SANDIA REPORT
SAND2021-11053
Printed September 2021

Sandia National Laboratories

Sensitivity Analysis Comparisons on Geologic Case Studies: An International Collaboration

Laura P. Swiler, Dirk-Alexander Becker, Dusty Brooks, Joan Govaerts, Lasse Koskinen, Pekka Kupiainen, Elmar Plischke, Klaus-Jürgen Röhlig, Elena Saveleva, Sabine M. Spiessl, Emily Stein, Valentina Svitelman

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

Over the past four years, an informal working group has developed to investigate existing sensitivity analysis methods, examine new methods, and identify best practices. The focus is on the use of sensitivity analysis in case studies involving geologic disposal of spent nuclear fuel or nuclear waste. To examine ideas and have applicable test cases for comparison purposes, we have developed multiple case studies. Four of these case studies are presented in this report: the GRS clay case, the SNL shale case, the Dessel case, and the IBRAE groundwater case. We present the different sensitivity analysis methods investigated by various groups, the results obtained by different groups and different implementations, and summarize our findings.

ACKNOWLEDGEMENTS

This report has benefited from the insights and assistance of many. The authors thankfully acknowledge Christi Leigh at Sandia National Laboratories for supporting this effort in its early stages. The authors are grateful to Laura Swiler, Klaus-Jürgen Röhlig (TU Clausthal) and Dirk-Alexander Becker (GRS) for their leadership. We also appreciate the feedback from Sergei Kucherenko with regards to metamodeling techniques and sensitivity analysis. This report has been a largely volunteer effort and we gratefully thank all the participants and reviewers for their time and effort.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
ANOVA	Analysis of variance
BE	Best estimate
BSPCE	Bayesian sparse polynomial chaos expansion
CC	Correlation coefficient
CDF	Cumulative distribution function
CHETLIN	Software code
CLAYPOS	Software code
COSI	Cosine-transformation-based algorithm for sensitivity indices
CUSUNORO	Cumulative sum of normalized reordered output
DCT	Discrete cosine transformation
DGSM	Derivative-based global sensitivity measure
DNN	Deep neural networks
DRZ	Disturbed rock zone
EASI	Effective algorithm for sensitivity indices
EES	Expected evolution scenario
eFAST	Extended Fourier amplitude sensitivity test
EGU	European geosciences union
ENSI	Swiss federal nuclear safety inspectorate
ESREL	European Safety and Reliability Conference
EXMAS	Software code
FANC	Belgian federal agency for nuclear control
FAST	Fourier amplitude sensitivity test
FCRD	Fuel cycle research and development
FF	Factor fixing
FFT	Fast Fourier transform
FORM	First-order-reliability-method
FP	Factor prioritization
GDSA	Geologic disposal safety assessment
GENESIS	Disposal project
GeRA	Software code
GP	Gaussian process
GRS	Gesellschaft für Anlagen- und Reaktorsicherheit
HDMR	High dimensional model representation
HLW	High level waste

Abbreviation	Definition
IAEA	International atomic energy agency
IBRAE	Nuclear safety institute of the Russian academy of sciences
IGSC	integration group for the safety case
IHLRWM	International High-Level Radioactive Waste Management Conference
INTESC	International Experiences in Safety Case for Geological Repositories
IRSN	Institute for Radiological Protection and Nuclear Safety (France)
KS	Kolmogorov-Smirnov
LILW	Low and intermediate level waste
MARS	Multivariate adaptive regression splines
MC	Monte-Carlo
MIM	Borgonovo moment independent importance measure
MOSEL	Investigation of modern methods of probabilistic sensitivity analysis of final repository performance assessment models
MTHM	Metric ton of heavy metal
NEA	Nuclear Energy Agency
NM	New Mexico
OECD	Organization for Economic Cooperation and Development
ONDRAF/NIRAS	Belgian National Agency for Radioactive Waste and enriched Fissile Material
PA	performance analysis
PAWN	Pianosi and Wagener Importance Measure
PCE	Polynomial chaos expansion
PC-PRCC	Partial (rank) correlation coefficient
PEAR	Pearson correlation coefficient
PFLOTRAN	Groundwater flow and transport code developed by DOE
POSIVA	Finnish company focused on final disposal solutions
PWR	Pressurized water reactor
QMC	Quasi Monte Carlo
RCC	Rank correlation coefficient
RI	Regionalized information
RBD	Random balanced design
RN	Radionuclide
RS-HDMR	Response Surface – High dimensional model representation
SA	Sensitivity analysis
SCK-CEN	Belgian nuclear research centre
SD	Structure discovery

Abbreviation	Definition
SFWST	Spent fuel and waste science technology
S_i	Sensitivity index first-order (also called the main effect index) for variable i. Sometimes also denoted as S_1 in the text.
S_{i2}	Sensitivity index second-order. This is the specific case of the more general S_α, where α represents the interaction of variable i with other variables j.
SIAM/ASA	Society of Industrial and Applied Mathematics/American Statistical Association
SNF	Spent nuclear fuel
SNL	Sandia National Laboratories
SORM	Second-order-reliability-method
SPEA/SPEAR	Spearman correlation coefficient
SRC	Standardized regression coefficient
SRRC	Standardized rank regression coefficient
SSG	Specific safety guide
SSR	Specific safety requirements
T_i	Sensitivity index total order (also called total effect index). Sometimes also denoted as S_T in the text.
TI	Trend identification
TN	Tennessee
TUC/TU Clausthal	Technische Universität Clausthal (Clausthal University of Technology)
UFD	The U.S. Dept. of Energy Used Fuel Disposition
UK	United Kingdom
UQ/SA	Uncertainty quantification and sensitivity analysis
URF	Underground research facility
USA	United States of America

This page is intentionally left blank

1. INTRODUCTION

Uncertainty quantification and sensitivity analyses (UQ/SA) carried out in the context of the safety assessment or performance assessment (PA) of a radioactive waste repository assist in understanding the system relationships, optimizing resources, and building confidence in the performance of the system and in meeting certain set requirements (e.g., compliance demonstrations). Uncertainty management is the process of continuous assessment of the uncertainties within a repository project, evaluation of their consequences, and of efforts for reducing, avoiding, or mitigating uncertainties.

Sensitivity analysis, the identification of important variables affecting uncertainty in the assessment results, is a critical activity in the uncertainty analysis process. [1] lists the following purposes and contexts in which methods of sensitivity analysis might prove useful:

to build and explore consequence models;
to explore the relationship between science and consequence models;
to support the elicitation of judgmental inputs to an analysis;
to develop efficient computational algorithms;
to design experiments;
to guide the making of inferences, forecasts and decisions;
to explore and build consensus;
to build understanding.

Hence, SA methods can be embedded in different stages of a decision-making context. It is essential, however, to distinguish between the repository system itself and the numerical model used to describe it. The latter is just a simplified image of the real system, depending on a limited number of input parameters that are related but, in general, not identical to the physical variables governing the system. The real uncertainties have to be mapped appropriately to uncertainties of the model parameters. Whenever doing numerical investigations one should not lose sight of the fact that one is actually investigating the model and not the real system.

1.1. Goal of the report

The purpose of this report is to provide summaries of a ***sensitivity analysis exercise*** developed by an international working group. The goal of the exercise is to gain a better understanding of the strengths and weaknesses of various SA methods, identify cost vs. performance tradeoffs of the methods, and highlight best practices and lessons learned. Multiple countries participated and demonstrated various SA methods on a series of case studies. For each case study, each group presented its results using different sensitivity analysis methods and/or different implementations of the same method. The breadth and scope of the case studies as well as the large variety of sensitivity analysis methods used provided a rich environment to study and compare results.

The case studies involve computational models addressing safety assessments for geologic disposal of radioactive waste. More detail about the cases is presented in subsequent chapters. We note that many studies have compared sensitivity analysis methods on analytic test functions. For example, [2] presents a variety of surrogate methods used in variance-based SA calculations and demonstrates results on canonical problems such as the Ishigami test function. [3] presents results

of finding "effective" dimensions of a model using sampling and SA approaches on three classes of analytic test functions. [4] presents density-based SA methods which extend the variance-based approaches that have become extremely popular over the past twenty years. And the recent paper by Puy et al. [5] provides results of an extensive benchmarking exercise comparing various implementations of variance-based SA estimators along with comparisons of their performance on ten test functions based on various factors such as input distribution type, number of runs, number of pairwise interactions, etc. We support and advocate for more of these studies. In this report, however, we focus on the demonstration of SA methods on realistic problems that are of interest to the radioactive waste management community. Demonstration on problems of realistic scope and scale highlights challenges and aspects that may not be faced on analytic test problems.

1.2. Safety Assessment for Geologic Disposal

The formal concept of a safety case for the long-term disposal of Spent Nuclear Fuel (SNF) and High Level Waste (HLW) in an engineered facility located in a deep geologic formation was first introduced by the Nuclear Energy Agency (NEA) [1] . Initial discussion and documentation on the topic as well as setting of standards continued in [7, 8, 9]. More recently, and in concert with the development of numerous national safety cases, there have been a number of international symposia, conferences, working groups, summary papers, and safety standards devoted to understanding, developing, and/or summarizing the nature, purpose, context, and elements of safety cases (e.g., [9 – 17]). In these recent summary and overview reports, it is observed that there is notable convergence in the understanding and development of safety case documents published by national and international organizations. The following excerpt from [14] (Section 3.1) provides a definition of a safety case that is current and consistent with the aforementioned documents:

The safety case is an integration of arguments and evidence that describe, quantify and substantiate the safety of the geological disposal facility and the associated level of confidence.

The concept of a post-closure safety assessment is described by the intergovernmental Nuclear Energy Agency [15] (Section 5.1) as an iterative set of assessments for evaluating the performance of a repository system and its potential impact that aims to provide reasonable assurance that the repository system will achieve sufficient safety and meet the relevant requirements for the protection of humans and the environment over a prolonged period. A safety assessment, (i) quantifies the repository system performance for all selected situations and (ii) evaluates the level of confidence (taking into account the identified uncertainties) in the estimated performance of the system. As examples, a description of the regulatory frameworks for the U.S. Waste Isolation Pilot Plant [18] and Yucca Mountain [19] are provided. The U.S. Dept. of Energy Used Fuel Disposition (UFD) campaign and Spent Fuel and Waste Science Technology campaign (SFWST) have published safety case overviews relevant to geologic disposal in the U.S. [20, 21].

At a high level, all countries describe uncertainty using qualitative and quantitative uncertainty management. In the framework of modeling for the purpose of safety or performance assessment, *uncertainty quantification* is described in the context of data representation, propagation, and analysis. That is, first one needs to identify and represent or characterize the uncertain inputs of a computational model (through distributions, intervals, empirical data). In this report, we note that inputs, parameters, factors, and variables all refer to the same thing: uncertain inputs to a model. Then, one needs a method to propagate input uncertainties through the model to obtain

uncertainties on output quantities and results of interest. This is frequently done by sampling several realizations for the input parameter sets and performing model calculations for each of them. Finally, the results need to be postprocessed; if the sampling followed a prescribed joint probability distribution for the model input representing its uncertainty, this is where statistical analyses are usually done. A related activity is often done concurrently: *sensitivity analysis*. One identifies the most important parameters affecting the results through sensitivity analysis. The goal of sensitivity analysis is to determine the contributions of individual uncertain analysis inputs to the total uncertainty in analysis results of interest [22, 23]. There is an extensive literature about the treatment of uncertainty for risk assessment and performance assessment. References [24 – 28] are representative of this broad field.

1.3. Sensitivity Analysis Working Group

Over the past five years, an informal working group has developed to investigate existing sensitivity analysis methods, examine new methods, and identify best practices. A series of annual meetings was held starting with a "Workshop on Handling Uncertainties" in September 2015 at Harwell, UK, followed by an "International Workshop on Sensitivity Analysis of Final Repository Systems" in Braunschweig, Germany in October 2016, and then by a "Quantification of Uncertainty Workshop" in August 2017 in Albuquerque, NM, USA. It was at this latter workshop in the U.S. where more collaborative work was initiated among the international participants and organizations, as discussed below.

A follow-on workshop in Brussels, Belgium in October 2018 led to a more formal establishment of this collaboration during the subsequent Integration Group for the Safety Case (IGSC) Symposium on the Safety Case in Rotterdam, The Netherlands in October 2018. The sensitivity analysis group is working under the auspices of Organization for Economic Cooperation and Development (OECD)/ NEA's Integration Group for the Safety Case (IGSC, https://www.oecd-nea.org/rwm/igsc/).

The Uncertainty Quantification workshop held in August 2017 in Albuquerque, USA, led to a joint sensitivity analysis exercise effort initiated in October 2017, with participation from GRS, Posiva, SCK-CEN, SNL, and TUC, and later joined by IBRAE as well as by ENSI, Fortum, IRSN and Nagra as observers. The Sensitivity Analysis subgroup discussed various case studies that could be examined from a safety assessment context for geologic disposal. These case studies represent different modelled systems as a function of time, with varying levels of detail, complexity, data uncertainty, and spatial extent. The Sensitivity Analysis subgroup identified seven test cases ranked in order of complexity and since then has compared analysis methods on progressively more complex models. Each case study owner provided input and output datasets and explanation of the case so that others could use the case to demonstrate their sensitivity analysis approaches. The group shared results at the next working group meeting—International SA-UQ Workshop—held in Brussels, Belgium, in 2018. These activities resulted in several papers summarizing interim results at the International SA-UQ workshop in Brussels in 2018. Further workshops were held in Berlin (November 2019) and online (March 2020, June 2020).

The concept of the study, the group's work, and interim results were presented at several occasions, e. g. at the Integration Group for the Safety Case (IGSC) Symposium held in Rotterdam in October 2018 [29 – 31]. Subsequently, a few members participated in the International High-Level

Radioactive Waste Management Conference held in (Knoxville TN, April 2019) [32, 33], the 2019 General Assembly of the European Geosciences Union (Vienna, April 2019) [34] and at the Ninth International Conference on Sensitivity Analysis of Model Output (Barcelona, October 2019) [35–38].

The focus of the SA working group is on the use of sensitivity analysis in case studies involving geologic disposal of spent nuclear fuel with the overall aim of providing guidelines for performing such analyses in the context of safety assessments or safety cases for geological repository facilities. To examine ideas and have applicable test cases for comparison purposes, we have selected multiple existing case studies. Four of these case studies which are based on less complex models are presented in this report: the GRS clay case, the SNL shale case, the Dessel case, and the IBRAE groundwater case. Multiple groups examined different sensitivity analysis techniques on each case. We present the different sensitivity analysis methods, the results obtained by different groups and different implementations, and summarize our findings.

When the working group started to collaborate, we identified several items of interest:

- Exercise planning: identification and description of various test cases, with participating organizations providing computational simulation run results (tables of runs including inputs and outputs).
- Exchange of ideas on SA theory, new developments, as well as practical challenges.
- Evaluation of SA method performance: demonstration of various SA methods on the case studies.
- Formulate SA guidelines: identification of best practices and lessons learned.
- Produce joint papers, attend conference and symposia.

Overall, this informal group has worked collaboratively to meet the goals outlined above. In addition to several meetings, we held many online videoconferences which are becoming slightly better in the era of limited travel. We anticipate this report to be the first of two volumes, where the second volume will describe SA results on more complicated cases, possibly together with a synthesis and recommendations on the application of SA in safety cases.

Chapter 2 details the SA methods investigated by the working group. They include correlation analysis, linear and rank regression, higher-order regression, variance-based indices (Sobol' indices), and graphical approaches (e. g. using CUSUNORO curves). The variance-based indices are calculated in a variety of ways. Traditionally, such calculations are based on specific sampling schemes. The working group, however, focused on existing (given) data and applied other methods such as EASI, COSI or approaches employing surrogates or metamodels such as Gaussian processes or polynomial chaos expansion.

Chapter 3 provides an overview of how the cases were selected. The four case studies are the GRS clay case, the SNL shale case, the Dessel case, and the IBRAE groundwater case. These are described in Chapters 4-7, respectively. The SA results by different teams and different methods are also presented in the chapter where the case is described. Finally, Chapter 8 provides a summary of the results and ideas about next steps.

2. SENSITIVITY ANALYSIS METHODS

There are a few general texts dealing with sensitivity analysis (SA) [39 – 43]. Sensitivity analysis and uncertainty analysis are often performed concurrently, but we emphasize they have different purposes. Uncertainty analysis involves the propagation of uncertainties on input parameters to the resulting uncertainty on output quantities. This is frequently done with sampling. In sensitivity analysis, the goal is to identify the most influential parameters affecting uncertainty in the simulation results. This is often done using samples generated as part of an uncertainty analysis process.

Uncertainty analysis is the description of the output of a simulation model under input uncertainty. In order to explore the whole input parameter space, Monte-Carlo methods (MC) are used. Monte Carlo methods sample the input uncertainty and run the model simulation for each input realization to obtain a sample approximation of the output distribution. There are other "forward uncertainty" propagation methods, including reliability methods such as the First-Order Reliability Method (FORM) and Second-Order Reliability Method (SORM) [44]. These methods are very efficient at finding tail probabilities. There are also methods such as polynomial chaos expansions and stochastic differential equations [45, 46], which propagate the uncertainty through the model, but these can be more intrusive to the codes and have generally not been adopted for performance assessment codes in the radioactive waste management community.

Sensitivity analysis uses tools for apportioning output properties to the different inputs. For sensitivity methods, one can roughly distinguish between [47]:

- Local analysis: The model is analyzed locally with respect to a reference/working point, to identify the steepest gradient as most important direction of change, apportioning local change to the different inputs.
- Screening methods: The model is analyzed with known bounds on the input parameters, to identify (screen out) input parameters with little influence, apportioning global change to the different inputs.
- Global methods: The model is analyzed with a known distribution of the inputs, apportioning the uncertainty in the output to the different inputs.

In this report, we focus on uncertainty analysis and global sensitivity methods as both work with specified input probabilities generated by Monte Carlo sampling approaches. For linear models, local and global analysis coincide. Moreover, screening methods may be applied locally, e.g. varying one factor at a time by a small amount, or globally, e.g. as in sequential bifurcation, where the model outputs from (carefully chosen) extreme input positions are compared. A table showing the various sensitivity analysis methods used in analyses of these cases is shown in Table 2-1, with a categorization of classes of methods. Note that in this report, the terms "parameters", "variables", "factors" and "inputs" are used interchangeably, and all refer to input parameters of a computational code.

Sensitivity analysis approaches		Factor Prioritization	Factor Fixing	Trend Identification	Structural Discovery	Regionalized Information	Number of model evaluations (where M is the number of input factors)
Graphical	Scatterplots	✓		✓		✓	>100
	Cumulative Sum of Normalized Reordered Output (CUSUNORO)	✓		✓	✓	✓	>100
Correlation & Regression analysis	Pearson correlation & Partial Correlation	✓		✓			>100
	Spearman Rank Correlation & Partial Rank Correlation	✓		✓			>100
	Regression coefficients (Linear, Rank, Stepwise)	✓		✓			>100
Variance-based	Sobol' indices	✓	✓		✓		>500xM
	Fourier Amplitude Sensitivity Test (FAST), extended FAST (eFAST)	✓	✓		✓		> 500xM
	Effective Algorithm for Sensitivity Indices, Cosine Sensitivity (EASI, COSI)	✓			✓		>500
	Random Balance Designs	✓			✓		>500
Moment-independent	Borgonovo's δ	✓	✓		✓	✓	>1000
	Pianosi and Wagener (PAWN)	✓	✓		✓	✓	>500xM

Table 2-1. Sensitivity Analysis Method Classification

The problem setting for global sensitivity analysis is twofold [48]: A description of the input uncertainty is needed, as well as a simulation model representing physical properties. In accordance with [39] we understand by SA addressing both aspects and considering the input uncertainty given by probability distributions. There are different serving objectives for

performing global sensitivity analyses. The methods can be framed into different sensitivity analysis settings [47, 49]:

- Factor prioritization (FP): Finding the most important input parameters.
- Factor fixing (FF): Finding the least important input parameters.
- Trend identification (TI): Identify monotonicity or convexity properties of the model.
- Structure discovery (SD): Uncover additivity, linearity, interactions.
- Regionalized information (RI): Finding active regions of input parameters.

To fix notation, we consider a (deterministic) simulation model $g: x = (x_1, \cdots, x_d) \mapsto y = g(x) = g(x_1, \cdots, x_d) = g(x_\alpha : x_{\neg\alpha})$ mapping d input parameters into a scalar output. We may split the inputs into groups where α is a subset of indices of input parameters (the group of interest) and ¬α is the complementary index group. The notation $(x_\alpha : x_{\neg\alpha})$, using Owen's smush operator [50], denotes the out-of-order concatenation of functional arguments, vectors and sample matrices. In case the physical model produces multivariate (e.g., spatio-temporal) output, we assume that g produces a scalar quantity of interest (maximum of a time series, value at a specific time/location).

We model the uncertainty by considering a *d*-dimensional random vector of inputs $X = (X_1, \cdots, X_d)$. Then the quantity of interest of the model output *Y=g(X)* is a scalar random variable. The probability densities of the input f_X and hence the marginal input densities f_i are assumed to be known. In case of independent inputs, $f(x) = \prod_{i=1}^{d} f_i(x_i)$ holds. The density of the output f_Y is generally not analytically known, its properties are estimated from an input sample promoted through the simulation model (Monte Carlo method). The associated cumulative distribution functions (CDFs) of an individual parameter *i*, of the full random vector of inputs *X* and of the output *Y* are denoted by F_i, F_X, F_Y, respectively. Sample matrices use the same notation as the random variables or vectors, i.e. $X \in R^{N \times d}$ is an input sample with N realizations (observations/runs).

Jim Gray of Microsoft Research coined the term Data-Driven Science for unifying empirical, theoretical and computational paradigms for scientific exploration. Within this framework, we can interpret simulation using Monte Carlo techniques as a way of getting from (possibly black-box) models to data. If we employ given data techniques for performing Global Sensitivity Analysis, then this step might be summarized as data-to-structure. With SA becoming part of a decision-making context, we obtain a route from simulation to data to information to knowledge to decisions.

In the radioactive waste management community, with time frames for different (national) radioactive waste disposal programs and license applications ranging in the order of decades, the use of data seems to be more persistent than the use of simulation models (with their dependence on implementation specifications such as programming languages, software packages, operating systems, hardware, etc.). A re-analysis of the data is possible even if the hardware on which the simulation model ran has long been put out-of-service.

Hence, this report draws special attention to sensitivity analysis methods that will work on given datasets (e.g. sample matrices with rows consisting of realizations of the inputs and the associated output results from a simulation model).

2.1. Correlation analysis

The Pearson correlation coefficient is a measure of the strength and direction of a linear relationship between two variables. Given a random pair of inputs/output (X, Y), the correlation between input factor i and the output is given by $\rho_i = \rho(X_i, Y) = \frac{1}{\sqrt{\mathbb{V}X_i \mathbb{V}Y}}(\mathbb{E}X_i Y - \mathbb{E}X_i \mathbb{E}Y)$ where $\mathbb{E}$ denotes the expectation and $\mathbb{V}$ the variance. Therefore, it compares the mean of the joint distribution with the product of the marginal means. In linear algebraic terms, it is the cosine of the angle between the sampled input vector of interest and the sampled simulated output.

2.1.1. Pearson

In the following, we will present key properties of sensitivity methods in standardized fashion, offering information boxes. The scope entry refers to the use of the method in different settings, depending on the framing of the analysis. Here we distinguish between factor fixing (FF), factor prioritization (FP), trend identification (TI), structure discovery (SD) and regionalized information (RI). The status entry is distinguishing between established methods (e.g. available in spreadsheet software or available in all sensitivity analysis software), emerging (available in dedicated sensitivity analysis software), early adopter and experimental. If the method works only under statistical independence of the random input variables this is remarked under input probability. If the method may be used to detect statistical independence between single input factors and the output factor (for use with FF) or if insignificance under this method is not independence this is found in the notes and caveat sections.

Sensitivity Method	Pearson (Product Moment) Correlation Coefficient		
Acronym	PEAR	**Symbol**	ρ
Scope	SD: Linear Dependence, Additivity; TI: Monotonicity; FP		
Type	Correlation/Regression		
Status	Established		
Given Data	Yes	Evaluation Costs	Least Squares Regression
Input Probability	No assumptions, one-dimensional technique		
Notes	Visual interpretation from scatterplots		
Caveats	Not a measure of independence Unclear interpretation in case of dependent inputs Dependence on probabilistic assumptions sometimes hidden		
References	[51, 52]		

2.1.2. Spearman

The Spearman correlation coefficient is similar to Pearson, but instead of measuring the correlation in the raw values, it measures correlations between the ranks of the values. If one replaces the inputs and outputs by their ranks (i.e. the smallest values are assigned to 1, the second smallest to 2, the largest but 1 to *N*-1, the largest to *N*), the inverse (empirical) cumulative distribution function of the output can be calculated, $V = \hat{F}_Y^{-1}(Y)$, where the empirical inverse CDF can be estimated as the index in the sorted sample (=rank) divided by sample size. The inverse CDFs of the inputs, $U_i = F_i^{-1}(X_i)$, are often known but may also be approximated by their empirical versions. One then performs the sensitivity analysis in these new coordinates which are now bounded between 0 and 1 and uniformly distributed to obtain transformation invariant measures. The most common of these is Spearman' rank correlation which is the correlation coefficient between U_i and V.

Sensitivity Method	Spearman Rank Correlation Coefficient		
Acronym	SPEA/SPEAR	**Symbol**	ρ^*
Scope	SD: Additivity; TI: Monotonicity, FP		
Type	Correlation/Regression		
Status	Established		
Given Data	Yes	Evaluation Costs	Least Squares Regression
Input Probability	No assumptions, one-dimensional technique		
Notes	Visual interpretation from dependograms Nonparametric technique Related measures like Kendell τ^* have not been considered		
Caveats	Not a measure of independence Unclear interpretation in case of dependent inputs Dependence on probabilistic assumptions sometimes hidden		
References	[52, 53]		

2.1.3. Partial

Partial correlation indicates the amount of association between two random variables x_a and y, controlling for the effect of other random variables, $x_{\neg a}$ which may confound a simple correlation analysis. For partial correlation, one builds two linear regression models (see below). The first is between x_a and $x_{\neg a}$ and the second is between y and $x_{\neg a}$. The Pearson correlation between the residuals of these two regressions is the partial correlation. This second regression removes the input factor of interest from the list of feature maps. The partial correlation coefficient provides a

measure of the linear relationship between y and x_a with the linear effects of the other variables $x_{\neg a}$ removed A version working on ranks is known as Partial Rank Correlation Coefficient.

Sensitivity Method	Partial (Rank) Correlation Coefficient		
Acronym	PCC/PRCC	**Symbol**	
Scope	SD: Linear Dependence, Additivity; TI: Monotonicity		
Type	Correlation/Regression		
Status	Established		
Given Data	Yes	Evaluation Costs	Least Squares Regression
Input Probability	No assumptions, multi-dimensional technique		
Notes			
Caveats	Not a measure of independence PRCC is a nonparametric technique Unclear interpretation in case of dependent inputs Dependence on probabilistic assumptions sometimes hidden		
References	[52, 53]		

2.2. Regression

Regression is a statistical procedure to estimate a functional relationship between a dependent variable (the output of interest) and independent variables (the input parameters). A general linear regression model (also called response surface) assumes a linear dependence between the inputs and the output which is of the form

$$\eta(Y) = \sum_i \beta_i \varphi_i(X) + \varepsilon$$

where η is the link function, φ_i are the feature maps (also called basis functions), β_i are the unknown regression coefficients and ε is a zero-mean error independent of the inputs X. Minimizing the mean square error yields a least square regression problem. For this, the images of the feature maps for the input sample are collected in a design matrix D and the least squares problem is solved by $= (\mathrm{D}^{\mathrm{T}}\mathrm{D})^{-1}\mathrm{D}^{\mathrm{T}}\eta(Y)$. Note that there are specialized solvers available, so one should never form this projection matrix explicitly. The prediction from the regression model is $\hat{y} = \mathrm{D}\beta$. Other minimization problems (e.g. using absolute sums instead of sums of squares) lead to slightly different results.

2.2.1. Linear

For standardized regression coefficients, one uses the standardization for link and features,

$\eta(y) = \frac{y-\mu_Y}{\sigma_Y}$ and $\varphi_i(x) = \frac{x_i-\mu_i}{\sigma_i}$ and reports β_i.

Sensitivity Method	Standardized Regression Coefficient		
Acronym	SRC	**Symbol**	β_i
Scope	SD: Linear Dependence, Additivity; TI: Monotonicity; FP		
Type	Correlation/Regression		
Status	Established		
Given Data	Yes	Evaluation Costs	Least Squares Regression
Input Probability	No assumptions, multi-dimensional technique		
Notes	Estimates the same value as PEAR under independence Visual interpretation from scatterplots		
Caveats	Not a measure of independence Unclear interpretation in case of dependent inputs Dependence on probabilistic assumptions sometimes hidden		
References	[54, 55]		

2.2.2. Rank

Analogous to the Spearman coefficient, one can perform linear regression with inputs and outputs being transformed to their ranks. Under independence of inputs, the Standardized Rank Regression Coefficients (SRRC) estimate the same value as Spearman rank correlation. Using ranks yields a nonparametric technique.

2.2.3. Stepwise

Up to now, we considered only regression models that included one, all, or all-but-one factors. Especially in a high-dimensional setting, one may also add a single factor at a time to the regression model, and monitor the change in the coefficients and the goodness-of-fit to decide if this factor is uninfluential (under a linear model assumption).

The coefficient of determination is one measure of goodness-of-fit that can be applied in a stepwise analysis. This coefficient is denoted R^2. For n simulations, let y_i denote the quantity of interest

from the i^{th} simulation, $\bar{y}$ denote the average of all y_i for $i = 1, 2, \ldots, n$, and $\hat{y}_i$ denote prediction of y_i from the linear regression. Then, R^2 is defined as:

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

In a stepwise analysis, R^2 can be re-calculated after each independent variable is added to the regression model. If we denote the R^2 calculated when the j^{th} independent variable is added to the model as ${R_j}^2$, then the incremental R^2 for the that variable is ${R^2}_j - {R^2}_{j-1}$ when j>1 and ${R^2}_j$ when $j = 1$. A high incremental R^2 for a variable indicates that the goodness-of-fit of the regression model improved when that variable was added to the model.

2.2.4. Higher Order

In the regression model, one can include higher order polynomial terms as additional feature maps, including also cross-product terms to capture interactions. Goodness-of-fit measures then provide feedback about the model properties. Note that power functions are numerically ill-conditioned, and one should use orthogonal function systems as a basis for feature maps (see below).

2.3. Variance-based indices

Variance-based indices apportion the output variance to the contributions from various input parameters. Under input independence each square integrable function can be decomposed into orthogonal functions (with respect to the input probability),

$$\begin{aligned} g(x) &= g_0 + \sum_{\mathrm{i}=1}^{\mathrm{d}} g_i(x_i) + \sum_{\mathrm{j}>\mathrm{i}} g_{i,j}(x_i, x_j) + \sum_{\mathrm{k}>\mathrm{j}} g_{i,j,k}(x_i, x_j, x_k) + \cdots + g_{1,2,\ldots,d}(x) \\ &= g_0 + \sum_{\mathrm{i}=1}^{\mathrm{d}} \sum_{|\alpha|=\mathrm{i}} g_\alpha(x_\alpha) \end{aligned}$$

such that the output variance can be decomposed into $\mathbb{V}[Y] = \sigma_Y^2 = \sum_{\mathrm{i}=1}^{\mathrm{d}} \sum_{|\alpha|=\mathrm{i}} \sigma_\alpha^2$ where the contribution of input group α is defined recursively via $\int \mathrm{g}_\alpha(x_\alpha)^2 f_\alpha(x_\alpha)\mathrm{d}x_\alpha = \sum_{\beta\subset\alpha} \sigma_\beta^2$. It also holds that $\mathbb{V}[\mathbb{E}[Y|X_\alpha]] = \int g_\alpha(x_\alpha)^2 f_\alpha(x_\alpha)\mathrm{d}x_\alpha$, the conditional output variance given the input group of interest is obtained by integration from this functional decomposition. This decomposition is called functional analysis of variance (ANOVA).

The relative contribution to the output variance is then the variance-based Sobol' effect for the index group α, $\mathrm{S}_\alpha = \frac{\sigma_\alpha^2}{\sigma_Y^2}$. The first order effect is given by $\mathrm{S_i} = \mathrm{S}_{\{\mathrm{i}\}}$ and the total effect is given by $\mathrm{T_i} = \sum_{\mathrm{i}\in\alpha} \mathrm{S}_\alpha$. Further notions for groups α in this context are the subset importance $\sum_{\beta\subset\alpha} \mathrm{S}_\beta$ and the superset importance $\sum_{\alpha\subset\beta} \mathrm{S}_\beta$. The main effective dimension is $\sum_{\mathrm{i}} \mathrm{T_i} = \sum_\alpha |\alpha| \mathrm{S}_\alpha$ where $|\alpha|$ is the number of indices in the group α. The Shapley value $\mathrm{Sh_i} = \sum_{\mathrm{i}\in\alpha} \frac{1}{|\alpha|} \mathrm{S}_\alpha$ attributes a fair share to each input factor with $\sum_{\mathrm{i}=1}^{\mathrm{d}} \mathrm{Sh_i} = 1$ [50].

The total effects are the gold standard in a FF setting, whereas the first order effects (also called correlation ratios) yield information on FP and additivity, and group effects give information on interactions in SD. The main effective dimension yields information on the complexity of the model; Shapley values offer further insights on interactions.

2.3.1. Sobol'

Estimation of the variance contribution via the functional ANOVA decomposition is cumbersome. Sobol' introduced a special design (later called pick'n'freeze) that allows one to estimate effectively first and total effects by computing the correlation between the pick'n'freeze sample output vector and the reference sample output vector: Given two independent input samples X^A and X^B and a group α of inputs of interest one has $\mathbb{V}[\mathbb{E}[Y|X_\alpha]] = \rho^2(g(X^A), g(X_\alpha^A : X_{\neg\alpha}^B))\mathbb{V}[Y]$. Group sizes of one give rise to first order effects, group sizes of d-1 yield total effects. Note that switching from first to total is accomplished by exchanging A and B sample blocks, without re-evaluation of the pick'n'freeze sample. Furthermore, Quasi-Monte Carlo Sampling techniques are used for decreasing the numerical errors in variance estimation.

Sensitivity Method	Sobol' Method for Variance-Based Sensitivity Effects, Ishigami-Homma-Saltelli method, Jansen estimator		
Acronym	Pick'n'freeze	**Symbol**	S_i,T_i
Scope	FP: Functional Dependence; SD: Additivity, Interactions; FF via total effects		
Type	Variance-based		
Status	EmergingT_i		
Given Data	No	Evaluation Costs	Sample block size times Dimension
Input Probability	Independent input distributions		
Notes	Best available estimation technique (within this class) still under scientific discussion Extension to group sensitivity possible Iterative enlargement of sample size possible T_i a gold standard for FF		
Caveats	S_i not a measure of independence Unclear interpretation in case of dependent inputs Variance used as measure of uncertainty For first order effects, algorithms with better convergence properties are available QMC for small sample size and large dimension might fail Quadratic decay in the coefficients only for continuous models		
References	[39 –41]		

2.3.2. FAST / EFAST

Another approach to variance decomposition is provided by Parseval's Theorem (a functional equivalence to the Pythagorean Theorem). Given an orthonormal functional basis, $\{\varphi_i: R^d \to R\}$, such that $g(x) = \sum_i \beta_i \varphi_i(X)$ then $\sigma_Y^2 = \sum_i \beta_i^2$ (this is sometimes called spectral energy, hence the methods are called spectral methods). If the decomposition also contains univariate, bi-variate, tri-variate, etc. input functions then the individual, pairwise, triple contributions from inputs or groups of inputs to the output variance can be identified.

The FAST method constructs its functional decomposition with the help of a Fourier transformation: Each input factor is assigned a frequency. The frequency contribution from the output in resonance with this input frequency is assigned to first order effects (sum of squares of the coefficients of the basic frequency and few of its higher harmonics) while the superposition principle allows for identification of interactions. For total effects, EFAST assigns a high frequency to the factor of interest while all other factors are set to a low frequency. Then all variance contributions above a threshold frequency can be assigned the factor of interest and its interaction with other terms. The base functions used here are all dependent on a single parameter which is an artificial time argument. We may therefore speak of the realizations being sampled signals. The FAST methods then separate the output sample variance into noise and signal part, the latter of which can be attributed to a specific input frequency.

Sensitivity Method	(Extended) Fourier Amplitude Sensitivity Test		
Acronym	(E)FAST	**Symbol**	S_i, T_i
Scope	FP: Functional Dependence; SD: Additivity, Interactions; FF via total effects		
Type	Variance-based		
Status	Emerging		
Given Data	No	Evaluation Costs	FFT, dependent on Input dimension for EFAST
Input Probability	Independent input distributions		
Notes	Iterative enlargement of sample size not available Nyquist Frequency of Shannon Sampling Theorem gives precision of algorithm		
Caveats	Frequency selection scheme produces sometimes samples of questionable quality Variance used as measure of uncertainty Quadratic decay in the coefficients only for continuous models		
References	[56 – 58]		

2.3.3. RBD

A simplification of FAST is the RBD method that uses just one basic frequency but permutes the associated model output for the different inputs. Undoing the permutation associated with factor, say i, on the output, one can identify via a FFT analysis the output variance in resonance with this factor. This gives estimates for first order effects.

Sensitivity Method	Random Balanced Design		
Acronym	RBD	**Symbol**	S_i
Scope	FP: Functional Dependence; SD: Additivity		
Type	Variance-based		
Status	Emerging		
Given Data	No	Evaluation Costs	Flexible sample size, FFT
Input Probability	Independent input distributions, also for independent groups of inputs		
Notes	Groups of inputs may be considered Iterative enlargement of sample size not available		
Caveats	Variance as measure of uncertainty Quadratic decay in the coefficients only for continuous models		
References	[59]		

2.3.4. EASI / COSI

Noting that the permutation needed for RBD can be constructed from an available sample, a given data technique was developed in [60] which constructs the permutation needed. Again, undoing the permutation associated with one input factor on the output, one can identify via a FFT analysis the output variance in resonance with this factor. This gives estimates for first order effects. As one deals with real signals, the FFT can be replaced with a discrete cosine transformation (DCT). Then the permutation is obtained by sorting the output using the input of interest as a key.

Sensitivity Method	Effective Algorithm for Sensitivity Indices, Cosine Sensitivity for Sensitivity Indices		
Acronym	EASI, COSI	**Symbol**	S_i
Scope	FP: Functional Dependence; SD: Additivity		
Type	Variance-based		
Status	Emerging		

Given Data	Yes	Evaluation Costs	FFT, DCT
Input Probability	No assumptions		
Notes	Higher order effects by sorting along search curves in higher dimensions Under input dependence still interpretable as a goodness-of-fit-measure for a functional dependence SALib (Python) lists EASI under RBD_FAST		
Caveats	Variance used as measure of uncertainty Quadratic decay in the coefficients only for continuous models		
References	[60, 61]		

2.3.5. Use of surrogates or metamodels to calculate Sobol' indices

A surrogate model or metamodel is a simplified model which is easily evaluated. Hence it takes the idea of a response surface a step further. One can substitute the metamodel in place of the computationally costly original model. Some techniques allow one to extract sensitivity information directly from the representation of the metamodel without further evaluations. Metamodels need to be trained, tested, and validated. The results obtained using metamodels are "conditional" to the training data, the training therefore becomes an important factor for the quality of metamodels. Some techniques allow one to train the metamodel in parallel with the build-up of the sample from the input/output realizations of the simulation model.

The function $\mathbb{V}[\mathbb{E}[Y|X_\alpha]]$ can be interpreted as the component of the output variance that can be explained by a functional dependence on X_α, or, using the variance decomposition formula, $\mathbb{V}[\mathbb{E}[Y|X_\alpha]] = \mathbb{V}[Y] - \mathbb{E}\big[\mathbb{V}[Y|X_\alpha]\big]$, as the average decrease in output variance when one receives perfect information on X_α. The relative term, when dividing by $\mathbb{V}[Y]$, is therefore a goodness-of-fit measure/coefficient of determination of the nonlinear regression curve/response surface $\varphi_\alpha(x_\alpha) = \mathbb{E}[Y|X_\alpha = x_\alpha]$. Thus, one can obtain variance-based sensitivity index estimates from any regression surface / surrogate / metamodeling technique.

2.3.5.1. PCE

Polynomial chaos expansion (PCE) methods approximate the functional dependence of the simulation response on uncertain model parameters by representing it as a linear combination of terms from an orthogonal polynomial basis. If one chooses univariate powers as a starting point for the regression model, then an orthogonalization with respect to a uniform input distribution yields shifted Legendre polynomials as an orthonormal basis. For interactions, all product terms should be included (the design matrix is then composed of the tensor product of the feature maps), however this might quickly exhaust the available sample size. Some sort of pruning is needed. High Dimensional Model Representation (HDMR) only considers interactions of order up to two, while other PCE methods use information criteria for truncating the number of important feature combinations.

Sensitivity Method	High Dimensional Model Representation, Polynomial Chaos Expansion		
Acronym	(RS)-HDMR, (BS)-PCE	**Symbol**	S_i, S_α, T_i,
Scope	FP: Functional Dependence; SD: Additivity, Interactions; FF via approx. total effects		
Type	Variance-based		
Status	Established, Early Adopter for BSPCE		
Given Data	Yes	Evaluation Costs	Multiple Least Squares
Input Probability	No assumptions		
Notes	Total effects may use $d-1$ dimensional regression model		
Caveats	Variance as measure of uncertainty Under input dependence, orthogonality in the functional basis may be lost Total effects are reported for a truncated basis (hence may miss high-order contributions)		
References	[45, 62, 63]		

2.3.5.2. Gaussian Processes

Gaussian processes are popular metamodels for computational simulations due to their ability to model complicated functional forms and to provide an uncertainty estimate of their predicted response value at a new input point. The core idea is based on geostatistical kriging models, where the responses at input points that are close should be correlated [64]. A Gaussian process (GP) is a stochastic process for which any finite set of *Y*-variables has a joint multivariate Gaussian distribution; a GP is fully specified by its mean and covariance function. Gaussian process regression not only provides conditional means, but also error bounds on the regression curves, hence enabling one to obtain confidence intervals for the sensitivity estimates.

Sensitivity Method	Metamodeling via Gaussian Processes / Kriging		
Acronym	GP	**Symbol**	
Scope	Metamodel		
Type			
Status	Established		
Given Data	Yes	Evaluation Costs	Inversion of Kernel Matrices

Input Probability	
Notes	Any non-given data method may be supplied with GP metamodel Variance-based sensitivity indicators can be computed efficiently Error bounds
Caveats	Additional to sampling error, a metamodeling error is encountered Choice of variogram / covariance / kernel function
References	[64–66]

2.3.6. Nearest Neighbors Methods

We can combine the ideas behind COSI (sorting the data) and pick'n'freeze design by noting that a pick'n'freeze design can be approximated by choosing the nearest neighbor (with respect to different input dimensions and combinations) from an available sample. For first order effects, this amounts to computing $\hat{s}_i = 1 - \frac{\sum_{j=1}^{N-1}\left(y_{\pi(j)} - y_{\pi(j+1)}\right)^2}{2\sum_{j=1}^{N}\left(y_j - \bar{y}\right)^2}$ where π denotes the permutation that sorts the input of interest. For higher order effects, one may obtain information on the neighborhood of a realization using the set of p-th nearest neighbors (of minimal distance to the realization under inspection), Euclidean minimal spanning trees connecting nearest neighbors — trees have one root and each node except the root has a parent node —, search curves approximating space filling curves or traveling salesperson solutions where each node (except start and end) of the curve has one previous and one next element (all distances are taken restricted to the input dimensions of interest). Then one may order the realizations with the help of these methods in higher dimensions and use a formula like the above one, using the squared difference of the outputs of between the current and the previous or next element.

Sensitivity Method	Nearest Neighbor Methods		
Acronym		**Symbol**	S_i, S_α, T_i,
Scope	FP, FF: Functional Dependence; SD: Additivity, Interactions		
Type	Variance-based		
Status	Experimental		
Given Data	Yes	Evaluation Costs	Sort/Nearest Neighbors
Input Probability			
Notes	Needs relatively large sample size		
Caveats	Does not work with QMC		
References	[67]		

2.3.7. Methods for Identifying Interactions

Note that the most common usage (current state-of-the-art) is to report main and total effects indices, but not higher order indices. If one has access to the higher order Sobol' indices, then candidates responsible for interactions can be identified. However, as theoretically expected, orthogonality in the tensor products of the polynomial chaos expansion is not necessarily respected in sample space, the results are error-prone. [68] suggest using first order effects of higher order conditional moments. This requires the lower dimensional regression curves to be estimated. [69] instead suggest removing the direct functional effects by passing to the second order finite difference of the output, analyzing $2y_{\pi(i)} - y_{\pi(i-1)} - y_{\pi(i+1)}$ instead of y where π denotes the permutation that orders the input of interest. Feeding these transformed outputs into a first-order variance sensitivity estimator gives therefore hints on which input factors participate in interactions in the simulation models, but one has no information on the composition of the groups of interacting factors. Indeed, a nearest neighbor estimator for the residual variance is provided by:

$$\frac{1}{6(N-2)}\sum_{i=2}^{N-1}\left(\left(y_{\pi(i-1)} - y_{\pi(i)}\right) + \left(y_{\pi(i+1)} - y_{\pi(i)}\right)\right)^2$$

Hence the analysis of the first-order effect of the second-order finite difference offers an indication of whether or not the conditional variance changes. This effect, if present, is also called heteroskedasticity. Under input independence, a change of the conditional variance is only possible when there are active interactions.

2.4. Moment-Independent Sensitivity Measures

What if variance is not an appropriate measure of uncertainty? One idea might be to extend the analysis to higher moments (skewness, kurtosis). However, then we face the problem of a multi-attribute decision problem, taking into account all these different sources of information. To avoid this, one can revert to functional distances. One can compute a distance between the joint input/output density and the product of the marginal densities, $d(f_i f_Y, f_{X_i,Y})$, replacing the difference of expectations, $\mathbb{E}[YX_i] - \mathbb{E}[Y]\mathbb{E}[X_i]$, in correlation analysis by an abstract notion of a distance between density functions.

Instead of the two dimensional distance, one may compute the expected distance between the output density conditional to the input X_i and the unconditional output density, $\mathbb{E}\left[d(f_Y, f_{Y|X_i})\right]$. Here, the inner distance measure (also called separation) is a random quantity in X_i. Therefore, the expected distance yields a sensitivity measure quantifying the distance to independence. Computation requires (kernel) density estimators of the unconditional and the conditional output density. Instead of considering distances between densities, distances between cumulative distribution functions or between characteristic functions are also under discussion.

2.4.1. Density-Based: Borgonovo's δ

Using the integrated absolute value as a distance, one has

$$\delta_i = \frac{1}{2}\iint \left|f_Y(y) - f_{Y|X_i=x_i}(y)\right| f_i(x_i)dx_i dy = \frac{1}{2}\iint \left|f_i(x_i)f_Y(y) - f_{X_i,Y}(x_i,y)\right| dx_i dy.$$

For this choice of a distance measure, the two formulations as expected conditional distance and

as the distance between the product of marginals and the joint density coincide. Moreover, this measure is invariant with respect to monotonic transformations of the inputs and outputs, hence it is a nonparametric technique.

Sensitivity Method	Borgonovo Moment Independent Importance Measure		
Acronym	MIM	**Symbol**	δ_i
Scope	FF: Functional Dependence; SD: Interactions; RI: via separations		
Type	Density based		
Status	Early Adopter		
Given Data	Yes	Evaluation Costs	Kernel Density
Input Probability			
Notes	Distance to stochastic independence Nonparametric technique Kuiper measure (range of CDF differences) may serve as proxy Improved properties with QMC sample		
Caveats	Conditional kernel density requires large sample size Large numerical noise		
References	[70, 74]		

2.4.2. Distribution-Based: Pianosi and Wagener

Using the Kolmogorov/Smirnov distance between unconditional and conditional cumulative distribution function as a separation measure, $\mathrm{KS}_i(x_i) = \max_y \left| F_Y(y) - F_{Y|X_i=x_i}(y) \right|$, one can eliminate the dependence on x_i by considering a statistic over these distance values. [71] suggests mean, median or max. [72, 73] remark that only the choice of using the mean leads to a sensitivity measure that allows interpretation as a value of information. In this case, given data estimators are available [72, 73].

Sensitivity Method	Pianosi and Wagener Importance Measure		
Acronym	PAWN	**Symbol**	
Scope	FF: Functional Dependence; SD: Interactions; RI: via separations		
Type	Distribution based		
Status	Early Adopter		
Given Data	Yes	Evaluation Costs	
Input Probability	Independent input distributions		

Notes	Given data estimators available Choice of the outer statistic and the number of conditioning intervals must be communicated when using PAWN.
Caveats	Large numerical noise due to outliers being picked up by the maximum in KS. Reference implementation is cost-intensive since it uses double-loop design. Different implementations use different approaches of splitting the variability range into subintervals. Splitting by parameter values could lead (e.g. in the cases with lognormal distributions) to the occurrence of the underpopulated subsamples, and correspondingly questionable sensitivity analysis results.
References	[71 – 74]

2.4.3. Distribution-Based: Flavor Measure

A numerically benign variant that works on CDF distances is currently attracting attention and is given by $\gamma_i = 6 \iint \left(F_Y(y) - F_{Y|X_i=x_i}(y)\right)^2 f_i(x_i) f_y(y) dx_i dy$. This is a transformation-invariant version of a Gini mean distance, a Cramér/von Mises distance or an energy distance, replacing the absolute maximum in the Kolmogorov/Smirnov measure by an integrated squared distance. [75] show that a given-data estimator of this sensitivity measure can be obtained as an average of quantile sensitivity measures for which on can employ a first-order variance estimation technique, replacing the output $y \in \mathrm{R}^N$ with the indicator function $\mathrm{I}\{y \leq y_{(j)}\}$ coding the output quantile $\frac{j}{N}$ where $y_{(j)}$ is the order statistic of y. With this technique, one circumvents the use of kernel density estimators. Moreover, with this quantile indicator function, there are strong links to both variance-based sensitivity and reliability methods. As it is transformation-invariant, it is a nonparametric technique.

2.5. Graphical Analysis

2.5.1. Scatterplots

Scatterplots allow for the presentation of univariate dependences. Here, the conditionalization used in the previously discussed methods is available as localization: The conditional mean is then a curve through the backbone of the point cloud (and a deviation from constancy indicates a dependence). The conditional variance measures the spread of the output realizations at a given point for the input of interest (and a deviation from a constant tube about the conditional mean signals heteroscedasticity that is a sign of interactions under input independence). A cut though the scatterplot along a vertical line gives rise to conditional densities of the output. If these differ from the unconditional output density, then this is a sign of pairwise interaction between the input of interest and the output.

Sensitivity Method	Scatterplot, Point Cloud		
Acronym		**Estimator**	
Scope	FP: Linear and Functional Dependence; SD: Interactions; RI: Regional Effects; TI		
Type	Visualization technique		
Status	Established		
Given Data	Yes	Evaluation Costs	Space for Presentation
Input Probability			
Caveats	Human eye tends to over-interpret structures		
References			

2.5.2. CUSUNORO Curves

One problem with scatterplots is the need to consider all inputs separately which can become tiresome for simulation models that feature high input dimensions. Hence, one tries to condense the information of one scatterplot into a curve, so that many scatterplots map into one graph with many curves. For this, [76] suggests a method that considers the conditional mean below this input value instead of conditional output mean at a given input value. Monitoring the change as this value passes through all available realizations then gives the CUSUNORO curve. Deviation from the zero-line signals importance (more precisely the mean squared gradient of this curve relates to first order effects), linear segments are related to flat conditional means, and sharp bends show discontinuities in the data.

Sensitivity Method	Cumulative Sum of Normalized Reordered Output		
Acronym	CUSUNORO	**Estimator**	
Scope	FP: Functional Dependence; TI: Curve Orientation; RI: Curve curvature		
Type	Variance-based		
Status	Early Adopter		
Given Data	Yes	Evaluation Costs	
Input Probability			
Notes	Variants for use with time series Significance ellipsoid gives hint on importance First order effects estimation replaces regression by interpolation		
Caveats	Variability of conditional mean, higher moments require tricks		
References	[76]		

2.5.3. Copula Distance, Rank CUSUNORO

Two common themes in sensitivity measures are the comparison of the joint distribution with the product of marginal and the transformation invariance. Combining both ideas one obtains the copula distance $[0,1] \to \left[-\frac{1}{2}, \frac{1}{2}\right]$, $u \mapsto u \cdot v - C(u,v)$ where $u = F_i(x_i)$, $v = \hat{F}_Y(y)$ are the (empirical marginal) uniform transformations, and $C(u,v) = F_{X_i,Y}\left(F_i^{-1}(u), \hat{F}_Y^{-1}(v)\right)$ is the empirical bivariate copula between X_i and Y. This copula is compared to the product copula that describes independence. Hence, any deviance of the copula-distance point-cloud from the zero line denotes dependence in the data. Plotting the Cusunoro curve with respect to the ranks forms a regression curve through these points [77].

2.5.4. Power Weighted Reordered Output

A way to include local information from the input into the output is to consider weighted means where the weights depend on the input locations. An idea related to extreme quantiles is to reshape the marginal input distribution $F_i(x_i)$ so that extreme values are obtaining more weights, considering maximum and minimum distributions of k independent copies of X_i have cdfs given by $F_i(x_i)^k$ and $1 - \left(1 - F_i(x_i)\right)^k$, respectively. Hence forming a weighted output mean with $\frac{1}{N}$ replaced by $\left(\frac{j}{N}\right)^k - \left(\frac{j-1}{N}\right)^k$ or $\left(\frac{N-j+1}{N}\right)^k - \left(\frac{N-j}{N}\right)^k$ for the j^{th} ranked input, respectively, gives information on the localization of information. We can map the interval $(0,1)$ continuously into these weights, so that one can produce a continuous set of curves from putting all weight on the minimum over the standard mean with equidistributed weights to all weight being put on the maximum.

2.6. Derivative-Based Global Sensitivity Measures

The gradient of a function carries information about the local behavior. Averaging over the gradient therefore gives a measure of global sensitivity. If the simulation is given by differential equations, then this derivative information is readily available. Otherwise, techniques of automatic differentiation might be used. There are theoretical links to the variance-based total effects. For more information, consider [78]. Especially, screening may benefit from using derivative-based sensitivity measures.

2.7. Global Sensitivity Methods and Settings

We have qualified the applicability of different sensitivity methods into settings. For an overview, we collect this information in a matrix format shown in Table 2-2.

	Factor Prioritization	Factor Fixing	Trend Identification	Structural Discovery	Regionalized Information
Pearson, Spearman, Standardized Regression Coefficients	☑		☑		
Sobol' Method, EFAST	☑	☑		☑	
RBD, EASI, COSI	☑			☑	
Moment Independent Methods	☑	☑		☑	☑
CUSUNORO Curves	☑		☑	☑	☑
Scatterplots	☑		☑		☑

Table 2-2. Purposes of SA Methods

2.8. Software Availability

Sensitivity analysis has established itself as a part of the scientific modeling and simulation workflow. While the software coverage for methods based on linear regression is generally good, variance-based and moment-independent sensitivity measures are not generally found in standard software products. Different data-analytic programming languages offer toolboxes and script collections. The coverage of methods differs; some packages are updated more often than others. There are a few ones which are mainly dedicated to uncertainty quantification.

In Table 2-3 below, we present software tools and a link to a website where appropriate. We note that some of these tools provide a set of functions or a library from which various elements can be called, while others are more self-encapsulated, standalone programs which perform an entire workflow from the sampling to surrogate modeling to sensitivity index calculations. There are a

few commercial products listed but most of these are publicly available. Note that this is not meant to be a comprehensive list but to show that there are a significant number of options for people wanting to use sensitivity analysis methods.

Toolbox/Package	Language	Website
UQLab	MatLab / Octave	www.uqlab.com
SAFE (Sensitivity Analysis for Everybody)	Matlab. Also R, Python	www.safetoolbox.info
GUI-HDMR	Matlab	www.gui-hdmr.de
Sensitivity	R	cran.r-project.org/web/packages/sensitivity/index.html
openTURNS	Python	openturns.github.io/www/
SALib	Python	salib.readthedocs.io/en/latest/
mads. Model Analysis and Decision Support	Julia/C	mads.lanl.gov
EcoLego	Set of toolboxes	www.ecolego.se
SobolGSA	C#	www.imperial.ac.uk/process-systems-engineering/research/free-software/sobolgsa-software/
SimLab	C++/R	ec.europa.eu/jrc/en/samo/simlab NOTE: SimLab is currently not available publicly but it was used by GRS in their analyses.
Dakota		dakota.sandia.gov
MUQ: MIT Uncertainty Quantification Library		http://muq.mit.edu/ More UQ focused, Bayesian calibration
NESSUS		www.swri.org/nessus More UQ, Reliability and failure estimation focused.
UQTk	C++/Python	http://www.sandia.gov/UQToolkit/
PSUADE		github.com/LLNL/psuade computing.llnl.gov/projects/psuade-uncertainty-quantification
SmartUQ		www.smartuq.com/software/sensitivity-analysis/

Table 2-3. UQ Software tools and packages available

This page is intentionally left blank

3. CALCULATION CASE SELECTION

When conceptualizing the sensitivity analysis exercise, an inventory of calculation cases which might be appropriate to address the aims of the exercise (see Section 1.1) was compiled. A prerequisite for including a case in the exercise was that the "case owner" was in a position to provide data files containing input parameter and output values for a reasonable number of realizations in order to allow the exercise participants to test their own SA methods and tools using these "existing data" without the need to perform model runs. Furthermore, it was important to create a set of cases covering a variety of features the SA methods and tools can be tested against. To this end, a questionnaire was created addressing these features. For each candidate case, information on the following issues was to be provided:

Phenomenological description. This includes a description of the system being modeled such as a repository, the waste forms and engineered barriers, the geology, physical and chemical processes being modeled, potential release pathways, etc.

Input characteristics. The input characteristics refer primarily to various aspects of the computational simulations, including the number of parameters being varied, the sampling methods used, and the number of runs that were generated. The input characteristics also include details about how uncertain parameters are described in the case study: are they continuous, categorical, or discrete inputs? Are there any inputs which vary in space and/or time? What probability distributions were used to represent and model the uncertain parameters? Were any uncertain parameters modeled with approaches other than probability distributions (e.g. with intervals, fuzzy sets, etc.)? Were there dependencies amongst the inputs and how were they treated (for example, with a joint probability distribution)? Finally, if a distinction between aleatory and epistemic uncertainty was made in the case study, it was requested that the case owners provide a listing of the epistemic vs. aleatory parameters and describe how the difference was handled in the case study.

Output characteristics: The output characteristics focus mainly on the quantities of interest, including a description of these quantities and details about whether the output quantities are scalar values or values varying in space and/or time (e.g. vector valued outputs).

Other aspects of the output that the case owners were asked to address include some analysis of the model output, such as known specific points concerning the model behavior which might be of importance for sensitivity analysis. Some examples include non-linearity, non-monotonicity, discontinuities, and regime changes at input thresholds. Listing parameter values for which simulation runs failed for physical or numerical reasons is also part of the output characterization. The case owner was also asked to suggest any issues such as possible bias or additional uncertainty introduced by a surrogate model or other issues the sensitivity analysts would find helpful. Finally, the case study owners were asked to provide some basic detail on the data they provided for the case study in terms of file organization and file formats.

Based on this information, seven cases were identified. It was agreed that each organization would address these models with their own methods and approaches. This report covers the first four cases.

This page is intentionally left blank

4. GRS CLAY CASE

4.1. Case Description

This model system, which was already used as a test model for sensitivity analysis in former projects [79, 80], describes the release of radionuclides from a generic repository for spent nuclear fuel (SNF) and high-level vitrified waste from reprocessing (HLW) in a Northern German clay formation. It is based on considerations made in the context of the project GENESIS [81] and is described in detail in [79]. The repository is assumed to be located in the middle of the Apt layer in the Lower Cretaceous Clay in Lower Saxony, see Figure 4-1.

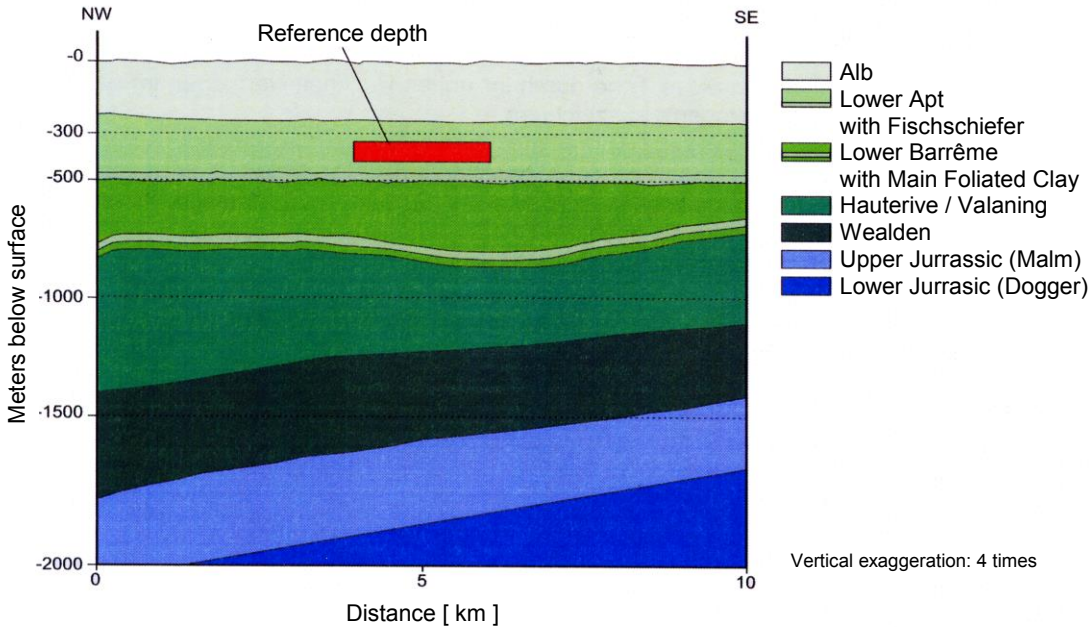

Figure 4-1. Cross-section of the model area in Northern Germany

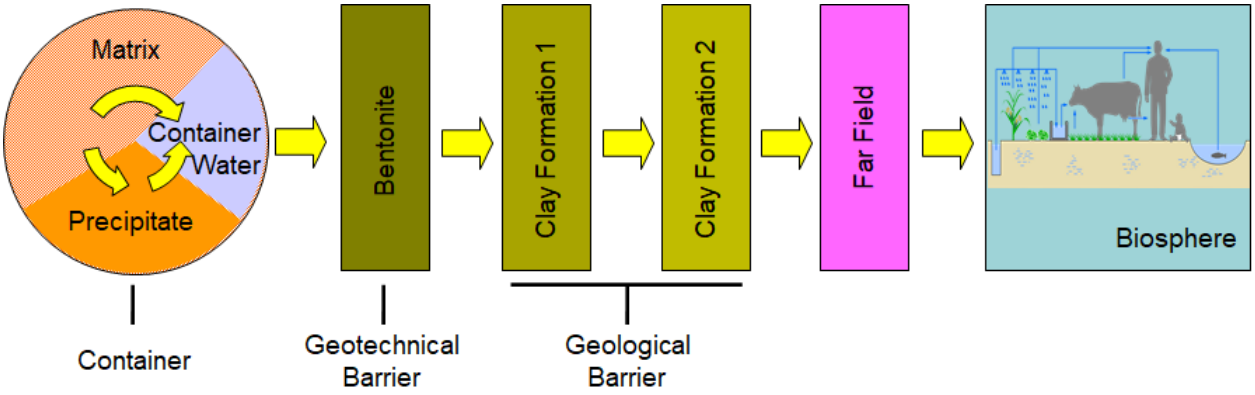

Figure 4-2. Schematic presentation of the model chain

The model comprises the near field with the waste containers, three clay layers (bentonite buffer, Apt and Alb), the far field and the biosphere, see Figure 4-2. The far field is modeled as a generic transport path of about 10 kilometers through a typical porous rock medium, including dilutive water flow. For biosphere modeling, standardized dose conversion factors are used, which take account of all relevant exposure paths for a population taking their water from a contaminated reservoir. The output is the annual effective dose to an adult human individual. As the parameters for far field and biosphere calculation were not varied in the probabilistic investigations, they are of no specific interest here.

The radionuclide mobilization and the (purely diffusive) transport through the clay layers were calculated with the near field code CLAYPOS (version 3.01, [79]). The far field represents a 1-D transport path through a generic porous medium and was modelled with the code module CHETLIN (version 4.1). For the biosphere (i. e. calculation of radiological consequences to humans from radionuclide flows) the code module EXMAS (version 2.1) was used. The simulation time was 10^8 years. Though it is unrealistic to assume that a real system could remain stable over such a long time frame, it was nevertheless decided to consider the total period, as the system's isolation capabilities turned out to be so effective that nothing interesting happens during the "normal" time frame of 1 million years. 100 points in time were evaluated.

Six near field parameters were selected for investigation. These are listed in Table 4-2.

Reference case name	**Description**
Waste form	Spent nuclear fuel in specific spent fuel element containers
Engineered barriers	Steel container, bentonite buffer
Repository description	1070 boreholes with 5 containers each, reference depth 400 m
Natural system	Lower Cretaceous Clays, formations Apt and Alb
Far Field	Generic 10-km transport path through a typical rock medium
Biosphere	Exposure pathways considered via pre-defined dose conversion factors, groundwater flow10^{-5} m³/yr
Conceptual release pathways	Purely diffusive transport through bentonite, Apt and Alp, stylized far field
Processes modeled	Degradation of waste matrix, radionuclide dissolution, solubility in container water, diffusive transport and retention in buffer and clay layers, 1-D advective-diffusive transport through far field
Software codes used	CLAYPOS3, CHETLIN4, EXMAS2
Reference for full description of case	Rübel, A. et al.: Development of Performance Assessment Methodologies. GRS-259, Braunschweig 2010

Table 4-1. GRS Clay Case Description

4.2. Description of inputs and outputs

The model has six input parameters, for which the names and distributions given in Table 4-2 were used. Samples of 4096 and 8192 randomly drawn sets of parameter values were provided for the investigations. All variables were treated independently and in the same way, no correlation was assumed.

Input parameter	Type	Range
DiffClay2: Diffusion constant clay formation 1	Log-uniform	8.3E-12 - 8.3E-10
DiffClay3: Diffusion constant clay formation 2	Log-uniform	8.3E-12 - 8.3E-10
KdBent: K_d value bentonite (U)	Log-uniform	4 - 400
KdClay2: K_d value clay formation 1 (U)	Log-uniform	2 - 200
KdClay3: K_d value clay formation 2 (U)	Log-uniform	2 - 200
PorClay: Porosity clay formation 2	Uniform	0.06 - 0.24

Table 4-2. GRS Clay Case Input Parameters

The same ranges are valid for both clay layers, but the actual values for each run are drawn independently. The element-specific diffusion coefficients in both clay formations as well as the porosity in formation 2 are valid for all elements except Cl, Se, Mo and I. For these four elements a fifth of the drawn values for diffusion coefficients and half of the value for the porosity were used due to anion exclusion. The given ranges for K_d values are valid for uranium; those of the other elements are coupled to these by formation-specific factors. H, C, Cl, Se, Rb and Pd have K_d equal to zero.

The investigated model output, as listed in Table 4-3, is the annual effective dose rate to an adult human individual, calculated using pre-defined biosphere dose conversion factors that take into account all relevant exposure paths.

Output Quantity	Type	Other
Annual effective dose rate to adult individual	time-dependent 194 points in time between 1.000E0 and 9.955E7 years	

Table 4-3. GRS Clay Case Output QoIs

The model does not include any problematic properties and is not expected to exhibit significant nonlinearities or discontinuities with regard to variation of the input parameters. The model was run many thousands of times, based on parameter samples of different sizes and drawn with different techniques. For the joint sensitivity analysis exercise the results obtained from random samples with sizes of 4096 and 8192 as well as an EFAST sample of size 8214 were provided.

4.3. Salient features and behavior of the model

The scatterplots for the six parameters of the clay system are presented in Figure 4-3 for a model time point 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.

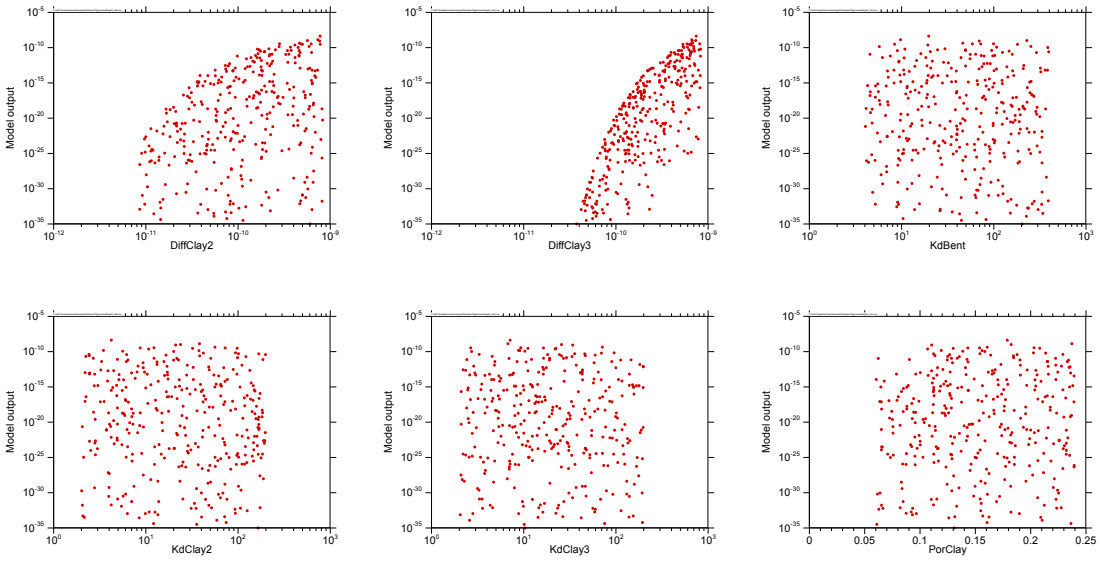

Figure 4-3. Scatterplots for the clay system, 10^6 years

The equivalent scatterplots for the system at the end of the calculated scenario are presented in Figure 4-4. 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 on the model output.

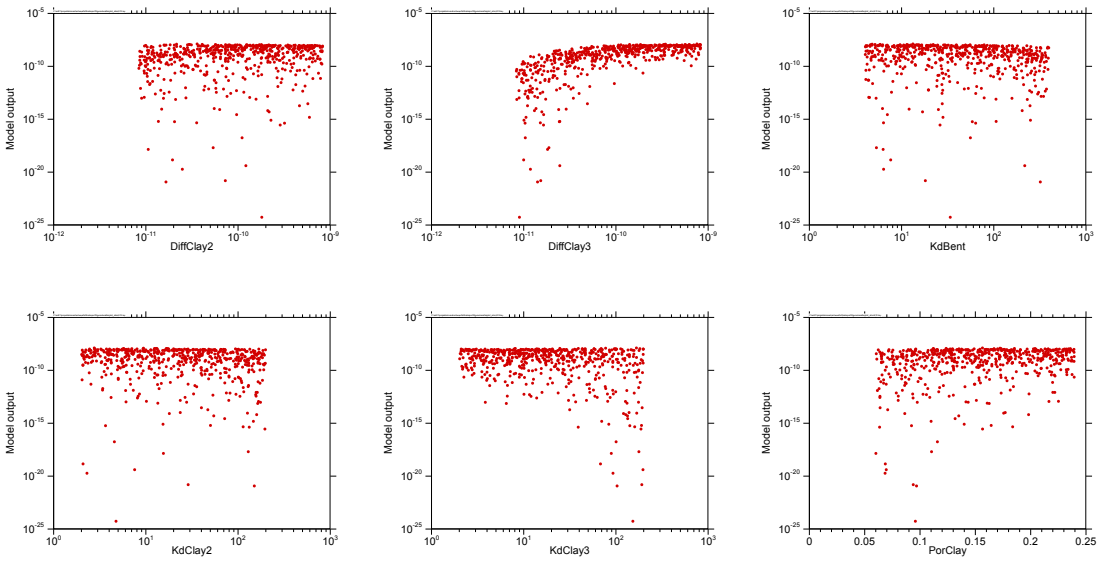

Figure 4-4. Scatterplots for the clay system, 10^8 years

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.

Figure 4-5 shows how the maximum, the mean and the median of the model output develop over time. The figure was made on the basis of 2048 random runs.

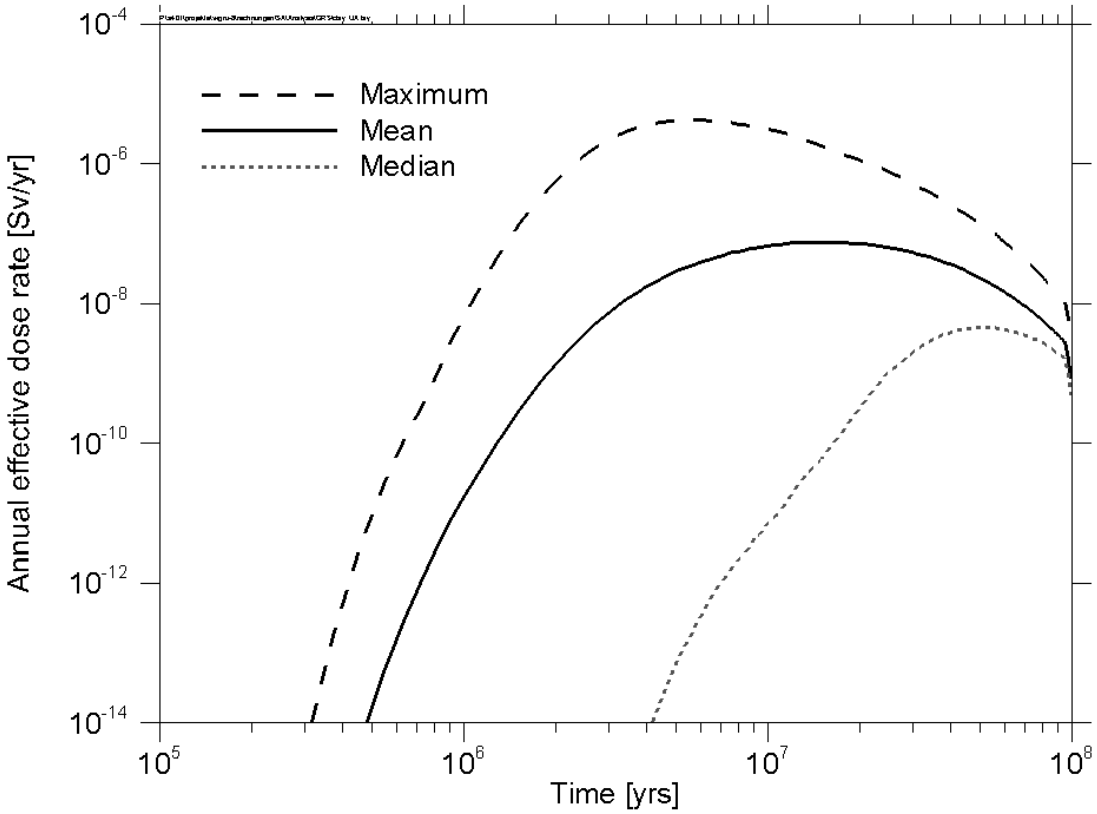

Figure 4-5. Time-development of maximum, mean and median of the clay system

4.4. Sensitivity Analysis Results

4.4.1. Results from GRS

GRS performed a lot of investigations with this model [80]. For the purpose of the report at hand, some basic investigations are presented as described in Table 4-4.

GRS Clay	**Description**
Sensitivity Analysis Method	Scatterplots, CUSUNORO, Linear correlation, linear regression, rank correlation, rank regression, EFAST, EASI, RS-HDMR [83, 84]
Sensitivity measures generated	Pearson CC, Spearman RCC, SRC, SRRC, S_i and T_i
Special considerations	(None)
Surrogate models used	Surrogate model included in RS-HDMR software
Transformations	(None)
QoIs addressed	Annual dose rate to an adult human individual
Number of samples used	Random 4096, Random 8192, EFAST 8214

Table 4-4. Sensitivity analysis of GRS clay model by GRS

Table 4-5 and Table 4-6 present the calculated sensitivity measures for two different points in time, namely 1 million years and 100 million years. For each sensitivity measure the three leading parameters are marked in red, green and blue. All indices agree about the parameter *DiffClay3* being the most important one at both points in time. On the second rank *DiffClay2* appears at 1 million years and *KdBent* at 100 million years. The further ranking depends on the method of analysis.

Input parameter	**Pearson CC**		**Spearman RCC**		**Regression SRC**		**Rank regression SRRC**	
	10^6 yr	10^8 yr	10^6 yr	10^8 yr	10^6 yr	10^8 yr	10^6 yr	10^8 yr
1. *DiffClay2*	0.206	0.111	0.270	0.195	0.204	0.116	0.287	0.208
2. *DiffClay3*	0.258	0.579	0.907	0.761	0.257	0.587	0.912	0.761
3. *KdBent*	-0.014	-0.265	0.002	-0.245	-0.013	-0.274	-0.001	-0.242
4. *KdClay2*	-0.038	-0.118	-0.012	-0.146	-0.026	-0.150	0.000	-0.121
5. *KdClay3*	-0.032	-0.131	-0.008	-0.222	-0.018	-0.146	-0.001	-0.217
6. *PorClay*	0.041	0.123	0.022	0.149	0.037	0.131	0.020	0.147

Table 4-5. Correlation-/regression-based SA of GRS Clay model by GRS (Random 4096)

Input parameter	S_i (EASI)		S_i (EFAST)		T_i (EFAST)		S_i (RS-HDMR)		T_i (RS-HDMR)	
	10^6 yr	10^8 yr	10^6 yr	10^8 yr	10^6 yr	10^8 yr	10^6 yr	10^8 yr	10^6 yr	10^8 yr
1. *DiffClay2*	0.047	0.050	0.078	0.063	0.829	0.193	0.052	0.056	0.587	0.109
2. *DiffClay3*	0.093	0.510	0.178	0.530	0.857	0.782	0.157	0.513	0.706	0.638
3. *KdBent*	0.004	0.083	0.000	0.095	0.040	0.213	0.001	0.088	0.015	0.153
4. *KdClay2*	0.004	0.025	0.000	0.020	0.030	0.066	0.001	0.019	0.016	0.040
5. *KdClay3*	0.004	0.026	0.000	0.024	0.042	0.098	0.001	0.021	0.012	0.060
6. *PorClay*	0.003	0.025	0.001	0.016	0.085	0.053	0.001	0.016	0.013	0.032

Table 4-6. Variance-based SA of GRS clay model by GRS (EASI: Random 4096, EFAST: 8214)

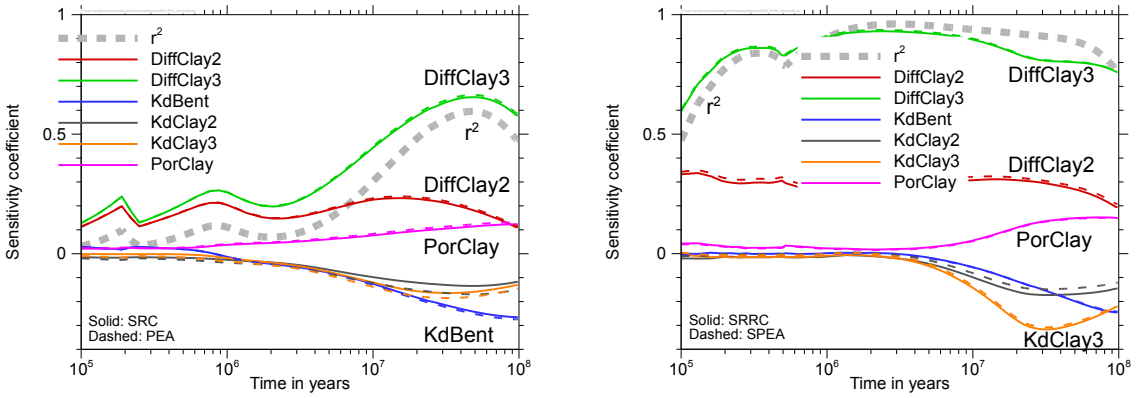

Figure 4-6. Regression- and correlation-based sensitivity analysis with direct (left) and rank-transformed (right) data for the clay model (Random 4096)

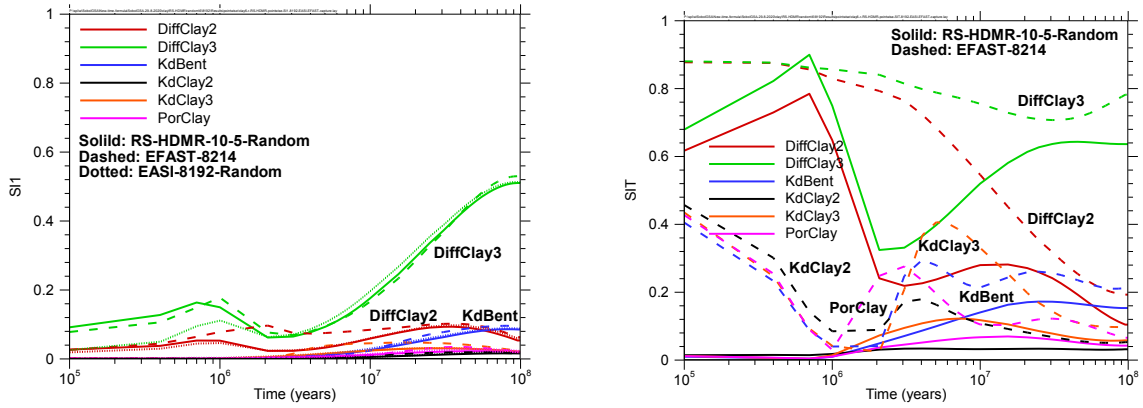

Figure 4-7. Variance-based sensitivity analysis for the clay model; first-order (left) and total-order (right) analysis (EASI and RS-HDMR: Random 8192, EFAST: 8214)

Figure 4-6 and Figure 4-7 show the time-development of the correlation-/regression-based sensitivity indices and the variance-based indices of first and total order. The methods agree satisfyingly about the sensitivities and show the same tendencies, although there are considerable

differences in detail at specific points in time. The total index curves obtained from the EFAST and RS-HDMR methods, however, are not in a good agreement, which shows that there are still shortcomings in the calculation schemes.

All applied sensitivity analysis methods agree qualitatively in that they find very low sensitivity of the model to all six input parameters in the time phase up to 1 million years. Only the diffusion coefficients in the clay layers seem to indicate a certain sensitivity, while the sensitivity indices for the other parameters are nearly zero. The variance-based first-order analysis does not recognize any significant sensitivity of parameters 4 to 6 in the "early" time frame. Interestingly however, the correlation or regression analysis detects a low systematic sensitivity for these parameters, if performed directly on the model output values, but it is not visible in their rank-based forms. This can be interpreted as a hint that there is a certain effect of these parameters only on the highest model output values, which is essentially mitigated by rank transformation.

In the very long time-frame, all methods detect an increasing sensitivity of the second parameter (*DiffClay3*). At the end of the model time, the first-order index of this parameter is calculated as ~0.5. Also, the other parameters, specifically parameter 3 (*KdBent*), gain more importance during the late time period.

At the time of 10^6 years, the sum of all S_i is far below 1, which means that there are considerable parameter interactions. The EFAST analysis shows that the total indices of parameters 1 and 2 are high, while the others are close to zero. From this, one can conclude that the model is dominated by interaction of these two parameters.

Physically, this can be understood, as the diffusion coefficients in both subsequent clay layers have to be high to produce a high radionuclide outflow at an "early" time. If in one of the layers there is low diffusion, this layer acts as a retarding barrier, considerably reducing the influence of the other one. At very late times, however, the system is dominated by parameter 2, which is the diffusion coefficient in the outer clay layer. The long time is in any case sufficient for the long-lived radionuclides to traverse the inner clay layer, so that the importance of the interaction with *DiffClay2* decreases considerably. The above-mentioned observation that at 10^6 years parameters 4 to 6 seem to have a small but detectable influence on the highest model output values can also be understood physically, as such values result from those rare situations that allow significant radionuclide release, and only these can be influenced at all by the non-dominating parameters.

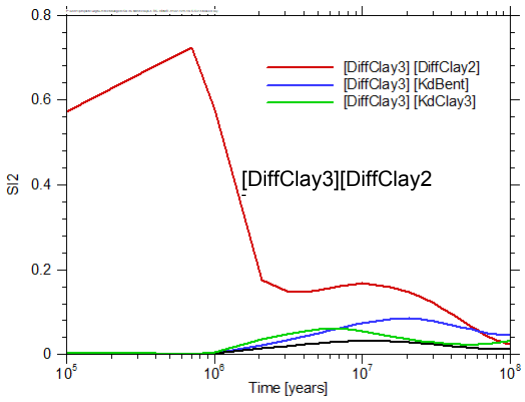

Figure 4-8. Variance-based sensitivity analysis for the clay model; second-order analysis (RS-HDMR: Random 8192)

Figure 4-8 shows the time-evolution of the 2nd-order sensitivity indices calculated with the RS-HDMR metamodeling method (to yield significant results this method needs careful adjustment of the polynomial orders taken into account, but this is a complex topic, which we do not discuss here, we used $k = 10$, $l = 5$). Obviously, the only interaction that plays a significant role is that between the diffusion coefficients in the two clay layers. This can be understood as it needs a high diffusion coefficient in layer 2 to allow the diffusion in layer 3 to become significant at all.

4.4.2. Results from SNL

GRS Clay	Description
Sensitivity Analysis Method	Surrogate Modeling, PEAR, SPEAR
Sensitivity measures generated	S_i, T_i, Pearson CC, Spearman RCC
Special considerations	
Surrogate models used	PCE (order 5), GP
Transformations	Rank
QoIs addressed	Annual dose rate to an adult human individual at 194 time points
Number of samples used	4096 (non-GP surrogate methods) or 1024 (GP)

Table 4-7. Sensitivity analysis of GRS Clay case by SNL

The sensitivity analysis performed by SNL on the GRS Clay case shown in Table 4-7 included calculation of the PEAR (Figure 4-9), SPEAR (Figure 4-10), PCC, and PRCC (Figure 4-11) sensitivity measures. A figure is not included for the PCC results because they are very similar to the PEAR results. Sobol' indices were also estimated using an order 5 PCE surrogate model fit to 4096 simulations or a GP model fit to 1024 simulations. The main Sobol' index estimates from the PCE surrogate model are plotted in Figure 4-12 and the total Sobol' index estimates from the PCE surrogate model are plotted in Figure 4-13.

At 10^6 years into the simulation, the sensitivity measures from all the sensitivity analysis methods applied by SNL are highest for *DiffClay3*. The independent variable with the second highest sensitivity measure is *DiffClay2*. None of the other independent variables appear to have a significant effect at 10^6 years. The methods identify *DiffClay3* and *KdBent* as the most important variables at the end of the simulation (10^8 years).

The difference between the S_i (Figure 4-12) and the T_i (Figure 4-13) measures indicates that there are significant interaction effects according to the PCE model. Estimates of dominant 2nd and 3rd order interaction effects are listed in Table 4-8. At 10^6 years, the dominant interaction is between the diffusion variables. The Sobol' index for this interaction is an order of magnitude higher than the other interaction indices. At 10^8 years, there is no dominant interaction, though interactions between the diffusion and K_d variables appear to be more important than interactions involving porosity.

The GP surrogate model construction does not work well with more than about 1000 samples, so the number of simulations used to fit this model were reduced to 1024 compared to the 4096 samples used for the other methods. Even with this restriction, it takes substantially longer to estimate sensitivity measures with the GP model than any of the other methods. The results from the GP model also did not offer additional insights beyond the interactions identified by the PCE model.

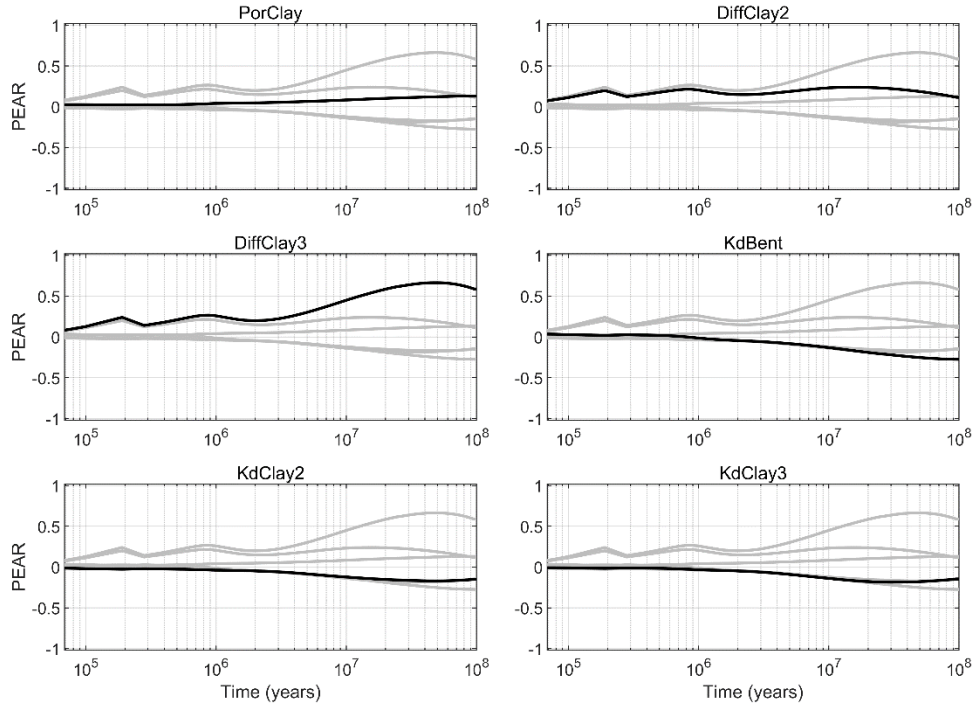

Figure 4-9. PEAR sensitivity results over time for the annual individual dose rate QoI from the SNL analysis of the GRS Clay case. Each axis in the plot highlights one of the independent variables

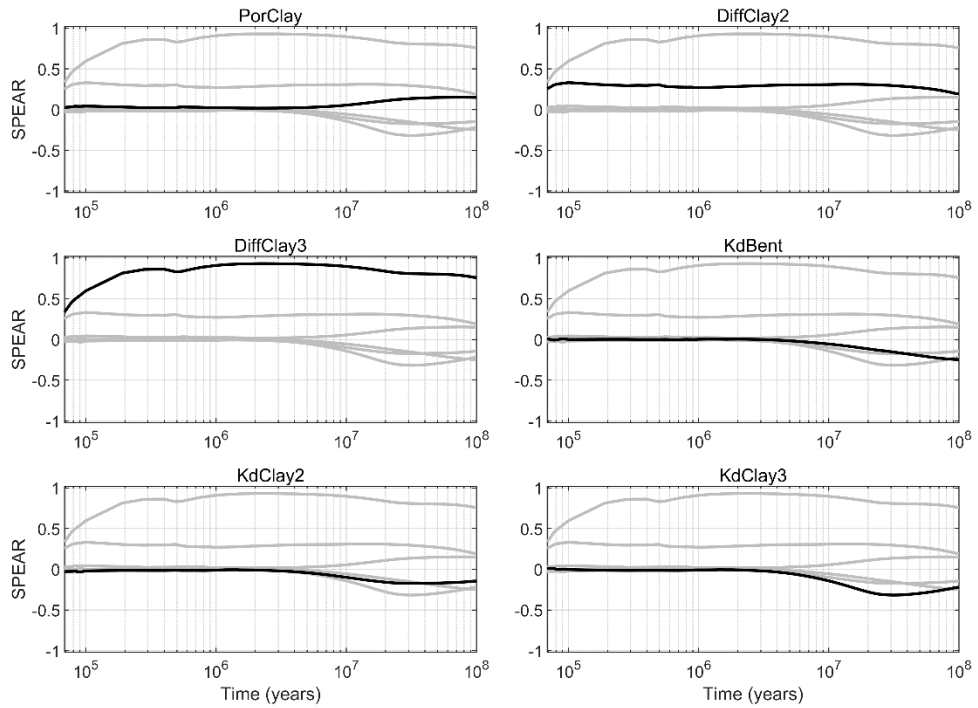

Figure 4-10. SPEAR sensitivity results over time for the annual individual dose rate QoI from the SNL analysis of the GRS Clay case. Each axis in the plot highlights one of the independent variables

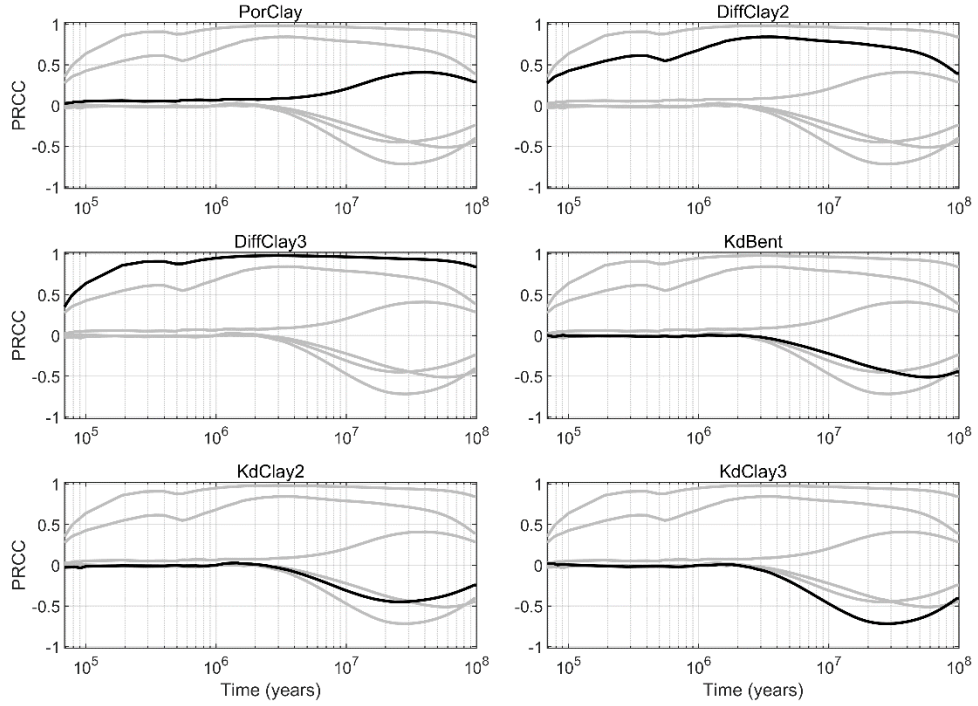

Figure 4-11. PRCC sensitivity results over time for the annual individual dose rate QoI from the SNL analysis of the GRS Clay case. Each axis in the plot highlights one of the independent variables

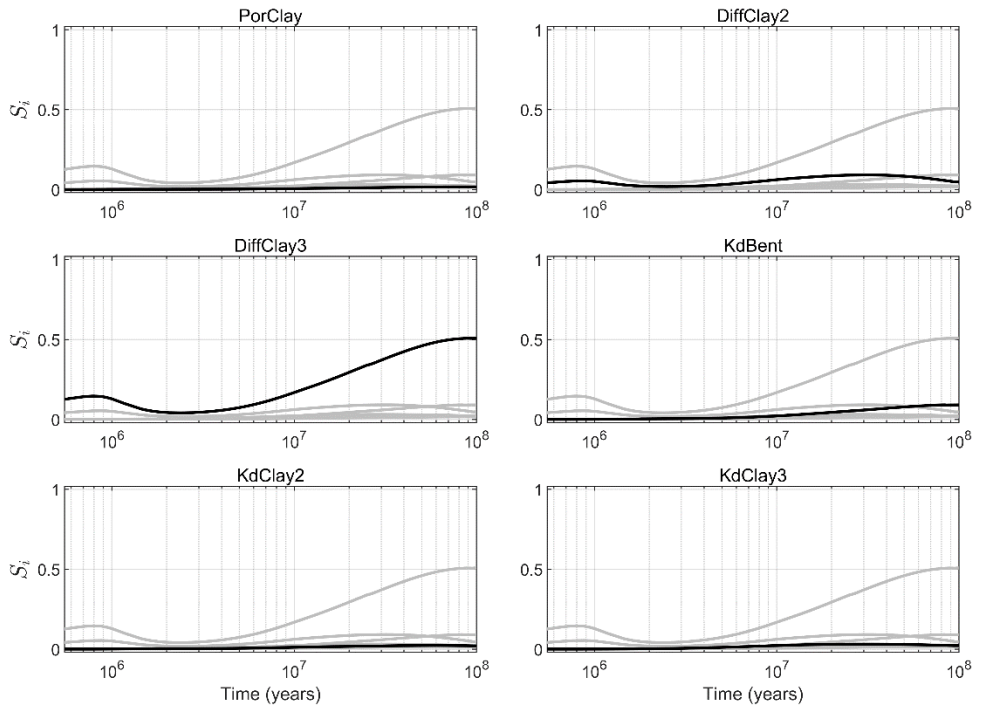

Figure 4-12. Main Sobol' index sensitivity results over time for the annual individual dose rate QoI estimated using an order 5 PCE surrogate model. Results are from the SNL analysis of the GRS Clay case. Each axis in the plot highlights one of the independent variables

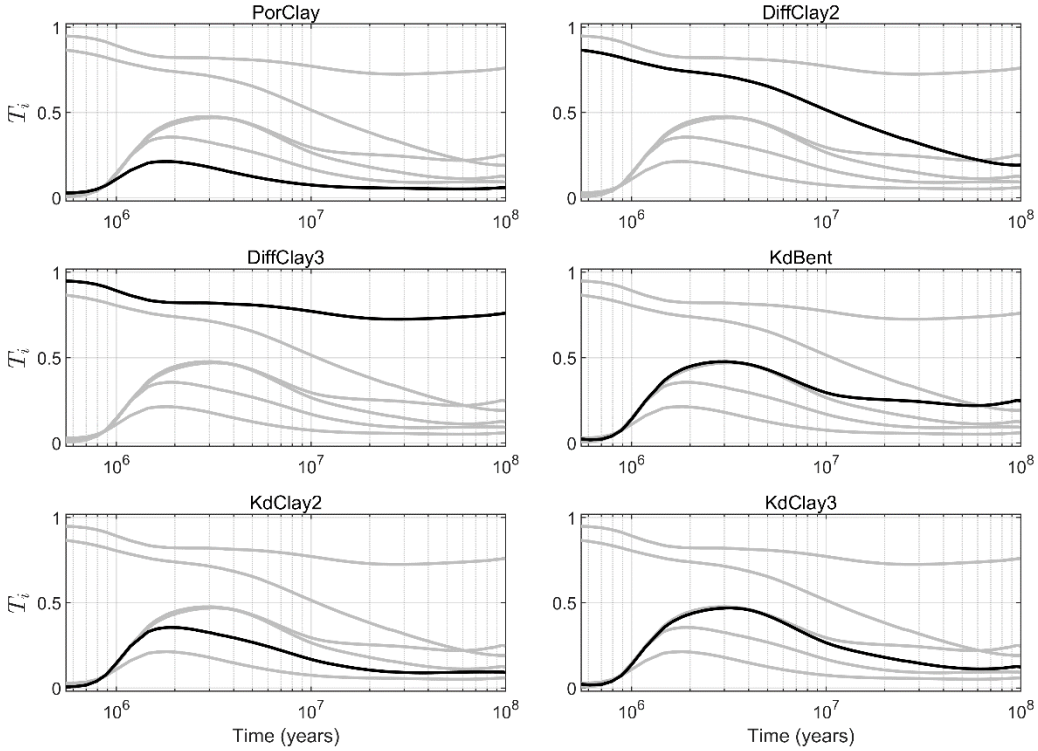

Figure 4-13. Total Sobol' index sensitivity results over time for the annual individual dose rate QoI estimated using an order 5 PCE surrogate model. Results are from the SNL analysis of the GRS Clay case

Interaction	S_i (PCE) 10^6 years	Interaction	S_i (PCE) 10^8 years
DiffClay2, *DiffClay3*	5.03E-01	*DiffClay3*, *KdBent*	4.31E-02
DiffClay2, *DiffClay3*, *KdClay2*	3.53E-02	*DiffClay3*, *KdClay3*	3.76E-02
DiffClay2, *DiffClay3*, *KdBent*	3.00E-02	*DiffClay2*, *DiffClay3*, *KdBent*	3.68E-02
PorClay, *DiffClay2*, *DiffClay3*	2.45E-02	*DiffClay2*, *DiffClay3*	1.90E-02
DiffClay2, *DiffClay3*, *KdClay3*	1.83E-02	*DiffClay3*, *KdBent*, *KdClay3*	1.45E-02

Table 4-8. Interaction terms from sensitivity analysis of GRS Clay case by SNL

4.4.3. Results from TUC

The sensitivity analysis methods used by TUC are shown in Table 4-9 with the results listed in Table 4-10.

Case Name	**Description**
Sensitivity Analysis Method	Linear regression, EASI, CUSUNORO, PCE, MIM
Sensitivity measures generated	SRC, S_i first order, S_{i2} second order, Borgonovo delta
Special considerations	Polynomial chaos uses harmonic base functions
Surrogate models used	
Transformations	Second order finite difference for interactions
QoIs addressed	SA for each time step
Number of samples used	4096 / all
Dataset	mc4096-6-siton-r*
Software	MatLab/Octave: in-house software

Table 4-9. Sensitivity analysis of GRS Clay by TU Clausthal

Input parameter	**CUSUNORO**	**SRC (at $t = 10^8$ yrs)**	**EASI (at $t = 10^8$ yrs)**	**LASI (Harmonic PCE)**
DiffClay2	Minor, nonlinear, +, effects at 33%	.11-	.05	Early interaction between 1 &2 (.20)
DiffClay3	Major, nonlinear, +, effects at 90% (early), 66% (late)	.58-	.52	Early interaction between 1 &2 (.20)
KdBent	Minor, linear, -	-.28+	.08	
KdClay2	Small, -	-.11+	.03	
KdClay3	Small, -	-.12+	.03	
PorClay	Small, +	.12+	.03	

Table 4-10. SA Results of GRS Clay by TU Clausthal

The Cusunoro curve [82] contains information from both linear analysis (local sign of changes and global monotonicity property), as well as nonlinear analysis ($\mathrm{SRC}^2 \neq S_i$) signals nonlinear contributions). All the methods agree on rankings.

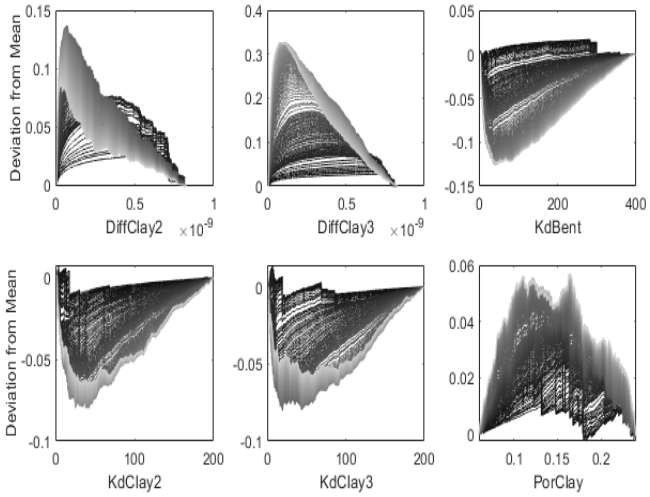

Figure 4-14. Time-Dependent CUSUNORO plots. Dark gray codes early, light gray late times.

Figure 4-14 demonstrates the use of CUSUNORO curves. Monotonicity properties are visible, detected by curves always above or below 0. Non-linear behavior is visible by sharp bends in the curves.

Moment-independent methods see *DiffClay3* as most important, followed by *DiffClay2* (up to $t = 10^7$ years). From 10^7 years, parameters 5 and 3 gain importance. As this increasing behavior stretches over the end of the simulation time, one needs to argue. As MIM do not differentiate between parameters reducing or amplifying output effects, we might see parameters beneficial to system performance.

For interactions, TUC analyzed the first order effect on second finite differences of the (reordered) output using the method sketched in Section 2.3.7. Here, heteroskedasticity (variable variance) is used as a proxy for interaction effects. Heteroskedasticity is a measure for the variability of the conditional variance, visible as inhomogeneity of the scatterplot. If for some fixed value of the parameter of interest the output values are widely spread, there is considerable influence of other parameters. If the scatter range changes over the parameter interval, this is a strong hint to the presence of interactions with other parameters. Figure 4-15 reports the rankings obtained from quantifying the sensitivity of the residual variance for each time-step. Lighter shades of gray indicate higher heteroskedasticity and with that a stronger suspicion of interactions. For the first 10 time-steps the ranking does not change from the default, most likely due to a zero output of the transport simulation. It also additionally suggests interactions involving parameters 1 and 2 (up to time $4 \cdot 10^7$ years) and late interactions involving parameters 2 and 3. However, it should be noted that this method suggests the presence of interactions, but not that these parameters actually interact.

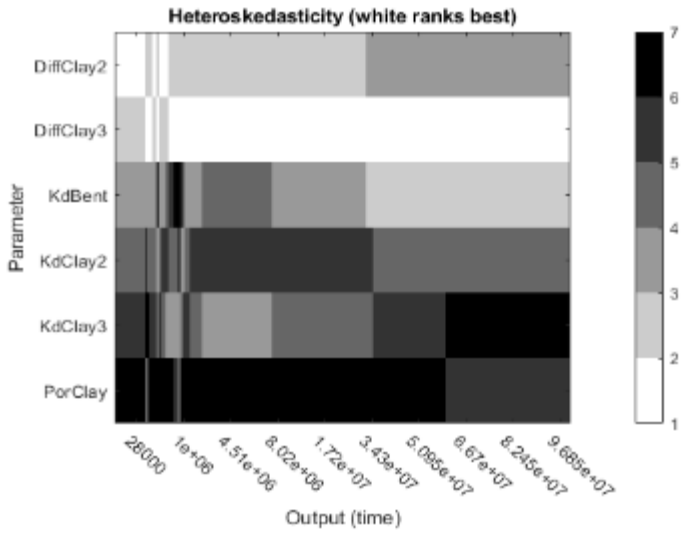

Figure 4-15. Heteroskedasticity: higher order effects may be due to interactions with parameters 2 and 5

From a linear analysis, the influence of parameter 6 is increasing at the end of the simulation time: Again, this needs careful arguments in commenting on possible extrapolation of the behavior.

4.4.4. Results from IBRAE

Case Name	**GRS Clay**
Sensitivity Analysis Method	PAWN, RBD-FAST, Correlation and regression
Sensitivity measures generated	PAWN index, S_i first order, Pearson correlation, Spearman rank correlation, partial correlation, partial rank correlation, regression and rank regression coefficients
Special considerations	Mean and median KS for PAWN. Number of conditioning intervals $n = 8$.
Surrogate models used	
Transformations	
QoIs addressed	SA for each time step, SA for peak value, SA for integral
Number of samples used	8192,4096
Dataset	mc8192-6-siton-r_SD_row.dat, mc4096-6-siton-r_SD_row.dat
Software	Python SaLib, Python in-house software

Table 4-11. Sensitivity analysis of GRS Clay by IBRAE

The sensitivity analysis by IBRAE as documented in Table 4-11 was performed using both available datasets (4096 and 8192 model runs). The summary of results is provided in Table 4-12. Figure 4-16 through Figure 4-21 demonstrate results by PAWN and RBD-FAST methods for different sample sizes. Note that RBD-FAST is identical to EASI in the software implementation

used. The results for the correlation- and regression-based method are mostly consistent with the results of other groups, so figures for them are not shown.

Overall sensitivity indices for *DiffClay3* are the highest, the second significant parameter is *DiffClay2*. However, the differences for the different methods and sample sizes occurred in the relative significance of the *DiffClay2* and *DiffClay3* at the "early" time steps. *DiffClay3* indices are higher according to the RBD method, while according to the PAWN method *DiffClay3* indices are equal to *DiffClay2* on the 8192 sample, and their mutual ranking fluctuates on the sample of 4096 realizations. Another ambiguity among different methods and sample sizes regards the ranking of the parameters of secondary significance (*DiffClay2*, *KdBent*, *KdClay3*) at the end of the simulation (10^8 years).

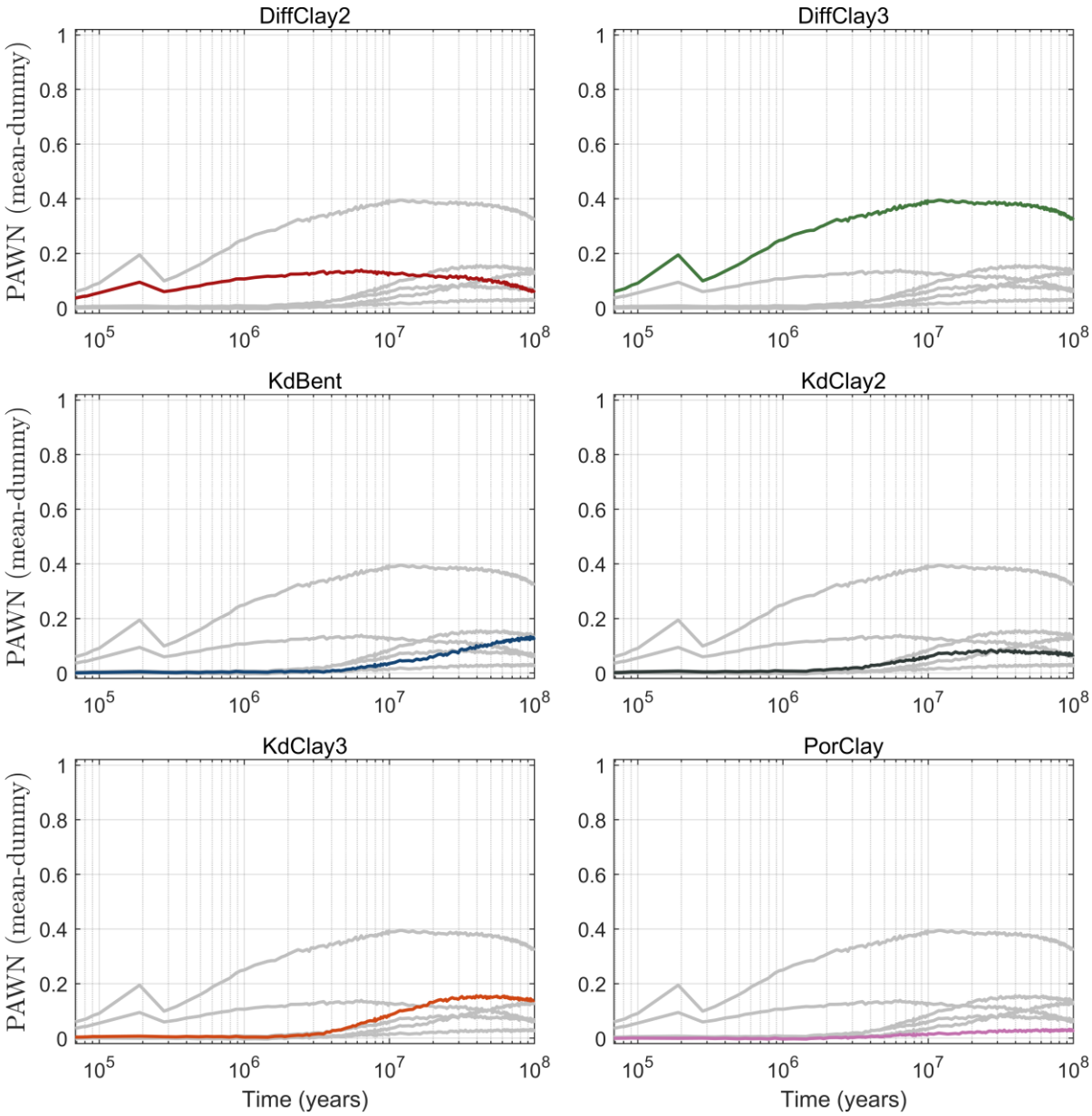

Figure 4-16. Sensitivity analysis of GRS Clay case by IBRAE: PAWN method (mean statistics, "dummy" parameter subtracted), 4096 realizations

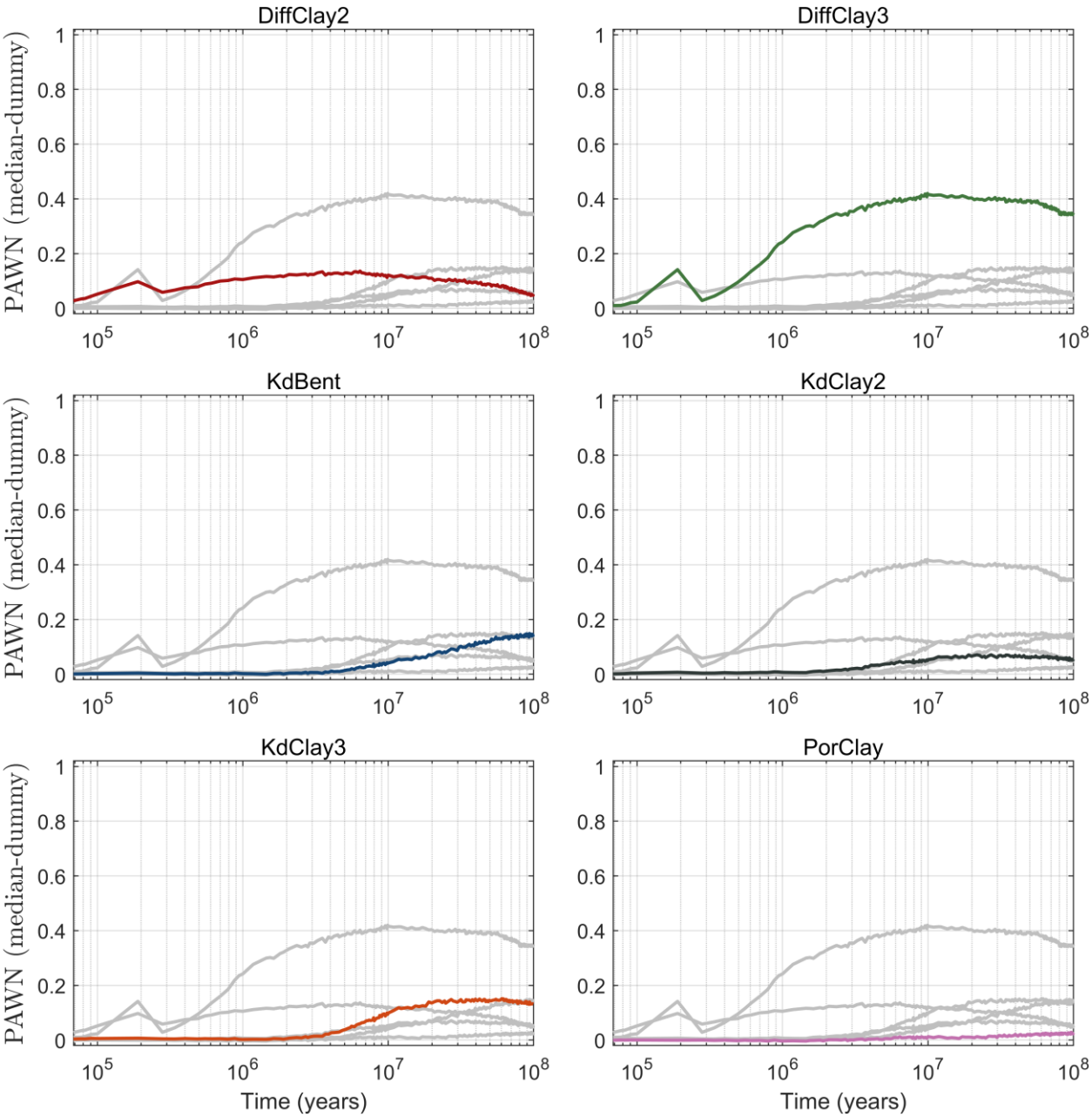

Figure 4-17. Sensitivity analysis of GRS Clay case by IBRAE: PAWN method (median statistics, "dummy" parameter subtracted), 4096 realizations

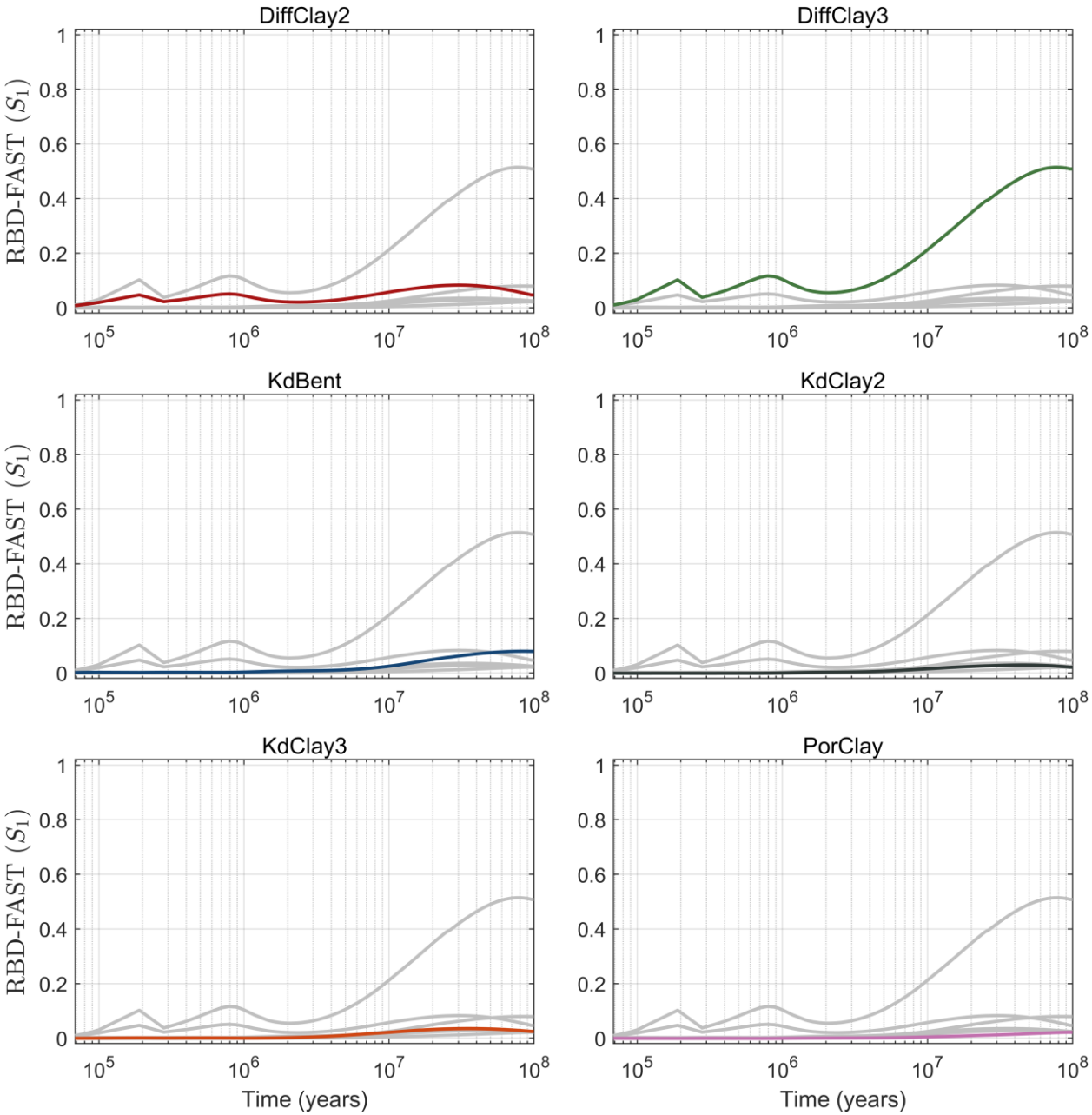

Figure 4-18. Sensitivity analysis of GRS Clay case by IBRAE: RBD-FAST method, 4096 realizations

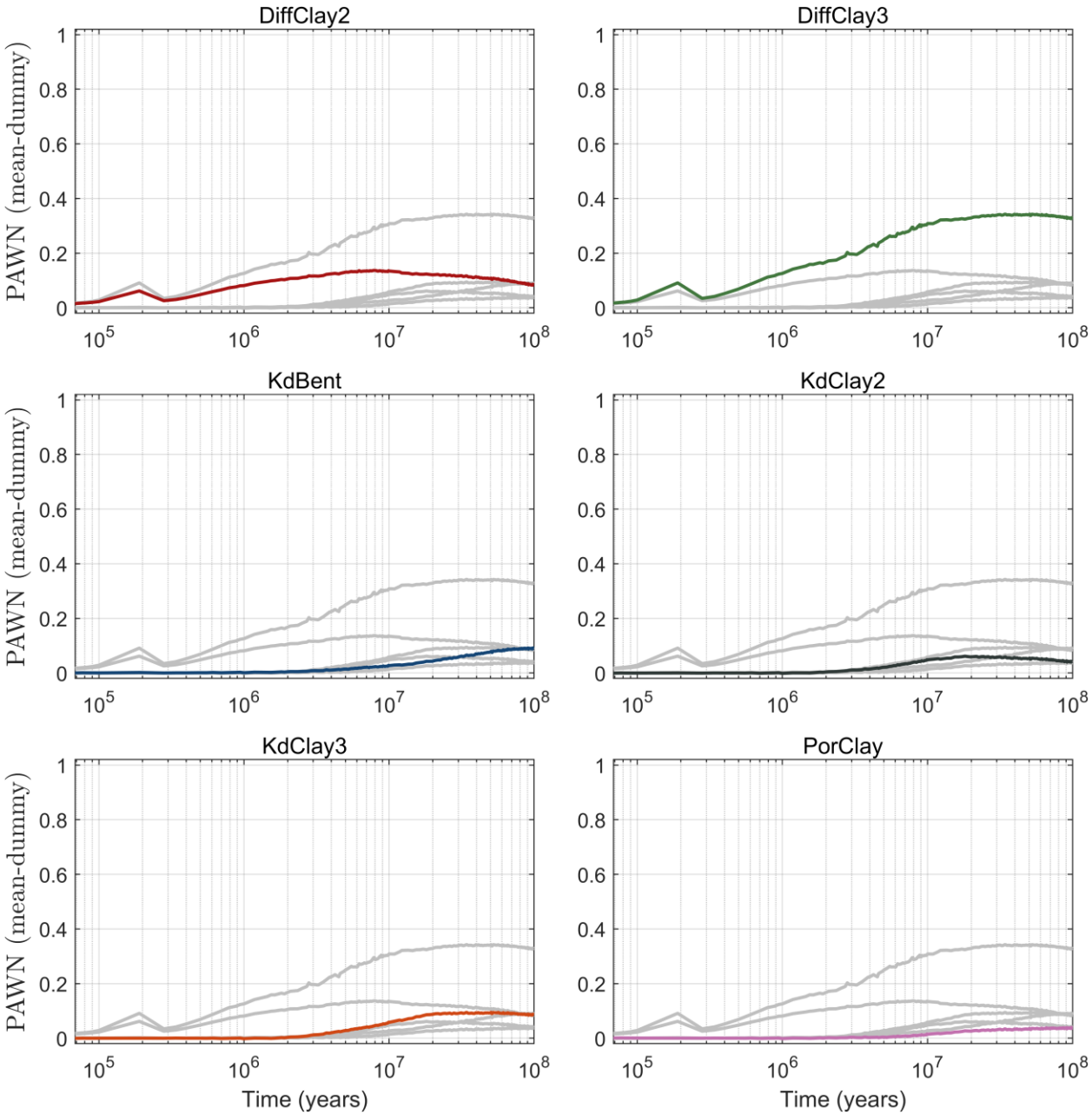

Figure 4-19. Sensitivity analysis of GRS Clay case by IBRAE: PAWN method (mean statistics, "dummy" parameter subtracted), 8192 realizations

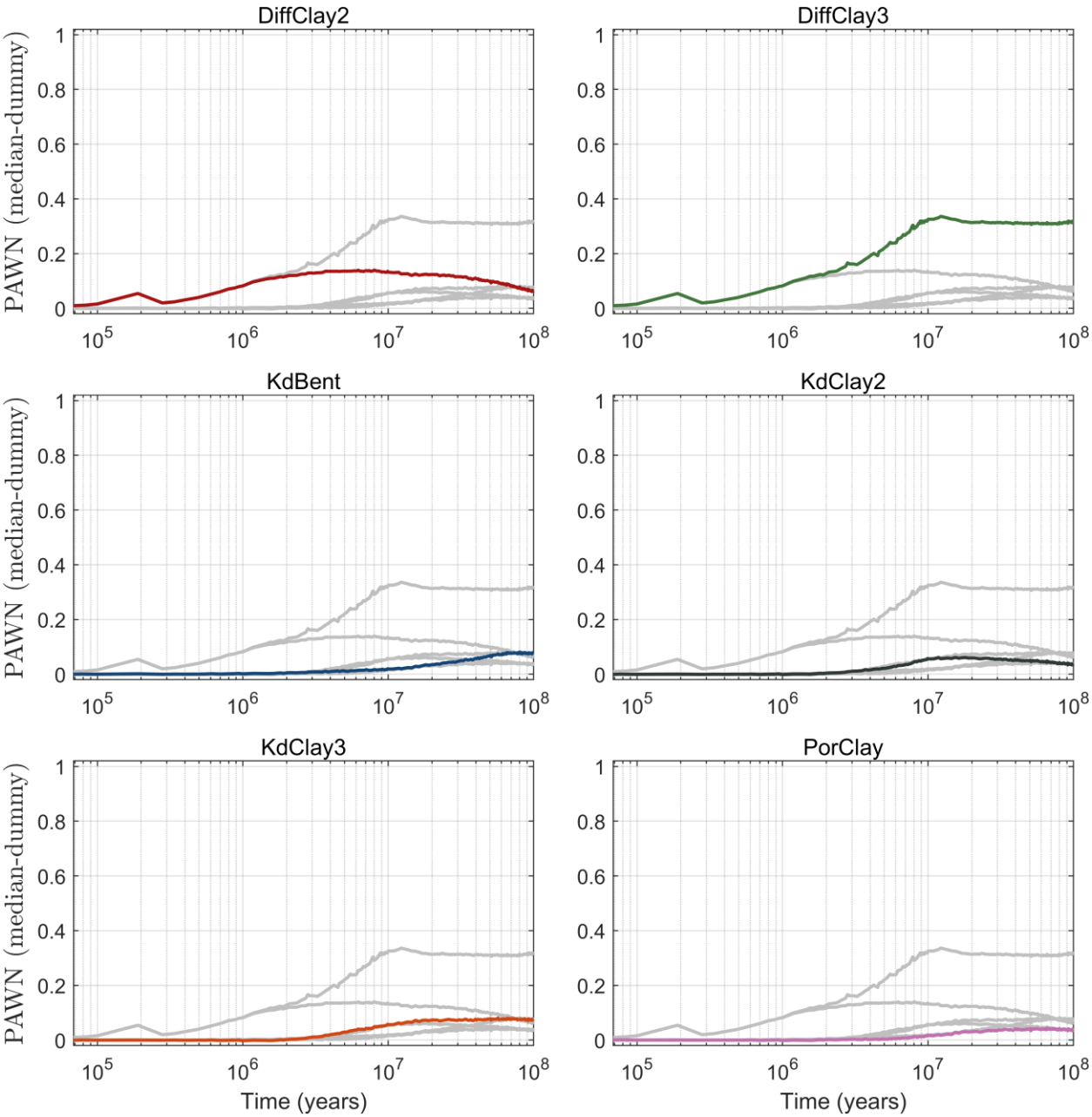

Figure 4-20. Sensitivity analysis of GRS Clay case by IBRAE: PAWN method (median statistics, "dummy" parameter subtracted), 8192 realizations

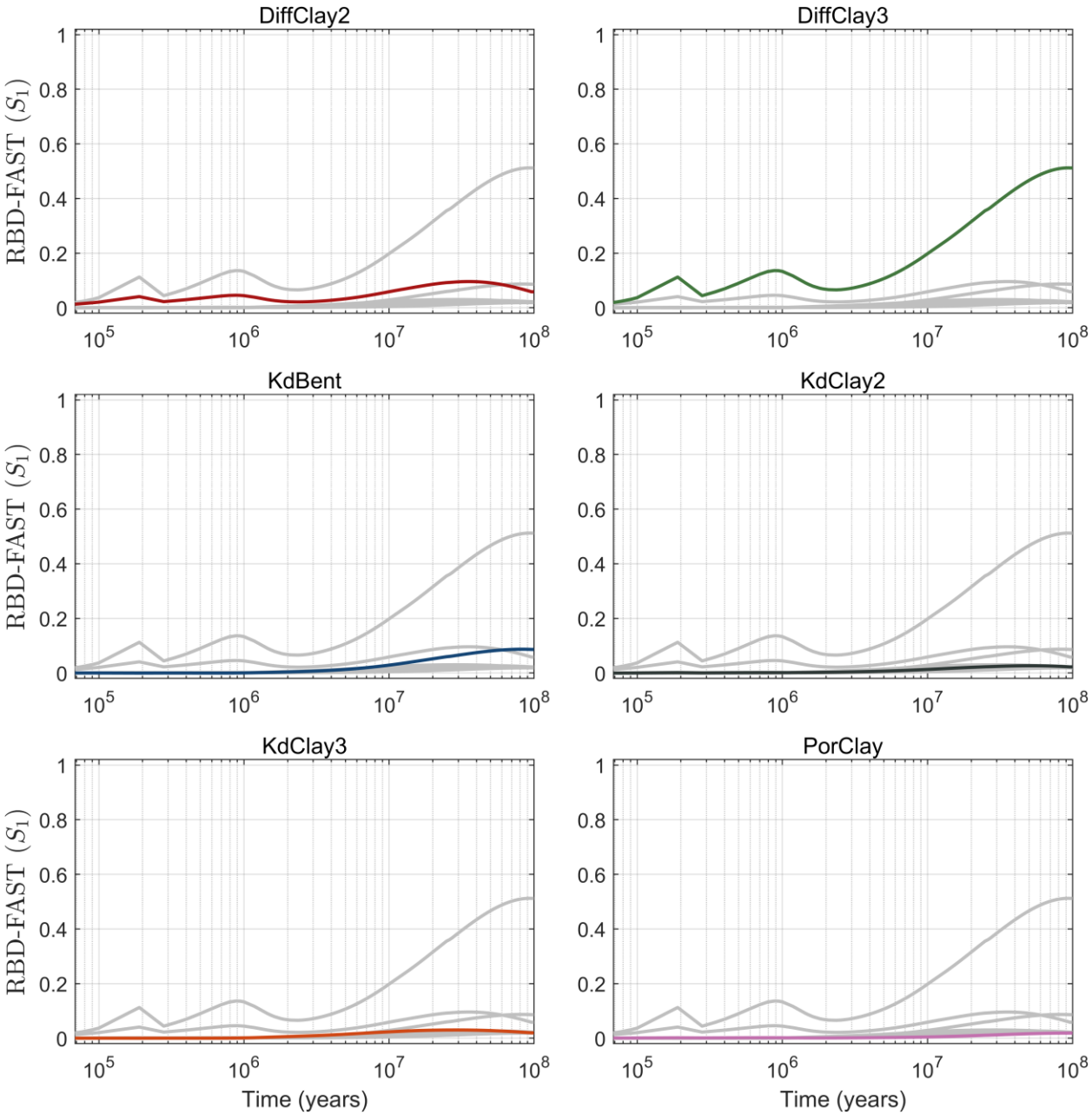

Figure 4-21. Sensitivity analysis of GRS Clay case by IBRAE: RBD-FAST method, 8192 realizations

Input parameter	PAWN	RBD-FAST	Correlation & regression methods
Parameter 1 *DiffClay2*	Overall small significance (~0.15 for integral and peak value). One of only 2 influencing parameters at early stages ($10^5<t<10^6$ years)	Overall small significance (~0.08 for integral and peak value). One of 2 present parameters at early stages ($10^5<t<10^6$ years)	Second significant parameter for integral and peak value by regression and partial rank correlation methods (+0.5÷0.75).
Parameter 2 *DiffClay3*	Most significant at $t>10^6$ years (0.25-0.45, 0.35 at the end of simulation), for integral and peak value.	Overall dominating significance (~0.1 at early stages, 0.5 at the end of simulation)	Most significant for integral and peak value by whole group of methods (+0.5÷1).
Parameter 3 *KdBentonit1*	Insignificant at $t<10^6$ years. Second significance for the end of simulation (0.2)	Insignificant at $t<10^6$ yreas. Second significance for the end of simulation (0.09)	Insignificant for integral and peak value by regression method, small significance by the rest (-0.1÷0.3)
Parameter 4 *KdClay2*	Insignificant at $t<10^6$ years. Small significance for the end of simulation (0.1)	Insignificant at $t<10^6$ years. Small significance for the end of simulation (0.05)	Insignificant for integral and peak value by regression method, small significance by the rest (-0.1÷0.25)
Parameter 5 *KdClay3*	Insignificant at $t<10^6$ years. Third significance for the end of simulation (0.18)	Insignificant at $t<10^6$ years. Small significance for the end of simulation (0.02)	Insignificant for integral and peak value by regression method, small significance by Pearson, Spearman, Partial correlation, Rank regression (-0.1÷0.25), medium significance by Partial rank correlation (-0.6)
Parameter 6 *PoroClay*	Insignificant at $t<10^6$ years. Small significance for the end of simulation (0.05)	Insignificant at $t<10^6$ years. Small significance for the end of simulation (0.02)	Insignificant for integral and peak value by regression method, small significance by Pearson, Spearman, Partial correlation, Rank regression (+0.1÷0.15), medium significance by Partial rank correlation (+0.55)

Table 4-12. SA Results of GRS Clay by IBRAE

4.4.5. Results from POSIVA

The overview of the analyses conducted by Posiva is presented in Table 4-13. The transformation applied for Quantity of Interest (annual dose rate to an adult human individual) in addition to rank transformation is to consider maximum (or peak) dose across all time points in each realization. For this transformation, the motivation is to understand the behavior of the upper bound of model output because the maximum dose rate is typically an interesting quantity occurring at different times in different realizations. The same analyses are conducted for the maximum dose rate distribution in order to study importance of input parameters on the maximum of the QoI also.

GRS Clay	**Description**
Sensitivity Analysis Method	Linear correlation, linear regression, rank correlation, rank regression, EASI
Sensitivity measures generated	PEAR, SPEAR, SRC, SRRC, S_i
Special considerations	-
Surrogate models used	-
Transformations	Rank, QoI maximum over time points
QoIs addressed	Annual dose rate to an adult human individual at 194 time points
Number of samples used	4096 and 8192
Software	Ecolego software and Sensitivity Analysis toolbox (https://www.ecolego.se/)

Table 4-13. The overview of the analysis details conducted by Posiva for the GRS Clay Case

The summary of results with the applied methodology is presented in Table 4-14 and Table 4-15 for the dataset with 4096 samples. The selected time points are 10^6 and 10^8 years based on time evolution of model output (see Section 3.3) and they represent the early phase, where the model output increases above zero and the last calculated time point. The highest model output values are captured by observing the maximum values of the output and applying the presented methods on the resulting distribution. These results are presented in the "Max" column because the time point varies in each realization.

Input parameter	**PEAR**			**SPEAR**			**PCC**		
	10^6 yr	10^8 yr	Max	10^6 yr	10^8 yr	Max	10^6 yr	10^8 yr	Max
1. *DiffClay2*	0.206	0.116	0.233	0.265	0.195	0.287	0.211	0.151	0.274
2. *DiffClay3*	0.258	0.587	0.502	0.889	0.761	0.809	0.263	0.625	0.529
3. *KdBent*	-0.014	-0.274	-0.165	0.001	-0.245	-0.195	-0.014	-0.345	-0.196
4. *KdClay2*	-0.038	-0.150	-0.135	-0.013	-0.146	-0.171	-0.027	-0.161	-0.134
5. *KdClay3*	-0.032	-0.146	-0.151	-0.010	-0.222	-0.307	-0.019	-0.179	-0.163
6. *PorClay*	0.041	0.131	0.098	0.022	0.149	0.147	0.040	0.168	0.114

Table 4-14. Part 1/2 of sensitivity analysis measures produced by Posiva for the GRS Clay Case. The coloring indicates the two parameters with the highest absolute values of each sensitivity measure (red and blue are applied for positive and negative values, respectively)

Input parameter	SRC			SRRC			S_i (EASI)		
	10^6 yr	10^8 yr	Max	10^6 yr	10^8 yr	Max	10^6 yr	10^8 yr	Max
1. *DiffClay2*	0.204	0.111	0.226	0.281	0.208	0.298	0.044	0.046	0.061
2. *DiffClay3*	0.257	0.579	0.495	0.894	0.761	0.812	0.090	0.508	0.256
3. *KdBent*	-0.013	-0.265	-0.159	-0.002	-0.242	-0.193	0.001	0.080	0.039
4. *KdClay2*	-0.026	-0.118	-0.107	0.000	-0.121	-0.146	0.001	0.022	0.019
5. *KdClay3*	-0.018	-0.131	-0.131	-0.003	-0.217	-0.299	0.001	0.023	0.025
6. *PorClay*	0.037	0.123	0.091	0.020	0.147	0.144	0.000	0.022	0.008

Table 4-15. Part 2/2 of sensitivity analysis measures produced by Posiva for the GRS Clay Case. The coloring indicates the two parameters with the highest absolute values of each sensitivity measure (red and blue are applied for positive and negative values, respectively)

Sensitivity analysis methods applied here are only correlation and regression analyses. The values of correlation measures as a function of time were studied in all time points but deemed not necessary to present for this summary. Also, similar behavior of the sensitivity methods applied here were observed with the dataset of 8192 samples and separate conclusions are not drawn.

Overall, the measures agree qualitatively in that they show relatively low sensitivity of the model to all six input parameters in the time phase up to 1 million years (*DiffClay2* and *DiffClay3* indicate only some sensitivity while the contribution of other parameters seems to be close to zero). Although the sensitivity indices show low values up to 1 million years, the correlation or regression analysis are (almost) systematic within the applied sensitivity measures. The coefficient of determination (so-called R^2) values in the "early" time frame stay relatively low with about 0.6 at maximum for PEAR, indicating low linear correlation between the output and input parameters.

At the end of the full time-frame (10^8years), all applied methods show an increasing sensitivity of the *DiffClay3* parameter compared to other parameters. Also, *KdBent* gains more importance during the late time period, being the parameter with the lowest negative coefficient against the output. In terms of first-order sensitivity indices (S_i), *DiffClay3* is clearly the most important parameter for the model output but because the sum of indices is largely below 1, the interactions between parameters are evident. Further analyses would be needed to examine the contribution of other parameters with the *DiffClay3* parameter to the model output.

The maximum of the output is clearly sensitive also to *DiffClay3* but also *DiffClay2* is important. From the negative sensitivity measures against the model output, contributions of *KdBent* and *KdClay3* are the most important but exact order of importance remains uncertain due changes in measures with non-transformed and rank-transformed model output.

4.4.6. Case Summary of Sensitivity Results

The system was analyzed by the different partners using a variety of sensitivity analysis methods. In view of identifying the most influential parameters, the results are in good agreement with each other and yield the following statements:

The diffusion coefficient in the outer clay formation (*DiffClay3*) has the highest influence on the model output. Most evaluations agree that this is valid over the total model time period.

For model times below 10^7 years, the second most influential parameter is the diffusion coefficient in the inner clay formation (*DiffClay2*).

For model times below 10^6 years there is no significant influence of any of the other parameters.

At late times, the influence of *DiffClay2* decreases while those of the other parameters increase.

Total indices reach their maximum values between 10^6 and 10^7 years. Parameter interactions seem to play a decreasing role at later times.

Physically, these findings can be understood as follows: The outer clay formation is the most effective barrier, as it has an extent of 250 m, which is more than three times the thickness of the inner clay formation (80 m) and 25 times the thickness of the bentonite buffer (10 m). The contaminant transport through the clay layers is carried by diffusion, so it is plausible that *DiffClay3*, the diffusion coefficient of the outer clay dominates the release of contaminants. The same argumentation holds for *DiffClay2* being the second most important parameter. On the first sight, it may seem surprising that the K_d values in the clay layers only play a minor role, but this is due to the fact that in all[1] simulations the effective dose rate is exclusively dominated by ^{129}I (at later times), which has very low sorption coefficients in all layers, and by ^{79}Se(at medium times), which does not sorb at all. Better sorbing radionuclides remain trapped in the clay layers for a very long time, and it does not make a difference at which point in the formation they finally end up. At any time, relevant influence of K_d values on the model output can only come from ^{129}I.

As mentioned in Section 4.4, from the comparison of direct and rank-based sensitivity analysis one can conclude that in the time frame between 1 million and 4 million years, there is an influence of the K_d values (and also the bentonite porosity) only on the highest model output values. This effect comes from those runs in which the radionuclides move relatively fast and cause a significant dose rate already at medium times. Then the K_d value of Iodine, although on a low level, has an additional influence on the time of appearance of the Iodine peak, and with that on the output value.

The linear sensitivity measures calculated by the different partners agree more or less exactly, as all use the same algorithms. The calculated values of variance-based sensitivity indices, however, depend on the method applied. First-order indices have been calculated using EASI/RBD-FAST, EFAST, PCE and RS-HDMR. The results are not identical but in fair agreement with each other as well as with those of the regression/correlation analysis. Total-order indices were calculated by GRS using EFAST and RS-HDMR as well as by SNL using the PCE surrogate model. While the EFAST and PCE results are in a fair agreement, those obtained with RS-HDMR differ significantly. This may, at least in part, be due to the fact that RS-HDMR does not really calculate the total-order indices but approximates them by summing up the first- and all second-order indices of a parameter, neglecting all higher orders of interaction. Indeed, the total-order indices calculated by RS-HDMR are always below those obtained with EFAST. Generally, it can be seen that the total-order indices of all parameters reach their maximum values between 1 million and 10 million years and decrease at later times. Obviously, in the late time frame the system is less controlled by

[1] A set of 512 random model runs was sanalysed in view of the contributions of all radionuclides to the effective dose rate. Among these, not a single run was found in which any other radionuclide than ^{129}I and ^{79}Se plays a non-negligible role.

parameter interactions but increasingly by the sole influence of *DiffClay3,* which reaches a first-order index of 0.5 at the end of the model time.

Two- and three-parameter interactions, quantified by second- and third-order indices, were calculated by SNL using the PCE surrogate model. While at the end of the modeling period, the values are rather low, the second-order index of *DiffClay2* and *DiffClay3* reaches about 0.5 at 1 million years. These results are confirmed by the calculations of GRS with RS-HDMR. The high influence of interaction of these parameters can easily be understood physically, as both diffusion coefficients have to be high to allow significant amounts of radionuclides to be released already after 1 million years.

The PAWN analysis performed by IBRAE tells a similar story about the system as the total-order analysis. Although PAWN is a density-based approach and conceptually different from the variance-based methods, it quantifies, in a certain sense, the total influence of a parameter. Unlike the total-order indices, however, the PAWN curves do not show a significant decrease at the end of the simulation time.

The investigations of TU Clausthal are well appropriate to provide a different view on the system sensitivities, less focused on the usual time curves of quantities that condense the sensitivity to a single value between 0 and 1 (or -1 and +1). CUSUNORO curves allow resolving the sensitivity over the parameter range. These plots show that for all parameters except *PorClay*, the lower part of the parameter range is most influential, and at least for the leading parameters *DiffClay3* and *DiffClay2,* the influential range is further shifted to the lower end with increasing time. This is understandable as high diffusion coefficients cause relatively fast transport, so that at late times, the main peak has already been released and variation of the diffusion coefficients between high values does not change anything essential. The moment-independent analysis performed by TU Clausthal confirms the general findings about the temporal evolution of sensitivities.

5. SNL SHALE CASE

The model describes a generic repository for commercial spent nuclear fuel in a shale host rock. Over the million-year simulation, the repository is undisturbed – no disruptive geological events and no human intrusion. The near field and far field are simulated in a single 3-D model domain, containing layered stratigraphy, the repository, and a household well (a simple biosphere) downgradient of the repository. The repository is in a thick, low permeability shale with higher permeability aquifers above and below the shale. It holds thousands of waste packages, each of which is a heat and radionuclide source. A waste package radionuclide source term is activated at the time of waste package breach (which depends on temperature and a sampled waste package degradation rate constant). Radionuclide transport away from the repository is primarily diffusive until radionuclides reach the aquifers, when advection driven by an applied pressure gradient becomes important. This case study is documented in [85].

5.1. Case Description

The SNL shale calculation case assumes a mined repository located approximately 500 m below land surface, accessed by vertical shafts, and containing 70,000 MTHM of commercial SNF. PA simulations use a half-symmetry model domain, in which approximately 15% (10,962 MTHM) of the 70,000 MTHM inventory is explicitly gridded. With the reflection boundary condition, 30% (21,924MTHM) of the 70,000 MTHM inventory is included in PA simulations.

The generic stratigraphic column for the shale calculation case consists of (from the bottom up): a 450 m thickness of indurated shale interrupted by a 30-m thick sandstone aquifer; a 75-m thick limestone aquifer; a 585 m thickness of sealing shale including a 90 m thickness of a silty shale unit; a 60-m thick sandstone aquifer; and a 30 m thickness of unconsolidated overburden.

In this calculation case, emplacement of pressurized water reactor (PWR) SNF occurs in a mined repository located in a deep, homogeneous, thickly bedded, essentially flat-lying stratum in a geologically simple and stable environment. The calculation case assumes horizontal, in-drift emplacement with 12-PWR waste packages elevated on plinths of compacted bentonite and drifts buffered and filled with compacted bentonite pellets and/or bricks in one or two layers. PA simulations assume that access drift and shafts are filled with compacted bentonite buffer (low permeability, high sorption capacity). The simulations employ a single bentonite buffer with material properties appropriate for a compacted mixture of 70% bentonite and 30% quartz sand [85 – 87].

SNL Shale	**Description**
Waste form	Spent nuclear fuel in 12-PWR canisters
Engineered barriers	Bentonite buffer with material properties appropriate for a compacted mixture of 70% bentonite and 30% quartz sand
Repository description	Repository has 515-m depth with 84 drifts, 4 shafts, and an emplacement footprint of 2.6 km^2. 4200 waste packages emplaced with 50 packages per drift and 20-m center-to-center spacing.

Natural system	450 m thick indurated shale interrupted by a 30 m thick sandstone aquifer; a 75 m thick limestone aquifer; a 585 m thickness of sealing shale including a 90 m thickness of silty shale unit; a 60 m thick sandstone aquifer; 30 m thickness of unconsolidated overburden.
Biosphere	Well water ingestion dose model
Conceptual release pathways	Advection, diffusion, element-based solubility, medium-specific sorption, decay in all phases
Processes modeled	Waste package degradation, waste form (UO_2) dissolution, equilibrium-controlled radionuclide sorption and precipitation/ dissolution, radioactive decay and ingrowth in all phases (aqueous, adsorbed, precipitate), coupled heat and fluid flow, and radionuclide transport via advection and diffusion.
Software codes used	PFLOTRAN, Dakota
Reference for full description of case	P. E. MARINER et al., *Advances in Geologic Disposal System Modeling and Shale Reference Cases*. SFWD-SFWST-2017-000044 / SAND2017-10304R. Sandia National Laboratories, Albuquerque, NM (2017) [85].

Table 5-1. SNL Shale Case Description

5.2. Description of Inputs and Outputs

Two hundred simulations were run in PFLOTRAN. Dakota was used to apply Latin hypercube sampling to the 10 uncertain variables listed in Table 5-2. The uncertain variables were assumed to be independent and there was only one sampling loop with no separation of uncertainties. The quantities of interest were the maximum (over time) ^{129}I concentration at three sandstone locations and three limestone locations as listed in Table 5-3. These output quantities are also denoted as response functions (RF) #1-#6. Sensitivity for these quantities of interest was performed using Dakota and Python.

Input parameter	**Distribution**	**Range**	**Units**
SNF Dissolution Rate (*rateSNF*)	Log uniform	10^{-8} - 10^{-6}	yr^{-1}
Mean Waste Package Degradation Rate (*rateWP*)	Log uniform	$10^{-5.5}$ - $10^{-4.5}$	yr^{-1}
Upper Sandstone k (*kSand*)	Log uniform	10^{-15} - 10^{-13}	m^2
Limestone k (*kLime*)	Log uniform	10^{-17} - 10^{-14}	m^2
Lower Sandstone k (*kLSand*)	Log uniform	10^{-14} - 10^{-12}	m^2
Buffer k (*kBuffer*)	Log uniform	10^{-20} - 10^{-16}	m^2
DRZ k (*kDRZ*)	Log uniform	10^{-18} - 10^{-16}	m^2
Host Rock (Shale) Porosity (*pShale*)	Uniform	0.1-0.25	--
Np K_d Buffer (*bNpKd*)	Log uniform	0.1-702	m^3kg^{-1}
Np K_d Shale (*sNpKd*)	Log uniform	0.047-20	m^3kg^{-1}

Table 5-2. SNL Shale Case Input Parameters [85]

Output QoI	Type	Other
sand_obs1, sand_obs2, sand_obs3	Scalars: Maximum ^{129}I concentration over time at three observation points in the sandstone aquifer. The first observation point is closest to the repository and the third is the farthest.	Also called RF1, RF2, and RF3.
lime_obs1, lime_obs2, lime_obs3	Scalars: Maximum ^{129}I concentration over time at three observation points in the limestone aquifer. The first observation point is closest to the repository and the third is the farthest.	RF4, RF5, and RF6.

Table 5-3. SNL Shale Case Output QoIs

5.3. Salient features and behavior of the model

Horsetails of the ^{129}I concentrations at the closest and farthest sandstone and limestone observation points from the repository are shown in Figure 5-1. At the observation points farthest from the repository, the ^{129}I concentration remains at the background concentration level, 10^{-20} [M], in approximately half of the realizations at most time steps; the median is visible late in the simulation for sand_obs3, but never visible in the plot for lime_obs3. The concentration curves are also monotonic, so the maximum concentrations, which are shown in scatterplots, are at the end of the simulation for all realizations. These simulations present a challenge for sensitivity analyses in which the maximum concentration over time is the quantity of interest; there are many realizations with the same low maximum concentration and the rest of the simulations have maximum concentrations that span many orders of magnitude.

The maximum ^{129}I concentrations for the sandstone aquifer observation points are plotted with the uncertain input variables in Figure 5-2. The neptunium K_d uncertain variables are not included since they do not affect ^{129}I concentrations. The maximum ^{129}I concentrations are calculated as the maximum over time. From these scatterplots, it appears that the shale porosity is the most important uncertain variable near the repository, with the upper sandstone k gaining importance farther from the repository.

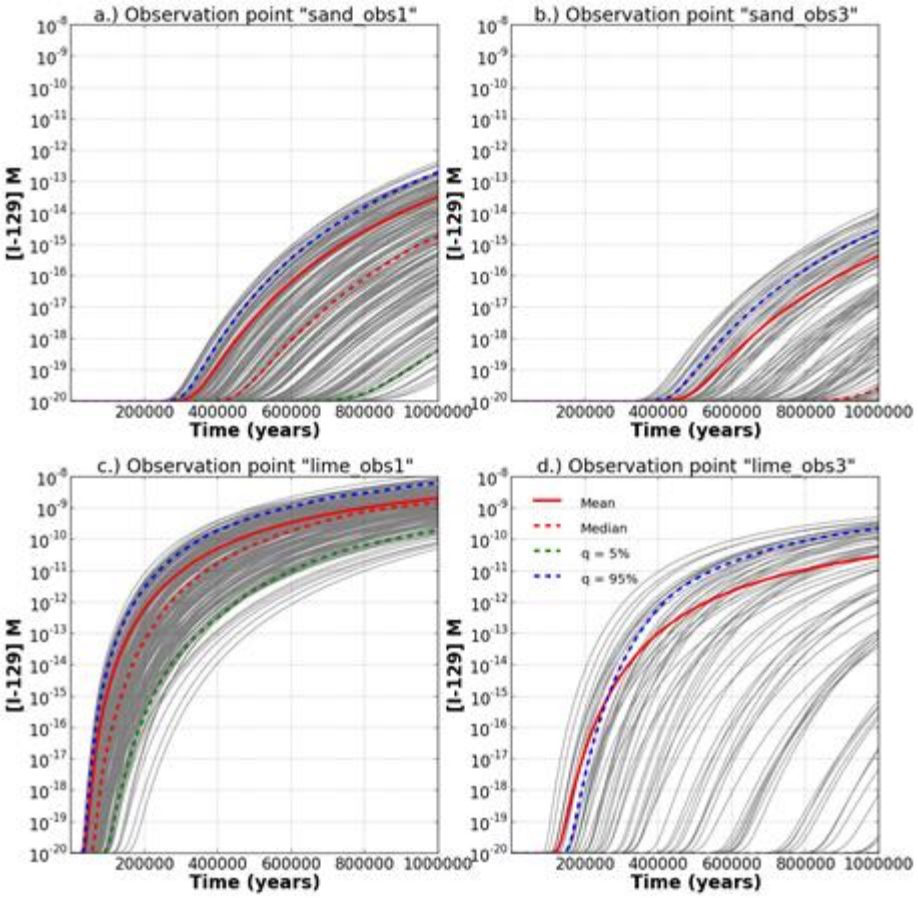

Figure 5-1. Horsetails for ^{129}I concentration at the closest (obs1) and furthest (obs3) points from the repository in the sandstone and limestone aquifers for the SNL Shale calculation case

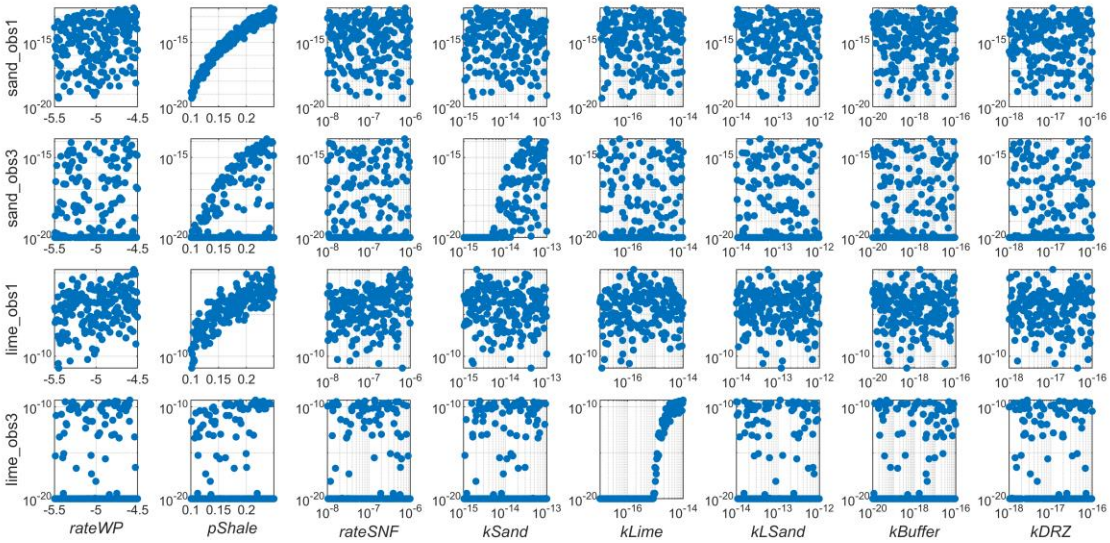

Figure 5-2. Scatterplots of ^{129}I concentrations for three observations points in the sandstone and limestone aquifers for the SNL Shale calculation case

5.4. Sensitivity Analysis Results

5.4.1. Results from GRS

GRS performed a number of investigations with both available datasets (50 and 200 realizations). The sample sizes are fairly small, so the sensitivity analysis should be interpreted carefully. Nevertheless, some obviously valid statements about sensitivities can be derived even from the smaller dataset.

Case Name	Description
Sensitivity Analysis Method	CUSUNORO, Linear correlation, linear regression, rank correlation, rank regression, EASI
Sensitivity measures generated	Pearson CC, Spearman RCC, SRC, SRRC, S_i
Special considerations	-
Surrogate models used	-
Transformations	-
QoIs addressed	Response functions 1 to 6
Number of samples used	50 / 200

Table 5-4. Sensitivity analysis of SNL shale model by GRS

The CUSUNORO analysis is presented in Figure 5-3. For RF1 it shows a clear deviation from the horizontal line only for the parameter *pShale*. Although, at least in the set with 200 runs, there seems to be a certain bend also in the curves for *kSand* and *rateWP*, this cannot be judged as significant. The situation is similar for RF2 and RF3, except that the direction of influence of *kSand* is obviously reversed. For RF4 the parameter *rateSNF* seems to gain some importance. For RF5 and RF6, *kLime* is the most conspicuous parameter, but its curves follow more or less straight lines, except a sharp bend around an input contribution of 0.8, corresponding to a permeability value of about 2.5E-15 m². This shape of the CUSUNORO curves suggests some change in the model behavior in a small range of values of *kLime*.

The number of simulations is too low for clearly identifying less significant sensitivities by using CUSUNORO.

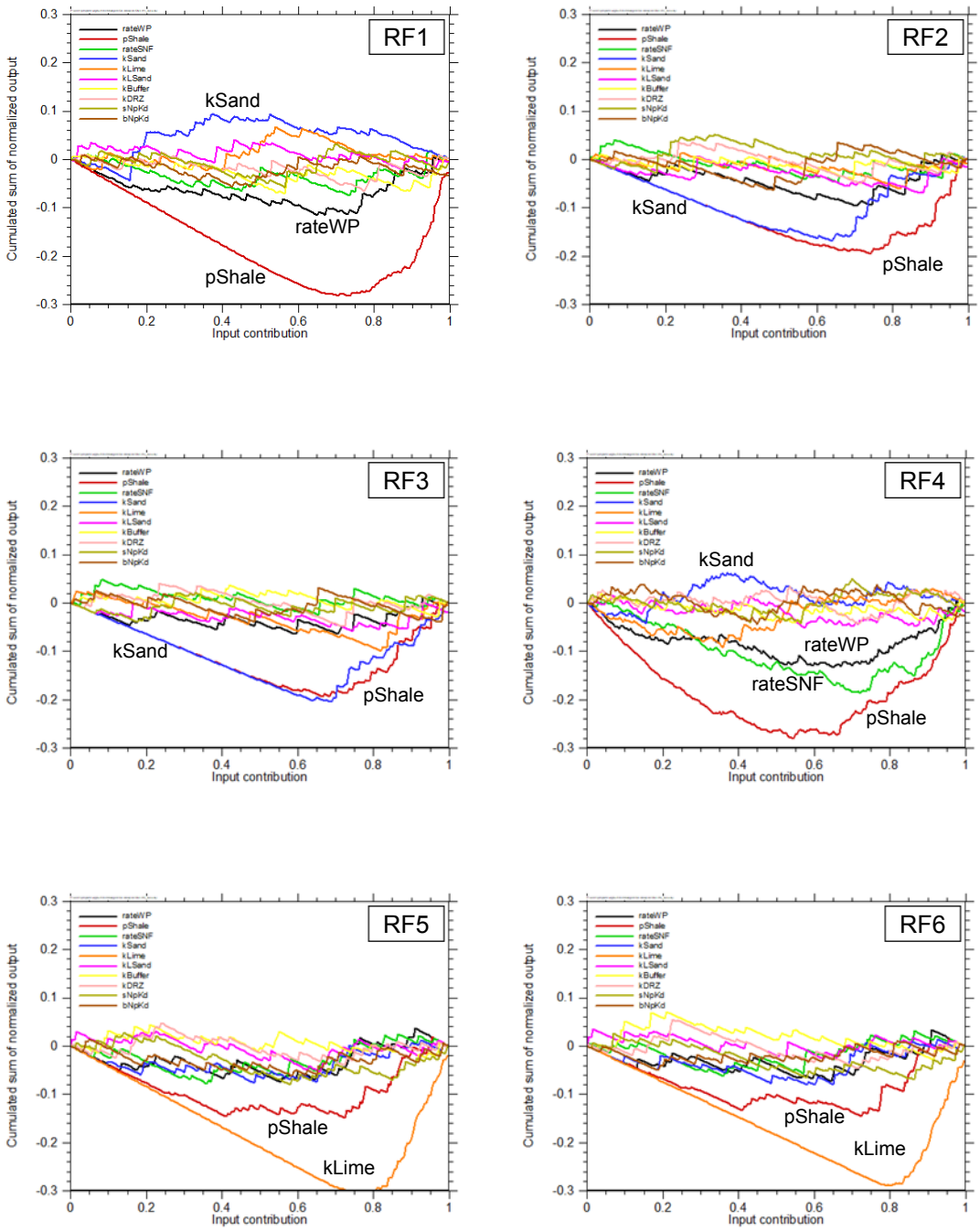

Figure 5-3. CUSUNORO analysis of the shale system

The results of the sensitivity analysis for RF1 using the correlation-/regression-based methods as well as the first-order sensitivity indices calculated using EASI are presented in Table 5-5.

Input parameter	Pearson CC		Spearman Rank CC		Regression SRC		Rank Regression SRRC		EASI S_i	
	50	200	50	200	50	200	50	200	50	200
rateWP	0.256	0.217	0.099	0.138	0.306	0.224	0.132	0.149	0.284	0.113
pShale	0.655	0.587	0.983	0.941	0.609	0.586	0.985	0.943	0.770	0.590
rateSNF	0.036	0.106	0.009	0.001	-0.009	0.095	0.020	-0.010	0.344	0.049
kSand	-0.190	-0.155	-0.116	-0.084	-0.121	-0.161	-0.064	-0.081	0.206	0.080
kLime	0.088	-0.042	0.041	-0.002	-0.108	-0.087	0.023	-0.006	0.265	0.071
kLSand	0.008	-0.003	0.027	-0.023	-0.120	-0.033	0.015	-0.015	0.176	0.032
kBuffer	0.583	0.168	-0.042	0.002	0.462	0.121	-0.001	0.010	0.511	0.086
kDRZ	0.171	0.090	-0.003	-0.010	0.050	0.063	0.001	-0.006	0.182	0.044
sNpKd	0.113	0.016	-0.018	-0.014	0.016	0.039	-0.014	-0.011	0.176	0.069
bNpKd	-0.060	0.097	0.011	-0.019	-0.173	0.111	0.006	-0.018	0.177	0.068

Table 5-5. SA Results of SNL shale model by GRS – RF1. The three leading parameters are marked in red, blue and green

For the set with 200 runs, all methods agree about the ranking of the first two parameters (*pShale* and *rateWP*). The third one is sometimes *kBuffer* and sometimes *kSand*. The value-based correlation or regression methods calculate very similar sensitivities for these two parameters. It is, however, conspicious that the rank-based methods seem to be much more unambiguous in this respect and rate *kSand* at least 8 times higher than *kBuffer*. This can be a hint that *kBuffer* acts mainly on the high model output values, which are essentially mitigated by rank transformation.

For the set with 50 runs, only the leading parameter *pShale* is identified sufficiently clearly, there is no agreement about the ranking of the other ones. However, the observation concerning the difference between direct and rank-based evaluation for *kBuffer* is basically the same as described above.

5.4.2. Results from SNL

Case Name	Description
Sensitivity Analysis Method	Surrogate modeling, PEAR, SPEAR, stepwise linear regression
Sensitivity measures generated	S_i, T_i, Pearson CC, Spearman RCC, SRC, incremental R^2
Special considerations	
Surrogate models used	GP, PCE (order 4)
Transformations	Scaling by 10^{13}
QoIs addressed	Maximum (over time) ^{129}I concentrations at 3 sandstone and 3 limestone observation points.
Number of samples used	All (200)

Table 5-6. Sensitivity analysis of SNL Shale case by SNL

The PCE surrogate model in Dakota could not estimate sensitivity measures for some of the quantities of interest due to low variance. This is because many of the maximum ^{129}I concentrations

are low (background) at observation points farther from the repository. To resolve this, the concentrations were all multiplied by 10^{13}. Scaling by 10^{20} was also performed and this did not significantly change the sensitivity measures. The PCE model was also fit using log-transformed ^{129}I concentrations.

At observation points in the sandstone aquifer, all methods identified *pShale*, *rateWP*, and *kSand* as the input variables with the greatest influence on maximum^{129}I concentration. Sensitivity to *pShale* decreases with distance from the repository, while sensitivity to *kSand* increases with distance. At the observation point closest to the repository (obs1), maximum ^{129}I concentration is negatively correlated with *kSand*; at observation points 2 and 3, it is positively correlated. Substituting *kLime* for *kSand*, behavior at the limestone observation points is very similar to behavior at the sandstone observation points, with *kLime* behaving analogously to *kSand.*

Correlations that are larger on rank-transformed data (simple rank, partial rank in the plots) than correlations on raw values (simple and partial in the plots), as for *kSand* at sand_obs2 and sand_obs3, suggest a monotonic but nonlinear correlation between input and output. Partial correlation coefficients (partial and partial rank in the plots) that are larger than a simple correlation coefficient (simple and simple rank in the plots), as for *rateWP* at sand_obsl and for *pShale* at sand_obs2 and sand_obs3, reveals a correlation that was partially masked by the influence of other inputs in the calculation of the simple correlation coefficient.

Stepwise linear regression and sensitivity indices provide a measure of the fraction of the variance in the output due to the variance in an input. For stepwise linear regression, this fraction is taken to be the difference between the R^2 obtained by the most recent addition of a variable to the regression and R^2 at the previous step of the regression (incremental R^2). Table 5-7 compares the fraction of the variance accounted for by stepwise linear regression with the fraction of the variance accounted for by main sensitivity indices obtained using either the GP or PCE surrogate model at sand_obs2. Individual input variables (without interactions) account for less than half of the variance in the maximum ^{129}I concentration at sand_obs2, regardless of the surrogate model used.

For *pShale*, *kSand*, and *rateWP*, the total sensitivity index is several times larger than the main sensitivity index, indicating the importance of parameter interactions. When the effects of parameter interactions are included, variance in the sampled input *pShale* accounts for 0.601 or 0.744 of the total variance in maximum ^{129}1 concentration at sand_obs2, almost twice that accounted for by main effects (of all input parameters) alone.

Variable	Incremental R^2	S_i (PCE)	S_i (GP)	T_i (PCE)	T_i (GP)
pShale	1.73E-01	1.79E-01	2.95E-01	6.01E-01	7.44E-01
kSand	6.20E-02	9.40E-02	1.33E-01	5.31E-01	5.75E-01
rateWP	2.10E-02	3.80E-02	2.00E-02	2.53E-01	1.59E-01
kLSand	1.80E-02	2.00E-03	0.00E+00	1.28E-01	0.00E+00
Fraction of variance accounted for:	2.75E-01	3.12E-01	4.49E-01		

Table 5-7. Sensitivity measures from sensitivity analysis of SNL Shale case by SNL

The sensitivity analysis results from each of the methods are shown for all quantities of interest in Figure 5-4 and Figure 5-5.

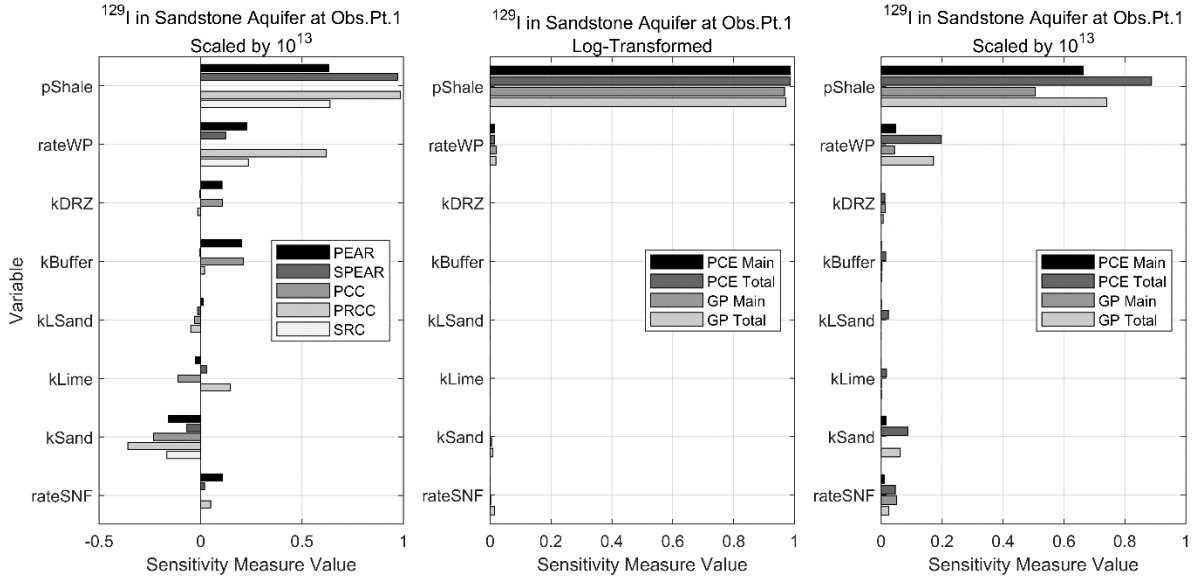

Figure 5-4. Sensitivity measures for the sandstone aquifer maximum ^{129}I concentration at observation point 1 from the SNL analysis of the SNL Shale calculation case

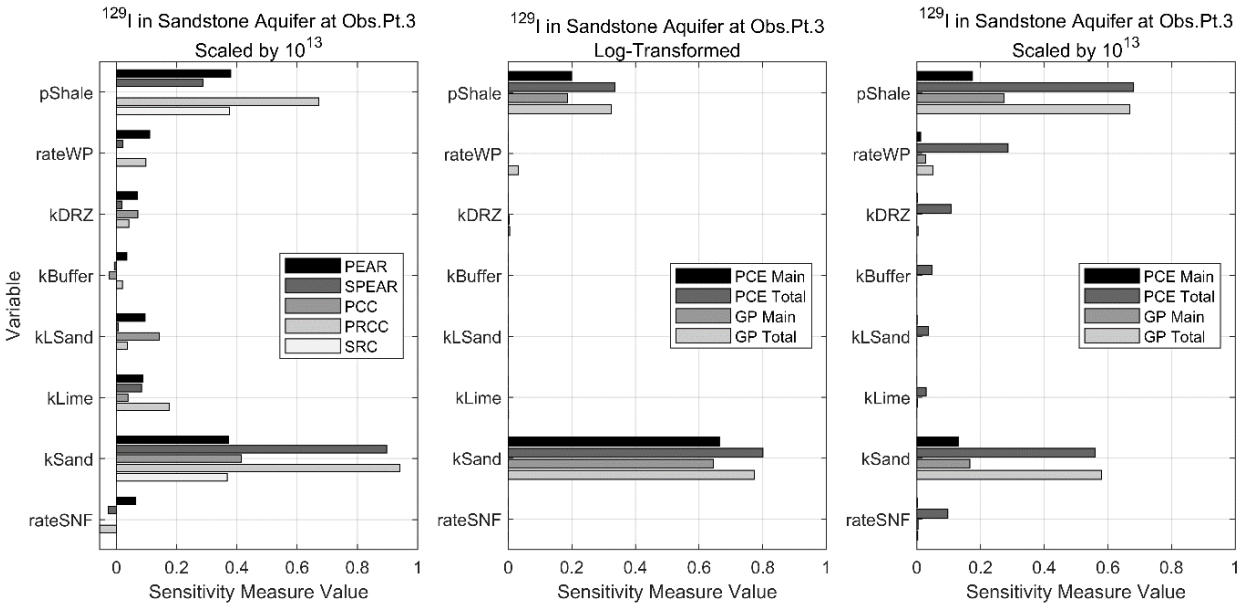

Figure 5-5. Sensitivity measures for the sandstone aquifer maximum ^{129}I concentration at observation point 3 from the SNL analysis of the SNL Shale calculation case

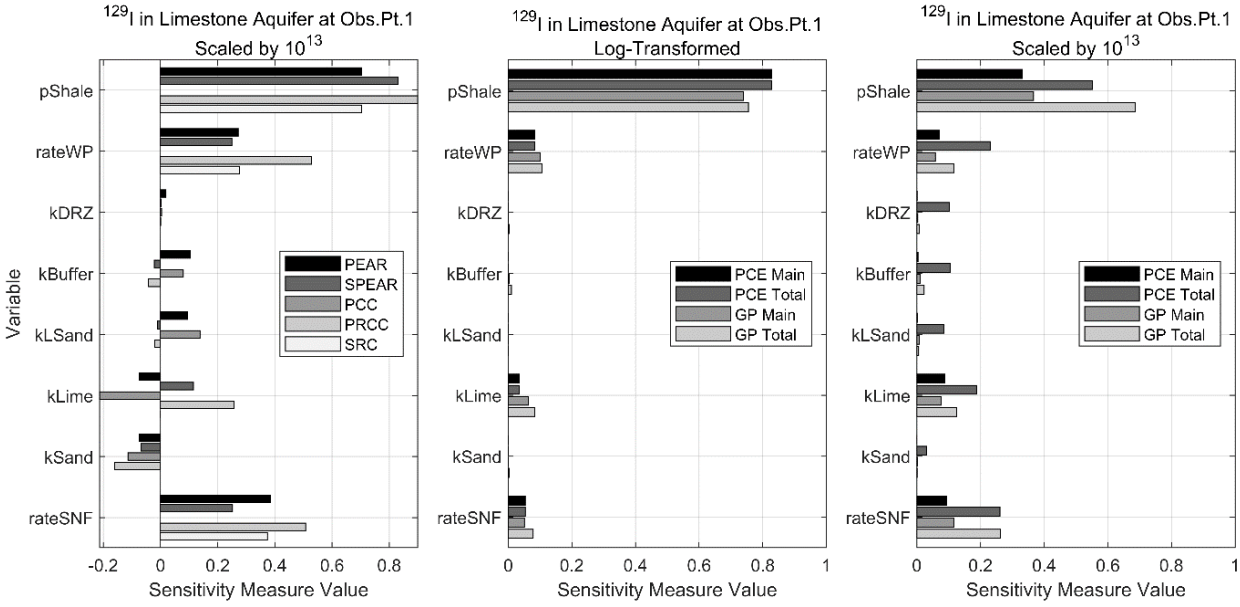

Figure 5-6. Sensitivity measures for the limestone aquifer maximum ^{129}I concentration at observation point 1 from the SNL analysis of the SNL Shale calculation case

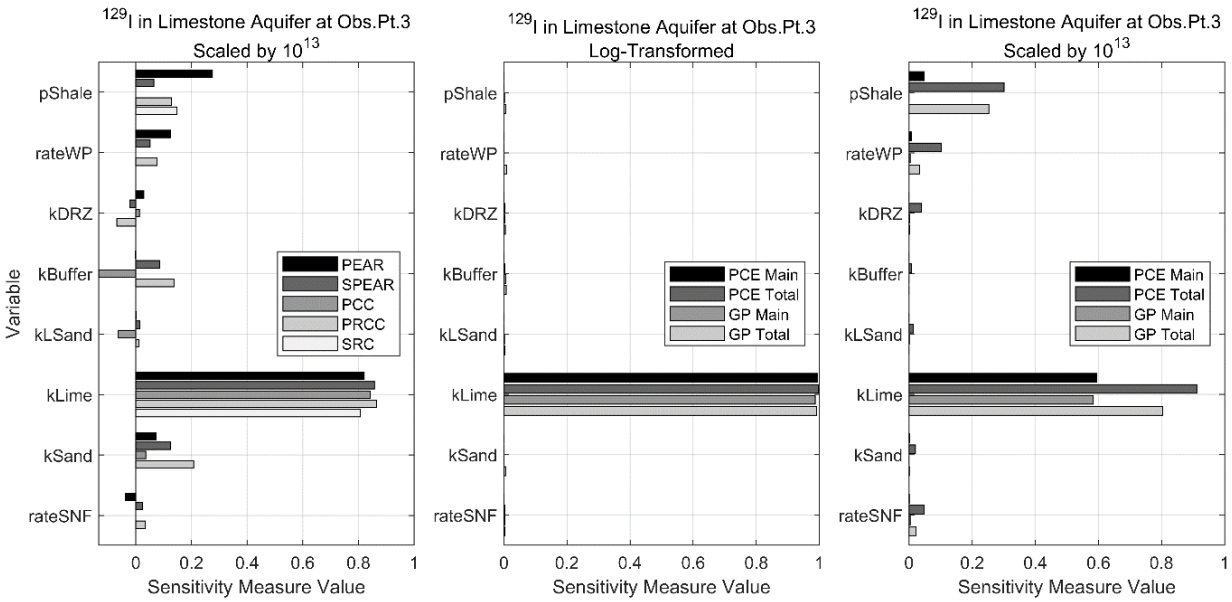

Figure 5-7. Sensitivity measures for the limestone aquifer maximum ^{129}I concentration at observation point 3 from the SNL analysis of the SNL Shale calculation case

5.4.3. Results from TUC

The sensitivity analysis methods used by TUC are shown in Table 5-8 with the results listed in Table 5-9.

Case Name	Description
Sensitivity Analysis Method	CUSUNORO, Copula distance, PCE, MIM
Sensitivity measures generated	S_i first order, S_{i2} second order
Special considerations	Polynomial chaos uses harmonic base functions
Surrogate models used	
Transformations	None, Log, Rank, Second order finite difference for interactions
QoIs addressed	SA for each output
Number of samples used	200 / all
Dataset	SNLShaleRep2_SAND2018-0517O_Rev1
Software	MatLab/Octave: in-house software

Table 5-8. Sensitivity analysis of SNL Shale by TU Clausthal

Input parameter	CUSUNORO	2nd order finite differences
Parameter 1 *rateWP*	Small influence in all responses (<10%)	
Parameter 2 *pShape*	Minor (20%-60%) influence in 4 responses	Cause of interactions in 4 responses
Parameter 3 *rateSNF*	Small influence in all responses (<10%)	
Parameter 4 *kSand*	Small influence in all responses (<20%)	
Parameter 5 *kLime*	Major (>60%) in two responses (limeobs2, 3), else small	Cause of interactions in 2 response (limeobs2,3)
Parameter 6 *kLSand*	Negligible influence in all responses (<5%)	
Parameter 7 *kBuffer*	Negligible influence in all responses (<5%)	
Parameter 8 *kDRZ*	Negligible influence in all responses (<5%)	
Parameter 9 *sNpKd*	Small influence in all responses (<10%)	
Parameter 10 *bNpKd*	Negligible influence in all responses (<5%)	

Table 5-9. SA Results of SNL Shale by TU Clausthal

Depending on model response analyzed, very different behaviors can be observed, see Table 5-9 and Figure 5-8. Some parameters are of minor influence throughout all outputs. While the analysis provides some clues on possible interactions, switching to a log-scale is without clear findings. This might be a sign that the interaction is multiplicative, such that the logarithm transfers the model into an essentially additive one with no interaction term present.

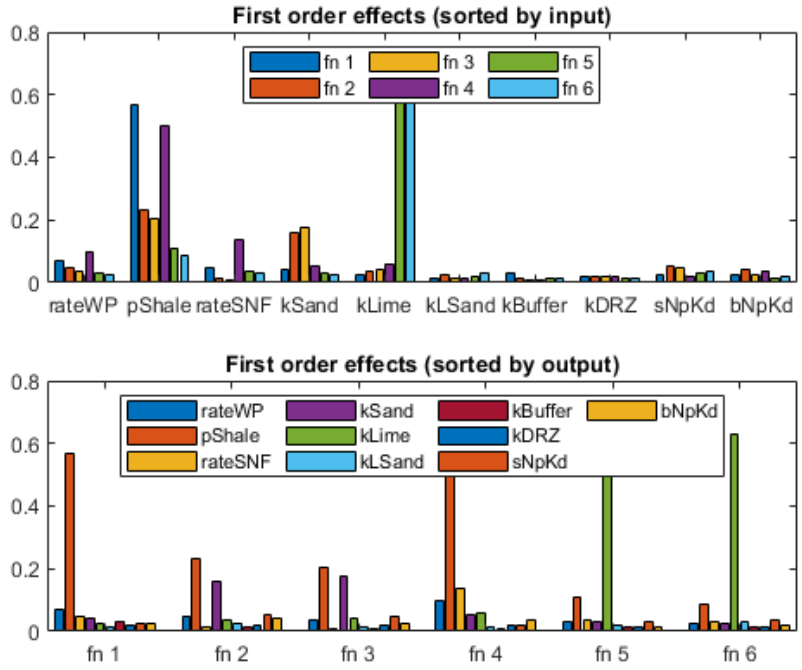

Figure 5-8. First order sensitivity measures for the maximum ^{129}I concentration. Sandstone aquifer observation points: fn 1 - fn 3, limestone aquifer observation points: fn 4 - fn 6

5.4.4. Results from IBRAE

Case Name	**SNL Shale**
Sensitivity Analysis Method	PAWN, RBD-FAST, Correlation and regression
Sensitivity measures generated	PAWN index, S_i first order, Pearson correlation, Spearman rank correlation, partial correlation, partial rank correlation, regression and rank regression coefficients
Special considerations	Mean and median KS for PAWN. Number of conditioning intervals $n = 8$ for larger sample and $n = 5$ for smaller sample.
Surrogate models used	
Transformations	
QoIs addressed	Response functions sand_obsi, lime_obsi $i = [1\ldots3]$ (maximum ^{129}I concentration at 3 sandstone and 3 limestone locations)
Number of samples used	200, 50
Dataset	SNLShaleRep2_SAND2018-0517O_Rev1.txt, SNLShaleRep1_SAND2017-11069O_Rev1.txt
Software	Python SaLib, Python in-house software

Table 5-10. Sensitivity analysis of SNL Shale by IBRAE

IBRAE performed the sensitivity analysis using both available datasets (50 and 200 realizations).

The results are shown in Table 5-11, Figure 5-9, and Figure 5-10. Figures for correlation-based and regression-based methods are omitted due to their similarity with the results of other groups.

In short, *pShale* is most significant for sand_obs1 and lime_obs1, *kSand* accompanies *pShale* for sand_obs2 and sand_obs3, *kLime* shows major influence for lime_obs2 and lime_obs3.

Overall results for both samples are almost similar, some discrepancies though occurred. In particular, the RBD variant of the FAST method with the smaller sample indicates *kBuffer* for sand_obs1, *rateWP* and *kSand* for lime_obs1, *rateSNF* for lime_obs2, *rateWP* and *rateSNF* for lime_obs3 as influencing while with larger sample these parameters are detected as insignificant.

PAWN approach on the contrary fails to detect the influence of the *pShale* for the sand_obs3 using smaller sample.

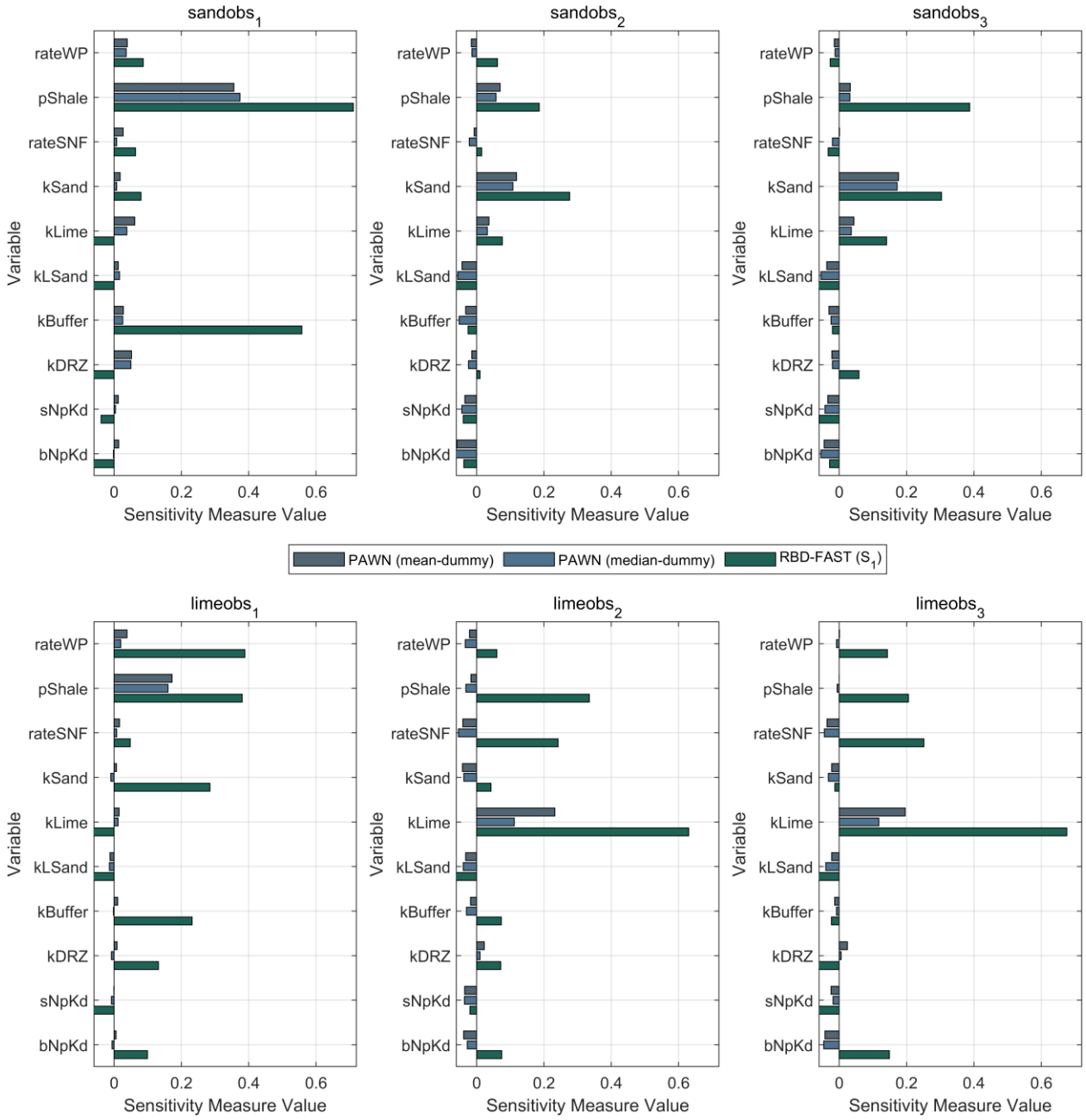

Figure 5-9. Sensitivity analysis of SNL Shale case (QoI 1-6) by IBRAE: RBD-FAST and PAWN methods (for PAWN "dummy" parameter is subtracted), 50 realizations

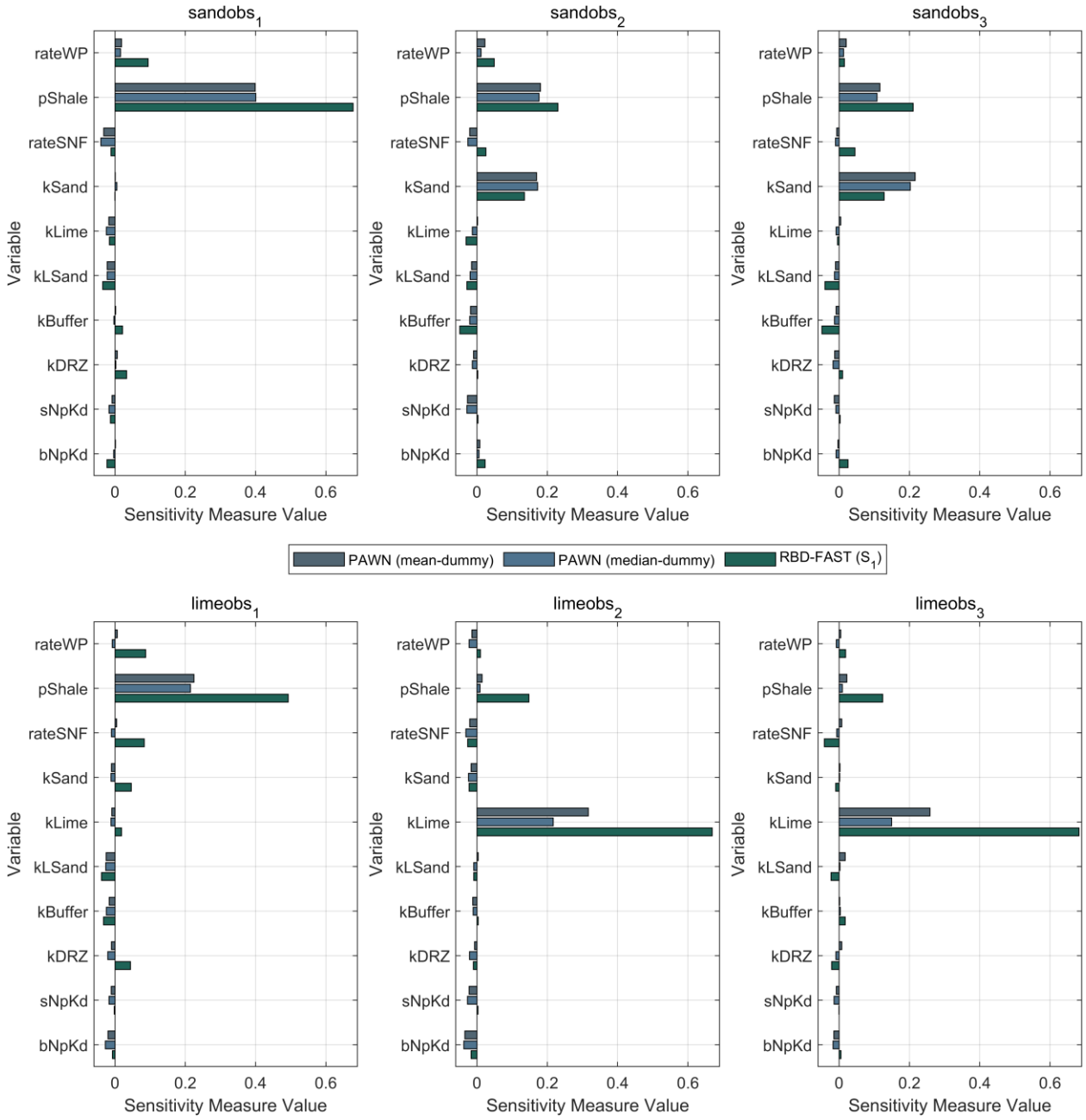

Figure 5-10. Sensitivity analysis of SNL Shale case (QoI 1-6) by IBRAE: RBD-FAST and PAWN methods (for PAWN "dummy" parameter is subtracted), 200 realizations

Input parameter	PAWN	RBD-FAST	Correlation & regression methods
rateWP	Small significance for sand_obs$_2$, sand_obs$_3$, lime_obs$_1$. Insignificant for the rest.	Small significance for sand_obs$_1$, sand_obs$_2$, sand_obs$_3$, lime_obs$_1$. Almost insignificant for lime_obs$_2$, lime_obs$_3$.	Medium influence for sand_obs1, lime_obs1 Small significance for the reset.
pShale	Most significant for sand_obs$_1$, sand_obs$_2$, lime_obs$_1$. Medium influence for sand_obs$_3$.	Most significant for sand_obs1-lime_obs$_1$. Small influence for lime_obs$_2$, lime_obs$_3$.	Major influence for sand_obs$_1$, sand_obs$_2$, sand_obs$_3$, lime_obs$_1$, small for lime_obs$_2$, lime_obs$_3$
rateSNF	insignificant for all responses	Insignificant for sand_obs$_1$, lime_obs$_2$, lime_obs$_3$. Small influence for sand_obs$_2$, sand_obs$_3$, lime_obs$_1$.	Only significant parameter by linear regression method. Medium influence for lime_obs$_1$ and small for the rest by other methods.
kSand	significant for responses sand_obs$_2$, sand_obs$_3$. Insignificant for the rest.	Insignificant for sand_obs$_1$, lime_obs$_2$, lime_obs$_3$. Medium influence for sand_obs$_2$, sand_obs$_3$, small influence for lime_obs$_1$.	Medium significance for sand_obs$_1$, sand_obs$_2$, sand_obs$_3$. Small influence for lime_obs$_1$, lime_obs$_2$, lime_obs$_3$.
kLime	Major influence for lime_obs$_2$. Insignificant for the rest.	Insignificant for sand_obs$_1$, sand_obs$_2$, sand_obs$_3$. Major influence for lime_obs$_2$, lime_obs$_3$, small influence for lime_obs$_1$.	Major influence for lime_obs$_2$, lime_obs$_3$, minor for the rest.
kLSand	Small significance for all responses.	Insignificant for all responses.	Almost insignificant for all responses.
kBuffer	Minor influence for response lime_obs$_2$, insignificant for the rest	Insignificant for sand_obs$_2$, sand_obs$_3$, lime_obs$_1$. Almost insignificant for sand_obs$_1$, lime_obs$_2$, lime_obs$_3$.	Almost insignificant for all responses.
kDRZ	insignificant for all responses	Insignificant for sand_obs$_2$, lime_obs$_2$, lime_obs$_3$. Small influence for sand_obs$_1$, sand_obs$_3$, lime_obs$_1$.	Almost insignificant for all responses.
sNpKd	insignificant for all responses	Insignificant for all responses.	Almost insignificant for all responses.
bNpKd	Minor significance for sand_obs$_2$, insignificant for the rest	Small significance for sand_obs$_2$, sand_obs$_3$. Insignificant for the rest.	Almost insignificant for all responses.

Table 5-11. SA Results of SNL Shale by IBRAE

5.4.5. Case Summary of Sensitivity Results

Recall that six QoIs were considered for the SNL Shale case. The QoIs were labeled as response functions 1 through 6. The first three response functions were the peak (maximum over time) ^{129}I concentrations at observation points 1, 2, and 3 respectively. These observations points are in the sandstone aquifer and are numbered based on increasing distance from the repository; response function 1 is the peak ^{129}I concentration at the sandstone aquifer observation point closest to the repository and response function 3 is the peak ^{129}I concentration at the sandstone aquifer observation point farthest from the repository. The last three QoIs are analogous, being peak ^{129}I concentration at observations within the limestone aquifer at increasing distances from the repository. SA results are summarized individually by QoI in this section to emphasize comparison between methods. However, some discussion is also included regarding differences in SA results between the different QoIs and how these differences relate to the repository. ^{129}I concentrations at observation points 2 and 5 are not detailed, since these results are consistent with shifting sensitivities between the nearest and farthest observation points relative to the repository.

Multiple participants applied linear regression models and calculated correlation coefficients (GRS, SNL, IBRAE) and CUSUNORO curves (GRS, TUC). Additionally, SNL applied stepwise linear regression and VBD via PCE and GP surrogate modeling, TUC applied Copula distance and MIM, and IBRAE applied PAWN and RBD-FAST. Scaling, rank transformations, and log transformation were also utilized in some of the analyses.

- Response Function 1: Peak ^{129}I at the observation point in the sandstone aquifer closest to the repository

  Methods applied by the participants agreed in ranking *pShale* as the parameter with the highest sensitivity metrics. This observation point is near the repository, meaning transport from the repository to the observation point is predominantly through the shale layer, so sensitivity to shale porosity makes sense. Methods typically ranked *kSand* and *rateWP* as the second and third most important parameters, however it was not always clear if this significance is meaningful. Sensitivity to *kSand* was noted by multiple participants to increase when rank-based methods were applied.

- Response Function 3: Peak ^{129}I at the observation point in the sandstone aquifer farthest from the repository

  This observation point is farther from the repository and this distance is reflected in changes in the SA results compared to observation point 1. Porosity of the shale still has an effect, but significant transport also occurs through the aquifer, so aquifer permeability (*kSand*) gains significance with distance. The relationship between *kSand* and the peak ^{129}I concentration also changes from a negative correlation at observation point 1 to a positive correlation at observation point 3. There was some inconsistency between methods at this observation point, with some (RBD-FAST, PCE and GP surrogates) ranking *pShale* higher than *kSand*, but most ranking *kSand* as the most important parameter.

  As shown in Figure 5-2 a large portion of the simulations result in very low ^{129}I concentrations at this observation point. This may be why methods disagree more on the top-

ranked parameter for this QoI. The SNL sensitivity results from surrogate modeling differ in their ranking of the top parameter and the importance of conjoint effects depending on whether scaling (which maintains high variance in the QoI) or log-transformation (which reduces variance in the QoI) was used. This is consistent with the linear methods, which also attribute much more of the variance to kSand when based on a rank transformation (SPEAR, PRCC) than without a rank transformation.

- Response Function 4: Peak ^{129}I at the observation point in the limestone aquifer closest to the repository

  Most methods ranked *pShale* highest at this observation point, however PCC (from SNL) ranked *kLime* the highest. The linear methods disagreed on the second and third most important parameters, *rateSNF* and *rateWP*. Surrogate model results based on a log-transformation rank *rateWP* as more important, whereas surrogate model results based on scaling rank *rateSNF* higher with more substantial interaction effects. The PAWN results (IBRAE) agree with other methods on ranking *pShale* the most important parameter but ranks the buffer permeability (*kBuffer*) as the second most important parameter.

- Response Function 6: Peak ^{129}I at the observation point in the limestone aquifer farthest from the repository)

  Permeability of the limestone aquifer (*kLime*) is consistently ranked the most important parameter at this observation point, across methods and participants. There is again disagreement on the parameters of secondary importance, with CUSUNORO (GRS), surrogate modeling on scaled concentrations (SNL), and RBD-FAST (IBRAE) identifying *pShale* as the second most important parameter. When surrogate modeling was applied to log-transformed concentrations, no other parameters had significant sensitivity indices.

In general, the sensitivity analyses across participants identified the same first most important parameter but differed on the importance of lower ranked parameters. This may be due to differences between methods in detecting variable significance, particularly because the QoIs in this case span many orders of magnitude. Variables of secondary importance may need to be identified by connecting less clear sensitivity analysis results with physical phenomenology. Results also tended to differ depending on which transformations were applied to the ^{129}I concentrations. Validation or fit metrics, in the context of surrogate modeling or regression analysis, may help select the most appropriate transformation to use for sensitivity analysis.

There is also a lack of consensus between participants on the delineation between sensitivity measure values that indicate secondary sensitivity versus sensitivity measure values that indicate negligible sensitivity. This highlights a need for development of consensus methods for testing or justifying conclusions regarding lower-ranked parameters.

This page is intentionally left blank

6. DESSEL CASE

This case study considers the near field model for the assessment of the long-term safety of a near-surface repository at Dessel (Belgium) for Intermediate and Low-Level radioactive waste. This model was used in the license application file that was submitted on 4 February 2019 by ONDRAF/NIRAS to the FANC, the Belgian regulatory body for Nuclear Control.

The implementation of the near field model is based on the expected evolution scenario (EES). The expected evolution of the disposal facility and its barriers is described in Chapter 14 of the safety report [88], and builds on the phenomenological knowledge basis of the engineered barriers in their environment described in Chapter 5 of the Safety Report [89]. The expected evolution of the disposal system over the relevant timeframes up to 2,000 years from t_0 builds on the characteristics and processes defining the long-term behavior of the engineered barriers and the transport of radionuclides.

The near field model is described in detail in the SCKCEN report ER-0336 [90]. Only the leaching of the ^{129}I radionuclide from conditioned waste in Type I monoliths is considered in this case study.

6.1. Case Description

In this exercise, the degradation of the earth cover and underlying cementitious barriers starts from t_0 + 650 y. This will be the starting point for the radionuclide (RN) leaching simulations and is representative for the part of the disposal facility that degrades earlier than expected.

The EES near field model considers a two-dimensional geometry representing half of a monolith stack as shown in Figure 6-1. In the vertical direction, the simulation domain consists of the:

Structural top slab and precast shielding slabs, combined in a redistributing layer
Stack of 6 type I monoliths
Module support slab
Columns and filled inspection rooms
Foundation slab
Embankment

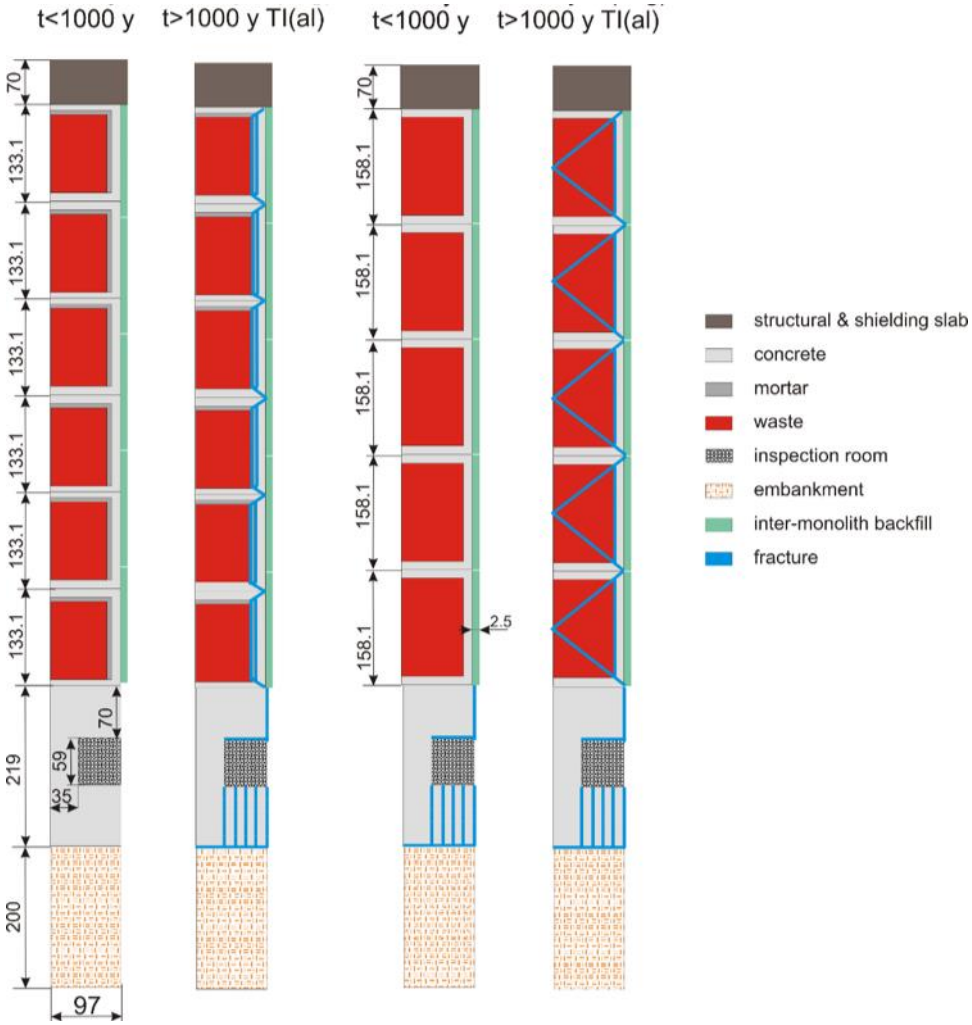

Figure 6-1. Geometry of a modelled domain (dimensions are in cm); Type I monoliths, Colors indicate different material types (backfilled inter-monolith spaces and fractures are denoted in resp. light green and blue color). The geometry is shown for the initial (left), and degraded state (*right*)

Dessel Case	**Description**
Waste form	Steel drums of homogeneously cemented LILW
Engineered barriers	Type I concrete monoliths (with backfill mortar),
Repository description	Near surface disposal facility
Natural system	Not considered
Biosphere	Not considered
Conceptual release pathways	Diffusion in concrete and mortar, advective transport in fracture and conductive sorbing media (inspection room & embankment)
Processes modeled	Advective-dispersive/diffusive transport
Software codes used	Comsol Multiphysics 3.5a
Reference for full description of case	[90]

Table 6-1. Dessel Case Description

6.2. Description of Inputs and Outputs

Twenty-two input parameters with uniform or log-uniform distributions. The ranges are given in Table 6-2 and Table 6-3, either for general or radionuclide specific parameters.

Based on the parameter ranges given in previous tables, 256 / 1024 parameters combinations were generated using the Matlab Sobolset method, with MatousekAffineOwen scrambling. The first 2^{22} points were skipped to avoid a poor burn-in of the Sobol' sequence. All evaluations presented here are based on the larger dataset. Comparisons for different sample sizes have been undertaken but are not reported here.

To compute a benchmark solution for the sensitivity indices, in particular the higher order, a larger dataset was generated for an analysis based on Sobol's method. The 24,000 samples were generated and analyzed with Matlab scripts based on the methods described in [91], which were obtained from https://ec.europa.eu/jrc/en/samo.

parameter	unit	min	max	distribution	EES value (BE)
drain_bio	mm/a	250	480	Uniform	480
flux_increase_time	a	30	700	Uniform	350
transmissivity_ratio	-	0.01	1	loguniform	0.04
w_fracture	mm	0.1	1	Uniform	0.3
D_concrete_caisson	m^2/s	5.14E-14	4.43E-10	loguniform	7E-11
D_concrete	m^2/s	5.14E-14	4.43E-10	loguniform	7E-11
D_mortar	m^2/s	1.85E-12	3.21E-10	loguniform	2E-11
D_waste	m^2/s	1.85E-12	6E-10	loguniform	6E-10
D_inspectionroom	m^2/s	5E-10	8E-10	uniform	8E-10
D_embankment	m^2/s	2E-11	1E-10	uniform	6E-11
poro_concrete	-	0.07	0.12	uniform	0.1
poro_mortar	-	0.07	0.12	uniform	0.1
poro_waste		0.2	0.3	uniform	0.25
poro_inspectionroom	-	0.4	0.6	uniform	0.55
poro_embankment		0.29	0.35	uniform	0.33
alpha_L_all	m	0.01	1	loguniform	0.1

Table 6-2. Used parameter ranges, distributions, and best estimate (BE)

parameter	unit	min	max	distribution	EES value (BE)
Kd_eff_I_con_caisson	L/kg	1E-5	57	log uniform	0.19
Kd_eff_I_con_module	L/kg	1E-5	57	log uniform	0.19
Kd_eff_I_mortar	L/kg	1E-5	290	log uniform	2.9
Kd_eff_I_waste	L/kg	1E-5	100	log uniform	1
Kd_eff_I_insp	L/kg	1E-5	57	log uniform	0.19
Kd_eff_I_emb	L/kg	1E-5	50	log uniform	0.5

Table 6-3. RN-specific (^{129}I) sorption parameter ranges, distributions and best estimate (BE)

Output Quantity	Type	Other
Activity flux (Bq/a)	Time series (200 values between 0 - 2000 years)	^{129}I activity flux out of the embankment to the geosphere

Table 6-4. Dessel Case Output QoIs

6.3. Salient features and behavior of the model

The statistics on the time-dependent output (^{129}I flux to the geosphere) are calculated and plotted in Figure 6-2. It shows that peak fluxes are occurring during the ~700 to 800-year time frame. Here, the uncertainty on the peak fluxes is the highest. Differences between extreme percentile values and the mean of the activity flux range over more than one order of magnitude, which could indicate a non-linear model response.

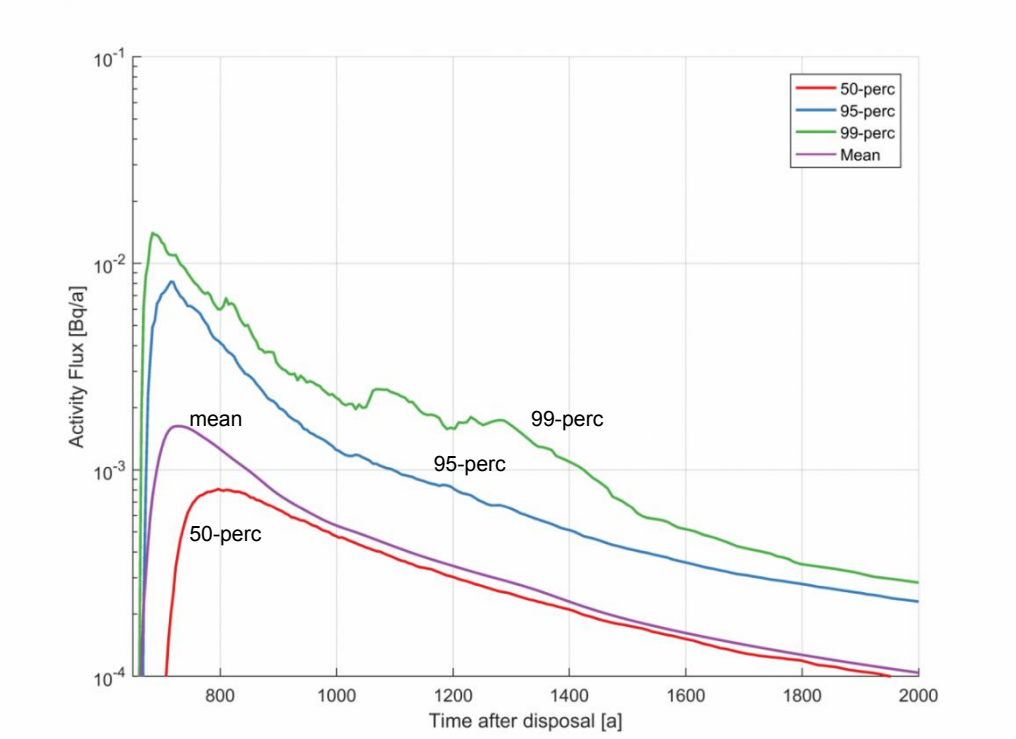

Figure 6-2. Uncertainty on the time-dependent ^{129}I radionuclide flux based on 1024 sample runs

6.4. Sensitivity Analysis Results

6.4.1. Results from GRS

GRS has analyzed the Dessel model with a number of sensitivity analysis methods, as shown in Table 6-5, for all points in time. Some results are compiled below.

Dessel model	**Description**
Sensitivity Analysis Method	CUSUNORO, Linear regression, rank correlation, rank regression, EASI
Sensitivity measures generated	SRC, SRRC, S_i
Special considerations	-
Surrogate models used	-
Transformations	Shifted log-transformation
QoIs addressed	^{129}I outflow
Number of samples used	1024

Table 6-5. Sensitivity analysis of Dessel model by GRS

Some CUSUNORO curves are presented in Figure 6-3. The following conclusions can be drawn from this analysis:

- Only 3 parameters have relevant influences at later times: *Kd_waste*, *Kd_emb* and *D_waste2*.
- There is no relevant influence of *Kd_waste* below the 10^{-2} fraction and of *Kd_emb* below the 10^{-1} fraction. This is understandable as very little sorption is as good as no sorption.
- *Kd_waste* has a mainly negative influence over most of the time. This is understandable as increasing K_d means decreasing outflow.
- At later times, *Kd_waste* has a positive influence at a small value interval between fractions 10^{-2} and 10^{-1}, for high values, however, the influence is again negative. This might be due to retarded release.
- Except from the earliest times, *Kd_emb* has a relevant influence only at values above 5; this influence is positive, but for medium times (around 1000 years) it changes to negative for the highest parameter values (> ~20). Little sorption is as good as no sorption, but it is not directly understandable why increasing K_d leads to increasing outflow here.
- *D_waste2* changes its direction of influence from positive to negative with time around 1400 years.

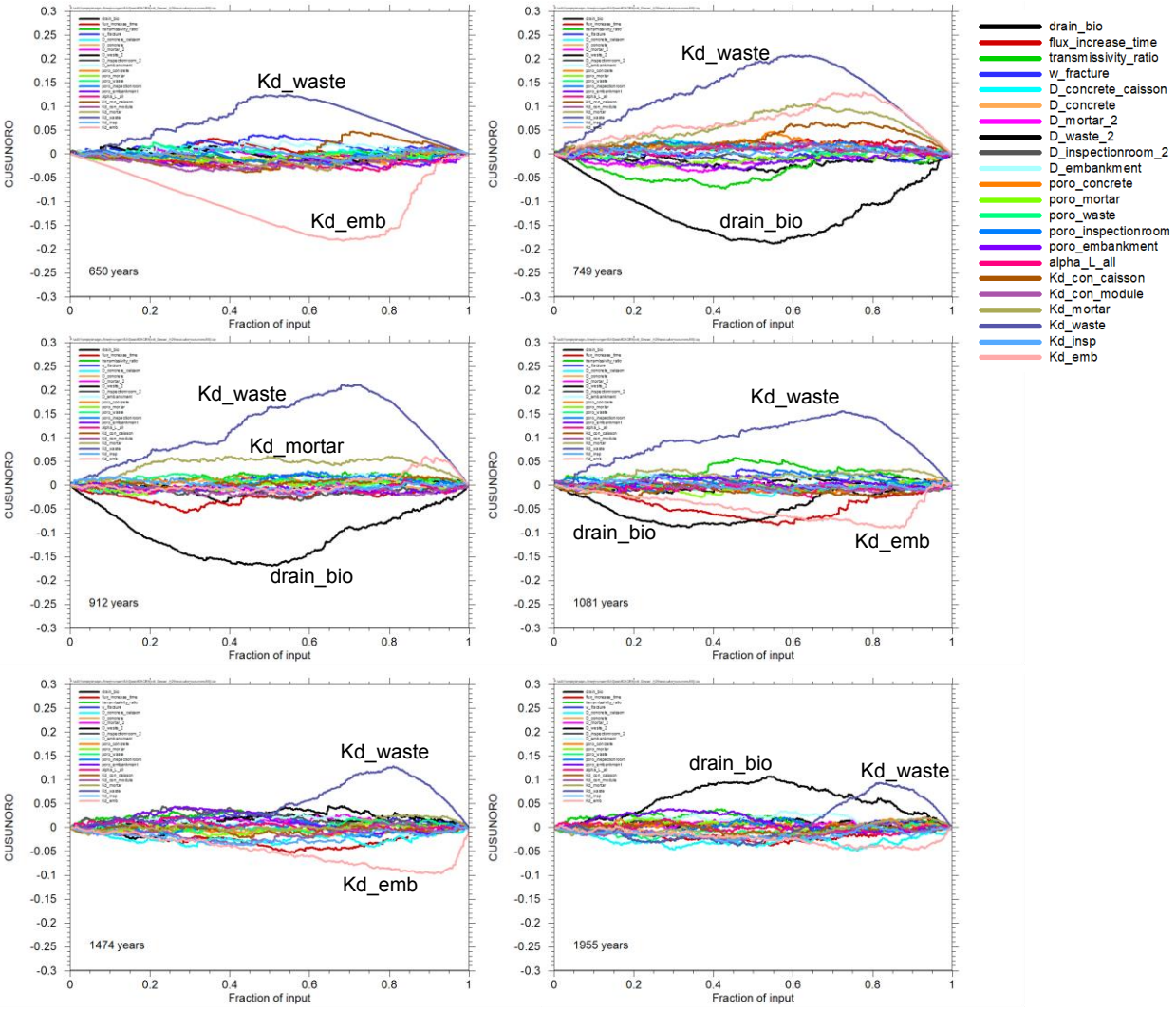

Figure 6-3. CUSUNORO analysis of Dessel model by GRS

In Table 6-6, the results of the regression-based as well as the variance-based analysis are presented. Correlation-based methods were not applied as, at least for uncorrelated input, its results can be expected to be very similar to those of the regression-based analysis and not to provide additional information. Instead, the regression analysis was additionally performed after applying the transformation $y \mapsto \log_2(1 + y/a)$. For high values ($y >> a$), this transformation is practically a log transformation, but unlike the latter, it avoids overweighting of low and very low values and even allows zeros. The parameter a is mapped to 1 and represents the "threshold" between low and high values. The transformation has a similar effect as the rank transformation as it mitigates the influences of very high values, but it is reversible and keeps all information.

Table 6-6 also contains the R^2 values for the regression investigations and the sum of all 1st order indices for the EASI investigation.

Input parameter	**Regression SRC**		**Rank Regression SRRC**		**SRC with transformation (a=10^{-4})**		**SRC with transformation (a=10^{-7})**		**EASI S_i**	
	797 yrs	**1401 yrs**	**797 yrs**	**1401 yrs**	**797 yrs**	**1401 yrs**	**797 yrs**	**1401 yrs**	**797 yrs**	**1401 yrs**
drain_bio	0.013	-0.034	0.014	-0.011	0.045	-0.030	0.050	-0.027	0.008	0.014
flux_increase_time	0.013	0.072	-0.102	0.081	-0.080	0.073	-0.136	0.057	0.017	0.022
transmissivity_ratio	-0.019	-0.015	0.063	-0.047	0.001	-0.021	0.020	0.001	0.024	0.014
w_fracture	-0.008	-0.024	-0.015	-0.027	-0.018	-0.033	-0.024	-0.051	0.013	0.020
D_concrete_caisson	-0.003	0.043	0.009	0.090	0.011	0.077	0.003	0.059	0.010	0.013
D_concrete	0.039	-0.020	0.018	0.021	0.034	-0.014	0.034	-0.028	0.009	0.006
D_mortar_2	0.018	-0.017	0.010	0.002	0.006	0.001	-0.010	-0.001	0.016	0.012
D_waste_2	0.324	0.024	0.324	-0.078	0.222	-0.095	0.140	-0.154	0.167	0.020
D_inspectionroom_2	-0.007	-0.063	0.020	-0.017	-0.005	-0.037	0.003	-0.018	0.010	0.017
D_embankment	0.010	-0.031	-0.014	-0.037	0.005	-0.022	0.011	-0.008	0.012	0.014
poro_concrete	-0.024	0.014	-0.048	-0.007	-0.039	0.006	-0.051	0.021	0.015	0.008
poro_mortar	0.020	0.019	0.014	0.004	0.011	0.002	0.001	-0.019	0.015	0.008
poro_waste	-0.027	-0.006	0.001	0.022	-0.022	0.009	-0.023	0.022	0.013	0.010
poro_inspectionroom	0.024	-0.009	-0.034	-0.013	-0.015	-0.012	-0.003	-0.024	0.011	0.019
poro_embankment	0.057	-0.061	0.003	-0.065	0.019	-0.067	-0.012	-0.054	0.014	0.022
alpha_L_all	0.000	-0.016	0.021	0.000	0.017	0.005	0.058	0.026	0.009	0.007
Kd_con_caisson	-0.103	-0.021	-0.083	0.034	-0.102	0.032	-0.065	0.066	0.022	0.015
Kd_con_module	-0.001	0.002	-0.001	0.051	-0.014	0.023	-0.004	0.052	0.022	0.008
Kd_mortar	-0.150	-0.060	-0.158	-0.003	-0.135	-0.040	-0.091	-0.014	0.063	0.021
Kd_waste	-0.286	-0.234	-0.485	-0.233	-0.451	-0.372	-0.392	-0.373	0.344	0.141
Kd_insp	-0.042	0.045	0.019	0.051	-0.050	0.040	-0.046	0.052	0.010	0.008
Kd_emb	-0.232	0.459	-0.261	0.103	-0.448	0.275	-0.687	0.158	0.158	0.134
r^2 / 1st order sum	0.298	0.300	0.510	0.101	0.527	0.266	0.721	0.228	0.982	0.553

Table 6-6. SA Results of Dessel model by GRS. The three leading parameters are marked in red, blue and green

The ranking of parameters does not seem very unique, but *Kd_waste*, *Kd_emb* and *D_waste_2* appear most often on the leading ranks. Several parameters switch their direction of influence between the early and the late point in time.

It can be seen that the R^2 value is increased by rank transformation as well as by the shifted log-transformation, but only at the early point in time. At the late point in time, it is decreased by all investigated types of transformation, even by a factor of three if rank transformation is performed. This might be a hint that there is some randomness in the small model output values at the late point in time.

Figure 6-4 presents a comparison between SRC and EASI analysis. The results agree qualitatively. One can see that:

D_waste2 changes its direction of influence from positive to negative with time around 1400 years (see CUSUNORO analysis).
Kd_emb changes its dominating direction of influence from negative to positive around 1000 years.

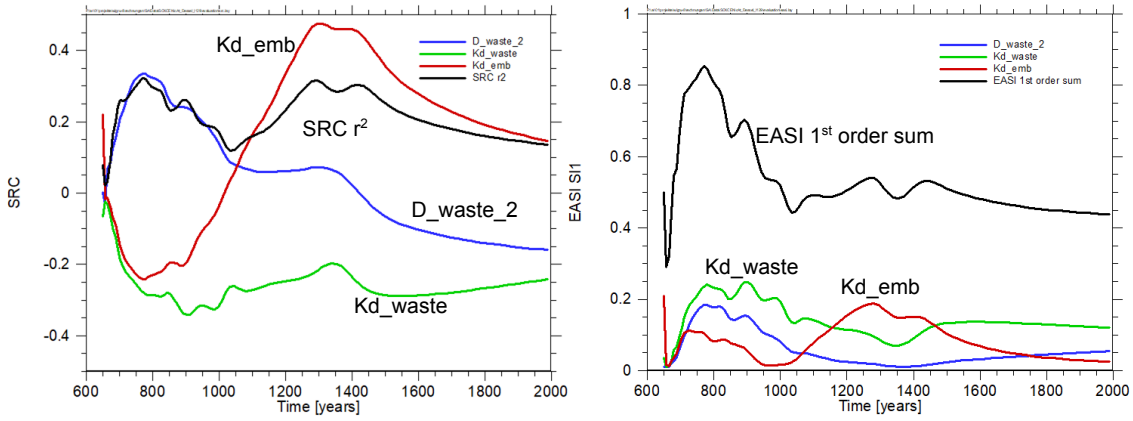

Figure 6-4. SRC (left) and EASI (right) analysis of the Dessel model

Figure 6-5 shows the effect of the transformation with different values of the transformation parameter *a*. This value separates "small" model output values (which are left more or less as they are) from "high" ones (which are practically logarithmized). The transformation smooths the curves and makes them more pronounced, R^2 is increased.

A transformation parameter of $a = 10^{-7}$ is clearly below the modus of the distribution of model output values, except for very early times. In effect, this transformation pronounces the differences between values in the sensitivity evaluation, compared to non-transformed data or data transformed with a higher *a* value. At the same time, it mitigates the differences between very high values. As the curve of *Kd_emb* seems most sensitive to the transformation parameter, one can conclude that its range of primary influence changes over time. At early times the influence of *Kd_emb* concerns mainly the medium values, at later times the higher values.

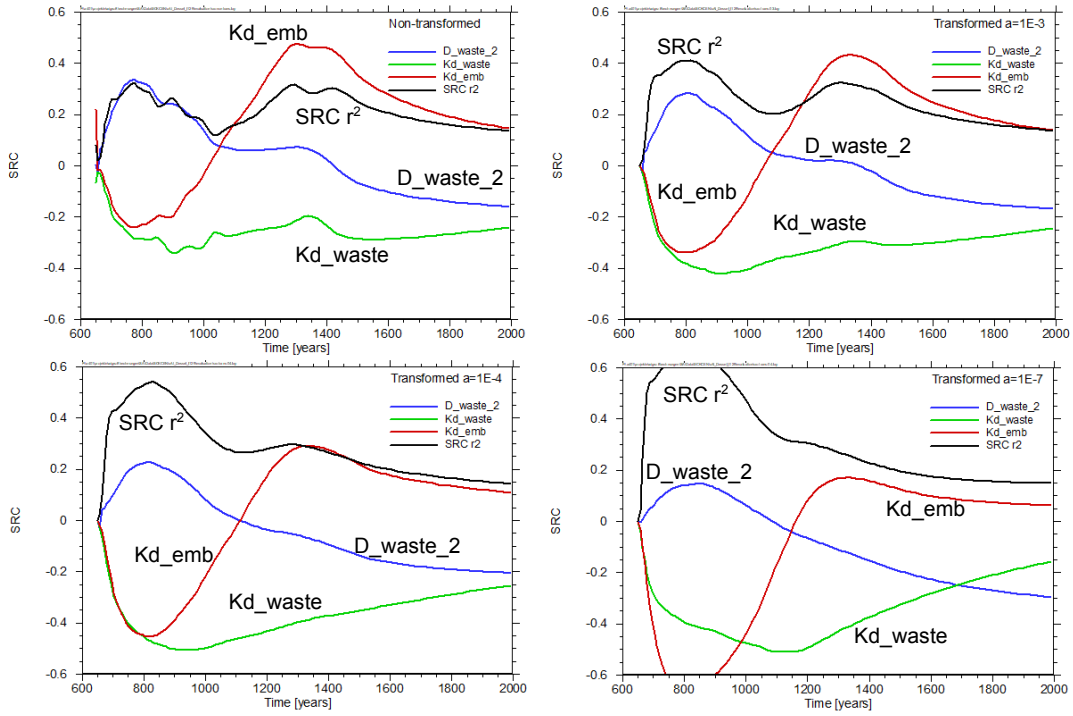

Figure 6-5. SRC analysis of the Dessel model with transformation

6.4.2. Results from SCK

SCKCEN has analyzed the Dessel model with a number of sensitivity analysis methods for all points in time, see Table 6-7. A selection of results is presented in Table 6-8 below.

Case Name	**Description**
Sensitivity Analysis Method	PEAR, SPEAR, EASI, Sobol' Method
Sensitivity measures generated	S_i, T_i, Pearson CC, Spearman RCC
Special considerations	
Surrogate models used	GP, PCE (order 2), DNN
Transformations	
QoIs addressed	Activity flux
Number of samples used	1024

Table 6-7. Sensitivity analysis of Dessel reference case by SCKCEN

Input parameter	SPEAR/EASI	Total order T_i (Sobol - 24k)
Kd_emb	Dominating 1200-1500 yrs	Important (800-1000 yrs)
Flux increase time	Dominating (<750 yrs)	Important (800-1000 yrs)
Kd_waste	Dominating 750-1200 yrs, 1500 yrs-	
D_waste_2	Moderate importance around 800 yrs	increasing importance towards the end of time range
Kd_con_module	Moderate importance in very early time frame (650-700 yrs)	
alpha_L	Moderate importance in very early time frame (650-700 yrs)	

Table 6-8. Summary of SA Results of Dessel Case by SCKCEN

6.4.2.1. Standard SA Techniques

In Figure 6-6 and Figure 6-7, the results of the regression-based (SRRC) as well as the variance-based analysis (EASI) are presented. Results of the correlation-based methods are not reported, as the results are very similar to those of the regression-based analysis. The six most influential parameters are chosen based on the ranking of their maximum SRRC across the time series. The same six parameters are plotted in the variance-based analysis, although the ranking might differ.

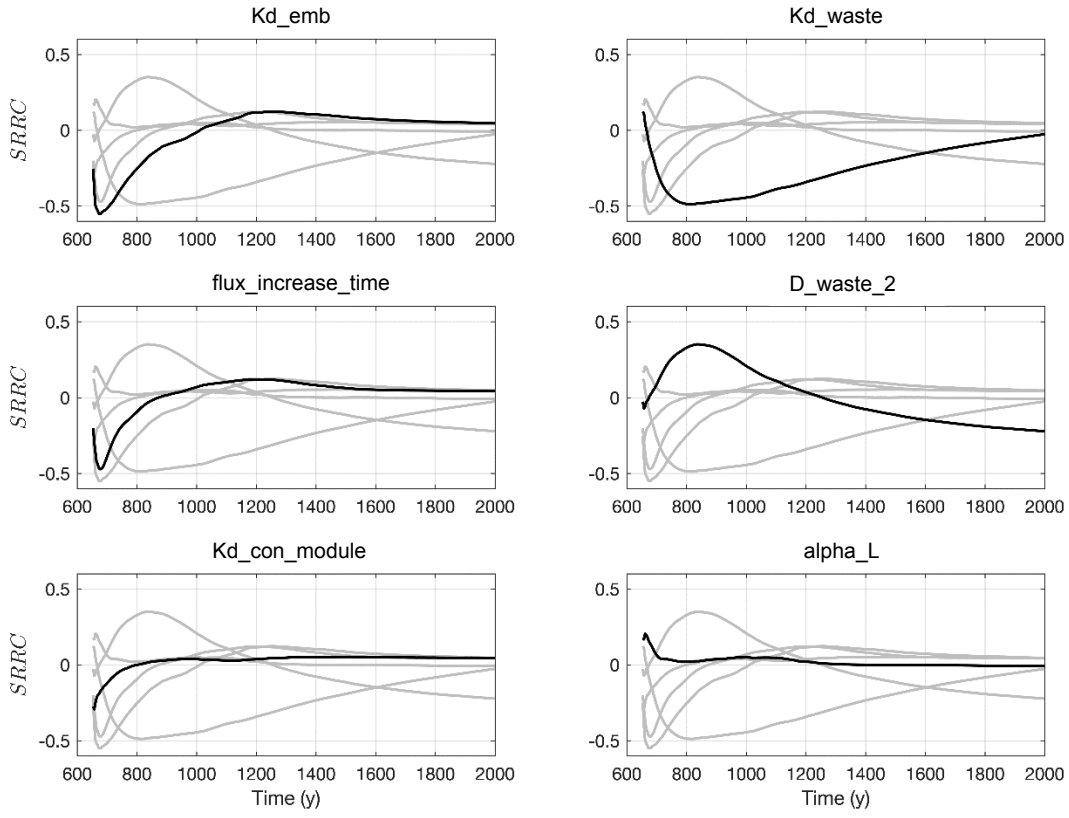

Figure 6-6. SRRC sensitivity results over time from the SCKCEN analysis of the Dessel case. Each axis in the plot highlights one of the independent variables

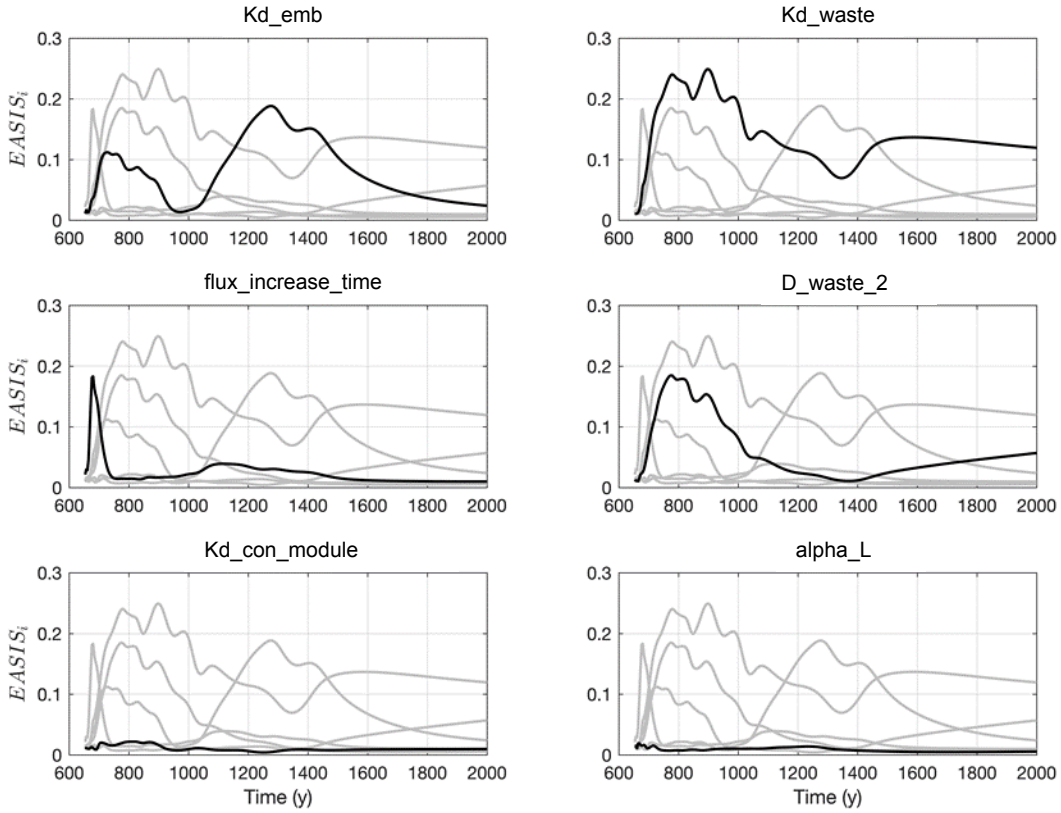

Figure 6-7. EASI first order sensitivity index over time from the SCKCEN analysis of the Dessel case. Each axis in the plot highlights one of the independent variables

6.4.2.2. Surrogate Model-Based Techniques

Apart from classical SA techniques, efforts were made to use state-of-the-art metamodelling techniques in order to compute first order- and total-effects.

The true first-order (S_i) and total-order (T_i) Sobol' indices were computed using the classical approach described by Saltelli et al. [91] and a total of N_{total} = 24,000 forward model runs ($N = 1000$ base samples, k = 22 parameters, $N_{total} = N * (k + 2)$).

The considered surrogate methods are polynomial chaos expansion (PCE), Gaussian processes (GP) and fully connected deep neural networks (DNN). The total number of forward runs used to construct the surrogate models are either 256 or 1024.

For the PCE surrogates, the chaospy toolbox was used (https://chaospy.readthedocs.io/en/master/). Both standard and sparse PCEs were tested, up to order 3 expansion. The PCE coefficients were fitted by least-square regression (point collocation method). The GPs were built using the GPflow package (https://gpflow.readthedocs.io/en/master/notebooks/intro.html). The selected GPs relied on an anisotropic exponential kernel a linear mean function, of which the parameters were estimated by marginal likelihood optimization). Cross-validation was used to select the optimal PCE expansion order and GP hyperparameters. The DNN surrogates were constructed using the

Keras interface (https://keras.io/) to Tensorflow (https://www.tensorflow.org). The DNNs had 3 hidden layers of 64, 128 and 256 neurons with a rectified linear unit (Relu) activation function and an output layer of 22 neurons with linear activation. During training of the DNNs, a l2 regularization of the DNN weights was used in addition to a mean square loss for data fitting.

The results of the different metamodeling techniques are in good agreement with each other. However, for the total-order sensitivity indices, some important interactions are not captured. This is discussed later.

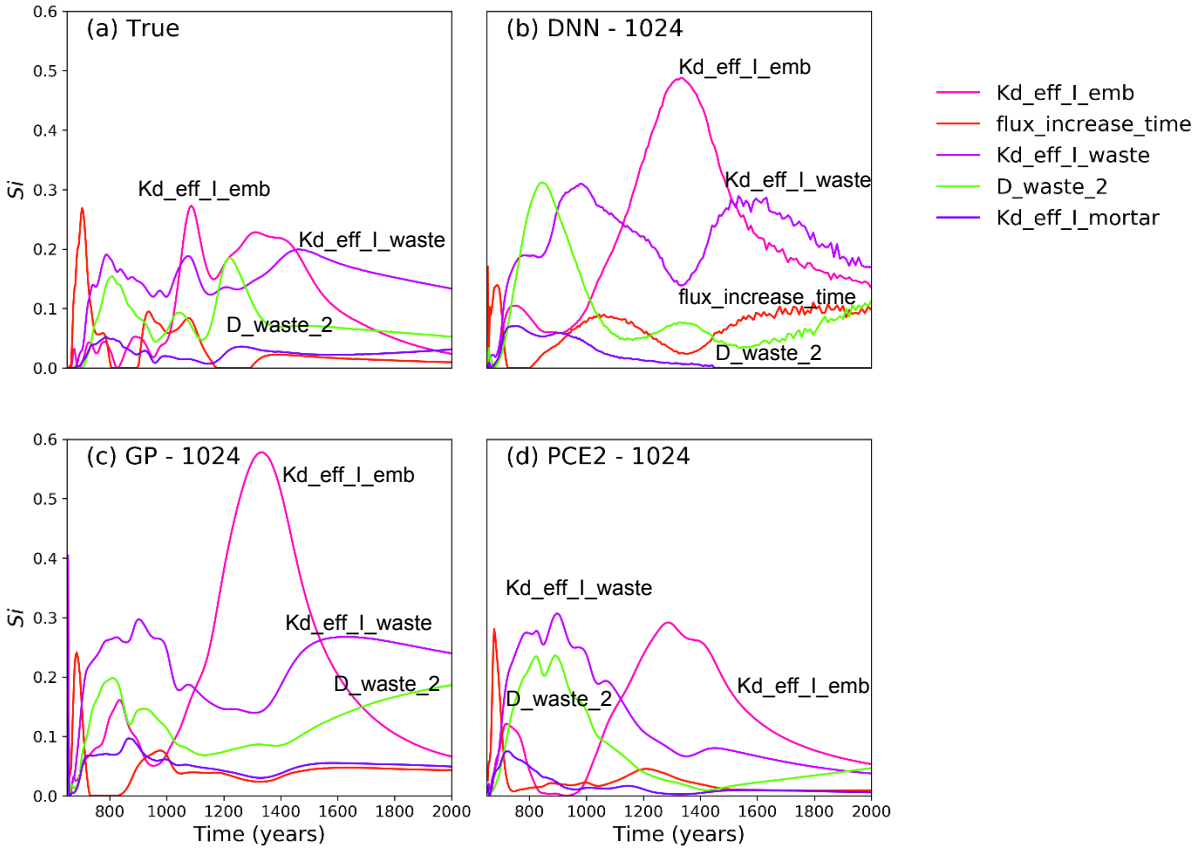

Figure 6-8. Results of first order sensitivity indices calculated by either the Sobol' method with 24,000 direct simulations ('True'), or by different metamodeling techniques based on 1024 sample - dataset

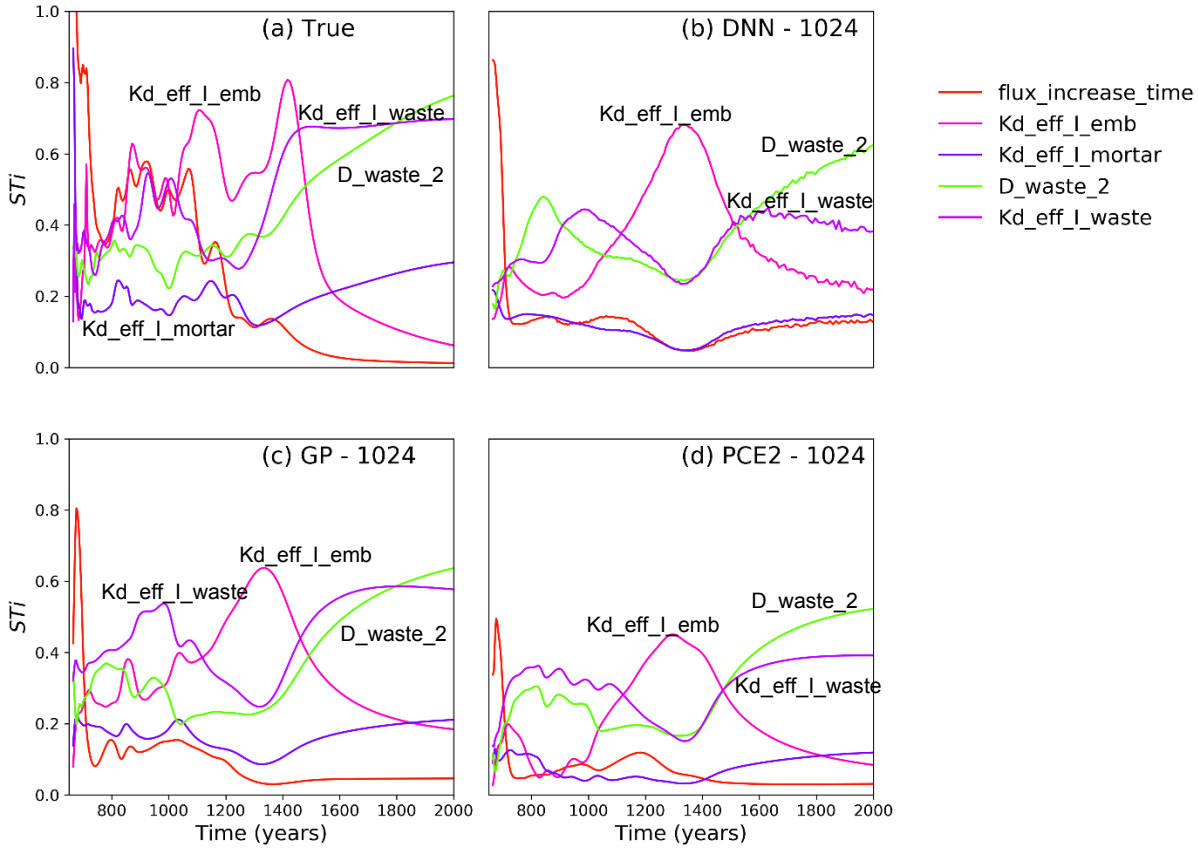

Figure 6-9. Results of total order sensitivity indices calculated by either the Sobol' method with 24,000 direct simulations ('True'), or by different metamodeling techniques based on 1024 sample dataset

6.4.3. Results from SNL

Case Name	Description
Sensitivity Analysis Method	Surrogate modeling, PEAR, SPEAR
Sensitivity measures generated	S_i, T_i, PEAR, SPEAR
Special considerations	
Surrogate models used	GP, PCE (order 2)
Transformations	
QoIs addressed	Activity flux
Number of samples used	256, 1024 (for comparison)

Table 6-9. Sensitivity analysis of Dessel reference case by SNL

Sensitivity measures from the different linear methods applied by SNL are plotted in Figure 6-10 through Figure 6-13. Sobol' indices calculated using an order 2 surrogate are plotted in Figure 6-14 and Figure 6-15. All of these analyses were performed using 256 samples and 1024 samples and it was determined that using 1024 samples reduces noise in correlation coefficients. Application of rank transformation further reduces noise in the correlation coefficients. Simple

rank (SPEAR, Figure 6-11) and partial rank correlation coefficients (PRCC, Figure 6-13) have nearly identical values. 2nd order PCE identifies the same important variables as other methods but indicates different timing of importance and possible importance of interactions.

The Gaussian process surrogate model did not appear to be appropriate for this dataset (see Figure 6-16).

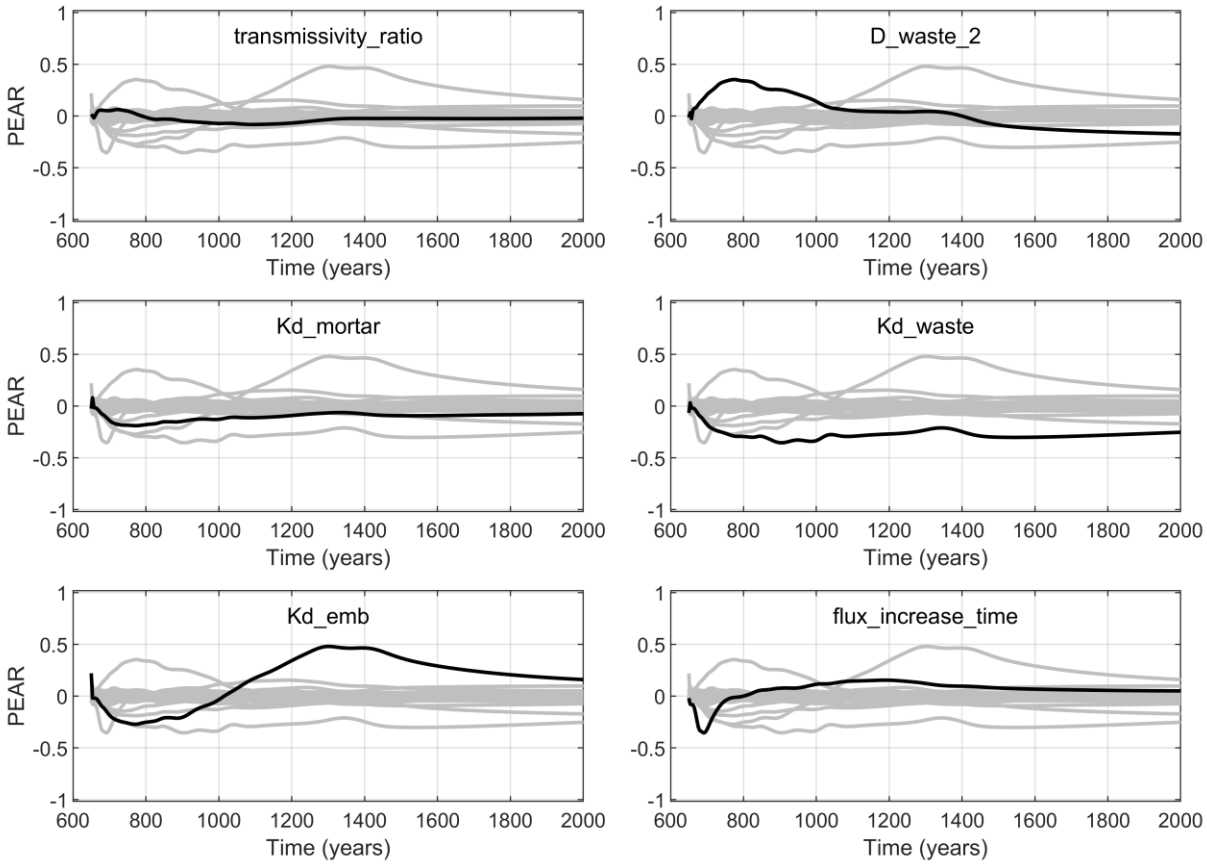

Figure 6-10. PEAR sensitivity results over time from the SNL analysis of the Dessel case. Each axis in the plot highlights one of the independent variables

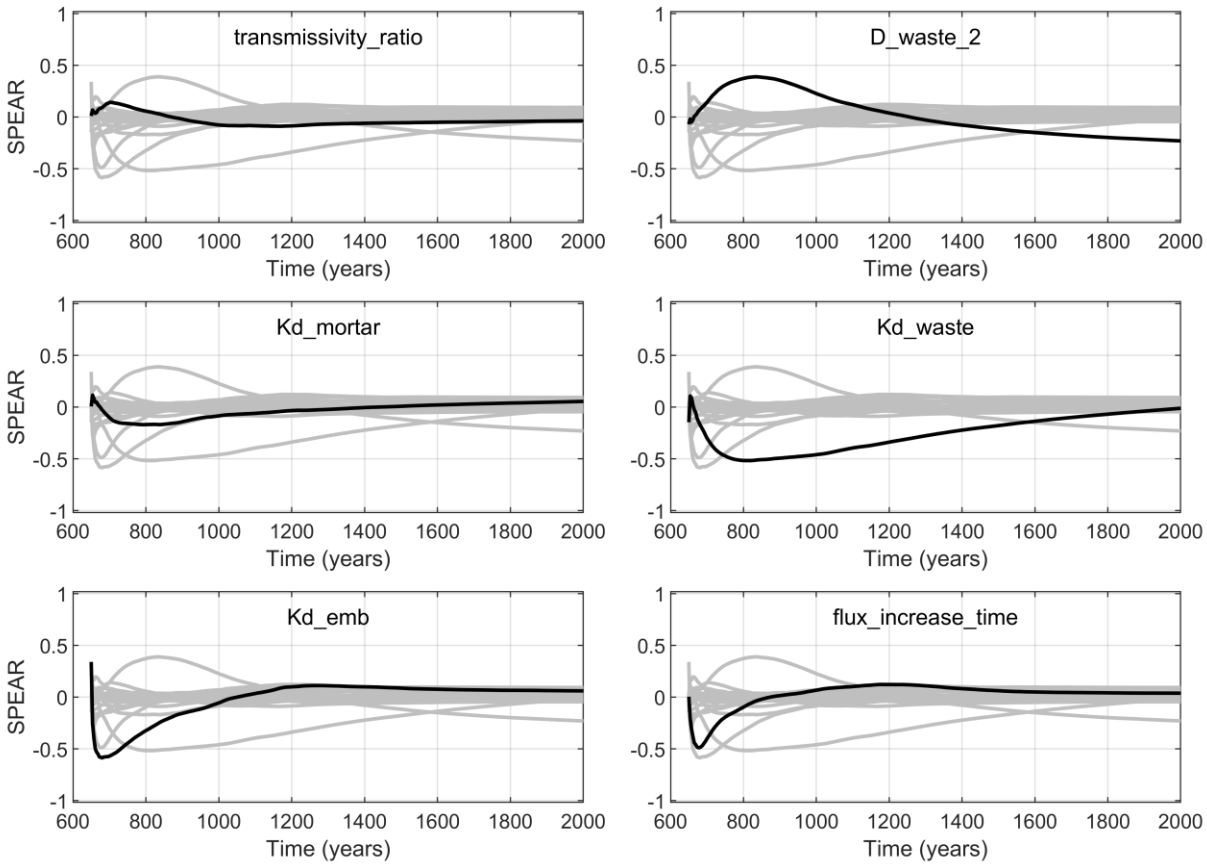

Figure 6-11. SPEAR sensitivity results over time from the SNL analysis of the Dessel case. Each axis in the plot highlights one of the independent variables

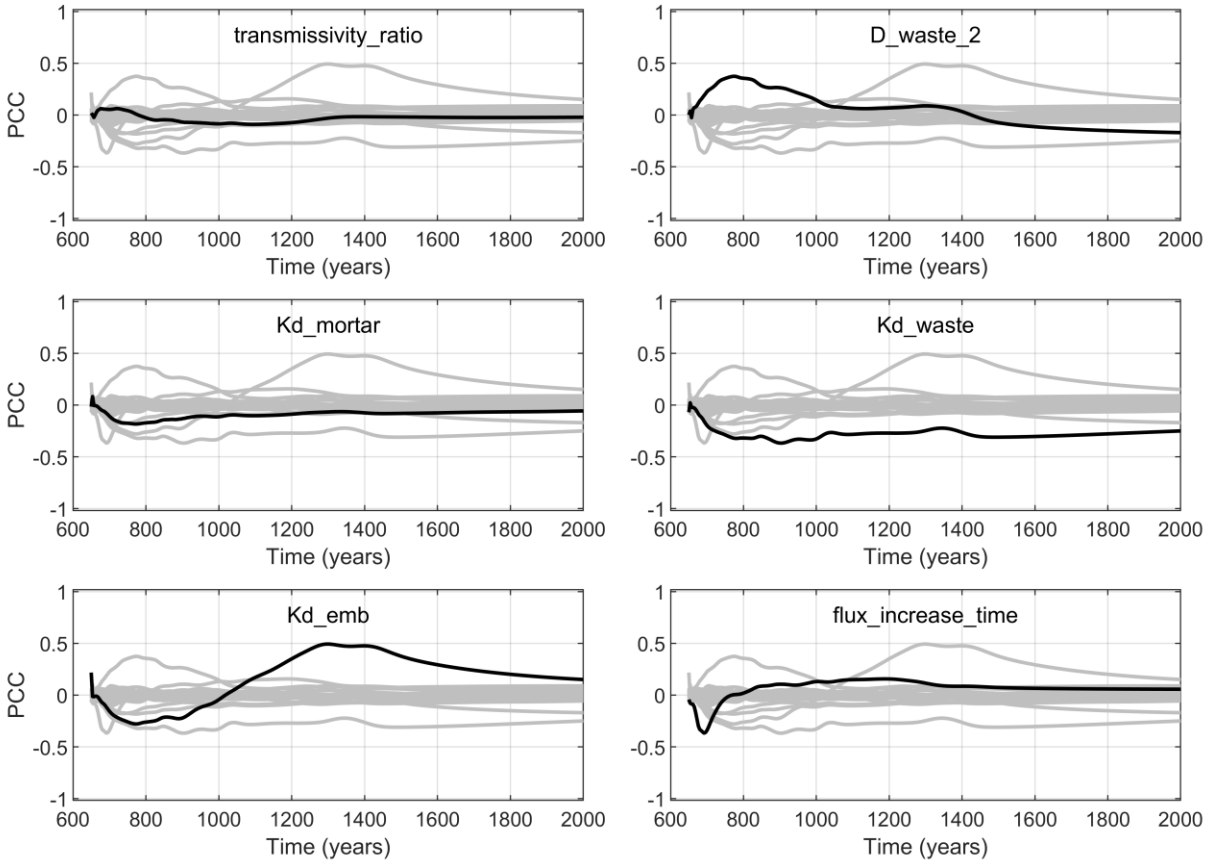

Figure 6-12. PCC sensitivity results over time from the SNL analysis of the Dessel case. Each axis in the plot highlights one of the independent variables

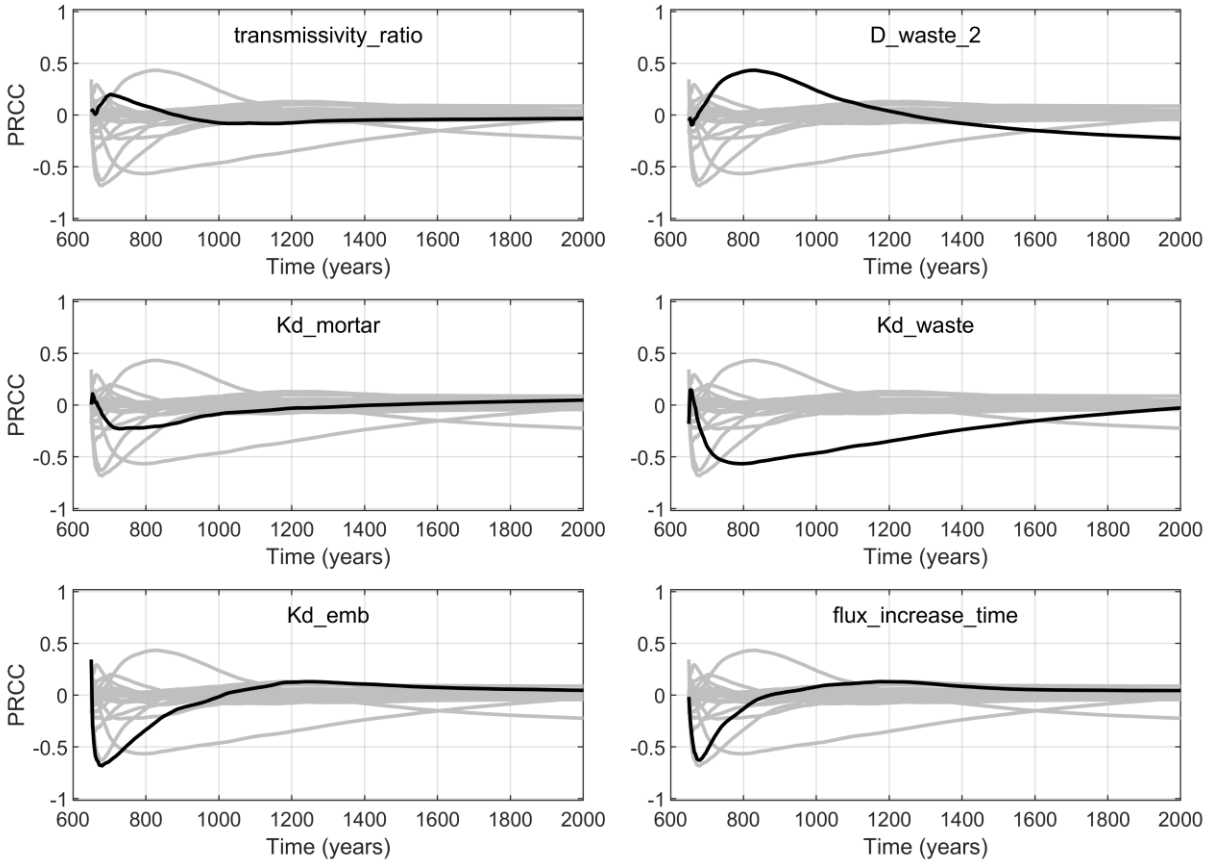

Figure 6-13. PRCC sensitivity results over time from the SNL analysis of the Dessel case. Each axis in the plot highlights one of the independent variables

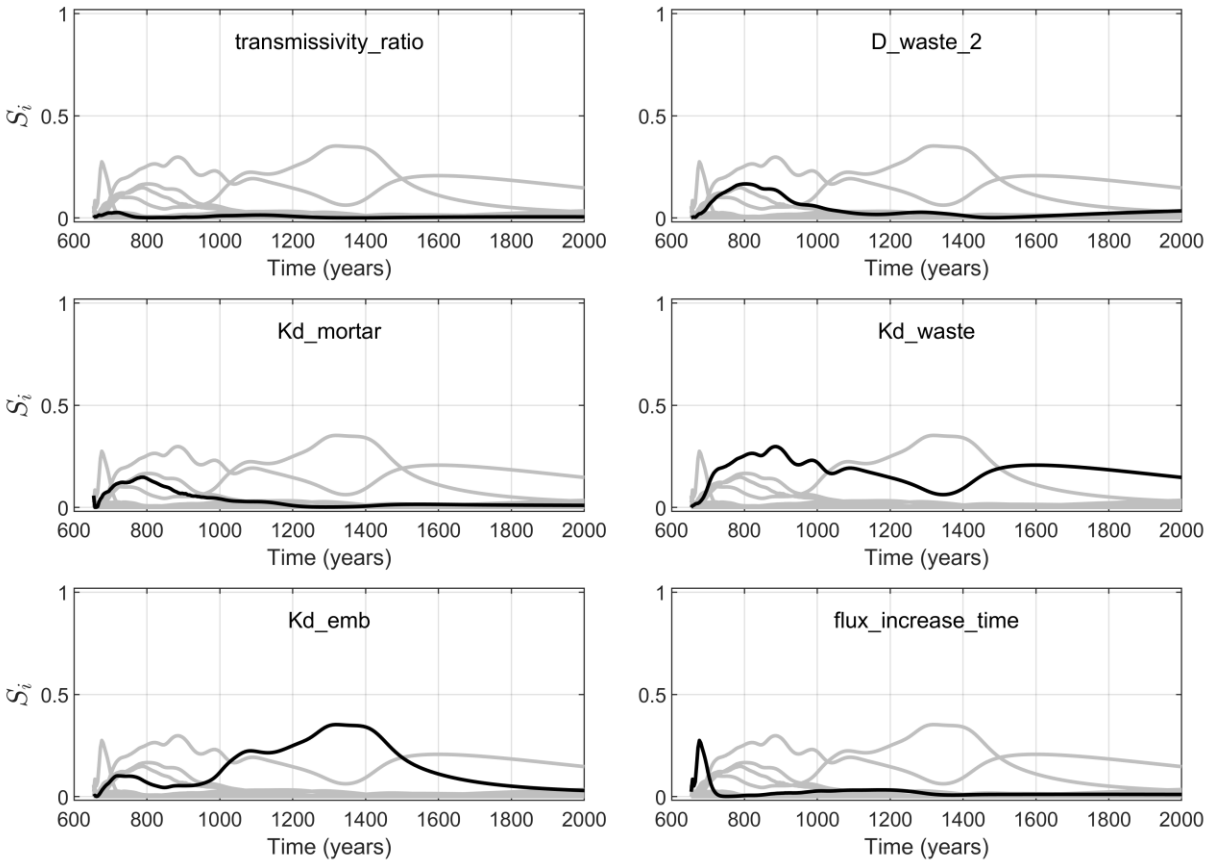

Figure 6-14. Main Sobol' index results from sensitivity analysis of the Dessel reference case by SNL estimated using a second order PCE surrogate. Each axis in the plot highlights one of the independent variables

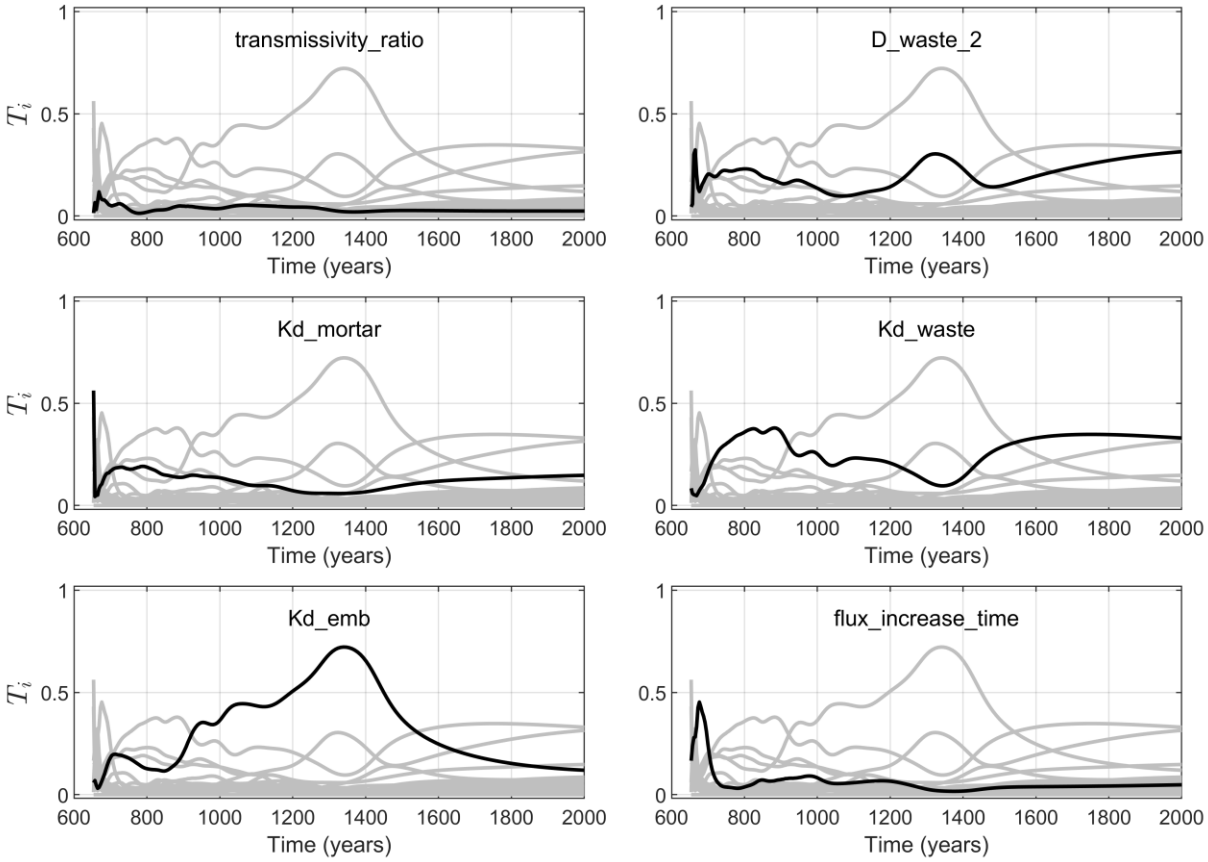

Figure 6-15. Total Sobol' index results from sensitivity analysis of the Dessel reference case by SNL estimated using a second order PCE surrogate. Each axis in the plot highlights one of the independent variables

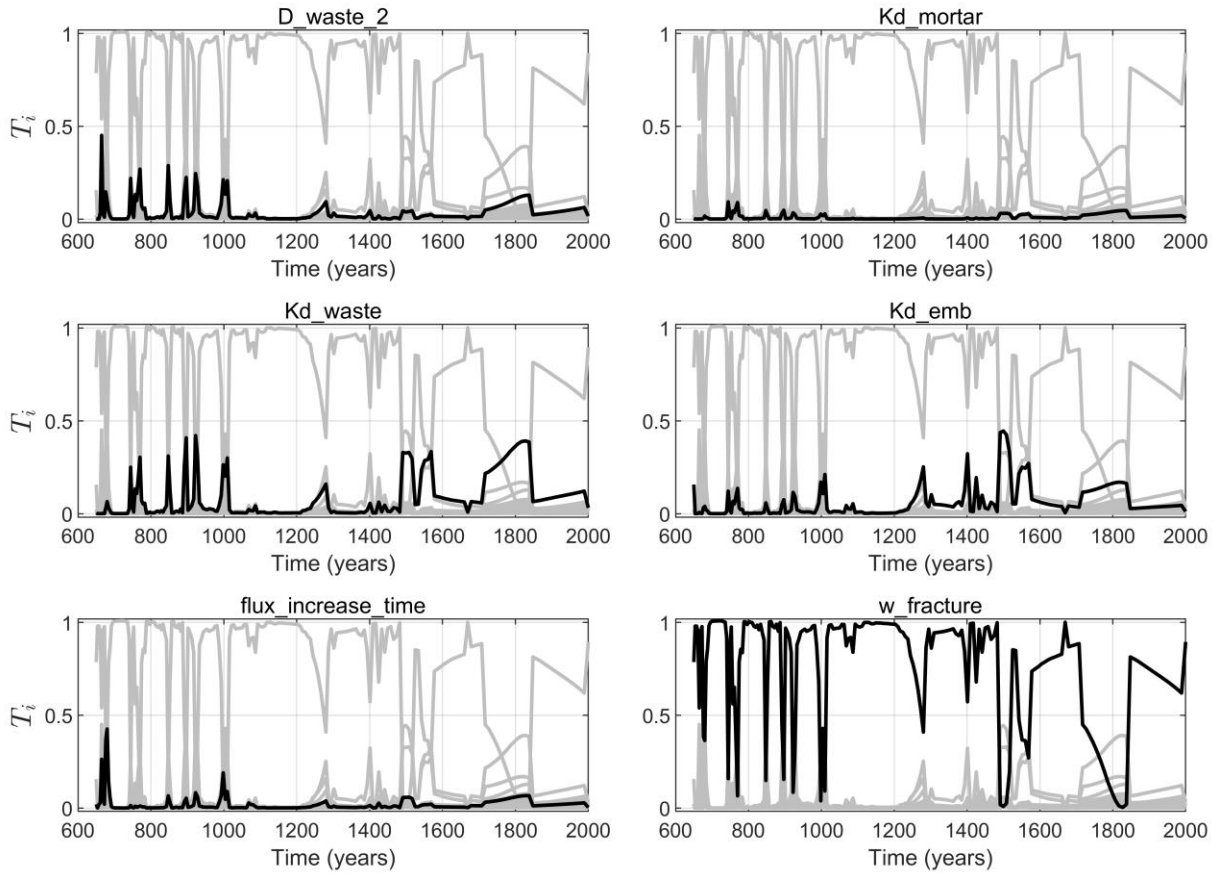

Figure 6-16. Total Sobol' index results from sensitivity analysis of the Dessel reference case by SNL estimated using a GP surrogate. Each axis in the plot highlights one of the independent variables

6.4.4. Results from TUC

The sensitivity analysis methods used by TUC are shown in Table 6-10 with the results listed in Table 6-11.

Case Name	**Description**
Sensitivity Analysis Method	CUSUNORO, EASI, PCE, MIM
Sensitivity measures generated	S_i first order, S_{i2} second order, CvM
Special considerations	Polynomial chaos uses harmonic base functions
Surrogate models used	
Transformations	Second order finite difference for interactions
QoIs addressed	Flux
Number of samples used	1024 (largest available)
Dataset	SA_TI_I129
Software	MatLab/Octave: in-house software

Table 6-10. Sensitivity analysis of Dessel Case by TU Clausthal

Input parameter	CUSUNORO	2nd order S_{ij}	Cramér / von Mises
	No major contributor		
Parameter 22 *Kd_emb*	Instant release, dominating 1200-1500 yrs	Interaction with *Kd_waste* and *D_waste_2*	Dominating <800 yrs, minor >800 yrs
Parameter 2 *Flux increase time*	Dominating (<750 yrs)		Minor <800 yrs
Parameter 20 *Kd_waste*	Dominating 750-1200 yrs, 1500 yrs-	*	Dominating >800 yrs
Parameter 8 *D_waste_2*	Minor	*	Minor >800 yrs
Parameter 19 *Kd_mortar*	Small (700-1000 yrs)		
Parameter 3 *Transmissivity ratio*	Small (<850 yrs)		

Table 6-11. SA Results of Dessel Case by TU Clausthal

Figure 6-17 shows the CUSUNORO curve at timestep $t = 1004$ years. Despite the large input dimension, only few of the input parameters are outside the significance ellipsoid for this sample size.

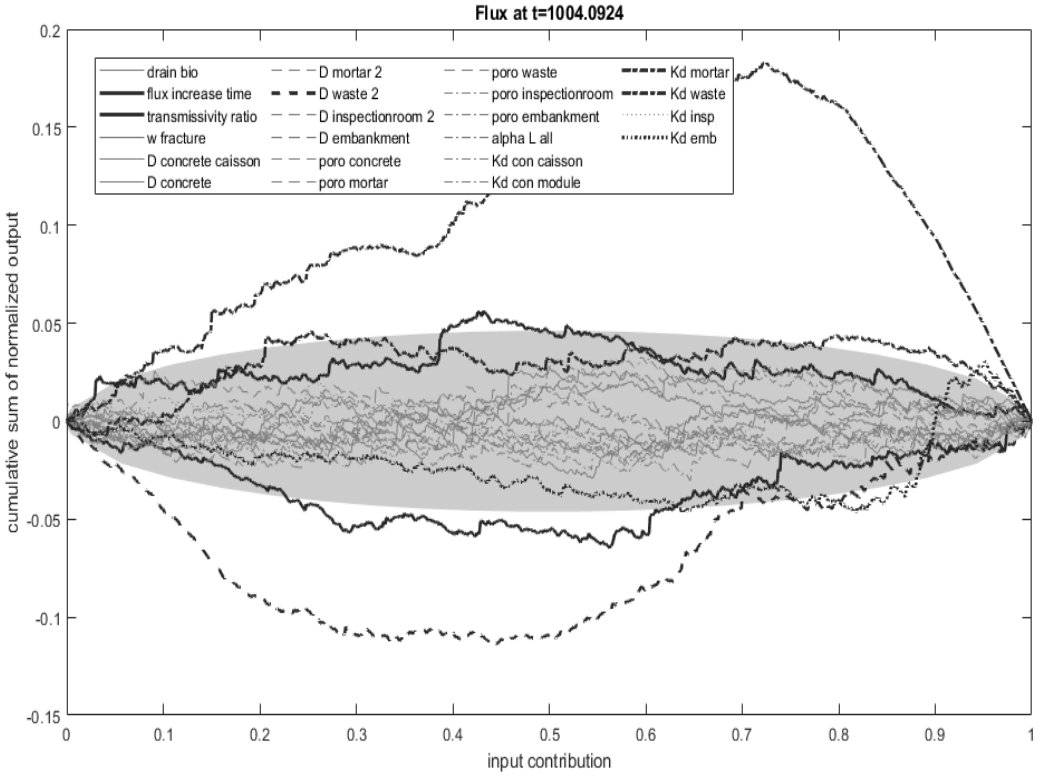

Figure 6-17. CUSUNORO curve with significance ellipsoid

6.4.5. Results from IBRAE

Case Name	Dessel
Sensitivity Analysis Method	PAWN, RBD-FAST
Sensitivity measures generated	PAWN index, S_i first order
Special considerations	Mean and median KS for PAWN. Number of conditioning intervals $n = 8$ was used.
Surrogate models used	
Transformations	
QoIs addressed	SA for each time step, SA for peak value, SA for integral
Number of samples used	256, 1024, 24000
Dataset	Results_SA_TI_I129_256_SobolSamples.h5, Results_SA_TI_I129_1024_SobolSamples.h5, Results_SA_TI_I129_Sobol24k.h5
Software	Python SaLib, Python in-house software

Table 6-12. Sensitivity analysis of Dessel Case by IBRAE

IBRAE performed sensitivity measures evaluation for the Dessel case using samples of 256, 1024 and 2400 realizations. The findings are listed in Table 6-13.

Figure 6-18 through Figure 6-26 show results for small (256), medium (1024 model runs), and the larger (24000 model runs) samples correspondingly.

The 5 most significant parameters are flux_increase_time (at the beginning of the simulation only), *Kd_waste*, *D_waste_2*, *Kd_emb*, *Kd_mortar*.

The differences can be seen in the rating of the parameters at the end of the simulation. For example, the influence of *Kd_con_caisson* and *Kd_mortar* parameters is medium according to PAWN but minor according to RBD. Also, *D_waste_2* at the end of the simulation is most significant by the PAWN method and only 3rd (after *Kd_waste* and *Kd_emb*) by RBD-FAST. On the contrary, at t≈1300 years influence of *Kd_emb* is significant by RDB-FAST and minor by PAWN.

Smaller samples allow detecting the most influential parameters, the ranking of the parameters of secondary significance though is less clear.

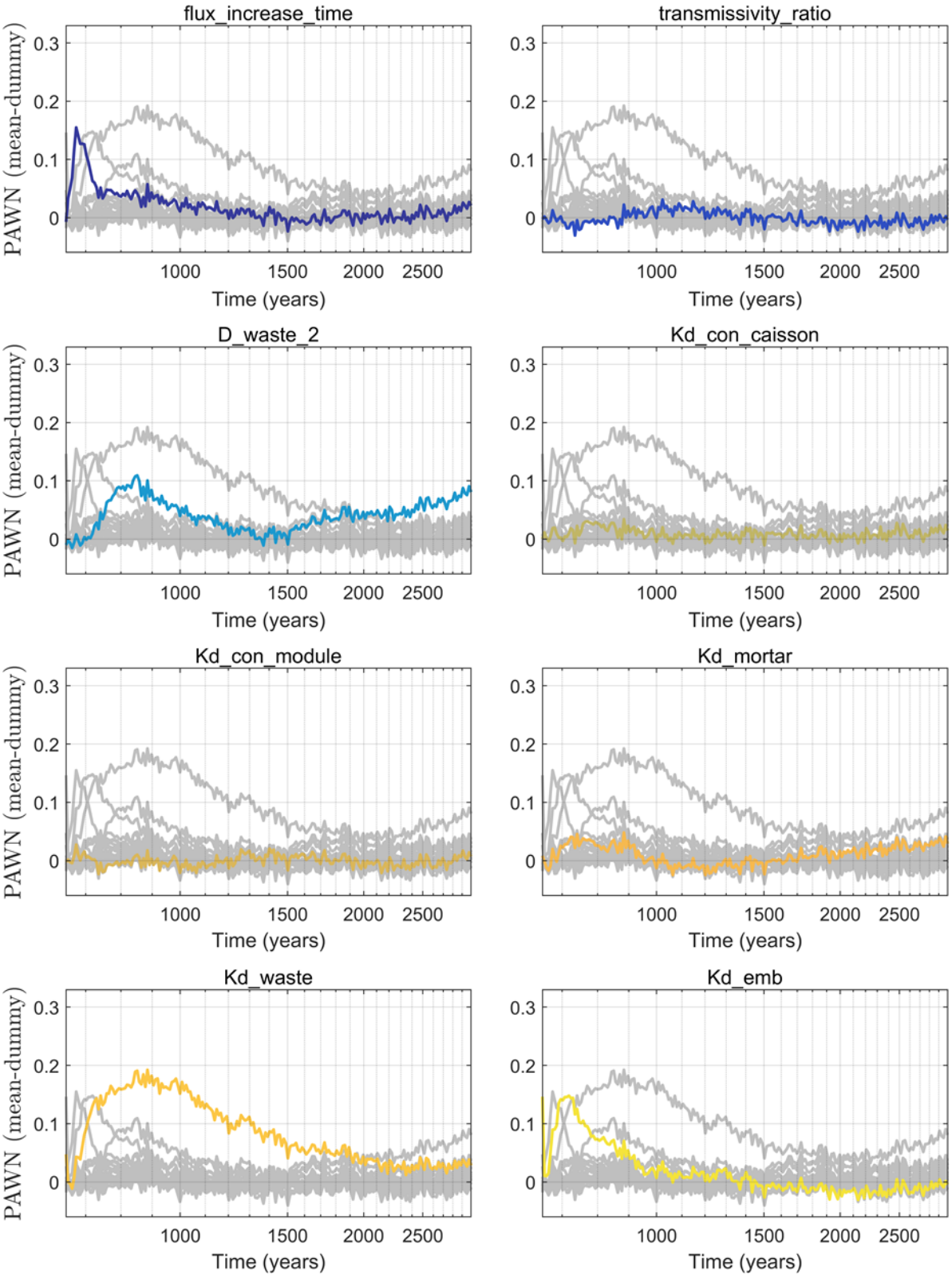

Figure 6-18. Sensitivity analysis of Dessel case by IBRAE: PAWN method (mean statistics, "dummy" parameter subtracted), 256 realizations

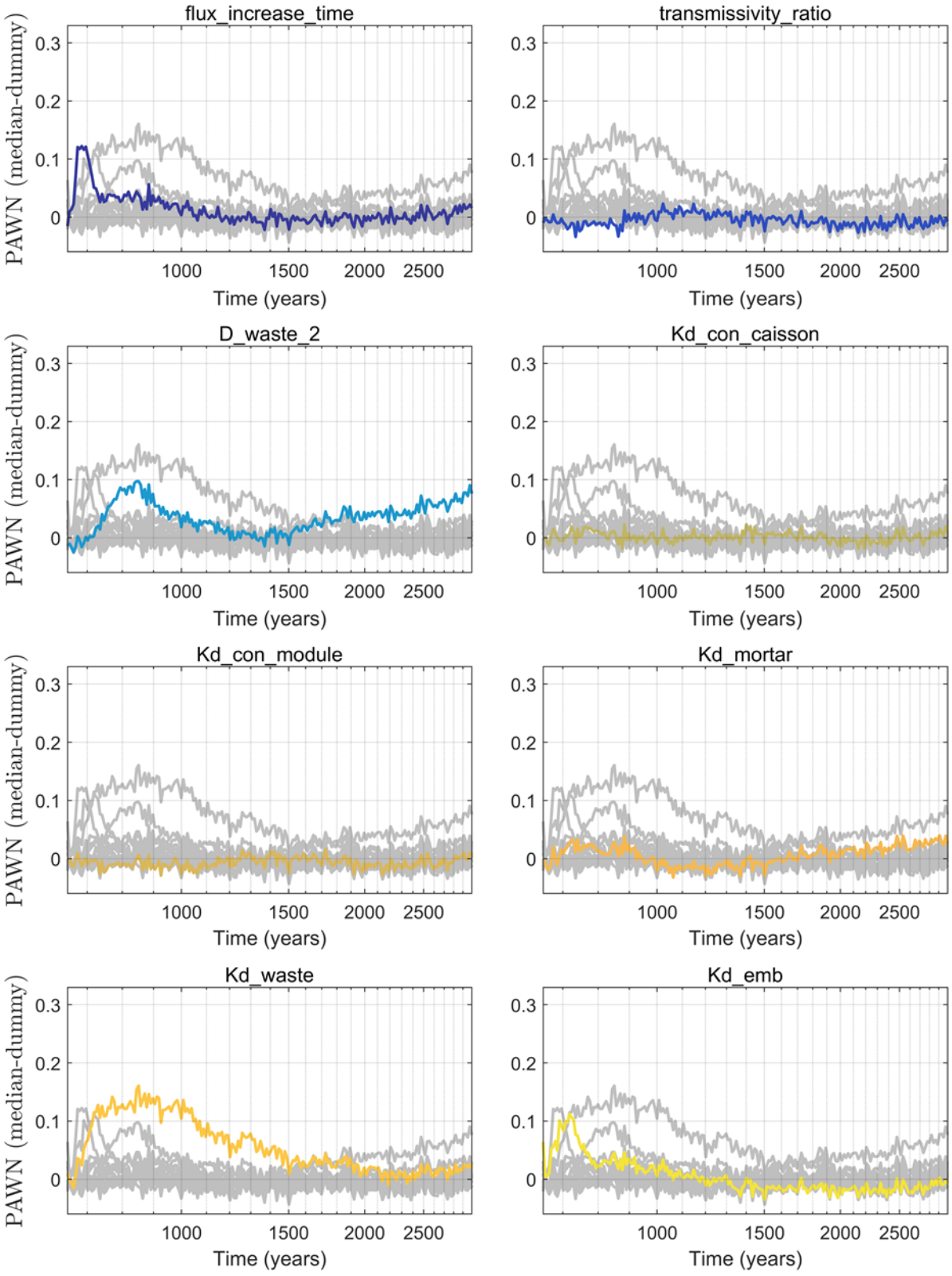

Figure 6-19. Sensitivity analysis of Dessel case by IBRAE: PAWN method (median statistics, "dummy" parameter subtracted), 256 realizations

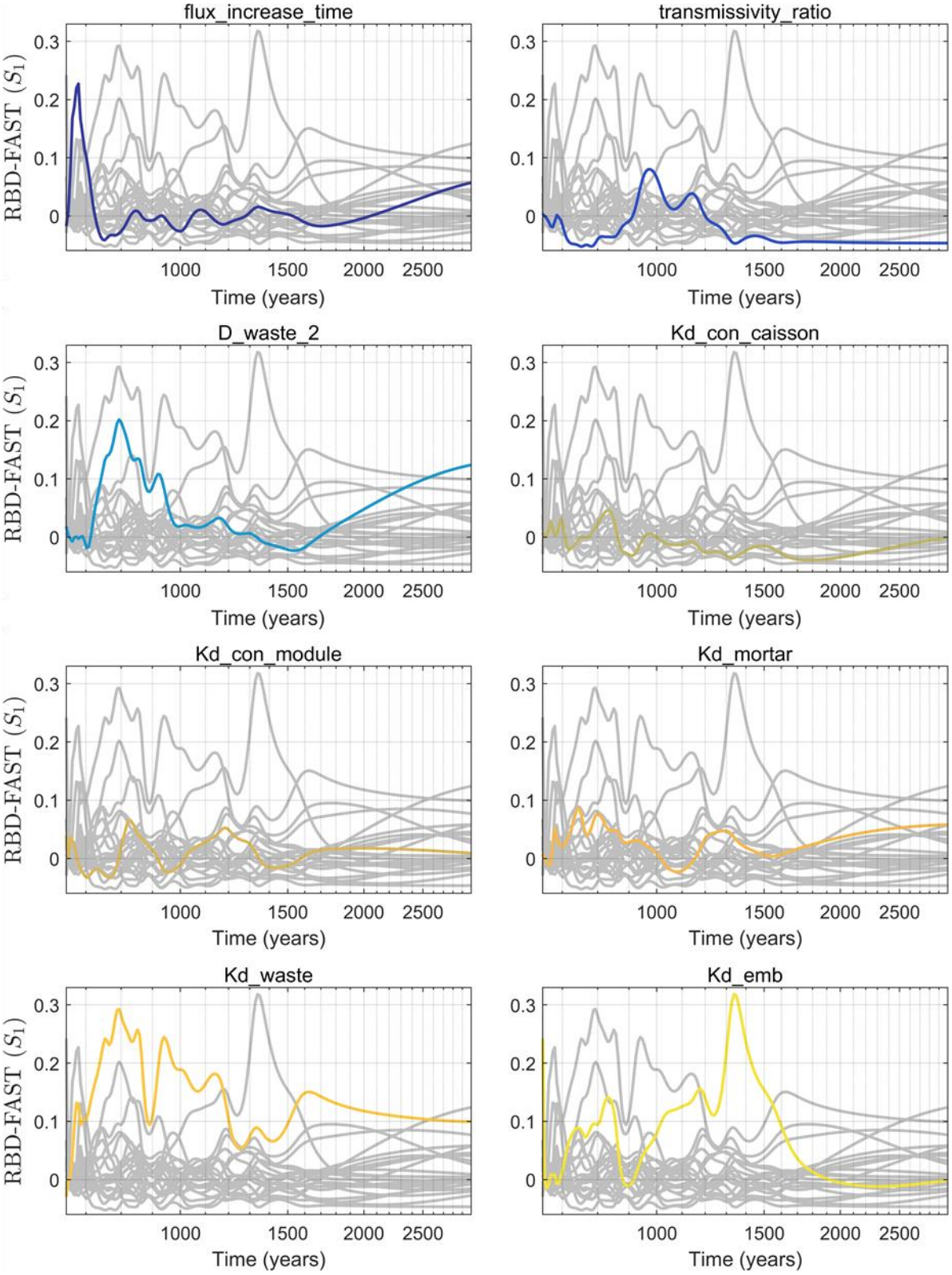

Figure 6-20. Sensitivity analysis of Dessel case by IBRAE: RBD-FAST method, 256 realizations

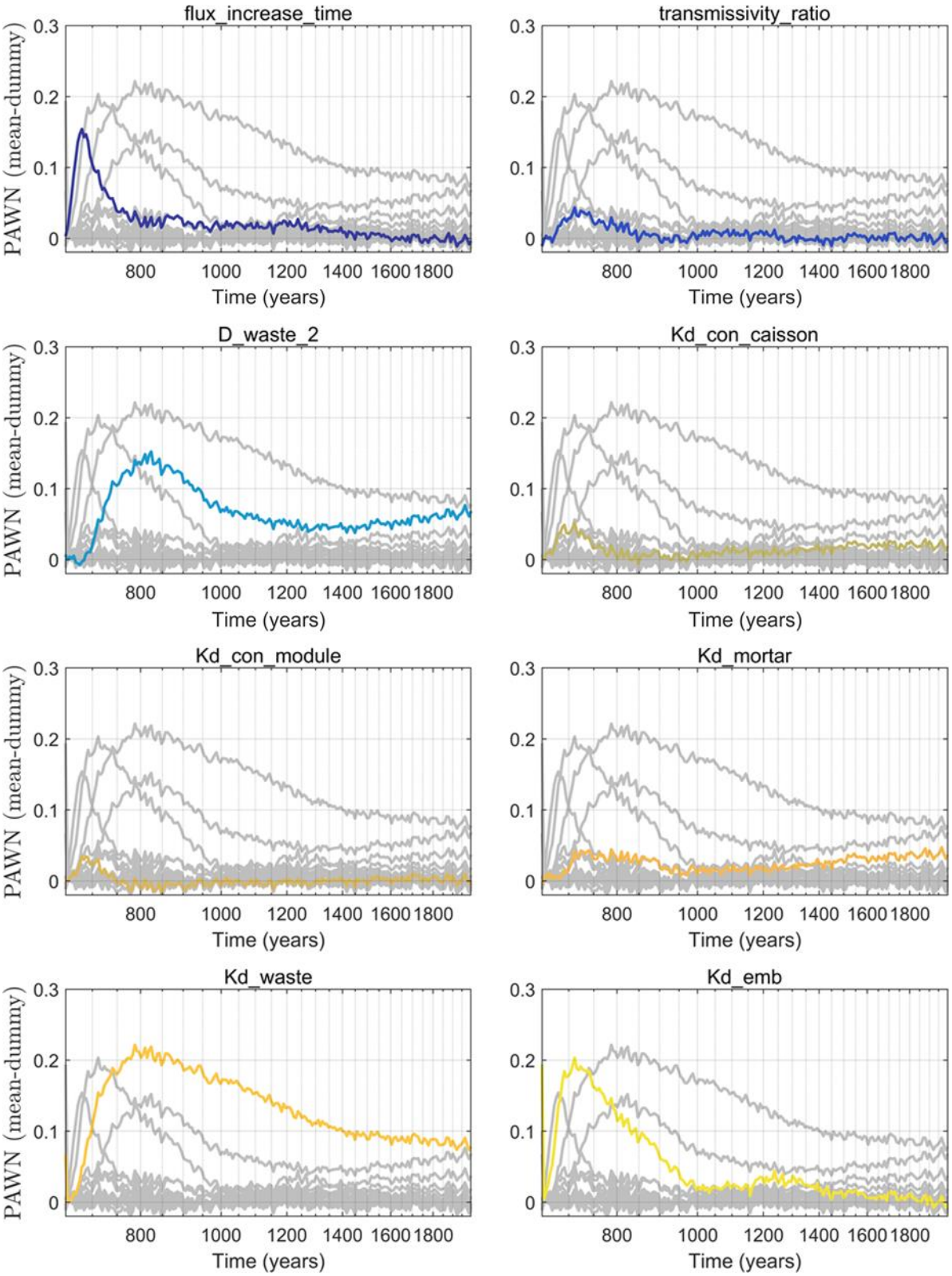

Figure 6-21. Sensitivity analysis of Dessel case by IBRAE: PAWN method (mean statistics, "dummy" parameter subtracted), 1024 realizations

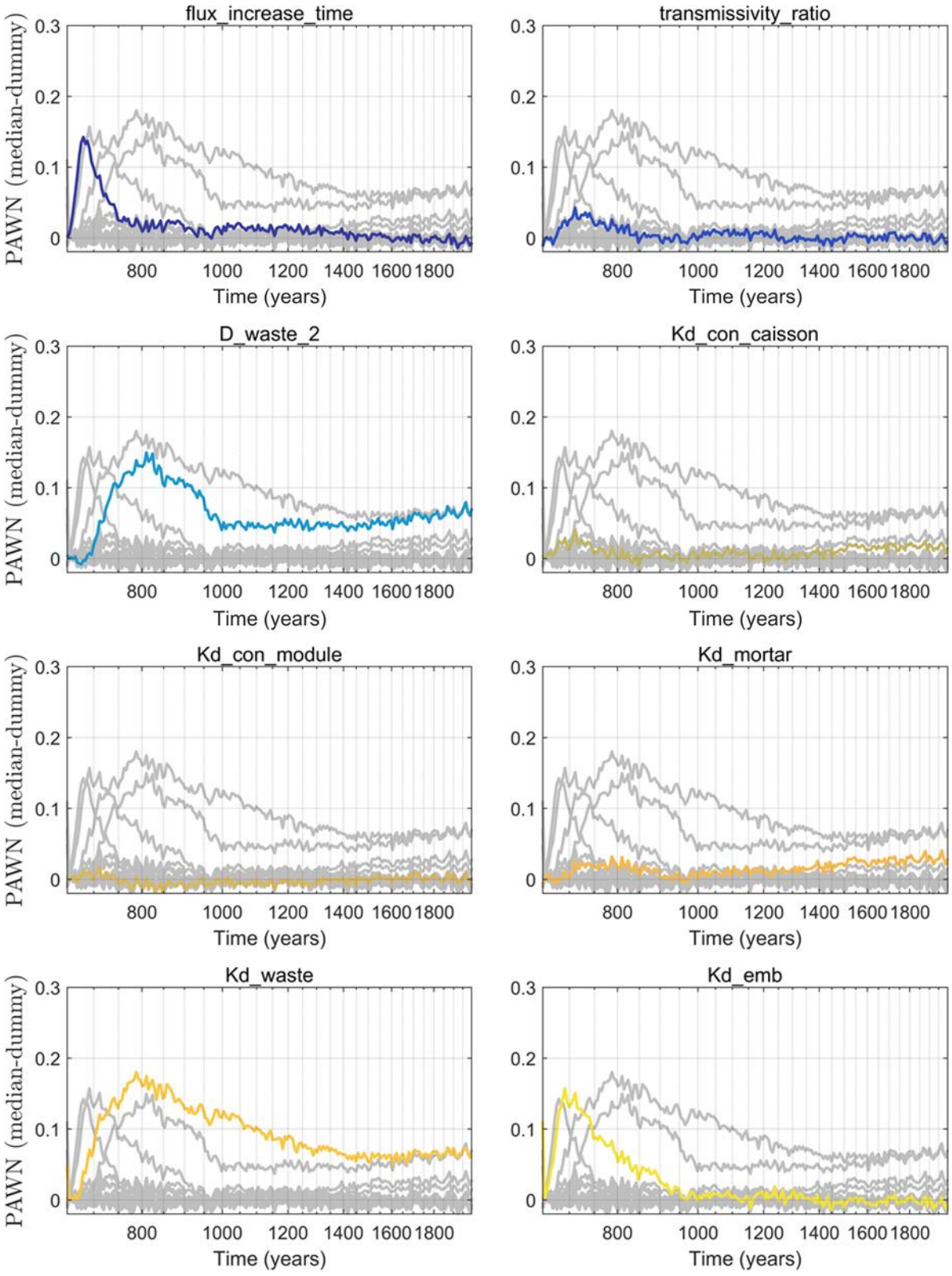

Figure 6-22. Sensitivity analysis of Dessel case by IBRAE: PAWN method (median statistics, "dummy" parameter subtracted), 1024 realizations

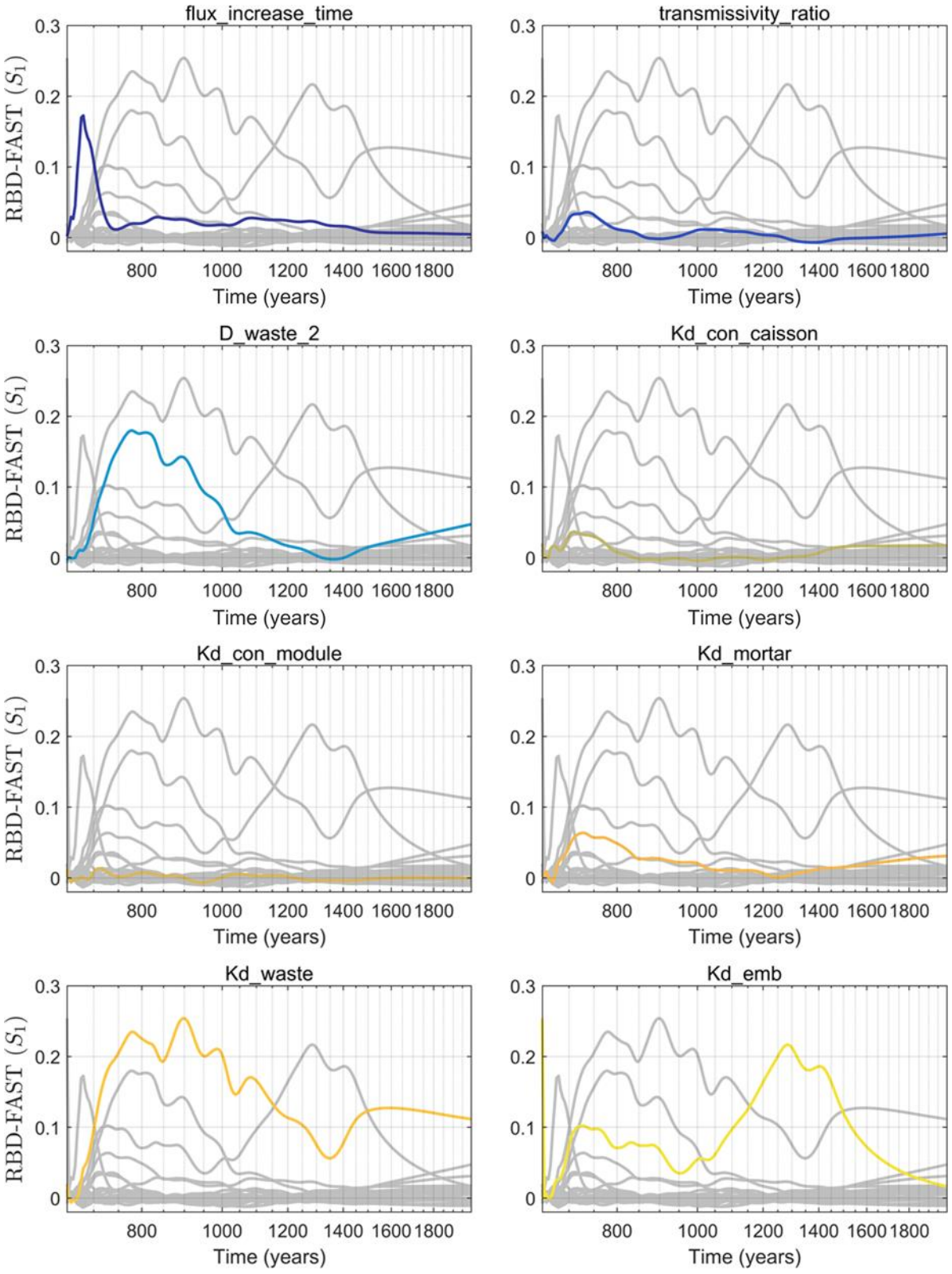

Figure 6-23. Sensitivity analysis of Dessel case by IBRAE: RBD-FAST method, 1024 realizations

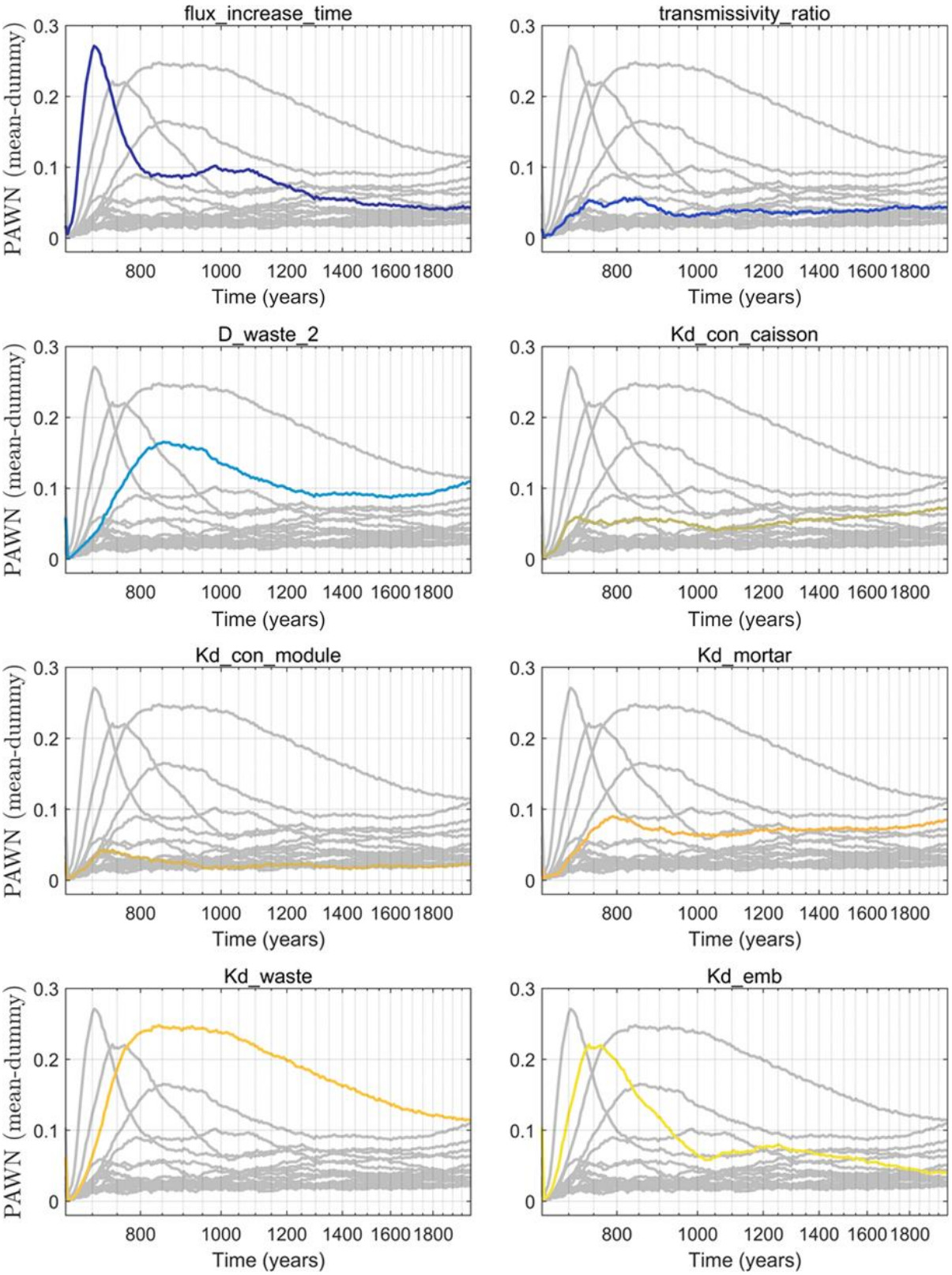

Figure 6-24. Sensitivity analysis of Dessel case by IBRAE: PAWN method (mean statistics, "dummy" parameter subtracted), 24000 realizations

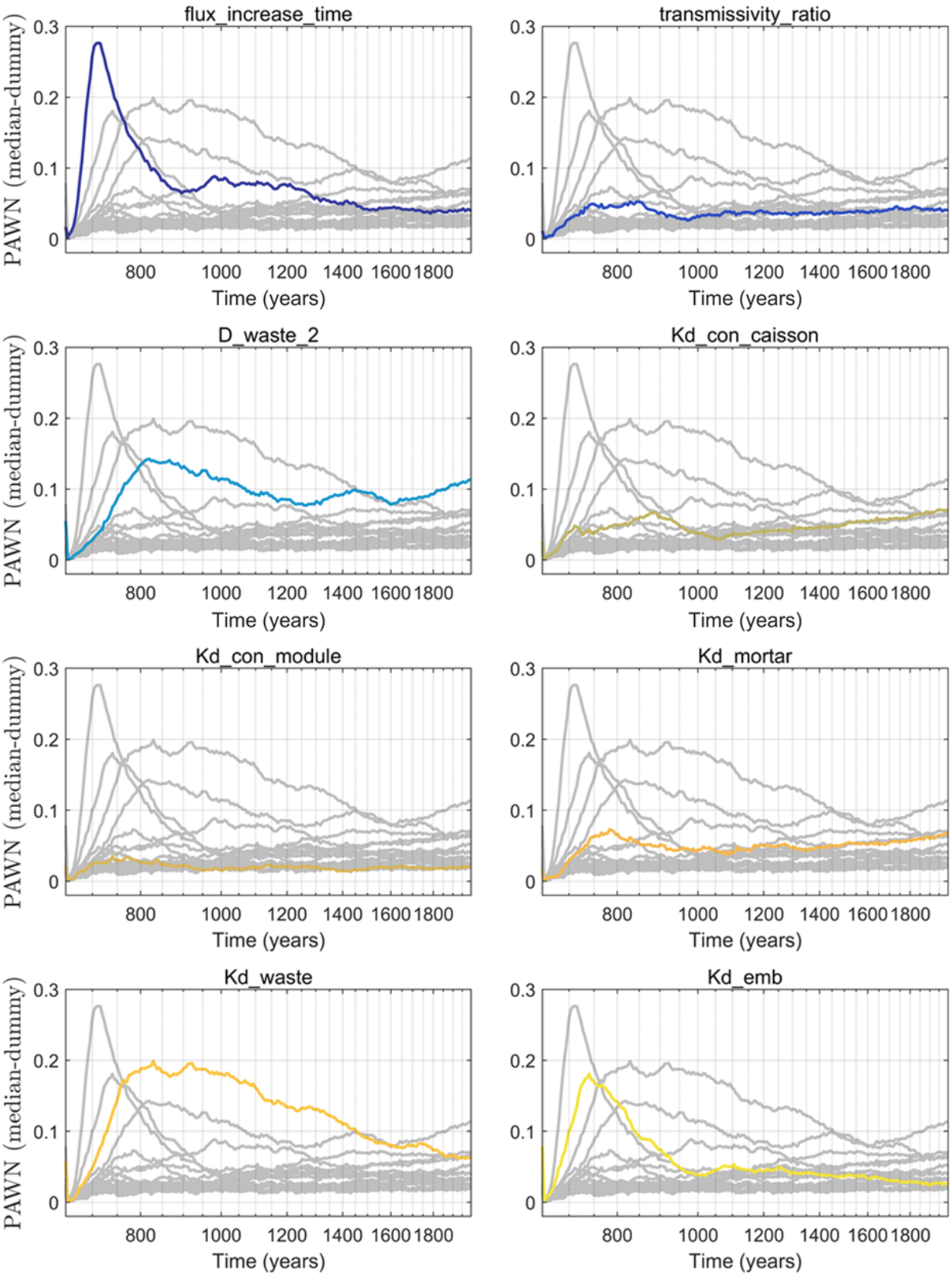

Figure 6-25. Sensitivity analysis of Dessel case by IBRAE: PAWN method (median statistics, "dummy" parameter subtracted), 24000 realizations

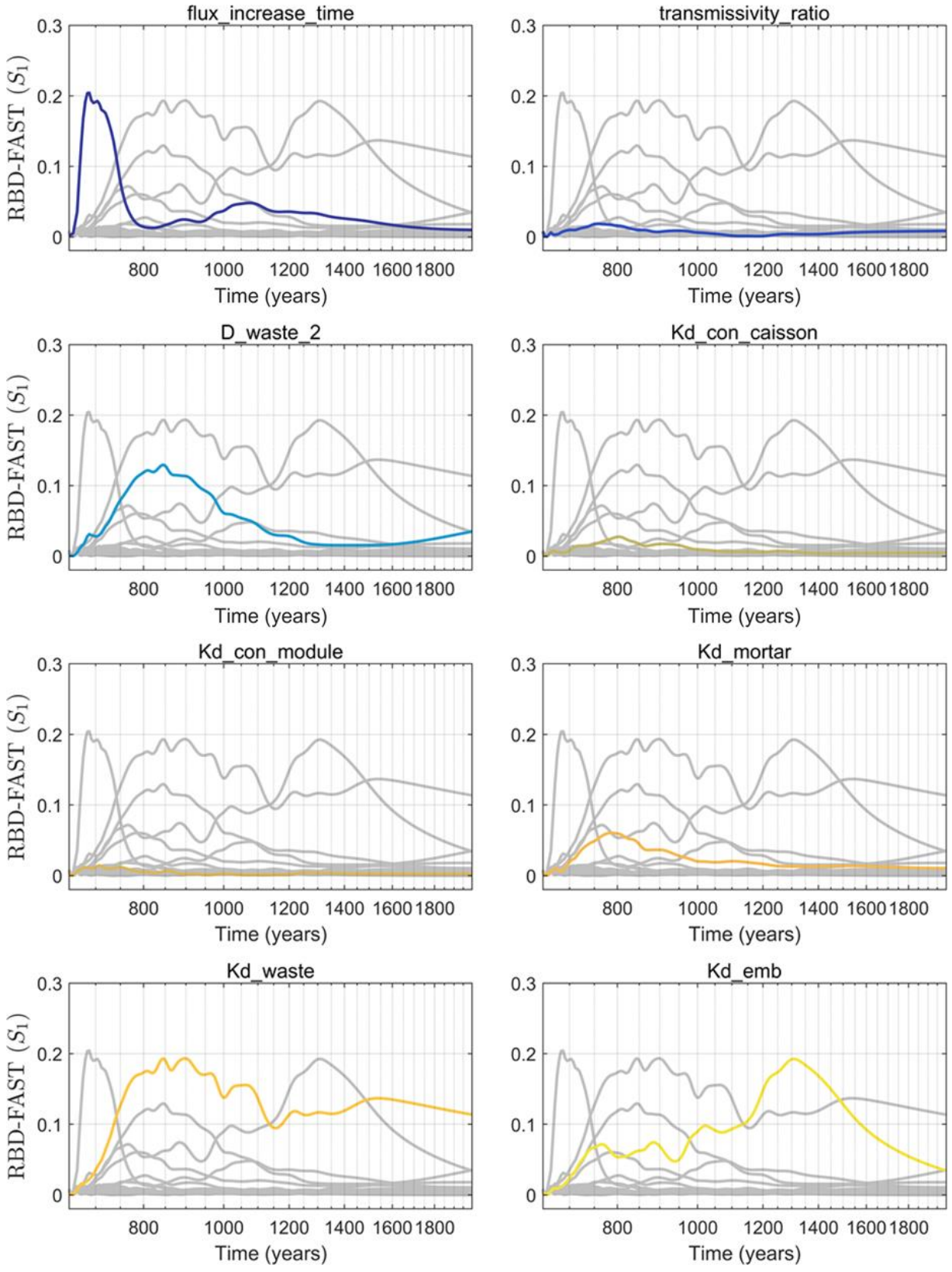

Figure 6-26. Sensitivity analysis of Dessel case by IBRAE: RBD-FAST method, 24000 realizations

Input parameter	PAWN	RBD-FAST
flux_increase_time	Dominating at $t \approx 700$ years, almost insignificant after	Dominating at $t \approx 700$ years, almost insignificant after
transmissivity_ratio	Mostly minor significance	Mostly minor significance
D_waste_2	Second at $700 < t < 900$ years, most influencing at the end of the simulation	Second at $700 < t < 900$ years and at the end of the simulation
Kd_con_caisson	Medium significance at t~900 years, small significance after that	Mostly minor significance
Kd_con_module	Mostly minor significance	Mostly minor significance
Kd_mortar	Medium significance at $700 < t < 900$ years and at the end of the simulation	Fourth at $700 < t < 900$ years
Kd_waste	Third at $t < 800$ years, dominating at $800 < t < 1500$ years	Dominating at $800 < t < 1100$ years, $t > 1500$ years
Kd_emb	Second at $t < 800$ years, small significance at $t > 1900$ years	Dominating at $1100 < t < 1500$ years, second at the end of the simulation
drain_bio, w_fracture, D_concrete_caisson, D_concrete, D_mortar_2, D_inspectionroom_2, D_embankment, poro_concrete, poro_mortar, poro_waste, poro_inspectionroom, poro_embankment, alpha_L_all, Kd_insp	Minor or insignificant	Minor or insignificant

Table 6-13. SA Results of Dessel Case by IBRAE

6.4.6. Case Summary of Sensitivity Results

In Figure 6-27, selected scatterplots are presented. These show the dependency of the model output on the values of the five most conspicuous parameters for several points in time. Obviously, the sensitivity of at least some of the parameters is distinctly time-dependent.

The shape of the scatterplots indicates possible threshold effects for sorption in the embankment and waste matrix (*Kd_waste* and *Kd_emb*), and non-monotonicity of the diffusion coefficient in the waste (D_waste2).

To analyze this in detail, the participants independently calculated the evolution over time of different sensitivity estimators using their available tools. While correlation- and regression-based methods yield values between -1 and 1, the sign indicating the direction of influence, variance-based methods are designed to calculate Sobol's sensitivity indices, which lie between 0 and 1. Table 6-14 gives an overview of the time-dependent evaluations, including the method of calculation in the case of variance-based sensitivity indices. The standardized regression coefficients (SRC) and sensitivity indices calculated using EASI (Effective Algorithm for Computing Global Sensitivity Indices, [92]) are in very good agreement between all participants, which is not too surprising as the mathematical algorithms applied are the same. This is a good mutual verification of the software tools.

More interesting from a scientific point of view is the comparison between the results for 1st-order S_i, calculated with EASI (SCKCEN/GRS), with PCE (SNL, [93], SCKCEN) and CUSUNORO slope (TUC, [94]) as presented in Figure 6-28. While the curves show similar qualitative behavior, there are considerable differences in detail. PCE computes higher index values for *Kd_waste* and, specifically, *Kd_emb*. This is probably due to the fact that PCE uses an internal surrogate model, which itself requires additional assumptions and parameters that have to be well adjusted. For this investigation, second-order polynomials were used.

Figure 6-29 presents the comparison of a direct and a rank-based sensitivity analysis, using the SRC and standardized rank regression coefficients (SRRC) methods. The rank transformation makes the time curves smoother and yields higher sensitivities for some of the parameters in the early phase. Typically, ranking leads to more pronounced differences of low values but less pronounced differences of high values. It can be concluded that the parameters flux_increase_time, *Kd_waste* and *Kd_emb* mainly influence the low model output values at early times. *Kd_emb* changes its direction of influence around 1000 years after disposal; after that time its influence becomes positive (higher value implies higher output) but seems to affect mainly the high model output values, as SRC is much higher than SRRC at medium times. The same conclusion can be drawn for *Kd_waste* at late times. Ranking input and output data did seem to improve the goodness-of-fit R^2 in the most interesting time-frame.

Sensitivity measure	SNL	SCKCEN	GRS	TUC	IBRAE
Standardized Regression Coefficients (SRC)		X	X		
Standardized Rank Regression Coefficients (SRRC)		X	X		
SRC with transformation			X		
Pearson's Correlation Coefficients (CC)	X				
Spearman' Rank Correlation Coefficients (RCC)	X				
Partial Correlation Coefficients (PCC)	X				
Partial Rank Correlation Coefficients (PRCC)	X	X			
1st order Sensitivity Index (S_i)	PCE	PCE/EASI	EASI	EASI/CSN	
2nd order Sensitivity Index (S_{i2})				Harmonic Regress.	
Total order Sensitivity Index (T_i)	PCE	PCE			

Table 6-14. Time-dependent sensitivity measures calculated by different participants

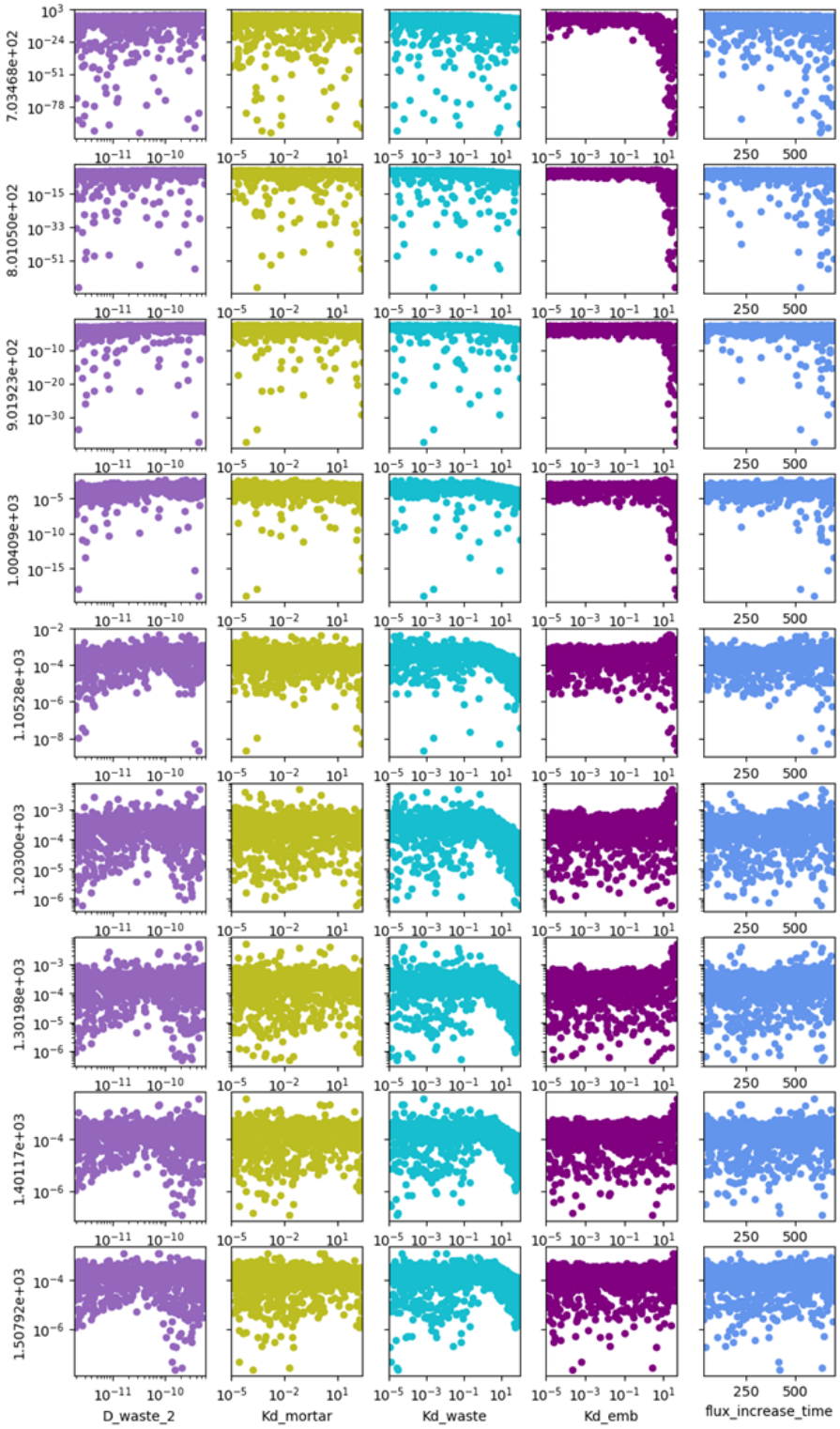

Figure 6-27. Scatterplots for the five most conspicuous parameters and several points in time (SNL)

The different sensitivity estimators agree about the most influential parameters, although, at each point in time, they calculate slightly different rankings. Parameter ranking, however, should not be regarded as the only relevant result of sensitivity analysis. In fact, a well-understood sensitivity analysis with different methods, each addressing its specific aspects, can essentially increase detailed understanding of model behavior and thus the confidence in the model.

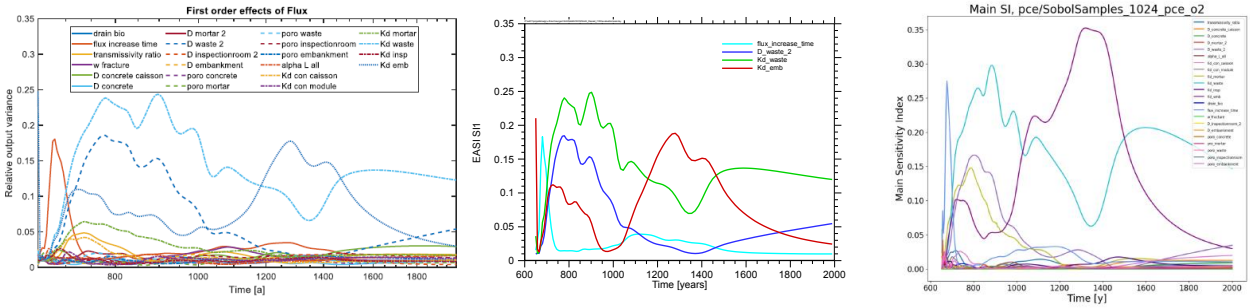

Figure 6-28. 1st-order sensitivity index, calculated with PCE, EASI and CUSUNORO. Left: SNL (PCE), middle: GRS (EASI), right: TUC (CUSUNORO slope)

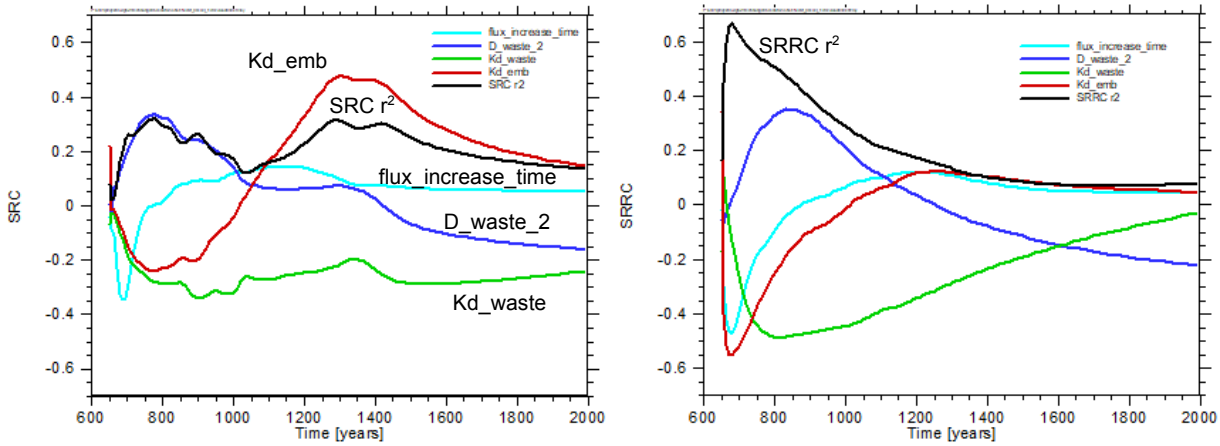

Figure 6-29. SRC vs. SRRC (GRS)

6.4.7. Interactions

If the functional dependence of the output upon the inputs is not explained by linear or additive functions (i.e. in form of a low goodness-of-fit R^2or a low sum of first order effects) then the missing part can be explained by the presence of interactions. First-order effects (and therefore also CUSUNORO plots) are not able to detect interactions. Figure 6-30 demonstrates this where, especially for the large output variance in the time span between 680 to 800 years, a large part remains unaccounted for by first order effects. In this figure, absolute rather than relative contributions to the variance are shown, which reveals additional information when considering sensitivities evolving over time. One may be tempted to screen-out parameters which do not contribute to first order effects but may be important in driving interactions.

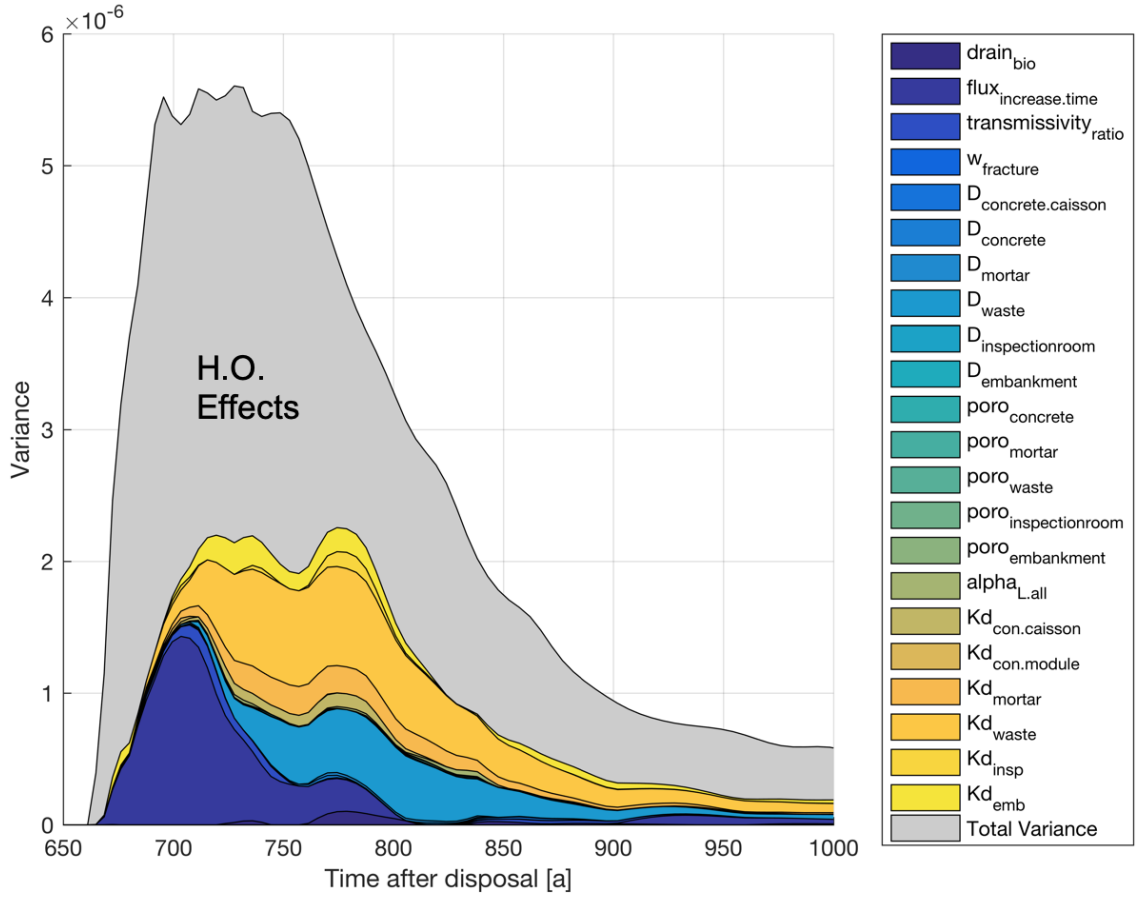

Figure 6-30. Sum of first order effects and absolute contribution to total output variance (SCKCEN)

Figure 6-31 compares results of total order sensitivity indices calculated by either the Sobol' method with 24,000 direct simulations ('True'), or by different metamodeling techniques, which were calibrated on the 1024 simulations – dataset. If factor fixing is the main objective of the SA, total effects need to be calculated.

The qualitative behavior of T_i in the critical time span closely follows that of first order effects. However, it appears that all metamodeling methods seem to miss important higher-order effects. The large-sample Sobol-method indicated high T_i values of *flux_increase_time* & *Kd_emb* in the period from 750 to 1000 years.

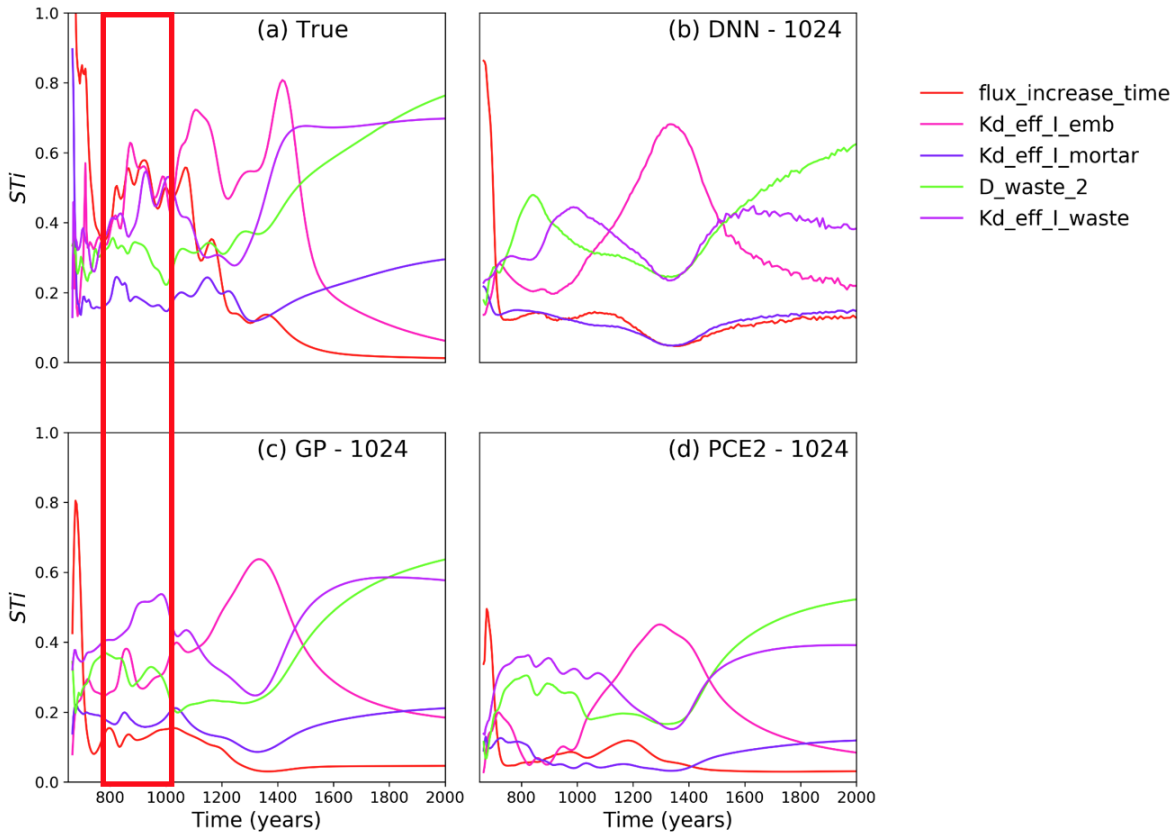

Figure 6-31. Results of total order sensitivity indices calculated by either the Sobol' method with 24,000 direct simulations ('True'), or by different metamodeling techniques

6.5. Interpretation of Sensitivity Analysis in the Light of Model Understanding (Model Owners Feedback)

6.5.1. Behavior of the Model with Respect to Sorption in the Waste Matrix (Kd_waste)

It is no surprise that *Kd_waste* is identified as the most important parameter by all methods. The sorption in the waste matrix determines the initial liquid concentrations, which scale almost directly with the advective outflux. The threshold effect can be explained by the back-up sorption capacity that the surrounding materials offer. As long as sorption in the mortar and concrete of the monolith caisson is larger than sorption in the waste matrix, the activity will quickly be transferred and attenuated in these surrounding, cementitious components.

6.5.2. Behavior of the Model with Respect to Sorption in the Embankment (Kd_emb)

Changing sign: Higher values of *Kd_emb* lead to lower outflow at early times (< 1000 years), but to higher outflow at late times. Why?

The embankment is the last component in the disposal facility which RNs encounter in their migration to the geosphere. The transport in the embankment is purely advective, similar to a one-dimensional plug flow. Higher K_d values lead to a higher retardation factor. A larger retardation will therefor result in a larger shift in time of the breakthrough curve. This is why at later times (after $1000a$), the influence changes its sign.

Penalizing combinations of parameters (e.g. fast increase in water flux and low retention in the waste matrix) result in sharp and early breakthrough curves (650 – 750 years). Only high *Kd_emb* values are able to significantly shift these steep breakthrough curves backwards in time, which explains the threshold effect at 10^{-1} L/kg.

6.5.3. Behavior of the Model with respect to diffusion in the Waste Matrix (Dp_waste)

In the scatterplots of Figure 6-27, non-monotonic behavior could be spotted for the *Dp_waste* parameter. Indeed, the diffusion coefficient in the waste matrix has a dual effect on the activity flux. On one hand, it accelerates the initial diffusion to the fracture on the right, which increases the peak activity flux once as water starts entering the system. On the other hand, it will also increase diffusion to the upper, lower and left direction where mortar is present, which accumulates and attenuates the activity. This effect would reduce the main peak flux and increase the tailing of the peak.

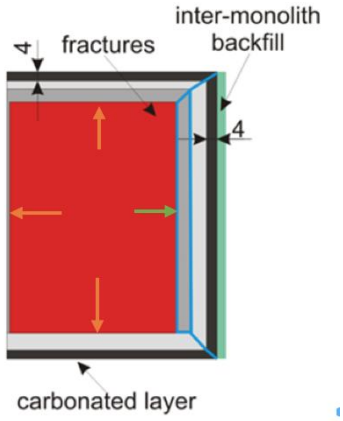

Figure 6-32. Non-monotonous effect of diffusion in the waste matrix. Diffusion in the orange direction will decrease peak flux, green will increase peak flux

6.6. Conclusions

For the studied Dessel example, the standard set of SA methods does not sufficiently explain the model behavior. To be able to detect such a situation, goodness-of-fit measures of the used sensitivity method must be reported to obtain information on its explanatory power for the case under consideration. Possible ways to address this issue are output transformations or moment-independent methods which take into account the whole distribution instead of point-estimators of conditional means or variances.

This page is intentionally left blank

7. IBRAE GROUNDWATER CASE

This model describes a cross-section groundwater flow in the heterogeneous geological media of Nizhnekansky massif (Krasnoyarsk territory) where construction of a site-specific underground research facility (URF) is considered. Site-specific means that the research facility for acquiring expertise to develop a radioactive waste repository is located in the same formation as the planned repository. The granitoid rock of Nizhnekansky massif indicates the presence of geological structural elements influencing significantly the permeability features such as dykes, faults, crushing zones (Figure 7-1). Boundary conditions are depicted in Figure 7-2. The modeling is performed using GeRa groundwater flow and transport modeling computer code [95].

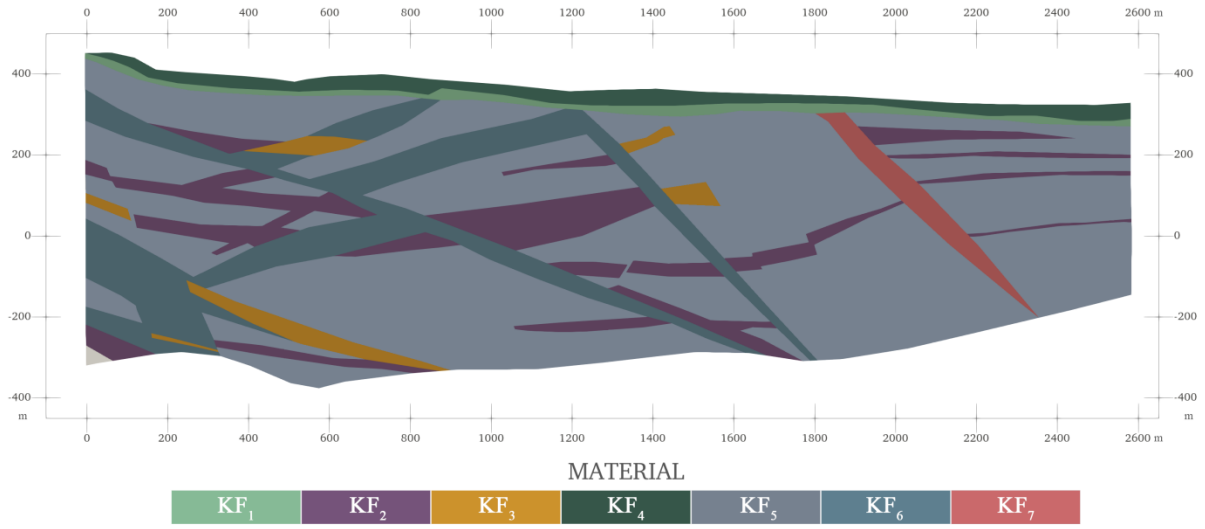

Figure 7-1. Geological model: hydraulic conductivities of structural elements

7.1. Case description

Reference case name	Description
Waste form	High-level radioactive waste
Engineered barriers	Stainless steel overpack, bentonite buffer
Repository description	450–500 m depth, 75 m deep boreholes containing waste packages
Natural system	500 m thick low permeability rock
Biosphere	Not considered
Conceptual release pathways	Not considered
Processes modeled	Single-phase fluid flow
Software codes used	GERA [95]
Reference for full description of case	[96, 97]

Table 7-1. IBRAE Case Description

7.2. Description of inputs and outputs

The groundwater flow model has 12 input parameters (Table 7-2): 7 hydraulic conductivities for geologic structural elements (quaternary deposits, weathering crust, fissured dikes, dikes, fissured gneiss, crushing zone, gneiss) and 5 boundary conditions (flow rates for 3 zones at the left boundary, infiltration, and drainage).

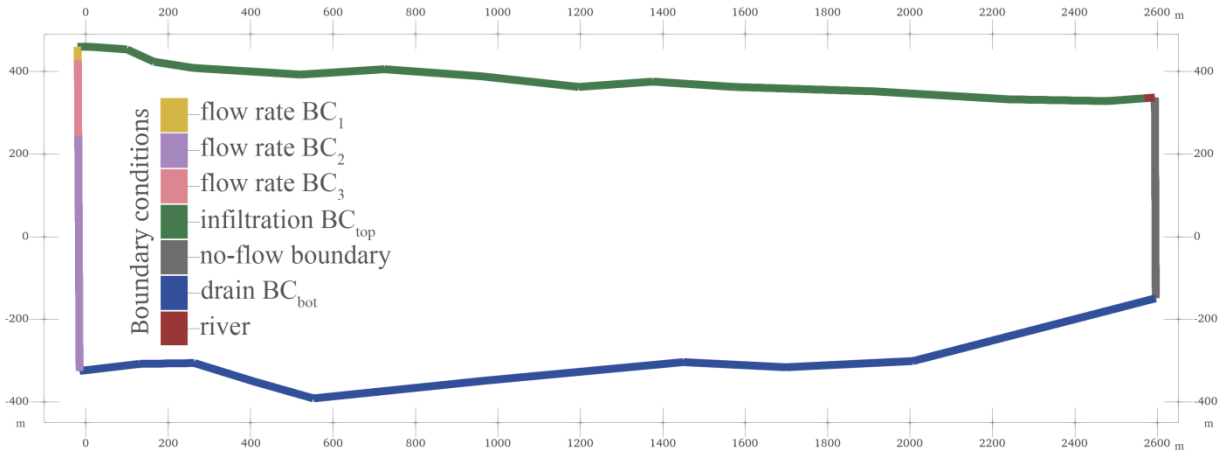

Figure 7-2. Groundwater flow model: boundary conditions

Samples were constructed by Saltelli modification of Sobol' sequence [98] presented in Python SAlib library [99]. This method treats parameters as independent, each variated within a given range. To study if there is a dependence of sensitivity analysis results on the number of samples were constructed 4 sample sets containing 140, 1400, 14,000, and 28,000 samples.

Input parameter	**Distribution (uniform)**	
Hydraulic conductivities [m/day]	min	Max
KF1	1.00E-03	1.00E+00
KF2	1.00E-06	1.00E-02
KF3	1.00E-06	3.50E-02
KF4	1.00E-03	5.00E-01
KF5	1.00E-06	1.00E-03
KF6	5.00E-03	1.00E-01
KF7	1.00E-02	1.00E+00
Boundary conditions [m^3/day]	min	Max
BC1	-1.00E-02	-1.00E-06
BC2	-1.00E-02	-1.00E-06
BC3	-1.00E-02	-1.00E-06
BCtop	1.00E-06	1.00E-04
BCbot	-1.00E-06	1.00E-03

Table 7-2. IBRAE Case Input Parameters

The model has an output composed of 37 values of hydraulic heads located corresponding to the locations of experimental observations later used in model calibration. Points 1-14 are from the

first borehole, 15-26 from the second, and 27-37 from the third. They can be considered separately as 37 outputs, although their behavior is similar (Figure 7-3).

Output QoI	Type	Other
Output 1	Vector – 37 hydraulic heads [m] located in different spatial points	Can be considered separately

Table 7-3. IBRAE Case Output QoIs

7.3. Salient features and behavior of the model

Distributions of outputs for different sample sets are shown in Figure 7-3. Each output is presented by its color. The distributions look normal-like, do not significantly depend on sample set size, and are rather similar for different outputs.

The result of the groundwater flow simulation (isolines of the hydraulic heads in the whole modeling space) is presented in Figure 7-4.

Scatterplots for first five outputs are shown in Figure 7-5.

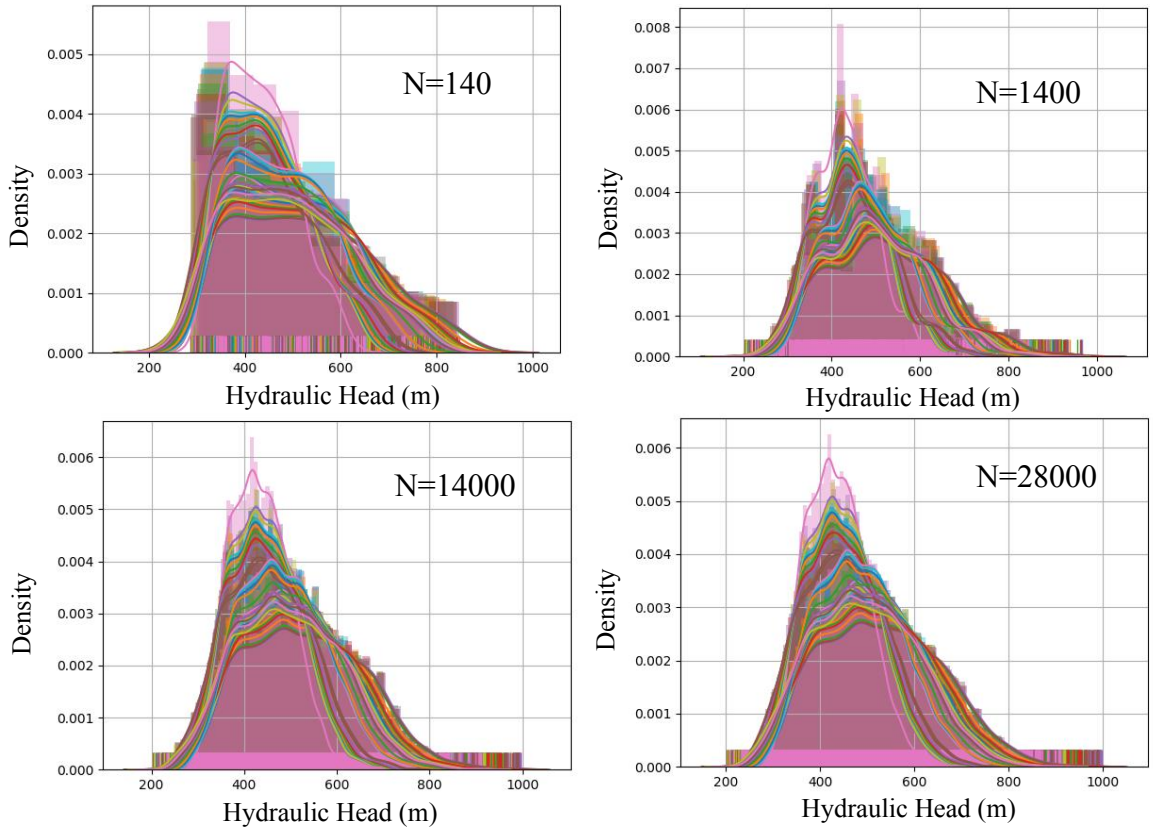

Figure 7-3. Output distributions for different sample sizes

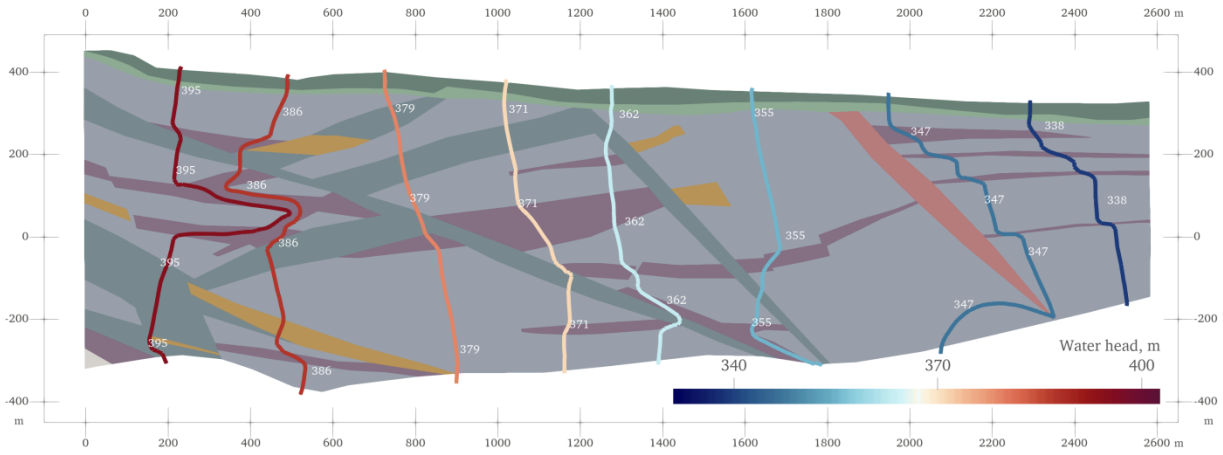

Figure 7-4. Simulation results

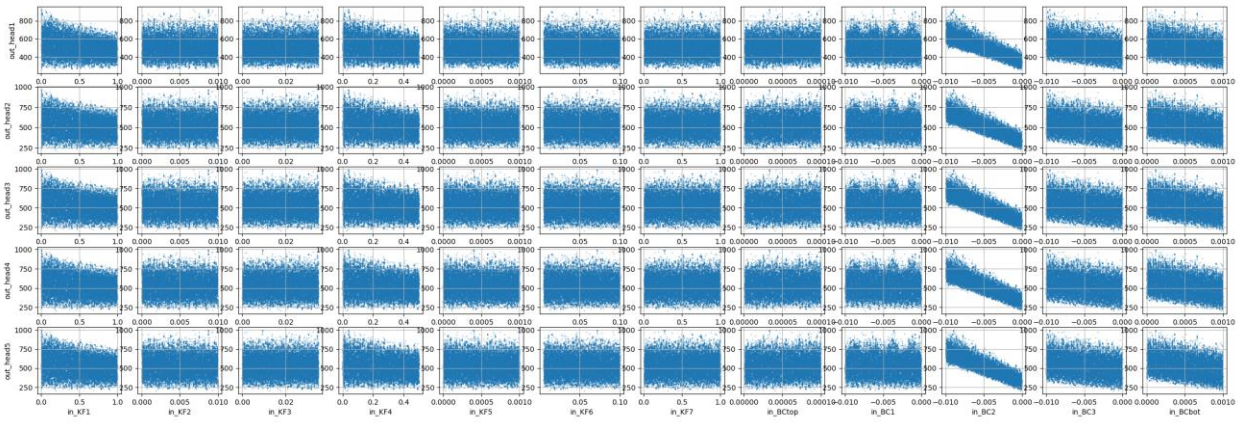

Figure 7-5. Scatterplots for first five outputs of groundwater model

7.4. Sensitivity Analysis Results

7.4.1. Results from GRS

Case Name	Description
Sensitivity Analysis Method	CUSUNORO, linear regression, rank regression, EASI
Sensitivity measures generated	SRC, SRRC, S_i
Special considerations	-
Surrogate models used	-
Transformations	-
QoIs addressed	37 hydraulic heads
Number of samples used	140 / 1400 / 14000 / 28000

Table 7-4. Sensitivity analysis of IBRAE case by GRS

The investigations were performed with all available sets of input and output. To keep the number of figures manageable, we present the results for the smallest and the largest dataset only.

The CUSUNORO plots of the 37 QoIs look very similar, at least for the set of 140 runs. As an example, Figure 7-6 compares the plots for the output quantities 1 and 37. Even the random details of the curves correspond, only a very sharp look at the figures reveals minor differences (for instance, between the pale orange and the pink curve around 0.7). Obviously, the different observation points reflect the same system properties, (nearly) only differing by a monotonic relationship.

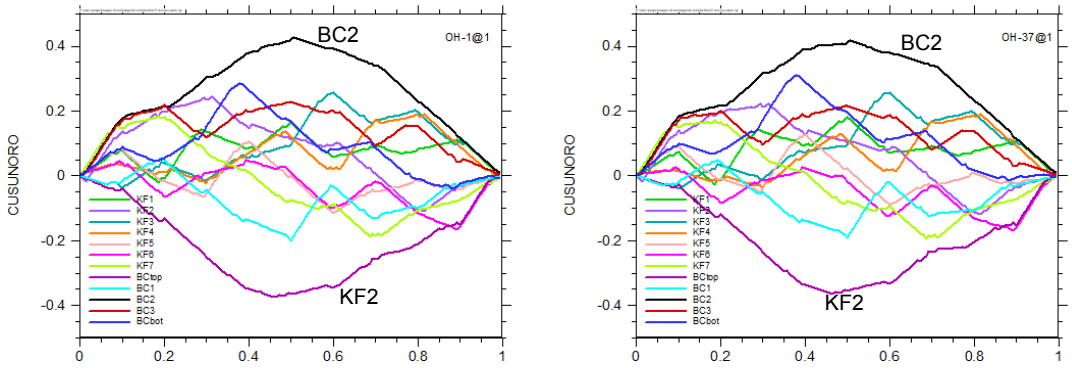

Figure 7-6. CUSUNORO analysis of IBRAE groundwater flow model by GRS (140 runs) for observation points 1 (left) and 37 (right)

The larger sets of runs, however, produce CUSUNORO plots that do not only look smoother but also seem to convey a different message. In Figure 7-7 and Figure 7-8, the plots for the output quantities 1 and 37, generated from the sets of 1400 runs and 28,000 runs, respectively, are presented. In comparison to the previous figure, it is conspicuous that the purple curve (*BCtop*) deviates only very slightly from the horizontal axis, indicating very low sensitivity. From the 140-runs-set it would have been identified as the second most important input parameter. Moreover, the 28,000-runs-curves show more distinct differences between the two output quantities. While *BC2* is clearly the most important parameter, the second one is *BC3* at observation point 1 but *BCbot* at observation point 37. Even so, the CUSUNORO plots look very similar for all observation points, also for the larger sets of runs.

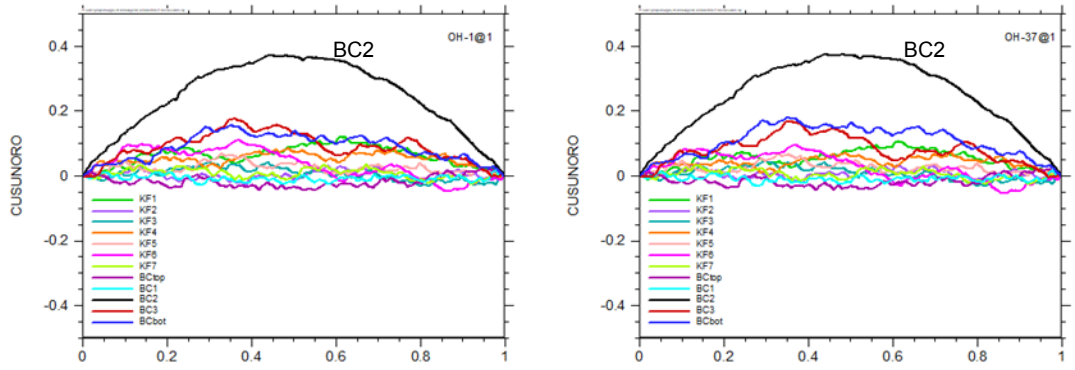

Figure 7-7. CUSUNORO analysis of IBRAE groundwater flow model by GRS (1400 runs) for observation points 1 (left) and 37 (right)

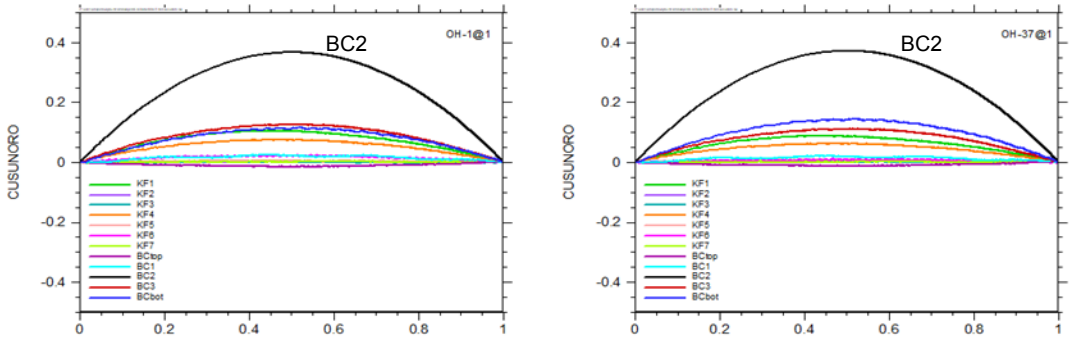

Figure 7-8. CUSUNORO analysis of IBRAE groundwater flow model by GRS (28,000 runs) for observation points 1 (left) and 37 (right)

From this example it can be concluded that a sample size of 140 is not large enough for a reliable sensitivity analysis of the groundwater flow model. Although the most important parameter *BC2* is identified correctly, *BCtop* is falsely identified as important and the curves of the other parameters are not adequate to allow any kind of statement. But even the CUSUNORO curves based on 1400 model runs do not allow clear identification and ranking of the sensitive parameters as there is still too much randomness.

Figure 7-9 shows the results of the direct and rank-based regression analysis for all 37 observation points, calculated from the 28,000-set of runs. Both sets of curves look nearly identical, which is a clear hint that the model is highly linear. This is confirmed by the R^2 value, which is nearly 1, no matter if a rank transformation is performed or not. The evaluation agrees with CUSUNORO insofar as it identifies the input parameter *BC2* as the clearly most important one, followed by *BCbot*, *BC3*, *KF1* and *KF4*. The other parameters play only minor roles. Moreover, these curves confirm that there is only little dependence of the sensitivities on the observation point. As the

system has been found to behave linearly, it is not necessary to perform a variance-based sensitivity analysis.

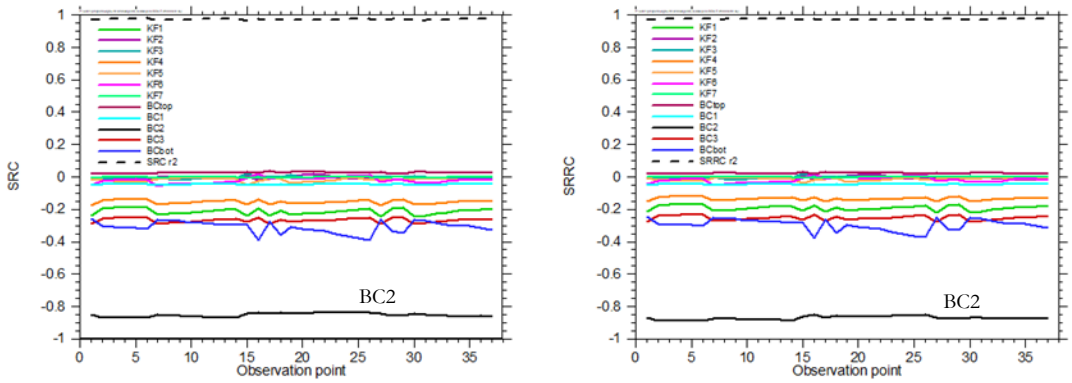

Figure 7-9. Direct (left) and rank-based regression analysis of IBRAE groundwater flow model by GRS (28,000 runs) for all observation points

7.4.2. Results from SNL

Case Name	Description
Sensitivity Analysis Method	Surrogate Modeling, PEAR, SPEAR
Sensitivity measures generated	S_i, T_i, Pearson CC, Spearman RCC
Special considerations	
Surrogate models used	PCE (order 2), MARS, Quadratic Regression. Linear Regression
Transformations	Rank
QoIs addressed	37 hydraulic heads
Number of samples used	140, 1400, 14000, 28000

Table 7-5. Sensitivity analysis of IBRAE case by SNL

The correlation coefficient sensitivity analysis results for the dataset of 1400 realizations are shown in Figure 7-10 (left). According to these results, *KF1*, *KF4*, *KF6*, and the *BC* parameters may be significant, however, *BC2* has the dominant effect. The use of a rank transformation does not appear to impact results. Differences between the partial correlation coefficients and the simple correlation coefficients suggest effects may be masked by other parameters.

The sensitivity indices calculated from linear, quadratic, MARS, and PCE (order 2) surrogate models (also for the dataset of 1400 realizations) are shown in Figure 7-10 (right). The results are essentially identical regardless of model type, indicating that no model is necessarily more appropriate for this data set than the others. Simpler (less costly) surrogate models or sensitivity measures are appropriate for this data set. Compared to the correlation results, the surrogate model results emphasize BC2 as the most important parameter with smaller effects associated with other parameters than in the correlation results. The total index values are not significantly different from the main index values, so none of the models detect significant interaction.

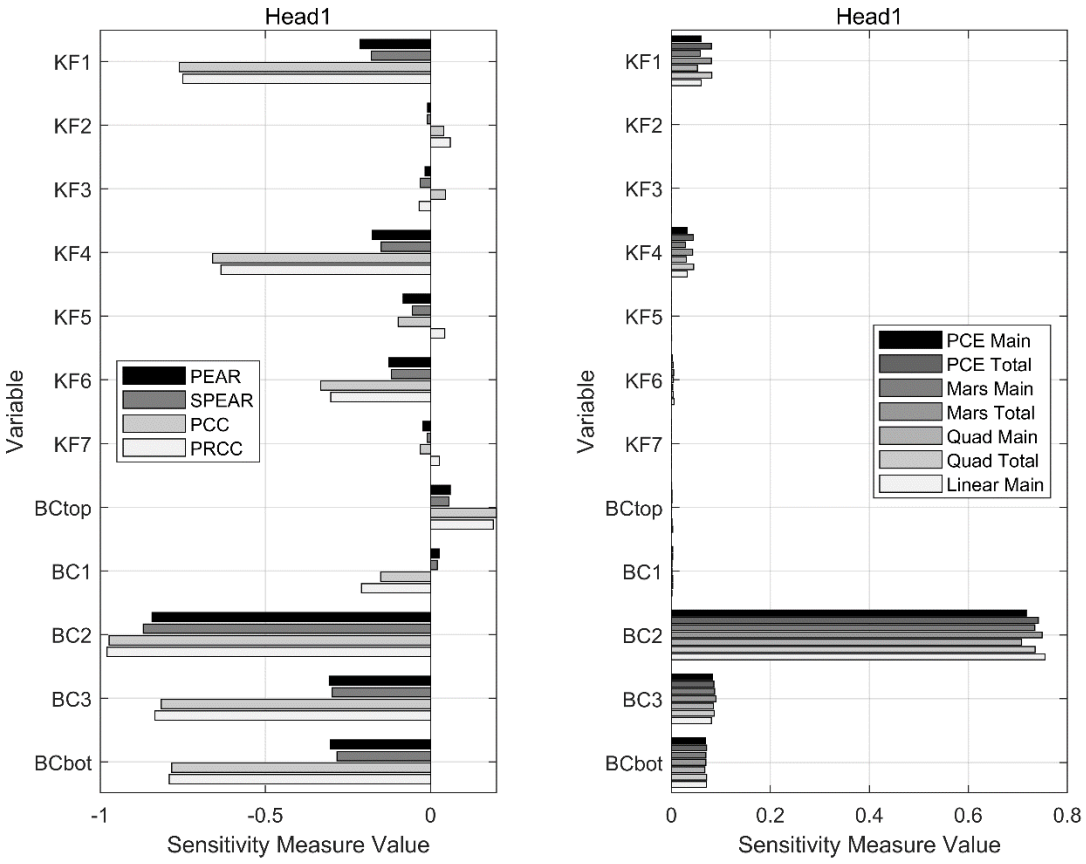

Figure 7-10. Sensitivity analysis results for hydraulic head 1 from the SNL analysis of the IBRAE reference case (1400 realizations)

This case study included simulation results for numerous increasing sample sizes, from 140 to 28,000. SNL analyzed the stability of sensitivity analysis results with increasing sample size (Figure 7-11 through Figure 7-15). While analyses on the smallest sample size could result in inconsistent rankings for parameters other than BC2, 1400 samples were found to be sufficient.

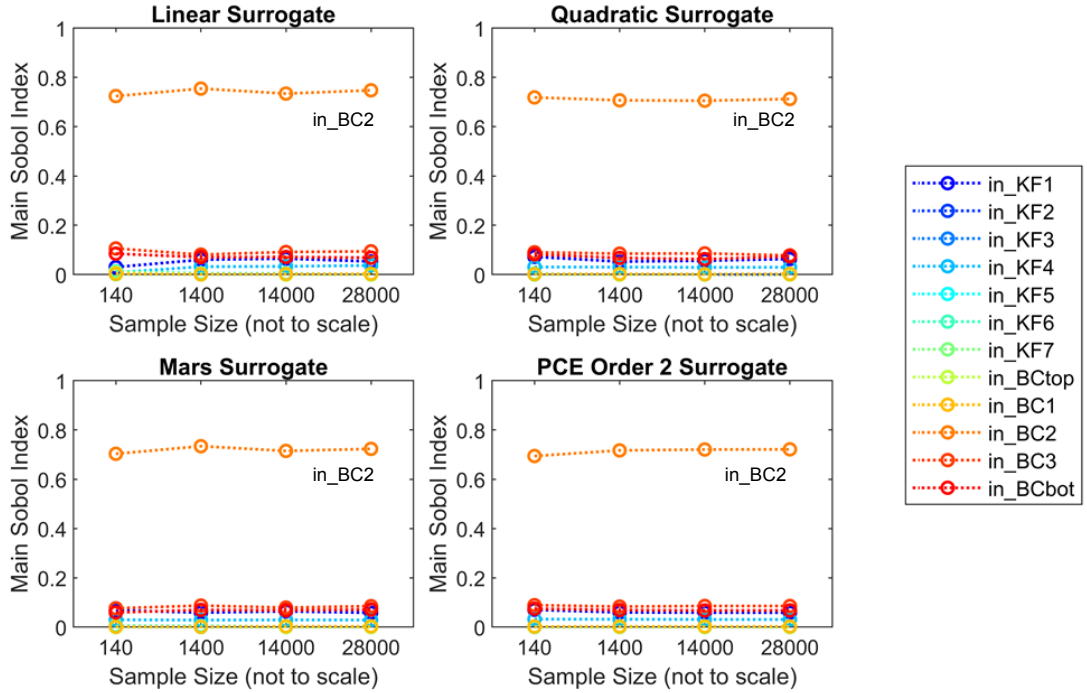

Figure 7-11. SNL results: Main Sobol' index for hydraulic head 1 obtained using increasing sample sizes (linear scale)

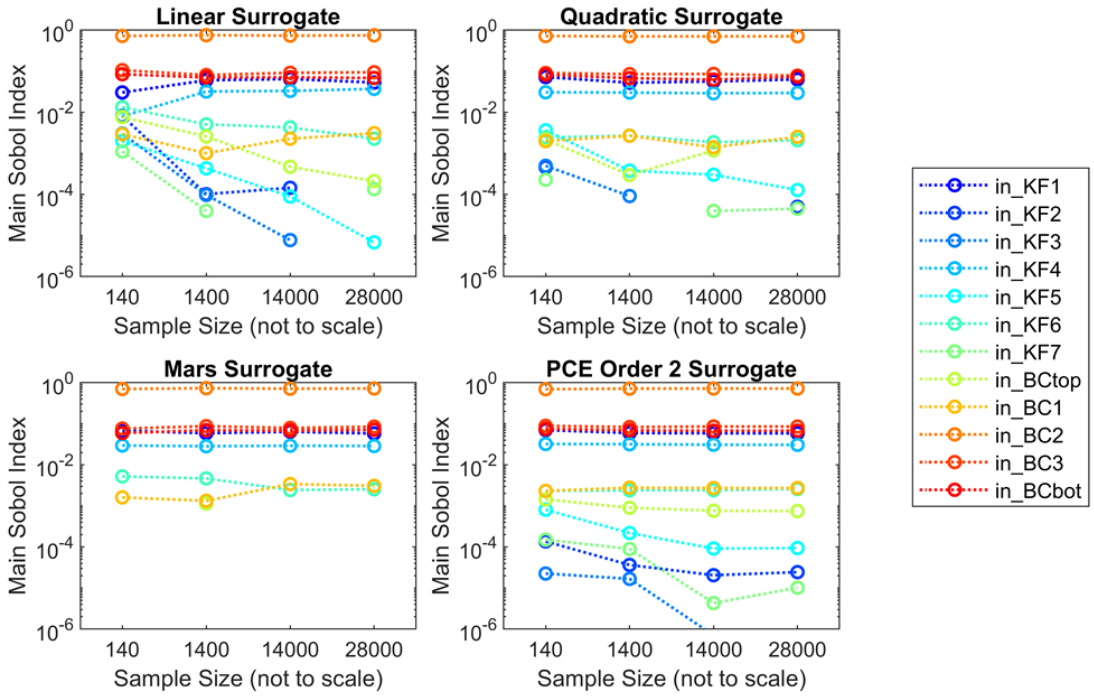

Figure 7-12. SNL results: Main Sobol' index for hydraulic head 1 obtained using increasing sample sizes (log scale)

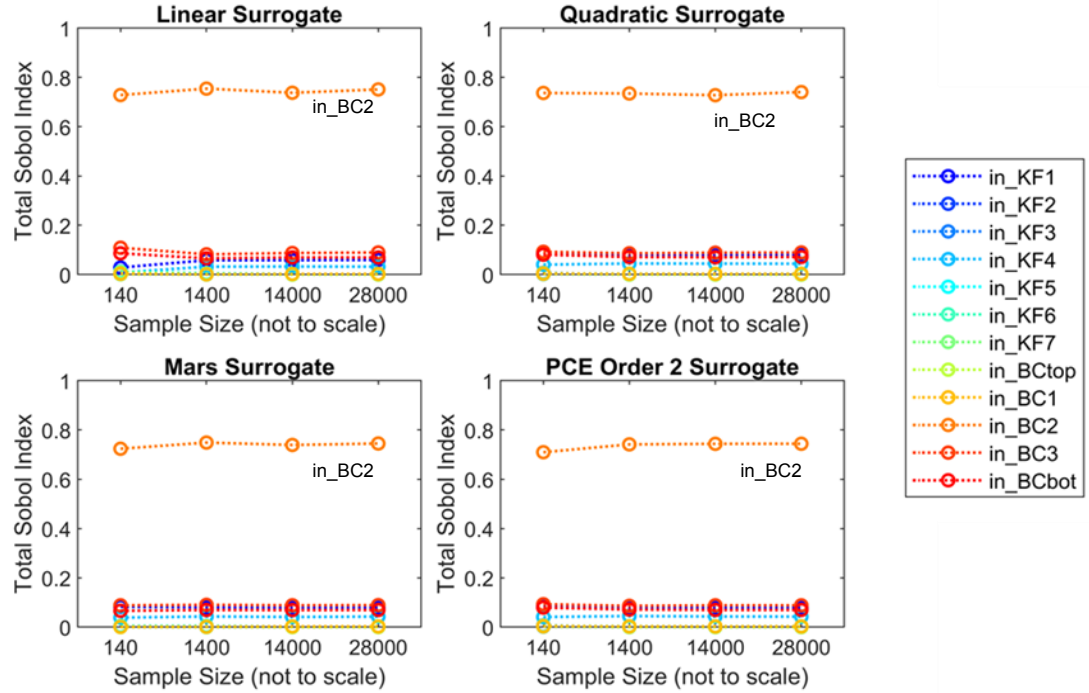

Figure 7-13. SNL results: Total Sobol' index for hydraulic head 1 obtained using increasing sample sizes (linear scale)

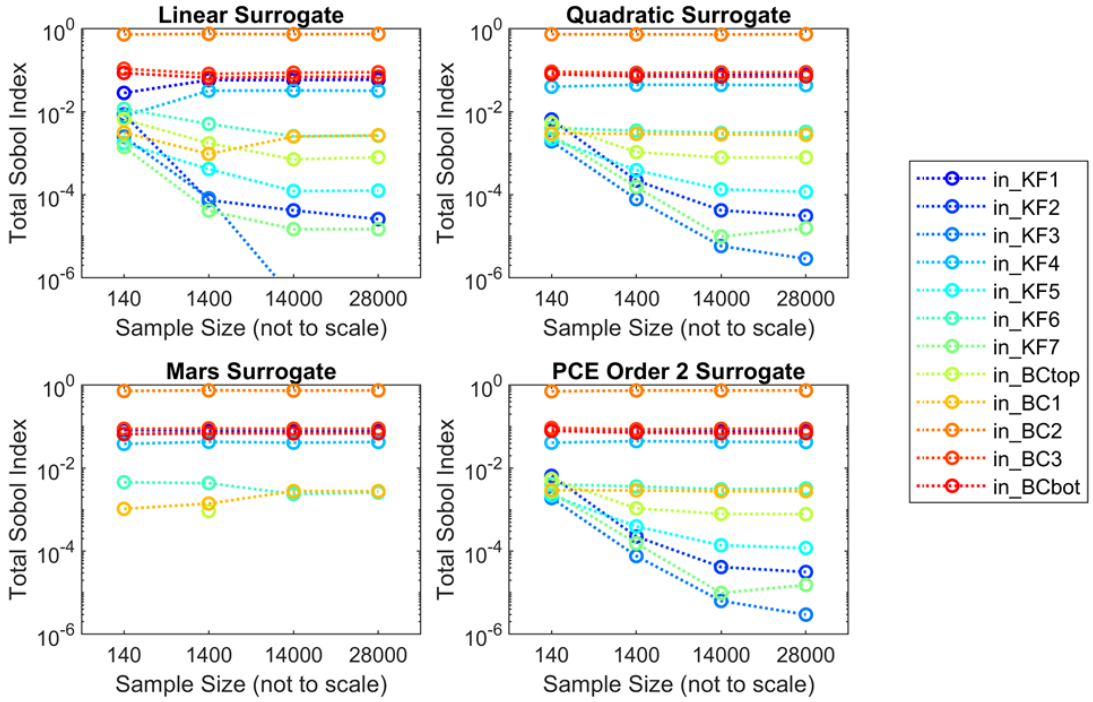

Figure 7-14. SNL results: Total Sobol' index for hydraulic head 1 obtained using increasing sample sizes (log scale)

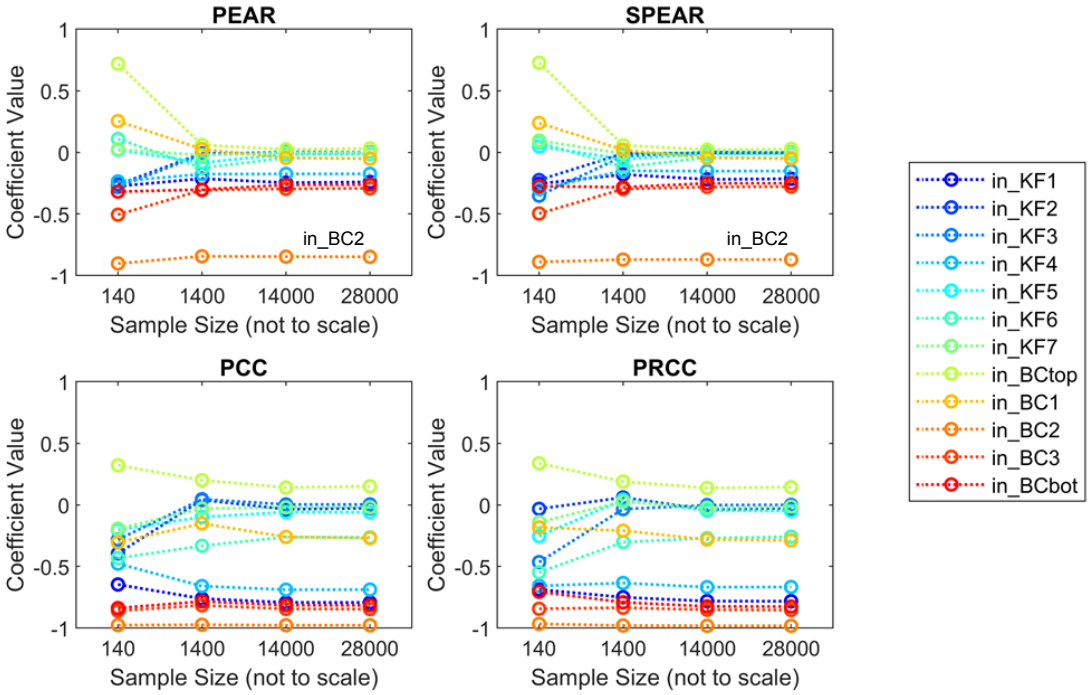

Figure 7-15. SNL results: Correlation sensitivity indices for hydraulic head 1 obtained using increasing sample sizes

7.4.3. Results from TUC

The sensitivity analysis methods used by TUC are shown in Table 7-6.

Case Name	**Description**
Sensitivity Analysis Method	CUSUNORO, Sobol'
Sensitivity measures generated	S_i first order, T_i total effect
Special considerations	Sample from Sobol' method, CUSUNORO uses independent blocks extracted from Sobol' sample
Surrogate models used	
Transformations	
QoIs addressed	All observation points show comparable behavior: head=1
Number of samples used	2×2000 , $(2 + 12) \times 2000$ for Sobol' S_i and T_i
Dataset	ibrae_geo_28000.h5
Software	MatLab/Octave: in-house software

Table 7-6. Sensitivity analysis of IBRAE Case by TU Clausthal

The analysis results are shown in Table 7-7.

Input parameter	First Order Effect S_i	Total Effect T_i	CUSUNORO
KF1	5.9%	7.9%	negative trend, minor influence
KF2			sign change with time, small
KF3			sign change with time, small
KF4	3.1%	4.3%	negative trend, minor influence
KF5			negative trend, small
KF6	0.1%	0.3%	sign change with time, small
KF7			
BCtop			positive trend, small
BC1	0.3%	0.3%	negative trend, small
BC2	72.5%	74.4%	negative trend, major influence
BC3	8.3%	8.9%	negative trend, small
BCbot	6.6%	6.9%	negative trend, small

Table 7-7. SA Results of IBRAE Case by TU Clausthal

The sum of first order effects is 0.968, the sum of first and total effects is 2: the behavior is mainly additive, the missing 4% of interactions are by pairwise interactions (otherwise the sum of first and total effects would be larger than 2).

7.4.4. Results from IBRAE

Case Name	IBRAE case
Sensitivity Analysis Method	PAWN, Sobol', RBD-FAST, Correlation and regression
Sensitivity measures generated	PAWN index, S_i first order, T_i total, Pearson correlation, Spearman rank correlation, partial correlation, partial rank correlation, regression and rank regression coefficients
Special considerations	Mean and median KS for PAWN. Number of conditioning intervals $n = 8$ for 28000, 14000,1400 samples and $n = 6$ for 140 sample.
Surrogate models used	
Transformations	
QoIs addressed	SA for each response
Number of samples used	28,000;14,000;1400;140
Dataset	ibrae_geo_28000.h5, ibrae_geo_14000.h5, ibrae_geo_1400.h5, ibrae_geo_140.h5,
Software	Python SaLib, Python in-house software

Table 7-8. Sensitivity analysis of IBRAE by IBRAE

This case was used in [96] as an illustration of the model development process with the assistance of sensitivity analysis and parameter optimization. Accordingly, the samples were initially generated specifically for Sobol' indices evaluation.

In this report PAWN, Sobol', RBD-FAST methods were applied to these samples. Correlation and regression methods were also briefly considered; however, investigations were focused on variance-based and moment-independent approaches.

The summary of the findings is provided in Table 7-9. Figure 7-16 illustrates slightly differing results for several different observation points. Figure 7-17 shows averaged by all observation points results for different sample sizes.

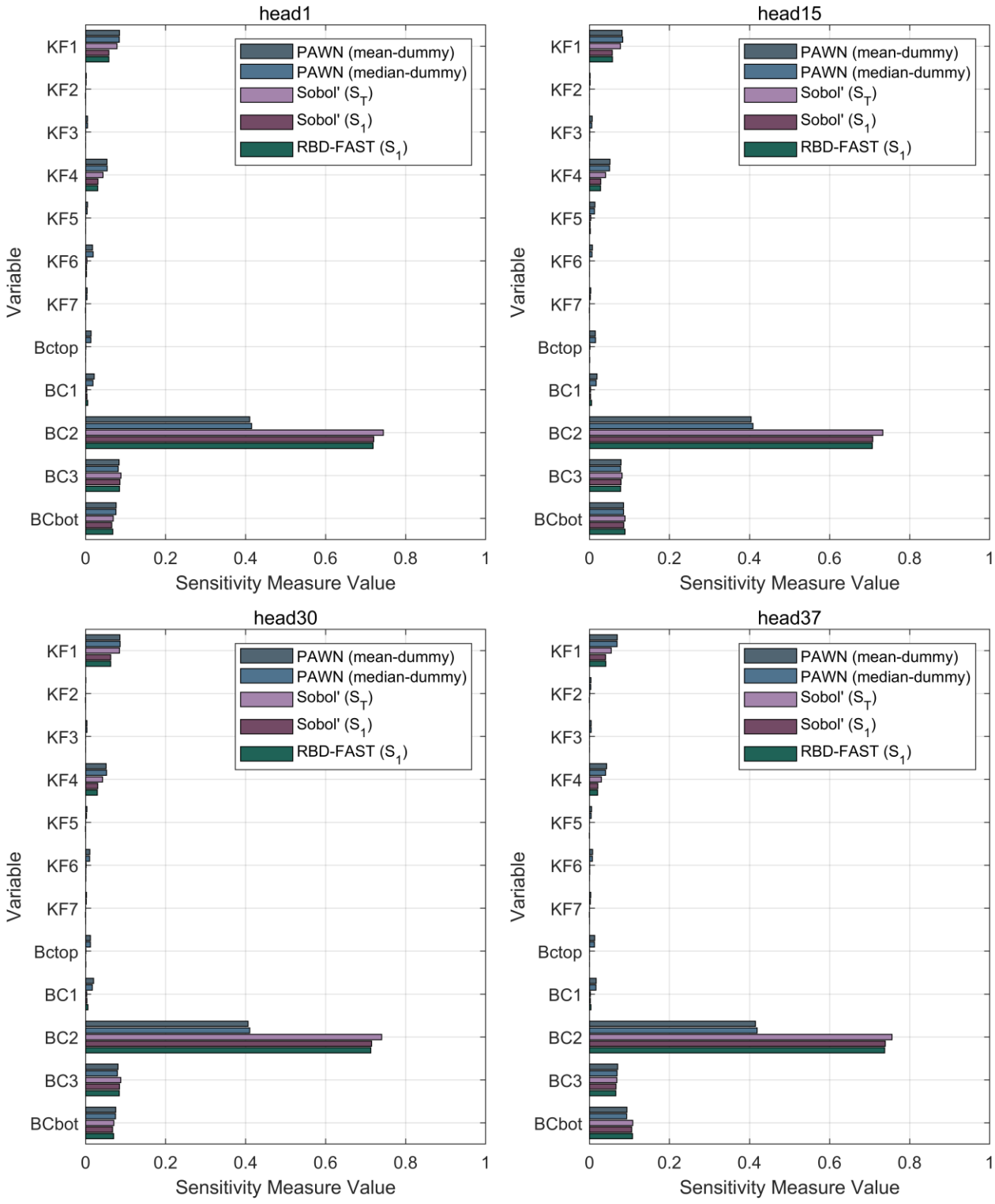

Figure 7-16. Sensitivity analysis of IBRAE Groundwater case by IBRAE: outputs head1, head15, head30, head37

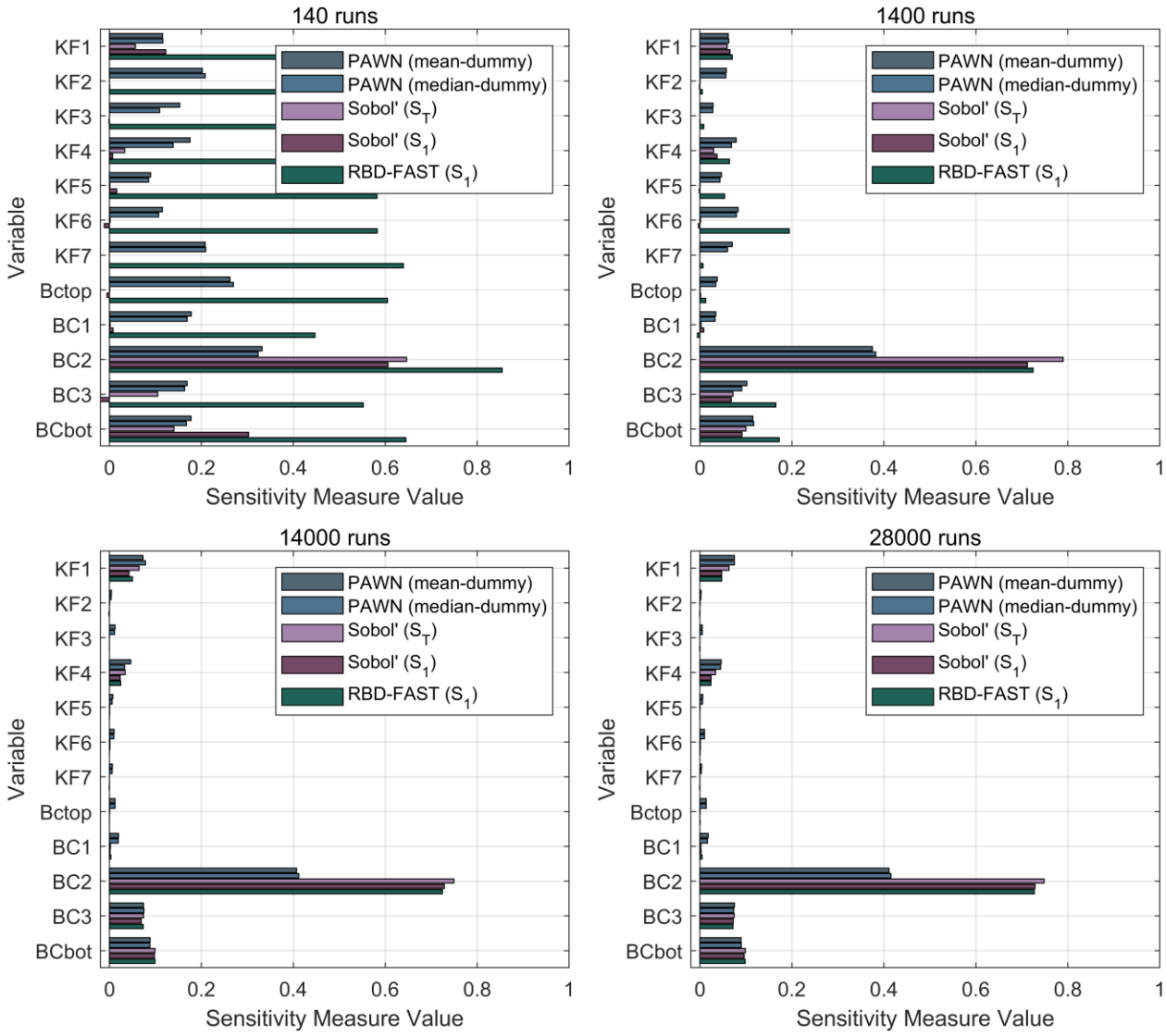

Figure 7-17. Sensitivity analysis of IBRAE Groundwater case by IBRAE: different number of model runs

Input parameter	PAWN	Sobol'	RBD-FAST	Correlation & regression methods
KF1	Small significance	Very small significance	Very small significance	Small (Pearson, Spearman, rank regression, partial correlation) or no influence (regression, partial rank correlation)
KF2	Insignificant	Insignificant	Insignificant	Small (partial correlation, partial rank correlation) or no influence (others)
KF3	Insignificant	Insignificant	Insignificant	Small (partial correlation, partial rank correlation) or no influence (others)
KF4	Small significance	Very small significance	Very small significance	Small (PEAR, SPEAR, rank regression) or no influence (others)
KF5	Insignificant	Insignificant	Very small significance	Small (PCC, PRCC) or no influence (others)
KF6	Insignificant	Insignificant	Insignificant	Small (PCC, PRCC) or no influence (others)
KF7	Insignificant	Insignificant	Insignificant	Small (PCC, PRCC) or no influence (others)
BC1	Insignificant	Insignificant	Insignificant	Small (PCC, PRCC) or no influence (others)
BC2	Most significant parameter	Most significant parameter	Most significant parameter	Falsely low influence by SRC on small sample, high by others
BC3	Small significance	Small significance	Small significance	High influence by PCC, small or no by others
BCtop	Insignificant	Insignificant	Insignificant	Falsely high influence by SRC on small sample, small (PCC, PRCC) or no influence(others).
BCbot	Small significance	Small significance	Small significance	Falsely high influence by SRC on small sample, small or no influence by others.

Table 7-9. SA Results of IBRAE case by IBRAE

7.4.5. Case Summary of Sensitivity Results

The model is relatively simple with a nearly additive and linear behavior. And results of analysis by different groups are mostly in good agreement.

BC2 is clearly the most important parameter. The physical interpretation of this result is that *BC2* defines the flow rate through the wider part of the left boundary.

The other 2 groups of parameters: *BCbot*, *BC3* and *KF1*, *KF4* (generally in the order listed) show small but detectable influence according to the whole variety of methods. Ranking for the parameters of secondary importance though may vary subtly among different observation points and methods. For example, the significance of *BCbot* is greater for the observation points that are closer to the bottom boundary of the model. Also, the PAWN method with medium statistics ranks *BC3* lower than *KF1* for approximately half of the observations.

Results on *BC1* and *BCtop* are slightly controversial: variance-based methods rate them as insignificant (with contribution less than 0.5%), while their contribution may falsely seem non-negligible by some other approaches, especially when using the smaller sample sets. Also, correlation and regression methods show the minor significance of *KF6*, while other methods rank this parameter as insignificant.

Results by CUSUNORO, RBD-FAST, PAWN methods on the smallest sample (140 model runs) allow detecting the most significant parameter *BC2* but seem not reliable enough for screening out insignificant parameters.

Sensitivity indices calculated by Sobol' method are mostly consistent (at least for the top five influential parameters) among different sample sizes and different surrogate models. Sobol' indices without the use of surrogate models, however, have visible numerical issues for samples with 140 and 1400 realizations such as slightly negative main indices S_i or conversely S_i slightly higher than total indices T_i for almost insignificant parameters. Meanwhile, results look fine for these samples with PCE surrogate model used, because the Sobol' indices are calculated analytically in this case. On the other end of the spectrum, there are no sufficient differences in the results for 14,000 realizations and 28,000 realizations. The 14,000 data set for 12 parameters seems more than enough.

This page is intentionally left blank

8. SUMMARY

8.1. The Sensitivity Analysis Exercise

This report has been produced by an informal, international working group working under the auspices of Organization for Economic Cooperation and Development (OECD)/ NEA's Integration Group for the Safety Case (IGSC, https://www.oecd-nea.org/rwm/igsc/) that is interested in uncertainty and particularly sensitivity analysis (SA) methods with a focus on performance or safety assessments for radioactive waste management. This working group has collaborated to identify test cases that are the basis of these studies. Participating organizations have provided computational simulation run results (tables of inputs and outputs from the runs) as a basis for the analyses presented in this report. The case studies to be addressed were identified based on a questionnaire developed at the 2017 Albuquerque workshop (see Section 1.1). The questionnaire aimed at categorizing the proposed case studies by their features and complexity in order to derive an appropriate strategy for the exercise described in this report. In particular, questions were asked about the following characteristics (see Chapter 3):

Phenomenological description. This includes a description of the system being modeled such as a repository, the waste forms and engineered barriers, the geology, physical and chemical processes being modeled, potential release pathways, etc.
Input characteristics. The input characteristics refer primarily to various aspects of the computational simulations, including the number of parameters being varied, the sampling methods used, and the number of runs that were generated. The input characteristics also include details about how uncertain parameters are described in the case study: are they continuous, categorical, or otherwise discrete inputs? Are there any inputs which vary in space and/or time? What probability distributions were used to represent and model the uncertain parameters? Are there correlations or other dependencies?
Output characteristics. The output characteristics focus mainly on the quantities of interest, including a description of these quantities, whether the outputs are scalar values or values varying in space and/or time, and if there are known features of the output to consider in the context of sensitivity analysis. Considerations include features such as non-linearity, non-monotonicity, discontinuities, or regime changes.

The four case studies highlighted in this report are: the GRS clay case, the SNL shale case, the Dessel case, and the IBRAE groundwater case. Some of the essentials concerning the input characteristics and the output features and behavior are summarized in Table 8-1. These four cases are less complex than the remaining three. This group plans to perform a second series of SA on the three remaining, more complex case studies in order to complete the exercise, an objective of which is to derive guidelines for the application of SA methods. The plan is to summarize the remaining cases in Volume 2, a companion report which will be issued later.

		GRS clay	SNL shale	Dessel	IBRAE groundwater
Input	Number of parameters	6	10	~20	12
	Variable type	Scalar, continuous	Scalar, continuous	Most are scalar, some will change at a given point in time. Different material-dependent values in the different domains Continuous	Scalar, continuous
	Dependencies	Inputs are independent	Inputs are independent	Inputs are independent	Inputs are independent
Output	Type	Time series	Vary in time and space	Time series	Scalar
	Specifics / complications	None known	None known	Proven non-monotonicity of diffusion coefficient Fracture flow Switch from diffusive to advective system	Response surface of the model many local minima

Table 8-1. Input and output characteristics of the four cases according to the questionnaire replies

We have surveyed sensitivity analysis methods in Chapter 2 and the detailed sensitivity analysis results on the four case studies are presented in Chapters 4-7. For each case study, multiple groups presented their results using different sensitivity analysis methods and/or different implementations of the same method. The breadth and scope of the case studies as well as the large variety of sensitivity analysis methods used (e.g. scatterplots, simple correlation coefficients, rank correlation coefficients, standardized regression coefficients, main and total effects variance-based Sobol' indices estimated by methods such as EASI, RBD-FAST, distribution-based methods such as PAWN, graphical methods like CUSUNORO, and others) provided a rich environment to study and compare results.

The case studies in this report highlighted some practical challenges such as standardization of data formats, standardization of graphical results and tables (as the inputs and outputs were extensive for some of the cases and we need better ways to quickly assess differences in results),

and the need for methods tailored to perform SA on time series data. With respect to the last item should be noted: most of the results involved some quantity over time (e.g. concentration as a function of time) and so a separate input-output sensitivity analysis was performed for each point in the time series. Some of the SA methods presented here can calculate full time series sensitivity indicators in a few seconds, but others require significant postprocessing. As many problems in this community involve time scales of thousands to millions of years, generating separate SA results per time point can add significant postprocessing computational cost to identify how the importance ranking for input parameters within a case study may change over time. The question then arises about the lessons and conclusions to be derived from such behavior. The SA results showed that changes in importance rankings were observed for the case studies, demonstrating value in this type of analysis. The issue of how to calculate sensitivity indices for time series data, and how to derive meaningful interpretations of the results, remains an open issue, possibly to be addressed by autocorrelation models, principal component analysis, or time-dependent surrogates. Development of more efficient and appropriate methods for time-dependent sensitivity analyses may be a worthwhile problem for future study.

8.2. Summary Findings for the Four Case Studies

8.2.1. GRS Case

This case study had time-varying results where the parameter importance ranking varied across time, but some of the parameters were consistently important. The diffusion coefficient in the outer clay formation (*DiffClay3*) has the highest influence on the model output. Most evaluations agree that this is valid over the total model time period. For model times below 10^7 years, the second most influential parameter is the diffusion coefficient in the inner clay formation (*DiffClay2*). For model times below 10^6 years there is no significant influence of any of the other parameters. At late times, the influence of *DiffClay2* decreases while those of the other parameters increase. Total indices reach their maximum values between 10^6 and 10^7 years. Parameter interactions seem to play a decreasing role at later times.

When comparing SA methods, the linear sensitivity measures (correlation and regression coefficients) calculated by the different partners agree more or less exactly, as all use the same algorithms. The calculated values of variance-based sensitivity indices, however, depend on the estimation method applied. First-order indices were calculated for this case using EASI/RBD-FAST, EFAST, PCE and RS-HDMR. The results are not identical but in fair agreement with each other as well as with those of the regression/correlation analysis. One item to highlight: while the EFAST and PCE results are in a fair agreement, those obtained with RS-HDMR differ significantly. This may, at least in part, be due to the fact that RS-HDMR does not really calculate the total-order indices but approximates them by summing up the first- and all second-order indices of a parameter, neglecting all higher orders of interaction. Indeed, the total-order indices calculated by RS-HDMR are always below those obtained with EFAST. Generally, it can be seen that the total-order indices of all parameters reach their maximum values between 1 million and 10 million years and decrease at later times for this case study.

8.2.2. SNL Shale Case

For this case study, there were peak ^{129}I concentration results at various locations in two layers: a sandstone aquifer layer and an aquifer limestone layer. Different parameters were important for the two layers. Peak ^{129}I results were calculated as the maximum ^{129}I concentration taken over time, so the QoIs and associated sensitivity analyses for this case study were not time-dependent. Multiple participants applied linear regression models and calculated correlation coefficients (GRS, SNL, IBRAE) and CUSUNORO curves (GRS, TUC). Additionally, SNL applied stepwise linear regression and VBD via PCE and GP surrogate modeling, TUC applied Copula distance and MIM, and IBRAE applied PAWN and RBD-FAST. Scaling, rank transformations, and log transformation were also utilized in some of the analyses.

Various SA methods agreed in ranking *pShale* (shale porosity) most important for response function 1, which was the Peak ^{129}I concentration at the observation point in the sandstone aquifer closest to the repository. The SA methods typically ranked *kSand* (sandstone permeability) and *rateWP* (mean waste package degradation rate) as the second and third most important parameters, however it was not always clear if this significance is meaningful. Sensitivity to *kSand* was noted by multiple participants to increase when rank-based methods were applied.

In contrast, there was some disagreement for response function 4, the Peak ^{129}I concentration at the observation point in the limestone aquifer closest to the repository. Most methods ranked *pShale* highest at this observation point, however PCC (from SNL) ranked *kLime* (limestone permeability) the highest. The linear methods disagreed on the second and third most important parameters, *rateSNF* (spent nuclear fuel dissolution rate) and *rateWP*. Surrogate model results based on a log transformation ranked *rateWP* as more important, whereas surrogate model results based on scaling ranked *rateSNF* higher with more substantial interaction effects. The PAWN results (IBRAE) agreed with other methods on ranking *pShale* the most important parameter but ranked the buffer permeability (*kBuffer*) as the second most important parameter. The other responses are discussed in more detail in Chapter 5.

In general, the sensitivity analyses across participants identified the same first most important parameter but differed on the importance of lower ranked parameters. This may be due to differences between methods in detecting variable significance, particularly because the QoIs in this case span many orders of magnitude. Variables of secondary importance may need to be identified by connecting less clear sensitivity analysis results with physical phenomenology, however the variables of secondary importance for this case seemed reasonable. Results also tended to differ depending on which transformations were applied to the ^{129}I concentrations. Validation or fit metrics, in the context of surrogate modeling or regression analysis, may help select the most appropriate transformation to use for sensitivity analysis.

There was also a lack of consensus between participants on the delineation between sensitivity measure values that indicate secondary sensitivity versus sensitivity measure values that indicate negligible sensitivity. This highlights a need for development of consensus methods for testing or justifying conclusions regarding lower-ranked parameters.

8.2.3. Dessel Case

In this case study, the parameter influences changed over time and the various sensitivity analysis methods tended to show this. Several SA methods were applied at various time points to show the change in sensitivity over time. Sorption in the embankment (*Kd_emb*), the time in which the infiltrating water flow ramps up (flux_increase_time) and the diffusion coefficient in the waste matrix (*D_waste_2*) clearly showed this behavior, which is expected given the setup of the model.

The sensitivity indices calculated using regression and correlation methods are in very good agreement between all participants. The ranking of parameters varied along the time range, but sorption parameters in waste and embankment (*Kd_waste_2*, *Kd_emb*), diffusion coefficient in the waste matrix (*D_waste_2*) and *flux_increase_time* are consistently identified as most important.

As in previous cases, the calculated values of variance-based sensitivity indices, depend on the estimation method applied. However, the results are fair agreement with each other as well as with those of the regression/correlation analysis.

A large part of the output variance in the time span between 680 to 800 years, remains unaccounted for by first order effects. The standard set of SA methods did not sufficiently explain this model behavior. However, it appears that all methods using surrogate models used to calculate total order effects did not capture this very well. The large-sample Sobol-analysis, performed as a benchmark for the surrogate-based methods, indicated high T_i values of *flux_increase_time* and *Kd_emb* in the period from 750 to 1000 years.

Special attention should be made to the behavior of the *D_waste_2* parameter, as discussed in §6.5, this parameter shows a non-monotonic behavior, however, this is not detected by the applied SA techniques.

8.2.4. IBRAE Case

The results from various methods (correlation coefficients, CUSUNORO, PCE-based Sobol' indices, FAST, PAWN, etc.) were consistent, pointing to one parameter dominating the results (flow rate through the widest part of the left boundary, *BC2*). The model is relatively simple with nearly additive behavior. The results of analysis by different groups are mostly in good agreement.

While boundary condition associated with the largest zone (*BC2*) is clearly the most important parameter, drainage boundary condition *BCbot*, flow rate through second-largest zone of the left boundary *BC3*, and hydraulic conductivities for two materials *KF1*, *KF4* show small but detectable influence according to the whole variety of methods. Results on smallest boundary part (*BC1*) and infiltration (*BCtop*) are slightly controversial: variance-based methods rate them as insignificant (with contribution less than 0.5%), while their contribution may falsely seem non-negligible when applying some other approaches, especially when using the small sample set (140 model runs).

The IBRAE case study included simulation results for numerous increasing sample sizes, from 140 to 28,000. Though such a large sample size is often not possible for repository cases, the relative simplicity of this groundwater model allowed for a very large sample size. This enabled participants to assess the stability of sensitivity analysis results with increasing sample size. While analyses on the smallest sample size could result in inconsistent rankings for parameters other

than, 1400 samples were found to be sufficient especially with the *BC2* use of surrogate models. Analysis of sample size sufficiency for sensitivity analysis results may not always be possible with existing simulation data depending on the sampling scheme that was used, but it could be considered when planning an uncertainty analysis.

8.3. Overall Summary

In summary, we found that the first order variance-based index estimates are now easily generated from observational data (i.e. existing data which were not generated by prescribed sampling schemes) using a variety of approaches and are one of the main SA approaches. Linear and rank correlation coefficients and regression approaches continue to be used and are informative. More advanced methods show results mostly consistent with simpler methods but there are important differences. Graphical methods such as CUSUNORO also provide additional visualization which can show influences over the range of a variable. It might be advisable to start an analysis by applying such methods.

We found consistency between the linear sensitivity measures (correlation and regression coefficients) calculated by the different partners but sometimes the variance-based sensitivity indices did not exactly agree with the linear sensitivity measures. Also, there were more differences in rankings seen across the several variance-based sensitivity indices, such as EASI/RBD-FAST, EFAST, PCE and RS-HDMR. Note that some of the variance-based methods rely on Monte Carlo sampling of the function (e.g. the simulation) using different sampling schemes while other methods rely on surrogate or metamodel approximations of the simulation. All the methods used in this study relied on fixed data sets generated by the case study owners: specialized sampling of the simulations was not possible. The various approximations and algorithm implementations used in the variance-based approaches may account for some of these differences. For surrogate methods, the surrogate type may play a role in the accuracy of the estimation of variance-based indices.

Differences in rankings were also seen depending on the use of data transformations. The appropriateness of specific transformations is not always obvious, and justification of such transformations may be an opportunity for future study. One key feature of parameters and responses in repository studies is that they vary by many orders of magnitude. In such cases, it may be appropriate to apply a log transformation. However, this may unduly weight smaller values. Validation or goodness-of-fit metrics, in the context of surrogate modeling or regression analysis, may help select the most appropriate transformation to use for sensitivity analysis or identify the best surrogate model. The appropriate use of the log (or any other) transform depends on the particular case study and usage; it remains an open issue and interpretation of SA results when transformations are used must be done carefully.

We often found that the sensitivity analyses across participants identified the same first most important parameter but differed on the importance of lower ranked parameters. Variables of secondary importance may need to be identified by connecting sensitivity analysis results with physical phenomenology. Also, most of the SA methods used in this report are factor prioritization methods: the linear sensitivity and variance-based indices perform well in identifying the most

important variables and their rankings. Some factor fixing approaches such as DGSM (derivative-based global sensitivity measure) are better suited to screening and identifying low rank variables which can be removed from the SA. We did not use the DGSM factor-fixing approach in these studies because derivatives were not available.

Parameter rankings obtained by Sobol' method are mostly consistent among different sample sizes and different surrogate models, however, there are often visible numerical issues for small sample sizes such as: negative main indices or conversely main indices slightly higher than total indices for parameters with minor or no significance, or sum of main indices more than one. This can be due to insufficient samples to accurately calculate the integrals defining the terms in the Sobol' index calculations and/or surrogate inaccuracies. When using surrogates, the methodology of Sobol' indices is applied to a model of reduced complexity: correct surrogate models are always approximations of the original models, they have higher smoothness than the original models, and typically higher order and unimportant interactions are eliminated in the surrogate models. For the PCE surrogate models, the Sobol' indices are always contained in the range [0,1] by the way the indices are constructed from the PCE coefficients. This is an attractive feature and may make PCE surrogates preferable in Sobol' index calculation. However, one always needs to be aware of surrogate inaccuracies.

With Sobol' index calculations based on surrogates, the surrogate model may induce approximation errors which may be significant. To check the reliability of the sensitivity indices based on surrogates, one can:

- Vary the number of samples and the tuning parameters of the surrogate model
- Perform the calculations with two or more methods of building surrogates and calculation methods to see if there is evidence of significant differences in parameter rankings.

We also note that the choice of sampling method is of paramount importance to the resulting accuracy of both surrogate models and values of Sobol' indices. However, the studies performed on these cases used the samples provided by the case owners and sampling method was not a focus of investigation in this report.

There is also a lack of consensus between participants on the delineation between sensitivity measure values that indicate secondary sensitivity versus sensitivity measure values that indicate negligible sensitivity. For example, some participants regarded a Sobol' main effect of 0.1 to be small whereas others might call a main effect of 0.2 as small with a threshold below 0.1 to be negligible. There also is not consensus on thresholds that should be used for secondary significance vs. negligible significant for correlation coefficients or other methods, and the thresholds for the various SA measures may differ. This highlights a need for development of consensus methods for testing or justifying conclusions regarding lower-ranked parameters.

Note that one goal of these case studies is to help the community identify best practices and lessons learned. We anticipate this report to be the first of two volumes, where the second volume will describe SA results on more complicated cases, together with a synthesis and recommendations on the application of SA in safety cases.

This page is intentionally left blank

REFERENCES

Chapter 1

[1] French, S. "Modelling, making inferences and making decisions: the roles of sensitivity analysis." Top no. 11 (2003): 229–251, doi:10.1007/BF02579043.

[2] Storlie, C., L. Swiler, J. Helton, and C. Sallaberry. "Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models." *Reliability Engineering & System Safety*, no. 94, (2009): 1735-1763, https://doi.org/10.1016/j.ress.2009.05.007.

[3] Kucherenko, S., B. Feil, , N. Shah,. and W. Mauntz. "The identification of model effective dimensions using global sensitivity analysis." *Reliability Engineering & System Safety*. Volume 96(4) (2011): 440–449 doi:10.1016/j.ress.2010.11.003

[4] Plischke, E., E. Borgonovo, and C. Smith. "Global sensitivity measures from given data." *European Journal of Operational Research*, no. 226 (2013): 536-550, https://doi.org/10.1016/j.ejor.2012.11.047.

[5] Puy, A., W. Becker, S. Lo Piano, and A. Saltelli. "The battle of total-order sensitivity estimators." https://arxiv.org/abs/2009.01147.

[6] Nuclear Energy Agency. *Confidence in the long-term safety of deep geological repositories: Its development and communication*. Paris, France: OECD/NEA; 1999.

[7] Nuclear Energy Agency. *Establishing and Communicating Confidence in the Safety of Deep Geologic Disposal: Approaches and Arguments*. Paris, France: OECD/NEA; 2002.

[8] Nuclear Energy Agency. *Post-closure Safety Case for Geological Repositories: Nature and Purpose*. NEA No. 3679. Paris, France: OECD/NEA; 2004.

[9] International Atomic Energy Agency. *Geologic Disposal of Radioactive Waste, Safety Requirements.* IAEA Safety Standards Series No. WS-R-4. Vienna, Austria: IAEA; 2006.

[10] Nuclear Energy Agency. *Safety Cases for Deep Geological Disposal of Radioactive Waste: Where Do We Stand?* Symposium Proceedings, Paris, France, 23-25 January 2007. NEA No. 6319. Paris, France: OECD/NEA; 2008.

[11] Nuclear Energy Agency. *International Experiences in Safety Case for Geological Repositories (INTESC): Outcomes of the INTESC Project*. NEA No. 6251. Paris, France: OECD/NEA; 2009.

[12] International Atomic Energy Agency. *Disposal of Radioactive Waste, Specific Safety Requirements*. IAEA Safety Standards Series No. SSR-5. Vienna, Austria: IAEA; 2011.

[13] International Atomic Energy Agency. *The Safety Case and Safety Assessment for the Disposal of Radioactive Waste, Specific Safety Guide*. IAEA Safety Standards Series No. SSG-23. Vienna, Austria: IAEA; 2012.

[14] Nuclear Energy Agency. *Methods for Safety Assessment of Geological Disposal Facilities for Radioactive Waste: Outcomes of the NEA MeSA Initiative*. NEA No. 6923. Paris, France: OECD/NEA; 2012.

[15] Nuclear Energy Agency. *The Nature and Purpose of the Post-Closure Safety Cases for Geological Repositories*. NEA/RWM/R(2013)1. Paris, France: OECD/NEA; 2013.

[16] Nuclear Energy Agency. *The Safety Case for Deep Geological Disposal of Radioactive Waste: 2013 State of the Art*. Symposium Proceedings, 7-9 October 2013, Paris, France. NEA/RWM/R(2013)9. Paris, France: OECD/NEA; 2014.

[17] Nuclear Energy Agency. IGSC Safety Case Symposium 2018 – Current Understanding and Future Direction for the Geological Disposal of Radioactive Waste, October 10-11, 2018, Rotterdam. Book of Abstracts. OECD/NEA; 2018.
[18] Howard, B.A., M.B. Crawford, D.A. Galson, and M.G. Marietta. "Regulatory basis for the Waste Isolation Pilot Plant performance assessment." *Reliability Engineering & System Safety* no, 69 (2000): 109–27, doi:10.1016/S0951-8320(00)00028-4.
[19] Rechard, R.P., T.A. Cotton, and M.D. Voegele. "Site selection and regulatory basis for the Yucca Mountain disposal system for spent nuclear fuel and high-level radioactive waste." *Reliability Engineering & System Safety* no. 122 (2014): 7–31,doi:10.1016/j.ress.2013.06.021.
[20] Freeze G.A., M.D. Voegele, P. Vaughn, J. Prouty, W.M. Nutt, E.L. Hardin et al. 2013. *Generic Deep Geologic Disposal Safety Case (SAND2013-0974P)*. FCRD-UFD-2012-000146.. Sandia National Laboratories, Albuquerque, NM.
[21] Freeze G.A., E.R. Stein, L.L. Price, R.J. MacKinnon, J.B. Tillman. 2016. *Deep Borehole Disposal Safety Analysis, (SAND2016-10949R)*. FCRD-UFD-2016-000075.. Sandia National Laboratories, Albuquerque, NM.
[22] Saltelli, A., K. Chan, E.M. Scott, and editors. 2000. *Sensitivity Analysis: Gauging the Worth of Scientific Models*. John Wiley & Sons, New York, NY..
[23] Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto. 2004. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons New York, NY.
[24] Kaplan, S., and B.J. Garrick. "On the Quantitative Definition of Risk." *Risk Analysis* no. 1 (1981): 11–27, doi:10.1111/j.1539-6924.1981.tb01350.x.
[25] Apostolakis, G. "The concept of probability in safety assessments of technological systems." *Science* no. 250 (1990): 1359–1364, doi:10.1126/science.2255906.
[26] Helton, J.C. "Treatment of Uncertainty in Performance Assessments for Complex Systems." *Risk Analysis* no. 14 (1994): 483–511, doi:10.1111/j.1539-6924.1994.tb00266.x.
[27] Paté-Cornell, M.E. "Uncertainties in risk analysis: Six levels of treatment." *Reliability Engineering & System Safety* no. 54 (1996): 95–111, doi:10.1016/S0951-8320(96)00067-1.
[28] Helton, J.C. "Quantification of margins and uncertainties: Conceptual and computational basis." *Reliability Engineering & System Safety* no. 96 (2011): 976–1013, doi:10.1016/j.ress.2011.03.017.
[29] Röhlig KJ, E. Plischke, D.A. Becker, E.R. Stein, J. Govaerts, B.J. Debusschere, L. Koskinen, P. Kupiainen, C.D. Leigh, P. Mariner, O. Nummi, B. Pastina, S.D. Sevougian, S. Spießl, L.P. Swiler, E. Weetjens, and T. Zeitler. 2019 "Sensitivity Analysis in Performance Assessment: Towards a Joint Approach." *The Integration Group for the Safety Case (IGSC) Symposium: Current Understanding and Future Direction for the Geological Disposal of Radioactive Waste*, October 2018, Rotterdam, NL.
[30] Stein E.R., L.P. Swiler, and D.S. Sevougian. 2019. "GDSA framework: Comparison of sensitivity analysis methods applied to a reference case repository in shale." *The Integration Group for the Safety Case (IGSC) Symposium: Current Understanding and Future Direction for the Geological Disposal of Radioactive Waste*, October 2018, Rotterdam, NL.
[31] Mariner, P., E.R. Stein, S.D. Sevougian, and E. Basurto. 2019. "Performance assessment of a reference nuclear waste repository in fractured crystalline rock: Connected vs.

unconnected fracture networks." *The Integration Group for the Safety Case (IGSC) Symposium: Current Understanding and Future Direction for the Geological Disposal of Radioactive Waste,* October 2018, Rotterdam, NL.
[32] Becker, D-A. 2019. "Sensitivity Analysis in Repository Performance Assessment: Findings from an International Exercise." *Proceedings of the International High-Level Radioactive Waste Management Conference* (IHLRWM), April2019, Knoxville, TN, American Nuclear Society.
[33] Stein, E.R., L.P. Swiler, and S.D. Sevougian. 2019. "Methods of Sensitivity Analysis in Geologic Disposal Safety Assessment (GDSA)." *Proceedings of the International High-Level Radioactive Waste Management Conference* (IHLRWM) April 2019, Knoxville, TN, American Nuclear Society.
[34] Spießl, S., and S. Kucerenko. "Noise-adjusted higher-order sensitivity indices of final repository models." *EGU General Assembly*, April 2019, Vienna, AT.
[35] Becker, D.A., B.J. Debusschere, J. Govaerts, L. Koskinen, S. Kucherenko, P. Kupiainen, C.D. Leigh, P. Mariner, B. Pastina, E. Plischke, K.J. Röhlig, S.D. Sevougian, S. Spießl, E.R. Stein, L.P. Swiler, E. Weetjens, and T. Zeitler. "A Joint Exercise on Sensitivity Analysis in Repository Performance Assessment." *Ninth International Conference on Sensitivity Analysis of Model Output*, October 2019, Barcelona, ESP, Universitat Oberta de Catalunya.
[36] Spießl, S,, and S. Kucherenko. "Estimation of higher-order sensitivity indices of final repository models by means of the RS-HDMR and BSPCE metamodeling approaches." *Ninth International Conference on Sensitivity Analysis of Model Output*, October 2019, Barcelona, ESP, Universitat Oberta de Catalunya.
[37] aSavaleva, S., V. Svitelman, P. Blionov, D. Valetov, and G. yNeyzhayev. "Coupling of sensitivity analysis and model calibration in radioactive waste disposal safety assessment." *Ninth International Conference on Sensitivity Analysis of Model Output*, October 2019, Barcelona, ESP, Universitat Oberta de Catalunya
[38] Plischke, E. "Has the Spell Been Broken? Estimating Variance-Based Sensitivity via Nearest Neighbors." *Ninth International Conference on Sensitivity Analysis of Model Output*, October 2019, Barcelona, ESP, Universitat Oberta de Catalunya

Chapter 2

[39] Saltelli, A., K. Chan, E.M. Scott, and editors. 2000. *Sensitivity Analysis: Gauging the Worth of Scientific Models*. John Wiley & Sons, New York, NY.
[40] Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto. 2004. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. John Wiley & Sons, New York, NY.
[41] Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, et al. 2008. *Global Sensitivity Analysis: The Primer*. John Wiley & Sons, New York, NY.
[42] Borgonovo, E. 2017. *Sensitivity Analysis: An Introduction for the Management Scientist.* Springer International Publishing.
[43] Ghanem, R., D, Higdon, H, Owhadi, and editors. 2017. *Handbook of Uncertainty Quantification.* Springer International Publishing, Switzerland.
[44] Haldar, A., and S. Mahadevan. 2000. *Probability, reliability, and statistical methods in engineering design.* John Wiley & Sons, New York, NY.

[45] Ghanem, R.G., and P.D. Spanos. 2002. *Stochastic finite elements: a spectral approach*. Springer Verlag, New York, NY.
[46] Xiu, D. 2010. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, Princeton, NJ.
[47] Borgonovo, E., and E. Plischke. "Sensitivity analysis: a review of recent advances." *European Journal of Operational Research* no. 248 (2016): 869–887.
[48] Kuhlmann, S., E. Plischke, K-J. Röhlig, and D-A. Becker. "Sensitivity analysis: Theory and practical application in safety cases." *The Safety Case for Deep Geological Disposal of Radioactive Waste: 2013 State of the Art.* Symposium Proceedings, Paris, France: 2013, p. 169–76.
[49] Saltelli, A., and S. Tarantola. "On the Relative Importance of Input Factors in Mathematical Models." *Journal of the American Statistical Association* no. 97 (2002): 702–709, doi:10.1198/016214502388618447.
[50] Owen, A.B., and C. Prieur. "On Shapley Value for Measuring Importance of Dependent Inputs." *SIAM/ASA J Uncertainty Quantification* no. 5 (2017): 986–1002, doi:10.1137/16M1097717.
[51] Ross, S.M. 2017. *Introductory Statistics*. Academic Press, London, UK.
[52] Helton, J.C. "Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal." *Reliability Engineering & System Safety* no. 42 (1993): 327–367, doi:10.1016/0951-8320(93)90097-I.
[53] Iman, R.L., M.J. Shortencarier, and J.D. Johnson. 1985. *FORTRAN 77 program and user's guide for the calculation of partial correlation and standardized regression coefficients (SAND85-0044).* NUREG/CR-4122. Sandia National Laboratories, Albuquerque, NM.
[54] Seber, G.A.F., and A.J. Lee. 2012. *Linear Regression Analysis 2nd edition.* John Wiley & Sons, Hoboken, NJ.
[55] Seber, G.A.F., and C.J. Wild. 2003. *Nonlinear Regression*. Wiley-Interscience, Hoboken, NJ.
[56] Saltelli, A., and R. Bolado. "An alternative way to compute Fourier amplitude sensitivity test (FAST)." *Computational Statistics & Data Analysis* no. 26 (1998): 445–460, doi:10.1016/S0167-9473(97)00043-1.
[57] Cukier, R.I., H.B. Levine, and K.E. Shuler. "Nonlinear sensitivity analysis of multiparameter model systems." *Journal of Computational Physics* no. 26 (1978): 1–42, doi:10.1016/0021-9991(78)90097-9.
[58] McRae, G.J., J.W. Tilden, and J.H. Seinfeld. "Global sensitivity analysis—a computational implementation of the Fourier Amplitude Sensitivity Test (FAST)." *Computers & Chemical Engineering* no. 6 (1982): 15–25, doi:10.1016/0098-1354(82)80003-3.
[59] Tarantola, S., D. Gatelli, and T.A. Mara. "Random balance designs for the estimation of first order global sensitivity indices." *Reliability Engineering & System Safety* no 91 (2006): 717–727, doi:10.1016/j.ress.2005.06.003.
[60] Plischke, E. "An effective algorithm for computing global sensitivity indices (EASI)." *Reliability Engineering & System Safety* no. 95 (2010): 354–360, doi:10.1016/j.ress.2009.11.005.

[61] Plischke, E. “How to compute variance-based sensitivity indicators with your spreadsheet software.” *Environmental Modelling & Software* no. 35 (2012): 188–191, doi:10.1016/j.envsoft.2012.03.004.
[62] Sudret, B. “Global sensitivity analysis using polynomial chaos expansions.” *Reliability Engineering & System Safety* no. 93 (2008): 964–979, doi:10.1016/j.ress.2007.04.002.
[63] Shao, Q., A. Younes, M. Fahs, and T.A. Mara. “Bayesian sparse polynomial chaos expansion for global sensitivity analysis.” *Computer Methods in Applied Mechanics and Engineering* no. 318 (2017): 474–496, doi:10.1016/j.cma.2017.01.033.
[64] Rasmussen, C.E., and C.K.I. Williams. 2006. *Gaussian processes for machine learning*. MIT Press, Cambridge, MA.
[65] Kleijnen, J.P.C. “Kriging metamodeling in simulation: A review.” *European Journal of Operational Research* no. 192 (2009): 707–716, doi:10.1016/j.ejor.2007.10.013.
[66] Gramacy, R.B. 2020. *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences 1st edition.* Chapman and Hall/CRC, New York, NY.
[67] Li, C., and S. Mahadevan. “An efficient modularized sample-based method to estimate the first-order Sobol′ index.” *Reliability Engineering & System Safety* no 153 (2016): 110–121, doi:10.1016/j.ress.2016.04.012.
[68] Plischke, E., and E. Borgonovo. “What about totals? Alternative approaches to factor fixing.” *Safety, Reliability and Risk Analysis: Beyond the Horizon-Proceedings of the European Safety and Reliability Conference*, ESREL, CRC Press. (2013): 3339–3344, doi:10.1201/b15938-507.
[69] Dhar, S.S., W. Bergsma, and A. Dassios. “Testing Independence of Covariates and Errors in Non-parametric Regression: Test of independence.” *Scandinavian Journal of Statistics* no. 45 (2018): 421–443, doi:10.1111/sjos.12301.
[70] Plischke, E., E. Borgonovo, and C.L. Smith. „Global sensitivity measures from given data.” *European Journal of Operational Research* no. 226 (2013): 536–550, doi:10.1016/j.ejor.2012.11.047.
[71] Pianosi, F., and T. Wagener. “A simple and efficient method for global sensitivity analysis based on cumulative distribution functions.” *Environmental Modelling & Software* no. 67 (2015): 1–11, doi:10.1016/j.envsoft.2015.01.004.
[72] Pianosi, F, and T. Wagener. “Distribution-based sensitivity analysis from a generic input-output sample.” *Environmental Modelling & Software* no. 108 (2018): 197-207,doi: 10.1016/j.envsoft.2018.07.019
[73] Borgonovo, E., G.B. Hazen, V.R.R. Jose, and E. Plischke. “Probabilistic sensitivity measures as information value.” *European Journal of Operational Research* no. 289 (2021): 595–610, doi:10.1016/j.ejor.2020.07.010.
[74] Plischke, E., and E. Borgonovo. “Fighting the Curse of Sparsity: Probabilistic Sensitivity Measures From Cumulative Distribution Functions.” *Risk Analysis* no. 40(12), (2020): p. 2639-2660, doi:https://doi.org/10.1111/risa.13571.
[75] Gamboa, F., T. Klein, and A. Lagnoux. “Sensitivity analysis based on Cramér–von Mises distance.” *SIAM-ASA Journal on Uncertainty Quantification* no. 6 (2018): 522–548, doi:10.1137/15M1025621.
[76] Plischke, E. “An adaptive correlation ratio method using the cumulative sum of the reordered output.” *Reliability Engineering & System Safety* no. 107 (2012): 149–156, doi:10.1016/j.ress.2011.12.007.

[77] Plischke, E., and E. Borgonovo. "Copula theory and probabilistic sensitivity analysis: Is there a connection?" *European Journal of Operational Research* no. 277 (2019): 1046–1059, doi:10.1016/j.ejor.2019.03.034.
[78] Sobol' I.M., and S. Kucherenko. "A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices." *Computer Physics Communications* no. 181 (2010): 1212–1217, doi:10.1016/j.cpc.2010.03.006.

Chapter 4

[79] Ruebel, A., D-A. Becker, E. Fein, A. Ionescu, T. Lauke, and J. Moenig, et al. 2010. *Development of performance assessment methodologies (GRS-259).* Braunschweig: Gesellschaft fuer Anlagen-und Reaktorsicherheit mbH (GRS).
[80] Spiessl, S., and D-A. Becker. 2017. *Investigation of modern methods of probabilistic sensitivity analysis of final repository performance assessment models (MOSEL) (GRS-412).* Braunschweig: Gesellschaft fuer Anlagen-und Reaktorsicherheit (GRS) gGmbH.
[81] Jobmann, M., P. Amelung, D. Billaux, M. Polster, H. Schmidt, and L. Uhlig. 2007. *Untersuchungen zur sicherheitstechnischen Auslegung eines generischen Endlagers im Tonstein in Deutschland – GENESIS – Anlagenband Geologie der Referenzregionen im Tonstein.* DBE Technology.
[82] Plischke, E. "An adaptive correlation ratio method using the cumulative sum of the reordered output." *Reliability Engineering & System Safety* no. 107 (2012): 149–156, doi:10.1016/j.ress.2011.12.007.
[83] Zuniga, M., S. Kucherenko, and N. Shah. Metamodelling with independent and dependent inputs." *Computer Physics Communications*, Volume 184, Issue 6. (2013): 1570-1580. https://doi.org/10.1016/j.cpc.2013.02.005.
[84] Spiessl, S. M., S. Kucherenko, D.-A. Becker, and O. Zaccheus, "Higher-order sensitivity analysis of a final repository model with discontinuous behavior using the RS-HDMR meta-modeling approach." *Reliability Engineering & System Safety,* 187 (2019): 149-158, https://doi.org/10.1016/j.ress.2018.12.004.

Chapter 5

[85] Mariner, P.E., E.R. Stein, J.M. Frederick, S.D. Sevougian, and G.E. Hammond. 2017. *Advances in Geologic Disposal System Modeling and Shale Reference Cases* (*SAND2017-10304 R).* SFWD-SFWST-2017-000044. Sandia National Laboratories, Albuquerque, NM.
[86] Stein, E. R., L. P. Swiler, and S. D. Sevougian. Methods of Sensitivity Analysis in Geologic Disposal Safety Assessment (GDSA) Framework. Paper presented at the *International High-Level Radioactive Waste Management*, Knoxville, TN, April 14-18, 2019.
[87] Swiler, L.P., Helton, J.C., Basurto, E., Brooks, D.M., Mariner, P.E., Moore, L.M., Mohanty, S., Sevougian, S.D., and E.R. Stein. 2019. *Status Report on Uncertainty*

Quantification and Sensitivity Analysis Tools in the Geologic Disposal Safety Assessment (GDSA) Framework (SAND2019-13835R). M2SF-19SN010304031. Sandia National Laboratories, Albuquerque, NM.

Chapter 6

[88] Nationale instelling voor radioactief afval en verrijkte splijtstoffen. Hoofdstuk 14 van het veiligheidsrapport voor de oppervlaktebergingsinrichting van categorie A-afval te Dessel: Veiligheidsevaluatie – Langetermijnveiligheid. NIROND-TR 2011-14 Versie 3. Brussel, Belgium: NIRAS; 2019.

[89] Nationale instelling voor radioactief afval en verrijkte splijtstoffen. Hoofdstuk 5 van het veiligheidsrapport voor de oppervlaktebergingsinrichting van categorie A-afval te Dessel: Kennis van de fenomenologie van de kunstmatige barrières in hun omgeving. NIROND-TR 2011-05 Versie 3. Brussel, Belgium: NIRAS; 2019.

[90] Perko, J., J. Govaerts, and D. Jacques. 2018. *Expected evolution scenario for the near surface radioactive waste disposal facility at Dessel, Belgium (ER-0336, OD-269)*. Mol, Belgium, SCK•CEN.

[91] Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index." *Computer Physics Communications* no. 181 (2010): 259–270, doi:10.1016/j.cpc.2009.09.018.

[92] Plischke, E. "An effective algorithm for computing global sensitivity indices (EASI)." *Reliability Engineering & System Safety* no. 95 (2010): 354–360, doi:10.1016/j.ress.2009.11.005.

[93] Adams, B.M., M.S. Ebeida, M.S. Eldred, G. Geraci, J.D. Jakeman, and K.A. Maupin, et al. 2018. *Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.8 User's Manual (SAND2014-4253*). Sandia National Laboratories, Albuquerque, NM.

[94] Plischke E. "An adaptive correlation ratio method using the cumulative sum of the reordered output." *Reliability Engineering & System Safety* no. 107, 2012, doi: 10.1016/j.ress.2011.12.007.

Chapter 7

[95] Bagaev, D., F. Grigoriev, I. Kapyrin, I. Konshin, V. Kramarenko, and A. Plenkin. "Improving Parallel Efficiency of a Complex Hydrogeological Problem Simulation in GeRa." *Supercomputing: 5th Russian Supercomputing Days*, RuSCDays 2019, September, 2019, Moscow, RU, Revised Selected Papers, Springer, Cham (2019): 265–277, doi:10.1007/978-3-030-36592-9_22.

[96] Neuvazhaev, G., A. Rastorguev, O. Morozov, I. Kapyrin, and F. Grigorev. "3D hydrogeological modeling of Deep Geological Disposal in the Nizhnekansky Rock massif." *EGU General Assembly Conference*. Abstracts no. EGU2020-21509. (2020), doi:10.5194/egusphere-egu2020-21509.

[97] Saveleva E, Svitelman V, Blinov P, Valetov D. "Sensitivity analysis and model calibration as a part of the model development process in radioactive waste disposal safety assessment." *Reliability Engineering & System Safety* no. 210(11), 2021. doi:10.1016/j.ress.2021.107521.

[98] Saltelli, A., P. Annoni, I. Azzini, F. Campolongo, M. Ratto, and S. Tarantola. "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index." *Computer Physics Communications* no.181 (2010): 259-270, doi:10.1016/j.cpc.2009.09.018.

[99] Herman, J., and W. Usher. "SALib: An open-source Python library for Sensitivity Analysis." *Journal of Open Source Software* no. 2 (2017): 97, doi:10.21105/joss.00097.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
T. Portone	1463	tporton@sandia.gov
L. P. Swiler	1463	lpswile@sandia.gov
D. Z. Turner	1463	dzturne@sandia.gov
A.C. Eckert	1544	acecker@sandia.gov
E. Basurto	8844	ebasurt@sandia.gov
T. LaForce	8844	tlaforc@sandia.gov
M. Nole	8844	mnole@sandia.gov
P. Mariner	8844	pmarine@sandia.gov
E. Stein	8844	ergiamb@sandia.gov
D. M. Brooks	8853	dbrooks@sandia.gov
M. J. Starr	8853	mjstarr@sandia.gov
Technical Library	01977	sanddocs@sandia.gov

Email—External [redacted]

Name	Company Email Address	Company Name
D-A. Becker	Dirk-Alexander.Becker@grs.de	GRS
K.-J. Rohlig	klaus.roehlig@tu-clausthal.de	TUC Clausthal University of Technology
E. Plischke	elmar.plischke@tu-clausthal.de	TUC Clausthal University of Technology
S. M Spiessl	Sabine.Spiessl@grs.de	GRS
J. Govearts	joan.govaerts@sckcen.be	SCK-CEN
L. Koskinen	Lasse.Koskinen@Posiva.fi	Posiva
P. Kupiainen	Pekka.Kupiainen@posiva.fi	Posiva
V. Svitelman	svitelman@ibrae.ac.ru	IBRAE
S. Kurcherenko	s.kucherenko@imperial.ac.uk	Imperial College
S. Mekki	Soufiane.MEKKI@oecd-nea.org	OECD Nuclear Energy Agency

This page left blank

This page left blank

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.