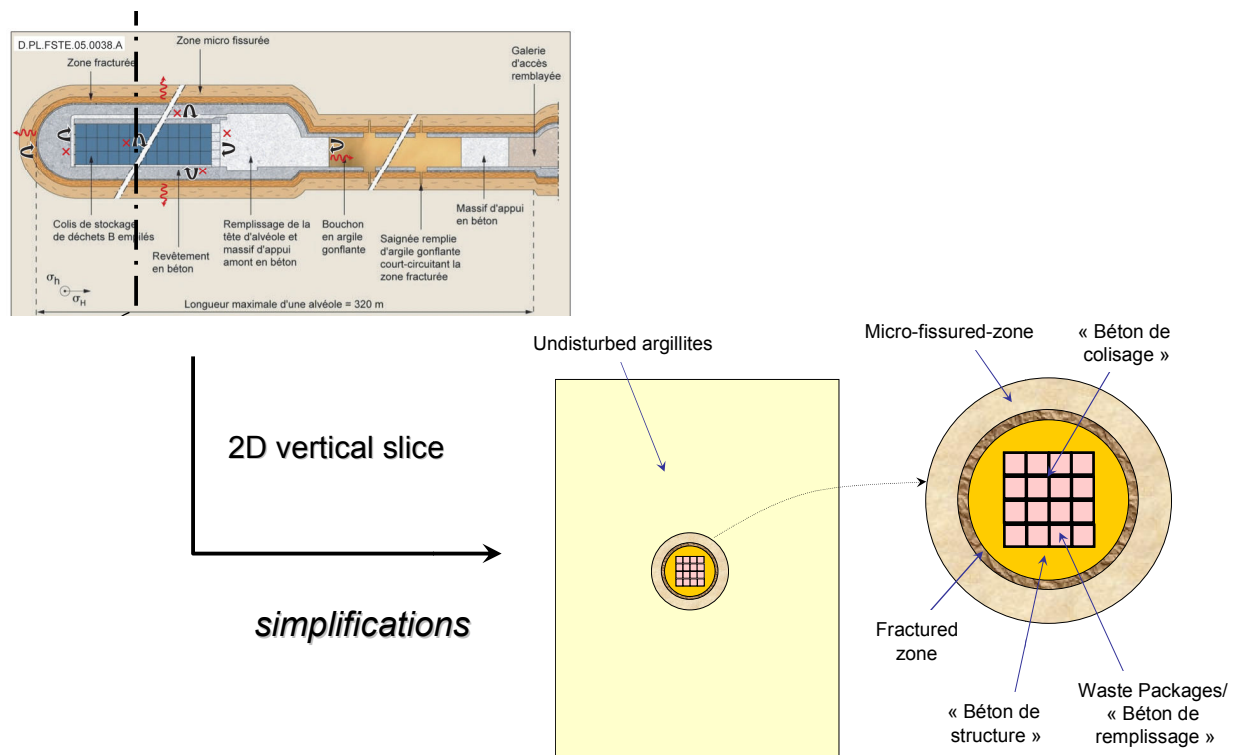


Figure 80 Repository geometry and components (i.e. materials).

The main assumptions to be taken into account in calculations are (see further details in [Pepin et al., 2007]):

- The disposal cell and its environment are considered as a homogeneous continuous porous medium, and it is supposed to be saturated of water from $t=0$; hence release of radionuclides from waste packages starts from $t=0$. Calculations are carried out until one million years.
- The migration of radionuclides in a porous medium in aqueous solution is considered as convective/diffusive/dispersive, taking also into account sorption (linear and reversible, K_d approach) and precipitation (solubility limit C_{sat}). The convective part is based on steady-state hydraulic results using a constant vertical head gradient through the host rock.

- The waste package is represented by two materials: (i) a specific concrete filling (“béton de remplissage”), in which the source term is distributed in an uniform way, there is no sorption and solubility is unlimited from $t=0$ and; (ii) an over-pack in a specific concrete (“béton de colisage”) with a high level of confining performance (low diffusivity, high K_d and low solubility limit) up to 10.000 years, and then degrading itself into filling concrete (“béton de structure”). Thus the material properties in the analysis will change at this time. The details for this are given in [Pepin et al., 2007].
- The ILW disposal cell is filled with a concrete material in which radionuclides are free to migrate with no sorption and with unlimited solubility, and thus, (like “béton de remplissage”, same deterministic input data), no uncertainty has to be considered.
- A sensible approach to the concrete geochemical characterization would be a variation in space and time of K_d and C_{sat} (solubility limit) taking into account the evolution across different states of concrete (sound, altered, degraded, neutralised conditions). Since the dynamics of propagation of the various alterations of concrete are not known, we consider the same geochemical conditions for the whole thickness of concrete (all the disposal cell) and its range of variation (uncertainty) takes into account the various alterations. This is applied by ranges of time.
- The mechanical disturbance is represented by the Excavated Damaged Zone (EDZ), which consists of a fractured zone with high permeability and of a micro-fissured zone. The geometrical boundaries of the micro-fissured zone do not change over time. The occurrence of a self-sealing phenomenon is integrated in the uncertainty.
- The chemical disturbance generated by the degradation of the concrete is included in the mechanical disturbance (EDZ) in terms of extension and in migration parameters (hydraulic, sorption and precipitation). The geochemical conditions considered in the EDZ are the same as in the undisturbed argillites.

4.6.2 Input uncertain parameters and outputs analysed

The pdfs used in this Benchmark are based on a large data set from an intensive characterization programme developed by Andra in combination with expert judgment and literature review. The pdfs include three sources of uncertainty: natural variability (especially for argillites), experimental measurement methods (laboratory + in situ tests) and up-scaling (extrapolation from small samples to repository scale). All pdfs are truncated. Three radionuclides are considered in this study: ^{129}I , ^{94}Nb and ^{79}Se . The whole set of input parameters pdfs are given in [Pepin et al., 2007], and their names and coded names are given in the following table.

Table 30 Names of input parameters

Name	Coded name
Radionuclide release rate in stainless steel (B1)	rB1ac
Radionuclide release rate in inconel (B1)	rB1in
Radionuclide release rate in coques (B5)	rB5co
Radionuclide release rate in stainless steel (B5)	rB5ac
Radionuclide release rate in inconel (B5)	rB5in
Effective diffusion coefficient in béton de colisage	DeWp
I distribution coefficient Kd(I) in béton de structure	KdIDC
Nb distribution coefficient Kd (Nb) in béton de structure	KdNbDC
Se distribution coefficient Kd (Se) in béton de structure	KdSeDC
Nb solubility limit in béton de structure before $2 \cdot 10^5$ y	Cs1NbDC
Nb solubility limit in béton de structure after $2 \cdot 10^5$ y	Cs2NbDC
Se solubility limit in béton de structure before $2 \cdot 10^5$ y	Cs1SeDC
Se solubility limit in béton de structure after $2 \cdot 10^5$ y	Cs2SeDC
Effective diffusion coefficient in the fractured zone	DeFZ
Diffusive porosity in the fractured zone	PorFZ
Effective diffusion coefficient in the micro fissured zone	DeMZ
Diffusive porosity in the micro fissured zone	PorMZ
Effective diffusion coefficient in the undisturbed argilites	DeUA
Diffusive porosity in the undisturbed argilites	PorUA
I distribution coefficient Kd(I) in the undisturbed argilites	KdIDC
Nb distribution coefficient Kd(Nb) in the undisturbed argilites	KdNbDC
Se distribution coefficient Kd(Se) in the undisturbed argilites	KdSeDC
Niobium solubility limit in the undisturbed argilites	CsNbUA
Selenium solubility limit in the undisturbed argilites	CsSeUA

Although the whole correlation structure was estimated for 8 input parameters and the constraints for the effective diffusion coefficients and porosities were set for 3 layers, only 5 correlation coefficients were effectively implemented in the sampling process, and no constraint was finally modelled; see [Pepin et al., 2007] for more details.

The output variables analysed were the molar flows (and their respective peaks and time to peaks) of the radionuclides considered getting out of the following layers

- the waste packages (WP, external envelope surface of waste packages pile),
- the disposal cell (DC, external envelope surface of « béton de structure »),

- the fractured zone (FZ, external envelope surface of fractured zone material),
- the micro-fissured zone (MZ, external envelope surface of micro-fissured zone material),
- the undisturbed argillites (UA15m), in the clay at a specific surface located at 15 meters from packages, and
- the undisturbed argillites (UA), at the top and bottom.

4.6.3 Study objectives

Although among the objectives of the original study was a comparison between the results obtained by means of a detailed 2-D code (Cast3M and Porflow, within the simulation platform Alliances, Andra) and a simplified 1-D code (Goldsim, JRC) in a probabilistic framework, only sensitivity analysis related results obtained by JRC are reported in section 4.7 of this document. In fact, the target in this section is to identify most relevant input parameters via regression analysis and draw conclusions about the use of this SA technique. Regressions used were based on the raw values and also on the ranks. SA indicators used were SRCs and SRRCs. A simple random sample of size 1000 was used in this study.

4.6.4 Main results

Eighteen dynamic output variables and 36 non-dynamic output variables are studied in this work. The dynamic variables are the flows getting out of the six layers over time for the three radionuclides considered. The thirty six non-dynamic variables are the peak flows for each layer and each radionuclide and the corresponding times to the peak flows. Twenty four input uncertain parameters are considered in the analysis (all the uncertain input parameters described in [Pepin et al., 2007], excluding permeabilities). All the results of the sensitivity analysis are collected in section 7.2.2 of [Práková et al., 2008].

Regression analyses based on the raw values and on the ranks have been performed for all dynamic and non-dynamic output variables. The first generic result shows that the quality of the regressions, measured in terms of their coefficients of determination (R^2), is much better when the ranks of the values are used instead of the raw values in most of the cases. This is the reason to base all our analyses on the study of the SRRCs rather than the SRCs. This means that, in most of the cases, it is convenient to analysis the dependence between outputs and inputs from the point of view of monotonic relations instead of from the point of view of linear relations. Results for dynamic outputs are always reported for the period 10 to 10^6 y.

Figure 81 shows the results of the regressions based on the ranks of the values for the molar flow of ^{129}I getting out of the six layers. It can be noted that parameters associated with the other radionuclides, such as $K_{d_{Se}UA}$ show up in ^{129}I plots. Certainly there is no physical reason for such dependence; this is just an artefact from including all input parameters in the regression analysis. The effect of the degradation of the concrete properties at 10^4 y can be seen in Figure 81 a) to d). A discontinuity in the curves associated to R^2 and to the SRRCs

associated to the most relevant input parameters may be seen in the first three figures, while in the fourth one a slope change does also happen at that time. No specific effect can be seen in the results for UA15m and for UA. The second general comment about the results for ^{129}I over time is about the existence of a minimum in the value of R^2 . In the six plots, at a given time, R^2 reaches a minimum. This happens at approx. 3000 y in WP, at approx. 15000 y in DC and so on. Only in UA this minimum does not happen. It is obvious the shift to later times of this minimum as we look at outer layers. This minimum is always located at a time when SRRCs associated to the most relevant input parameters switch their sign.

Figure 81 does also show that, regarding ^{129}I , the most important parameters are the distribution coefficients of this radionuclide in the béton de structure ($K_d\text{DC}$) and in the undisturbed argilites ($K_d\text{UA}$), and the effective diffusion coefficient in the béton de colisage (DeWP) and in the undisturbed argilites (DeUA), though their importance change over time and from one to another layer.

The ^{129}I flow getting out of the waste packages (Figure 81a) is most affected at early times (before 10^4 y) by $K_d\text{DC}$ and DeWP . After that time $K_d\text{UA}$ becomes the most important parameter, followed by $K_d\text{DC}$ and DeUA . It is also remarkable the change of the sign of the SRRCs for these input parameters over time. A similar behaviour can be seen in Figure 81b) and Figure 81c) regarding the second and the third layers.

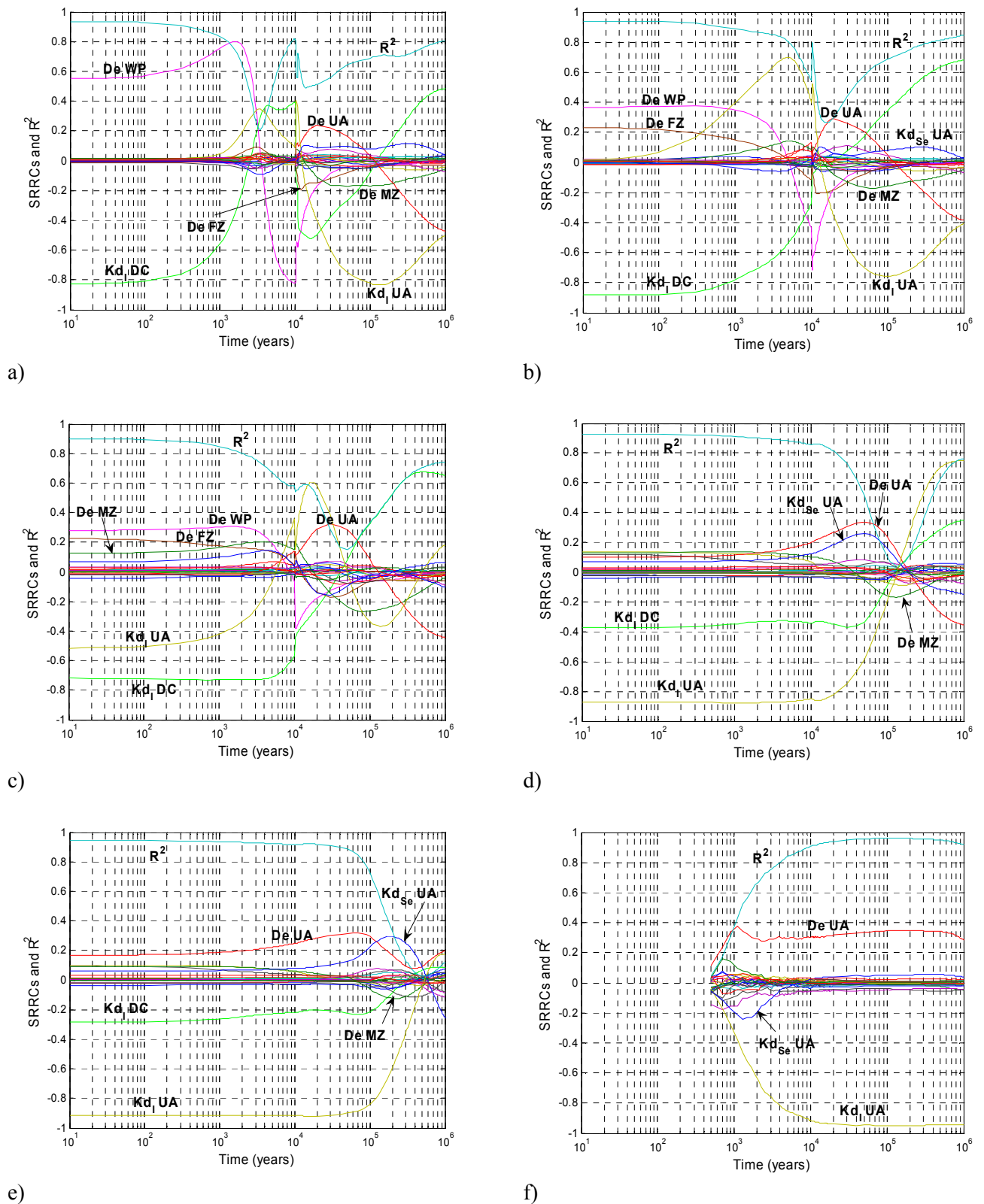


Figure 81 Evolution of SA indicators for ^{129}I molar flows getting out of a) waste package, b) disposal cell, c) fractured zone, d) micro-fissured zone, e) undisturbed argillites (15 m), f) undisturbed argillites .

Regarding the ^{129}I flow getting out of the micro fissured zone (Figure 81d), $K_d\text{UA}$ becomes the most important input parameter at all times, followed by $K_d\text{DC}$ and DeUA . A change in the sign of the three SRRCs happens at approx. $1.5 \cdot 10^5$ y. In the next layer, $K_d\text{UA}$ becomes even more predominant. In the last layer, $K_d\text{UA}$ becomes almost the only important parameter, followed by DeUA . All flows getting out of this layer were null before 500 y, that is why SRRCs are not represented for earlier times (Figure 81f).

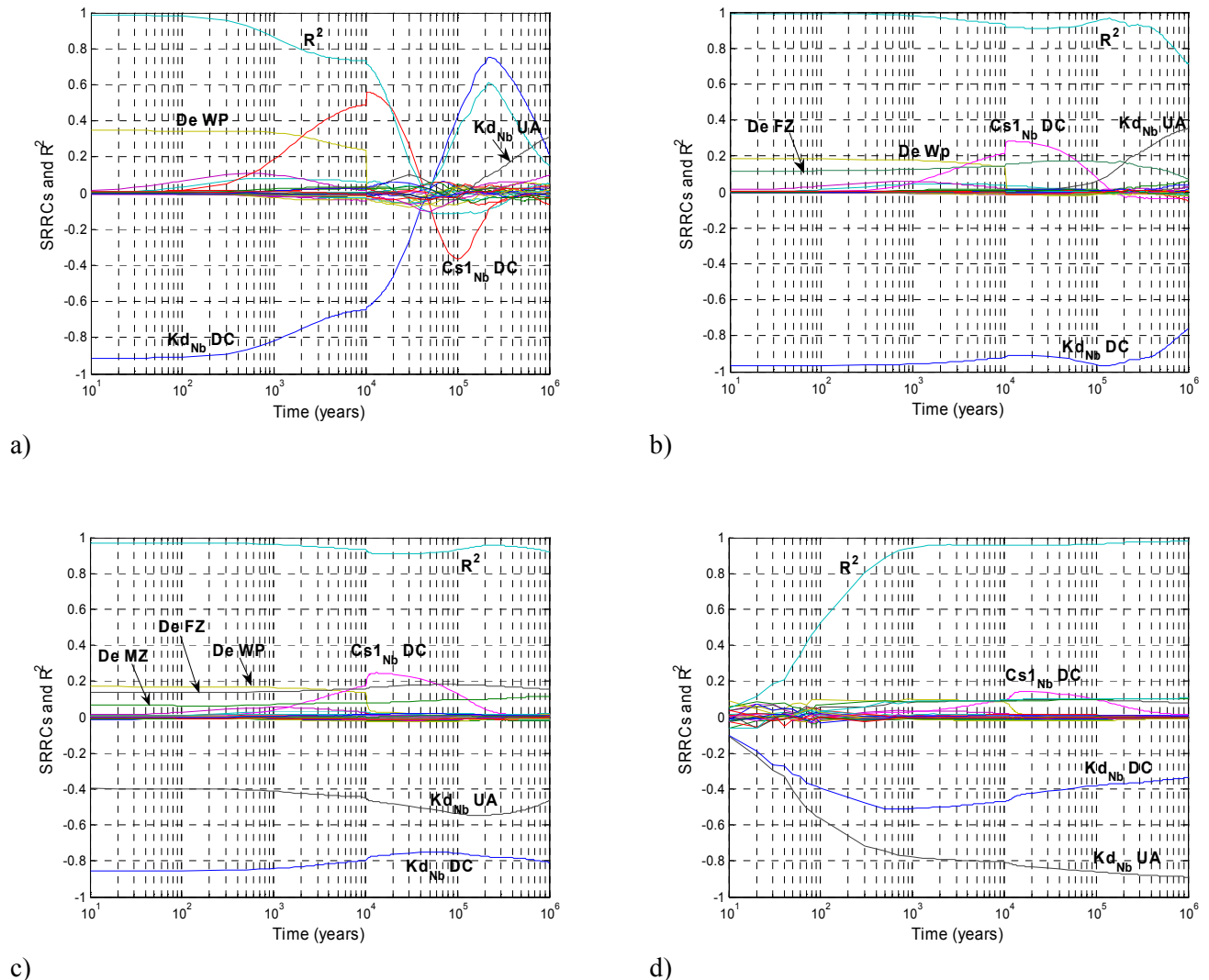


Figure 82 Evolution of SA indicators for ^{94}Nb molar flows getting out of a) waste package, b) disposal cell, c) fractured zone, d) micro-fissured zone.

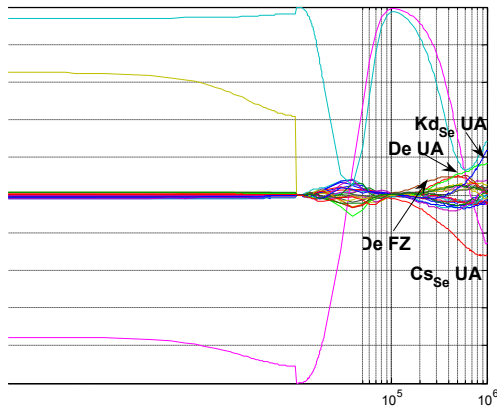
As in the case of ^{129}I , the effect of the degradation of the concrete properties at 10^4 y on the ^{94}Nb flows can be seen in Figure 82a) to Figure 82c). The effect does not affect so much ^{94}Nb flows getting out of the micro fissured zone. It is also remarkable the very high quality of the regressions obtained for the second, third and fourth layers (R^2 close to 1) for most of the times. This is mostly due to the fact that SRRCs associated to most important input parameters do not change their signs over time and they have very high absolute values. In the first

layer, the two most important input parameters change their SRRCs' signs at approx. $5 \cdot 10^4$ y. Results for ^{94}Nb in UA15m and in UA are not reported here because in the former case results are very similar to those obtained in MZ, while in the latter case all flows are null and no SA results could be obtained.

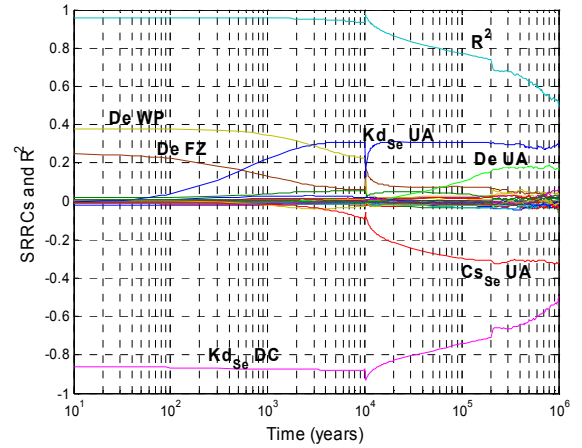
The most important parameters regarding the flow of ^{94}Nb are the distribution coefficients of this radionuclide in the béton de structure (K_{NbDC}) and in the undisturbed argilites (K_{NbUA}), the effective diffusion coefficient in the béton de colisage (DeWP) and the Nb solubility limit ($\text{Cs}_{1\text{NbDC}}$), though DeWP is moderately relevant only in the first layer at early times. $\text{Cs}_{1\text{NbDC}}$ is moderately important at intermediate times in the first layer and its importance is dramatically reduced as we look at outer layers. K_{NbDC} is the most important input parameter in the first three layers, and it loses its primacy in favour of K_{NbUA} only in the last layers.

As in the case of ^{129}I and ^{94}Nb , the effect of the degradation of the concrete properties at 10^4 y on the ^{79}Se flows can be seen in Figure 83a to Figure 83e. The effect does not affect ^{79}Se flows getting out of the sixth layer (external boundary of the undisturbed argilites). It is also remarkable the very high quality of the regressions obtained for the first five layers (R^2 close to 1) at early times (before 10^4 y). There is a general decrease in the quality of the regressions after that time in all layers except in the sixth one and, partially, in the first one. This is mostly due to the fact that SRRCs associated to most important input parameters decrease at late times. Only for the flow getting out of WP and of UA after 10^4 y R^2 remains high.

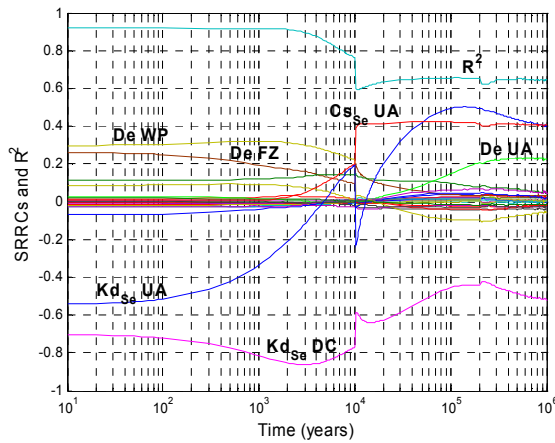
Figure 83 shows that, regarding ^{79}Se , the most important parameters are the distribution coefficients of this radionuclide in the béton de structure (K_{SeDC}) and in the undisturbed argilites (K_{SeUA}), the effective diffusion coefficient in the béton de colisage (DeWP) and in the undisturbed argilites (DeUA), and the solubility limit of Se in the undisturbed argilites (Cs_{SeUA}). K_{SeDC} and K_{SeUA} are the two most important input parameters affecting ^{79}Se results. K_{SeDC} is the most important one in the first three layers while it switches to the second position in the three outer layers, in favour of K_{SeUA} . DeWP is moderately important only for the flow of ^{79}Se getting out of the first three layers until 10^4 y. It is completely irrelevant after that time and has no impact on the results for the flow getting out of the three outer layers. K_{SeDC} and K_{SeUA} . Cs_{SeUA} becomes moderately important in all layers only after 10^4 y.



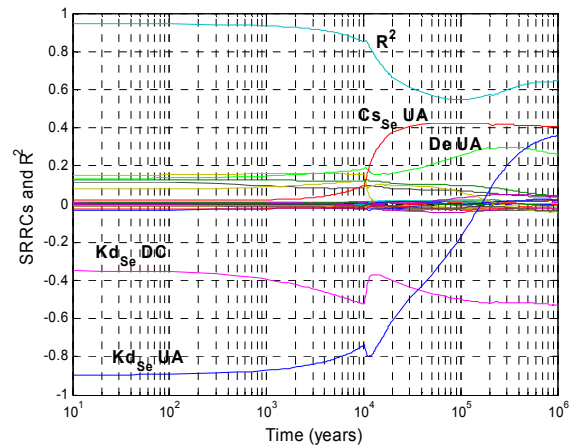
a)



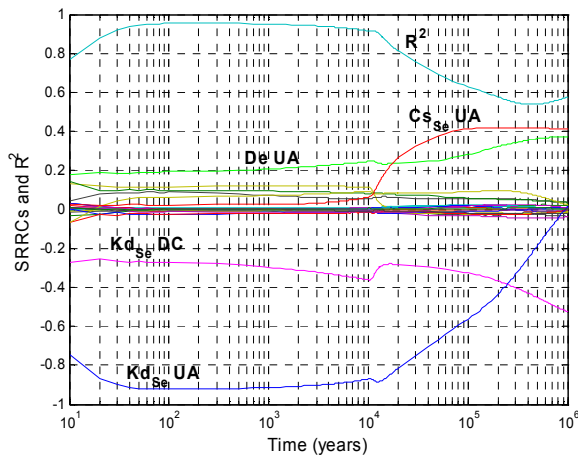
b)



c)



d)



e)

f)

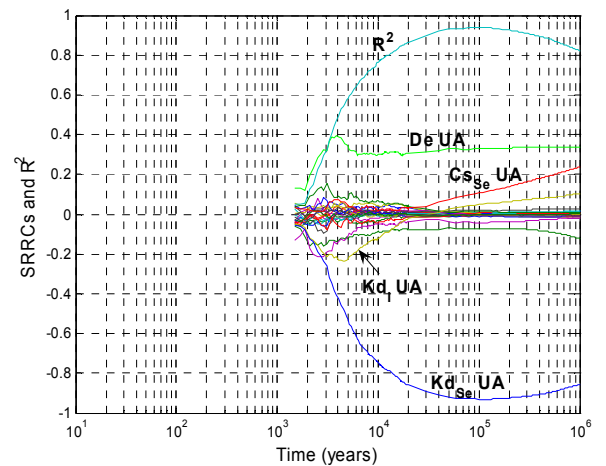


Figure 83 Evolution of SA indicators for ^{79}Se molar flows getting out of a) waste package, b) disposal cell, c) fractured zone, d) micro-fissured zone, e) undisturbed argillites (15 m), f) undisturbed argillites.

Regarding ^{129}I peak flow, $K_d\text{DC}$ is the most important parameter in the inner layers, while $K_d\text{UA}$ followed by DeUA are the most important parameters in the outer layers. DeWP is the most important parameter in WP. Regarding the time to the peak, $K_d\text{DC}$ and DeWP are the most important parameters in the internal layers while $K_d\text{UA}$ is the only parameter that really matters in the outer layers.

Regarding ^{94}Nb peak flows, only the K_d s have a relevant impact, $K_{d\text{Nb}}\text{DC}$ in the inner layers and $K_{d\text{Nb}}\text{UA}$ in the outer layers. The parameter that affects most the time to the peak flow of ^{94}Nb in the four intermediate layers is $\text{Cs}_{1\text{Nb}}\text{DC}$, followed by $K_{d\text{Nb}}\text{DC}$. For the first layer, results are neither reliable nor significant (too small R^2) and in the last layer flows were null at all times.

The parameter that affects most ^{79}Se peak flow in the first five layers is $K_{d\text{Se}}\text{DC}$. In intermediate layers, this parameter is followed in order of importance by $\text{Cs}_{\text{Se}}\text{UA}$, DeUA and $\text{Ks}_{\text{Se}}\text{UA}$. $\text{Ks}_{\text{Se}}\text{UA}$ and DeUA are the most important parameters in the last layer. The time to the peak flow is controlled by $\text{Cs}_{\text{Se}}\text{UA}$ in WP, by $K_{d\text{Se}}\text{DC}$ in the next two layers and by $K_{d\text{Se}}\text{UA}$ in the three outer layers.

4.6.5 Conclusions of the JRC application case (French clay site)

After this study, the main conclusions about the use of regression based SA indicators (R^2 , SRCs and SRRCs) are

Very poor quality results (small R^2) were obtained when working on the raw values. The quality dramatically increased when regressions were based on the ranks (of both, inputs and outputs). This scale (ranks), which is adequate to analyse monotony relations, seems much more appropriate to analyse the data of this study than the linear scale.

Regressions based on the logarithmic transformation were not used due to the existence of null values in many output variables.

Relevant drops in the values of R^2 were observed in many time dependent output variables. These drops were related in most of the cases to sign changes (from positive to negative or vice versa) in the sensitivity indicators (SRRCs) associated to the most important parameters. Regression based SA techniques are adequate to analyse the system behaviour at those times even though R^2 values are small. Lack of monotony must be acknowledged in those cases and input parameters whose sensitivity indices change sign remain the most important ones.

The inclusion of input parameters that definitely could not affect the results but whose associated SRRCs were not insignificant are considered computational artefacts that could not be explained. The advice in such cases is to exclude those input parameters from the analysis.

4.7 Lessons learnt

The application of different probabilistic sensitivity analysis methods to realistic repository models showed some interesting properties of such systems and specific problems that cannot be investigated using simple analytical models. The investigations of the individual project participants were focused at analysing repository systems of different types in different host rocks, each with its own peculiarities. These investigations revealed specific problems that can arise in the sensitivity analysis of such systems.

The Dutch salt system was analysed by varying a number of input parameters that determine the closure behaviour of the compacted salt plug and thus the release of contaminants from the near field. At some point in time, the plug reaches a status of practical tightness, and the near field calculation is terminated. Therefore, if the plug permeability is analysed as a time-dependent model output, more and more values are missing with increasing model time. Some specific strategy is necessary to handle this problem. Simply ignoring the missing values leads to a decrease of the sample size available for evaluation and to less reliable results. It was found that the best strategy is to replace the missing values by the last available value of each simulation.

For the Dutch clay system it was found that the model output, as well as the input parameters, varies over several orders of magnitude so that an evaluation on a linear scale is highly influenced by only a few simulations. Therefore, a logarithmic transformation was applied both to input and output parameters, which lead to a better performance of the analysis. A similar but even harder problem arose in the case of the German salt system. About 85 % of the possible parameter combinations lead to exact zero output, because the near field is closed by convergence before the waste gets in contact with brine. In some of the other simulations, however, the output becomes relatively high. A logarithmic transformation does not work in this case, as it is not applicable to zero values. Therefore, a specific transformation was applied that seemed more adequate for the situation. The performance of the sensitivity analysis was considerably increased by this measure. Transformation of model input and output seems to be a topic worth of being further investigated.

In the work done for the Spanish granite system it was found very useful to include some "derived parameters" in the sensitivity analysis. These parameters are combinations of the random input parameters of the model that are suspected to have a strong effect on the system behaviour (on the basis of the experts experience), such as the retarded travel time of a radionuclide in the geosphere. It was found that the influence of the "derived parameters" on the models results was most relevant. Derived parameters should be identified at the beginning of the sensitivity analysis and treated in the same way as the random input variables through the analysis, keeping in mind that they are correlated with some input parameters.

For the Spanish granite system the different statistics used in the sensitivity analysis identify the same set of relevant parameters with a very similar ranking of importance. "Derived pa-

rameters" were at the top of the ranking of importance of the different parameters for all the statistics considered.

The biosphere system investigated by Facilia depends on a high number of parameters 26 of which were considered uncertain and selected for the analysis. This is still a high number of input parameters for a sensitivity analysis, at least a variance-based one. It could be demonstrated how different kinds of sensitivity analysis methods, including Morris screening, perform on such a system.

As a general finding, the correlation- and regression-based sensitivity analysis methods performed well on most of the investigated systems. As the theory predicts, in most cases the rank-based evaluation performed better than the value-based evaluation, since the systems show a monotonic but non-linear behaviour. The coefficient of determination (R^2) becomes usually higher and the results provide a clearer parameter ranking if a rank transformation is performed on the input and output parameters.

Non-parametrical statistics (Smirnov test, Mann-Whitney test and t test) were applied to some of the systems. The parameter rankings obtained by these tests sometimes differ from those calculated with regression or correlation methods, which is due to the principally different mathematical ideas they are based on, emphasising different model peculiarities. It has been found that is more useful to present the results of the Monte Carlo filtering in terms of the statistics instead of the p-values, because the statistics tend to a stable value when the number of runs increases while the p-values tend to zero. The t statistic produces in some cases anomalous results because the hypothesis of applicability is not fulfilled. Monte Carlo filtering can be done using only Smirnov and Mann-Whitney tests for values and ranks, respectively.

Variance-based sensitivity analysis (EFAST) was applied to two of the systems. It performed not as well as initially expected, at least in the case of the German salt system. This is due to the wide range of possible output values including a high number of exact zeros. When applied to a transformed output EFAST performed better on that system.

Additionally to the mentioned numerical methods of sensitivity analysis some specific graphical methods have been applied to some of the model systems. These are scatter plots, cob-web plots and Contribution to the Sample Mean (CSM) plots. It has been found that they provide a good means for analysing and presenting the sensitivity of the model to the individual input parameters by a quick view. The CSM statistic allows selecting the parameters whose contribution to the mean is controlled by a small region of values, and the CSM plots are used to identify exactly such region of values. The information provided by CSM plots can be useful to guide R&D efforts to reduce the uncertainty in that particular region of values.

Different sampling methods (Simple Random Sampling, SRS and Latin Hypercube Sampling, LHS) were investigated in the Spanish investigations, but no significant influence of the sam-

pling strategy was identified. SRS is preferred for practical reasons: several samples can be combined to obtain a sample of greater size. Moreover, the Spanish model is fast enough to allow an analysis of the influence of the sample size. Sample sizes up to 25 000 were investigated and it was found that the results of correlation- and regression-based methods became sufficiently robust with samples of about 5000.

Random input parameters that are known for sure that can not have any effect on the output variable (selenium solubility can have no effect on the doses due to I129, for instance) can be used to derive empirical values for the minimum value of a statistic that is statistically significant (empirical thresholds).). When theoretical thresholds can be calculated there is a good agreement between theoretical and empirical thresholds.

5. Summary

Probabilistic sensitivity analysis is a powerful means for analysing the sensitivity of calculation models to uncertain input parameters, which is an important task in the field of final repository performance assessment. The work described in this report was aimed at investigating different mathematical methods of probabilistic sensitivity analysis in this context.

For this purpose, three steps of work were performed. The first step was a comprehensive review of sensitivity analysis methods that are available for application. In the second step a number of these methods, especially those, which had scarcely been applied to final repository systems so far, were tested with simple analytical models in order to analyse their behaviour under known circumstances. The third step was dedicated to applying different sensitivity analysis methods to realistic repository models, finding out how they work in practice and how they react to repository-specific model properties.

In the first step, a number of methods of different types were identified:

- **Screening methods.** These methods are designed to identify non-relevant input parameters at low computational costs, not taking into account the realistic distribution functions of the uncertain parameters. The most important method of this kind is the Morris screening method.
- **Graphical methods.** Such methods aim at evaluating a set of calculation runs with statistically distributed input parameters by means of a descriptive graphical presentation. Such techniques allow a quick qualitative assessment of the sensitivity, which is often more helpful than a calculated numerical result. Useful graphical presentations are scatter plots, cobweb plots and CSM plots.
- **Monte-Carlo-based methods.** These sensitivity analysis methods use a Monte-Carlo sample of input parameter values, drawn under consideration of the applicable probability density functions (pdfs). The model results are evaluated using specific mathematical procedures, based, for example, on calculation of correlation or regression coefficients. These techniques work best on models with a close-to-linear behaviour, which can often be better approximated by performing a rank transformation on the model input and output. Another kind of Monte-Carlo-based sensitivity analysis is called Monte-Carlo Filtering. It comprises the Smirnov and Mann-Whitney two-sample-tests, are used for identifying statistically significant relations between different regions of input parameters and output variables.
- **Variance-based methods.** These sensitivity analysis methods also use a sample of input parameters according to their pdfs, but, depending on the specific method, the sampling has to follow an appropriate scheme. The general idea of these methods is to decompose the total variance of the model output in contributions that are due to the individual input parameters and to their interactions of any order. All such methods calculate the same sensitivity measures, but using different mathematical techniques. The

most important methods are the Sobol' method and the Fourier Amplitude Sensitivity Test (FAST) with its extension EFAST.

Since variance-based sensitivity analysis has scarcely been applied in the context of final repository safety analysis, it was considered a specific task within the work described here, to do experiments and practical tests with this kind of techniques.

The goal of the second step of work was to analyse and compare different methods of sensitivity analysis by applying them to simple analytical models with a known behaviour, which even allow an exact theoretical determination of the sensitivity measures under consideration. As a more realistic standardised example the Level E test case, which describes the radionuclide release from a simple hypothetical repository, was included. The work was done in form of a benchmark exercise, focused on variance-based methods.

A lot of insight into the internals of variance-based sensitivity analysis has been gained during the course of the benchmark exercise: First of all, it was noticed that for the standard algorithms the different implementations seem to be very stable and produce results with only subtle differences. In some situations, however, the results depend quite substantially on the implementation and/or the choices of the user.

The interest in methods that do not require a particular sampling scheme (cheap methods) has arisen as of late in the benchmark exercise so that it was noticed with satisfaction that the results obtained with them are comparable to specialised methods. The advantage of the cheap methods is that they can be applied to the samples of realisations obtained for the probabilistic assessment. These results of the benchmark exercise 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.
- For a sensitivity analysis of a model with dependent inputs with methods requiring special sampling schemes care must be taken that the sampling scheme also satisfies the input distribution.
- Algorithms with fixed maximal harmonic or fixed number of intervals per partition may not capture discontinuities and may produce systematic errors by under- or over-estimating the sensitivity indices.
- Random Balance Design shows no advantages when compared with a cheap method like EASL.
- For small sensitivity indices nearly all methods show bad convergence properties. Here, IHS and Sobol' methods are positive exceptions to the rule.

Finally, in the third step of the work, different sensitivity analysis techniques of all kinds were tested in practical situations by applying them to different realistic repository models, describing the concepts of different countries. The following application cases were chosen:

- A model for a rock salt repository in the Netherlands, implemented using the code EMOS-ECN,
- A model for a repository in argillaceous rock in the Netherlands, implemented using the code PORFLOW,
- A model for a granite repository in Spain, implemented using the code GoldSim,
- A model for a rock salt repository in northern Germany, implemented using the code package EMOS,
- A biosphere model developed for the Olkiluoto Island in Finland,
- A model for a French clay site, implemented using the code GoldSim.

It was found that some of the investigated models show peculiarities that can cause problems with some of the sensitivity analysis methods. A typical property of many repository system models is that the calculation results vary over several orders of magnitude, maybe even including zero. Since the variance is a statistical measure based on a linear scale, it is not best adequate for describing the variability of such models. The variance is then typically dominated by a few values out of a set of several thousands. This reduces the robustness of the variance-based sensitivity analysis. If the results show a lot of very low or zero values among a few higher ones, the analysis can be improved by performing an adequate data transformation.

From the work described in this report it can be concluded that probabilistic sensitivity analysis can reveal some very interesting and useful information about the global behaviour of models with uncertain parameters and is therefore a valuable tool in analysing the long-term safety of final repositories in the context of the safety case. Monte-Carlo-based methods are easily applicable and require only a moderate number of runs, but the results should be used with care. It should always be kept in mind that correlation- and regression-based sensitivity measures are not appropriate if the system shows a highly nonlinear or even non-monotonic behaviour. Variance-based techniques, on the other hand, are adequate also for nonlinear and non-monotonic systems, but often require a high computational effort and the robustness of the results may be poor. Graphical methods are a very good means of sensitivity analysis and sometimes even reveal dependencies that are not seen from the calculated sensitivity measures.

Different sensitivity analysis methods applied to a model often widely agree about the ranking of parameters, confirming the robustness of the SA, but sometimes they do not. For a detailed understanding of the system behaviour several methods should be applied. The different methods are complementary and each one provides specific information. Future work should concentrate on the question, which methods applied to which kind of system yields the most meaningful results in view of the safety case.

6. References

- Archer, G.; Saltelli, A.; Sobol', I.M. (1997): Sensitivity measures, anova-like techniques and the use of bootstrap, *Journal of Statistical Computation and Simulation*, 58, pp. 99-120.
- Badea, A.; Bolado, R. (2008): Review of sensitivity analysis methods and experience. Milestone M2.1.D.4. PAMINA IP.
<http://www.ip-pamina.eu/downloads/pamina.m2.1.d.4.pdf>
- Barnichon, J.D.; Neerdael, B.; Grupa, J.; Vervoort, A. (2000): CORA Project TRUCK-II. SCK-CEN Report 00/C070033/JDB/BNe/P-2, R-3409, January 2000.
- Becker, A.; Fischer, H.; Hofer, E.; Kloos, M.; Krzykacz, B.; Martens, K.-H.; Röhlig, K.-J. (1997): Evaluation of Elements Responsible for the Effective Engaged Dose Rates Associated with the Final Storage of Radioactive Waste: EVEREST Project, Volume 3a: Salt Formation, Site in Germany. Contract No. FI2W-CT90-0017, EUR 17449/3a EN, 1997
- Becker, D.-A.; Spießl, S. (2008): Variance-Based Sensitivity Analysis for the Long-Term Safety of an Underground Radioactive Waste Repository in Rock Salt. PSAM9, May 2008.
- Bettonvil, B.; Kleijnen, J.P.C. (1996): Searching for important factors in simulation models with many factors: sequential bifurcations, *European Journal of Operational Research*, 96, pp. 180-194.
- Bolado, R.; Badea, A. (2008): JRC's contribution to the benchmark based on synthetic PA cases. Technical report, PAMINA Project, Sixth Framework Programme, European Commission, 2008.
- Bolado, R.; Castaing, W.; Tarantola, S. (2009): Contribution to the sample mean plot for graphical and numerical sensitivity analysis. *Reliability Engineering & System Safety* 94: 1041-1049.
- Box, G.E.P.; Draper, N.R. (1987): *Empirical Model Building and Response Surfaces*. Wiley Series in probability and Mathematical Statistics. John Wiley & Sons, Inc.
- Broed, R. (2007): Landscape model configuration for biosphere analysis of selected cases in TILA-99 and in KBS-3H safety evaluation, 2007. Posiva Working Report 2007-108. Posiva Oy, Olkiluoto, Finland.
- Conover, W.J. (1980): *Practical Nonparametric Statistics*. Second edition. Applied Probability and Statistics. John Wiley and Sons, Inc.

- Cooke, R.M.; Van Noortwijk, J.M. (2000): Graphical Methods. In 'Sensitivity analysis', Saltelli, A., Chan, K. and Scott, E.M. (Editors). Wiley Series in probability and statistics. John Wiley & Sons Ltd.
- Cukier, R.I.; Fortuin, C.M.; Shuler, K.E.; Petschek, A.G.; Schaibly, J.H. (1973): Study of the Sensitivity of Coupled Reaction systems to Uncertainties in Rate Coefficients. I. Theory. Journal of Chemical Physics, Vol. 59 (8), pp. 3873-3878.
- Cukier, R.I.; Schaibly, J.K.; Shuler, K.E. (1975): Study of the Sensitivity of Coupled Reaction systems to Uncertainties in Rate Coefficients. III. Analysis of the Approximations. Journal of Chemical Physics, Vol. 63 (3), pp. 1140-1149.
- Cukier, R.I.; Levine, H.B.; Shuler, K.E. (1978): Review: Nonlinear sensitivity analysis of multiparameter model systems, Journal of Computational Physics, 26, pp. 1-42.
- Curti, E. (2003): Glass dissolution parameters: Update for "Entsorgungsnachweis". PSI Bericht Nr. 03-18, November 2003.
- De Bock, C.; Londe, L.; Blümling, P.; Rothfuchs, T.; Breen, B.; Huertas Ilera, F. (2004): Input Data and Functional Requirements Buffer Construction Technology. ESDRED Deliverable 1 WP1/Module 1, 23 December 2004.
- Ekström, P.A.: Eikos A (2005): Simulation Toolbox for Sensitivity Analysis. [http://www.luthagen.org/ekstrom/docs/Eikos A Simulation toolbox for Sensitivity Analysis.pdf](http://www.luthagen.org/ekstrom/docs/Eikos_A_Simulation_toolbox_for_Sensitivity_Analysis.pdf) , 2005
- Godon, N.; Gin, S.; Minet, Y.; Grambow, B.; Lemmens, K.; Aertsens, M. (2008): Reference report on the state of the art of glass properties and glass alteration during long term storage and under disposal conditions. NF-PRO Deliverable 1.3.1 of RTDC 1, Part 1. April 2008.
- Grambow, B. et. al. (2007): Final synthesis report RTDC 1: Dissolution and release from the waste matrix. NF-PRO Deliverable 1.6.3 of RTDC 1. May 2007.
- Grupa, J.B.; Houkema, M. (2000): Terughaalbare opberging van radioactief afval in diepe zout - en kleilagen - Modellen voor een veiligheidsstudie. NRG Report 21082/00.33017/P, Petten, June 2000.
- Heijdra, J.J.; Hamilton, L.F.M.; Prij, J.; Slagter, W. (1995): Transient Convergence and Compaction of Crushed salt as Incorporated in the Computer Code EMOS. ECN-I-95-029, ECN, Petten, November 1995.
- Helton, J.C.; Johnson, J.D.; Sallaberry, C.J.; Storlie, C.B. (2006): Survey of Sampling-based Methods for Uncertainty and Sensitivity Analysis. Reliability Engineering and System Safety, Vol. 91, pp. 1175-1209.

- Homma, T.; Saltelli, A. (1996): Importance measures in global sensitivity analysis of nonlinear models, *Reliability Engineering and System Safety* 52, pp. 1-17.
- Ikonen A.T.K., Leppänen V., Gunia M., Broed R. (2008): Handling of climate and future human actions as scenario uncertainty within terrain and ecosystems as a part of biosphere assessment for the Olkiluoto site. In: Strand, P., Brown, J. & Jølle, T. (eds.). *Proceedings of the International Conference on Radioecology & Environmental Radioactivity*, 15-20 June 2008, Bergen, Norway. Oral and Oral Poster Presentations.
- Koda, M., McRae, G. J., Seinfeld, J. H. (1979): Automatic sensitivity analysis of kinetic mechanisms, *International Journal of Chemical Kinetics*, 11, pp. 427 – 444.
- Krzykacz, B. (1990): 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.
- Levandowski, D.; Cooke, R.M.; Duintjer Tebbens, R.J. (2007): Sample-based estimation of correlation ratio with polynomial approximation. *ACM Transactions on Modeling and Computer Simulation*, 18(1):3:1-3:16, 2007.
- Minet, Y.; Lartigue, J.E. (2008): Modelling of integral modular material experiments using the Hytec geochemical code - Proposition of a source term model. NF-PRO Deliverable 1.3.8 of RTDC 1. CEA, February 2008.
- Morris, M.D. (1991): Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2), pp. 161-174.
- NEA - Nuclear Energy Agency (1989): PSACoin level E intercomparison. Technical report, OECD, Paris, 1989.
- Pearson, K. (1905): On the General Theory of Skew Correlation and Non-linear Regression, volume XIV of *Mathematical Contributions to the Theory of Evolution*. Drapers' Company Research Memoirs, Cambridge University Press, Cambridge, UK, 1905.
- Pepin, G.; Plas, F.; Prváková, S.; Nilsson, K.-F. (2007): Milestone 4.3.1 First Benchmark specification for the uncertainty analysis based on the example of the French clay site. PAMINA milestone M4.3.1, April 2007.
- Plischke, E. (2009): An effective algorithm for computing global sensitivity indices (EASI). *Reliability Engineering&System Safety*, 2009. Submitted Manuscript.
- Plischke, E.; Röhligh, K.-J.; Badea, A.; Bolado, R.; Ekström, P.-A.; Hotzel, S. (2009): Sensitivity analysis benchmark based on the use of analytic and synthetic PA cases. Milestone M2.1.D.11. PAMINA IP.

- Plischke, E.; Röhlig, K.J. (2008): 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.
- Poley, A.D. (2000): Concept ontwerp terughaalbare opslag van radioactief afval in diepe boorgaten in steenzout (Torad-B). NRG report 21514/00.31223/C, May 2000.
- PORFLOW (1999): A software tool for multiphase flow, heat and mass transport in fractured porous media. User's manual version 4.00. ACRi, California, 1999.
- Prado-Herrero, P. SimLab and GTM1. An External Model Example – The PSACOIN Level E. <http://sensitivity-analysis.jrc.it/tutorial/SimLab%2520and%2520GTM1-TT1.pdf>, 2005.
- Prváková, S.; Bolado, R.; Badea, A.; Nilsson, K.-F.; Pepin, G.; Treille, E. (2008): First benchmark for the uncertainty analysis based on the example of the French clay site. RESULTS. PAMINA milestone M4.3.2, February 2008.
- Saltelli, A. (2002): Making Best Use of Model Evaluations to compute Sensitivity Indices. Computer Physics Communications, 145, pp. 280-297.
- Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. (2008): Global Sensitivity Analysis – The Primer. John Wiley&Sons, Chichester, 2008.
- Saltelli A.; Tarantola, S. (2002): 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.
- Saltelli, A.; Tarantola, S.; Campolongo, F.; Ratto, M. (2004): Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models John Wiley & Sons.
- Saltelli, A.; Tarantola, S.; Chan, K.P.S. (1997): Sensitivity Analysis of Model Output, An Improvement of the FAST Method, EUR 17338EN, 1997, Luxembourg.
- Saltelli, A.; Tarantola, S.; Chan, K. (1999): A Quantitative, Model Independent Method for Global Sensitivity Analysis of Model Output. Technometrics, 41(1), pp. 39-56.
- Saltelli, A.; Chan, K.; Scott, E. (2000): Sensitivity Analysis. John Wiley&Sons, Chichester, 2000.
- Schaibly, J.K.; Shuler, K.E. (1975): Study of the Sensitivity of Coupled Reaction systems to Uncertainties in Rate Coefficients. II. Applications. Journal of Chemical Physics, 59 (8), pp. 3879-3888.
- Schröder, T.J. (2008a): PAMINA task 2.1.D – Update model parameters for sensitivity analysis. NRG Note 2.1952.21/08.88799, Petten, 11 April 2008.

- Schröder, T.J. (2008b): EMOS versie ccm2 voor PAMINA WP2.1D: aanpassingen van de porositeit-permeabiliteit relatie en de compactiesnelheid in EMOS ccm1. NRG Note 2.1952.21/08.89397, Petten, 4 June 2008.
- Sillen, X. (2002): BENIPA Project - Plastic clay reference case: Data for radionuclide transport calculations, Rev.1. SCK•CEN, 27 August 2002.
- Simlab 2.2 Reference Manual (2003), JRC Ispra, Italy.
- Sobol', I.M. (1993): Sensitivity estimates for Nonlinear Mathematical Models. MMCE, 1 (4), 407-414.
- Sobol', I.; Tarantola, S.; Gatelli, D.; Kucherenko, S.; Mauntz, W. (2007): Estimating the approximation error when fixing unessential factors in global sensitivity analysis. Reliability Engineering&System Safety, 92:957-960, 2007.
- Spiers C.J., et al. (2005): Buffer/backfill in salt (long-term compaction, additives, precompacted bricks). NF-PRO Meeting, October 2005.
- Storck, R.; Buhmann, D.; Hirsekorn, R.P.; Nies, A.; Rausch, H. (1990): EMOS: Programmpaket zur Langzeitsicherheitsanalyse eines Endlagers für radioaktive Abfälle - Version 4. GSF-Bericht /90, TL/90, September, 1990.
- Stuart, A.; Ord, K.; Arnold, S. (2009): Kendall's advanced theory of statistics: Classical inference and the linear model, volume 2A. John Wiley&Sons, Hoboken, NJ, 2009.
- Tarantola, S.; Gatelli, D.; Mara, T.A. (2006): Random Balance Designs for the Estimation of First Order Global Sensitivity Indices. Reliability Engineering and System Safety 91, pp. 717-727.
- Zhang, X.; Grupa, J. (2006): NF-Pro deliverable 3.5.7 – Compaction behaviour and permeability of low porosity compacted salt grit (dry and wet). NRG Report 21146/06.77412/P, December 2006.

