

**Investigation of Modern  
Methods of Probabilistic  
Sensitivity Analysis of Final  
Repository Performance  
Assessment Models  
(MOSEL)**



Investigation of Modern  
Methods of Probabilistic  
Sensitivity Analysis of Final  
Repository Performance  
Assessment Models  
(MOSEL)

Sabine Spießl  
Dirk-Alexander Becker

June 2017

**Remark:**

The work underlying report was conducted with financial support from the German Federal Ministry for Economic Affairs and Energy (BMWi) under sign 02E10941.

The authors are responsible for the contents of this report.

The report reflects the perception and opinion of the contractor and does not necessarily agree with the opinion of the sponsor.

**Key Words**

Final Disposal, Performance Assessment, Probabilistic Analysis, Radioactive Waste, Sensitivity Analysis, Uncertainty Management

## **Abstract**

Sensitivity analysis is a mathematical means for analysing the sensitivities of a computational model to variations of its input parameters. Thus, it is a tool for managing parameter uncertainties. It is often performed probabilistically as global sensitivity analysis, running the model a large number of times with different parameter value combinations. Going along with the increase of computer capabilities, global sensitivity analysis has been a field of mathematical research for some decades.

In the field of final repository modelling, probabilistic analysis is regarded a key element of a modern safety case. An appropriate uncertainty and sensitivity analysis can help identify parameters that need further dedicated research to reduce the overall uncertainty, generally leads to better system understanding and can thus contribute to building confidence in the models.

The purpose of the project described here was to systematically investigate different numerical and graphical techniques of sensitivity analysis with typical repository models, which produce a distinctly right-skewed and tailed output distribution and can exhibit a highly nonlinear, non-monotonic or even non-continuous behaviour.

For the investigations presented here, three test models were defined that describe generic, but typical repository systems. A number of numerical and graphical sensitivity analysis methods were selected for investigation and, in part, modified or adapted. Different sampling methods were applied to produce various parameter samples of different sizes and many individual runs with the test models were performed. The results were evaluated with the different methods of sensitivity analysis. On this basis the methods were compared and assessed.

This report gives an overview of the background and the applied methods. The results obtained for three typical test models are presented and explained; conclusions in view of practical applications are drawn. At the end, a recommendation for executing sensitivity analysis in a practical study is given



## Table of Contents

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
1.1	Background .....	1
1.2	Management of uncertainties.....	1
1.2.1	Identification and quantification of inherent uncertainties .....	2
1.2.2	Uncertainty and sensitivity analysis .....	3
1.2.3	Time-dependent versus maximum analysis .....	4
1.2.4	Assessment of uncertainties and drawing consequences .....	4
1.3	Sensitivity analysis and its role in the safety case .....	5
<b>2</b>	<b>Overview of methods for sensitivity analysis.....</b>	<b>9</b>
2.1	Basic terms and notations.....	9
2.2	Monte-Carlo (MC) methods .....	9
2.3	Sampling .....	11
2.3.1	Random sampling.....	11
2.3.2	Stratified sampling .....	11
2.3.3	Quasi-random (low-discrepancy) sampling .....	13
2.3.4	Periodic sampling .....	14
2.3.5	Correlation.....	15
2.3.6	Comparison of sampling methods.....	16
2.4	Uncertainty analysis.....	19
2.4.1	Graphical uncertainty analysis .....	19
2.4.2	Numerical uncertainty analysis .....	21
2.5	Graphical methods for sensitivity analysis.....	23
2.5.1	Scatterplots.....	23
2.5.2	Contribution to the sample mean (CSM) .....	23
2.5.3	Contribution to the sample variance (CSV) .....	24
2.5.4	Cumulated sums of normalised re-ordered output (CUSUNORO) .....	24
2.5.5	Conditional cobweb plots and mean rank plots .....	25
2.6	Correlation- and regression-based methods .....	26

2.6.1	Pearson's correlation coefficients .....	26
2.6.2	Standardised regression coefficients (SRC).....	27
2.6.3	Partial correlation coefficients (PCC) .....	28
2.6.4	Rank transformation .....	29
2.7	Variance-based methods.....	30
2.7.1	The Sobol' method.....	32
2.7.2	(Extended) Fourier Amplitude Sensitivity Test (FAST/EFAST).....	32
2.7.3	Random Balance Design (RBD) .....	33
2.7.4	Effective Algorithm for computing global Sensitivity Indices (EASI).....	34
2.7.5	Metamodelling and State-Dependent Parameter method (SDP).....	34
2.8	Non-parametric methods .....	35
2.8.1	Two-sample tests.....	36
<b>3</b>	<b>Description of the test systems .....</b>	<b>37</b>
3.1	Repository for Spent Fuel and HLW in clay.....	37
3.2	Repository for LILW in rock salt .....	38
3.3	Repository for Spent Fuel and HLW in rock salt (SAM-GBS) .....	41
<b>4</b>	<b>Graphical sensitivity analysis .....</b>	<b>45</b>
4.1	Scatterplots.....	45
4.1.1	Scatterplots for the clay system .....	45
4.1.2	Scatterplots for the LILW system .....	47
4.2	CSM, CUSUNORO and CSV plots .....	50
4.2.1	CSM plots for the clay system.....	50
4.2.2	CSM plots for the LILW system.....	52
4.2.3	CSM versus CUSUNORO .....	54
4.2.4	CSV plots .....	55
4.3	Summary .....	56
<b>5</b>	<b>Correlation- and regression-based sensitivity analysis .....</b>	<b>59</b>
5.1	Comparison of correlation and regression .....	59
5.2	The principle message.....	60

5.3	Direct evaluation of the model output (SRC) .....	61
5.3.1	The clay system.....	61
5.3.2	The LILW system.....	61
5.4	Rank-based evaluation (SRRC).....	62
5.4.1	The clay system.....	62
5.4.2	The LILW system.....	63
5.5	Summary .....	64
<b>6</b>	<b>Variance-based sensitivity analysis .....</b>	<b>65</b>
6.1	Sobol' method.....	66
6.1.1	The clay system.....	66
6.1.2	The LILW system.....	67
6.2	EFAST .....	69
6.2.1	The clay system.....	69
6.2.2	The LILW system.....	70
6.3	RBD.....	71
6.3.1	The clay system.....	72
6.3.2	The LILW system.....	72
6.4	EASI .....	73
6.4.1	The clay system.....	73
6.4.2	The LILW system.....	74
6.5	SDP .....	75
6.5.1	The clay system.....	75
6.5.2	The LILW system.....	75
6.6	Comparison with regression-based sensitivity analysis.....	76
6.7	Summary .....	78
<b>7</b>	<b>Influence of the sampling scheme.....</b>	<b>81</b>
7.1	Random sampling versus Latin Hypercube Sampling .....	81
7.2	Random sampling versus quasi-random LpTau sampling.....	83
7.3	Summary .....	84

<b>8</b>	<b>Detailed sensitivity investigations.....</b>	<b>85</b>
8.1	Influence of number of parameters .....	85
8.2	Hidden sensitivities .....	88
8.3	Handling of two-split output by a dynamic two sample test .....	91
8.4	Summary .....	93
<b>9</b>	<b>Output transformation.....</b>	<b>95</b>
9.1	Definition of transformations .....	96
9.2	Selection of the threshold .....	98
9.3	Sensitivity analysis results .....	99
9.3.1	The SAM-GBS model .....	99
9.3.2	The LILW model .....	102
9.4	Summary .....	105
<b>10</b>	<b>Handling of parameter dependencies .....</b>	<b>107</b>
<b>11</b>	<b>Summary, recommendations and outlook.....</b>	<b>109</b>
11.1	Project summary and general conclusions.....	109
11.1.1	Background and project setup .....	109
11.1.2	Graphical sensitivity analysis .....	110
11.1.3	Correlation- and regression-based sensitivity analysis.....	111
11.1.4	Variance-based sensitivity analysis .....	111
11.1.5	The role of the sampling scheme .....	112
11.1.6	Dependence on the set of parameters .....	113
11.1.7	Two-sample tests.....	113
11.1.8	Output transformation .....	114
11.2	Recommendation for proceeding in practical cases .....	115
11.3	Related work and outlook .....	120
	<b>List of References.....</b>	<b>123</b>
	<b>List of Figures .....</b>	<b>129</b>
	<b>List of Tables.....</b>	<b>133</b>

<b>A</b>	<b>Data of the test systems.....</b>	<b>135</b>
A.1	Repository for Spent Fuel and HLW in clay.....	135
A.2	Repository for Low- and Intermediate-Level waste (LILW) in rock salt...	140



# **1 Introduction**

## **1.1 Background**

The long-term performance of repositories for radioactive waste has to be analysed using adequate computer models. Such models are made up by mapping the existing or planned real repository structure to a simplified system of coupled components that can be described by numerical algorithms, taking into account various effects like radionuclide mobilisation, gas production, pressure build-up, fluid flow, sorption, dissolution and precipitation or other effects or processes that might be considered important in the considered case. The resulting model typically depends on a number of input parameters. Such a model cannot describe the real evolution of the system, but if set up thoroughly, it should at least be able to yield results that allow an assessment of the long-term safety of the system, provided that the parameters are given the “right” values. This, however, is a demanding problem, as a single model parameter normally represents, in a stylised manner, a more or less complex group of physical or chemical influences to the real system, which are not known in detail and cannot be quantified exactly. Such parameters can be measured in situ or estimated by experts, but are always subject to a range of uncertainty. Sometimes a conservative choice is possible, which means that the parameter is set to a value on the pessimistic end of its range. In many cases, however, a uniquely pessimistic choice is not possible. Additionally, over-conservatism can destroy the proof of safety. Therefore, conservatism can be applied only to a limited extent.

## **1.2 Management of uncertainties**

A proper management of uncertainties is an essential part of any safety case, since a meaningful statement on the long-term safety of a final repository system is only possible if its uncertainty can be quantified reliably. With respect to an overall assessment process, management of uncertainties comprises several steps, each of which requires thorough investigation.

### 1.2.1 Identification and quantification of inherent uncertainties

Once identified, all uncertainties will be assessed and screened with regard to their relevance for the safety case. While some of them can be ruled out for obvious low relevance, others can be easily avoided using conservative assumptions. Though this is a somewhat subjective process, it is nevertheless necessary in order to reduce the set of uncertainties on the “input side” of the assessment process to a manageable number. In this report, these uncertainties are called *inherent uncertainties*. They have to be taken as they are; the assessment process has no direct influence on the initial uncertainties, although it can trigger specific research or development work in order to reduce them.

With respect to their origin, the inherent uncertainties of a final repository safety assessment are often classified in three categories:

- *Scenario uncertainties* reflect our lack of knowledge about the future development of the repository and its environment.
- *Model uncertainties* result from the fact that a model can never describe the reality exactly.
- *Parameter uncertainties* refer to the numerical input data of the applied models.

While parameter uncertainties can be directly described by a mathematical formulation, scenario and model uncertainties are, in principle, non-numerical. If they are to be included in a formal uncertainty assessment process, they have to be mapped to numerical values in an adequate manner.

With respect to their nature, uncertainties can be ideally distinguished in two types:

- *Epistemic uncertainties* are due to our limited knowledge of the natural conditions and processes. In principle, they can be reduced by performing adequate research.
- *Aleatory uncertainties* are due to influences that are or have to be considered random, according to our present-day understanding of the underlying processes. Aleatory uncertainties cannot be reduced.

In practice, a proper distinction between these types is not always easy or can even be impossible. Many uncertainties are an overlap of both types, one or the other dominat-

ing. Moreover, it can be a matter of the scale of detail, whether an uncertainty appears epistemic or aleatory.

Numerical uncertainties can be described using *probability density functions* (pdf). The integral of a pdf over a limited interval of values gives the total probability that the unknown actual parameter value is in this interval. The integral of a pdf over the total parameter range is always 1.

Proper identification, assessment and quantification of the inherent uncertainties is a very complex and time-consuming task within the safety case, requiring a lot of specific expert knowledge in physics, chemistry, hydrodynamics, geology, climatology, biology and other natural sciences, but also in mathematics. The correct choice of pdf types depends on the degree of knowledge about the quantity in question, which has to be properly analysed.

Inherent uncertainties may be subject to mutual influences. If, for example, the porosity of a porous medium is low one can expect that the permeability of the medium will also be low (or vice versa), but this is not sure and a strict formula to connect these values cannot be given. Such statistical interdependencies often occur if the underlying physical processes are not completely understood or hard to describe in detail. For mathematical purposes, the interdependencies are normally expressed as statistical correlation, although this concept does not cover all kinds of dependency.

### **1.2.2 Uncertainty and sensitivity analysis**

For numerical assessment of long-term safety a model is necessary, which is assumed to represent the actual system development in a representative manner. Such a model has a number of input parameters and produces one or several output values that allow, in comparison with specific reference values, statements about the safety of the system or assessment of the performance of a subsystem. Such values are called safety indicators or performance indicators, respectively (/BEC 03/, /BEC 09a/).

Any safety statement derived from a model is subject to some uncertainty, which results from the inherent uncertainties. To quantify the uncertainty of the model output one has to evaluate the model a number of times with the input parameters varied according to the inherent uncertainties and to analyse the model output statistically. This

procedure is called uncertainty analysis. Uncertainty analysis refers to the model output alone with no reference to the inherent uncertainties.

A more sophisticated question is, to what extent a specific inherent uncertainty or a group of inherent uncertainties affects the overall uncertainty of the safety statement. There are many mathematical approaches to this problem, addressing different aspects and based on different concepts. This is called sensitivity analysis.

Uncertainty and sensitivity analysis of model output is a purely mathematical procedure. One always has to keep in mind that any statement concerning uncertainty or sensitivity that has been derived in this manner is a statement about the model under consideration, and not a statement about the real system.

### **1.2.3 Time-dependent versus maximum analysis**

In the context of repository safety, it is often more interesting for the performance assessor how the – real or hypothetic – system under consideration will contaminate the environment maximally, no matter at which time this maximum arises, than what happens at a specific point in time. Therefore, also uncertainty and sensitivity analysis can be performed either for the calculated contaminant releases or annual doses at specific points in time or for the run maxima. The results answer different questions and can therefore be completely different. An extreme situation would be that the maximum occurs as a sharp peak and some specific parameter considerably influences the time at which this happens, but not at all the height of the maximum. This parameter would not show up in a maximum-related sensitivity analysis, but could be one of the most important in a time-related analysis.

### **1.2.4 Assessment of uncertainties and drawing consequences**

The uncertainty and sensitivity measures that have been calculated are numerical numbers that do not provide any added value to uncertainty management by themselves. These values have to be assessed in the context of the real system under consideration, which again requires expert knowledge. Therefore, it is necessary that numerical uncertainty and sensitivity results be not simply listed without comment and maybe used for assembling parameter ranking lists, but presented and explained with relation to the system under consideration. A proper interpretation of the obtained re-

sults will often lead to an improved understanding of the system behaviour and consequently to more confidence in the modelling results. If a plausible explanation is not found this can be a hint to some essential lack of understanding or even to an error in the model.

### **1.3 Sensitivity analysis and its role in the safety case**

Uncertainty and sensitivity analysis, in the context addressed here, is the systematic investigation of the influences of parameter uncertainties to the results of model calculations. While the goal of uncertainty analysis is the assessment of the uncertainty of the calculated results under influence of the entirety of all existing parameter uncertainties, sensitivity analysis aims at investigating the sensitivity of the model to variations of a single parameter or a group of interacting parameters within its range of uncertainty.

Sensitivity analysis can be performed as a *local* or *global* analysis by means of *deterministic* or *probabilistic* methods. Local sensitivity analysis addresses the behaviour of the model under variation of the investigated parameter alone while all others are kept constant. In contrast, global sensitivity analysis means that all parameters that are considered uncertain are varied together, nevertheless focusing the evaluation to each parameter individually. Approaches of the former type cannot reveal the effects of interactions of different parameters and describe the model behaviour only at a fixed point in the parameter space. Therefore, global sensitivity analysis is normally more meaningful. In a deterministic investigation the model is evaluated with only a few, specifically selected parameter combinations and the results are compared to each other in order to analyse the influences of this variation in detail. Such investigations can provide improved understanding of the model behaviour. For systematic sensitivity quantification, however, a high number of model runs is necessary, exploring the parameter space (or a subspace) as homogeneously as possible, and the results must be evaluated by some adequate mathematical or graphic procedure. This is called a probabilistic approach, because it takes account of the (assumed) probabilities of parameter combinations. Often, local is equalised with deterministic, global with probabilistic sensitivity analysis, but these terms refer to different aspects, and a global deterministic approach is as well conceivable as a local probabilistic one.

The report at hand deals with global probabilistic sensitivity analysis. For a probabilistic investigation all uncertain parameters have to be identified and their uncertainty has to

be quantified using adequate *probability density functions* (pdf). Then the model is executed a high number of times with parameter values that are drawn according to their assumed probabilities. The parameter matrix, containing one set of values for each model run, is called *sample*, the *sample size* is the number of contained sets, i. e. the number of runs to be executed. A characteristic output variable is collected from all runs. The sensitivity analysis procedure finally quantifies or visualises the sensitivity of this output variable to variations of the input parameters.

Probabilistic sensitivity analysis has been a field of specific mathematical interest for the last two or three decades, triggered by the constantly increasing performance of computers. A complex model that would require weeks to be evaluated just a few hundred times 30 years ago can be executed tens of thousands of times within a few hours on a modern parallel system. Therefore, mathematical concepts have attracted interest that used to be impossible to be applied in practice, and a variety of sophisticated sensitivity analysis methods have been developed. Mathematicians tend to test their methods using specifically designed models or test functions. This, however, does not seem to reflect correctly the performance of these methods if applied to complex performance assessment models with their typical properties like extreme nonlinearities and a wide, non-gaussian distribution of possible output values.

In the context of long-term performance assessment for radioactive waste repositories probabilistic uncertainty and sensitivity analysis is increasingly considered as a powerful and indispensable tool for building confidence in the safety case. One has to keep in mind, however, that such methods can never make statements about the real repository system but only about the numerical model under consideration, including the applied pdfs of the input parameters. While a proper uncertainty analysis yields simple and mathematically unique results concerning the overall uncertainty of the model output under the influence of the parameter uncertainties, sensitivity analysis provides more sophisticated insights to the model behaviour. Primarily, it acts as a tool for identifying those parameters that are most important for the model in that sense that their uncertainty dominates the uncertainty of the model output. It has turned out, however, that a detailed sensitivity analysis is often very valuable for improving the general understanding of the model, recognising implausible behaviour and identifying model errors. Sensitivity analysis for repository models has been under investigation for some time and was already addressed within the project PAMINA (/BEC 09b/). Comprehen-

sive sensitivity analyses for the American repository projects WIPP and Yucca Mountain have been performed by Helton and others (/HEL 98/, HEL 09/).

The goal of the work documented here is to test different classical and modern methods of sensitivity analysis with realistic long-term performance assessment models, to assess their usability in practice and to develop a sensible way of proceeding. A short theoretical overview is given in chapter 2. In chapter 3 the test systems used for investigations are described. Chapters 4 to 6 contain the preliminary results of the work performed so far, and in chapter 7 some preliminary conclusions are drawn.



## 2 Overview of methods for sensitivity analysis

This chapter gives an overview of the main theoretical concepts underlying the applied methods and techniques. This is meant for the reader's orientation, but not as a comprehensive description of the theory, which can be found in the indicated literature. A good survey of methods is also given in /HEL 06/.

### 2.1 Basic terms and notations

This section briefly explains some essential terms and notations used in this report.

In the context of this report, a *model* is assumed to be a set of numerical instructions that transforms a defined number of  $k$  input values  $x_1, \dots, x_k$  uniquely into one output value  $y$ . The input values are called *parameters*. Dependent on its degree of uncertainty and/or available information in form of data or general knowledge, the distribution of each parameter is characterised by its *probability density function* (pdf). The integral of the pdf over the total parameter interval is always 1. Common distribution types are uniform, log-uniform, normal, log-normal or triangular distributions. The resulting distribution of the output variable is then defined by the model.

A set of input parameter values is interpreted as a realisation of a set of probabilistic parameters  $X_1, \dots, X_k$ . The model output is represented by the probabilistic variable  $Y$ .  $X$  and  $Y$  are called probabilistic as they are represented by values that are distributed according to characteristic probabilities. A *sample* of size  $N$  consists of a collection of  $N$  complete sets of parameter values for the model under consideration, i. e.  $N$  realisations of the parameter vector  $X$ , generated by some sampling method. The model can be described using a function  $f$ :

$$Y = f(X) = f(X_1, X_2, \dots, X_k) \tag{2.1}$$

where  $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,N})$  and  $Y = (y_1, y_2, \dots, y_N)$ ;  $i = 1, 2, \dots, k$ .

### 2.2 Monte-Carlo (MC) methods

The probabilistic methods applied in the investigations belong to the field of Monte-Carlo (MC) methods. Such methods are based on the idea of carrying out a number of model evaluations with input data chosen randomly with respect to their probability

density functions. MC methods have been originally developed for multi-dimensional numerical integration, but can be used to solve various kinds of computational problems (/NIE 92/). Different techniques of statistical sampling are available for utilisation within MC methods to approximate solutions to problems (/MAR 08/).

Various types of Monte-Carlo methods have been developed to improve convergence and/or accuracy and to reduce CPU cost and time. These different types are characterised by the type of sampling schemes and/or other information used in the analysis or in the process of finding a solution to a given problem.

The plain Monte-Carlo, also called crude or traditional (/LIJ 00/) or just Monte-Carlo methods are the basic version of Monte-Carlo methods in which random sampling schemes (see chapter 2.3.1) are used.

To increase the efficiency of the Monte-Carlo method, techniques for variance reduction have been established. A means for reaching this goal is utilisation of stratified sampling techniques (e.g. Latin Hypercube Sampling, see chapter 2.3.2.1) (/NIE 92/).

In Quasi-Monte-Carlo (QMC) methods, the samples are generated with quasi-random sampling schemes. These sampling schemes use specifically designed deterministic sequences with better uniformity properties (/NIE 92/, /OEK 02/ and /GOE 09/). At first sight, such sequences look random, but they are designed to avoid the typical clustering of points generated by a memory-less random process. The according sampling algorithms are designed to minimise the discrepancy, i.e. the deviation from an ideal uniform distribution. The resulting sequences of numbers are therefore called low-discrepancy sequences. Quasi-Monte-Carlo methods can provide more accurate results than plain Monte-Carlo methods with reduced CPU cost and time (/NIE 92/).

A disadvantage of Quasi-Monte-Carlo methods is that they may produce worse results for systems with high dimensions. Theoretically – though not necessarily in practice – the superiority of QMC compared to plain MC decreases rapidly with the dimension of the parameter space. To overcome these limitations, hybrid Monte-Carlo methods have been introduced. Hybrid Monte-Carlo methods employ sampling schemes utilising QMC sampling and another type of sampling scheme (/OEK 02/).

## 2.3 Sampling

In this section the most important sampling techniques are shortly explained.

### 2.3.1 Random sampling

Random sampling is best adequate for all evaluations that require totally independent sample points. This is necessary, for example, to derive proper and unbiased statistical statements about the model under consideration. A pure uncertainty analysis with determination of statistical measures like mean or quantiles, confidence bounds or probability of limit exceedance, should therefore be done using random sampling.

To perform a proper random sampling a random number generator is needed. It is, however, not an easy task for a computer to generate a random sequence. A hardware random number generator that makes use of some random physical effect like semiconductor noise provides a possibility to generate genuine random numbers. Normally, however, algorithms for pseudo-random number generating (PRNG) are applied. Starting with a given numerical value, called *seed*, such algorithms generate sequences of numbers in a certain interval that look random but are actually fully deterministic and reproducible. The values produced by a PRNG should not only be distributed uniformly but should not show any sign of “non-randomness”, which is not easy to define. One important – but, obviously, not sufficient – indicator for the quality of a PRNG is the period length, which gives the total length of the sequence until it starts from the beginning, repeating all numbers in exactly the same order. Different PRNGs of different quality are available (/DUT 09/). For MC methods, however, random sampling is often not optimal, because instead of providing a homogeneous coverage of the parameter space it tends to building clusters and gaps. This is due to the fact that random sampling does not – and should not! – have any memory, which could lead to preference of “empty” regions in the parameter space.

### 2.3.2 Stratified sampling

Generally, stratified sampling means that the statistical population from which the sample is to be drawn is exhaustively partitioned into non-overlapping subsets (strata) with specific common characteristics. Random sampling is then performed independently for each of the subsets. In the context of Monte-Carlo methods, where the population is

a continuous, multi-dimensional parameter space, strata can be defined by dividing each parameter interval into subintervals and combining these adequately. If only one parameter is to be drawn  $N$  times, one can divide its total interval into  $N$  subintervals of equal probability  $1/N$  and then take one value randomly from each subinterval, which makes sure that the whole interval is homogeneously covered by the sample. For high-dimensional parameter spaces, however, this simple approach does not work, because the required sample size would become extremely high. With, e.g., ten parameters even a very rough division of two subintervals per parameter would require more than 1000 sample points. It is therefore inevitable that many combinations of parameter subintervals remain unconsidered in the sample.

### **2.3.2.1 Latin Hypercube Sampling (LHS)**

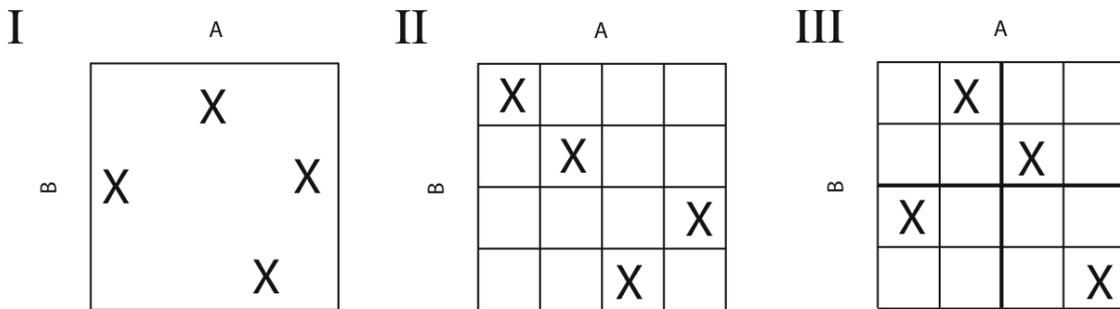
The Latin Hypercube Sampling (LHS) method is a form of stratified random Monte Carlo sampling scheme, first suggested by /MCK 79/.

A Latin Square is a quadratic grid filled with different symbols, each occurring exactly once per row and column. The generalisation of this concept to a multi-dimensional space is called a Latin Hypercube. This principle can be used to define a sampling algorithm that provides a more homogeneous coverage of the parameter space than random sampling. A true Latin Hypercube with its full combinatorial complexity, however, is not needed for this purpose; only one symbol is regarded.

First, the range of each of the  $k$  parameters is subdivided in  $N$  subintervals of equal probability. Then it has to be made sure by an adequate algorithm that each  $k-1$ -dimensional hyperplane is met by exactly one sample point. The latter requirement could be easily fulfilled by simply following a diagonal through the hypercube, which would provide a rather poor coverage of the total space. Therefore, the subintervals should at least be shuffled randomly for each of the parameters before combining them. A possibility to further improve the sampling is to subdivide the total parameter space into larger cells of equal probability and to add the requirement that each of these cells is met equally often. This is called orthogonal sampling.

The concept is illustrated in Fig. 2.1 with a simple example of two parameters and four sample points. With random sampling (I) the points are randomly distributed. Simple LHS (II) guarantees that each line and row of the parameter space is met exactly once.

Orthogonal sampling (III) additionally provides a homogeneous density of points in each 4x4-cell.



**Fig. 2.1** The principle of Latin Hypercube Sampling (taken from Wikipedia)

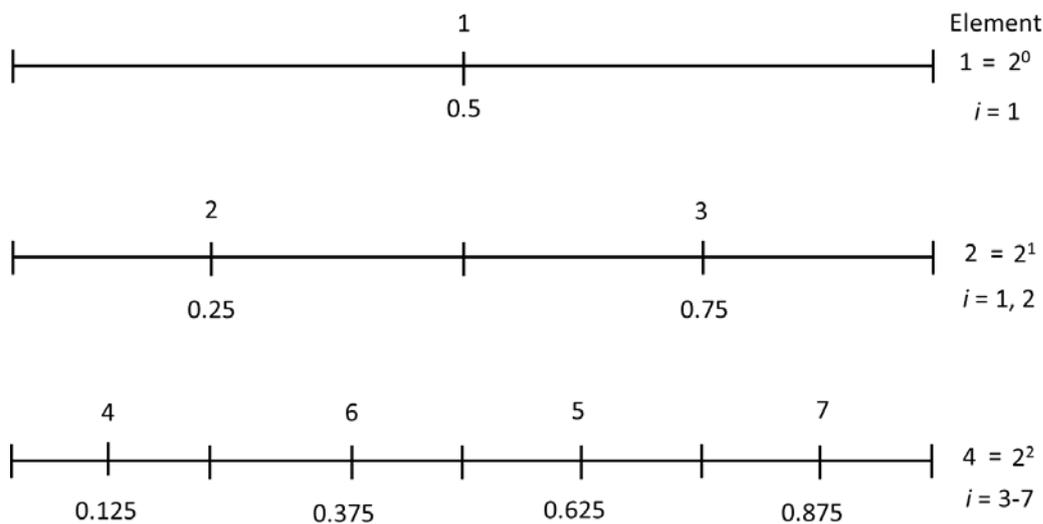
### 2.3.3 Quasi-random (low-discrepancy) sampling

In view of faster convergence and higher accuracy of a Monte Carlo simulation, for specific investigations uniformity of the distribution of sample points may be more important than 'true' randomness (/NIE 92/). With Quasi-Random Sampling (QRS) schemes, more uniformity can be put into the samples. Such schemes use deterministic formulas to generate sequences of numbers with better uniformity properties (/KIP 08/). Quasi-random sequences should nevertheless provide homogeneity in each dimension and at any stage of the sampling process, which is why, at the first sight, they look random. Unlike random sampling, however, a QRS algorithm makes use of a memory, which prevents it from building the typical clusters and gaps. QRS algorithms aim at minimising the mathematical discrepancy, which is a measure for the deviation of a given distribution from the ideal one. Therefore, the sampling sequences generated by such algorithms are sometimes called low-discrepancy (LD) sequences.

In particular for problems with few parameters, LD sequences can help to obtain reliable sensitivity analysis results with low computational cost (e.g., /MOR 94/). For high-dimensional problems, the performance depends on which type of LD sequences is used and/or how the points of the sequences interacts with the studied problem (/LEM 04/).

### 2.3.3.1 Sobol' LpTau sequences

For the investigations with quasi-random sampling described in this report Sobol' LpTau sequences were used (/SOB 67/). These sequences are designed to cover the  $k$ -dimensional parameter space as well as any of its lower-dimensional projections as homogeneously as possible. The Sobol' sampling establishes successively finer partitions of the  $[0,1]$  interval on the base of two and then rearranges the coordinates in each dimension. The generated sample is then transformed to the desired intervals and distribution for each parameter. Fig. 2.2 illustrates this principle in one dimension for a uniform distribution. Details on calculating Sobol' LpTau sequences in higher dimensions can be found, e.g., in /SHU 94/.



**Fig. 2.2** Illustration of the basic idea of the Sobol' LpTau sequences in  $[0,1]$

### 2.3.4 Periodic sampling

Some sensitivity analysis methods require specific sampling schemes. The basic idea of a family of methods is to introduce some periodicities in the sample and to rediscover them in the calculation results by performing a Fourier analysis. These methods have the drawback of working only with specifically drawn samples. Such methods – (E)FAST and RBD – are described and investigated later in this report. At this point only some general considerations on periodic sampling are given.

As with all sampling methods, a largely homogeneous coverage of the parameter space in all dimensions is desirable. As periodic sampling anyway tends to leaving

large gaps in the multi-dimensional parameter space, one should at least try to reach homogeneous distributions for the individual parameters. Best homogeneity can be reached by choosing a linear sequence that changes its direction when reaching the interval end. This technique leads to a triangle-shaped sampling curve

### 2.3.5 Correlation

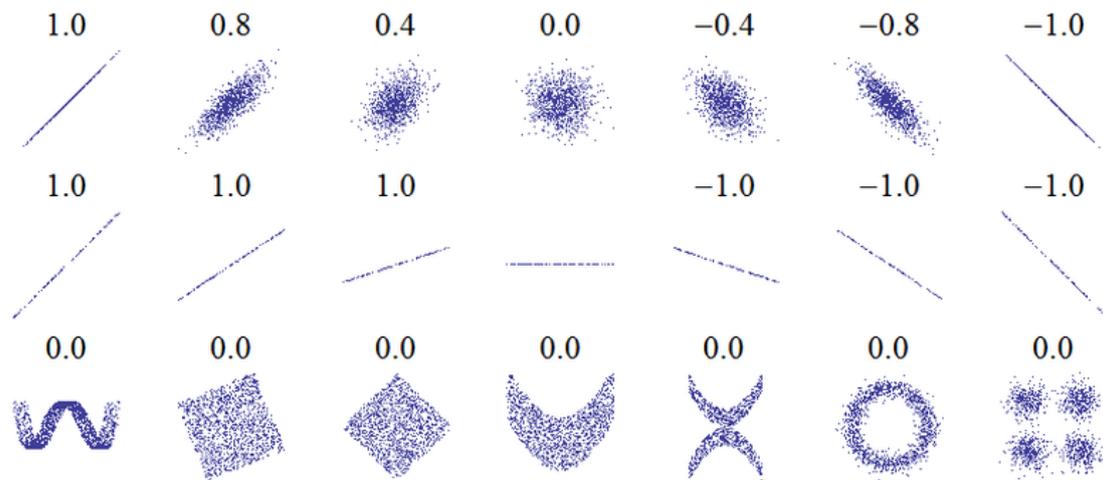
The parameters of a probabilistic investigation are not always independent of each other. If there is a strong mathematical relation between two parameters, then they can be represented by one single parameter by regarding the relation as part of the model. Often, however, parameters are statistically dependent, which means that a specific tendency of one parameter causes a tendency of another one.

Two statistical parameters  $X$  and  $Y$  are called correlated if their correlation coefficient

$$\rho_{X,Y} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}, \quad (2.2)$$

is different to 0, otherwise they are uncorrelated. In the above formula  $\text{cov}(X, Y)$  means the covariance of  $X$  and  $Y$  and  $E(\cdot)$  the expectation value. The correlation coefficient is always between -1 and 1, the sign indicating the direction of correlation. Absolute values of 1 mean a strong linear or inverse linear correlation. If the correlation coefficient is calculated for the rank-transformed values, which means that each value is mapped to its number in a size-ordered list, it is called rank correlation coefficient.

Parameters can be statistically dependent even if they are uncorrelated, i. e. their correlation coefficient is 0. Fig. 2.3 shows different statistical dependencies of two parameters. Note that the distributions in the lower line are all uncorrelated.



**Fig. 2.3** Possible statistical dependencies of two parameters with their correlation coefficients. For the case presented in the middle, the correlation coefficient is undefined, since there is no variation in  $Y$  (from Wikipedia: [https://en.wikipedia.org/wiki/Correlation\\_and\\_dependence](https://en.wikipedia.org/wiki/Correlation_and_dependence)).

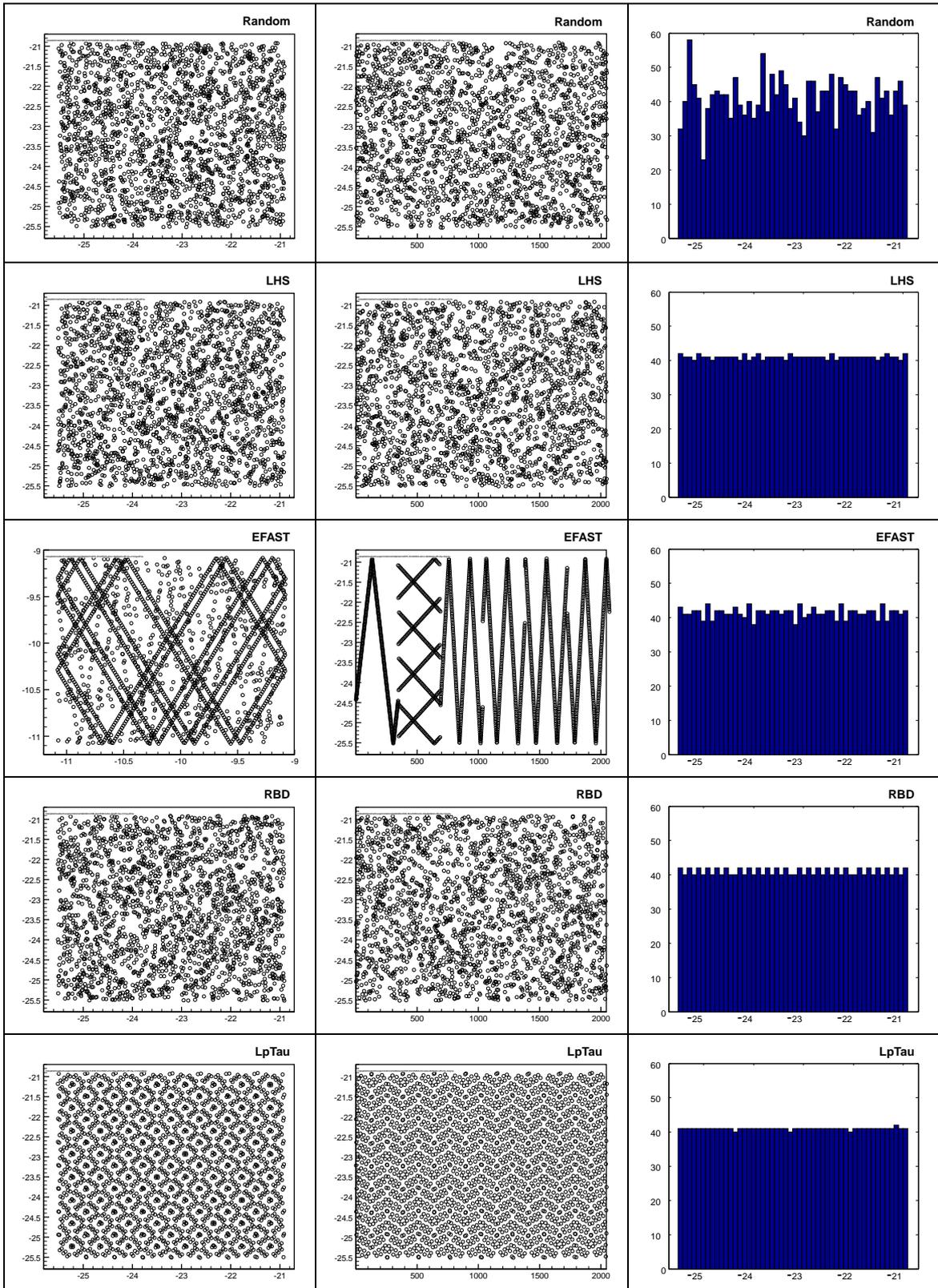
In reality, input parameters of a PA model are often statistically dependent on each other in some manner. This can be a sign for lack of understanding of the physical details, but has nevertheless to be taken into account, which is normally done by assuming a required correlation matrix. Calculating the correlation coefficients of all pairs of parameters should approximate this matrix as closely as possible. Therefore, the sampling procedure has to take account of the correlation matrix.

An approved algorithm for implementing a given correlation matrix into a sample is the Iman-Conover method (/IMA 82, MIL 05/). This method preserves the marginal distributions and works with any sampling scheme, except schemes that rely on specific paths through the multi-dimensional parameter space (like periodic sampling). The drawn parameter values remain unchanged, only their pairing is affected by the algorithm. The Iman-Conover method is designed to generate a sample that has the same rank correlation matrix as the desired, correlated reference distribution. Usually, the linear correlation matrix is then at least approximately equal to the desired one.

### 2.3.6 Comparison of sampling methods

Five sampling methods have been used to produce the results discussed in the report: Random sampling, LHS, EFAST sampling, RBD sampling and Sobol' LpTau sequenc-

es. Fig. 2.4 shows a graphical comparison of these five methods on the basis of two uniformly distributed, uncorrelated parameters  $p_1$  and  $p_2$ . Three diagrams have been made for each sampling method: a frequency histogram of  $p_1$ , a scatterplot of  $p_1$  vs.  $p_2$  and a scatterplot of  $p_1$  vs. simulation number.



**Fig. 2.4** Comparison of sampling methods by means of three diagrams:  $p_1$  vs.  $p_2$  (left),  $p_1$  vs. number of simulation (middle),  $p_1$  frequency histogram (right)

On the first sight, random sampling, LHS and RBD sampling look random, while there are clear regular structures in EFAST and LpTau sampling. Obviously, EFAST sampling produces the least homogeneous coverage of the two-dimensional parameter space, leaving large white regions. RBD sampling seems to be superior to EFAST sampling in view of one- and two-dimensional homogeneity. LpTau provides the most, random sampling the least homogeneous p1 marginal distribution. Although the regular patterns in the LpTau scatterplots seem to indicate systematically under- and overrepresented regions in the two-dimensional parameter space, in fact this sample provides the most homogeneous coverage, as the white regions are smaller than with any other method.

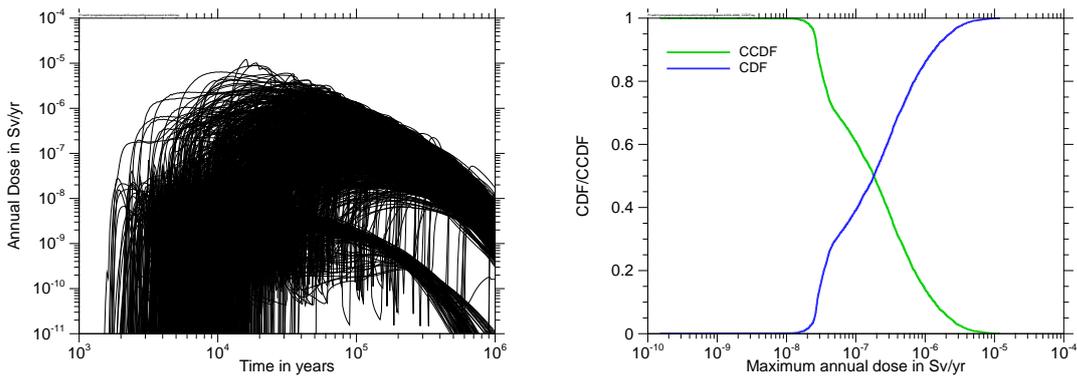
## **2.4 Uncertainty analysis**

Although the subject of this report is sensitivity analysis, this chapter is dedicated to uncertainty analysis, since it is an essential part of each probabilistic analysis.

In the context of this report, uncertainty analysis means quantifying the overall uncertainty of the results of a PA model calculation under the given uncertainties of the model input parameters. Proper quantification of the input uncertainties, i. e. establishing the correct pdfs, is a pre-condition for obtaining reliable results and a demanding task, but not a part of the uncertainty analysis itself. It is always assumed that the utilised sample represents the actual uncertainty of all input parameters. Under this precondition, it does not matter how many input parameters are taken into account, since, in contrast to sensitivity analysis, uncertainty analysis does not establish any references from the model output to the model input. The only question of interest is the variation of the model output itself.

### **2.4.1 Graphical uncertainty analysis**

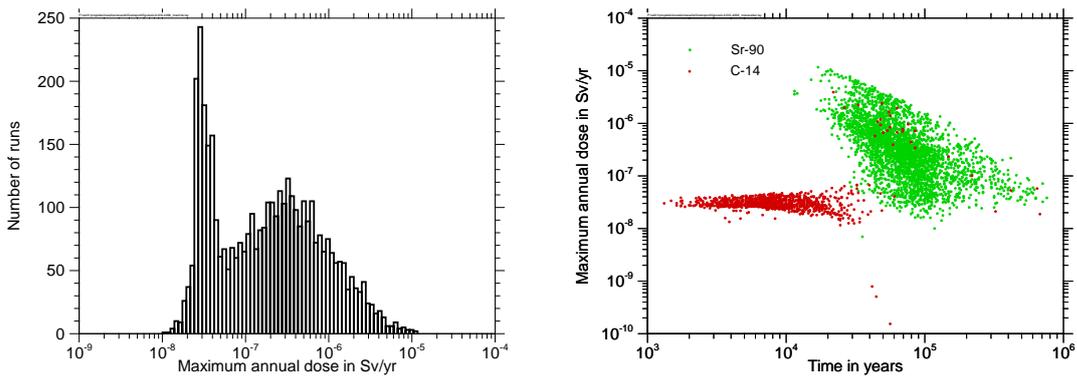
We assume that we have a time-dependent model, so that each model run produces a time curve of some model output quantity of interest. The overall uncertainty of the model results can be visualised by plotting all time curves in one single plot. Such a figure is not adequate to show or even compare the model behaviour in individual runs, but can give a graphic impression of the uncertainty as it shows the fluctuation ranges and can draw attention to extreme or outlying runs. An example with 1000 curves is shown in Fig. 2.5.



**Fig. 2.5** Graphical uncertainty analysis: plot of all runs (left) and CDF/CCDF (right)

Histograms are another possibility of illustrating the uncertainty of calculation results graphically. For a histogram the output values for a specific point in time or for the absolute maxima are grouped in an adequate number of sub-intervals that are (normally) equidistant on a linear or logarithmic scale, each interval (“bin”) is plotted as a column with a height proportional to the number of values in the interval. If the number of bins is selected properly, the shape of the envelope qualitatively approximates the (unknown) pdf of the distribution. In order to produce a more unique illustration one can simply count the values that do not exceed a given threshold  $x$ , normalise it to the total number of values and plot it versus  $x$ . Such a curve always starts at 0 at the lower end of the interval and ends at 1 at the upper end. This plot is called cumulated density function (CDF). It is often preferred to present it in its complementary form (CCDF), starting at 1 and ending at 0. An example is shown in Fig. 2.5.

Scatterplots can be used to graphically illustrate any two-dimensional distribution. If the run maxima of a set of calculations are to be evaluated, it is helpful to plot the values of the maxima versus the times of occurrence. Such a maximum scatterplot gives a visual impression not only of the distribution of the maximum values, but also of their dependency on time. Plots of this kind can be specifically helpful if an additional piece of information, for example the radionuclide responsible for the maximum, is coded by colouring the scatter points, see Fig. 2.6 (in the figure, the colours indicate the second most relevant radionuclide as in the used dataset the most relevant is always Sn-126).



**Fig. 2.6** Graphical uncertainty analysis: histogram (left) and colour-coded maximum scatterplot (right)

### 2.4.2 Numerical uncertainty analysis

For a numerical uncertainty analysis one can calculate various statistical measures for selected points in time. If sufficient supporting points are selected, the measures themselves can be plotted as time curves. In the context of repository performance assessment, however, the quantity under consideration, e. g. a safety indicator like the individual dose rate (/BEC 09a/), is normally supposed to be as low as possible. Therefore, it can be interesting to determine the absolute maximum of each run and to analyse these values instead of the values at a fixed point in time.

In the following a short overview of the most common uncertainty measures is given. For the details we refer to standard statistics literature, good introductions and references can be found on Wikipedia (<http://en.wikipedia.org>) under the relevant keywords.

Statistically distributed values, represented by a statistic variable  $X$ , are characterised by their **expectation**  $E(X) = \mu$ . For discrete values this is identical to the **arithmetic mean** of the values, which gives a rough impression of the order of magnitude of the values, but does not say anything about their distribution. For widely spread values, which are typically produced by PA models, the mean is dominated by the highest values.

Some statistical measures that can provide valuable information about the distribution are based on the central moments

$$\mu_k = E((X - \mu)^k). \quad (2.3)$$

Obviously, the first central moment is always 0. The second central moment is called **variance**:

$$\text{Var}(X) = E((X - \mu)^2). \quad (2.4)$$

The square root of the variance is called **standard deviation**. This formula is valid for calculation of the standard variation of a given distribution. The **sample standard deviation** of the values is an estimator for the standard deviation of the true distribution, calculated from a limited sample. Since the sample can never be complete and the true expectation of the distribution is unknown, the corrected formula

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (2.5)$$

should be used for calculation, which can be proven to be an unbiased estimator. The standard deviation is a numerical measure for uncertainty, but it should be kept in mind that for non-normal distributions it might say little about the real variation of the data.

The **median**  $m$  of a distribution marks its middle, i. e. that value in the parameter interval at which the integral of the pdf reaches 1/2:

$$\int_{-\infty}^m \text{pdf}(x) dx = \frac{1}{2}. \quad (2.6)$$

A finite set of data is separated by the median in two equally populated parts. Compared to the arithmetic mean, the median is less sensitive to the influences of extreme values and outliers and is often more significant for comparing a single value with the entirety of all data.

As a generalisation of the same concept, a **quantile**  $q$  is the value that separates the distribution in two parts at a specific fraction:

$$\int_{-\infty}^q \text{pdf}(x) dx = r, \quad 0 < r < 1. \quad (2.7)$$

If  $r$  is given as a percentage, the quantile is sometimes called the  $r^{\text{th}}$  **percentile** of the distribution. The median is the same as the 50%-quantile or the 50<sup>th</sup> percentile.

## 2.5 Graphical methods for sensitivity analysis

Graphical methods of sensitivity analysis provide a good means for getting a quick visual impression of the sensitivities of a system. In many respects, such methods are more appropriate to convey the basic messages about the sensitivity. Moreover, they can visualise aspects that cannot be derived from a numerical sensitivity measure. Sensitivity graphs are easy to understand and in most cases, and they can be made with low computational effort. In this chapter, several types of graphs are explained.

### 2.5.1 Scatterplots

A scatterplot is a plot of unconnected data points on a 2D-area. This is an appropriate means to present statistically distributed data. A scatterplot of the model output versus the value of a specific parameter provides a direct view of how the parameter influences the result. Such a plot can convey detailed information that cannot be derived from other types of plots. Moreover, it is possible to colour the points according to some additional piece of information, which can reveal specific interesting relationships. On the other hand, scatterplots are inadequate or unclear for investigating and comparing several parameters in one plot, or for showing the time-evolution of the system sensitivities.

### 2.5.2 Contribution to the sample mean (CSM)

The CSM plot (/SIN 93/, /BOL 09/) is a specific kind of 2D-plot that allows a quick assessment and comparison of the sensitivities of a system to variation of different parameters. For each parameter to be presented a curve, starting at (0,0) and ending at (1,1), is plotted on a unit square according to the following instructions:

From a probabilistic set of calculation runs the model output values for a specific point in time or the run maximum are paired with the values of the parameter under investigation. These pairs are sorted according to increasing parameter values. Then the cumulated contribution to the sample mean is calculated for each pair. That means that for any parameter value the model output values are added up to the relevant pair and divided by the number of runs as well as by the arithmetic mean of all values. This procedure automatically yields 1 for the last pair, which is defined by the highest parameter value. The CSM is plotted versus the fraction of the number of runs.

If the model output does not depend at all on the parameter under investigation, the cumulated sum of output values will increase by comparable amounts all over the interval, which results in a curve very close to the diagonal. If, however, low parameter values predominantly cause low model output values, the cumulated sum increases slowly in the lower part of the diagram and more steeply in the end. Therefore, a strong sensitivity is indicated by a CSM curve that deviates significantly from the diagonal. Actually, however, not the deviation from the diagonal is the primary indication of sensitivity, but the curvature. Negative curvature means an inverse influence of the parameter on the model output, i. e. increasing parameter values cause decreasing model output. If the influence is non-monotonic the CSM curve has an inflection point.

CSM plots can easily be generated and are appropriate for comparing the influences of several parameters, or the time development. They also allow spotting parameter ranges of higher or lower sensitivity by analysing the curvature.

### **2.5.3 Contribution to the sample variance (CSV)**

The idea of CSV is the same as that of CSM, but using the variance instead of the mean. Since the variance is likewise a linear measure, it can be split in the same way as the mean. The message of a CSV curve is similar to that of CSM, but less direct and mathematically harder to understand. CSV curves are often more pronounced, but less smooth than CSM curves.

### **2.5.4 Cumulated sums of normalised re-ordered output (CUSUNORO)**

CSM does not work properly with non-positive model output and not at all if the mean is zero. Moreover, it is not invariant against a constant offset. The CUSUNORO plot

/PLI 12/ is an enhancement of the CSM plot, which avoids these drawbacks. It is applicable on any data set and yields identical curves for data sets that only differ by a constant.

Instead of the sum of the output values themselves, CUSUNORO uses the sum of their deviations from the mean, normalised to the square root of the sum of squares of these differences. The result is a curve that starts at (0,0) and ends at (1,0). High sensitivity is indicated by a significant deviation of the curve from the x-axis. Like in the CSM plot the degree of sensitivity is actually indicated by the curvature.

### **2.5.5 Conditional cobweb plots and mean rank plots**

The following methods are not applied in this report but can provide a quick graphical visualisation of sensitivities and are therefore shortly mentioned here for completeness.

An often seen kind of plot is based on the idea to display the value ranges of all input parameters and the model output on parallel vertical axes and to represent each individual model run by a line connecting the associated values. If only a specific selection of runs is plotted, for instance those 10 % with the highest model output values, one can directly see which parameter values are preferably associated with this condition. Such a plot is called a conditional cobweb plot. The effects are best visualised with no more than a few hundred lines.

The mean rank plot (/COR 12/) provides basically the same message, but avoids the irritating multiplicity of intersecting lines. First, all input parameters and the model output are ranked. The parameter ranks are shown side by side on axes in the y-direction. As in the case of a cobweb plot the x-axis has no meaning but can be used for grouping the parameters according to their meaning (e. g., near field – far field – biosphere parameters). The model runs are then divided into classes according to some property, for instance on the basis of the model output ranks. The classes are represented as lines in individual colours, showing the mean rank for each parameter. The lines themselves are unnecessary but can help focus the view. For a parameter with low sensitivity the rank values will be more or less uniquely distributed over the total range, regardless of the output rank, and all lines will intersect the respective axis near the middle. On the other hand, a broad distribution of the intersecting points is a hint to high sensitivity (/KUH 15/).

## 2.6 Correlation- and regression-based methods

For numerical evaluation of the sensitivities of a numerical model on the basis of a probabilistic set of model runs, a variety of different methods have been developed. A classical and often applied approach is to approximate the influence of a parameter to the model output as well as possible by a linear relationship and to derive a measure of sensitivity from this approximation. Due to their nature, such methods work best if the actual model behaviour is close to linear, with respect to the parameter under consideration.

There are two different mathematical approaches to this idea, one of which is based on correlation, the other on regression.

### 2.6.1 Pearson's correlation coefficients

For two statistical variables,  $X$  and  $Y$ , with  $n$  realisations  $x_i, y_i, i = 1 \dots n$ , the empirical linear correlation coefficient  $\rho(X, Y)$ , or Pearson's correlation coefficient, can be calculated as

$$\rho(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 \sum_{i=1}^n (y_i - \bar{Y})^2}}, \quad (2.8)$$

where  $\bar{X}$  and  $\bar{Y}$  denote the arithmetic means of  $X$  and  $Y$ . In case of a strong positive or negative linear relationship between  $X$  and  $Y$  the correlation coefficient is +1 or -1, respectively, regardless of the gradient. An absolute value below 1 indicates a statistical tendency, which is the less pronounced the closer the value is to 0. As, however, the correlation coefficient requires a linear relationship, a value of 0 does not exclude a non-linear statistical relationship.

For a model with a number of input parameters  $X_k$  and the model output  $Y$ , the correlation coefficients  $\rho_k = \rho(X_k, Y)$  provide a measure of sensitivity, which is the more reliable the closer to linear the model behaves. It should be kept in mind, however, that a high absolute value does not mean that a small variation of parameter  $X_k$  causes a high change of the model output. It is rather an indication that the parameter in question dominates the influences on the model output, not allowing for much scatter due to other influences.

## 2.6.2 Standardised regression coefficients (SRC)

Linear regression is based on the attempt to approximate the relationship between the statistical Variables  $X$  and  $Y$  as well as possible by a linear function:

$$\hat{Y} = b_0 + \sum_{j=1}^k b_j X_j . \quad (2.9)$$

For every set of parameter values  $x_{ji}$ ,  $i = 1 \dots n$ , an absolute error  $\varepsilon_i$  results:

$$y_i = b_0 + \sum_{j=1}^k b_j x_{ji} + \varepsilon_i = \hat{y}_i + \varepsilon_i . \quad (2.10)$$

The coefficients  $b_j$ ,  $j = 1 \dots k$ , can be determined to minimise the total error according to the least squares:

$$\sum_{i=1}^n \varepsilon_i^2 = \min . \quad (2.11)$$

Then the coefficients  $b_1 \dots b_k$  are a measure of the sensitivity of the model output  $Y$  against variations of the model input parameters  $X_1 \dots X_k$ . In order to allow for a unique assessment of these coefficients, they are transformed such that they get the expectation 0 and the standard deviation 1:

$$\frac{\hat{Y} - \bar{Y}}{\hat{s}} = \sum_{j=1}^k \frac{b_j \hat{s}_j}{\hat{s}} \cdot \frac{X_j - \bar{X}_j}{\hat{s}_j} \quad (2.12)$$

with

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i , \quad \bar{X} = \frac{1}{n} \sum_{i=1}^n x_{ji} \quad (2.13)$$

and

$$\hat{s} = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - \bar{Y})^2}{n-1}} , \quad \hat{s}_j = \sqrt{\sum_{i=1}^n \frac{(x_{ji} - \bar{X}_j)^2}{n-1}} . \quad (2.14)$$

The coefficients  $(b_j \hat{s}_j) / \hat{s}$  are called standardised regression coefficients (SRC). Their values are always between -1 and 1, negative values indicating inverse dependency.

Although in practice, the SRC values are often similar to Pearson's correlation coefficients, their mathematical meaning is different. While the correlation coefficient indicates, how much the model is dominated by the parameter, a high SRC means a high gradient of the regression line.

The coefficient of model determination is defined as

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{Y})^2}{\sum_{i=1}^n (y_i - \bar{Y})^2} . \quad (2.15)$$

It represents the fraction of the variation of the model output that can be declared by linear regression. That means,  $R^2 = 1$  for a strong linear relationship and  $R^2 = 0$  for a model that cannot at all be declared by a linear regression (i.e., the best linear regression is a simple constant).  $R^2$  is equal to the squared multiple correlation coefficient of the output with all input variables. It provides a measure for the degree of linearity of the model under consideration. As a rule of thumb,  $R^2$  should be at least 0.5 to allow for a meaningful sensitivity analysis with linear methods.

### 2.6.3 Partial correlation coefficients (PCC)

If the input parameters  $X_1 \dots X_k$ , or some of them, are correlated among themselves, which may have been brought about on purpose due to some assumed statistical physical coupling, there is a common influence to the model output  $Y$ . The sensitivity analysis methods described so far cannot resolve this coupling but describe the total influence of an input parameter to the output, including the indirect influence via other parameters. In order to separate these influences, one can use the following regression ansatz:

$$\hat{X}_j = c_{j0} + \sum_{p=1, p \neq j}^k c_{jp} X_p , \quad \hat{Y} = b_0 + \sum_{p=1, p \neq j}^k b_p X_p . \quad (2.16)$$

This describes the influences of all input parameters different from  $X_j$  to  $X_j$  as well as to the model output  $Y$ . The partial correlation coefficient (PCC) of  $X_j$  and  $Y$  is defined as

the correlation coefficient of the errors  $(X_j - \hat{X}_j)$  and  $(Y - \hat{Y})$ . It can be seen as a measure for the sensitivity of the model output to the input parameter  $X_j$  alone, regardless of indirect influences via other parameters. It can be shown that these coefficients are identical to the SRCs if the parameters are uncorrelated.

#### 2.6.4 Rank transformation

As mentioned, correlation- and regression-based methods for sensitivity analysis are founded on the assumption that there is a close-to-linear relationship between the input parameters and the model output. For complex models, this is often not the case, which can be seen by a low value of  $R^2$ . In many cases, however, the relationship is at least monotonic. A monotonic relationship can be transformed into a linear one by performing a rank transformation. This is done by sorting the values of each input parameter and the model output according to their size and replacing each value by its rank in the respective list. If calculated for the ranks,  $R^2$  often – but not always! – increases considerably compared to its value calculated from the original values.

All sensitivity measures explained in this chapter can as well be calculated on the basis of the ranks instead of the values. On the one hand, the rank-based measures are often more significant and allow for a more unique sensitivity ranking of the input parameters. On the other hand, the rank transformation annihilates details of the model behaviour and therefore causes some loss of quantitative meaning of the results.

Tab. 2.1 shows the corresponding value- and rank-based sensitivity measures with their naming and abbreviations used in this report.

**Tab. 2.1** Naming of correlation- and regression-based methods of sensitivity analysis

Principle/Method/Concept	Value-based	Rank-based
Correlation	Pearson's correlation coefficients (PEAR)	Spearman's rank correlation coefficients (SPEA)
Regression	Standardised regression coefficients (SRC)	Standardised rank correlation coefficients (SRRC)
Partial correlation	Partial correlation coefficients (PCC)	Partial rank correlation coefficients (PRCC)

## 2.7 Variance-based methods

The variance of a statistically distributed variable  $X$  is (cf. eq. (2.4))

$$\text{Var}(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2. \quad (2.17)$$

Based on the variance, a sensitivity measure can be defined that reflects the global sensitivity quantitatively, regardless of the linearity or non-linearity of the model:

$$S_j = \frac{\text{Var}_{X_j}[E(Y|X_j = \text{const.})]}{\text{Var}(Y)}. \quad (2.18)$$

In this notation,  $E(Y|X_j = \text{const.})$  means the expectation of the model output  $Y$  under the side condition that the input parameter  $X_j$  is kept constant. This expectation value varies with  $X_j$ , and the variance of its distribution is divided by the total variance of  $Y$ . The values  $S_j$  are called first-order sensitivity indices.

It can be shown that, as long as the input parameters  $X_i$  are independent of each other, the model output  $Y = f(X_1, \dots, X_k)$  can be uniquely decomposed in terms of increasing dimensionality:

$$f(X_1, \dots, X_k) = f_0 + \sum_{j=1}^k f_j(X_j) + \sum_{j_1 < j_2}^k f_{j_1 j_2}(X_{j_1}, X_{j_2}) + \dots + f_{12\dots k}(X_1, \dots, X_k) \quad (2.19)$$

with functions that depend only on the parameters in their index:

$$f_0 = E(Y) ,$$

$$f_j(X_j) = E(Y|X_j = \text{const.}) - f_0 , \quad (2.20)$$

$$f_{j_1 j_2}(X_{j_1}, X_{j_2}) = E(Y|X_{j_1} = \text{const.}, X_{j_2} = \text{const.}) - f_{j_1}(X_{j_1}) - f_{j_2}(X_{j_2}) - f_0$$

and so on. This is denoted as high-dimensional model representation (HDMR). The total variance of the model output  $Y$  can be decomposed uniquely as

$$\text{Var}(Y) = \sum_{j_1=1}^k V_{j_1} + \sum_{j_1 < j_2}^k V_{j_1 j_2} + \dots + V_{12\dots k} \quad (2.21)$$

where

$$V_{j_1} = \text{Var}_{X_{j_1}} [E(Y|X_{j_1} = \text{const.})] , \quad (2.22)$$

$$V_{j_1 j_2} = \text{Var}_{X_{j_1}, X_{j_2}} [E(Y|X_{j_1} = \text{const.}, X_{j_2} = \text{const.})] - V_{j_1} - V_{j_2}$$

and so on. Higher-order sensitivity indices  $S_{j_1 j_2}$  (and so on) are defined as the ratios of the respective variance components to the total variance of  $Y$ , so that

$$\sum_{j_1=1}^k S_{j_1} + \sum_{j_1 < j_2}^k S_{j_1 j_2} + \dots + S_{12\dots k} = 1 . \quad (2.23)$$

While the first-order indices (SI1) are a measure for the sensitivity of the model to one input parameter alone, the higher-order indices characterise the combined influence of two or more parameters.

While it is often difficult to calculate the individual higher-order indices, there are some methods for determining what is called the total-order sensitivity indices (TSI). For the model parameter  $X_j$  the TSI is defined as the sum of all first- and higher-order sensitivity indices that contain the index  $j$ . It is a measure for the total influence of the parameter on the model output, alone and via any interactions.

The calculation of the variance-based sensitivity indices from a set of Monte-Carlo-runs is not a straightforward task, but a number of computational methods have been developed for this purpose. These methods yield more or less accurate approximations of

the first-, higher- or total-order sensitivity indices. Some methods need specific sampling schemes. In the following, some methods are shortly described without going into the details.

### **2.7.1 The Sobol' method**

The method originally proposed by Sobol' (/SOB 93/) and further developed by others (/SOB 01/, /SAL 02/) calculates the sensitivity indices using specific estimators. For this purpose, two independent samples of size  $n$  are used as a starting point. A number of combined sample matrices are constructed, each by replacing a single column of the first sample with the respective column of the second one. The model has to be evaluated for each combination. Therefore, the number of necessary model runs increases essentially with the number of parameters under consideration.

The Sobol' method is the most direct way to calculate the sensitivity indices of any order as well as the TSIs. It can handle parameter correlations without problems. On the other hand, due to the high number of model runs, the method is very computationally expensive.

### **2.7.2 (Extended) Fourier Amplitude Sensitivity Test (FAST/EFAST)**

FAST (/CUK 78/, /SAL 97, /SAL 98/) has been developed in order to allow the calculation of first-order sensitivity indices with essentially reduced computational effort. The general idea is the following: instead of using a randomly drawn or otherwise constructed sample, the  $k$ -dimensional parameter space is scanned using periodical functions. The periodicity can be rediscovered in the model output, the more pronounced the higher the sensitivity is. If each input parameter is assigned an individual frequency, one can distinguish the influences by performing a Fourier analysis of the model output. It can be shown that this procedure yields the sensitivity indices. Since the model is, in general, non-linear, each frequency has to be analysed together with a sufficient number of harmonics. Therefore, the frequencies have to be selected thoroughly to avoid overlap of main harmonics. The highest frequency that has to be considered determines the minimal number of model runs to be executed, because the maximum frequency that can be identified in a discrete sample of  $n$  values is  $(n - 1)/2$ .

A variant of the standard FAST method is called Extended FAST (EFAST, /SAL 97/). This method works with only two different basic frequencies. The higher one is assigned to one selected parameter, the lower frequency is used for all other parameters. The Fourier analysis of the model output then yields two numbers: the first reflects the influence of the selected parameter alone and the second measures the common influence of all parameters except the selected one. The first number is an estimator for the SI1 of the selected parameter, and the second evaluates the common influence of all parameters different from the selected one. If subtracted from 1, this yields an estimator for the TSI. By this procedure, SI1 and TSI are determined in one step, but only for one of the parameters. Therefore, each parameter is considered in an own investigation with a number of model runs that results from dividing the intended total sample size by the number of parameters. The EFAST sampling scheme takes account of this by automatically assigning the high frequency successively to each parameter for a number of runs.

Principally, it would be possible to calculate second- or higher-order sensitivity indices with FAST. That would require a common frequency for all parameters whose interaction is to be analysed.

### 2.7.3 Random Balance Design (RBD)

A disadvantage of FAST/EFAST is that the sampling yields a rather inhomogeneous coverage of the parameter space. To avoid this problem, and nevertheless use a Fourier-based technique, the Random Balance Design (RBD) method has been invented based on the idea to use only one frequency, but to hide the periodicity for all variables except one by shuffling the runs appropriately (/TAR 06/). This is realised in the following way: For an investigation with  $n$  runs one starts with an empty sample matrix of  $k$  lines and  $n$  columns. A random permutation  $\pi_1$  of the lines is applied; then the values of the variable  $x_1$  are calculated according to a periodic scheme and inserted in the first column. If the permutation is reversed, the values look randomly distributed. In the next step the second column is filled with values for the variable  $x_2$  in the same way, using a different permutation  $\pi_2$  and so forth. The same frequency (normally 1) is used for all variables. As a result one gets a sample that covers the parameter space satisfyingly and does not show visual periodicity. Nevertheless, if any of the permutations  $\pi_1 \dots \pi_k$  is applied to the model output, the periodicity of the respective input parameter will show up the more distinct the higher the model sensitivity is. All other variables, how-

ever, remain shuffled and therefore have only a random influence to the model output. The first-order sensitivity index can be computed via Discrete Fourier Transformation (DFT).

RBD calculates all first-order indices from one set of model runs. Higher- or total-order indices, however, cannot be calculated. Although the sample is generally more homogeneous than a FAST/EFAST sample, a specific sampling is still needed.

#### **2.7.4 Effective Algorithm for computing global Sensitivity Indices (EASI)**

The idea of EASI (/PLI 10/) is basically the same as that applied in RBD, but with an essential difference. While RBD uses random permutations to generate the sample from periodical data, EASI goes the opposite way. It uses a randomly or otherwise drawn sample and determines the needed permutations such that a periodicity occurs in the values of one variable in each case. These permutations, successively applied to the model output, bring about the same periodicity, so that it can be analysed by DFT. There is a small noise superposed to the pure periodicity, but as long as the frequency spectrum of this noise is randomly distributed, it has a negligible influence to the Fourier analysis.

Like RBD, EASI calculates the first-order sensitivity indices. The essential advantage, however, is that nearly any sample can be used, for instance a random sample, a stratified sample or a quasi-Monte-Carlo sample. It should only be kept in mind that a systematic coupling of parameters can distort the Fourier evaluation by inducing unwanted periodicity. This freedom in sampling strategy is a specific feature of EASI in comparison with all other variance-based techniques mentioned so far. It allows not only for selecting an optimal sampling strategy but also for re-using existing model runs as a part of an investigation with a higher number of runs.

#### **2.7.5 Metamodelling and State-Dependent Parameter method (SDP)**

The approach shortly described in this section is based on the idea of metamodelling. This means that the original computational model, which calculates a variety of effects and solves differential equations, is replaced by a simple, computationally cheap meta-model that simulates the behaviour of the original model as well as possible. The sensitivities are then calculated from the metamodel. The SDP approach is an integrated

technique, which generates and mathematically analyses the metamodel in one step (/RAT 07/).

While the approaches described so far make direct use of the variance decomposition according to eq. (2.21), the SDP method estimates the first-order components  $f_j(X_j)$  of the HDMR (eq. (2.19)) using recursive filtering and Fixed Interval Smoothing (FIS) algorithms to fit SDP models to the input-output mapping. It is assumed that the effects of all terms of order higher than 1 can be estimated by a Gaussian white noise with mean 0. Then the variances of the functions  $f_j$  are calculated to estimate the main effects.

The SDP approach is established upon comparing the HDMR representation with a State-Dependent Regression (SDR) approach. The 'time dependency' of the SDR model represents the sequence of the Monte Carlo simulations of the computational model. For the estimation of the SDP it is essential to map the variability of these parameters in some stochastic way: it is assumed that the evolution of each SDP can be described by one type of a Generalized Random Walk (GRW) process on non-stationary random sequences, for example with the integrated random walk (IRW) process. It was observed that this process generates good results as it enforces the smooth properties of a cubic spline upon the estimated SDP relationship. If the IRW process is used, the SDPs can be approximated. The first-order sensitivity indices can then be estimated from the different smoothed estimates of the HDMR terms.

In contrast to many classical variance-based methods, the SDP method can be applied using any sample.

## **2.8 Non-parametric methods**

In contrast to parametric statistics, non-parametric statistics do not make use of parametrised probability distributions. The parameters of non-parametric statistical models are not defined a priori and fitted to the data, but derived from the data directly.

Some methods of sensitivity analysis are based on non-parametric approaches.

### 2.8.1 Two-sample tests

Two-sample tests compare the parameter distributions of two independent samples. The null hypothesis is that both belong to the same population, i.e. they are equally distributed. The null hypothesis is tested against the alternative hypothesis that the distributions of both samples differ significantly.

In order to assess the sensitivity of a model to its input parameters, one can separate the total sample of parameters and model output values in two sub-samples according to the output value. For instance, the 90 %-quantile of the output distribution can be used as the sample separation criterion, which means that those 10 percent of the model runs that produce the highest output values are taken as one sample and the remaining 90 percent as the second sample. The distributions of the corresponding values of one specific input parameter are then compared using some statistical test. If a significant difference between these distributions is found, it is obvious that the model is sensitive to the parameter under investigation.

In the case of the Smirnov test (SMIR) the null hypothesis of equal distributions of both samples is rejected if the maximum difference of the empirical distributions exceeds a certain quantile of a test distribution. Other statistical tests that can be used are, e.g., the Mann-Whitney-U-test or the Cramér-von Mises test. Although such a test yields, strictly speaking, a yes/no answer, in practice there is some quantitative result, which allows a ranking of the parameters. Such rankings, however, can considerably differ from those calculated with other SA methods and it should be kept in mind how they have been established.

### **3 Description of the test systems**

The purpose of the project described in this report was to investigate different methods of sensitivity analysis applied to typical final repository system models with a variety of inherent parameter uncertainties. According to experiences with numerical long-term safety assessment, such models can exhibit a number of specific properties that may cause problems in the context of sensitivity analysis. In order to encompass a wide variety of such particularities, three different model systems were used. Two of these systems are fully generic, but based on experiences with realistic systems or concepts. The third system is based on an obsolete concept for disposal of SF and HLW in rock salt. In the following, the main features and numerical properties of the test systems are described.

#### **3.1 Repository for Spent Fuel and HLW in clay**

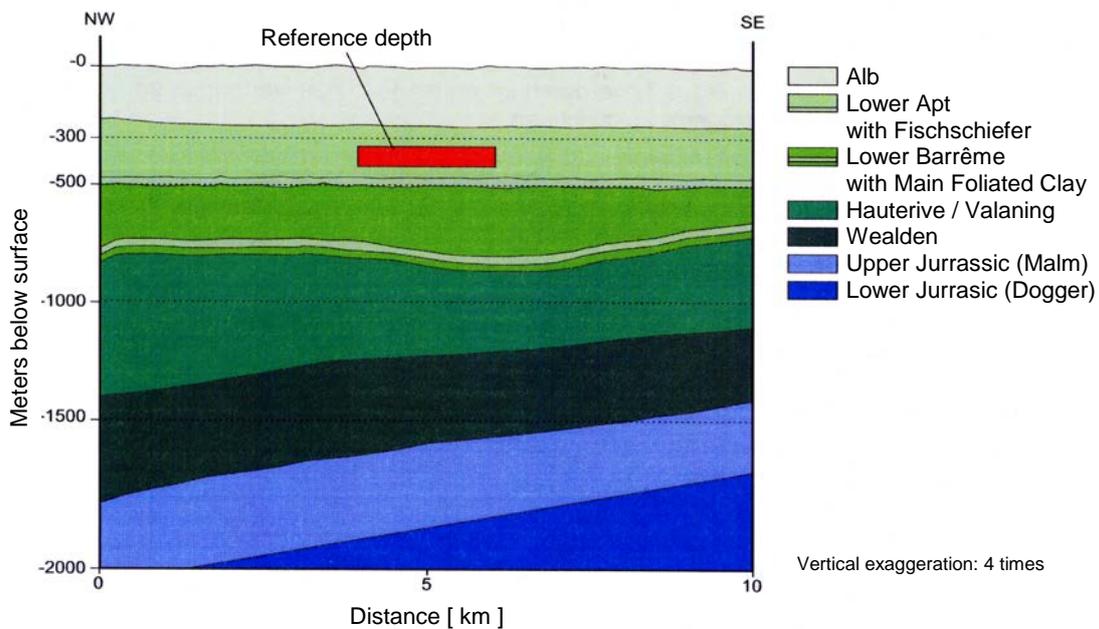
This model system describes a generic repository in a Northern German clay formation. It is based on considerations made in the context of the project GENESIS (/JOB 07/) and was already investigated in the project TONI (/RUE 07/). The repository is assumed to be located in the middle of the Apt layer in the Lower Cretaceous Clay in Lower Saxony, see Fig. 3.1.

The model comprises the near field with the waste containers, three clay layers (bentonite buffer, Apt and Alb) and the biosphere. The radionuclide mobilisation and the (purely diffusive) transport through the clay layers was calculated with the near field code CLAYPOS (version 3.01), the far field with the code module CHETLIN (version 4.1) and for the biosphere the code module EXMAS (version 2.1) was used.

Six parameters were selected for investigation. These are listed in Tab. 3.1.

**Tab. 3.1** Parameters of the clay system model

Parameter	Type of pdf	Range
<i>DiffClay2:</i> Diffusion constant clay formation 1	Log-uniform	8.3E-12 - 8.3E-10
<i>DiffClay3:</i> Diffusion constant clay formation 2	Log-uniform	8.3E-12 - 8.3E-10
<i>KdBent:</i> Kd value bentonite (U)	Log-uniform	4 - 400
<i>KdClay2:</i> Kd value clay formation 1 (U)	Log-uniform	2 - 200
<i>KdClay3:</i> Kd value clay formation 2 (U)	Log-uniform	2 - 200
<i>PorClay:</i> Porosity clay formation 2	Uniform	0.06 - 0.24



**Fig. 3.1** Cross-section of the model area in Northern Germany

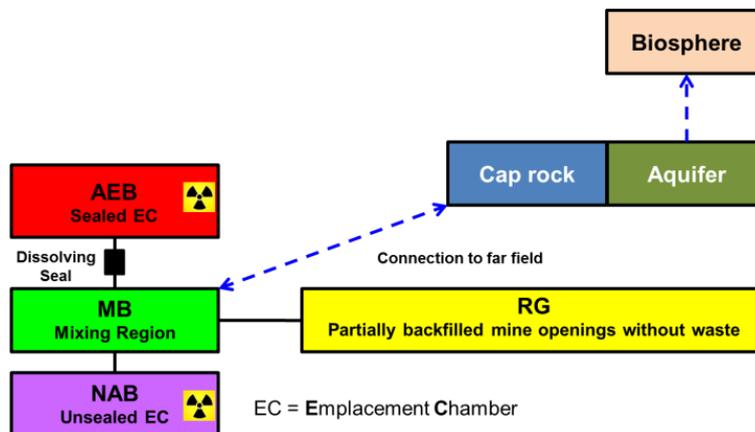
### 3.2 Repository for LILW in rock salt

The second model, which was used for most of the investigations, represents a repository for Low- and Intermediate-Level Waste (LILW) installed in an abandoned former salt production mine. This model is generic, but was motivated by a model developed for a real German site of that type. A salt mine that has not been specifically designed and built for waste disposal is typically subject to various unfavourable effects, and

measures have to be taken to avoid or mitigate unwanted processes. Therefore, the model describing such a system can become rather complex and exhibit a highly non-linear or even non-continuous behaviour that is hard to predict.

The LILW model was established as a test model for sensitivity analysis reflecting all relevant properties of the model for the real site and leaving aside only such details that are irrelevant for the numerical particularities. In Fig. 3.2 a scheme of the model structure is presented. The near field part consists of sub-models for two emplacement chambers, one of which is sealed from the rest of the mine, a mixing region and a large area of mine openings without waste. The model exit, located in the mixing region, is coupled to a far field model, which calculates the contaminant transport through a pathway in the cap rock and the aquifer, and finally to a biosphere code determining the radiation exposure of man in the form of the annual effective dose to an adult human individual.

In reality, there would be a slow intrusion of brine to the mine building, causing a gradually increasing rise of the fluid level and a time-dependent dissolution of contaminants. For simplicity reasons, however, it is assumed that the non-sealed part of the mine is filled up with brine instantaneously at some specific point in time. From then on, the brine seeps slowly into the sealed emplacement area AEB.



**Fig. 3.2** Schematic view of the components of the LILW system

The most peculiar feature of the model is the seal between AEB and MB. It represents a sealing construction made of salt concrete, which is subject to chemical dissolution (corrosion) by magnesium-containing brine. As the corrosion front is assumed to proceed slowly through the material, the seal loses its isolation capability nearly suddenly at some point in time, resulting in an essential change of the model behaviour and usu-

ally a considerable increase of the radiation exposure. The time of seal failure depends on a number of parameters, so that, analysed at a specific point in time, the model behaves nearly non-continuous with respect to some of these parameters.

The model was set up to be calculated with the RepoTREND package (/REI 16a/, /REI 17/), using the modules LOPOS, GeoTREND-SP and BioTREND.

The LILW model depends on a number of parameters, 20 of which were selected for statistical variation. These are defined in Tab. 3.2. Most investigations, however, were conducted with subsets of 11, 7 or 6 parameters while the others were fixed to their standard value. In order to distinguish between these cases, the model is addressed as LILW20, LILW11, LILW7 or LILW6 in this report.

**Tab. 3.2** Parameters of the LILW system model. The colours indicate the subsets of parameters that were used for the different investigations (LILW6 = red, LILW7 = red + green, LILW11 = red + green + blue, LILW20 = all)

Parameter	Type of pdf	Range or pdf parameters	Standard value
<i>GasEntryP</i> : Gas entry pressure	Uniform	0 - 2.5	2.0
<i>IniPermSeal</i> : Initial permeability of dissolving seal	Log-normal	$\mu=41.0605$ $\sigma=1.9809$	$1.0 \cdot 10^{-18}$
<i>RefConv</i> : Reference convergence rate	Log-uniform	$1.0 \cdot 10^{-5}$ - $1.0 \cdot 10^{-4}$	$4.0 \cdot 10^{-5}$
<i>AEBConv</i> : Factor of local convergence variation in AEB	Log-uniform	0.05 - 5.0	1.0
<i>GasCorrPE</i> : Organics corrosion rate	Log-normal	$\mu=12.6642$ $\sigma=1.1177$	$1.0 \cdot 10^{-5}$
<i>TBrine</i> : Time of brine intrusion	Log-normal	$\mu=8.8857$ $\sigma=0.6933$	7500
<i>BrineMgSat</i> : Relative magnesium saturation of brine	Triangular	0 - 0.1 - 1.0	0.1
<i>RGConv</i> : Factor of local convergence variation in RG	Log-uniform	0.25 - 2.5	1.0
<i>GasCorrFe</i> : Metal corrosion rate	Log-normal	$\mu=-6.6728$ $\sigma=1.1177$	$4.0 \cdot 10^{-3}$
<i>AEBGasProd</i> : Proportion of the material involved in gas production in AEB	Triangular	0.1 - 0.8 - 1.0	0.8
<i>NABGasProd</i> : Proportion of the material involved in gas production in NAB	Triangular	0.1 - 0.8 - 1.0	0.8
<i>RGGasProd</i> : Proportion of the material involved in gas	Triangular	0.1 - 0.8 - 1.0	0.8

production in RG			
<i>NABConv:</i> Factor of local convergence variation in NAB	Log-uniform	0.05 - 5.0	0.2
<i>MBConv:</i> Factor of local convergence variation in MB	Log-uniform	0.075 - 0.75	0.2
<i>DiffCoeff:</i> Diffusion coefficient	Log-uniform	$1.0 \cdot 10^{-10}$ - $1.0 \cdot 10^{-8}$	$1.0 \cdot 10^{-9}$
<i>RefPor:</i> Reference porosity	Triangular	0.15 - 0.3 - 0.4	0.3
<i>FacDisp:</i> Longitudinal dispersion length	Triangular	0.5 - 1.0 - 2.0	0
<i>ConvFak:</i> Variation factor for sheeting	Uniform	0 - 2.0	1.0
<i>PorDebris:</i> Porosity of debris from sheeting	Uniform	0.25 - 0.5	0.4
<i>C14Inv:</i> Variation factor for C-14 inventory	Uniform	0 - 2.5	1.0

### 3.3 Repository for Spent Fuel and HLW in rock salt (SAM-GBS)

As test model for a repository for spent fuel and high-level waste an old concept was chosen, which had been designed for the emplacement of waste containers in boreholes and drifts in a common emplacement area. It was used in the SAM project (/BUH 91/) and was already then evaluated statistically with variation of 31 parameters. These are listed in Tab. 3.3. In 2004, the model was used for a first comparison of sensitivity analysis techniques within the project ARTE (/BUH 04/). For this purpose, a new set of model runs was executed using a random sample of size 3000 and applying the EMOS modules REPOS (version 6.04), CHETLIN (version 3.01) and EXCON (version 2.02) (/STO 96/). This existing set of model results was also used for some new investigations in the context of MOSEL.

The prominent property of the SAM-GBS model is the high probability of zero output. If the waste-containing parts of the mine are separated from the rest due to convergence before brine can get in contact with the wastes, no contaminants are mobilised and a zero release is calculated. This situation, which is beneficiary for repository safety but causes problems with sensitivity analysis, is typical for repositories for heat-generating waste in rock salt. The SAM-GBS model produced about 85 % zero output runs.

**Tab. 3.3** Parameters of the SAM-GBS system model

Parameter	Type of pdf	Range or pdf parameters	Unit
<i>TBrine</i> Time of brine intrusion	Log-normal	$\mu=45$ $\sigma=1.24$	a
<i>VIncl</i> Volume of brine inclusion	Log-Hist.	11 - 1000	m <sup>3</sup>
<i>TLifeLWR</i> Container lifetime LWR-BE	Log-normal	$\mu=500$ $\sigma=0,28$	a
<i>FConvHAW</i> Factor convergence reduction (HAW)	Log-uniform	0,001 - 0,1	-
<i>FConMAW</i> Factor convergence reduction (MAW)	Log-uniform	0,1 - 1	-
<i>KRef</i> Reference convergence rate	Log-normal	$\mu=1 \cdot 10^{-2}$ $\sigma=0,92$	1/a
<i>EActSalt</i> Activation energy (salt)	Uniform	5800 - 7200	1/K
<i>ExpPP</i> Exponent in perm.-por.-relation	Log-normal	$\mu=4,5$ $\sigma=0,13$	-
<i>PermSeal</i> Seal permeability	Log-normal	$\mu=7 \cdot 10^{-16}$ $\sigma=0,78$	m <sup>2</sup>
<i>FPerm</i> Permeability increase during inflow	Log-uniform	1 - 100	-
<i>FBrineGas</i> Factor brine transport / gas transport	Log-uniform	120 - $1,2 \cdot 10^5$	-
<i>FBrineConc</i> Factor brine transp. / conc. difference	Log-uniform	$2,56 \cdot 10^9$ - $2,56 \cdot 10^{11}$	-
<i>FBrineTemp</i> Factor brine transp. / temp. difference	Log-uniform	$9,59 \cdot 10^8$ - $9,59 \cdot 10^{10}$	-
<i>CDiff</i> Diffusion coefficient	Log-normal	$\mu=6,7 \cdot 10^{-8}$ $\sigma=1,35$	m <sup>2</sup> /s
<i>ConcMg</i> Magnesium concentration in brine	Uniform	0 - 88	kg/m <sup>3</sup>
<i>RmCem</i> Mobilisation rate cemented containers	Log-normal	$\mu=1 \cdot 10^{-2}$ $\sigma=0,36$	1/a
<i>Rm1LWR</i> Mobilisation rate 1 LWR-BE	Log-normal	$\mu=1 \cdot 10^{-1}$ $\sigma=0,92$	1/a
<i>Rm2LWR</i> Mobilisation rate 2 LWR-BE	Log-normal	$\mu=1,6 \cdot 10^{-3}$ $\sigma=0,92$	1/a
<i>RmRkHTR</i> Mobilisation rate RK from HTR-BE	Log-uniform	$4 \cdot 10^{-4}$ - 0,01	1/a
<i>RmRpHTR</i> Mobilisation rate RP from HTR-BE	Log-normal	$\mu=2 \cdot 10^{-3}$ $\sigma=0,55$	1/a
<i>CLGlass</i> Mobilisation: elution coefficient glass	Log-normal	$\mu=3,6 \cdot 10^{-1}$ $\sigma=0,85$	kg/m <sup>2</sup> /a
<i>EActGlass</i> Mobilisation: activation energy glass	Log-normal	$\mu=75$ $\sigma=0,50$	kJ/mol

<i>LimSolAcid</i> Solubility limits acidic (Sr)	Log-uniform	$1 \cdot 10^{-6} - 1$	mol/l
<i>LimSolNeutr</i> Solubility limits neutral (Sr)	Log-uniform	$1 \cdot 10^{-6} - 1$	mol/l
<i>LimSolAlk</i> Solubility limits alkaline (Sr)	Log-uniform	$1 \cdot 10^{-6} - 1 \cdot 10^{-4}$	mol/l
<i>LimSolUnk</i> Solubility limits unknown (Sr)	Log-uniform	$1 \cdot 10^{-6} - 1$	mol/l
<i>KdCemSr</i> Kd-values cement (Sr)	Log-uniform	0,04 - 4	m <sup>3</sup> /kg
<i>KdCemZr</i> Kd-values cement (Zr)	Log-uniform	100 - 10000	m <sup>3</sup> /kg
<i>KdGeo</i> Kd-values geosphere	Log-uniform	$5 \cdot 10^{-4} - 5 \cdot 10^{-2}$	m <sup>3</sup> /kg
<i>FDisp</i> Variation factor for long. dispersion	Log-normal	$\mu=1$ $\sigma=0,916$	-
<i>FDil</i> Dilution by flow rate	Log-uniform	0,41 - 2,44	-



## **4 Graphical sensitivity analysis**

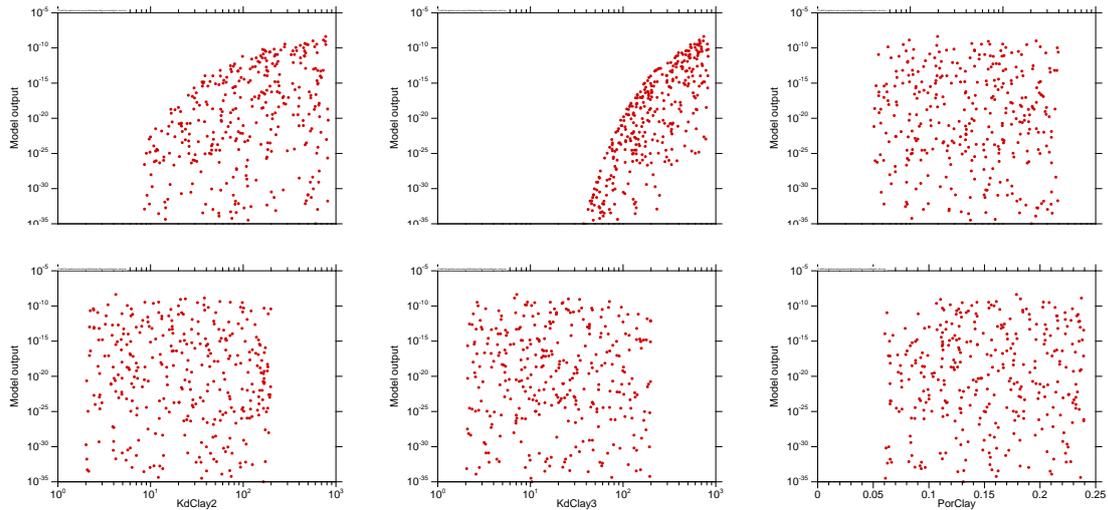
This chapter is dedicated to presenting and interpreting results achieved with graphical methods of sensitivity analysis applied to the test systems. The methods themselves have been introduced in chapter 2.5.

### **4.1 Scatterplots**

Scatterplots of the model output versus the value of one of the model parameters are easily made from the results of a probabilistic calculation and provide a direct view of the influence of a parameter to the model output, even with low sample sizes. Therefore, such plots are predestined as a starting point for sensitivity analysis. For a time-dependent model, scatterplots can either be made for a specific point in time or for the run maxima.

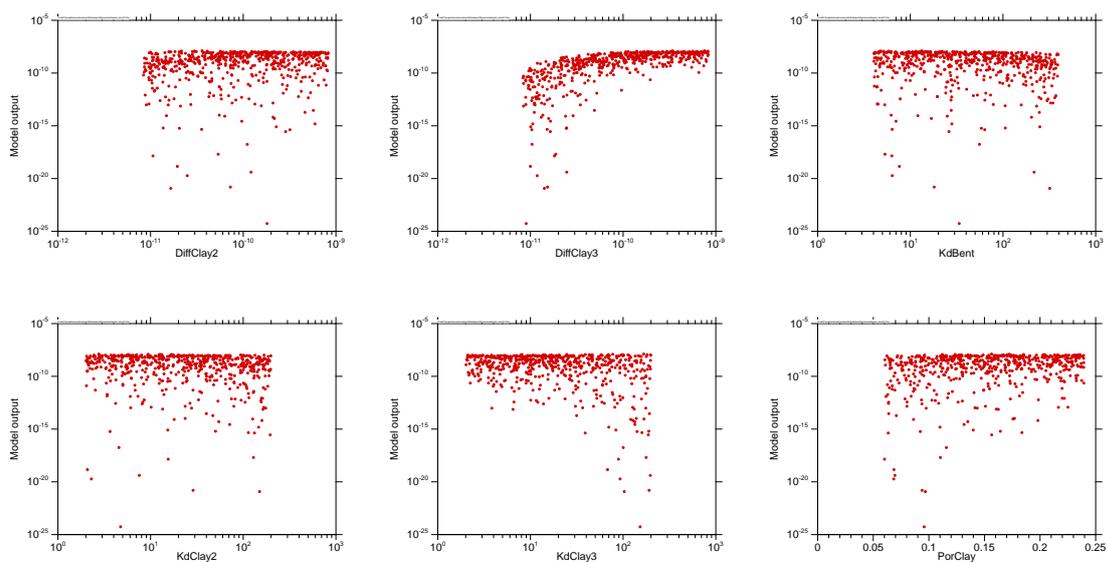
#### **4.1.1 Scatterplots for the clay system**

The scatterplots for the six parameters of the clay system are presented in Fig. 4.1 for a model time of  $10^6$  years on the basis of 512 runs (random sample). Even with this small sample size it can be seen that the model output, the annual dose rate in Sv/year, is distributed over many orders of magnitude and obviously, only two of the parameters – the diffusion constants in the two clay regions – have a clear influence.



**Fig. 4.1** Scatterplots for the clay system,  $10^6$  years

The equivalent scatterplots for the system at the end of the calculated scenario are presented in Fig. 4.2. The points concentrate below the value of  $3 \cdot 10^{-6}$  Sv/yr, which seems to be an upper limit. All parameters seem to have a certain influence to the model output. Since in most model runs the absolute maximum is reached at the end of the scenario, there are only minor differences between the scatterplots presented in Fig. 4.2 and those obtained for the run maxima. By this reason, the latter are not shown here.

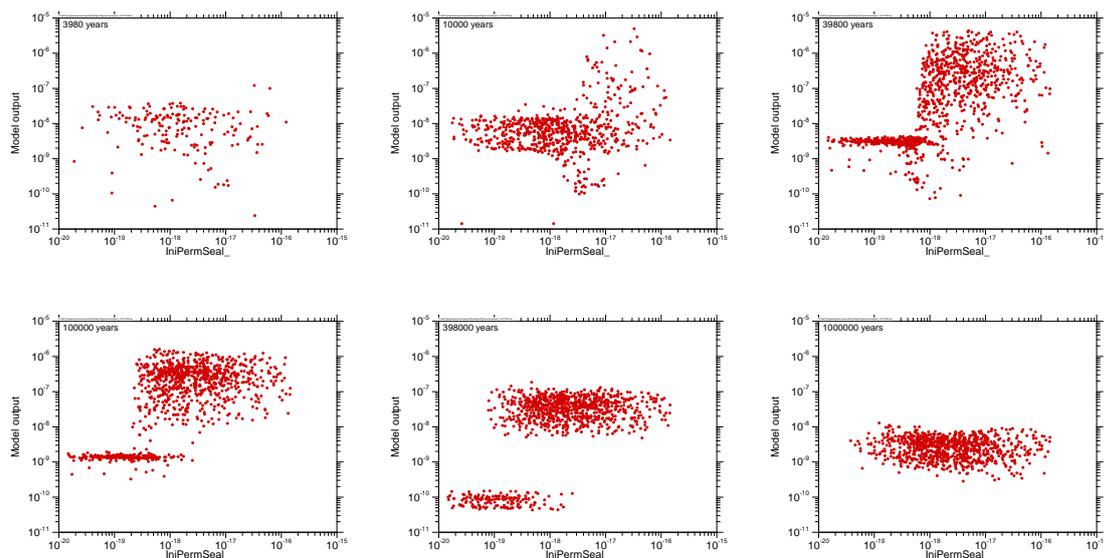


**Fig. 4.2** Scatterplots for the clay system,  $10^8$  years

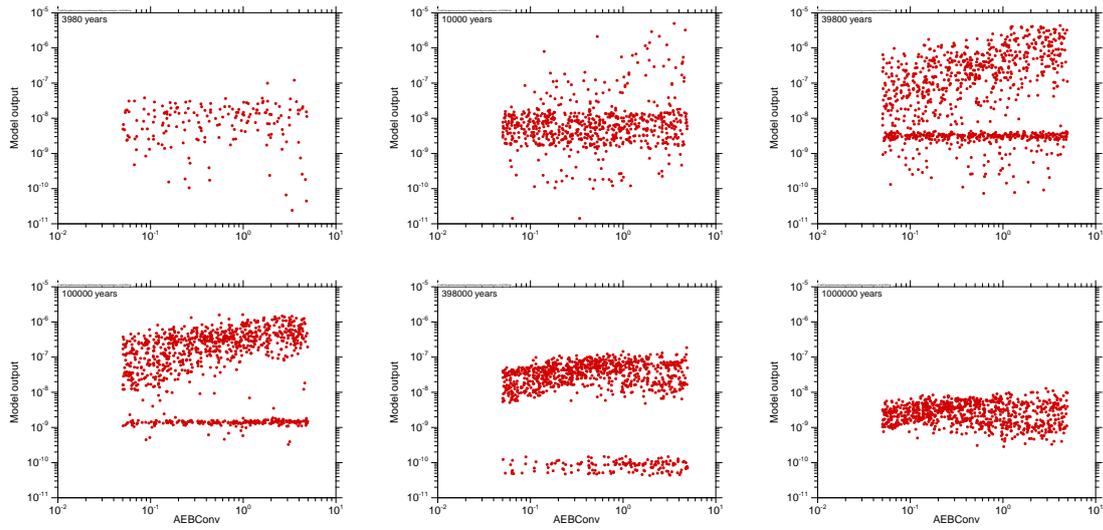
The scatterplots seem to suggest that during the relevant assessment period of 1 million years the system is predominantly controlled by the diffusion in the clay regions, while the sorption and the clay porosity gain some importance only in the very late phase, in which the model is no longer valid for describing the reality and which is interesting only for model understanding. This, however, might be a misinterpretation, as simple scatterplots are not appropriate for showing coupled influences of parameters. Since the individual dots look all the same and indicate the value of only one parameter, it is not discernible whether or not there is a tendency to produce higher or lower model output if two or more parameters act together.

#### 4.1.2 Scatterplots for the LILW system

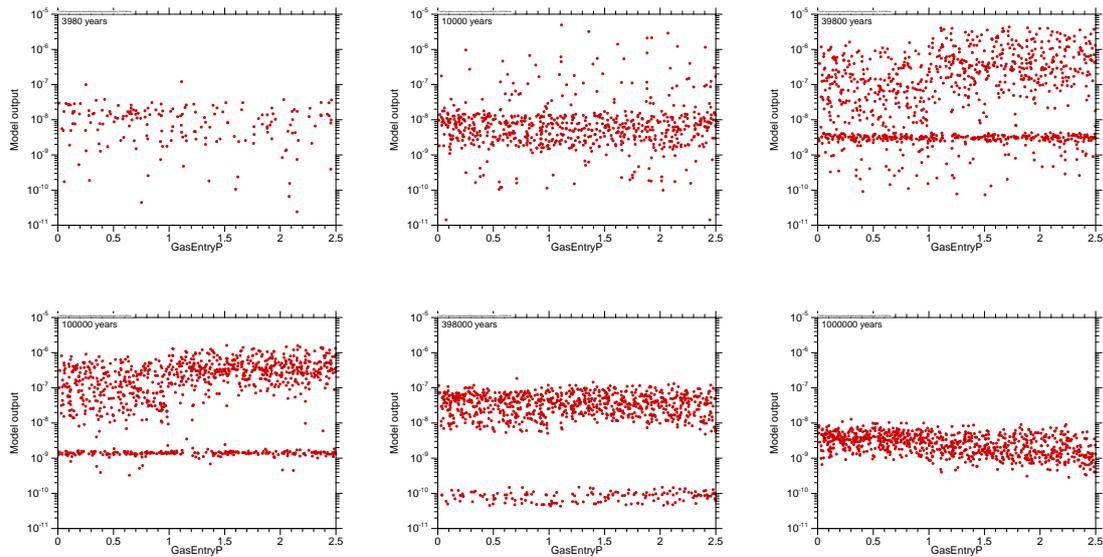
The LILW model is more complex than the clay model and yields scatterplots that distinctly evolve with time. In Fig. 4.3 to Fig. 4.6 this is shown for four influential parameters. The plots are based on 1000 runs, calculated with the LILW6 model. A clear split in two separate clouds of dots starts to become visible around 10 000 years, which is due to the failure of the seal. Those runs in which the seal has already failed at the presented time yield a higher output than those in which it has not. Obviously, the parameter *IniPermSeal* (the initial permeability of the seal) has a dominant influence on this behaviour.



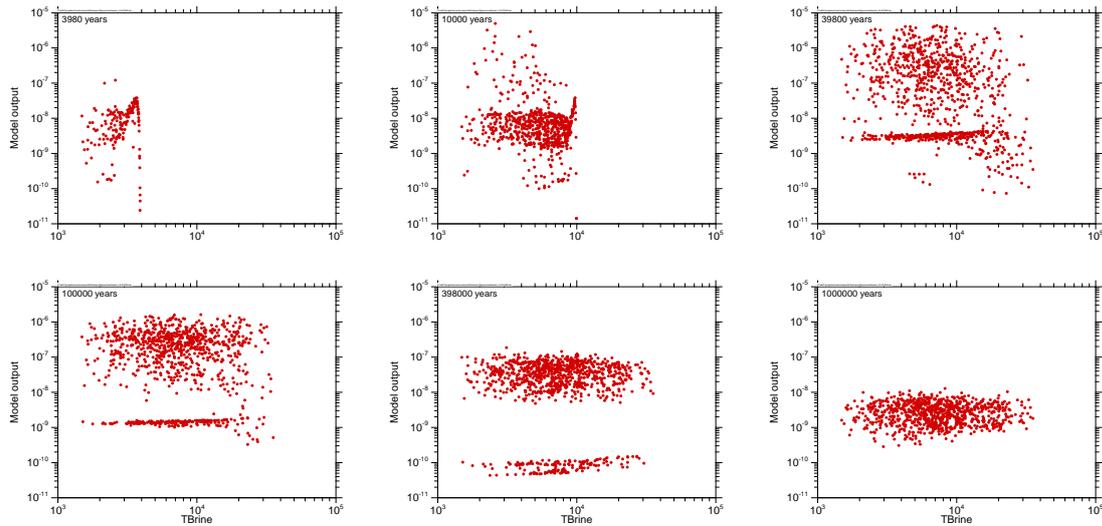
**Fig. 4.3** Time-development of the scatterplot for the parameter *IniPermSeal* (LILW6 model)



**Fig. 4.4** Time-development of the scatterplot for the parameter *AEBConv* (LILW6 model)



**Fig. 4.5** Time-development of the scatterplot for the parameter *GasEntryP* (LILW6 model)



**Fig. 4.6** Time-development of the scatterplot for the parameter *TBrine* (LILW6 model)

With permeabilities higher than  $10^{-18}$  m<sup>2</sup> the seal fails relatively early and the number of runs in which it has not yet failed decreases with time. After 1 million years, there are still a few cases in which the seal remains intact, but as the annual dose generally decreases with time, these cannot be seen on the selected scale.

The parameter *AEBConv* obviously has no significant influence on the model results as long as the seal is intact, but causes a further increase after seal failure. This is understandable because the parameter characterises the convergence of the sealed emplacement area, which becomes, of course, relevant only after seal failure. Also the parameter *GasEntryP* does only affect the upper cloud of dots. The parameter defines the gas pressure necessary for the gas to displace liquids from the pore space in the mine sealings and to allow the gas to escape. Due to specific model properties, the behaviour changes significantly at a value of 1.0, resulting in a higher release from the sealed emplacement area AEB. This can be seen in the scatterplots, especially those for medium times, from a kind of step at that value. The parameter *TBrine* is the time at which the mine is assumed to be suddenly filled with brine. The model output is zero before this time. While the parameter has a visible positive effect (higher parameter value – higher model output) on the lower cloud of points, which means before seal failure, there is nearly no effect on the upper cloud of points. As the model assumes that all radionuclides dissolve completely and instantaneously as soon as they get in contact with brine, due to the ongoing convergence a later brine intrusion leads to a higher

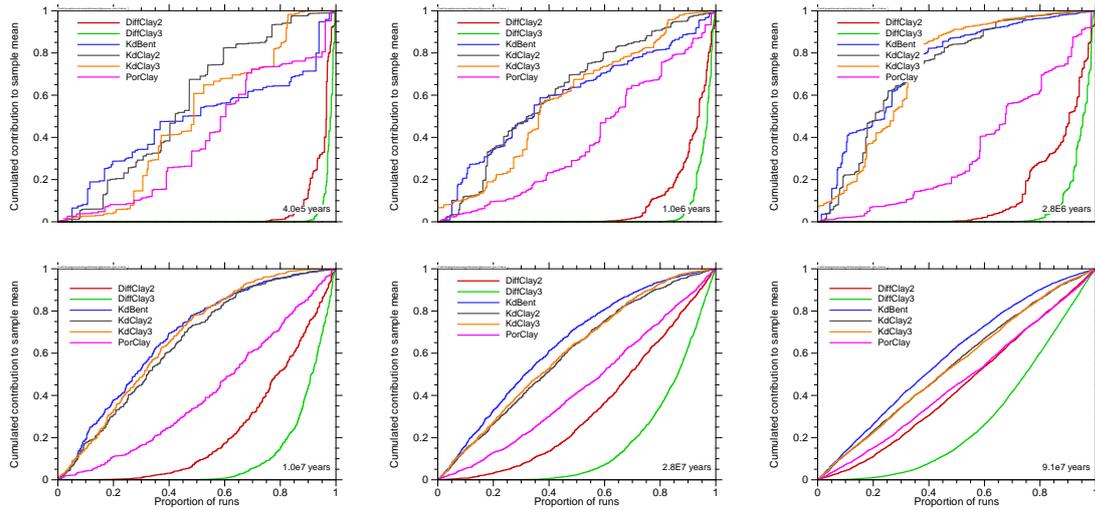
concentration in the non-sealed emplacement area and with that to a slightly higher outflow of radionuclides. Obviously, this effect is not relevant for the sealed emplacement area.

## **4.2 CSM, CUSUNORO and CSV plots**

CSM plots provide a quick graphical view of the sensitivities. This is demonstrated in the following by examples of the clay and the LILW system.

### **4.2.1 CSM plots for the clay system**

In Fig. 4.7 CSM plots for the clay system are presented for several points in time, based on a random sample of 4096 runs. It does not make sense to present the plots for earlier times than 400 000 years, because at such times only very few parameter combinations produce a non-zero output at all. Even after model times of several millions of years there are still a lot of zero or very low results, which is the reason for the uneven, step-like shape of the curves. The contribution to the sample mean is dominated by a few higher values, which are, in general, caused by parameters different from the presented one. This results in randomly distributed jumps in the CSM curves. Only for very late times of 10 million years or more, the curves look satisfyingly smooth.



**Fig. 4.7** CSM plots of the clay system

To understand the message of the plots, we first look at the lowest presented time of 400 000 years. The CSM curves for the parameters *KdBent*, *KdClay2*, *KdClay3* and *PorClay* show randomly distributed steps as explained above and progress horizontally in between. That means that most runs produce zero or very low model output, and the few ones that do not appear with no obvious relation to the parameter value. The model behaviour must therefore be dominated by other influences. These CSM curves are too uneven to be of relevance. The curves for *DiffClay2* and *DiffClay3*, however, remain on the zero level for the lower 80 or 90 % of the respective parameter values and the runs that really contribute to the mean, producing the steps, are concentrated at the upper end of the parameter intervals, which is a strong indication for significant sensitivity of the model to these parameters.

For later times, increasingly more runs produce non-zero output, so that the steps become smaller and the curves appear smoother. Beginning from about 1 million years, there is a clearly visible deviation from the diagonal even for the CSM curves for *KdBent*, *KdClay2*, *KdClay3* and *PorClay*. For the *Kd*-values the higher contributions to the mean concentrate at the lower end of the interval, which means that low *Kd*-values tend to produce higher dose rates. For the clay porosity, however, the CSM curve proceeds below the diagonal, meaning that higher porosities lead to higher dose rates.

#### 4.2.2 CSM plots for the LILW system

The time evolution of the CSM plots for the LILW6 system is presented in Fig. 4.8 on the basis of an investigation with 3000 runs and a random sample. Again, at early times the model output is often 0, so that the respective curves look uneven and stepped. For the earliest time of 3300 years, there is, understandably, a clear influence of the parameter *TBrine*. Only the lower 10 % of the values can lead to a non-zero output at all. The sensitivity to *TBrine* is still high at 10 000 years, but the importance of this parameter decreases quickly. Since a non-zero output is only possible if *TBrine* is smaller than the point in time under consideration, the CSM curve of *TBrine* visibly reflects the percentage of such calculation cases as the point where it reaches the value of 1: after 3300 years it is about 10 %, after 10 000 years 66 % and after 33 000 years 95 %. For the later times, there are no zero outputs.

The CSM curve for *IniPermSeal* seems to start only at a certain percentage, which is 40 % for 33 000 years, 20 % for 100 000 years, 10 % for 330 000 years and 5 % for 1 million years. This means that the lowest values of this parameter systematically yield very low model output, which is due to the fact that the seal has not failed by the respective time. *IniPermSeal* seems to have a strong influence on the failure of the seal, which is understandable, because it characterises the initial permeability of the seal and the flow of corrosive fluid.

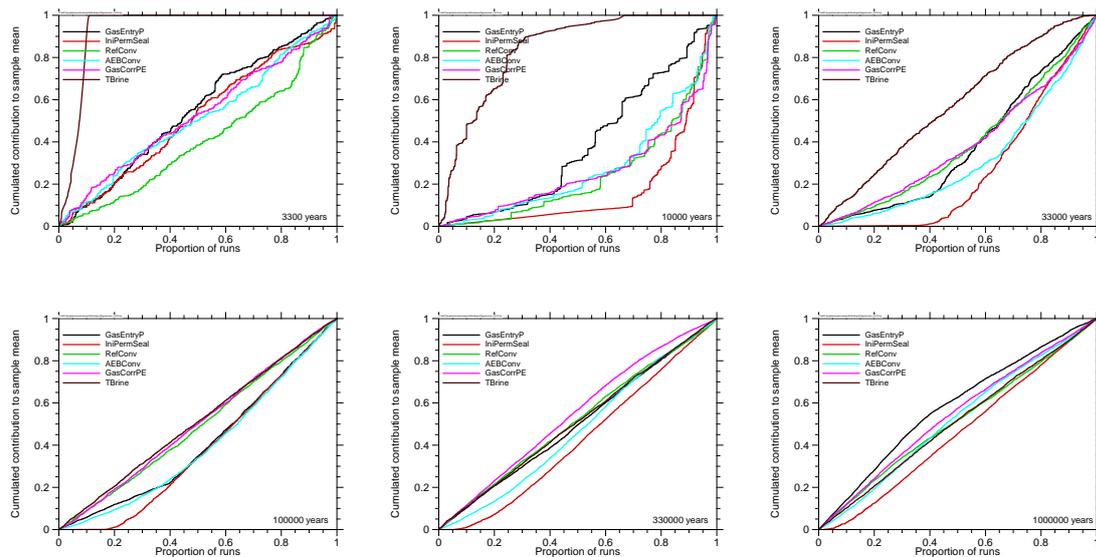
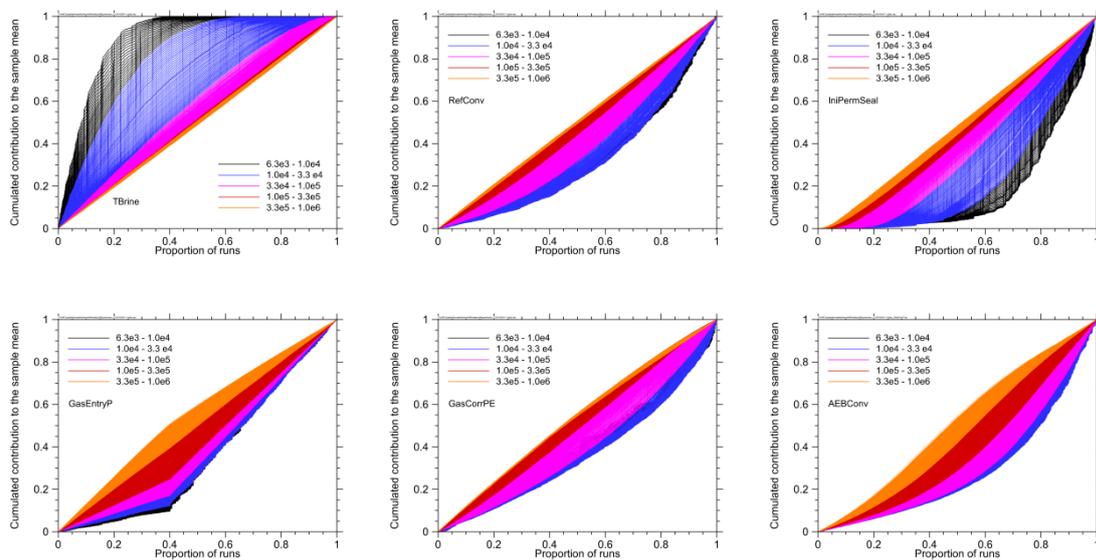


Fig. 4.8 CSM plots for the LILW6 system

The CSM curve for *GasEntryP* has a sharp bend at a proportion of 40 %, which is a hint that something remarkable happens if its value exceeds the 0.4-quantile. This corresponds with the value of 1.0, which is indeed the threshold at which the system behaviour changes as explained in the previous chapter.

Fig. 4.9 shows CSM plots for the LILW11 system. Here only one parameter is presented per plot, but the time-development is indicated by colours. The plots have been made on the basis of an investigation with 16 384 runs and a quasirandom sample. Due to the different number of varied parameters the curves are not fully comparable to those of Fig. 4.8, but this is of minor importance. As a result of the high number of runs, the curves look much smoother.



**Fig. 4.9** Time-development of CSM curves for the LILW11 system

This kind of presentation is appropriate for showing how the sensitivity of the system to a parameter evolves over time. It can be seen, for example, that the sensitivity to *TBrine* is high in the early phase but decreases quickly. The CSM curves for the parameters *GasEntryP*, *GasCorrPE* and *AEBConv* proceed below the diagonal for early times and above the diagonal for late times. They change their main curvature from concave to convex during the model time. This means that these parameters change their direction of influence to the model output: all of them have a positive influence (increasing parameter value – increasing dose rate) in the early phase but a negative influence (decreasing parameter value – increasing dose rate) at late times. The late CSM curves for *AEBConv* even have an inflection point, which means that in its low

range, the parameter has a positive influence but in its high range a negative influence. The CSM curves of the parameter *GasEntryP* show no curvature at all, except from the sharp bend at the 0.4-quantile, corresponding to the value of 1.0 Mpa. That means that there is practically no continuous sensitivity of the model to this parameter, neither below nor above the threshold value. While for early times, a gas entry pressure above the threshold leads to a higher dose rate than a value below the threshold, the opposite is the case for late times. This is due to the fact that with a high gas entry pressure the inventory of the sealed emplacement area AEB is released faster, so that it is widely depleted in the late phase.

#### 4.2.3 CSM versus CUSUNORO

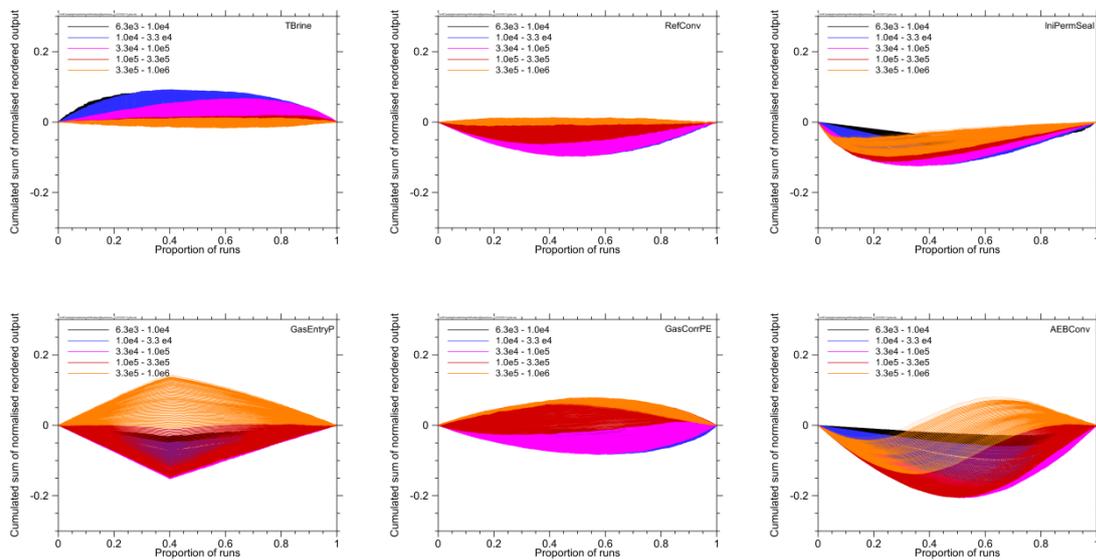
The CSM plot is easy to make and to understand, but has a few drawbacks:

- The curves are oriented around the diagonal, which makes it a bit hard to recognise deviations.
- Due to the non-rectangular projection to the x-axis conspicuous features are hard to assign to a specific quantile.
- Magnification by confining the range of y-values in order to better present curve details is not possible.
- Since the calculation is related to the mean of all values and not to their variation, a common additive offset changes the shape of the CSM curves. The higher the absolute value of the mean, the less pronounced appear the curves. If, for instance, the model output were a temperature, the shape of the CSM curves would depend on the measuring unit, and if Kelvin were used, deviations from the diagonal would possibly be barely visible.
- Should the mean of the values be zero, CSM plots cannot be made.

To overcome all of these drawbacks, the CUSUNORO plot has been proposed by Plischke (/PLI 12/). For this kind of presentation the values are normalised to the standard deviation instead of the mean. The curves proceed around the x-axis, starting at (0,0) and ending at (1,0).

In Fig. 4.10 CUSUNORO plots are presented for the same model calculation and in the same way as the CSM plots in Fig. 4.9. While the principle statement is similar, there are some differences. The dominance of the parameter *TBrine* in the early phase ap-

appears less pronounced in the CUSUNORO plot. Moreover, the quantile at which the influence of *TBrine* stops cannot be seen as clearly as in the CSM plot. On the other hand, a slight positive curvature of the very late curves is discernible, and could be optically magnified by using a different y-scale, which is nearly invisible in the CSM plot. The CUSUNORO curves for *AEBConv* show even more clearly than the CSM curves that for late times this parameter changes its direction of influence over its interval. While the curvature is positive for low values, it becomes negative for high values. Physically, this means that if the convergence rate of the sealed emplacement area is low, a slight increase will cause a higher dose rate at late times, but if it is high, a further increase will reduce the dose rate.



**Fig. 4.10** Time-development of CUSUNORO curves for the LILW11 system

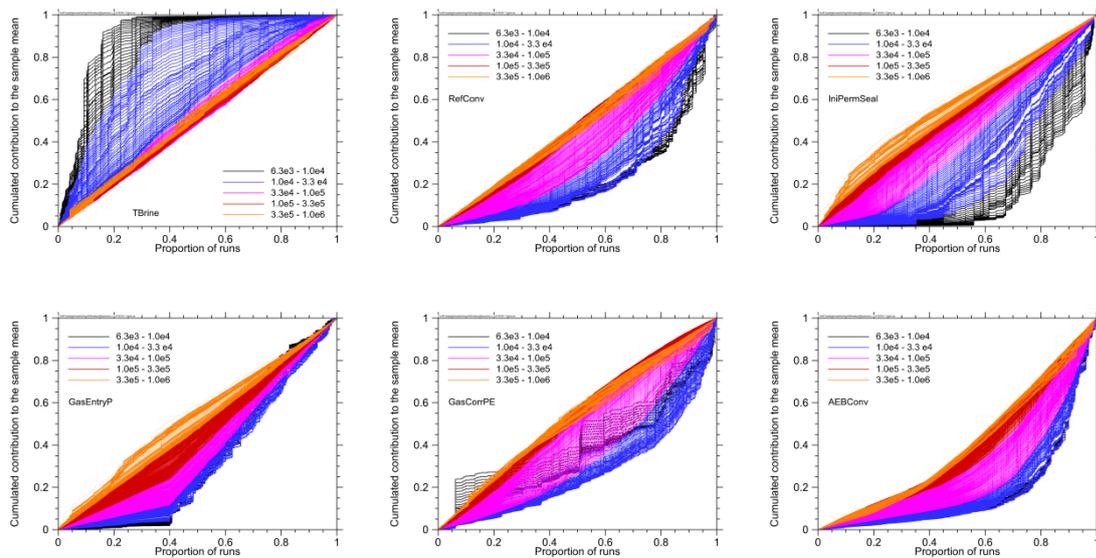
Generally, CUSUNORO plots seem less appropriate for this kind of presentation as the curves for increasing points in time are more likely to intersect and cover each other.

#### 4.2.4 CSV plots

Following the principle of the CSM plot, but plotting the contribution to the total sample variance instead of the sample mean, one gets a similar kind of plots. Due to the quadratic nature of the variance, CSV curves are generally more pronounced but less smooth and robust than CSM curves. For an output distribution over several orders of magnitude with many low and a few higher values, as it is typical for a repository mod-

el, the variance is often dominated by a relatively low number of individual values. If these dominating values are distributed more or less randomly, the CSV curve looks uneven. In Fig. 4.11 the time-development of the CSV curves for the LILW11 model is presented in the same way as for CSM in Fig. 4.9, and despite the high number of model evaluations (16384) many of the curves, especially those for early points in time, which are based on data sets with many zeros, show distinct jumps.

Due to this property, amongst other things, CSV seems less appropriate for a quick visual assessment of the sensitivities of the system, especially with a low number of model runs. An advantage, however, is that in contrast to CSM, the CSV curves are invariant to a constant offset and can be made also for negative model output or zero mean. Only a zero total variance would be forbidden, but this would mean a strictly constant model output, which is anyway of no interest for sensitivity analysis.



**Fig. 4.11** Time-development of CSV curves for the LILW11 system

### 4.3 Summary

A variety of graphical methods for sensitivity analysis are available for providing a quick and descriptive visual impression of the sensitivities of a model. Scatterplots, CSM, CSV and CUSUNORO plots were applied to the test systems.

Scatterplots are the most direct method of investigating the influence of a parameter to the model output. They provide the most detailed impression of the parameter influ-

ence and allow visual assessment of the sensitivity. Therefore, scatterplots are most valuable for gaining system understanding and screening the variables in view of further sensitivity analysis. A drawback is that a big number of plots are needed, as each individual plot represents one specific parameter and one specific evaluation point. Combining several scatterplots for demonstrating the time development or comparing different parameters is possible by using different colours, but will normally lead to confusing and unclear figures.

CSM plots provide an illustrative visualisation of parameter sensitivities and are easy to understand and communicate. As for each parameter and each evaluation point a continuous curve is produced, several of them can be clearly presented in one figure. So these plots are much better than scatterplots for comparing different parameters or different points in time. A drawback of CSM is that the curves are not invariant to an additive shift in the model output, and cannot be made at all if the mean happens to be 0. Moreover, as the scale is fixed, curve details cannot be magnified.

CSV plots supply a similar – though different – message as CSM, but are generally harder to understand. The curves are often more pronounced and less smooth. Unlike CSM, CSV plots are invariant to additive shift and can be made for any non-constant set of model output values, regardless of its mean.

CUSUNORO plots provide basically the same information as CSM plots, avoiding the mentioned disadvantages. These plots are not only invariant to additive shift, but use a scalable, rectangular presentation, which allows detail studies. A drawback is that they are harder to explain and understand.

CSM, CSV and CUSUNORO plots are the only methods investigated in the project that allow identification of sections of the parameter range with higher or lower sensitivity. An inflection point is a sign for a change of the direction of influence at some specific parameter value.

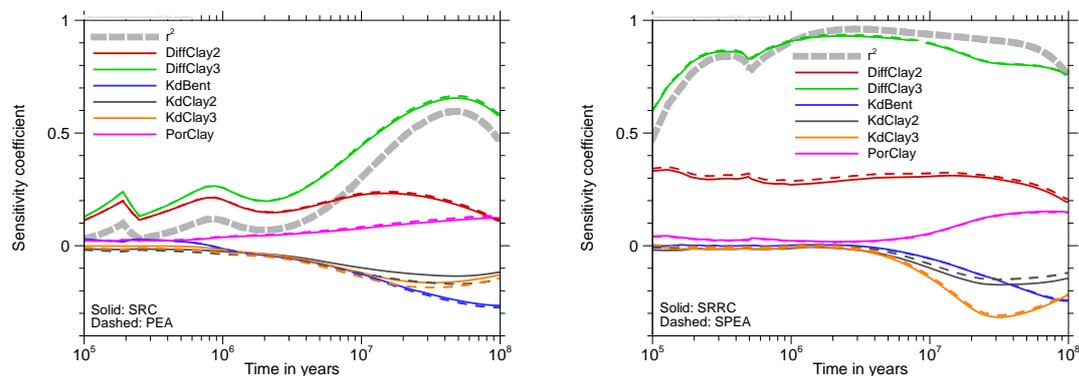


## 5 Correlation- and regression-based sensitivity analysis

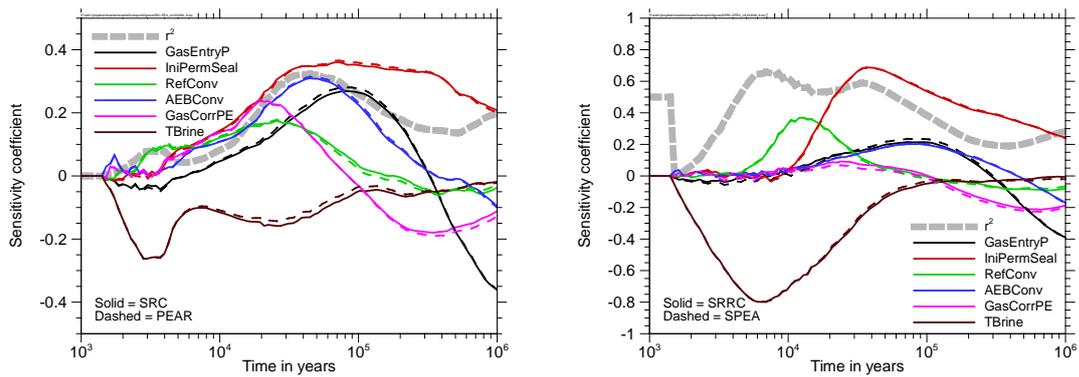
Since correlation- and/or regression-based methods have been the most common approach to sensitivity analysis for a long time, intensive investigations with such methods were made using the clay model and the LILW model in order to provide a starting point for comparison with other methods. In this chapter the results are presented and explained.

### 5.1 Comparison of correlation and regression

The first question to answer is how the correlation-based sensitivity measures differ from the regression-based ones. These are compared in Fig. 5.1 and Fig. 5.2 for the clay model and the LILW6 model. The figures show the time curves for the mentioned sensitivity measures calculated for both systems model runs, based on random samples of sizes 4096 (clay) and 3000 (LILW6). All examples show that the difference between the correlation-based and the regression-based analysis is small for all times, no matter if the direct or rank-based evaluation is performed. This may be a bit surprising, because correlation and regression are different mathematical concepts, and by theory, a strong correlation between input and output of a model one does not necessarily mean a high regression coefficient, and vice versa. Obviously, however, this difference is of low significance for practical sensitivity analysis. By this reason, we confine ourselves to considering the regression-based types of sensitivity analysis in the following.



**Fig. 5.1** Regression- and correlation-based sensitivity analysis with direct (left) and rank-transformed (right) data for the clay model



**Fig. 5.2** Regression- and correlation-based sensitivity analysis with direct (left) and rank-transformed (right) data for the LILW6 model

## 5.2 The principle message

The time curves of SRC or SRRC provide information on how the sensitivity of the system under consideration to the individual parameters evolves with time. In contrast to the graphical methods described in the previous chapter, however, this method cannot resolve the variation ranges of the parameters. Only one value per parameter is calculated for each point in time, which can only be representative if the model answer to variation of the parameter is more or less homogeneous over the total range of variation, i.e. the model behaviour is close to linear. Consequently, a low SRC or SRRC value does not necessarily mean a low sensitivity.

As an example, we consider the SRC and SRRC curves for AEBConv (Fig. 5.2). The values become very small and finally negative at late times. By comparing this to the CUSUNORO curves for this parameter (Fig. 4.10), one can see that the reason is not a vanishing sensitivity but the fact that the direction of influence changes over the range of parameter values. The late CUSUNORO lines are distinctly curved, indicating a high sensitivity, but they change their curvature from positive to negative, which means a non-monotonic behaviour. For SRC the positive and negative contributions compensate each other, resulting in a near-zero value. For the latest times, the negative contribution even predominates, which can also be verified with the CUSUNORO plot. This kind of information cannot be derived from the SRC or SRRC curves, which seem to indicate a low sensitivity at late times. Also the curves for *GasEntryP* and *GasCorrPE*

cross the zero line at some time, but in these cases, that really means that there is no sensitivity at some point in time, which can be seen at the CSM or CUSUNORO plots.

### 5.3 Direct evaluation of the model output (SRC)

The coefficient of model determination  $R^2$  provides a measure for assessing how well the model is determined by linear regression. While a value of 1 would mean a strong multilinear relationship, zero indicates that linear regression does not at all approximate the model. It is often recommended as a rule of thumb that for regression-based sensitivity analysis to work properly  $R^2$  should be greater than 0.5. In Fig. 5.1 and Fig. 5.2 the time-development of  $R^2$  is presented as a light-grey, long-dashed curve. It can be seen that for both model systems, if the direct SRC evaluation is applied, it remains below 0.5 for most of the model time and is even below 0.2 for parts of the time interval. Nevertheless, the SRC evaluation seems to identify at least the main sensitivities correctly.

#### 5.3.1 The clay system

The left part of Fig. 5.1 shows that  $R^2$  is low during the “realistic” time period and rises above 0.2 only after 7 million years. Nevertheless, some sensitivity statements can be derived from the SRC evaluation. In agreement with the graphical sensitivity analysis presented in chapter 4, the parameters *DiffClay3* and *DiffClay2* are identified as the most and second most influential ones over the total range of time. The SRC values of the other parameters are very close to zero during the first million years of model time and only then start to assume significant nonzero values. It becomes clear that the parameters *DiffClay3*, *DiffClay2* and *PorClay* have a positive influence while *KdBent*, *KdClay3* and *KdClay2* have a negative one, as expected: higher Kd means higher sorption and with it less release of contaminants. All calculated sensitivities become more pronounced with time and also  $R^2$  rises. This is due to the fact that for most parameter combinations a significant contaminant release starts only at a late time.

#### 5.3.2 The LILW system

The complex character of the LILW system is reflected in the SRC curves, which look rather uneven and over a wide range of time do not seem to allow proper assessing of sensitivities and ranking of parameters. Only the parameter *TBrine* clearly dominates

during the first few thousand years. From about 20 000 years on,  $R^2$  rises above 0.2 and the parameter ranking becomes a bit more unique. As mentioned above, the SRC values for the parameters *AEBCConv*, *GasEntryP* and *GasCorrPE* change their sign during the model time, which is consistent with the findings from the CSM and CUSUNORO plots.

## 5.4 Rank-based evaluation (SRRC)

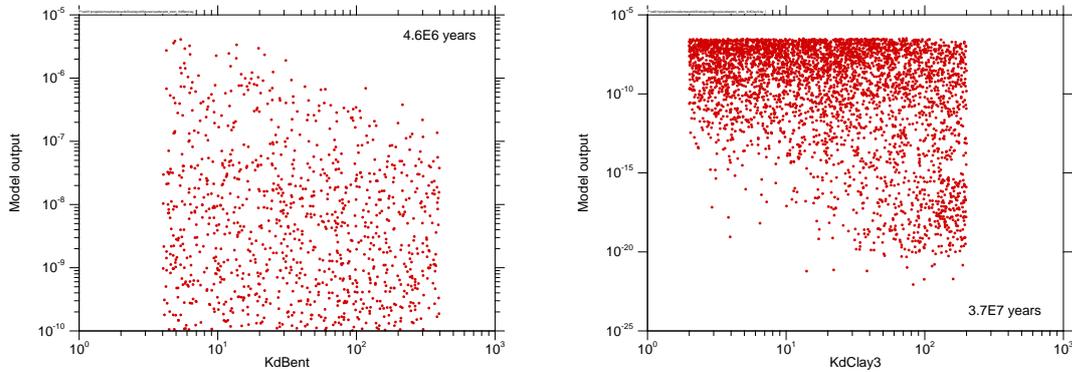
For both models, it can be seen that  $R^2$  essentially rises if a rank transformation is performed. It is generally assumed that the regression-based sensitivity analysis then becomes more meaningful.

### 5.4.1 The clay system

For the clay model the increase of  $R^2$  caused by rank transformation is extreme. It even reaches nearly 1 in a time range where it is below 0.1 for the direct evaluation. Such a pronounced difference in linear model determination, induced by rank transformation, is a hint at a strongly nonlinear but monotonic model behaviour. In this case, it can be explained by the fact that, in the time range of a few million years, the release of contaminants typically increases in more and more model runs, resulting in model output values logarithmically distributed evenly over many orders of magnitude, as can be seen from the scatterplots in Fig. 4.1.

In fact, the SRRC time curves provide a more unique parameter ranking than the SRC curves, especially in the “realistic” phase. The parameter *DiffClay3* appears as clearly most dominating over the total time range, followed by *DiffClay2*. The other parameters show no practical influence at all, initially, and their SRRC curves only start to diverge from the zero line after 5 million years. Interestingly, this happens later than in the direct SRC evaluation. At very late times, however, SRRC yields a different parameter ranking than SRC and assigns a higher influence to the parameter *KdClay3*. In order to explain this, Fig. 5.3 presents two scatterplots, visualising the influences of *KdBent* at 4.6 million years and of *KdClay3* at 37 million years. It can be clearly seen that *KdBent*, at the earlier time, mainly affects the highest model output values while *KdClay3*, at the late time, mainly influences the lowest values. A variation of the relatively few high values hardly changes the value ranking and the rank regression, but, as these values are orders of magnitude higher than the lower ones, can have a considerable influence on

the direct regression. Therefore, sensitivities of this type are better detected by SRC. On the other hand, variations of very low values are of minor relevance for direct regression but get more importance on a rank basis.



**Fig. 5.3** Scatterplots of the clay model, presenting parameters that influence the highest (left) and the lowest (right) values

From the SRC and SRRC time curves it can be concluded that during the time frame between 0.5 and 5 million years the influence of *KdBent*, *KdClay2*, *KdClay3* and *PorClay* affects mainly the highest model output values. At late times, these parameters exert an increasing influence to the lower model output values, which is especially valid for *KdClay3*.

#### 5.4.2 The LILW system

For the LILW model the increase of  $R^2$  caused by rank transformation is less pronounced. Around the time of 400 000 years it changes merely from 0.14 to 0.19. Especially for the early time phase up to about 10 000 years, however, there is a considerable increase of  $R^2$ , and compared to the SRC evaluation SRRC seems to produce a more unique ranking. According to SRRC, there is practically no influence during this time frame from other parameters than *TBrine* and *RefConv*, while SRC calculated comparably significant values also for the other parameters. Again, this is an indication that these parameters mainly influence the high values, which changes the regression coefficient but not the ranking.

At late times the SRRC curves are very similar to those calculated with SRC, but look a bit smoother. The linear model determination, measured by  $R^2$ , is only slightly better for

the rank-transformed data. Obviously, the model linearity cannot be significantly increased by rank transformation in this phase.

## 5.5 Summary

Correlation- or regression-based sensitivity analysis emerged to be an appropriate approach for identifying the main sensitivities of both investigated test models. Significant differences between the results obtained with the correlation-based and the regression-based methods were not found; therefore, it seems sufficient to apply one or the other. For the purpose of this report we use SRC/SRRC.

As only one value per parameter is calculated for each point in time, the method is not adequate to resolve variations of the influence to the model output over the parameter interval. Consequently, a value near zero does not necessarily mean a weak influence of the parameter.

Linear regression performs best for models with a close-to-linear behaviour. This can be checked by calculating the coefficient of model determination  $R^2$ . A high value of  $R^2$ , however, is not a precondition for the method to yield meaningful results. The main sensitivities of both test models were identified with SRC even where  $R^2$  is below 0.1.

Rank transformation can significantly increase  $R^2$ . By theory, this is the case for nonlinear but monotonic models. Therefore, the rank-based variant SRRC is often considered more appropriate for nonlinear systems. This, however, is not necessarily the case. If the model output is roughly logarithmically distributed over several orders of magnitude and a parameter mainly influences the highest values, this is better detected by SRC than by SRRC, because the ranking of the few high values merely changes. On the other hand, influences of a parameter to the lowest values are generally better detected by SRRC.

When applying SRRC it should be kept in mind that the rank transformation is a non-reversible transformation. Inevitably, some information about the model sensitivity gets lost if a rank-based evaluation is performed.

## 6 Variance-based sensitivity analysis

In contrast to regression- or correlation-based sensitivity analysis, variance-based sensitivity analysis does not, by theory, require any linear or close-to-linear behaviour of the model under consideration. A variety of methods are available to calculate variance-based sensitivity indices, a few of which were applied to the test models. The methods have been theoretically described in chapter 2.7.

Not all methods, however, are designed to calculate all first-, higher- and total-order indices. The first-order sensitivity indices (SI1) quantify the influence of the variation of one parameter alone to the variation of the model output. Second- and higher-order indices quantify the combined influence of two or more parameters. The total-order indices (SIT), on the other hand, quantify the influence of a specific parameter in coaction with all others.

The sum of all first- and higher-order indices is, by theory, 1 (although in practice, due to shortcomings of the applied methods, the calculated indices sometimes sum up to a higher value). The sum of all first-order indices provides a measure for how well the system is determined by independent influences of the individual parameters and how well the sensitivities are described by first-order analysis.

In the context of MOSEL, we investigated a selection of methods. The main focus was laid on the first-order indices, as these can be calculated by all methods and convey a well understandable message. Second- and higher-order indices might provide interesting insights in the model behaviour in specific situations, but the number of such indices is high and an easily applicable method for calculating them was not available. Therefore, this report does not provide any results on second- or higher-order sensitivity indices. Additional information, however, can sometimes be taken from the total-order indices, which some methods are able to calculate with justifiable effort.

Tab. 6.1 contains an overview of the investigations that were performed. Some basic results are presented in the following. As the influence of the number of varied parameters is the subject of chapter 8.1, here we concentrate on results for the clay model and the LILW6 model.

**Tab. 6.1** Investigations performed with variance-based methods

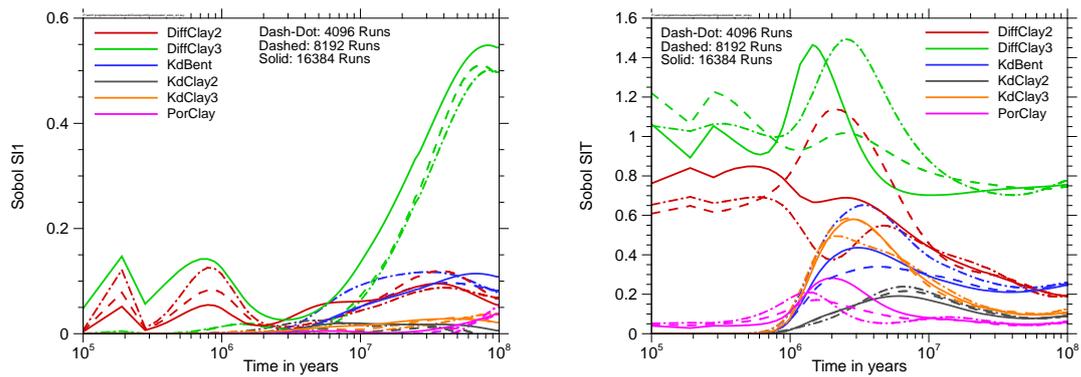
Method	Clay	LILW6	LILW7	LILW11	LILW20
Sobol'	SI1, SIT	SI1, SIT	SI1, SIT	SI1, SIT	-
EFAST	SI1, SIT	SI1, SIT	SI1, SIT	-	-
RBD	SI1	SI1	SI1	SI1	-
EASI	SI1	SI1	SI1	SI1	SI1
SDP	SI1	SI1	SI1	SI1	SI1

## 6.1 Sobol' method

The Sobol' method is the oldest method for calculating the variance-based sensitivity indices. It requires a specific kind of sampling and a rather high number of runs, but principally allows calculation of sensitivity indices of any order. Several variations and computational improvements of the procedure originally proposed by Sobol' have been developed (/SOB 01/, /SAL 02/). In this project, the method was investigated in the form implemented in the SimLab 3 software. The results presented in the following were calculated using SimLab 3.2.6.

### 6.1.1 The clay system

In Fig. 6.1 the time-developments of the Sobol' sensitivity indices of first and total order are presented for the clay system, calculated from 4096, 8192 and 16384 runs. Especially for the total-order indices, there are still big differences between the corresponding curves for 8192 and 16384 runs, which means that even the set with the highest number of runs does not provide stable results. Moreover, the calculated total-order index for *DiffClay3* exceeds 1, which should not be possible, by theory. Obviously, the applied algorithm is not adequate to provide reliable results.



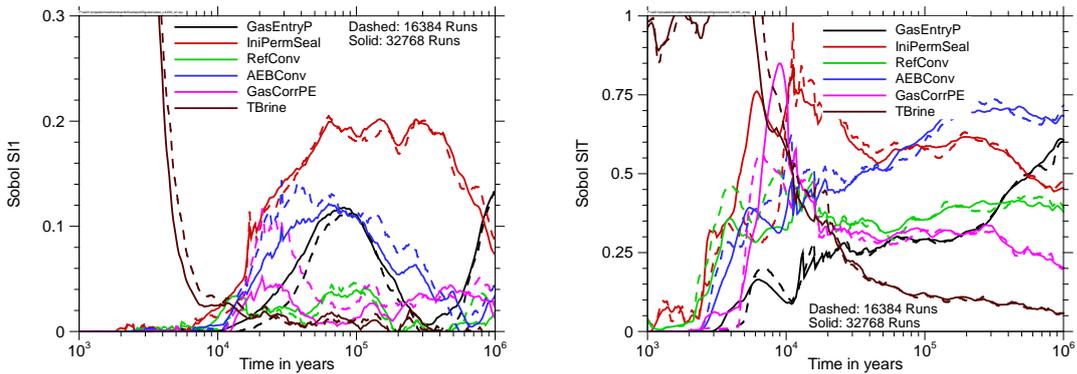
**Fig. 6.1** Sobol' indices of first (left) and total order (right) for the clay model

Nevertheless, the main findings from the SRC-/SRRC-analysis seem to be confirmed by the Sobol' analysis: *DiffClay3* is by far the most influential parameter nearly over the total time period, followed by *DiffClay2*, while *KdBent* achieves some considerable influence only in the very late phase, and finally even exceeds that of *KdClay2*. An interesting result, however, is that around 3 million years all first-order indices are very small, their sum being smaller than 0.1. The calculated total-order indices, on the other hand, exhibit the most confuse and unstable behaviour around this time. This can be interpreted in that way that at that time the system is dominated by multiple parameter interactions. At very late times, when the sum of all SI1 rises and reaches a maximum of about 0.8, the convergence of the SIT-curves is better, but the total-order analysis as a whole does not seem reliable.

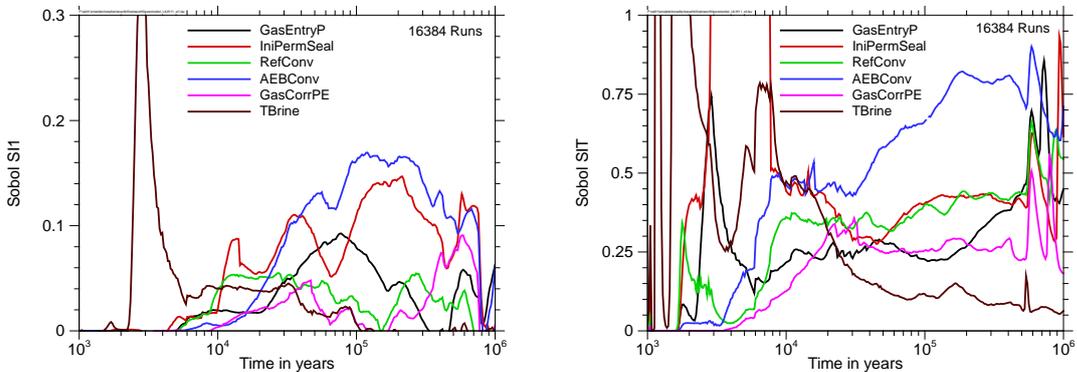
### 6.1.2 The LILW system

The time-developments of the first- and total-order sensitivity indices, calculated with the Sobol' method for the LILW6 model, are presented in Fig. 6.2. The indices were calculated for 301 points in time, resulting in curves that look very uneven and do not reach satisfying agreement even between sample sizes as high as 16384 and 32768. Although the Sobol' analysis identifies the main sensitivities in basic agreement with the SRC/SRRC results, this investigation does not seem to be reliable in this case. At least, some information about the total sensitivity indices can be deduced. This, however, does not seem to apply any longer if more parameters are taken into account. In Fig. 6.3 the time-dependent Sobol' evaluation for the LILW11 model is presented for a sample size of 16384. For better comparability with the results for the LILW6 model on-

ly the six most relevant parameters are depicted. The curves exhibit a very unsteady progress and hardly allow clear statements about the sensitivity. The calculated SI1 are generally lower than for LILW6, which is understandable as the total output variance is distributed to contributions of more parameters, but they even assume negative values (not depicted in the figure), which is forbidden by theory and does not make sense. At least, the SIT curves seem more or less stable in the medium time frame between 30 000 and 500 000 years and there is also a fair agreement with those for the LILW6 model in this time frame.



**Fig. 6.2** Sobol' indices of first (left) and total order (right) for the LILW6 model



**Fig. 6.3** Sobol' indices of first (left) and total order (right) for the LILW11 model (only the 6 most relevant parameters are presented)

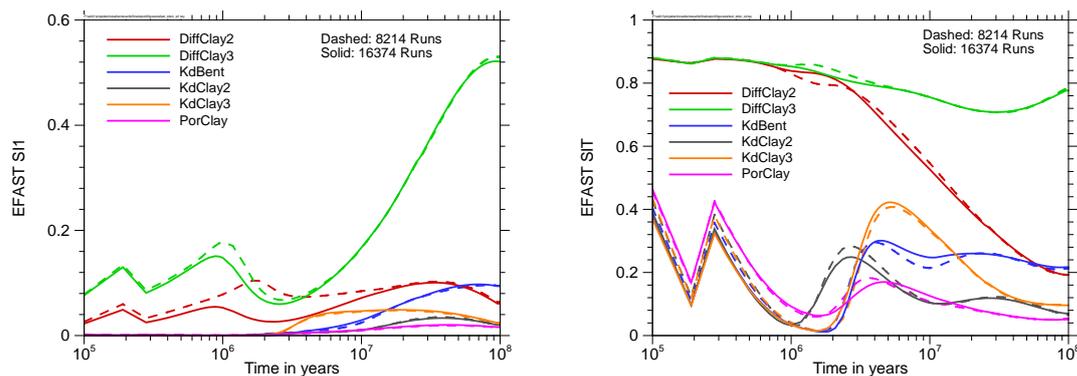
Especially in view of the high computational effort for the Sobol' analysis in the form as implemented in SimLab 3, a practical use cannot be derived from these investigations.

## 6.2 EFAST

The EFAST method is a specific variant of the FAST method, which is able to determine the first- and total-order sensitivity indices in one step. EFAST requires a specific, periodic sampling, which is non-extendible. Although the sampling scheme is basically deterministic, a random element can be introduced by a phase shift. The evaluations were made with SimLab 3.2.6.

### 6.2.1 The clay system

EFAST results for the clay model, calculated with sample sizes of 8214 and 16374, are presented in Fig. 6.4. The curves look smooth and plausible and seem to be fairly convergent for the investigated number of runs. The first-order analysis is well comparable to the Sobol' results (see Fig. 6.1, left) and also confirms the results of SRC/SRRC. The total-order EFAST analysis produces much clearer results than Sobol'. In the early phase, *DiffClay2* and *DiffClay3* have high and nearly identical SIT values. These parameters seem to dominate the system in mutual interaction without relevant contributions from the other parameters. The SIT values calculated for these are probably due to the low number of runs with relevant output in the early phase, and actually originate from a pseudo-interaction with *DiffClay2* and *DiffClay3*.



**Fig. 6.4** EFAST indices of first (left) and total order (right) for the clay model

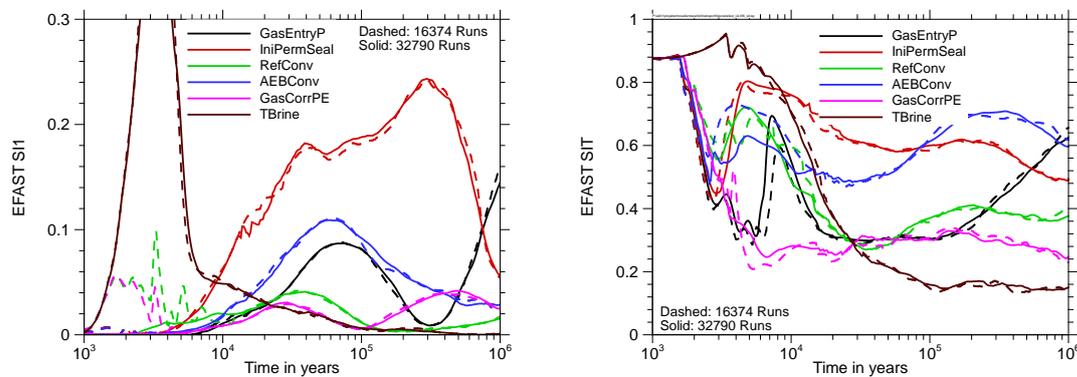
From 1 million years on, the total-order index of *DiffClay2* decreases while those of *KdBent*, *KdClay2*, *KdClay3* and *PorClay* increase and that of *DiffClay3* remains relatively high. In the late phase the system is clearly dominated by *DiffClay3* in interaction with all other parameters. In the time phase below 1 million years the model output is

normally very low and therefore controlled by few runs with higher output values. As a consequence, a stable evaluation cannot be expected and the curve progressions should not be taken too seriously.

For the relatively simple clay system, the EFAST method seems to provide a reliable approach to variance-based sensitivity analysis, which is able to calculate the first-order as well as the total-order sensitivity indices. The results are consistent with those of earlier investigations with the clay model (/SPI 12c/).

## 6.2.2 The LILW system

EFAST results for the first- and total-order sensitivity indices of the LILW6 model are presented in Fig. 6.5. Compared to the Sobol' results the curves look smoother and the convergence between sample sizes of about 16 000 and 32 000<sup>1</sup> is better, so that, at the first sight, EFAST seems to provide more stable and reliable results for variance-based sensitivity analysis of the LILW model.

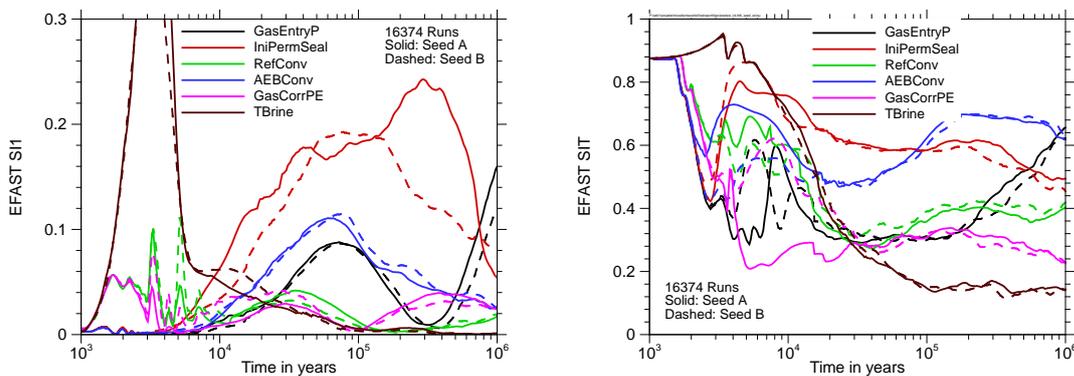


**Fig. 6.5** EFAST indices of first (left) and total order (right) for the LILW6 model

In Fig. 6.6, however, two EFAST evaluations are compared that only differ by the random phase shift, which was realised by using two different seeds for the random number generator. There are essential differences between the curves, especially with regard to the parameter *IniPermSeal*. At 300 000 years the two S1 calculations even

<sup>1</sup> The EFAST sampling scheme as realised in the SimLab software generates specific sample sizes; therefore, they slightly differ from the sample sizes used for the Sobol analysis

differ by a factor of 2. The EFAST sampling follows specific periodic paths through the parameter space, interrupted by jumps induced by the phase shift (see Fig. 2.4). Since these jumps in the sample are the only difference between the two sets of model runs, the evaluation is obviously susceptible to them. The highly nonlinear model can react to a jump in the input parameters with a considerable change in the output and thus influence the results of the evaluation. The parameter *IniPermSeal*, which is most affected by this susceptibility, is the one that predominantly controls the quasi-non-continuous behaviour of the LILW model due to seal failure. One can conclude that EFAST is not best compatible with extremely nonlinear models. Moreover, the EFAST sampling provides a very inhomogeneous coverage of the parameter space (see Fig. 2.4), so that, even with high sample sizes, specific parameter combinations are never met. This may also be a reason for the relatively low robustness of the EFAST evaluation.



**Fig. 6.6** EFAST indices of first (left) and total order (right) for the LILW6 model, calculated with two different random seeds

### 6.3 RBD

Like EFAST, the Random Balance Design (RBD) technique is also based on periodic sampling and Fourier analysis of the output, but the periodicity is hidden by applying random permutations of the parameter values (see Chapter 2.7.3), resulting in a more homogeneous coverage of the parameter space. The method is only capable of calculating first-order indices.

### 6.3.1 The clay system

The SI1 time curves obtained with RBD for the clay model are presented and compared to the corresponding results of EFAST and Sobol' in Fig. 6.7. The convergence between sample sizes of 8192 and 16384 is quite good and the curves are in fair agreement with those produced by the other methods. During the early phase RBD yields a lower estimation of the sensitivity index for *DiffClay3* than the other methods, but as mentioned above, the model itself is not very robust for early times and the sensitivity analysis results are of low relevance.

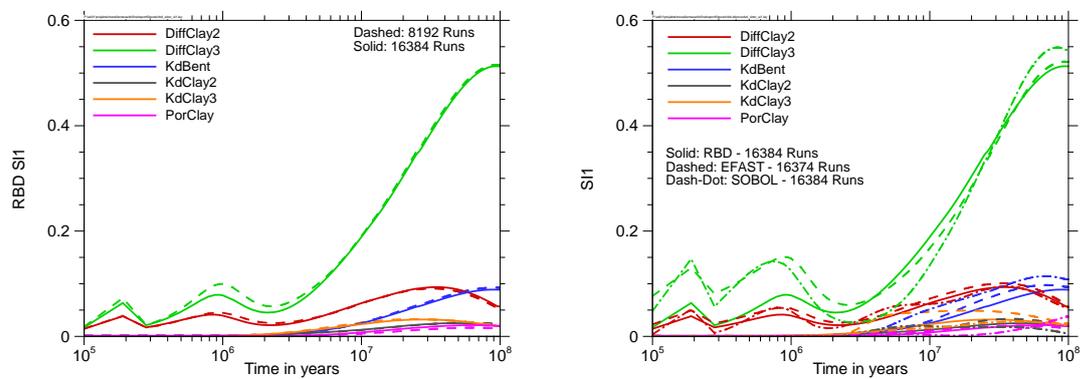
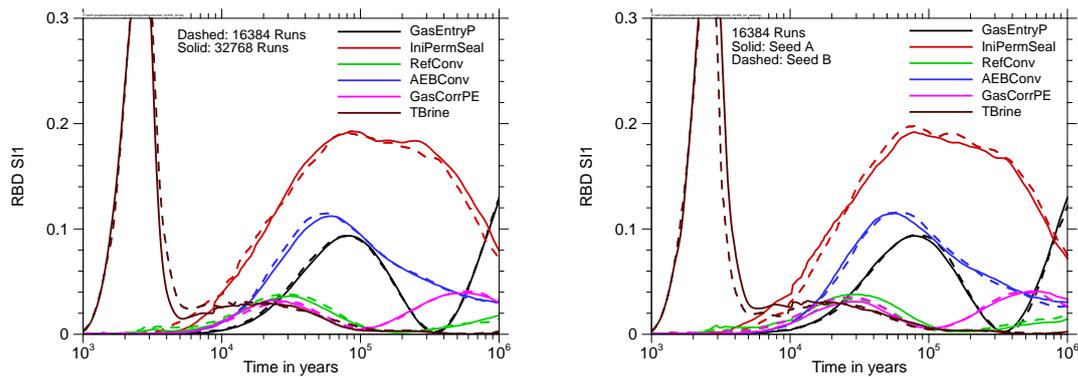


Fig. 6.7 RBD first-order indices for the clay model

### 6.3.2 The LILW system

The RBD results for the LILW6 model are presented in Fig. 6.8. The RBD sampling is subject to a random element, represented by the applied permutations. As for the EFAST sampling, the influence of this random element was investigated by changing the seed of the random number generator. The results are compared on the right side of the figure. The agreement is fair and much better than for the EFAST results (Fig. 6.6). Obviously, the RBD method is less susceptible to strong nonlinearities. This is probably, at least in part, due to the fact that the RBD sampling provides a more homogeneous coverage of the parameter space.



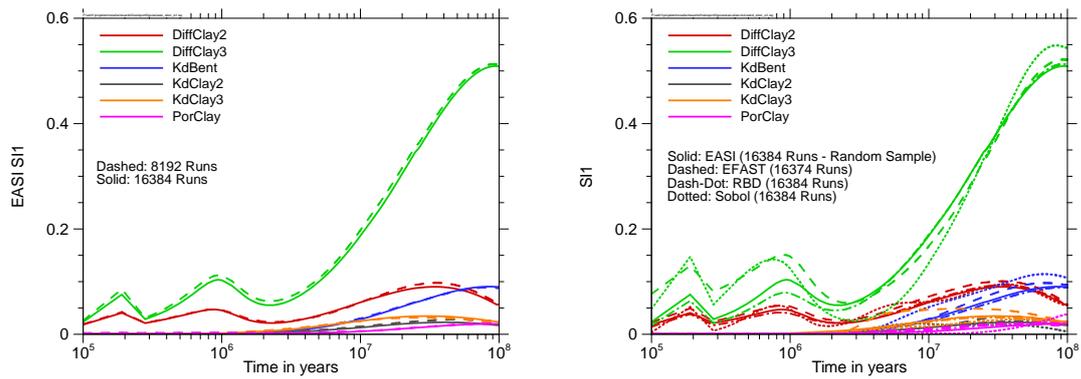
**Fig. 6.8** RBD first-order indices for the LILW6 model, compared for different sample sizes (left) and different random seeds (right)

## 6.4 EASI

In contrast to the variance-based methods presented so far, EASI does not need a specific sampling but can principally be applied to model results calculated with any sample. The only requirement is that the sample should be free of systematic couplings between the values of the individual parameters, as any kind of order in the sample can lead to unwanted side-effects. The influence of the sampling scheme is the subject of chapter 7. For the investigations presented here, random samples were used.

### 6.4.1 The clay system

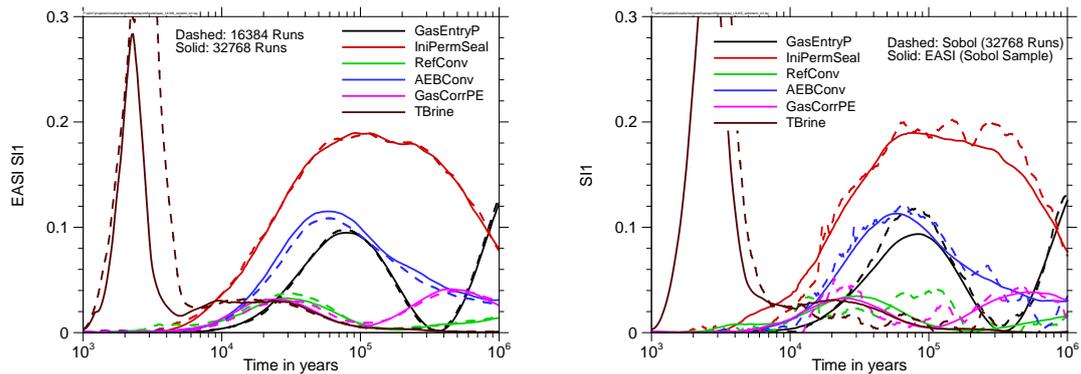
The SI1 time curves obtained with EASI for the clay model are presented and compared to the corresponding results of the other variance-based methods in Fig. 6.9. The convergence between the sample sizes of 8192 and 16384 is good, and the agreement with the other methods is fair. There is a close agreement especially with the RBD results, which is probably due to the fact that both methods use basically the same idea and differ only with respect to the sampling.



**Fig. 6.9** EASI first-order indices for the clay model

### 6.4.2 The LILW system

EASI results for the LILW6 model are presented in Fig. 6.10. The convergence between the sample sizes of 16384 and 32768 is good, except from the very early phase, where the model output is dominated by relatively few values. In the right part of the figure a comparison with the Sobol' calculation is made (see Fig. 6.2). Here the Sobol' sample was also used for calculating EASI. Interestingly, EASI produces smoother curves and obviously more stable results than Sobol', even if a sample is used that was specifically drawn for Sobol'.



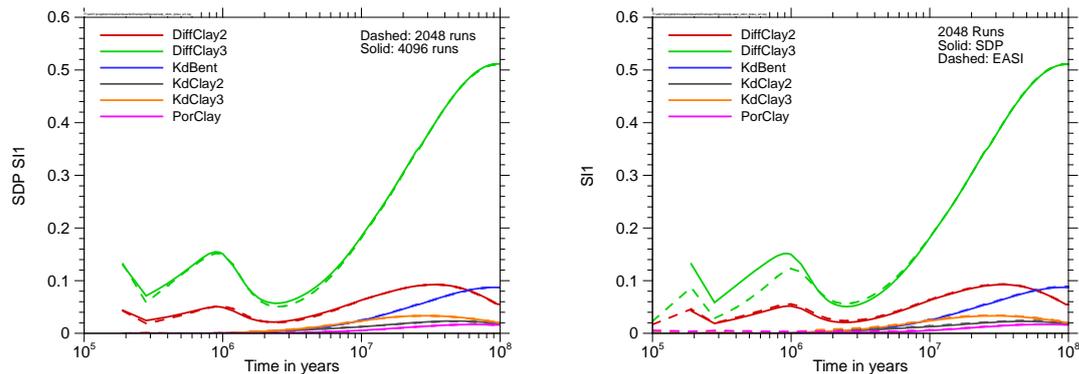
**Fig. 6.10** EASI results for the LILW6 model, compared for different sample sizes (left) and with the Sobol' calculation (right, the Sobol' sample was also used for EASI)

## 6.5 SDP

The SDP method is an integrated metamodeling approach and can be calculated with a specific software tool by Ratto and Young (/RAT 07/). Like EASI it is only able to calculate first-order indices and does not require a specific sampling scheme.

### 6.5.1 The clay system

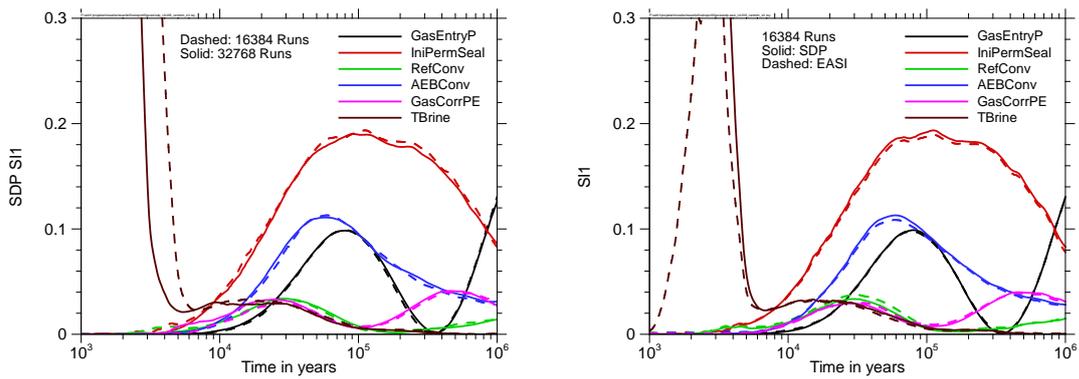
The SDP results for the clay system are presented in Fig. 6.11. On the left side of the figure, sample sizes of 2048 and 4096 are compared, on the right side compares the results to those of EASI. There is obviously a fair convergence already with the relatively low sample size of 2048. Moreover, SDP seems to calculate nearly the same results as EASI, except from the parameter *DiffClay3* in the early time phase, which is probably due to the low number of non-zero results in this phase.



**Fig. 6.11** SDP results for the clay model on the basis of random samples, compared for different sample sizes (left) and with EASI (right)

### 6.5.2 The LILW system

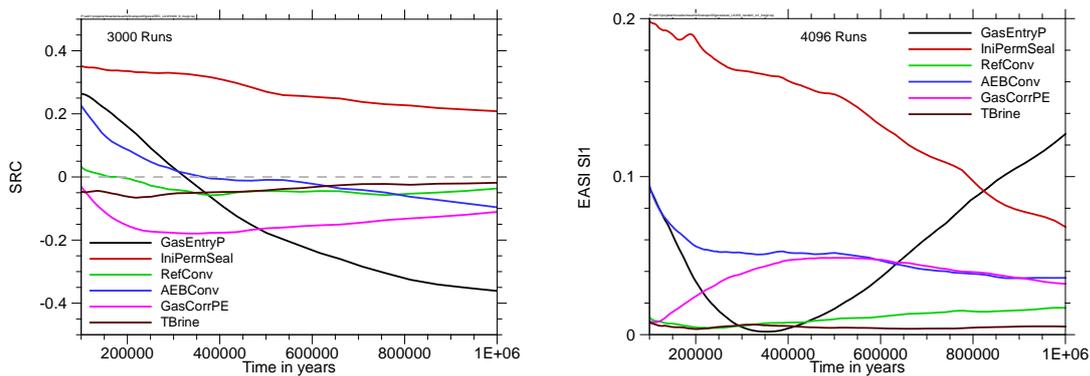
The SDP results for the LILW6 model are presented in Fig. 6.12. The curves look smooth and like in the clay case, the convergence is quite good. In the right part of the figure, the SDP results are compared to those obtained with EASI. Again, except from the very early phase, both methods seem to yield very similar results. This is specifically interesting in view of the computing times for SDP (about one day) and for EASI (a few seconds).



**Fig. 6.12** SDP results for the LILW6 model on the basis of random samples, compared for different sample sizes (left) and with EASI (right)

### 6.6 Comparison with regression-based sensitivity analysis

Basically, the results of variance-based sensitivity analysis confirm those of the regression-based analysis; striking contradictions were not found. The variance-based analysis even provides less information as it cannot resolve the direction of influence. Nevertheless, there is some added value of this kind of evaluation, which can be seen in Fig. 6.13. The figure shows a time detail of correlation-based and variance-based sensitivity evaluation of the LILW6 system.



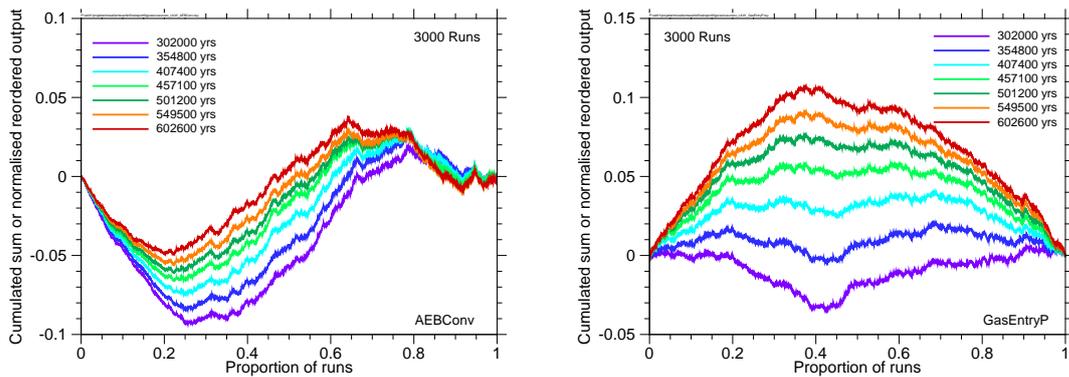
**Fig. 6.13** Comparison of correlation-based (3000 runs, left) and variance-based (4096 runs, right) sensitivity analysis for the LILW6 system

In the time phase between 300 000 and 600 000 years, the SRC of the parameter *AEBConv* is very close to zero, which could be misinterpreted as a low relevance of the

parameter. During this time phase the absolute SRC values of the parameters *RefConv*, *TBrine* and especially *GasCorrPE* are much higher than that of *AEBCConv*. The influence of *AEBCConv*, however, is not at all irrelevant as the EASI evaluation shows. The SI1 of *AEBCConv* is in fact nearly constant, but clearly different from zero, similar to that of *GasCorrPE* and much higher than those of *RefConv* and *TBrine*. The reason for this difference is that the model is non-monotonic with respect to *AEBCConv* and during the considered time phase the positive and negative influences compensate each other in the linear evaluation. By contrast, the SRC of *GasEntryP* intersects the zero line at about 320 000 years, and the SI1 value is indeed zero at the same time. This means that there is in fact no significant influence of *GasEntryP* to the model output around 350 000 years.

The CUSUNORO plots in Fig. 6.14 confirm these findings. While the curves for *AEBCConv* reverse their curvature, which means that the direction of influence changes over the parameter interval, those for *GasEntryP* are more or less uniquely curved, but change the curvature sign over time and the curve for 350 000 years is nearly flat.

This investigation shows that in fact there can be an added value of a variance-based sensitivity analysis compared to a regression-based one in that it is able to identify non-monotonic influences.



**Fig. 6.14** CUSUNORO plots of the LILW6 model for the time period between 300 000 and 600 000 years for the parameters *AEBCConv* (left) and *GasEntryP* (right)

## 6.7 Summary

In the context of the project described here, five methods for calculating variance-based sensitivity indices were tested with the clay system and the LILW system: Sobol', EFAST, RBD, EASI and SDP. While Sobol' and EFAST yield the first- and total-order indices, RBD, EASI and SDP are only able to calculate the first-order indices.

As all methods estimate the same sensitivity measures, their results should, by theory, be similar, and one would expect that with increasing number of model runs they converge to the "true" values. It was, however, found that, even with 16 000 to 32 000 runs and merely 6 parameters, the convergence was partly unsatisfying, especially for the highly nonlinear LILW6 model. Although all methods principally show the same tendencies, their calculated time curves do not quite agree in detail.

The Sobol' method in the investigated form (as implemented in SimLab 3.2) produced the least stable results and does not inspire much confidence. Therefore, the method was excluded from further investigations in this project. This, however, must not be misunderstood as a concluding assessment of the Sobol' method in principle.

The EFAST method produced fairly convergent and plausible results for the clay model. For the LILW6 model, however, the obtained results look less stable and, even for high sample sizes, seem to depend on the random seed used for the sampling. Obviously, EFAST is less appropriate for models with strong nonlinearities or even discontinuities. As Sobol' was excluded, it is, however, the only remaining method that allows determination of total-order indices.

RBD calculated fairly convergent, smooth and plausible time curves for the first-order indices of both models. For the highly nonlinear LILW6 model the method seems more appropriate than EFAST, but it does not provide total-order indices and also requires a specific sampling.

For both models, the results obtained with EASI are very similar to the RBD results. EASI, however, works with any sampling scheme, so that one is not bound to use model calculations specifically made for the intended evaluation. If applied with a Sobol' sample, EASI produces obviously "better" results than the Sobol' method itself. Moreover, EASI is very quick.

SDP is an integrated metamodeling approach. Like EASI, it can be applied to any set of probabilistic model calculations on the basis of any sample. For the LILW6 model, if calculated on the basis of the same sample, the SDP results are in close agreement with the EASI results. SDP, however, uses a complicated algorithm and requires a specific software program, which needs a lot of time for the evaluation, so that SDP is by orders of magnitude slower than EASI.

The results obtained in the project do not imply any preference of either RBD or SDP over EASI. An essential drawback of all three methods, however, is that they are unable to calculate total-order indices.

Principally, variance-based sensitivity analysis seems appropriate to analyse the sensitivities of the investigated models. Compared to regression-based sensitivity analysis, the first-order indices often do not seem to provide additional qualitative information, but possibly give a better impression of the quantitative influence of the individual parameters for nonlinear models. Especially non-monotonic influences are better detected by variance-based analysis. Total-order indices can provide insight to the interactions of the parameters.

It has to be kept in mind, however, that typically the output of repository models is roughly logarithmically distributed over several orders of magnitude, and variance is not best appropriate to quantify variations on such a scale, as it gives an overweight to the highest values. This problem is further addressed in chapter 9.



## **7 Influence of the sampling scheme**

In chapter 2.3 five different sampling strategies were introduced and compared in view of homogeneity. Some sensitivity analysis methods, like Sobol' and EFAST, require their specific sampling scheme. One has to decide about the application of such methods at the very beginning of the probabilistic analysis, which can lead to results that are then inappropriate – or at least less appropriate – for other evaluations. Moreover, such samples are normally not extendible, which means that for an investigation with a higher number of runs a totally new sample has to be drawn and existing results from the preceding investigation cannot be re-used. It is therefore generally preferable to resort to analysis methods that work with any sample.

For a pure uncertainty analysis, random sampling is principally the best choice as it allows unbiased statistical statements about the model output. Sensitivity analysis, however, does not aim at statistics, but at quantifying the relations between the model input and output. For this purpose a homogeneous coverage of the parameter space is more important than randomness. In this chapter we investigate how the applied sampling scheme influences the results of sensitivity analysis.

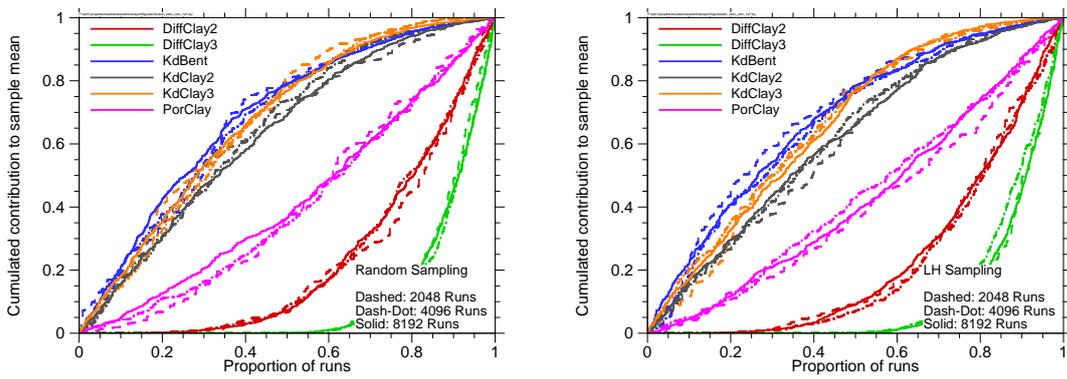
As Sobol', EFAST and RBD anyway require their specific sampling, these methods are not appropriate for comparing sampling schemes. Therefore, in the following only results obtained with methods that can be applied with any sampling scheme are used.

### **7.1 Random sampling versus Latin Hypercube Sampling**

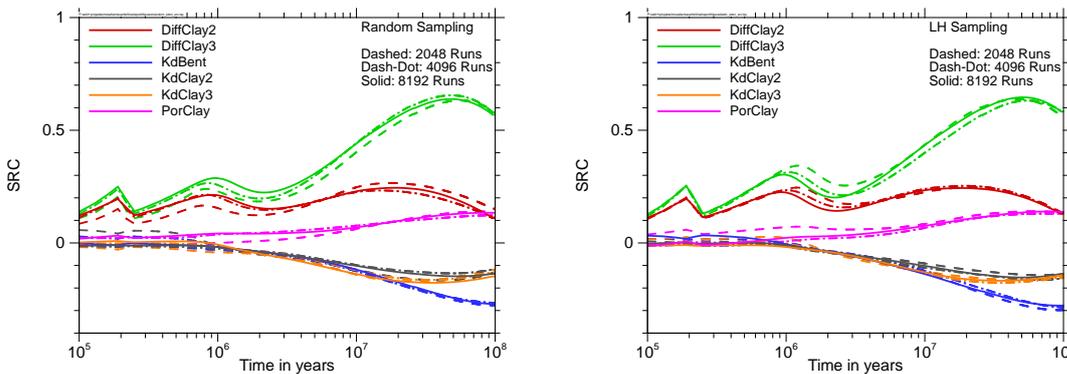
Latin Hypercube sampling (LHS, see chapter 2.3.2.1) is a well-known and often recommended sampling method, which is sometimes assumed to be superior to random sampling for sensitivity analysis, due to its property to produce more balanced marginal distributions (see Fig. 2.4, right). This, however, does not automatically mean a more homogeneous coverage of the multi-dimensional parameter space.

To test LHS in comparison with random sampling, both methods were used for drawing samples of sizes 2048, 4096 and 8192 for the clay model system, and some sensitivity analysis evaluations were performed. These are presented in the following.

Fig. 7.1 shows CSM curves for  $1.0 \cdot 10^7$  years, calculated on the basis of random and Latin Hypercube samples with 2048, 4096 and 8192 runs. There is no evidence for better performance of one or the other method with respect to convergence of the results. There are still clearly visible differences of about the same dimension between the corresponding curves for all three sample sizes in both cases. Principally the same conclusion is to be drawn from the SRC evaluation shown in Fig. 7.2. LHS seems to yield, if at all, only slightly better convergent time curves than random sampling.



**Fig. 7.1** CSM curves for the clay system at  $1.0 \cdot 10^7$  years, calculated on the basis of random samples (left) and Latin Hypercube samples (right) of different sizes

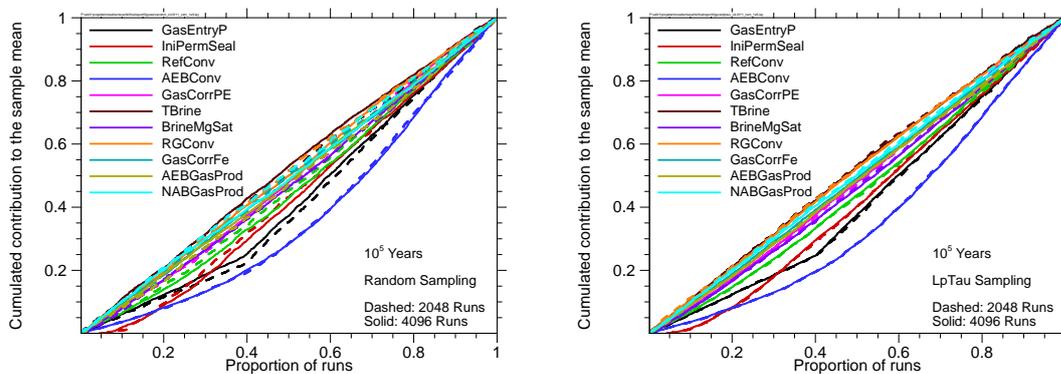


**Fig. 7.2** Time-dependent SRC evaluation for the clay system on the basis of random samples (left) and Latin Hypercube samples (right) of different sizes

## 7.2 Random sampling versus quasi-random LpTau sampling

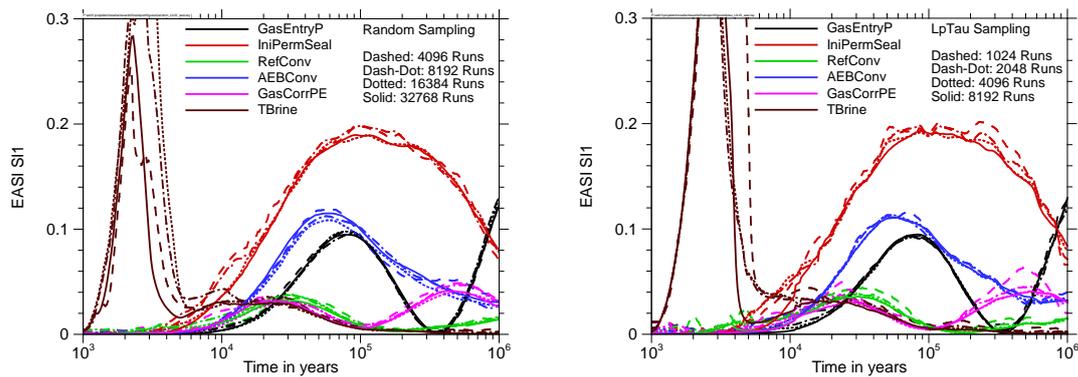
Quasi-random sampling means a sampling strategy that is fully deterministic and specifically designed to cover the parameter space as homogeneously as possible (see chapter 2.3.3). For our investigations we chose the Sobol' LpTau method and the LILW system.

In Fig. 7.3 CSM curves for the time  $10^5$  years are compared. The curves were generated from results of the LILW11 model, obtained with random and LpTau samples of sizes 2048 and 4096. While there are still visible differences between the corresponding curves for the random samples, those for the LpTau samples are in much closer agreement. Obviously, the convergence with increasing sample size is better if LpTau sampling is applied. This is the more remarkable as it is sometimes claimed that quasi-random sequences are superior to random sequences only in low-dimensional spaces. If, however, some of the parameters only have a weak influence on the model output, they do not “fully count”. A better criterion is the average dimension, which is equal to the sum of all total-order sensitivity indices ( $/SOB 16/$ ). Without a mathematical proof, Sobol' claims that quasi-random sampling outperforms random sampling as long as the average dimension is less than about 3.



**Fig. 7.3** CSM plots for the LILW11 system, based on random (left) and LpTau (right) samples of sizes 2048 and 4096

In Fig. 7.4 EASI time curves for the LILW6 system are presented, calculated on the basis of random and LpTau samples of different sizes, but the LpTau samples are by a factor of 4 smaller than the random samples. Nevertheless, the convergence with increasing samples sizes is similar, maybe even slightly better, for the LpTau curves.



**Fig. 7.4** Time-dependent EASI evaluation for the LILW6 system on the basis of random samples (left) and LpTau samples (right) of different sizes

### 7.3 Summary

From the investigations on sampling schemes, it can be concluded that Latin Hypercube sampling does not provide a significant advantage to random sampling with respect to the required number of runs for sensitivity analysis. This statement is valid for the investigated clay system, but since it is a relatively simple model with no strong nonlinearities and only six parameters, there is no reason to assume that LHS is significantly superior to random sampling for more complex models.

Quasi-random LpTau sampling, however, is obviously superior to random sampling in that it produces convergent results at significantly lower sample sizes. For the LILW6 model LpTau seems to yield comparable convergence at sample sizes that are by a factor of 4 smaller than for random sampling. This means that with LpTau sampling only one fourth of the model runs are necessary that are needed for similar robustness with random sampling. One should keep in mind, however, that this depends on the number of influential parameters. The average dimension, which can be calculated as the sum of all total-order indices, is a measure for this. As a rule of thumb, if it does not exceed 3, one can expect advantages from quasi-random sampling.

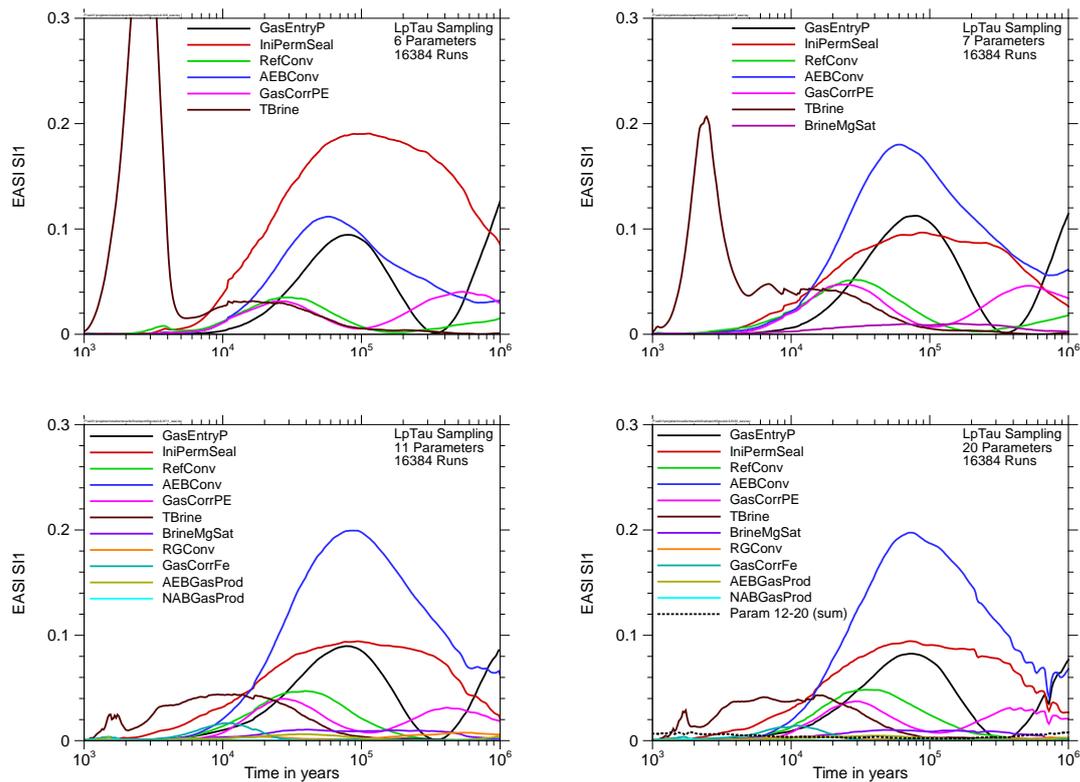
## 8 Detailed sensitivity investigations

In this chapter we present specific investigations performed on the LILW system. It turned out that, due to its complexity, the sensitivity analysis results exhibit some details that are hard to understand at the first sight.

### 8.1 Influence of number of parameters

In Fig. 8.1 the EASI time curves for the LILW model with 6, 7, 11 and 20 parameters (LILW6 to LILW20) are presented. All calculations are based on 16384 runs from LpTau samples. One should recall that all four models are numerically identical and only differ by the number of varied parameters. The parameters that were not varied were fixed to their reference value (see Tab. 3.2). A general observation is that the total variance of the model output is dominated by only a few parameters while the many remaining ones have only a negligible influence. This is an often-reported property of numerical models describing practical problems (/SAL 08/). Moreover, there are three facts that attract attention:

1. The corresponding curves for the LILW6 model and the LILW7 model, especially those of the parameters *IniPermSeal* and *AEBConv*, look conspicuously different. These two parameters obviously exchange their ranking positions. While under the migration from 6 to 7 parameters the SI1 value of *IniPermSeal* is reduced by a factor of 2 in its maximum, that of *AEBConv* increases by a similar factor. The curves of *GasEntryP*, *RefConv* and *GasCorrPE* show slighter differences. The differences are surprising as the variable additionally varied seems to have no significant influence at all; its SI1 reaches a maximum of only about 0.01, so that one would expect that it hardly matters whether this parameter is varied over its uncertainty interval or fixed to any value within. This, however, is obviously a wrong conclusion, see chapter 8.2.

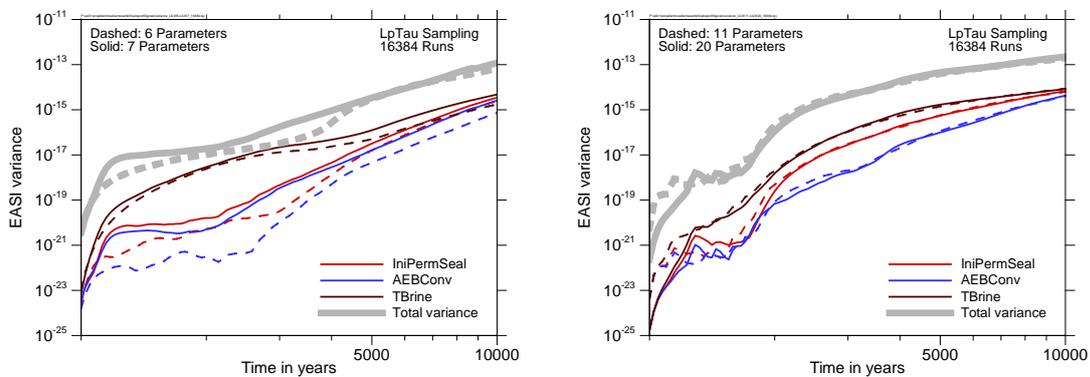


**Fig. 8.1** Time-dependent EASI evaluation for the LILW model with 6, 7, 11 and 20 parameters; for LILW20 the sum of the SI1 values of the parameters 12 – 20 is shown as a black dotted line

2. The high peak of the SI1 curve of the parameter *TBrine* at very early times seems to be considerably reduced if additional parameters are taken into account. While for LILW6 its early maximum reaches 0.44 (not visible in 8.1), it is only 0.21 for LILW7 and decreases further to values below 0.03 for LILW11 and LILW20. At such early times there are only very few cases with a non-zero model output, and this depends on *TBrine* alone, since contaminant release can and will only arise after the time defined by *TBrine*. Other parameters can only influence the value of the model output if it is allowed by *TBrine*. These are higher-order effects. It is therefore well understandable that the only parameter with a significant first-order effect at very early times is *TBrine*. As there are so few non-zero values at early times, the statistical basis is small and the numerical evaluation is considerably influenced by random effects. These cannot actually be assigned to specific parameters but appear like a blurred interaction of all parameters in producing the output variance. Numerically, this leads to increased higher-order contributions. Consequently, the first-order effect of *TBrine* seems to decrease if

more parameters are taken into account. To better visualise this, Fig. 8.2 shows the time development of the total variance and the absolute contributions of three parameters during the early phase for different numbers of parameters taken into account. The first-order index calculates as the quotient of the parameter's contribution and the total variance, which can be seen in the logarithmic plot as the distance between the curves. The total variance increases by 8 to 9 orders of magnitude between 1000 and 10000 years, which is due to the rapid increase of non-zero results.

Therefore, one should not put too much confidence in the early-phase evaluation.



**Fig. 8.2** Total variance and contributions of three parameters during the early time phase; LILW6 vs. LILW7 (left) and LILW11 vs. LILW20 (right)

3. The curves seem to become more unstable with increasing number of varied parameters. This is mainly visible between the curves for the LILW11 model and the LILW20 model, which hardly differ except from this effect. Although none of the additional parameters perceptibly contributes to the total output variance, they all together seem to have some more or less random-dominated influence – indicated by a dotted black line in 8.1 – that visibly disturbs the evaluation for the more important parameters.

From these investigations one can conclude that

- too many varied parameters that have no significant influence can disturb the stability of sensitivity analysis results,
- even if a parameter looks insignificant by its own, it can have essential influences on the sensitivities of the others,

- if the statistical basis is weak, a high number of insignificant parameters can mask actual sensitivities.

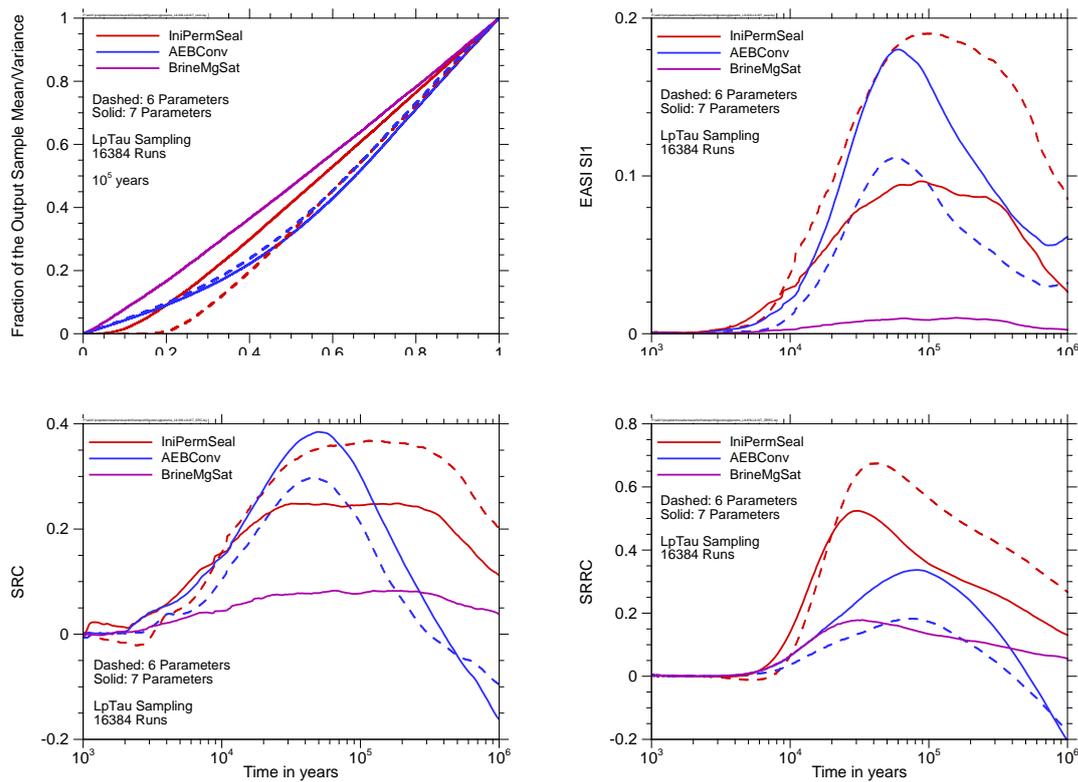
Therefore, the number of parameters taken into account in sensitivity analysis should be thoroughly considered.

## 8.2 Hidden sensitivities

The conspicuous differences in the time development of the sensitivity indices calculated for the LILW model with 6 and 7 parameters indicate that there is some “hidden sensitivity”. This is shown more clearly in Fig. 8.3 by means of different sensitivity plots. For better clarity, only the two most influential parameters *IniPermSeal* and *AEBCConv* and the parameter *BrineMgSat*, which is the one additionally varied in the LILW7 model, are presented in the plots.

In all four plots substantial differences between the curves for the LILW6 and the LILW7 model are visible. These are obviously caused by the variation of the parameter *MgBrineSat*, which itself has only a low sensitivity according to all methods.

All four plots show the same tendency: If the additional parameter is varied the sensitivity of the system to *IniPermSeal* decreases, while the sensitivity to *AEBCConv* increases. The methods do not agree, however, about the sensitivity ranking of the parameters. According to EASI, the additional parameter causes an exchange of the ranking positions of the two leading parameters over nearly the total time period, which is only in part the case according to SRC and not at all according to SRRC. However, as explained in chapter 6.6, a low SRC or SRRC value of the parameter *AEBCConv* does not necessarily mean a low sensitivity.

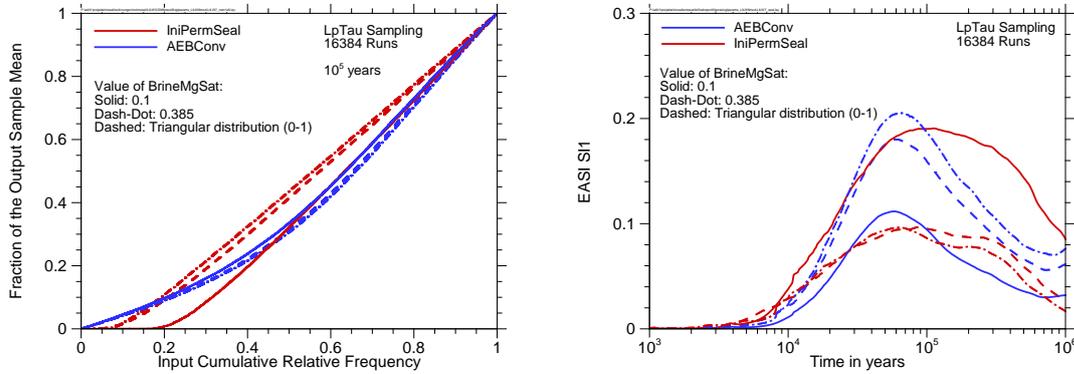


**Fig. 8.3** Comparison of sensitivity plots for the LILW6 and the LILW7 model: CSM (top left), EASI (top right), SRC (bottom left), SRRC (bottom right)

Physically, *MgBrineSat* characterises the relative saturation of magnesium in the corrosive brine in the repository mine. As magnesium is the critical element for corrosion of the seal material, the magnesium content of the brine can influence the corrosion speed and the time of seal failure. It is assumed that, according to the conditions in an abandoned mine for salt and potassium production, the brine has a chemical composition between saturated NaCl solution and IP21 solution. The relative magnesium saturation is therefore related to the magnesium content of IP21, so that the range of possible values is between 0 and 1. The reference value is assumed to be 0.1, which is also taken as the modulus of the non-symmetric triangular distribution.

One has to note that the reference value of *BrineMgSat* is distinctly different from the expectation value of the distribution, which is 0.37. Consequently, the LILW6 model uses the constant value 0.1 for all runs while the LILW7 model uses values between 0 and 1 with an average of 0.37. Therefore, a calculation was made with a modified LILW6 model, using not the reference value but a fixed value close to the expectation

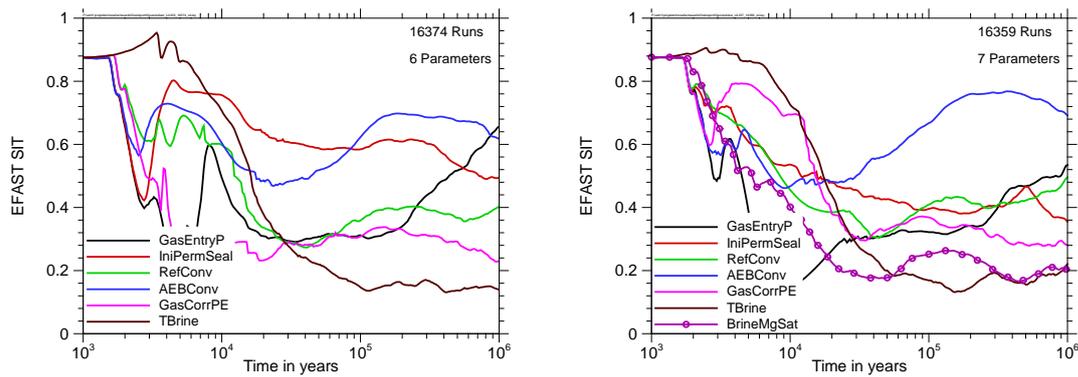
of the distribution (by technical reasons, the value 0.385 was used). The results are presented in Fig. 8.4 as CSM plots and EASI time curves.



**Fig. 8.4** Comparison of sensitivity plots for the original and the modified LILW6 model with those of the LILW7 model: CSM (left) and EASI (right)

The Fig. shows that there are only minor differences between the calculated sensitivities of the modified LILW6 model and the LILW7 model. Obviously, the increase of the mean of values for *BrineMgSat* by more than a factor of 3 is responsible for the change of the sensitivity analysis results. This can be understood from the physical point of view: Higher values of *BrineMgSat* lead to earlier seal failure, so that the probability that the seal has failed at a specific point in time is generally much higher than with the small reference value of *BrineMgSat*. Consequently, the value of the second parameter that influences this probability, the initial seal permeability *IniPermSeal*, becomes less important. Once the seal has failed, *IniPermSeal* has only little influence to the release of contaminants, which is then dominated by the convergence rate of the sealed emplacement area, *AEBCConv*. At the time 10<sup>5</sup> years, for instance, the seal has failed in about 85 % of the runs of LILW7, but only in 75 % of the runs of LILW6.

Thus, the influence of *BrineMgSat*, although surprising at the first sight, is plausible. Nevertheless one would expect this influence to be deducible from the calculated sensitivities for the parameter itself, which does not seem to be the case. As, according to the understanding, the sensitivity is due to an interaction with the parameter *IniPermSeal*, it might be visible in a second-order analysis, which was not performed due to technical reasons. A total-order analysis, however, was made with EFAST. In Fig. 8.5 the EFAST total-order results are compared between the LILW6 model and the LILW7 model.



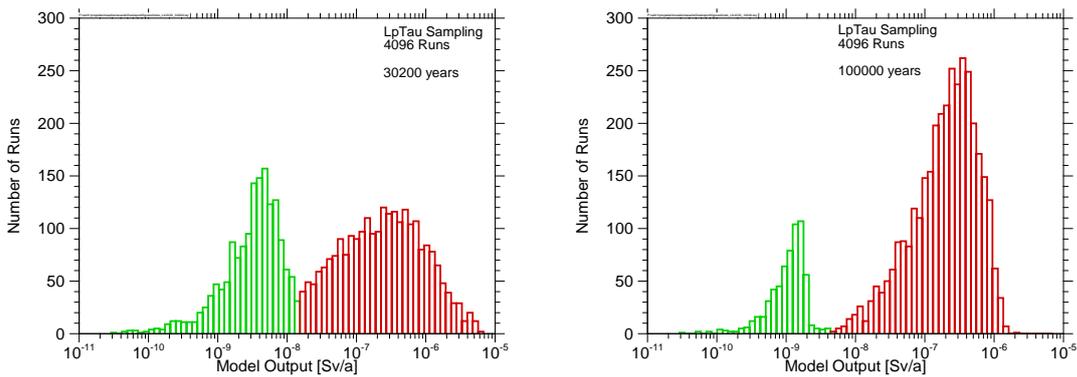
**Fig. 8.5** EFAST total-order analysis of the LILW6 (left) and the LILW7 (right) model; for better visibility the curve for *BrineMgSat* is marked by circles

The results for the early time phase must not be taken seriously because of the low number of non-zero runs. From about 20 000 years on, however, they should be satisfyingly reliable. In agreement with the explanation given above, it is confirmed that the effect of *IniPermSeal* decreases and that of *AEBConv* increases if *BrineMgSat* is additionally varied, but the analysis does not show any conspicuousness with respect to the total-order index of this parameter itself. Its total-order index is about 0.2 for most of the model time, which is a relatively low value for total-order.

From this investigation we conclude that a complex model can have relevant sensitivities that show up neither in correlation-based sensitivity analysis nor in first- or total-order variance-based sensitivity analysis.

### 8.3 Handling of two-split output by a dynamic two sample test

The annual dose calculated by the LILW model increases typically by about two orders of magnitude after the seal failure. On the considered time scale, this happens nearly instantly at some point in time, depending on the parameters. Therefore, at any specific point in time, in a certain percentage of model runs the seal has already failed, and this percentage increases with time. Consequently, the distribution of model output values is typically two-split. The histograms presented in Fig. 8.6 clearly show this separation. They were made for the LILW20 model.



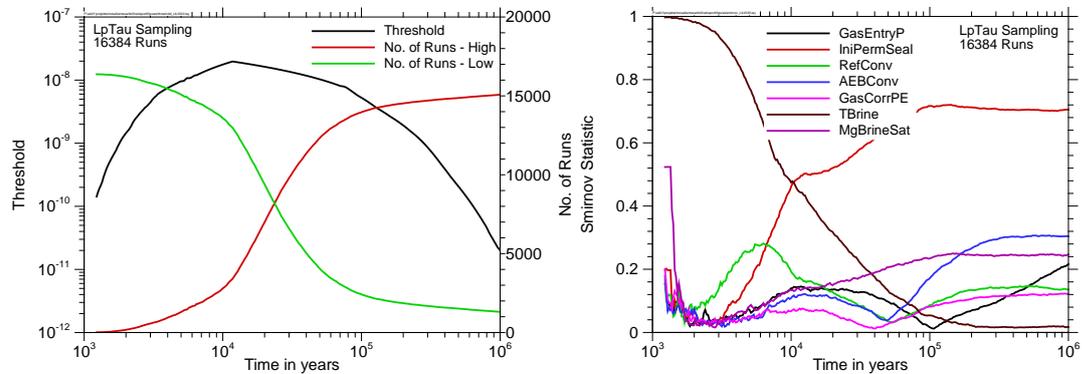
**Fig. 8.6** Histograms for the LILW20 model for two points in time; calculation cases with (presumably) intact seal are presented in green, those with failed seal in red

Actually, there will be a certain overlap of the two sub-distributions, but this cannot be resolved by means of the available model output, since the model does not provide information about the status of the seal. Therefore, the threshold was determined by visual inspection, accepting that it cannot be quite exact.

This specific property makes the model a bit problematic for sensitivity analysis and is probably the main reason for the difficulties that arise with the standard methods. An interesting question is, of course, which parameters are most responsible for the output value to lie in the “lower” or in the “higher” part of the histogram.

In order to analyse this question, we developed the following idea: Since the output distribution is two-split “by nature”, it seems predestined for a two-sample test (see chapter 2.8.1). Therefore, we applied a modified version of the Smirnov sensitivity test. In the standard version of this test, the output data are separated in two sub-samples at the 90<sup>th</sup> percentile of the distribution; then the corresponding distributions of the input parameters are compared. If there is a significant difference, the respective parameter has a notable influence on whether or not the model output value will be among the highest 10 %. As in this case, the question is whether the output value will be in the lower or in the upper part of the split distribution, we used a dynamic, time-dependent threshold value instead of a rigid separation at the 90<sup>th</sup> percentile. The temporal progression of the threshold value, which was determined manually by visual inspection, is presented in Fig. 8.7 (left) together with the numbers of model runs that were thus as-

signed to the lower and the upper part of the distribution. The right part of the figure shows the results of the Smirnov test performed on this basis.



**Fig. 8.7** Time-dependent dynamic Smirnov sensitivity evaluation of the LILW20 model with: threshold value and numbers of runs (left) and results of the Smirnov test (right, only the 7 most important parameters are presented)

This evaluation was expected to indicate which parameters are most influential to the question whether or not the seal has failed at each point in time. In fact, the results differ from those obtained with other methods in that they show a clear dominance of *IniPermSeal* and actually a relatively high influence of *BrineMgSat*, which was not identified as relevant by any other method. In the time range between about 40 000 and 160 000 years this parameter is even the second most important one according to this kind of analysis. This is a plausible result in view of the explanations given in chapter 8.2.

## 8.4 Summary

The investigations described in this chapter have shown that for a complex nonlinear model like the LILW system a specifically designed, detailed sensitivity analysis can reveal interesting facts about the system that cannot be seen at standard sensitivity analysis results.

Generally, the relevant sensitivities are confined to a low number of parameters, and in view of stable and robust results it seems to be a good idea to keep the number of varied parameters as small as possible. There is, however, some danger that a quick screening of the parameter sensitivities by means of standard methods, including total-

order analysis, leads to exclusion of parameters that in fact have interesting influences to the model results. Analysing such influences can help increase the model understanding.

If some parameter is, from a physical point of view, suspected to be influential but does not show up in the sensitivity analysis it can be helpful to perform several probabilistic sets of model runs with and without the respective parameter being varied, or with several different fixed values of the parameter, and to apply the same sensitivity analysis methods to all of the sets individually. If the results are similar, the influence of the parameter is obviously actually low; otherwise one should try to explain the differences and, if possible, perform specifically designed investigations.

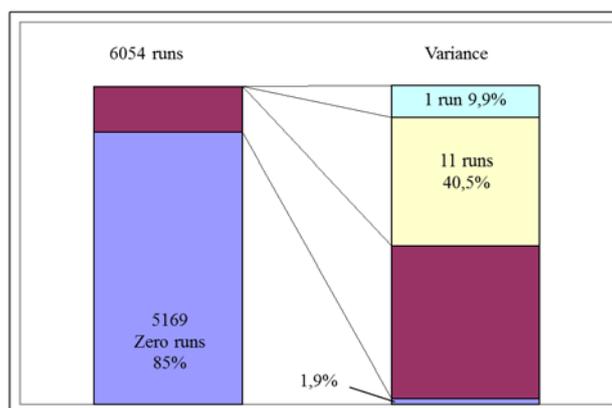
A model like the LILW model that produces distinctly split output distributions seems predestined to be analysed with a two-sample test like the Smirnov test. Such tests analyse the influences of the individual parameters to the tendency of the model to run in one or the other direction. This makes sense specifically if the separation of the two subsets of model output values is not made according to a fix and more or less arbitrary threshold, but on the basis of the model output itself. For a time-dependent analysis the threshold has to be calculated for each point in time separately.

## 9 Output transformation

The output of repository models is typically distributed over several orders of magnitude. While the highest values are most relevant for assessing the safety of the system, the lower ones are of minor interest. Especially models for repositories in rock salt can even yield zero results. This is often a consequence of a model situation in which all waste-containing parts of the mine are closed by convergence before brine could get in contact with waste.

On a linear scale, the distribution of the model output values is typically distinctly right-skewed, which means that there is a clear concentration of values near zero and a long tail of relatively few higher values. Correlation-, regression and variance-based methods of sensitivity analysis are mathematically designed to work on a linear scale as they make use of linear concepts. Therefore, such evaluations are often dominated by a few high values, so that a sensitivity analysis performed on a linear scale does not necessarily provide the most meaningful results.

The dominance of very few values is especially pronounced for variance-based sensitivity analysis, as illustrated in Fig. 9.1 for the SAM-GBS model. This model was used since it was of specific interest how the transformation technique works on a set of output values with a high number of zeros. It produces about 85 % zero output runs, which together make only 1.9 % of the total variance. On the other hand, the highest one of more than 6000 runs makes nearly 10 % of the total variance and the 0.2 % of all runs with the highest results are responsible for more than 50 % of the total variance. In variance-based sensitivity analysis, such disproportion can not only lead to results of low robustness and stability, but also to an unwanted strong overweighing of the highest values.



**Fig. 9.1** Contributions to the total variance of the results of the SAM-GBS system

Sometimes, it is recommended in such cases to use the logarithm of the model output for analysis. This can be a sensible approach if the model output is actually distributed over a limited interval on the logarithmic scale, but for a typical repository model a simple logarithmic transformation might not solve the problem but turn it into the opposite. Zeros are impossible to take into account on a logarithmic scale, and moreover, repository models sometimes yield output values that are very low and physically meaningless but determined by numerical effects. As a consequence, instead of the high values the lowest ones are overvalued and can distort the sensitivity analysis.

One transformation that is in wide use in sensitivity analysis, especially for correlation- and regression-based methods, is the rank transformation, which replaces each value with its rank in an ordered list (see chapter 2.6.4). This is automatically applied by the Spearman's rank correlation method and the SRRC method; it is, however, unusual to apply it for variance-based sensitivity analysis. The rank transformation linearises the relationship between model input and output, which can also give overweight to small values. An additional drawback of the rank transformation is that it is not reversible and destroys information contained in the original data.

Therefore, instead of the logarithmic or rank transformation, we look for transformations that are better adequate to the typical output distributions of repository models.

## 9.1 Definition of transformations

The basic idea of applying a transformation is that different model output values be differently weighed, according to magnitude, in sensitivity analysis. While the undue

overweight of high values is to be mitigated, the transformation should not, at the same time, give such an overweight to low values. A transformation that fulfils this requirement can only be defined if a distinction between “low” and “high” values is given. In order to provide such a distinction, we introduce a threshold  $a$ , which has to be defined sensibly. The transformation, however, shall not handle “low” and “high” values differently, but be designed as a continuous, bijective function. Thus it is made sure that the transformation is uniquely reversible and does not destroy information as the rank transformation does.

Assuming that the model output is never negative, we require that the transformation shall

- be strictly monotonic,
- map 0 to 0,
- map  $a$  to 1,
- map values  $\ll a$  to values near 0,
- map values  $\gg a$  to moderately increasing values above 1.

Such a transformation maps all values below  $a$  to the interval between 0 and 1.

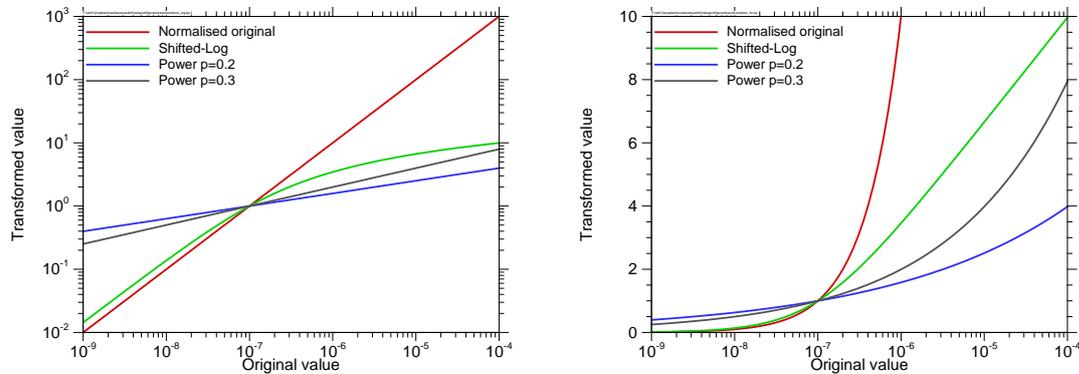
We made experiments with two types of transformations of this kind:

- shifted-logarithm:  $y \rightarrow \log_2(1 + y/a)$ ,
- power:  $y \rightarrow (y/a)^p$ .

The shifted-logarithm transformation uses the logarithm to the basis 2 in order to automatically map the value  $a$  to 1. For the power transformation, a positive exponent significantly below 1 should be chosen to fulfil the requirements. For the investigations presented here, exponent values of 0.2 and 0.3 were used.

The courses of the three transformation functions are visualised in Fig. 9.2 in comparison with the original data, divided by the threshold value, which was set to  $10^{-7}$  for demonstration. One can see that the shifted-logarithm transformation leaves low values nearly unchanged and behaves like the logarithm for high values. The power transfor-

mations, however, have a notable effect also to low, non-zero values, transferring them closer to 1.



**Fig. 9.2** Shifted-log transformation and power transformation with two different exponents, visualised with a logarithmic (left) and a linear y-axis (right)

## 9.2 Selection of the threshold

As all transformations depend on the threshold parameter  $\alpha$ , a proper selection of its value is essential. One possibility is to orient it at physical – or physiological – facts. If, for example, the model calculates the annual dose to a human individual one can choose a threshold that is related to the permissible limit, for instance, 1 % of it. This would mean that all values above 1 % of the limit are considered as “high” and are given an increasing weight in the sensitivity analysis.

If, however, the repository performance is so good that a threshold oriented at the permissible limit is never exceeded or even remotely reached, a transformation with such a threshold value will hardly make sense. In this case, it seems better to choose a threshold value that is oriented at the actual model output. One possibility is to set  $\alpha$  equal to the median of all output values. As the transformation is monotonic, it does not change the ranking so that the median of the transformed data is automatically 1, which is a nice mathematical property. With this choice, the upper half of the values is treated as “high”, the lower half as “low”. We call this a median-centred transformation. Another idea is to calculate the threshold such that after the transformation the arithmetic mean of the values becomes 1. This technique automatically leads to sensitivity

analysis results that are centred on the limit between “low” and “high”, but the calculation is numerically difficult.

If a non-constant threshold value is used for a time-dependent analysis, however, the threshold has to be determined individually for each point in time, and when interpreting the sensitivity analysis results one has to be aware that the threshold can vary considerably over time.

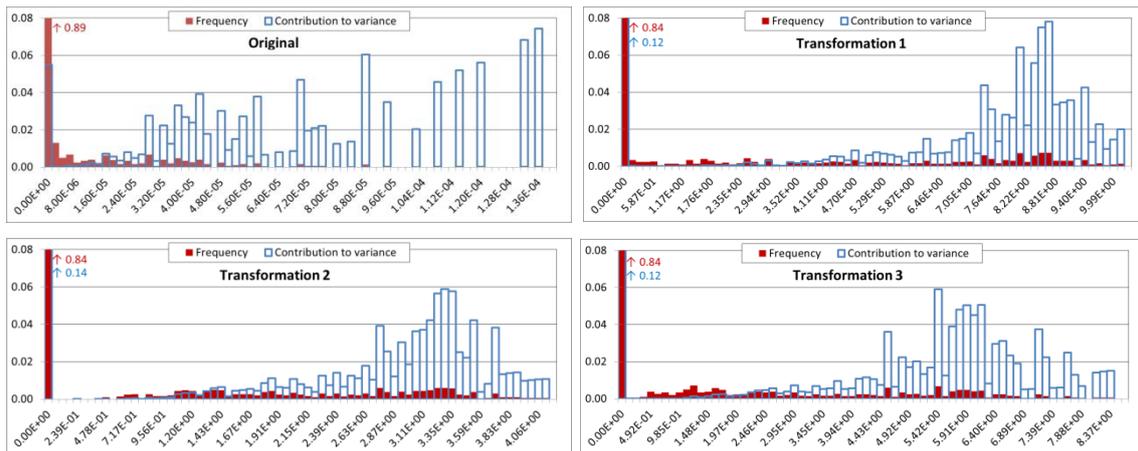
For the investigations the approaches with a constant threshold value and with a threshold equal to the median were tested.

### **9.3 Sensitivity analysis results**

The effects of the transformations to sensitivity analysis are demonstrated in the following with two models. The evaluations for the obsolete SAM-GBS model were done using an old set of calculation results.

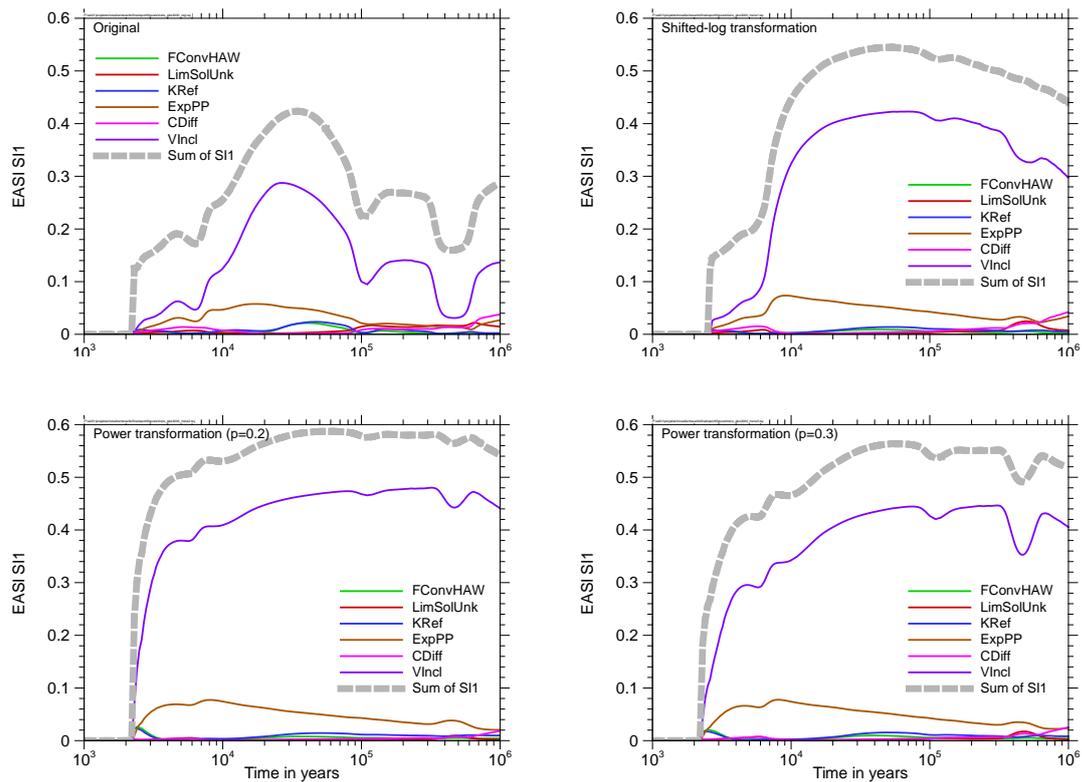
#### **9.3.1 The SAM-GBS model**

Fig. 9.3 shows how the three transformations change the distribution of the contributions to the total variance. The histograms are based on 3000 runs of the SAM-GBS model. While for the original data the variance is clearly dominated by very few high results – each of the five highest values contributes about as much to the variance as all zero and near-zero results together –, in the transformed data the contributions to the variance look much more balanced with regard to the relation of bin population and contribution to the variance, no matter which transformation is used.



**Fig. 9.3** Histograms for frequency and contribution to the total variance of 3000 runs of the SAM-GBS model (run maxima); calculated from the original data and after application of three transformations (1: shifted-log, 2: power  $p=0.2$ , 3: power  $p=0.3$ ) with  $a = 10^{-7}$  Sv/yr

Fig. 9.4 shows how the transformations work on a variance-based sensitivity analysis with EASI. As in the figure above,  $a = 10^{-7}$  Sv/yr was used as the threshold value. This is oriented at the German legal requirement that the radiation exposure to a human individual, originating from the repository, shall not exceed  $10^{-5}$  Sv/yr for probable scenarios (/BMU 10/). The threshold was chosen as 1 % of this value as proposed above.



**Fig. 9.4** First-order sensitivity indices calculated using EASI for the SAM-GBS model from the original and transformed data ( $\alpha = 10^{-7}$  Sv/yr); only the 6 most relevant of 31 parameters and the sum of all S11 are presented

The parameter *VIncl*, which represents the volume of an assumed brine inclusion in the host rock, dominates clearly in all evaluations. The dominance appears, however, much more pronounced, if any of the three transformations is applied. In the time phase between 380 000 and 560 000 years the first-order index of *VIncl*, calculated from the original data, even decreases below 0.04, exceeding those of three other parameters by a mere factor of two. In the transformed evaluations, there is only a slight dent visible in the S11 curves in this phase. Moreover, the parameter *ExpPP* is clearly the second most dominant in the transformed evaluations nearly for the total model time. Generally, the transformations seem to increase the sensitivity indices for important parameters and decrease those for the non-important parameters. Thus the evaluation becomes clearer.

Unlike the power transformations, the shifted-log transformation hardly changes the sensitivities in the very early phase, compared to the original data. This is due to the

fact that in this time phase the model results are generally very low and the shifted-log transformation has little effect on low data. The power transformations, however, bring these low values closer to 1 and thus reduce the total variance. This leads to higher sensitivity indices for those parameters that mainly influence the highest of these low values. A similar effect can be seen at the very end of the model time, when the calculated annual doses become lower again. So the application of a power transformation can produce clearer sensitivity statements especially if predominantly low values are calculated by the model.

Looking at the light-grey curves in Fig. 9.4 one realises that all three transformations considerably increase the sum of all S11 values. This means that the higher-order influences are reduced; the transformed model output is obviously less determined by parameter interactions than the original values.

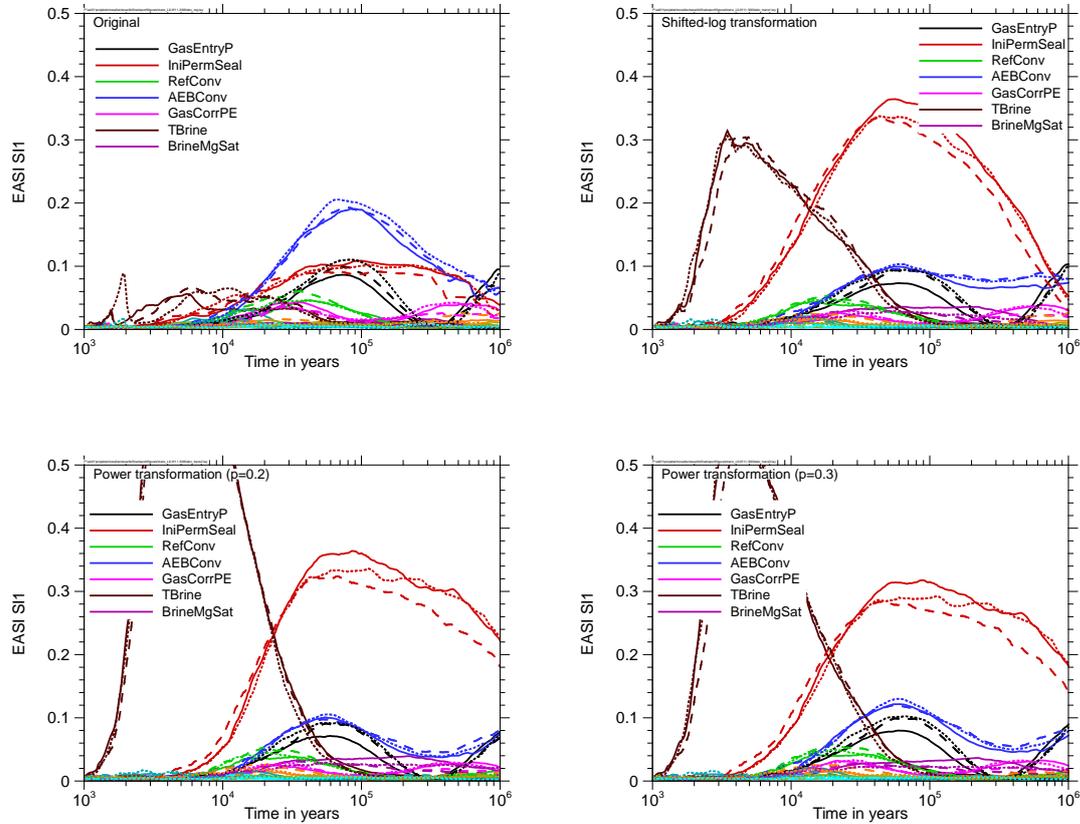
### 9.3.2 The LILW model

Fig. 9.5 shows the effects of the three transformations to the EASI evaluation of the LILW11 model. In view of the generally lower output of this model a threshold value of  $10^{-8}$  Sv/yr, that is one order of magnitude lower than for the SAM-GBS model, was used. The different line styles represent the results obtained with three different random samples of size 3000 each. The sample size was deliberately chosen low for this investigation in order to test the capability of transformations to enhance robustness.

Principally, the same effects as described for the SAM-GBS model can be observed. The transformations generally increase the calculated first-order effects of the most relevant parameters and slightly decrease those of the less relevant ones. Specifically, the effect on *TBrine* is eye-catching. Its S11 is considerably increased in the early and medium phase, and especially the power transformations lead to calculation of first-order indices up to 0.75 (not visible in the figure).

The most interesting effect, however, is that, compared to the evaluation with the original data, all three transformations lead to an exchange of the rankings of *IniPermSeal* and *AEBConv*, as it was the case in the investigations with different numbers of parameters. The S11 curves for the transformed LILW11 model output resemble more the curves for the original LILW6 model (see Fig. 6.10) than those for the original LILW11

model. Moreover, the parameter *BrineMgSat*, which is responsible for the difference, becomes at least visible in the transformed evaluation.



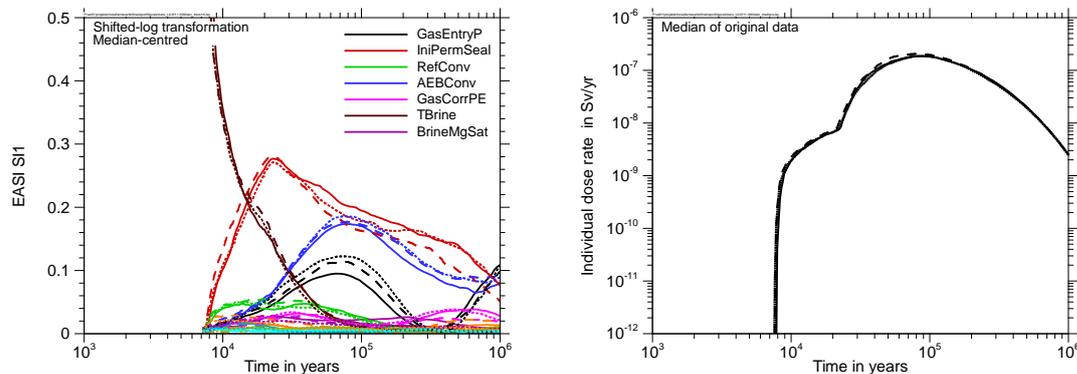
**Fig. 9.5** First-order sensitivity indices calculated using EASI with three different random samples (3000 runs each) for the LILW11 model from the original and transformed data ( $a = 10^{-8}$  Sv/yr); in view of clarity only 7 parameters are shown in the legend

The decrease of the sensitivity of *AEBConv* by the effect of transformation is plausible, in that it mainly influences the magnitude of the annual dose after seal failure, while *IniPermSeal* mainly acts on the time of failure. The highest model output values are likely to be caused, among other effects, by a high *AEBConv*. In the original-data analysis these few values have a high weight, which is considerably reduced by the transformation. So logically, the calculated sensitivity of *AEBConv* decreases under application of any of the transformations, while those of *IniPermSeal* and *BrineMgSat* increase, as they together are most responsible for the decision whether the seal has failed or not, so whether the model output generally tends to be higher or lower. So ob-

viously, transformation can help to detect such hidden sensitivities without the need to execute two sets of model runs with different numbers of varied parameters.

When comparing the results for the three different random samples in Fig. 9.5, one notes that these are not significantly closer to each other in the evaluations of transformed data than in the evaluation of the original data. Obviously, a transformation of the investigated type cannot be expected to significantly enhance the robustness of the evaluation in view of variations due to random influences of the sample.

In Fig. 9.6 the time-dependent results of an evaluation with a shifted-log transformation are presented, in which the median of the original data was used as the threshold value for each point in time. The time-development of the median is also shown.



**Fig. 9.6** First-order sensitivity indices calculated using EASI with three different random samples for the LILW11 model from the original and transformed data (left); a median-centred shifted-log transformation was used

One can see that only after 7500 years the median starts to achieve relevant values, which is due to the fact that before this time many calculation results are zero and the others very low. In the beginning, *TBrine* appears as the only parameter that has any influence at all and consequently its SI1 is very high. Beginning with about 10 000 years the sensitivity analysis results are qualitatively similar to those with a fixed threshold value of 10<sup>-8</sup> Sv/yr, but appear less pronounced. For most of the time the median is higher than the fixed value so that fewer values are considered “high” than in the previous investigation.

## 9.4 Summary

On a linear scale, the output of a repository performance assessment model is typically distinctly right-skewed and tailed and therefore not best appropriate for variance-based sensitivity analysis as it gives a strong overweight to few high values. A simple log-transformation will normally not solve this problem but turn it into its opposite.

Therefore, we defined two types of bijective transformations in order to mitigate the overweighing of high values without at the same time giving overweight to very low values: the shifted-log transformation and the power transformation, the latter with an exponent closer to 0 than to 1 – we tested exponents of 0.2 and 0.3.

The transformations make use of a threshold value that discriminates “high” from “low” values. This can be chosen oriented at physical, physiological or legal data or calculated automatically to fulfil some mathematical requirement. The latter, however, may lead to results that are hard to interpret.

Applied to the data to be analysed by variance-based sensitivity analysis, the transformations seem to lead to clearer and more unique results. The sum of all first-order indices increased in all investigated cases if any of the transformations was applied. This shows that the transformations reduced the influence of parameter interactions, making a first-order analysis more meaningful.

Generally, sensitivities that are relevant over the total parameter ranges seem to become more pronounced by the transformation while less relevant ones decrease even further. An analysis on non-transformed data, however, will probably better identify parameters that specifically influence the highest model results.

Output transformation can yield hints to hidden sensitivities as investigated in chapter 8.2. By comparing the results of the sensitivity analysis results for non-transformed and transformed data one can possibly achieve valuable insight to the model behaviour without the need to calculate several sets of model runs with different samples.



## 10 Handling of parameter dependencies

Complex models often depend on a large number of input parameters many of which are interconnected by physical or technical reasons or because his accords to the intentions of the investigator. Insofar as such couplings can be expressed by strict functional relations, this is a pure technical problem. The modeller can either select one of the coupled parameters for statistical variation and make sure by appropriate technical measures that the others are varied accordingly, or introduce a new parameter on which all the coupled ones depend; their values are calculated before each individual model run from the sample values for the new parameter. The new parameter is then used as the actual parameter for uncertainty and sensitivity analysis. The latter technique is applied in RepoTREND (/REI 16a/, /REI 17/) as it allows more direct assignment of the real uncertainties to the parameters under investigation and, moreover, has technical advantages (/REI 16b/).

Statistical couplings between two or more of the model input parameters can occur if there is some reason to assume that a statistical tendency of one parameter causes a specific tendency of another one. For example, a high porosity of a medium will normally induce a high permeability for fluids and gases, but this is not a strict coupling. Although correlation is only a special case of statistical dependency, observed or assumed statistical parameter couplings are often taken into account in probabilistic investigations by introducing more or less arbitrarily chosen, non-zero correlation coefficients between the concerned parameters. Many sampling techniques allow defining a correlation matrix before the sampling and then generate samples that approximately reflect the desired correlation coefficients.

While the described approach may be adequate for uncertainty analysis, statistical dependencies are a known problem in sensitivity analysis. The reason is that an identified sensitivity cannot clearly be assigned to one or the other parameter if both are correlated. Moreover, sensitivity analysis methods, for instance FAST/EFAST, do not allow correlated input. A procedure for variance-based sensitivity analysis with correlated input is described in /LIG 10/.

A statistical dependency between model parameters is normally due to some common influence that has not been modelled itself. Often this influence is not exactly known and is therefore simplified as a parameter correlation with some estimated value of the correlation coefficient.

The feeling that a pair of parameters should be correlated is often a hint at a lack of understanding of physical effects. In many cases, however, it is more or less impossible to resolve this by research, as an exact description would require a model on a much more detailed scale. In the mentioned example of a correlation between porosity and permeability the common influence originates from the microstructure of the porous medium and is hard to be analysed in detail.

For strongly correlated parameters any method of sensitivity analysis will calculate more or less the same sensitivities, so that it is hard to assess which parameter is actually responsible for the variation of model results. Therefore, one should avoid statistical dependencies where possible. This can be done by simply ignoring the dependency and using a sample with statistically independent parameter values. In this case one gets results that assign the sensitivities uniquely to the parameters, but are valid for a modelling situation that does not represent the actual physical conditions exactly as intended. One could perform two sensitivity analyses, one with a correlated and the other with a non-correlated sample, and compare the results.

A better, but methodically more challenging approach is to ascribe the coupled influence to two (or even more) statistically independent parameters, which together determine the correlated ones. In the case of porosity and permeability, one could introduce a simple model that links the permeability to the porosity, depending on some additional uncertain parameter, which itself is independent of the porosity. This can be a simple factor or a more complex relationship. In this way, parameter correlation can be avoided and more meaningful sensitivity analysis results are obtained, showing which influences to the results originate from the uncertainty of the porosity itself and which are due to the uncertainty of the link to the permeability.

## **11 Summary, recommendations and outlook**

In this chapter a summarising overview of the main project results and conclusions is given at first. Then, as a final outcome of the project, a recommendation is formulated how to proceed to execute a meaningful sensitivity analysis in a safety case. Finally, some open questions are identified and related research activities are mentioned.

### **11.1 Project summary and general conclusions**

#### **11.1.1 Background and project setup**

For assessing the safety of an existing or planned repository system, numerical models are needed. Such models, which can describe the total system as a whole or parts of it, are used to generate safety statements. As any model makes use of a number of uncertain input parameters every safety statement itself is subject to some uncertainty, which has to be quantified in order to allow proper assessment of the repository safety. This is done by uncertainty analysis.

Sensitivity analysis is a mathematical means for detecting, separating and quantifying the influences of the uncertainties of the individual parameters and parameter interactions to the overall uncertainty of the model output. Global sensitivity analysis means that all parameter uncertainties are taken into account together, each covering its total range. This is normally done as a probabilistic analysis by executing a large number of model runs with different combinations of parameter values. The goal of the project described here was to investigate different classical and newly-developed mathematical methods of global sensitivity analysis with typical repository models and to develop some recommendations for performing a sensible and meaningful global sensitivity analysis in the context of a repository safety case.

Sensitivity analysis is not an end in itself. A general finding of this project and previous investigations is that sensitivity analysis is a valuable tool for getting deeper insight in the behaviour of a simulation model and helps understand and communicate how the model works. This can enhance confidence in the model or even identify weaknesses and mistakes. A simple parameter ranking table however, derived from some more or less arbitrarily selected standard sensitivity analysis method, does not perform very well in these tasks.

Therefore, in the MOSEL project a variety of sensitivity analysis methods were intensely tested with three numerical models that describe realistic repository systems. These models simulate the integrated system including all components from the waste containers to the biosphere. In all cases the model output is the annual dose to an adult human individual. The models were either taken from former studies or specifically designed to represent typical properties of repository models. The distributions of the input parameters were selected according to expert judgement.

All three models produce distinctly right-skewed and tailed output distributions spanning several orders of magnitude, which is typical for integrated repository system models. One model – the LILW model – additionally exhibits a strong nonlinear, nearly non-continuous behaviour due to the sudden failure of a seal in the near field. This is a characteristic property of a model made for a repository in an existing salt mine like the Morsleben repository (/BEC 09c/). One model – the SAM-GBS model – produces a high proportion of zero output over the total time range, which is typical for repositories for heat-generating waste in rock salt, due to early closure of parts of the mine. These repository-specific properties can evoke unexpected sensitivity analysis effects that need to be investigated thoroughly.

All models were run very many times using different samples of different sizes up to about 32 000. Different sampling schemes were also applied.

A number of graphical and numerical methods for global sensitivity analysis were applied to the test systems. These methods can be grouped in three types: correlation-/regression-based methods, variance-based methods and non-parametric methods.

### **11.1.2 Graphical sensitivity analysis**

It was found that graphical methods of sensitivity analysis are generally very useful for gaining system understanding. Scatterplots of the model output versus the values of individual parameters contain much information, but are a bit unhandy and difficult to condense to a concise statement about the sensitivity. Good experiences were made with CSM plots, which allow, unlike the considered numerical methods, identification of parameter subranges with increased influence to the model output, provided that the used sample size is high enough to produce sufficiently smooth CSM curves. An improved variant of the CSM plot is the CUSUNORO plot, which basically provides the

same information about the system but is independent of an additive offset and generally better to present. For getting an impression of the time-development of the sensitivity, CSM or CUSUNORO plots for many points in time can be presented in one plot, colour-coding increasing time ranges. This kind of presentation was found to be very descriptive.

### **11.1.3 Correlation- and regression-based sensitivity analysis**

Correlation- and regression-based methods make use of an assumed linearity of the model and have been in wide use for many years. These methods exist in their direct form acting on the original model output and in a rank-based form, which performs a rank transformation on the data before executing the regression or correlation analysis. The rank transformation generally enhances the linearity of monotonic models and is therefore assumed to improve the performance of this type of methods. The investigations confirmed that the rank-based evaluation can considerably enhance the coefficient of model determination  $R^2$  and lead to clearer results, but we found that even with an  $R^2$  value of 0.2 or less the main sensitivities seem to be reliably identified by the direct version of the method. If applying a rank transformation, one has to keep in mind that this transformation is not bijective and destroys some information contained in the original data. If, for example, a parameter variation leads to increase of the highest model output values by a factor of 2 or even 10, a rank-based evaluation cannot detect these influences as long as the value ranking remains unchanged. While influences to the highest values are better detected by the direct evaluation, even with a low  $R^2$ , a rank-based evaluation is more adequate for analysing influences to low model output values. Applying both methods and comparing the results can be helpful for identifying model particularities.

Although correlation and regression are different mathematical concepts, the investigations did not show noteworthy differences between both kinds of evaluation. Any reason for applying one or the other cannot be derived from the project results.

### **11.1.4 Variance-based sensitivity analysis**

Variance-based sensitivity analysis is a concept developed for finding sensitivities in distinctly nonlinear and even non-monotonic models. Instead of trying to linearise the relationship, the contributions of the variances of the parameters to the variance of the

model output are calculated. While the first-order sensitivity indices quantify the contributions of the individual parameters alone, second- and higher-order indices characterise the influences of two- and multi-parameter interactions. In theory, the indices of all orders and for all parameters add up to 1. The sum of all indices describing first- or higher-order influences of a certain parameter is the total-order sensitivity index of this parameter.

Different methods are available for calculating the variance-based sensitivity indices. Five selected methods were applied to the test systems in a number of individual investigations: Sobol', EFAST, RBD, EASI and SDP. While all these methods are capable of calculating the first-order indices, only the Sobol' method calculates higher-order indices. Since, however, the results obtained with the available implementation of the Sobol' method were unsatisfying, it was excluded from further investigation. EFAST is at least capable of calculating total-order indices. As regards the first-order indices, the most convincing results were achieved with EASI. This method is mathematically smart and numerically effective, works with any sample and seems to generate robust results.

The variance-based methods seem to yield a message qualitatively similar to that of the regression-based SRC analysis; striking contradictions were not found. The variance-based analysis even provides less information as it cannot resolve the direction of influence. Variance-based sensitivity analysis, however, was found to be superior to regression-based analysis if there are significant non-monotonic influences to the model output. In such a case, SRC can pretend a low relevance, which the variance-based analysis does not.

### **11.1.5 The role of the sampling scheme**

The selection of the sampling scheme to be applied for drawing the sample for a probabilistic set of runs turned out to be essential for the stability and robustness of the results. As in uncertainty analysis unbiased statistical statements are desirable, random sampling is principally the best choice for this purpose. For sensitivity analysis, however, this is generally not the case, as random sampling tends to develop clusters and gaps, resulting in a rather inhomogeneous coverage of the parameter space. Stratified sampling schemes like the Latin Hypercube sampling (LHS) are designed to reach a better homogeneity. Quasirandom low-discrepancy sequences are deterministic se-

quences that, on the first sight, resemble random sequences but are specifically designed to cover an interval of values as homogeneously as possible at any sequence length. The quasirandom LpTau sampling scheme is based on this idea.

The project results did not provide evidence for a significant superiority of LHS to random sampling with regard to sensitivity analysis. The numbers of runs necessary for reaching convergent results were similar with both techniques. Clearly better performance, however, was found for the quasirandom LpTau sampling scheme. To reach comparable convergence, about four times as many runs were necessary with random sampling compared to LpTau sampling. Although the superiority of quasirandom sampling to random sampling is theoretically confined to low dimensionalities, it could be confirmed even for the LILW11 model with 11 parameters. As a rule of thumb, if the average dimension, which can be calculated as the sum of all total-order indices, does not exceed 3, one can expect advantages from quasi-random sampling.

#### **11.1.6 Dependence on the set of parameters**

The sensitivity analysis results for the LILW model were found to depend considerably on the selection of parameters that are taken into account. One parameter, the relative saturation of magnesium in the brine, was identified to be responsible for these differences. Although this parameter itself does not turn out to be of significant relevance, no matter if calculated with correlation- or variance-based methods, clearly different results for at least two of the other parameters were achieved depending on whether not its uncertainty was taken into account in the analysis. Although the origin of the differences can be well understood by analysing the functioning of the system, it may be surprising that the standard methods of sensitivity analysis do not reveal the obviously essential influence of one specific parameter. There seems to be some danger that relevant model properties are overlooked if one exclusively relies on such standard methods.

#### **11.1.7 Two-sample tests**

The idea of two-sample tests like the Smirnov test, which belong to the non-parametric methods of sensitivity analysis, is to analyse the sensitivity of a model to a parameter by splitting the total set of model output values in two subsets according to some criterion, transferring this split to the parameter sample and comparing the respective distri-

butions of the parameter in the two sub-samples. If they differ significantly, there must be some influence of the parameter to the assignment of the model output to one or the other subset.

If the separation is done using a rigid criterion like the 90<sup>th</sup> percentile of the model output distribution, as it is done in the standard version of the Smirnov sensitivity test, the results of such a test may be difficult to interpret. Then the test produces sensitivity statements similar to those derived from a conditional cobweb plot or a mean rank plot (see chapter 2.5.5). For a model with an inherent tendency to produce a two-split output distribution, however, this technique can yield valuable information about the sensitivities with regard to the model output falling in the lower or the upper part, provided that the split criterion is adapted accordingly. Due to the nearly sudden failure of the seal in the LILW system, the model has exactly this property, typically producing a bimodal output distribution. By this reason the model seems predestined to be investigated with a two-sample test. This was done using a dynamic threshold, which was determined manually for each point in time. In fact, the results of the analysis showed a considerable influence of the magnesium saturation, which had not been found with the standard methods.

### **11.1.8 Output transformation**

In order to better adapt the model output to the scale of evaluation and to avoid overvaluation of a few high values, one can apply an appropriate transformation to the model output before performing the sensitivity analysis. This can make sense specifically for variance-based sensitivity analysis as the total variance of a typical output distribution of a repository model, which is distinctly right-skewed and tailed on a linear scale, is often clearly dominated by very few individual runs. The well-proven rank-based methods of sensitivity analysis automatically apply a rank transformation, which linearises a monotonic relation, but is not bijective and therefore destroys some amount of information contained in the original model output. Sometimes it is recommended to analyse the output on a logarithmic scale, which means to apply a log transformation, but this can end up in a strong overvaluation of very low values and a complete failure with zero values.

An alternative is to use a monotonic, continuous and bijective transformation that maps “low” values to the interval between 0 and 1 and “high” values to moderately increasing

values above 1. For the distinction a threshold value must be defined, which can be oriented at physical or legal arguments or at formal properties of the distribution. Two transformations of this type were tested: the power transformation and the shifted-log transformation. It was found that a variance-based sensitivity analysis performed on the transformed data can produce results that allow a clearer assessment of the sensitivities and, in comparison with the direct evaluation, even reveal hidden sensitivities without the necessity to calculate a new set of runs.

## **11.2 Recommendation for proceeding in practical cases**

The goal of this chapter is to provide the practitioner with a guideline for performing a sensible sensitivity analysis of a repository model in the context of a safety case. This is, of course, not a universally valid code of practice but a recommendation on the basis of the outcome of the project, which can require revision after gaining further experiences, maybe with methods or tools that were not available for the project.

### **Step 1: Identification of uncertain parameters**

Before starting any kind of probabilistic analysis one has to identify the relevant uncertain parameters of the model and to properly quantify their uncertainties by means of appropriate probability density functions (pdf). This can be a demanding task, which has not been a goal of the project described here.

If statistical couplings between certain parameters are assumed, this often means that there is some kind of common influence from some parameter that itself is not directly taken into account in the model. The preferable way to handle this situation is to resolve the coupling by introducing a new uncertain parameter that is actually subject to the uncertainty and to calculate the model input values from this. It is even conceivable to introduce two or even more uncertain but statistically independent parameters that together determine the program input values. Using correlated parameters should remain an exception as it unavoidably leads to coupled sensitivities that cannot clearly be assigned to one or the other parameter. If necessary, one can consider executing the sensitivity analysis once with and once without taking account of the correlation.

For sensitivity analysis, one should generally include as few uncertainties as possible but as many as necessary. In case of doubt one should, as a start, rather include too

many than too few parameters and leave the final selection to step 2. For the following it is assumed that properly parametrised pdfs are available for all parameters.

Generally, one should be aware that sensitivity analysis is a stylised mathematical approach and avoid drawing overhasty conclusions from the results.

### **Step 2: Graphical screening by means of scatterplots**

Scatterplots are easily made and provide a good visual impression of the sensitivities. They can therefore be used for a first screening. A relatively small sample size of some 1000 is normally sufficient for this task, for which quasirandom sampling is recommended, as it is likely to lead to better robustness of the results than random sampling. Scatterplots should be made for each parameter and a number of points in time, covering the total time interval of interest, as well as for the run maxima. Although a big number of scatterplots can become necessary following this proposal, their creation and visual assessment is not too time-consuming if supported by an appropriate computational tool.

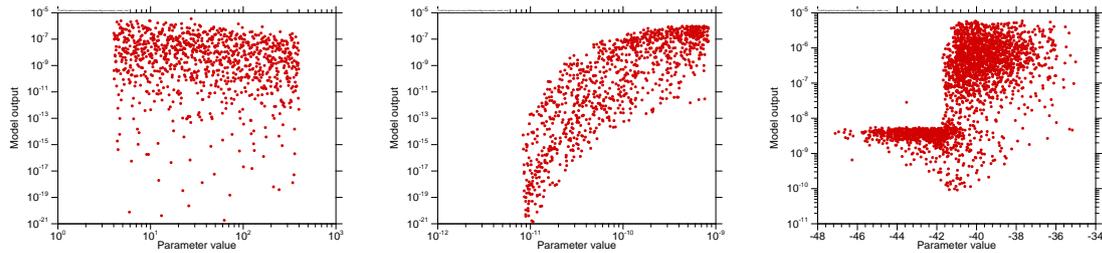
One has to keep in mind that higher-order sensitivities due to parameter couplings might not become visible in one-parameter scatterplots.

Based on the results of the screening, a final selection of parameters to be taken into account in the sensitivity analysis has to be made by excluding those with little visual effects to the output. The number of parameters should not exceed 10, if possible.

### **Step 3: Assessing the kind of influence**

One can use the scatterplots to roughly assess the kind of influence of the individual parameters to the model output. This can be done using three categories, see Fig. 11.1 for examples:

- a) basically linear (on a linear or logarithmic scale),
- b) clearly nonlinear but monotonic,
- c) extremely nonlinear, non-monotonic, split or otherwise conspicuous.



**Fig. 11.1** Example scatterplots of types a (left), b (middle) and c (right)

The scatterplots for all relevant parameters should be assessed in this way for a number of points in time and for the run maxima. By this assessment one gets a rough impression of the complexity of the model. The recommended further procedure depends on how complex the system seems to be.

#### **Step 4: Simple sensitivity analysis**

If the scatterplots are mainly of type a, maybe a few ones of type b, then a simple sensitivity analysis is probably sufficient. If a more detailed sensitivity analysis is required anyway, one can skip this step, as it is covered by the following one.

One should decide whether there are some points in time of specific interest or an analysis of the run maxima is sufficient.

A simple sensitivity analysis can be made on the basis of the existing probabilistic results that were produced for the scatterplots. If, however, there are a significant number of parameters visually assessed as irrelevant, more robust results can be expected from a new set of runs including only the final selection of parameters. A sample size of some or a few thousand will normally suffice. Again, quasirandom sampling is recommended. Random sampling can also be used, but might require a higher sample size to reach stable sensitivity analysis results.

We recommend presenting a graphical analysis, for example by means of CSM or CUSUNORO plots. CUSUNORO is mathematically smarter, but CSM might be easier to communicate.

In case the output distribution spans many orders of magnitude as in the examples of Fig. 11.1, one might consider applying a shifted-log transformation. Practically this

means that the evaluation is made on the logarithmic scale but without overvaluing meaningless low values.

We propose to calculate three sensitivity measures: the standardised regression coefficients (SRC), the standardised rank regression coefficients (SRRC) and the first-order sensitivity indices (SI1). For the latter, the EASI method seems best appropriate. For SRC and SRRC the coefficient of linear model determination,  $R^2$  should be checked. While a high value (near 1) indicates a good performance of the method, even a value of 0.2 or less does not mean that the method fails. The  $R^2$  value for SRRC should not be lower than that for SRC. Moreover, if the sum of all calculated SI1 is close to 1, the system is well-determined by first-order effects. A low value, however, is an indication for considerable influences of parameter interactions, which cannot be resolved by the methods applied so far.

If all three measures agree about the parameter ranking, the results seem plausible, explain the model sensitivities satisfyingly and no parameters of type c were identified by the scatterplots, the sensitivity analysis can be stopped at this point. Otherwise, one should proceed to step 5.

### **Step 5: Detailed sensitivity analysis**

This step should be performed if there seem to be major nonlinear influences, visible at scatterplots of type b or even c, if step 4 did not produce fully understandable and explicative results or if by any reason deeper system understanding is expected to be achieved from the sensitivity analysis.

If the model is time-dependent, the detailed sensitivity analysis should also be performed as a time-dependent analysis. Therefore, the model output should be available as a time series. The analysis should be based on a higher number of runs, but as a start, the existing results can be used. If quasirandom or random sampling was applied, the sample can be extended later.

We recommend producing time-dependent CSM or CUSUNORO plots as they are a good means for getting a visual impression of the time-development of the sensitivities. These plots also allow detecting non-monotonic influences. Moreover they can reveal parts of the parameter intervals with a specifically high or low sensitivity.

For a time-dependent model, time curves of the sensitivity measures are helpful for understanding the system development and should therefore always be made in a detailed analysis. A comparison of SRC, SRRC and SI1 time curves can reveal interesting details about the system. Qualitative differences between SRC and SI1 can indicate non-monotonic influences. Qualitative differences between SRC and SRRC can be a hint at different influences to low and high values. Generally, influences to low values are better detected by SRRC, influences to high values by SRC. Applying a shifted-log or power transformation can also be helpful for analysing such different influences, but it should not be done without cause.

In order to assess the stability of the achieved results one can calculate the same curves from different sets of model runs. One possibility is to start with a relatively small sample size and increase it successively by factors of two. Comparison of the curves shows how many runs are needed to reach satisfyingly stable results.

Generally, it should be avoided to simply present curve plots without interpretation as this might be more confusing than helpful. Only such results should be presented that have been well understood and from which plausible messages about the system and its time evolution can be derived.

If the detailed sensitivity analysis yields unsatisfying or suspicious results or leaves any other open questions, one can proceed to step 6.

### **Step 6: Expert sensitivity analysis**

This step requires deeper expert knowledge and should only be performed by investigators who actually have such knowledge. General instructions cannot be given as the procedure depends on the specific problems to be addressed. Some ideas, which were triggered by the findings of this project, are given in the following.

A low sum of first-order indices is a hint at considerable parameter interactions. Performing a shifted-log or power transformation can reduce the effects of such interactions to the model results and with it increase the significance of first-order analysis. A detailed analysis of interactions, however, requires a systematic higher-order sensitivity analysis. This requires a well-working tool, which was not available for this project. Total-order analysis can also yield some information, but is not always easy to interpret.

Total-order indices can be calculated using EFAST, but this requires a specific set of runs and does not seem to perform too well with highly nonlinear systems.

Suspicious parameters that are expected to have some influence but do not show up in the standard sensitivity measures can be investigated by performing two sets of calculations, with and without variation of the parameter in question. It is possible that the results essentially differ.

An appropriate transformation of the model output can lead to clearer results especially of variance-based sensitivity analysis. We propose to use a shifted-log or power transformation with a problem-adequate threshold value. It was found that application of such a transformation and comparison of the results with those for the non-transformed output can even identify influential parameters with low sensitivity measures, without the need of an extra set of calculations.

A two-split, bimodal output distribution predestines a model to be analysed with a two-sample test. We applied the Smirnov test with a dynamic, time-dependent threshold. Such an analysis identifies the most important influences to the decision whether the model output value appears in the upper or the lower part of the distribution and can therefore provide valuable information about the system functioning.

### **11.3 Related work and outlook**

The work in the project MOSEL was accompanied and supported by other research activities. The project NUMSA (/KUH 15/), executed by TU Clausthal, was carried out in close cooperation with MOSEL and made, in parts, use of the same models, but was focused on numerical effectivity as well as at methods and aspects that were not in detail considered in MOSEL. A specific, generic model for a repository for high-level waste in rock salt was defined that produces a considerable proportion of zero output, in order to investigate how such situations can be handled in sensitivity analysis.

Metamodeling is a mathematical technique to simplify a complex numerical model by simulating its behaviour as close as possible by a multi-parameter function. For this purpose, the original model is run a number of times with different parameter combinations and the metamodel function is calibrated using the results, which can be done by interpolation or approximation. In view of sensitivity analysis, the question arises

whether one can save model runs and computing time by replacing the original model by an appropriately calibrated metamodel and then analysing the uncertainties of the metamodel. One approach that makes use of this concept is the SDP method investigated in chapter 6.5, but it was found that, if applied to repository models, this method is not at all numerically efficient and needs a lot of computing time. Therefore, a specific investigation was undertaken in cooperation with the Institute of Scientific Computing of the TU Braunschweig (/RAN 15/). Several metamodels for the LILW6 model based on interpolation or approximation were developed, calibrated with a set of original model results and tested with a larger set. Then the sensitivity indices were calculated from the metamodels and compared with those obtained in the MOSEL project. As a result, it was found that a metamodel approach can reduce the computational effort for sensitivity analysis, but the applied metamodeling techniques do not seem to be best appropriate for this purpose. Generally, the metamodeling approaches seem promising, but need further improvement in view of sensitivity analysis.

To make the investigated methods and techniques of sensitivity analysis available for application in repository safety assessment with the program package RepoTREND, an appropriate computational tool was needed. As a basis for such a tool, the software program SimLab seemed appropriate, which was developed by the Joint Research Centre (JRC) in Ispra/Italy. SimLab 2, however, is technically obsolete and cannot be integrated in other programs. The newer version SimLab 3, which was designed as a program library, has been withdrawn by the developers due to some errors and shortcomings. Therefore, a completely new development was initiated in cooperation between GRS and JRC. The new concept is based on scripts written in the statistical programming language R, as this powerful open-source tool is popular among mathematicians, verified R-scripts are available for all standard methods of sensitivity analysis and it is relatively easy to write own scripts or adapt existing ones to specific needs. The new software package SimLab 4 (/JRC 16/), which was developed by JRC with support from GRS, can be used as a stand-alone tool or as a program library for use with own developments. It is written in C++ but provides an interface to R. This concept makes it very flexible for adaptation to individual needs or integration of new methods.

On the basis of SimLab 4, a new framework program for preparation, execution and evaluation of statistical investigations with the RepoTREND program package, called RepSTAR (/BEC 16/), was developed in cooperation with the project ADEMOS. This

framework consists of routines for sampling and data support of statistical runs as well as the interactive tool RepoSUN for executing probabilistic uncertainty and sensitivity analysis of the results.

A well working method for higher-order variance-based sensitivity analysis was not available for the investigations of the MOSEL project, and total-order indices could only be calculated with the EFAST method, which needs a specific non-extendible sampling and does not seem to perform too well on highly nonlinear models. Some orienting experiments, however, were undertaken with the SobolGSA software of the Imperial College London (/ZUN 13/, /KUC 13/). This program makes use of a metamodeling approach in combination with a modified Sobol' analysis. It works with any samples and is able to calculate first-, second- and total-order sensitivity indices. As the results look promising, an integration of this concept in the RepoSUN tool seems sensible.

In MOSEL, only a limited selection of sensitivity analysis methods have been tested. It is possible that other approaches like non-parametric regression or density-based sensitivity analysis provide additional insight to the behaviour of complex systems or reveal sensitivities that are not recognised by the investigated methods. Therefore, it seems advisable to permanently follow the development in the field of sensitivity analysis, to test new methods and to update the proceeding recommendation if necessary.

### **Acknowledgement**

The authors wish to thank Klaus-Jürgen Röhlig, Elmar Plischke and Sebastian Kuhlmann from the Technical University of Clausthal for their permanent interest, various fruitful discussions and valuable input.

Financial support was granted by the German Federal Ministry for Economic Affairs and Energy (BMWi) under sign 02E10941.

## List of References

- /BEC 03/ Becker, D.-A.; Buhmann, D.; Storck, R.; Alonso, J.; Cormenzana, J.-L.; Hugl, M.; van Gemert, F.; O’Sullivan, P.; Laciok, A.; Marivoet, J.; Sillen, X.; Nordman, H.; Vieno, T.; Niemeyer, M.: Testing of Safety and Performance Indicators (SPIN), EUR 19965 EN, European Commission, Brussels 2003.
- /BEC 09a/ Becker, D.-A. (Editor); Cormenzana, J. L.; Delos, A.; Duro, L.; Grupa, J.; Hart, J.; Landa, J.; Marivoet, J.; Orzechowski, J.; Schröder, T.J.; Vokal, A.; Weber, J.; Weetjens, E.; Wolf, J.: Safety Indicators and Performance Indicators PAMINA Deliverable (D-N°:3.4.2), European Commission, 2009.
- /BEC 09b/ Becker, D.-A.; Spießl, S.; Röhlig, K.-J. ; Plischke, E.; Bolado-Lavin, R.; Badea (JRC), A.; Cormenzana, J. L.; Schröder, T. J.; Hart, J.; Avila, R.; Ekström, P. A.; Broed, R.: Evaluation of Approaches to Sensitivity Analysis. PAMINA Deliverable (D-N°:2.1.D.1), European Commission 2009.
- /BEC 09c/ Becker, D.-A.; Buhmann, D.; Mönig, J.; Noseck, U.; Rübél, A.; Spießl, S.: Endlager Morsleben – Sicherheitsanalyse für das verfüllte und verschlossene Endlager mit dem Programmpaket EMOS. Planfeststellungsverfahren zur Stilllegung des Endlagers für Radioaktive Abfälle Morsleben, Verfahrensunterlage P 278. Bundesamt für Strahlenschutz 2009 (in German).  
[http://www.endlager-morsleben.de/SharedDocs/Downloads/Morsleben/DE/planfeststellungsverfahren/lfdnr106-278-00-v01-p278.pdf?\\_\\_blob=publicationFile&v=1](http://www.endlager-morsleben.de/SharedDocs/Downloads/Morsleben/DE/planfeststellungsverfahren/lfdnr106-278-00-v01-p278.pdf?__blob=publicationFile&v=1).
- /BEC 12/ Becker, D.-A.; Spiessl, S.: Particularities in Sensitivity Analysis of Numerical Models for Long-Term Safety of Final Repositories. Proceedings of PSAM11/ESREL2012, IAPSAM&ESRA, June 2012.
- /BEC 14/ Becker, D.-A.: Improvement of the Reliability and Robustness of Variance-Based Sensitivity Analysis of Final Repository Models by Application of Output Transformation. Proceedings of PSAM12, IAPSAM, 2014.
- /BEC 16/ Becker, D.-A.: RepoSTAR – Ein Codepaket zur Steuerung und Auswertung statistischer Rechenläufe mit dem Programmpaket RepoTREND. GRS-411, Braunschweig 2016 (in German).

- /BMU 10/ Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit: Safety Requirements Governing the Final Disposal of Heat-Generating Radioactive Waste. 30<sup>th</sup> September 2010.
- /BOL 09/ Bolado-Lavin, R.; Castaings, W.; Tarantola, S.: Contribution to the sample mean plot for graphical and numerical sensitivity analysis. RESS 94 (2009) 1041–1049.
- /BUH 91/ Buhmann, D.; Nies, A.; Storck, R.: Analyse der Langzeitsicherheit von Endlagerkonzepten für wärmeerzeugende radioaktive Abfälle. GSF-Bericht 27/91, Braunschweig 1991 (in German).
- /BUH 04/ Buhmann, D.; Becker, D.-A.; Keesmann, S.; Kröhn, K.-P.; Rübél, A.: Aktualisierung sicherheitsanalytischer Rechenprogramme für Teilsysteme eines Endlagers. GRS-200, Braunschweig 2004 (in German).
- /COR 12/ Cormenzana, J. L.: Graphic representation alternative to conditional cobweb plots. PSAM11/ESREL2012, IAPSAM&ESRA, June 2012.
- /CUK 78/ Cukier, R. I.; Levine, H. B.; Shuler, K.E.: Nonlinear sensitivity analysis of multiparameter model systems, J. Comput. Phys.26 (1) (1978) 1–42.
- /DUT 09/ Dutang, C. Savicky, P. and Wuertz, D. 2009. Introduction to random number generation - quick introduction of randtoolbox - full presentation of the RNG tools.  
[https://r-forge.r-project.org/scm/viewvc.php/\\*checkout\\*/pkg/randtoolbox/inst/doc/randtoolbox.pdf?revision=4622&root=rmetrics&pathrev=4622](https://r-forge.r-project.org/scm/viewvc.php/*checkout*/pkg/randtoolbox/inst/doc/randtoolbox.pdf?revision=4622&root=rmetrics&pathrev=4622).
- /GOE 09/ Göncü, A. Monte Carlo and quasi-Monte Carlo methods in financial derivative pricing. PhD, Florida State University, 2009.
- /HEL 98/ Helton, J. C.: Uncertainty and sensitivity analysis in performance assessment for the Waste Isolation Pilot Plant. Computer Physics Communications 117 (1999) 156-180.

- /HEL 06/ Helton, J. C.; Johnson, J. D.; Sallaberry, C. J.; Storlie, C. B.: Survey of sampling-based methods for uncertainty and sensitivity analysis. RESS 91 (2006), 1175–1209.
- /HEL 09/ Helton, J. C.; Sallaberry, C. J.: Computational implementation of sampling-based approaches to the calculation of expected dose in performance assessments for the proposed high-level radioactive waste repository at Yucca Mountain, Nevada. RESS 94 (2009), 699–721.
- /JOB 07/ Jobmann, M.; Amelung, P.; Billaux, D.; Polster, M.; Schmidt, H.; Uhlig, L.: Untersuchungen zur sicherheitstechnischen Auslegung eines generischen Endlagers im Tonstein in Deutschland. GENESIS Abschlussbericht, Anlagenband: Geologie der Referenzregionen im Tonstein, DBE Technology, 2007.
- /JRC 16/ Joint Research Centre: SimLab4 Manual. JRC, Ispra 2016.
- /KUC 13/ Kucherenko, S.: SobolHDMR: a general-purpose modeling software. Methods Mol Biol. (2013) 1073: 191-224.
- /KUH 15/ Kuhlmann, S.; Plischke, E.; Röhlig, K-J.: Optimierung der numerischen Effizienz von Verfahren zur Sensitivitätsanalyse im Hinblick auf Modelle zur Analyse der Langzeitsicherheit von Endlagern. Abschlussbericht. TU Clausthal, September 2015.
- /LIG 10/ Li, G.; Rabitz, H.; Yelvington, P. E.; Oluwole, O. O.; Bacon, F.; and Kolb, C. E.: Global Sensitivity Analysis for Systems with Independent and/or Correlated Inputs. J. Phys. Chem. A 2010, 114, 6022–6032.
- /LIJ 00/ Li, J. X.; Winker, P.: Time series simulation with quasi Monte Carlo methods, 2000.  
<http://fmwww.bc.edu/cef00/papers/paper151.pdf>.
- /NIE 92/ Niederreiter, H.: Random number generation and quasi-Monte Carlo methods. Society for Industry and Applied Mathematics (SIAM) Press, Regional Conference Series in Applied Math., Philadelphia, PA, 1992.

- /OEK 02/ Ökten, G.: Random sampling from low-discrepancy sequences: Applications to option pricing. *Mathematical and Computer Modelling*, 35 (2002), 1221-1234.
- /PLI 10/ Plischke, E.: An effective algorithm for computing global sensitivity indices (EASI). *RESS 95* (2010) 354–360.
- /PLI 12/ Plischke, E.: An adaptive correlation ratio method using the cumulative sum of the reordered output. *RESS 107* (2012), 149–156.
- /RAN 15/ Rang, J.; Matthies, H. G.; Spiessl, S. M., Becker, D.-A.: Metamodels for a final repository model. Final report. Institute of Scientific Computing, T Braunschweig, October 2015.
- /RAT 07/ Ratto, M.; Pagano, A.; Young, P.: State Dependent Parameter Metamodeling and Sensitivity Analysis. *Computer Physics Communications*, 177, 863–876.
- /REI 16a/ Reiche, T.: RepoTREND – Das Programmpaket zur integrierten Langzeitsicherheitsanalyse von Endlagersystemen. GRS-413, Braunschweig 2016 (in German).
- /REI 16b/ Reiche, T.; Becker, D.-A.: RepoSTAR – New Framework for Statistic Runs for Uncertainty and Sensitivity Analysis of a Radioactive Waste Repository Model. Presentation at SAMO 2016, La Réunion, December 2016.  
<https://samo2016.sciencesconf.org/118783>.
- /REI 17/ Reiche, T.: RepoTREND – A Program Package for Safety Analysis of Radioactive Waste Repository Systems. American Nuclear Society, IHLRWM 2017, Charlotte, NC, April 2017.
- /RUE 07/ Rübél, A.; Becker, D.-A.; Fein, E.: Radionuclide transport modelling to assess the safety of repositories in clays. GRS-228, Braunschweig 2007.
- /SAL 97/ Saltelli A.; Tarantola S.; Chan K.: Sensitivity analysis of model output: an improvement of the FAST method, EUR Report 17338 EN, 1997.

- /SAL 98/ Saltelli, A.; Bolado, R.: An alternative way to compute Fourier amplitude sensitivity test (FAST). *Computational Statistics & Data Analysis* 26 (1998) 445-460.
- /SAL 02/ Saltelli, A.: Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications* 145 (2002) 280–297.
- /SAL 08/ Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S.: *Global Sensitivity Analysis: The Primer*. Wiley and Sons, 2008.
- /SIN 93/ Sinclair J.: Response to the PSACOIN Level S exercise. PSACOIN Level S intercomparison. Organisation for Economic Cooperation and Development, Nuclear Energy Agency, 1993.
- /SOB 93/ Sobol', I.M.: Sensitivity estimates for nonlinear mathematical models. *MMCE*, 1 (4) (1993) 407–414.
- /SOB 01/ Sobol', I.M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation* 55 (2001) 271–280.
- /SOB 16/ Sobol', I. M.: Average Dimension. Presentation at SAMO 2016, La Réunion, December 2016.  
[https://samo2016.sciencesconf.org/data/program/KS\\_Sobol.pdf](https://samo2016.sciencesconf.org/data/program/KS_Sobol.pdf).
- /SPI 12a/ Spiessl, S. M.; Becker, D.-A.: A Metamodel Approach to Sensitivity Analysis of a PA Model for a Generic HLW Repository in Clay. Proceedings of PSAM11/ESREL2012, IAPSAM&ESRA, June 2012.
- /SPI 12b/ Spiessl, S.M., Becker, D.-A.: Performance of Six Sampling Methods for Global Sensitivity Analysis on the Basis of a Performance Assessment Model for a Generic High Level Radioactive Waste Repository in Clay. Poster, Uncertainty in Computer Models (UCM) 2012 conference, Sheffield, July 2-4 2012,  
<http://www.mucom.ac.uk/UCM2012/Forms/Downloads/Posters/Spiessl.pdf>.

- /SPI 12c/ Spiessl, S. M.; Becker, D.-A.; Rübél, A.: EFAST Analysis Applied to a PA Model for a Generic HLW Repository in Clay. RESS 107 (2012), 190 - 204.
- /SPI 14/ Spiessl, S. M.; Becker, D.-A.: Investigation of Different Sampling and Sensitivity Analysis Methods Applied to a Complex Model for a Final Repository for Radioactive Waste. Proceedings of PSAM12, IAPSAM 2014.
- /STO 96/ Storck, R.; Buhmann, D.; Hirsekorn, R.-P.; Kühle, T.; Lührmann, L.: Das Programmpaket EMOS zur Analyse der Langzeitsicherheit eines Endlagers für radioaktive Abfälle, Version 5. GRS-122, Braunschweig 1996 (in German).
- /TAR 06/ Tarantola, S.; Gatelli, D.; Mara, T.: Random balance designs for the estimation of first order global sensitivity indices. RESS 91 (2006), 717 – 727.
- /ZUN 13/ Zuniga, M.; Kucherenko, S.; Shah, N.: Metamodelling with independent and dependent inputs. Computer Physics Communications, 184, 6 (2013) 1570-1580.

## List of Figures

Fig. 2.1	The principle of Latin Hypercube Sampling (taken from Wikipedia) .....	13
Fig. 2.2	Illustration of the basic idea of the Sobol' LpTau sequences in [0,1] .....	14
Fig. 2.3	Possible statistical dependencies of two parameters with their correlation coefficients. For the case presented in the middle, the correlation coefficient is undefined, since there is no variation in $Y$ (from Wikipedia: <a href="https://en.wikipedia.org/wiki/Correlation_and_dependence">https://en.wikipedia.org/wiki/Correlation_and_dependence</a> ).....	16
Fig. 2.4	Comparison of sampling methods by means of three diagrams: $p_1$ vs. $p_2$ (left), $p_1$ vs. number of simulation (middle), $p_1$ frequency histogram (right).....	18
Fig. 2.5	Graphical uncertainty analysis: plot of all runs (left) and CDF/CCDF (right) .....	20
Fig. 2.6	Graphical uncertainty analysis: histogram (left) and colour-coded maximum scatterplot (right) .....	21
Fig. 3.1	Cross-section of the model area in Northern Germany .....	38
Fig. 3.2	Schematic view of the components of the LILW system.....	39
Fig. 4.1	Scatterplots for the clay system, 106 years.....	46
Fig. 4.2	Scatterplots for the clay system, 108 years.....	46
Fig. 4.3	Time-development of the scatterplot for the parameter IniPermSeal (LILW6 model) .....	47
Fig. 4.4	Time-development of the scatterplot for the parameter AEBConv (LILW6 model) .....	48
Fig. 4.5	Time-development of the scatterplot for the parameter GasEntryP (LILW6 model) .....	48
Fig. 4.6	Time-development of the scatterplot for the parameter TBrine (LILW6 model) .....	49
Fig. 4.7	CSM plots of the clay system.....	51
Fig. 4.8	CSM plots for the LILW6 system.....	52
Fig. 4.9	Time-development of CSM curves for the LILW11 system.....	53
Fig. 4.10	Time-development of CUSUNORO curves for the LILW11 system.....	55

Fig. 4.11	Time-development of CSV curves for the LILW11 system .....	56
Fig. 5.1	Regression- and correlation-based sensitivity analysis with direct (left) and rank-transformed (right) data for the clay model.....	59
Fig. 5.2	Regression- and correlation-based sensitivity analysis with direct (left) and rank-transformed (right) data for the LILW6 model.....	60
Fig. 5.3	Scatterplots of the clay model, presenting parameters that influence the highest (left) and the lowest (right) values .....	63
Fig. 6.1	Sobol' indices of first (left) and total order (right) for the clay model .....	67
Fig. 6.2	Sobol' indices of first (left) and total order (right) for the LILW6 model .....	68
Fig. 6.3	Sobol' indices of first (left) and total order (right) for the LILW11 model (only the 6 most relevant parameters are presented) .....	68
Fig. 6.4	EFAST indices of first (left) and total order (right) for the clay model.....	69
Fig. 6.5	EFAST indices of first (left) and total order (right) for the LILW6 model.....	70
Fig. 6.6	EFAST indices of first (left) and total order (right) for the LILW6 model, calculated with two different random seeds .....	71
Fig. 6.7	RBD first-order indices for the clay model .....	72
Fig. 6.8	RBD first-order indices for the LILW6 model, compared for different sample sizes (left) and different random seeds (right).....	73
Fig. 6.9	EASI first-order indices for the clay model .....	74
Fig. 6.10	EASI results for the LILW6 model, compared for different sample sizes (left) and with the Sobol' calculation (right, the Sobol' sample was also used for EASI).....	74
Fig. 6.11	SDP results for the clay model on the basis of random samples, compared for different sample sizes (left) and with EASI (right) .....	75
Fig. 6.12	SDP results for the LILW6 model on the basis of random samples, compared for different sample sizes (left) and with EASI (right) .....	76
Fig. 6.13	Comparison of correlation-based (3000 runs, left) and variance-based (4096 runs, right) sensitivity analysis for the LILW6 system .....	76
Fig. 6.14	CUSUNORO plots of the LILW6 model for the time period between 300 000 and 600 000 years for the parameters AEBCConv (left) and GasEntryP (right) .....	77

Fig. 7.1	CSM curves for the clay system at 1.0·10 <sup>7</sup> years, calculated on the basis of random samples (left) and Latin Hypercube samples (right) of different sizes .....	82
Fig. 7.2	Time-dependent SRC evaluation for the clay system on the basis of random samples (left) and Latin Hypercube samples (right) of different sizes .....	82
Fig. 7.3	CSM plots for the LILW11 system, based on random (left) and LpTau (right) samples of sizes 2048 and 4096 .....	83
Fig. 7.4	Time-dependent EASI evaluation for the LILW6 system on the basis of random samples (left) and LpTau samples (right) of different sizes .....	84
Fig. 8.1	Time-dependent EASI evaluation for the LILW model with 6, 7, 11 and 20 parameters; for LILW20 the sum of the SI1 values of the parameters 12 – 20 is shown as a black dotted line.....	86
Fig. 8.2	Total variance and contributions of three parameters during the early time phase; LILW6 vs. LILW7 (left) and LILW11 vs. LILW20 (right) .....	87
Fig. 8.3	Comparison of sensitivity plots for the LILW6 and the LILW7 model: CSM (top left), EASI (top right), SRC (bottom left), SRRRC (bottom right) .....	89
Fig. 8.4	Comparison of sensitivity plots for the original and the modified LILW6 model with those of the LILW7 model: CSM (left) and EASI (right) .....	90
Fig. 8.5	EFAST total-order analysis of the LILW6 (left) and the LILW7 (right) model; for better visibility the curve for BrineMgSat is marked by circles .....	91
Fig. 8.6	Histograms for the LILW20 model for two points in time; calculation cases with (presumably) intact seal are presented in green, those with failed seal in red .....	92
Fig. 8.7	Time-dependent dynamic Smirnov sensitivity evaluation of the LILW20 model with: threshold value and numbers of runs (left) and results of the Smirnov test (right, only the 7 most important parameters are presented) .....	93
Fig. 9.1	Contributions to the total variance of the results of the SAM-GBS system .....	96
Fig. 9.2	Shifted-log transformation and power transformation with two different exponents, visualised with a logarithmic (left) and a linear y-axis (right).....	98

Fig. 9.3	Histograms for frequency and contribution to the total variance of 3000 runs of the SAM-GBS model (run maxima); calculated from the original data and after application of three transformations (1: shifted-log, 2: power $p=0.2$ , 3: power $p=0.3$ ) with $\alpha = 10^{-7}$ Sv/yr .....	100
Fig. 9.4	First-order sensitivity indices calculated using EASI for the SAM-GBS model from the original and transformed data ( $\alpha = 10^{-7}$ Sv/yr); only the 6 most relevant of 31 parameters and the sum of all SI1 are presented .....	101
Fig. 9.5	First-order sensitivity indices calculated using EASI with three different random samples (3000 runs each) for the LILW11 model from the original and transformed data ( $\alpha = 10^{-8}$ Sv/yr); in view of clarity only 7 parameters are shown in the legend .....	103
Fig. 9.6	First-order sensitivity indices calculated using EASI with three different random samples for the LILW11 model from the original and transformed data (left); a median-centred shifted-log transformation was used .....	104
Fig. 11.1	Example scatterplots of types a (left), b (middle) and c (right).....	117

## List of Tables

Tab. 2.1	Naming of correlation- and regression-based methods of sensitivity analysis .....	30
Tab. 3.1	Parameters of the clay system model .....	38
Tab. 3.2	Parameters of the LILW system model. The colours indicate the subsets of parameters that were used for the different investigations (LILW6 = red, LILW7 = red + green, LILW11 = red + green + blue, LILW20 = all) .....	40
Tab. 3.3	Parameters of the SAM-GBS system model .....	42
Tab. 6.1	Investigations performed with variance-based methods.....	66
Tab. A.1	Geometry data for the clay model .....	135
Tab. A.2	BSK-3 container data.....	135
Tab. A.3	Radionuclide inventory.....	135
Tab. A.4	Mobilisation rates for different fuel compartments.....	137
Tab. A.5	Relative inventory in the different fuel compartments in per cent .....	137
Tab. A.6	Solubility limits in [ mol·l <sup>-1</sup> ] (“high” denotes no solubility limit) .....	137
Tab. A.7	Distribution coefficients for the Lower Cretaceous Clays (all elements not included here are assumed to have K <sub>d</sub> = 0).....	139
Tab. A.8	Dose conversion factors (DCF) in Sv·y <sup>-1</sup> / Bq·m <sup>-3</sup> .....	139
Tab. A.9	Near field segment data.....	140
Tab. A.10	Global near field data.....	142
Tab. A.11	Waste package data .....	143
Tab. A.12	Radionuclide inventories.....	143
Tab. A.13	Far field data.....	144
Tab. A.14	Sorption distribution coefficients (K <sub>d</sub> ) in m <sup>3</sup> /kg .....	144



## A Data of the test systems

In the following tables the basic data of the two main test model systems are given. As no new model calculations were performed for the SAM-GBS system, the data of this system are not given here, see /BUH 91/.

### A.1 Repository for Spent Fuel and HLW in clay

For the details of this system see /RUE 07/.

**Tab. A.1** Geometry data for the clay model

Parameter		Value
Borehole length	[ m ]	50
Numbers of containers per borehole	[ - ]	5
Area of the hexagonal reference section	[ m <sup>2</sup> ]	1 913.05
Width of the CLAYPOS model	[ m ]	43.74
Transport length in the bentonite	[ m ]	10
Transport length in the Apt	[ m ]	80
Transport length in the Alb	[ m ]	250

**Tab. A.2** BSK-3 container data

Parameter		Value
Length	[ m ]	4.90
Diameter	[ m ]	0.43
Wall thickness	[ m ]	0.05
Void volume	[ m <sup>3</sup> ]	0.30
Mass	[ kg ]	1 660
Inventory	[ t <sub>HM</sub> ]	1.6
Life-time	[ y ]	2 500

**Tab. A.3** Radionuclide inventory

Nuclide	Half-life [ y ]	Container [ Bq ]	Repository [ Bq ]
C-14	$5.730 \cdot 10^{+3}$	$3.020 \cdot 10^{+10}$	$1.616 \cdot 10^{+14}$
Cl-36	$3.000 \cdot 10^{+5}$	$5.493 \cdot 10^{+08}$	$2.939 \cdot 10^{+12}$

Nuclide	Half-life [ y ]	Container [ Bq ]	Repository [ Bq ]
Co-60	$5.272 \cdot 10^0$	$1.634 \cdot 10^{14}$	$8.743 \cdot 10^{17}$
Ni-59	$7.500 \cdot 10^4$	$9.626 \cdot 10^{10}$	$5.150 \cdot 10^{14}$
Ni-63	$1.000 \cdot 10^2$	$1.371 \cdot 10^{13}$	$7.333 \cdot 10^{16}$
Se-79	$1.100 \cdot 10^6$	$2.796 \cdot 10^{10}$	$1.496 \cdot 10^{14}$
Sr-90	$2.864 \cdot 10^3$	$3.756 \cdot 10^{15}$	$2.009 \cdot 10^{19}$
Zr-93	$1.500 \cdot 10^6$	$1.341 \cdot 10^{11}$	$7.172 \cdot 10^{14}$
Nb-94	$2.000 \cdot 10^4$	$1.013 \cdot 10^{05}$	$5.421 \cdot 10^{08}$
Mo-93	$3.500 \cdot 10^3$	$7.148 \cdot 10^{07}$	$3.824 \cdot 10^{11}$
Tc-99	$2.100 \cdot 10^5$	$4.066 \cdot 10^{11}$	$2.175 \cdot 10^{15}$
Sn-126	$2.345 \cdot 10^5$	$4.964 \cdot 10^{10}$	$2.656 \cdot 10^{14}$
I-129	$1.570 \cdot 10^7$	$3.222 \cdot 10^{09}$	$1.724 \cdot 10^{13}$
Cs-135	$2.000 \cdot 10^6$	$3.481 \cdot 10^{10}$	$1.863 \cdot 10^{14}$
Cs-137	$3.017 \cdot 10^1$	$6.324 \cdot 10^{15}$	$3.383 \cdot 10^{19}$
Sm-151	$9.300 \cdot 10^1$	$1.092 \cdot 10^{14}$	$5.844 \cdot 10^{17}$
Ra-226	$1.600 \cdot 10^3$	$1.691 \cdot 10^{04}$	$9.048 \cdot 10^{07}$
Th-229	$7.880 \cdot 10^3$	$1.032 \cdot 10^{04}$	$5.519 \cdot 10^{07}$
Th-230	$7.540 \cdot 10^4$	$6.280 \cdot 10^{06}$	$3.360 \cdot 10^{10}$
Th-232	$1.405 \cdot 10^{10}$	$8.118 \cdot 10^{00}$	$4.343 \cdot 10^{04}$
Pa-231	$3.276 \cdot 10^4$	$2.029 \cdot 10^{06}$	$1.085 \cdot 10^{10}$
U-233	$1.592 \cdot 10^5$	$5.795 \cdot 10^{06}$	$3.100 \cdot 10^{10}$
U-234	$2.455 \cdot 10^5$	$6.609 \cdot 10^{10}$	$3.536 \cdot 10^{14}$
U-235	$7.038 \cdot 10^8$	$7.969 \cdot 10^{08}$	$4.263 \cdot 10^{12}$
U-236	$2.342 \cdot 10^7$	$1.348 \cdot 10^{10}$	$7.210 \cdot 10^{13}$
U-238	$4.468 \cdot 10^9$	$1.953 \cdot 10^{10}$	$1.045 \cdot 10^{14}$
Np-237	$2.144 \cdot 10^6$	$2.153 \cdot 10^{10}$	$1.152 \cdot 10^{14}$
Pu-238	$8.774 \cdot 10^1$	$4.747 \cdot 10^{14}$	$2.540 \cdot 10^{18}$
Pu-239	$2.411 \cdot 10^4$	$3.465 \cdot 10^{13}$	$1.854 \cdot 10^{17}$
Pu-240	$6.563 \cdot 10^3$	$8.160 \cdot 10^{13}$	$4.366 \cdot 10^{17}$
Pu-241	$1.435 \cdot 10^1$	$1.387 \cdot 10^{16}$	$7.419 \cdot 10^{19}$
Pu-242	$3.750 \cdot 10^5$	$6.035 \cdot 10^{11}$	$3.229 \cdot 10^{15}$
Am-241	$4.322 \cdot 10^2$	$3.277 \cdot 10^{14}$	$1.753 \cdot 10^{18}$
Am-242m	$1.410 \cdot 10^3$	$3.228 \cdot 10^{12}$	$1.727 \cdot 10^{16}$

Nuclide	Half-life [ y ]	Container [ Bq ]	Repository [ Bq ]
Am-243	$7.37 \cdot 10^{+3}$	$5.740 \cdot 10^{+12}$	$3.071 \cdot 10^{+16}$
Cm-243	$2.91 \cdot 10^{+1}$	$1.447 \cdot 10^{+12}$	$7.741 \cdot 10^{+15}$
Cm-244	$1.810 \cdot 10^{+1}$	$8.996 \cdot 10^{+14}$	$4.813 \cdot 10^{+18}$
Cm-245	$8.500 \cdot 10^{+3}$	$3.484 \cdot 10^{+11}$	$1.864 \cdot 10^{+15}$

**Tab. A.4** Mobilisation rates for different fuel compartments

	Fuel matrix	Metal parts	Gas space
Mobilisation rate [ $y^{-1}$ ]	$1.0 \cdot 10^{-6}$	$3.6 \cdot 10^{-3}$	instantaneous

**Tab. A.5** Relative inventory in the different fuel compartments in per cent

Element	Metal parts	Fuel matrix	Gas space
C	72.20	26.41	1.39
Cl	0.00	94.00	6.00
Ni, Mo, Nb	99.50	0.47	0.03
Sn	0.00	98.00	2.00
I, Se	0.00	97.00	3.00
Cs	0.00	96.00	4.00
Rb, H	0.00	95.00	5.00
Sr, Sm, Pb	0.00	99.90	0.10
Zr	9.40	86.07	4.53
Tc	0.10	99.89	0.01
Pd, Cm, Am, Pu, Pa, U, Th, Ac, Np, Ra	0.00	99.99	0.01

**Tab. A.6** Solubility limits in [  $\text{mol} \cdot \text{l}^{-1}$  ] (“high” denotes no solubility limit)

Element	Reference	Minimum	Maximum
H	high	high	high
Be	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	high
C <sub>org</sub>	high	high	high
Cl	high	high	high
Ca	$1 \cdot 10^{-2}$	$1 \cdot 10^{-2}$	$1 \cdot 10^{-2}$
Ni	$3 \cdot 10^{-5}$	$1 \cdot 10^{-5}$	$8 \cdot 10^{-5}$
Se	$5 \cdot 10^{-9}$	$2 \cdot 10^{-11}$	$1 \cdot 10^{-5}$

Element	Reference	Minimum	Maximum
Sr	$2 \cdot 10^{-5}$	$3 \cdot 10^{-6}$	$1 \cdot 10^{-4}$
Zr	$2 \cdot 10^{-9}$	$3 \cdot 10^{-11}$	$2 \cdot 10^{-9}$
Nb	$3 \cdot 10^{-5}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-4}$
Mo	$1 \cdot 10^{-6}$	$1 \cdot 10^{-6}$	$1 \cdot 10^{-5}$
Tc	$4 \cdot 10^{-9}$	$1 \cdot 10^{-9}$	$1 \cdot 10^{-8}$
Pd	$5 \cdot 10^{-8}$	$1 \cdot 10^{-10}$	$2 \cdot 10^{-7}$
Ag	$3 \cdot 10^{-6}$	$1 \cdot 10^{-10}$	$3 \cdot 10^{-6}$
Sn	$1 \cdot 10^{-8}$	$5 \cdot 10^{-9}$	$1 \cdot 10^{-7}$
I	high	high	high
Cs	high	high	high
Sm	$5 \cdot 10^{-7}$	$3 \cdot 10^{-7}$	$9 \cdot 10^{-7}$
Ho	$5 \cdot 10^{-7}$	$3 \cdot 10^{-7}$	$9 \cdot 10^{-7}$
Pb	$2 \cdot 10^{-6}$	$2 \cdot 10^{-8}$	$8 \cdot 10^{-5}$
Po	high	high	high
Ra	$2 \cdot 10^{-11}$	$4 \cdot 10^{-12}$	$5 \cdot 10^{-8}$
Ac	$1 \cdot 10^{-6}$	$5 \cdot 10^{-8}$	$3 \cdot 10^{-5}$
Th	$7 \cdot 10^{-7}$	$2 \cdot 10^{-7}$	$3 \cdot 10^{-6}$
Pa	$1 \cdot 10^{-8}$	$1 \cdot 10^{-8}$	$1 \cdot 10^{-5}$
U	$3 \cdot 10^{-9}$	$3 \cdot 10^{-10}$	$5 \cdot 10^{-7}$
Np	$5 \cdot 10^{-9}$	$3 \cdot 10^{-9}$	$1 \cdot 10^{-8}$
Pu	$5 \cdot 10^{-8}$	$3 \cdot 10^{-9}$	$1 \cdot 10^{-6}$
Am	$1 \cdot 10^{-6}$	$5 \cdot 10^{-8}$	$3 \cdot 10^{-5}$
Cm	$1 \cdot 10^{-6}$	$5 \cdot 10^{-8}$	$3 \cdot 10^{-5}$

**Tab. A.7** Distribution coefficients for the Lower Cretaceous Clays (all elements not included here are assumed to have  $K_d = 0$ )

Element	$K_d$ [ $m^3/kg$ ]	Element	$K_d$ [ $m^3/kg$ ]	Element	$K_d$ [ $m^3/kg$ ]
Tc	$5 \cdot 10^{-4}$	C	$2 \cdot 10^{-2}$	U	0.012
Se	$5 \cdot 10^{-4}$	Sr	$2 \cdot 10^{-4}$	Ra	$3 \cdot 10^{-3}$
Zr	$8 \cdot 10^{-2}$	Np	$6 \cdot 10^{-2}$	Ni	$1 \cdot 10^{-3}$
Nb	1	Pu	0.5	Th	$8 \cdot 10^{-2}$
Cs	$5 \cdot 10^{-2}$	Am/Cm	0.5	Pa	1
I	0	Pb	$6 \cdot 10^{-4}$	Ac	$4 \cdot 10^{-2}$

For the element-specific porosity a value of 0.16 was used for all elements except Chloride, Selenium, Molybdenum and Iodine. For these four elements anion exclusion by a factor of two was assumed and consequently a porosity of 0.08 was used.

For the pore diffusion coefficient a value of  $2.6 \cdot 10^{-10} m^2 \cdot s^{-1}$  was used for all radionuclides.

For calculating the annual dose to a human individual the biosphere dose conversion factors given in Tab. A.8 were used.

**Tab. A.8** Dose conversion factors (DCF) in  $Sv \cdot y^{-1} / Bq \cdot m^{-3}$

Activation and fission products		Th- and Np- series		U- and Ac-series	
Nuclide	DCF	Nuclide	DCF	Nuclide	DCF
C-14	$4.6 \cdot 10^{-8}$	Cm-248	$5.0 \cdot 10^{-6}$	Cm-246	$8.0 \cdot 10^{-7}$
Cl-36	$3.5 \cdot 10^{-8}$	Pu-244	$3.0 \cdot 10^{-6}$	Pu-242	$9.4 \cdot 10^{-7}$
Ca-41	$2.0 \cdot 10^{-9}$	Cm-244	$3.8 \cdot 10^{-7}$	AM-242	$7.6 \cdot 10^{-7}$
Co-60	$3.9 \cdot 10^{-6}$	Pu-240	$9.6 \cdot 10^{-7}$	U-238	$7.1 \cdot 10^{-7}$
Ni-59	$4.9 \cdot 10^{-9}$	U-236	$5.6 \cdot 10^{-7}$	Pu-238	$7.5 \cdot 10^{-7}$
Ni-63	$1.1 \cdot 10^{-9}$	Th-232	$1.1 \cdot 10^{-4}$	Th-234	$4.8 \cdot 10^{-9}$
Se-79	$3.4 \cdot 10^{-7}$	Ra-228	$2.4 \cdot 10^{-6}$	U-234	$1.4 \cdot 10^{-6}$
Rb-87	$1.3 \cdot 10^{-7}$	U-232	$5.4 \cdot 10^{-6}$	Th-230	$3.7 \cdot 10^{-5}$
Sr-90	$1.8 \cdot 10^{-7}$	Th-228	$1.3 \cdot 10^{-6}$	Ra-226	$3.0 \cdot 10^{-5}$
Zr-93	$3.7 \cdot 10^{-8}$	Cm-245	$1.4 \cdot 10^{-6}$	Pb-210	$2.3 \cdot 10^{-6}$
Mo-93	$3.2 \cdot 10^{-7}$	Pu-241	$1.8 \cdot 10^{-8}$	Po-210	$4.9 \cdot 10^{-6}$

Activation and fission products		Th- and Np- series		U- and Ac-series	
Nuclide	DCF	Nuclide	DCF	Nuclide	DCF
Nb-94	$3.1 \cdot 10^{-6}$	Am-241	$8.0 \cdot 10^{-7}$	Cm-247	$2.9 \cdot 10^{-6}$
Tc-99	$8.8 \cdot 10^{-9}$	Np-237	$4.7 \cdot 10^{-6}$	Am-243	$2.0 \cdot 10^{-6}$
Pd-107	$1.9 \cdot 10^{-9}$	U-233	$3.9 \cdot 10^{-6}$	Pu-239	$9.8 \cdot 10^{-7}$
Sn-126	$1.6 \cdot 10^{-5}$	Pa-233	$8.8 \cdot 10^{-9}$	U-235	$3.3 \cdot 10^{-6}$
I-129	$5.6 \cdot 10^{-7}$	Th-229	$1.7 \cdot 10^{-5}$	Pa-231	$4.0 \cdot 10^{-5}$
Cs-135	$5.7 \cdot 10^{-8}$	Ra-225	$1.1 \cdot 10^{-7}$	Ac-227	$1.0 \cdot 10^{-5}$
Cs-137	$9.5 \cdot 10^{-7}$	Ac-225	$3.7 \cdot 10^{-8}$	Th-227	$1.9 \cdot 10^{-8}$
Sm-151	$3.2 \cdot 10^{-10}$			Ra-223	$1.1 \cdot 10^{-7}$

## A.2 Repository for Low- and Intermediate-Level waste (LILW) in rock salt

Only the basic near field and far field data of the LILW system are given in the following. The biosphere data are identical to those described in A.1.

**Tab. A.9** Near field segment data

		EXIT	AEB	MB	RG	NAB
Height	m	10	10	100	160	5
Width	m	10	50	20	60	30
Length	m	100	250	20	300	200
Height of seal	m		2		10	10
Width of seal	m		4		10	10
Length of seal	m		25		10	10
Additional volume	m <sup>3</sup>			7000		
Porosity of add. volume	-			0.4		
Local convergence factor	-		1	0.2	1	0.2
Seal permeability	m <sup>2</sup>				1.0E-14	1.0E-14
Seal porosity	-				0.2	0.2
Vol. of backfill w/o waste	m <sup>3</sup>		34000		1750000	25000
Initial backfill porosity	-		0.39		0.4	0.4
gas storage / pore volume	-				0.2	0.3

Gas-prod. prop. of material	-		0.8		0.8	0.8
Level of lowest seal	m		12			
Porosity	-	0.3				
Non-backf. vol. below AE	m <sup>3</sup>		40000			
Backfilled vol. below AE	m <sup>3</sup>		100000			
Non-backf. vol. adj. AE	m <sup>3</sup>		1000			
Backfilled vol. adj. AE	m <sup>3</sup>		2500			
Non-backf. vol. above AE	m <sup>3</sup>		250000			
Backfilled vol. above AE	m <sup>3</sup>		250000			
Height of voids above AE	m		132			

**Tab. A.10** Global near field data

General		
Fluid density	kg/m <sup>3</sup>	1200
Fluid viscosity	Pa s	2.4E-03
Rock temperature	K	300
Hydrostatic pressure	MPa	5.0
Rock pressure	MPa	10
Stress exponent	-	5.0
Diffusion coefficient	m <sup>2</sup> /s	1.0E-09
Dispersion length	m	0
Convergence		
Reference convergence rate	1/a	4.0E-05
Reference porosity	-	0.3
Stress exponent for backfill support	-	5.0
Permeability-porosity parameter	-	7.46E-09
Permeability-porosity exponent	-	5.253
Final porosity	-	1.0E-05
Specific		
Initial backfill porosity	-	0.4
Sheeting factor	-	1.0
Initial porosity of sheeting debris	-	0.4
Initial porosity of seals	-	0.2
Initial permeability of seals	m <sup>2</sup>	1.0E-18
Volumetric dissolution capacity of brine	-	0.2
Maximum permeability increase	-	1.0E+04
viscosity of brine in unaltered seal	Pa s	2.0E-03
Gas production		
Specific gas production (metal)	mol/kg	24
Specific gas production (cellulose)	mol/kg	33.3
Specific gas production (organics)	mol/kg	143
Corrosion rate (metal)	1/a	4.0E-03
Corrosion rate (cellulose)	1/a	0
Corrosion rate (organics)	1/a	1.0E-05

**Tab. A.11** Waste package data

		NAB	AEB	RG
Matrix mass	kg	0	0	0
Container mass	kg	4.0E+05	1.0E+07	2.0E+06
Waste volume	m <sup>3</sup>	2.0E+03	3.0E+04	0
Converging void volume	m <sup>3</sup>	1.5E+03	1.5E+04	0
Non-converging void volume	m <sup>3</sup>	0	0	0
Mass of gas producing substances (1)	kg	1.0E+05	1.5E+06	4.0E+05
Mass of gas producing substances (2)	kg	3.0E+05	3.5E+06	1.0E+03
Interim storage time	a	0	0	0

**Tab. A.12** Radionuclide inventories

Radionuclide	NAB	AEB	Radionuclide	NAB	AEB
H-3	8.0E+08	2.0E+12	Cm-248	0.0E+00	0.0E+00
Cc-14	1.4E+06	4.0E+07	Pu-244	0.0E+00	0.0E+00
Cl-36	5.0E+06	8.0E+08	Cm-244	5.0E+05	2.0E+09
Ca-41	5.0E+04	7.0E+07	Pu-240	3.0E+06	1.0E+09
Co-60	2.0E+14	8.0E+12	U-236	2.0E+04	8.0E+06
Ni-59	6.0E+08	1.5E+11	Th-232	0.0E+00	3.0E+05
Ni-63	3.0E+10	2.0E+13	Cm-245	2.0E+02	2.0E+05
Se-79	3.0E+05	2.0E+08	Pu-241	2.0E+08	2.0E+11
Rb-87	3.0E+05	5.0E+03	Am-241	3.0E+07	9.0E+10
Sr-90	2.0E+09	5.5E+12	Np-237	4.0E+04	1.0E+07
Zr-93	9.0E+06	9.0E+09	U-233	5.0E+03	3.0E+03
Nb-94	4.0E+07	2.0E+10	Th-229	5.0E+03	2.0E+03
Mo-93	4.5E+05	9.0E+07	Cm-246	3.0E+00	1.0E+05
Tc-99	3.0E+08	7.5E+10	Pu-242	2.0E+04	1.0E+07
Pd-107	3.0E+04	7.0E+07	U-238	2.0E+04	2.0E+07
Sn-126	4.0E+05	2.0E+08	U-234	7.0E+04	7.0E+07
I-129	2.5E+04	9.0E+06	Th-230	1.0E+01	6.0E+05
Cs-135	5.0E+05	3.0E+08	Ra-226	0.0E+00	1.0E+10
Cs-137	5.0E+13	7.5E+13	Cm-247	0.0E+00	0.0E+00
Sm-151	5.0E+08	9.0E+10	Am-243	8.0E+03	2.0E+07

**Tab. A.13** Far field data

		Overburden	Aquifer
Layer		Grauer Salztou	Oberkreide
Length of transport path	m	0.8	0.2
Cross section of transport path	m <sup>2</sup>	150	12500
Groundwater volume flow rate	m <sup>3</sup> /a	0	15000
Dispersion length	m	4	4
Molecular diffusion coefficient	m <sup>2</sup> /a	0.016	0.016
Porosity	-	0.3	0.3
Density	kg/m <sup>3</sup>	2500	2500

**Tab. A.14** Sorption distribution coefficients (Kd) in m<sup>3</sup>/kg

Element	Overburden	Aquifer	Element	Overburden	Aquifer
C	0	0	Sn	0	0
Cc	0.003	0	I	0	0
Cl	0	0	Cs	0.003	0
Ca	1.40E-04	0	Sm	0.7	0
Co	0.014	0	Cm	0.7	0
Ni	0.014	0	Am	0.4	0.1
Se	3.50E-04	0	Pu	0.4	0.1
Rb	0.035	0	Np	0.042	0.01
Sr	1.40E-04	0	Pa	0.7	0.6
Zr	0.7	0	Th	0.7	0.2
Nb	0.7	0	U	0.2	0.002
Mo	0	0	Ra	0.004	0.04
Tc	3.50E-04	0	Pb	0.028	1
Pd	0.014	0	Po	0	1.2
Sb	0	0	Ac	0.49	0.3

**Gesellschaft für Anlagen-  
und Reaktorsicherheit  
(GRS) gGmbH**

Schwertnergasse 1  
**50667 Köln**

Telefon +49 221 2068-0

Telefax +49 221 2068-888

Boltzmannstraße 14

**85748 Garching b. München**

Telefon +49 89 32004-0

Telefax +49 89 32004-300

Kurfürstendamm 200

**10719 Berlin**

Telefon +49 30 88589-0

Telefax +49 30 88589-111

Theodor-Heuss-Straße 4

**38122 Braunschweig**

Telefon +49 531 8012-0

Telefax +49 531 8012-200

[www.grs.de](http://www.grs.de)

**ISBN 978-3-944161-94-5**