Deliverable 4.7

Report describing numerical improvement and developments and their application to treat uncertainty when dealing with coupled processes

Work Package 4: Development/improvement of numerical methods & tools for modelling coupled processes (DONUT)

The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847593.

Document information

Project Acronym	**EURAD**
Project Title	**European Joint Programme on Radioactive Waste Management**
Project Type	**European Joint Programme (EJP)**
EC grant agreement No.	**847593**
Project starting / end date	**1st June 2019 – 30 May 2024**
Work Package No.	**4**
Work Package Title	**Development and Improvement Of Numerical methods and Tools for modelling coupled processes**
Work Package Acronym	**DONUT**
Deliverable No.	**D4.7**
Deliverable Title	**Report describing numerical improvement and developments and their application to treat uncertainty when dealing with coupled processes.**
Lead Beneficiary	**ANDRA**
Contractual Delivery Date	**M52**
Actual Delivery Date	**16/05/2024**
Type	**Report**
Dissemination level	**PU**
Editors	**Attila Baksay (TS Enercon) and Francis Claret (BRGM) compiled a multi authors contribution. For each contribution compiled in this document an author list is associated**

To be cited as:

Baksay A., Claret, F. (2024). Report describing numerical improvement and developments and their application to treat uncertainty when dealing with coupled processes. Final version as of 16.05.2024 of deliverable D4.7 of the HORIZON 2020 project EURAD. EC Grant agreement no: 847593.

Disclaimer

Acknowledgement

This document is a deliverable of the European Joint Programme on Radioactive Waste Management (EURAD). EURAD has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847593.

Status of deliverable		
	By	Date
Delivered (Lead Beneficiary)	Andra	16.05.2024
Verified (WP Leader)	F. Claret	15.05.2024
Reviewed (Reviewers)	D. Becker, D. Diaconu, Dr. I. Ivanov, J. Grupa, Dr. Z. Lahodová	06.05.2024
Approved (PMO)	B. Grambow	15.05.2024
Submitted to EC	Andra (Coordinator)	17.05.2024

Table of content

Executive Summary

The main objective of the Development/improvement of numerical methods and tools for modelling coupled processes (DONUT) WP was to improve numerical methods and tools for modelling multi-physical Thermo-Hydro-Mechanical-Chemical coupled processes (THMC) occurring in radioactive waste disposal is a major and permanent issue to support its optimization of design.

Considering the complexities and nonlinearities inherent in THMC process simulations, Task 4 of DONUT WP concentrated on advancing innovative numerical methods to address uncertainty and sensitivity analysis in complex coupled models, along with developing mathematical representations of uncertain parameters for coupled processes.

This document is a compilation of the results of the work carried out in the framework of DONUT WP4. All participating institutions have contributed to this document by briefly describing their efforts to address uncertainties and sensitivity analysis. The short introduction of the efforts is presented here and the detailed reports are presented in the document chapters.

Task 4 saw collaborative efforts between various institutions. One collaboration involved the Technical University of Liberec and the Institute of Geonics of the Czech Academy of Science, which utilized parallel Bayesian inversion and multilevel Monte Carlo method (MLMC) libraries to solve an inverse problem related to a hydro-mechanical model. They integrated these methods into safety studies for near-field transport, enhancing their applicability through the addition of a neural network for numerical homogenization of discrete fractures in 2D models.

Another collaboration by UJV, GRS, and HZDR focused on complex uncertainty treatment and sensitivity analysis. They conducted coordinated data collection, uncertainty quantification, and sensitivity analysis methods applied to geochemical and transport applications. Data on the impact of geochemical conditions on the transport of selected radionuclides collected form an experiment to use it in a defined RN model system. The tool for probabilistic uncertainty and sensitivity analysis for the flow and transport simulations was developed. To model different scale heterogeneities a development was used that was relying on actual rock measurements.

SCK CEN's efforts cantered on accelerating reactive transport model (RTM) simulations using machine learning (ML), investigating the replacement of RTMs with ML methods for nonlinear regression. Their findings showed significant speedups of about 25-30 times against single-threaded RTM calculations, while maintaining high accuracy.

TS Enercon and WSP Golder collaborated on developing a coupled physical-chemical modelling approach integral to long-term Total System Performance Assessment (TSPA). Their aim was to enhance understanding of how coupled processes affect repository near-field evolution and to provide tools for sensitivity and uncertainty analysis of the TSPA model.

UDC from Spain focused on innovative numerical methods for adjoint-state methods, sensitivity analyses, and uncertainty quantification. They explored both continuous and discrete adjoint state methods, along with various sensitivity analysis techniques such as Sobol, Morris, and VARS methods.

Tampere University delved into sensitivity analysis to validate models used in planning and safety assessment of geological repositories. They utilized variance-based methods like Sobol and RBD Fast, and considered the feasibility of using surrogate models for sensitivity analysis across different case studies.

BRGM worked on modelling the pore water composition's influence on sorption processes in deep geological repositories. They conducted uncertainty propagation studies using polynomial chaos expansions, analysing the effects of perturbed model input parameters and associated thermodynamic databases on pore water chemistry modelling.

Multi-fidelity Techniques and Uncertainty Quantification for Mechanical and Transport Processes in Fractured Rock

[1]Březina, J., [1]Exner, P., [1]Špetlík, M., [1]Stebel, J.,

[2]Béreš, M., [2]Bérešová, S., [2]Sysala, S.

[1]Technical University of Liberec, Czech Republic

[2]Institute of Geonics of the Czech Academy of Sciences, Czech Republic

Abstract

The report summarizes the results achieved by teams from the Technical University of Liberec and the Institute of Geonics of the Czech Academy of Sciences, as part of Task 4 within the DONUT workpackage. We introduce a parallel Bayesian inversion library based on the delayed acceptance Metropolis-Hastings method for accelerated sampling using suitable surrogate models. Then we apply the library to an inverse problem, specifically the identification of the parameters to the hydro-mechanical model describing formation of the excavation disturbed zone. Additionally, we introduce a multilevel Monte Carlo method (MLMC) library, which has been effectively applied to test transport problems. We present a neural network developed for fast numerical homogenization of small-scale discrete fractures in the 2D case. This represents a crucial advancement for applying MLMC to models with discrete fracture networks. Finally, we integrate developed methods into a benchmark safety study for near-field transport.

Significance Statement

The majority of the safety analysis methods for deep repositories of radioactive waste are based on the Monte Carlo method, which requires evaluation of many forward simulations. Addressing computational complexity, simplified models are employed with uncertainties in parameters assessed indirectly or subjectively from observations and full models. As a result, outcomes are compromised in terms of accuracy and uncertainty quantification. To surmount these challenges, our research focuses on three core aspects:

Employing Physically-Based Forward Models. Our approach integrates stochastic continuum and discrete fracture networks in flow, transport, and mechanical models for describing crystalline rock masses. The discrete fracture-matrix (DFM) method, distinct from discrete fracture network (DFN) approaches, aims to capture small-scale features by the stochastic continuum and provide models of variable accuracy.

Apply Bayesian Inversion as an objective and robust technique to identify model parameters and their uncertainties (posterior distribution) based on observed data and prior knowledge. The developed SurrDAMH library combines a full forward model with its surrogate model, accelerating the sampling process while preserving posterior distribution quality. Parallel sampling techniques further enhance speed. The library's efficiency was demonstrated on several test problems and then applied to the identification of the parameters of the excavation disturbed zone (EDZ). The nonlinear and non-smooth hydro-mechanical model proves the simple surrogate models insufficient, yet through parallelization and a simpler 2D model, we derive good posteriors. Further development of model-specific surrogates would allow wider usage of the Bayes inversion.

Control Approximation Errors. The multilevel Monte Carlo (MLMC) method takes advantage of the models of varying accuracy and cost to provide computationally efficient estimation of quantities of interest within prescribed accuracy. Our parallel implementation has been demonstrated to provide a significant advantage against classical Monte Carlo on the test Darcy flow and transport problems. This motivated the development of fast numerical homogenization for DFM models using deep-learning neural networks. That homogenizaton technique was successfully applied to 2D models. In the near future, mastering the homogenization of 3D DFM transport models in combination with MLMC would provide a much more flexible, precise, and efficient alternative to the widely used stochastic DFN models.

In conclusion, the combination of the developed techniques is demonstrated in a near-field study of the interaction between natural fractures and EDZ.

Table of Content

List of Figures

Glossary

DAMH	Metropolis-Hastings algorithm with delayed acceptance
DFM	Discrete fracture-matrix (model)
EDZ	Excavation disturbed zone
MCMC	Markov Chain Monte Carlo
MLMC	Multilevel Monte Carlo
NN	Neural network
PDF	Probability density function
QoI	Quantity of interest
RBF	Radial basis function
TSX	Tunnel sealing experiment

1. Introduction

The report is organized as follows. In Section 2, we begin with the principles of Bayesian inversion and overview of the SurrDAMH Python library developed during the project. The library leverages the Metropolis-Hastings algorithm with delayed acceptance (DAMH) to accelerate the sampling process. Next, in Section 2.1, we apply Bayesian inversion to estimate the parameter distribution for the excavation disturbed zone's (EDZ) poroelastic model. Specifically, we characterize the permeability of the intact rock and its increase in the vicinity of the tunnel that is induced by the excavation process. Given the complexity of the poroelastic model, Section 2.2 is dedicated to an in-depth examination of various surrogate models within the context of the DAMH algorithm.

To address uncertainty propagation in fractured rock, we introduce the multilevel Monte Carlo (MLMC) method in Section 3. In Section 3.1, we present a newly developed MLMC Python library with a parallel and adaptive implementation of the method. For an efficient application of MLMC, we explore the potential of using neural networks (NN) for rapid numerical homogenization of the fractured media in Section 3.2.

Section 4 presents a benchmark problem for uncertainty propagation in a near-field transport model as a step toward a more realistic safety assessment.

2. Bayesian Inversion

Bayesian inversion stands as a more informative alternative to traditional deterministic inversion methods, frequently used in repository modeling, see (Axelsson, Béreš, and Blaheta 2021), (Haslinger, Blaheta, and Mäkinen 2021) or (Malík and Kolcun 2023). Unlike deterministic methods, which yield only point estimates of model parameters, Bayesian inversion provides a probability distribution for these parameters. This distribution, known as the posterior distribution, incorporates both the observational data and prior knowledge, offering a comprehensive understanding of parameter uncertainties. Another advantage is the much better robustness in scenarios involving complex forward models and noisy observational data. However, it's worth noting that Bayesian methods are less established, harder to interpret, and computationally more demanding, which can pose challenges in their application. The SurrDAMH Python library was developed to remedy some of these issues.

We consider the additive noise model

$$Y = G(U) + Z,$$

where $G: \mathbb{R}^n \to \mathbb{R}^m$ is the forward model. The multivariate random variables Y, U, and Z represent the observed data ($Y \in \mathbb{R}^m$), unknown parameters ($U \in \mathbb{R}^n$), and observational noise ($Z \in \mathbb{R}^m$), respectively. Given a particular observation vector $y \in \mathbb{R}^m$, the posterior probability density function (PDF) of the parameter vector u is the conditional PDF $f_{U|Y}(u|y)$, which can be expressed using Bayes' theorem as:

$$f_{U|Y}(u|y) \propto f_{Y|U}(y|u) f_U(u) = f_Z\big(y - G(u)\big) f_U(u),$$

where f_U and f_Z (also called the data likelihood) are the PDFs of U and Z, respectively. The equality like symbol means "up to the normalization constant". In this sense the sentence could be appended: likelihoods are the PDFs of U and Z, respectively, up to the normalization constant.

Samples from the posterior distribution are usually generated via Markov Chain Monte Carlo (MCMC) methods. (Brook et al. 2011) The obtained samples are subsequently utilized to estimate the moments (e.g., expected value, variance) of the posterior distribution, to approximate its marginal distributions, etc.

However, when dealing with computationally intensive forward models, efficient variants of MCMC methods become crucial for feasible posterior sampling.

As the primary MCMC method, we used the delayed-acceptance Metropolis-Hastings algorithm (DAMH, see (Cui et al. 2011) in combination with suitable surrogate models. Importantly, the forward model is not replaced by the surrogate (this would lead to sampling from approximate distribution); rather, the

surrogate is leveraged to reduce the number of required forward model evaluations. Specifically, at each step of the DAMH algorithm, the proposed sample is first evaluated using the cheap surrogate model, and only if it is likely to be accepted, it is evaluated using the expensive exact forward model.

We implemented these sampling methods using our in-house Python library, SurrDAMH (Bérešová 2022b), designed for the numerical realization of Bayesian inversion. The library's core concept involves building and refining a surrogate model during the sampling process.

SurrDAMH is able to run multiple instances of forward model solvers in parallel and use separate processes to simultaneously generate multiple Markov chains. These parallel samplers share a single surrogate model. Given that forward model evaluations typically consume most of the computational time, the number of solvers is generally set to be lower than the number of generated Markov chains for optimal resource utilization.

The sampling procedure comprises several phases:

1. The initial phase utilizes the basic Metropolis-Hastings algorithm, with the obtained snapshots serving to construct an initial surrogate model.
2. The main phase employs the DAMH algorithm, during which the surrogate model is iteratively refined and used to speed up the sampling process.
3. Upon achieving sufficient surrogate model accuracy, further updates are halted, and the DAMH algorithm continues without them.

The sampling processes use a Gaussian random walk proposal distribution $\mathcal{N}_n(u, C)$. For each sampling phase, it is possible to specify the covariance matrix C, surrogate model type, and a stopping criterion. The adaptive selection of the proposal distribution's covariance matrix C has also been implemented, a feature essential for efficient sampling from complex posterior distributions, see (Bérešová 2022a) for details.

2.1 TSX experiment and Bayesian inversion

Direct measurement of the permeability field in the EDZ at tunnel scale is challenging due to the heterogeneity of the rock mass. The use of mine-by measurements in combination with inverse models provides an alternative. In this section, we explore the application of Bayesian inversion to the data from the continuous measurement of pore pressures during and after excavation, performed as part of the Tunnel Sealing Experiment (TSX) conducted at the Underground Research Laboratory in Canada in 1996-1997, see (Chandler et al. 2002).

From existing parallel tunnels, one vertical and one horizontal borehole were drilled orthogonally toward the tunnel before its excavation. The boreholes were equipped with continuous measurements of hydraulic pressures at different distances from the tunnel. The tunnel was excavated in 17 days, leading to sudden increase of the pore pressure in the vertical borehole and decrease in the horizontal borehole, as expected due to stress changes. Since the rock permeability was low, very slow relaxation of the pore pressure was observed and logged over one year without reaching the initial values.

The data were already used as part of the DECOVALEX-THMC project Task A, see (Rutqvist et al. 2009), where plastic models and classical inversion were used to determine certain model parameters, while others were taken from direct measurement. However, these calculations did not account for uncertainties, a gap that Bayesian inversion can fill.

In our forward model, a linear poroelastic model of a 2D cut of the tunnel neighborhood is computed in software Flow123d (Březina et al. 2011-2023). A Dirichlet boundary condition for the pressure and the displacement is considered on the exterior boundary. At the beginning, zero total force on the interior boundary (the tunnel wall) is prescribed so that it compensates for the initial stress. The excavation is then simulated through the interior boundary condition where the hydraulic pressure is linearly decreased to the atmospheric pressure while the total force increases to values corresponding to the

initial stress. The principle unknowns: the displacement, the hydraulic pressure head, and the velocity, are discretized using a finite element method, mechanical and hydraulical equations are split and solved iteratively. Additionally, the nonlinear hydraulic conductivity according to (Rutqvist et al. 2009) is considered:

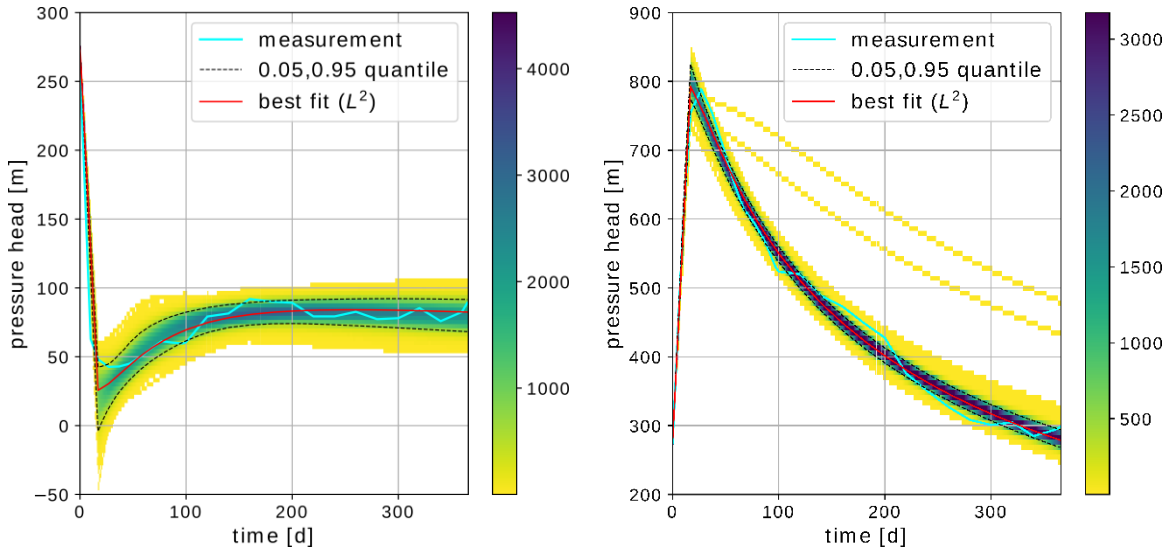
$$k = \frac{\rho g}{\mu}\left[\mathring{\kappa}_r + \Delta\mathring{\kappa}_m \exp\left(\mathring{\beta}\sigma_m\right)\right]\exp\left(\mathring{\gamma}\Delta\sigma_{VM}\right), \tag{1}$$

with an exponential dependence on the mechanical stress, σ_m being the mean stress and σ_{VM} the von Mises stress, whereas $\mathring{\kappa}_r$ is the residual permeability under high mean stress and $\Delta\mathring{\kappa}_m$, $\mathring{\beta}$, $\mathring{\gamma}$ are fitting constants.

In Bayesian inversion, 8 free parameters were selected: the storativity, the Young modulus, the initial stress in x and y direction, and 4 hydraulic conductivity parameters including $\mathring{\kappa}_r$. The prior lognormal distributions are chosen such that the mean is close to the parameter values found in (Rutqvist et al. 2009). The library SurrDAMH is used for computing the Bayesian inversion and other in-house codes make it possible to run everything in parallel on a cluster of Metacentrum (www.metacentrum.cz).

In the beginning, the Bayesian inversion was run with the proposed forward model and measured data from 4 observation points (two on the side of the tunnel and two above it). Similarly to the results in (Rutqvist et al. 2009), no good fit was found because the model is too simple to capture all the effects both in the vertical and horizontal direction. Therefore the inversion was run for each borehole data separately and results for the closer observation points H1 (on a side) and V1 (above) are presented below.

So far, the best results have been achieved with an adaptive proposal distribution step that optimizes the random walk through the parameter space to reach the chosen target acceptance ratio (number of accepted/rejected samples). Setting the target acceptance ratio to 0.5 results in a sufficient amount of accepted samples while covering a significant part of the parameter space around the prior estimate. Without this kind of adaptivity, the acceptance ratio could hardly reach 0.1 with a reasonable constant step size. A comparison of observed simulation results with measured data in the boreholes H1 and V1 is shown in Figure 1, note the different scale of the y-axes.

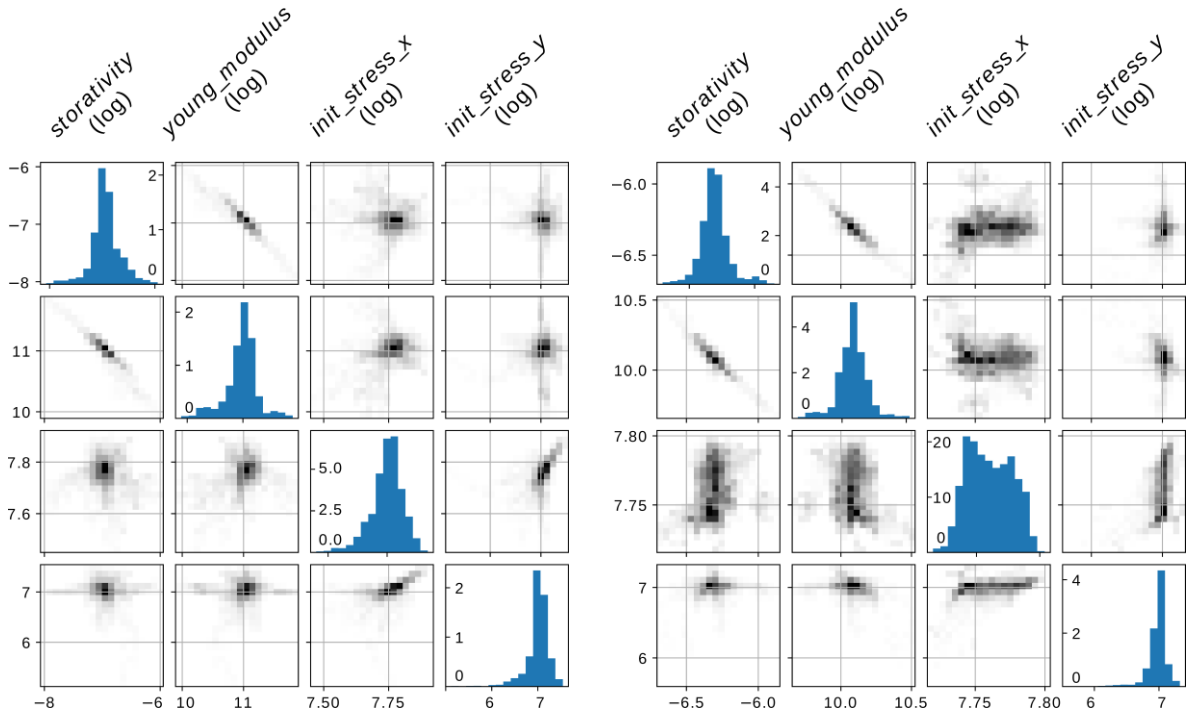


(a) Observation point H1

(b) Observation point V1

Figure 1 - Temporal frequency plot histograms of the simulated pressure evolution for posterior samples in comparison with measured data.

A very good fit can be seen in the case of the V1 observation point both in the excavation phase including the large pressure peak and in the relaxation phase. The measured data are well covered by 95% of samples with modus sample being in the middle. In the case of the H1 observation point, the measured data are also well covered by samples but the peak at the excavation phase end is missed and the pressure at that point can even turn negative for some samples (also seen in (Rutqvist et al. 2009)).



Observation point H1. Observation point V1.

Figure 2 - Marginal posterior PDF for selected model parameters and their pairs. The rows correspond to the model parameters in the columns.

One of the crucial parameters of the Bayesian inversion is the random noise. The noise's standard deviation was estimated to be 20 m and 25 m for H1 and V1, respectively. The noise is naturally supposed to be correlated (in time) but this conflicts with the fact that the model solution is not smooth at the excavation phase end. Setting any correlation in the noise always resulted in a bad fit around the pressure peak for both H1 and V1 cases. Further tests in this matter are underway in which the forward model is complemented with a smoothing term that actually simulates the delayed response of the pressure measurement sensors.

In Figure 2, histograms in selected 1D-2D parameter space slices are shown for both H1 and V1 observation points. At first glance, a strong correlation between the storativity and the Young modulus can be seen in both H1 and V1: the stiffer the rock is, the smaller the chance there is for pores to open and increase the storativity. The pressure at V1 is more sensitive to the initial stress, especially in the x direction (compare the initial stress scale in graphs). On the other hand, the initial stress parameters seem to be independent of the other parameters.

In Figure 3, selected 2D histograms for permeability parameters in the V1 case. The parameter $\mathring{\kappa}_r$ related to the permeability of the intact rock shows a positive correlation to the storativity and negative to the Young modulus: in the case of higher storativity, it can be assumed that the pore space is larger which would also cause higher permeability; the opposite correlation to the Young modulus is then in accordance to the correlation between the storativity and the Young modulus as seen in Figure 2. The rest of the permeability parameters seem to be largely independent of the mechanical parameters.

The presented results were computed without using any surrogate models. When trying to apply both polynomial or radial basis function surrogate models available in SurrDAMH, the Markov chains often

stagnated during the DAMH phase due to a high amount of pre-rejected samples. This problem is studied in deeper detail in Section 2.2.

Finally let us note, that the hydraulic conductivity defined in (1) is unbounded, and unrealistically high values were observed in some of the samples. Some of these unphysical samples even showed a good fit. The relationship (1) is very sensitive to the parameters in the two exponential terms and there is no information during the Bayesian inversion to reject these samples. Two approaches to this matter are currently being tested: changing (1) with a different term with some bounds introduced and/or adding the permeability measurement data from (Rutqvist et al. 2009) to the observed data which would influence the likelihood.

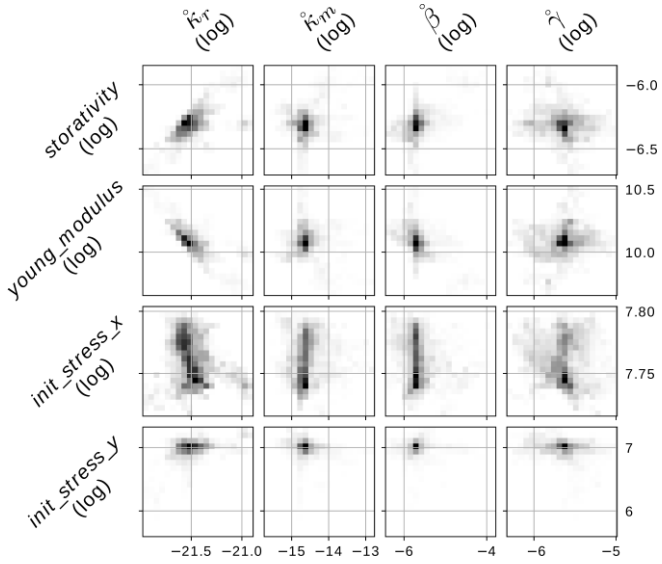


Figure 3 - Mariginal posterior PDF for permeability parameters versus mechanical parameters in the case of V1.

2.2 Surrogates for Hydro-Mechanical Model

This section aims to assess the performance of different surrogate models to the forward model of the TSX experiment used in the previous section. We evaluate the approximation properties of the polynomials and radial basis function (RBF) surrogates implemented in SurrDAMH and compare them to an alternative approach using a neural network. This study is motivated by the observed poor performance of the DAMH algorithm at the TSX inversion problem when combined with SurrDAMH surrogates.

Each component $g: \mathbb{R}^n \to \mathbb{R}$ of the forward model G is approximated separately. We consider the following kinds of surrogates.

Polynomial surrogate model. The function g is approximated by its orthogonal projection

$$\tilde{g}(u) = \sum_{i=1}^{P} \alpha_i \, p_i(u)$$

to the space of polynomials P^n, using Hermite polynomials p_i as a basis. The projection is performed with respect to a dot product weighted by the posterior distribution. Roughly speaking, the surrogate model accuracy will be higher for samples of higher acceptance probability.

RBF surrogate model. For a selected set of samples $u^{(k)}$ and their corresponding values $G\left(u^{(k)}\right)$, the RBF surrogate has the form

$$\tilde{g}(u) = \sum_{k=1}^{K} \alpha_k \, \phi\left(\left\|u - u^{(k)}\right\|_2\right) + \sum_{i=1}^{P} \alpha_i \, p_i(u), \tag{2}$$

where $\phi: \mathbb{R} \to \mathbb{R}$ is a chosen radial kernel function. The weights α_k and β_i are determined in terms of the least square method. In particular, we shall consider thin-plate-spline kernel $\phi(r) = r^2 \log r$.

Neural network model. Since the hydro-mechanical model described in Section 2.1 uses a non-smooth dependency of the permeability on the stress given by (1), its approximation by the smooth polynomial and RBF surrogates could be inadequate. Therefore, we explore an NN surrogate with three hidden layers of size 100 and a ReLU activation function. Additionally, we approximate all of the components at once, i.e. we approximate the original function $G: \mathbb{R}^n \to \mathbb{R}^m$.

The polynomial, RBF, and NN surrogates were trained on the dataset consisting of $15{,}123$ samples and had their precision evaluated at $136{,}107$ samples. These samples were obtained using a simple Metropolis-Hastings method for the Bayesian inversion of the H1 observation, as shown in Figure 1a. The behavior of the trained surrogates in the DAMH algorithm was then simulated using the whole MCMC chain.

Figure 4 -Figure 6 display the performance of the compared surrogates. On the left side of each figure, the log-likelihood $\log(f_Z)$ for the exact model is plotted against that of the surrogate model. Samples with higher acceptance probability are located in the top-right corner. Clearly, points are closer to the diagonal line for the NN surrogate, indicating a more accurate approximation. On the right side, the histograms of the simulated acceptance rate of the surrogate proposed sample are displayed. The higher the values the better, an ideal surrogate will have an acceptance rate constant of 1. The NN surrogate not only exhibits a slightly better acceptance rate but also effectively avoids very poor approximations. This advantage is particularly pronounced when dealing with small training datasets, which are often encountered in the early phases of the DAMH algorithm.

These results suggest that NN surrogates could be much more robust, but their behavior within the DAMH algorithm has yet to be verified.

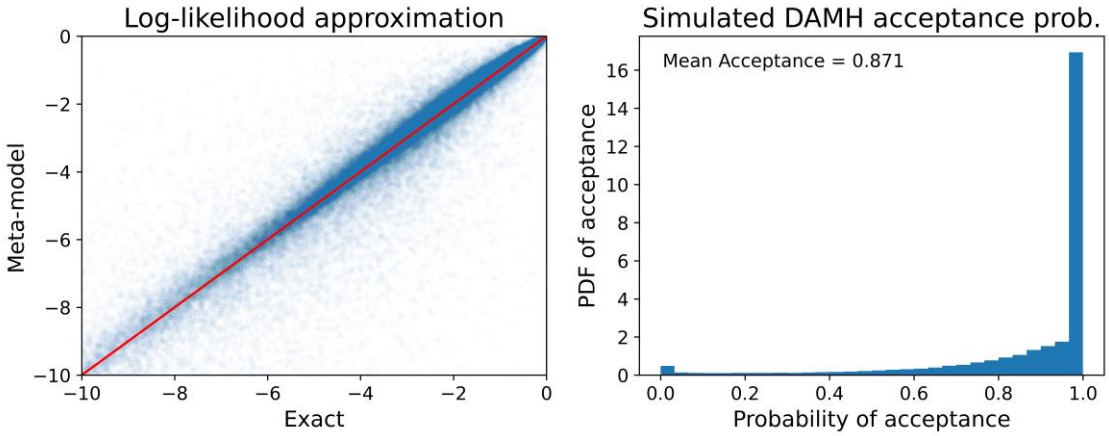


Figure 4 - Polynomials up to the degree 5.

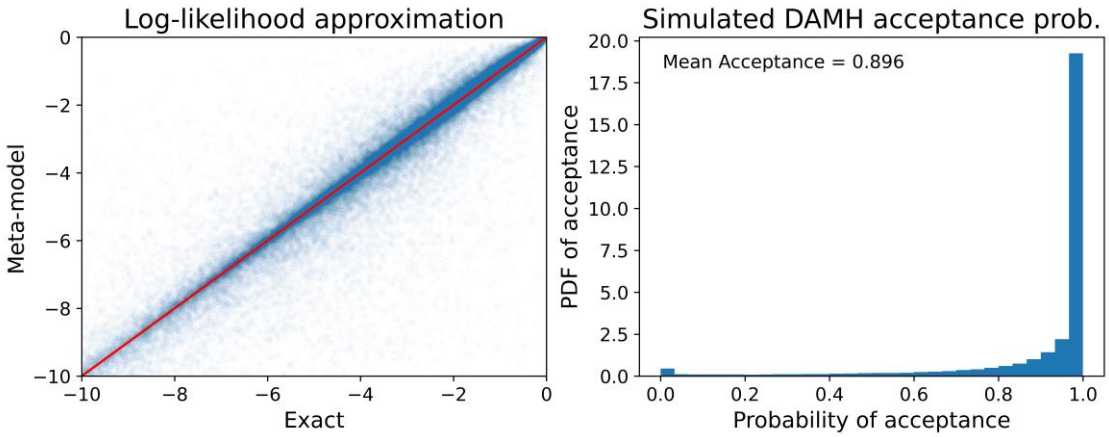


Figure 5 - Thin-plate-spline kernel RBF with the polynomials up 5-th degree.

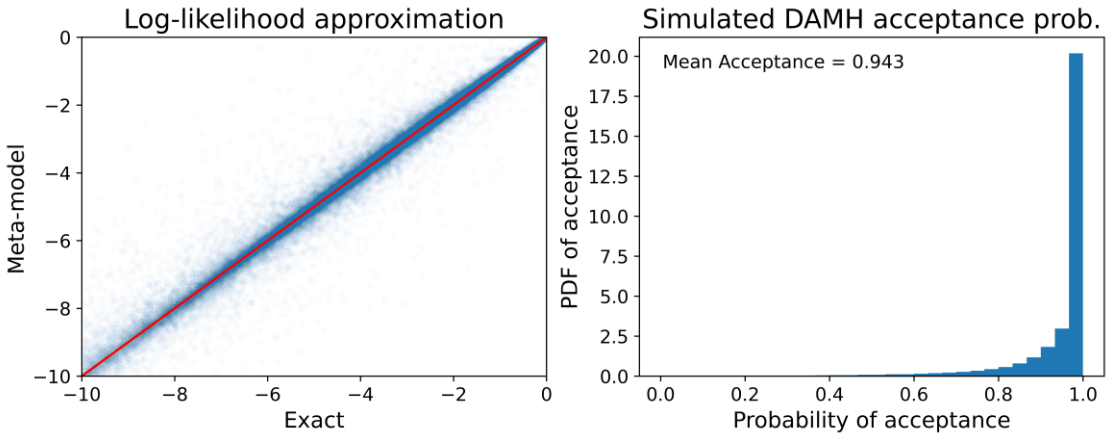


Figure 6- NN with three hidden layers of size 100 with ReLU activation functions.

3. Multilevel Monte Carlo Method and Neural Networks

Sensitivity analysis methods and uncertainty propagation through forward models are pivotal in the context of safety analysis for deep repositories. Due to the extensive number of model evaluations required by the traditional Monte Carlo method, simplified or surrogate models are often employed, thereby introducing approximation errors. The MLMC method, introduced in (Giles 2015), provides a more efficient and accurate point estimate of the mean value of a quantity of interest (QoI). Unlike traditional Monte Carlo methods that rely on repeated runs of a single model, MLMC uses a hierarchy of progressively less accurate models. These models aim to approximate the same QoI but at a reduced computational cost. Many samples are collected for less accurate models, while much fewer are collected for highly accurate models.

Formally, let $P(\omega)$ be our QoI, where ω represents a realization of random parameters. In our cases, $P(\omega)$ is a random variable that is a solution of a partial differential equation. Although $P(\omega)$ is not computationally accessible, it is feasible to compute its approximation $P_L(\omega)$, a finite element method solution in our case. Then, let $P_1, \dots, P_{L-1}$ be a sequence approximating P_L. These approximations are more accurate from P_1 to P_L but are more computationally expensive. Then, the expected value of P_L is:

$$\mathbb{E}[P_L] = \mathbb{E}[P_1] + \sum_{l=2}^{L} \mathbb{E}[P_l - P_{l-1}]\,. \tag{3}$$

The MLMC estimates an expectation as follows:

$$Y = \langle P_1(\omega^1)\rangle_{N_1} + \sum_{l=2}^{L} \langle\, P_l(\omega^l) - P_{l-1}(\omega^l)\rangle_{N_l}, \tag{4}$$

where L is the number of levels, N_l represents the number of simulation samples on level l. The total cost and variance of the estimate:

$$C = \sum_{l=1}^{L} N_l\, C_l \quad V = \sum_{l=1}^{L} \frac{V_l}{N_l} \tag{5}$$

are given in terms of C_l, V_l representing the cost and the variance of differences $P_l - P_{l-1}$. If the cost C_l decays fast enough for the small levels in comparison to the increasing variance V_l, the optimal number of vector of level samples N_l can be calculated from C_l, V_l, providing the total cost comparable to few evaluations of the finest level model.

3.1 MLMC Software Library

The Python software library MLMC (Březina and Špetlík 2022) has been gradually developed. The most recently released version, 1.0.2 (Jul 12, 2022), provides users with the following main features:

- User-defined simulation with output QoI of user-defined structure.
- Symbolic calculation with quantities of interest and their estimation using collected samples.
- Simulation samples are scheduled iteratively. V_l and C_l are continuously estimated to correct the optimal N_l.
- Scheduled samples can run in parallel locally or on clusters with the PBS scheduler system.
- Sample data can be stored in HDF5 files, which is particularly beneficial when working with large datasets. Results are post-processed in small chunks that fit into internal memory.
- In addition to estimating expectations of QoI, there are direct methods for calculating moments or a covariance matrix. Also, the algorithm for PDF reconstruction using the maximal entropy method is available.

The current version of the library provides the capabilities necessary for conducting reliable MLMC experiments.

3.2 Deep Neural Networks for Numerical Homogenization of Fractured Rock

Constructing the coarse-level approximations, a prerequisite for employing the MLMC method, becomes intricate for the discrete fracture-matrix (DFM) models. While the small-scale fractures cannot be resolved directly, capturing their effect remains imperative to achieve accurate coarse approximations. To this end, numerical homogenization can be employed, but this approach makes the whole MLMC calculation impractically slow. Therefore, a fast meta-model for numerical homogenization becomes an essential component of the MLMC framework for DFM models. In this regard, we are going to utilize deep-learning meta-models.

Generating an approximation with a step H from an approximation with a step $h < H$ commences by extracting the discrete fractures larger than H and constructing the coarse mesh. The remaining

homogenization mesh, consisting of fine-level elements and small-scale discrete fractures, is restricted to the overlapping homogenization blocks, as illustrated in Figure 7. Subsequently, the effective conductivity tensor is calculated for each block and the tensor field is interpolated from a regular grid of blocks to the elements of the coarse mesh.

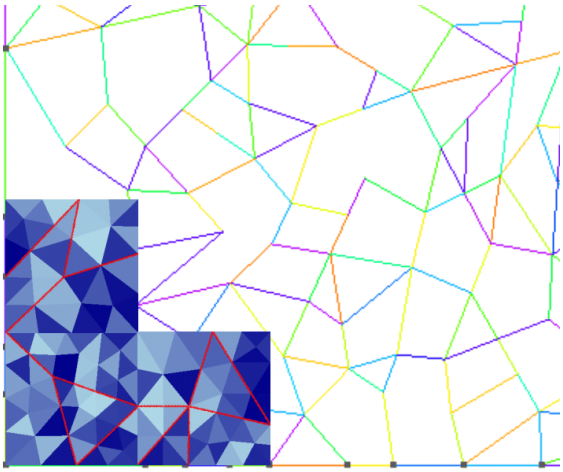
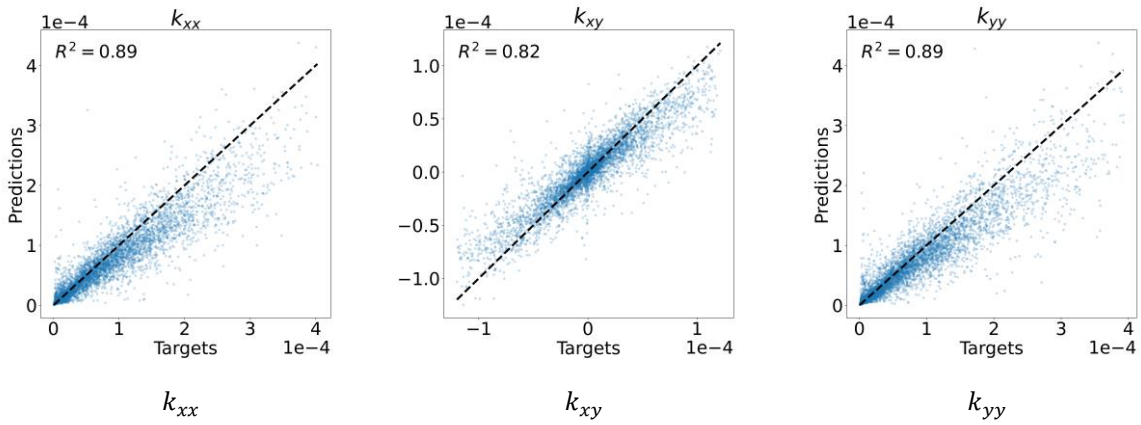


Figure 7 - Fine level fractures and corresponding rasterized homogenization blocks.

The described numerical homogenization procedure is notably slower compared to solving the fine-level problem, rendering it unsuitable for the efficiency needs of MLMC. Once the homogenization blocks and their effective hydraulic conductivity are obtained, the implementation of a data-driven meta-model offers a route to accelerate computations. Our deep learning meta-model comprises a convolutional neural network followed by a feed-forward neural network. Convolutional NNs (Goodfellow et al. 2016, ch. 9) typically process data on regular grids, akin to images. In our context, we prescribe an input spatial random field of conductivity tensors on unstructured meshes and adopt rasterization to project the field onto a regular grid.

To achieve this, we utilize the Datashader library (Cottam et al. 2022) for nearest-neighbor interpolation into a 256×256 matrix of grayscale pixels. As we interpolate a tensor field, three independent components of the 2D tensor are assigned to every pixel. A Convolutional NN functions as an input field feature extractor, followed by a feed-forward NN with three output neurons, serving to perform the regression task. Employing supervised learning, our deep learning meta-model addresses the regression problem by learning the unknown mapping from input random field to output effective tensors through collected pairs of input and corresponding model outputs.

Our meta-model, trained on a dataset encompassing samples from three distinct mesh steps (a total of 32,000 samples), yields sufficiently accurate results for each step's subsets. Figure 8 showcases results for a specific subset.



k_{xx} k_{xy} k_{yy}

Figure 8 - Meta-model accuracy for each independent element of conductivity tensor

In further research, we would like to employ the trained meta-model within MLMC to investigate its impact on computational cost, variance decrease, and mean estimation.

4. EDZ Transport Benchmark

A simplified safety case calculation has been conducted as a benchmark for integrating the developed tools and techniques. The excavation disturbed zone (EDZ) emerges as a result of both the excavation of repository galleries and subsequent poroelastic relaxation due to water drainage. These processes induce mechanical changes that alter the micro-crack network, significantly increasing the effective permeability of the disturbed rock. Despite its narrow profile, the EDZ's extent is comparable to that of repository-scale faults, potentially enhancing the connectivity of the natural fracture network and reducing contaminant breakthrough times.

To capture the interaction between the EDZ and fracture network, we utilized the 3D DFM transport model with EDZ parameters sampled from the posterior of the TSX inversion problem, and a random fracture network generated for the Forsmark site data (Öhman and Follin 2010), the diffusion and source term parameters were taken from the reference calculations for the Czech waste package system. We considered a parametric geometry model of a single repository storage tunnel. The maximal concentration of the tracer on the model boundary was used as the performance indicator, i.e., the quantity of interest. The Figure 9 displays an estimated PDF of the safety indicator using 1000 Monte Carlo samples. We compare cases with and without the impact of the EDZ and three positions of tracer sources. While the impact of the EDZ is likely overestimated due to the combination of parameters from different localities, the results demonstrate the ability of the stochastic model to capture this difference along with the variance primarily caused by the random fractures. A subtler difference between source positions is attributed to the impact of a single fixed fault intersecting the tunnel in combination with a predominant flow direction.

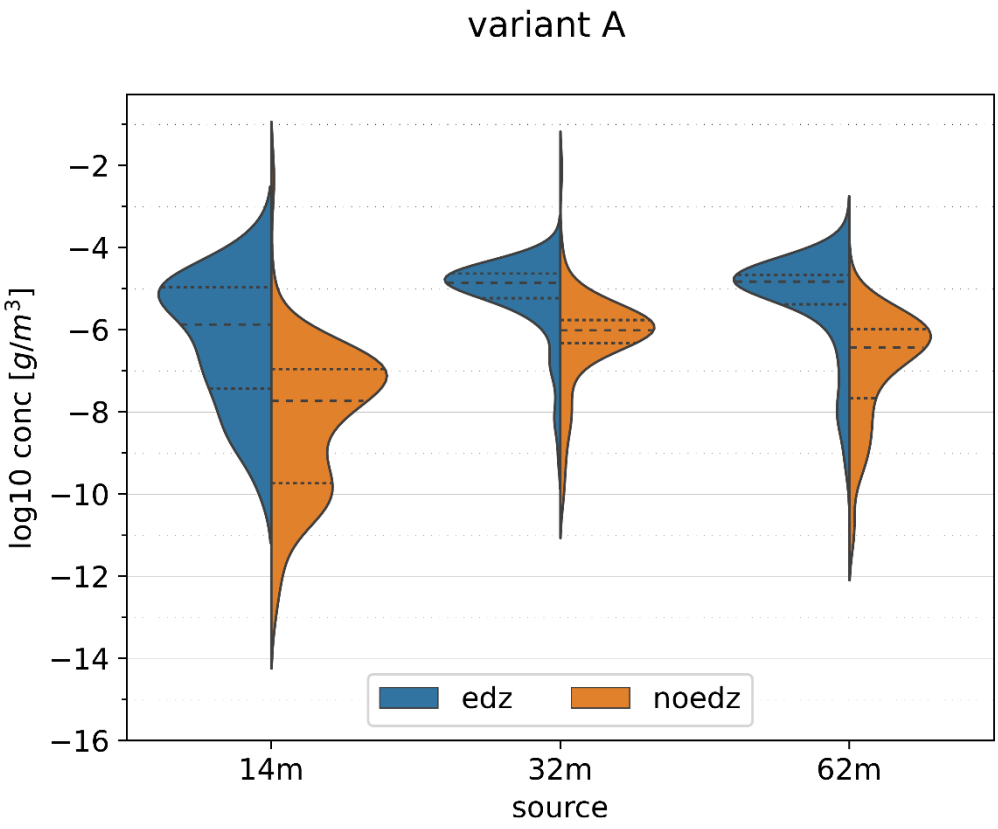


Figure 9 - Estimated PDF of the safety indicator, maximal concentration on the boundary, with and without increased EDZ permeability and different position of the source container.

Acknowledgements

The participation of the TUL and IGN teams within the EURAD project has been cofinanced by The Czech Radioactive Waste Repository Authority (SÚRAO) under grant agreement number SO2020-017.

Code Source

http://flow123d.github.io - Homepage of the Flow123d project.

http://github.com/GeoMop/MLMC - Homepage of the MLMC Python library.

https://github.com/dom0015/surrDAMH/tree/Version1 - Homepage of the surrDAMH library.

References

Axelsson, O., Béreš, M., Blaheta, R., 2021. Computational Methods for Boundary Optimal Control and Identification Problems. Mathematics and Computers in Simulation 189, 276-290.

Bérešová, S., 2022a. Bayesian Approach to the Identification of Parameters of Differential Equations. Ph.D. Dissertation, VSB - Technical University of Ostrava, Ostrava, Czechia, 122 pp.

Bérešová, S., 2022b. surrDAMH. Repository. https://github.com/dom0015/ surrDAMH/tree/Version1.

Březina, J., Stebel, J., Exner, P., Hybš, J., 2011-2023. Flow123d. http://flow123d.github.io, repository: http://github.com/flow123d/flow123d.

Březina, J., Špetlík, M., 2022. MLMC Python library. http://github.com/GeoMop/MLMC.

Chandler, N. A., Cournut, A., Dixon, D., 2002. The Five Year Report of the Tunnel Sealing Experiment: An International Project of AECL, JNC, ANDRA and WIPP. Pinawa, Manitoba (Canada): Atomic Energy of Canada Limited.

Cottam, J., Bednar, J.A., Crist, J., Wang, P., 2022. Datashader. https://datashader.org/.

Cui, T., Fox, C., O'Sullivan, M.J., 2011. Bayesian Calibration of a Large-Scale Geothermal Reservoir Model by a New Adaptive Delayed Acceptance Metropolis Hastings Algorithm: ADAPTIVE DELAYED ACCEPTANCE METROPOLIS-HASTINGS ALGORITHM. Water Resources Research 47(10). https://doi.org/10.1029/2010WR010352.

Giles, M.B., 2015. Multilevel Monte Carlo Methods. Acta Numerica 24, 259-328. https://doi.org/10.1017/S096249291500001X.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.

Haslinger, J., Blaheta, R., Mäkinen, R.A.E., 2021. Parameter Identification for Heterogeneous Materials by Optimal Control Approach with Flux Cost Functionals. Mathematics and Computers in Simulation 189, 55-68.

Malík, J., Kolcun, A., 2023. Determination of the initial stress tensor from deformation of underground opening - theoretical background and applications. In: Chleboun, J., Kůs, P., Papež, J., Rozložník, M., Segeth, K. and Šístek, J. (eds.): Programs and Algorithms of Numerical Mathematics. Proceedings of Seminar. Jablonec nad Nisou, June 19-24, 2022. Institute of Mathematics CAS, Prague, 107-122.

Öhman, J., Follin, S., 2010. Site investigation SFR – Hydrogeological modelling of SFR – Model version 0.2, No. R-10-03, Svensk Kärnbränslehantering AB.

Rutqvist, J., Börgesson, L., Chijimatsu, M., Hernelind, J., Jing, L., Kobayashi, A., Nguyen, S., 2009. Modeling of Damage, Permeability Changes and Pressure Responses During Excavation of the TSX Tunnel in Granitic Rock at URL, Canada. Environmental Geology 57(6), 1263-74. https://doi.org/10.1007/s00254-008-1515-6.

Brooks, S. 2011: Handbook of Markov Chain Monte Carlo, ISBN: 1420079417

Characterization and quantification of uncertainties in RN migration in crystalline formations

[1]Jankovsky, F., Zuma, M., [2]Pospiech, S., [2]Brendler, V., [3]Noseck, U., [3]Schneider, A., [3]Becker, D.-A.

[1]ÚJV Řež, a. s., Czech Republic

[2]Institute for Resource Ecology, HZDR, Germany

[3]GRS gGmbH, Braunschweig, Germany

Abstract

A joint collaboration by UJV, GRS, and HZDR has been performed to carry out a complex uncertainty treatment and sensitivity analysis effort based on a coordinated data collection, uncertainty quantification and testing sensitivity analysis methods, applied to geochemical and transport applications. The work consisted of three different parts. Firstly, a block experiment in an artificial fracture was performed to investigate the impact of geochemical conditions on the transport of selected radionuclides and also to be used as a reference model to investigate the impact of uncertainties of hydrogeological and geochemical data on the RN transport in a defined model system. Secondly, a tool for probabilistic uncertainty and sensitivity analysis by means of flow and transport simulations in a fractured medium was developed. It was realized by an interface between the transport code d³f++ and RepoSTAR (a tool box for controlling and evaluating probabilistic runs). Thirdly, a workflow from raw data (rock measurements) to model which enables to include different types of host rocks with natural small scale (sub-cm) to meso scale (sub-m) heterogeneities was developed to characterise the geochemical parameter along fluid migration pathways. All three tasks were related to each other which is described in detail..

Significance Statement

Simulations of THMC process couplings of radioactive waste disposal involves many input models and data whose level of couplings vary in space and time and could be weak and strong. Effects of uncertainties of input data and models on final indicators are difficult to take into account as there are strong non-linearities, and classical methods have thus some limitations. Relevant methods and tools should be defined and set up to assess (i) effect of all kinds of uncertainties of results indicators (uncertainty analysis) and (ii) relevant input data and models which manage the result (sensitivity analysis). A special challenge is the mineralogical heterogeneity affecting, among others, the distribution coefficients of radionuclides and thus their retardation.

About developments of mathematical representations of uncertain parameters applied for coupled processes, a joint programme is outlined, since UJV, GRS, and HZDR collaborated to carry out a complex uncertainty treatment and sensitivity analysis effort based on a coordinated data collection, uncertainty quantification and testing sensitivity analysis methods, applied to geochemical and transport applications.

The work for including hydrogeological and geochemical uncertainties into the flow and transport modelling in co-operation between the partners UJV, HZDR and GRS consisted of three different parts. Firstly, it was decided to use the Czech block experiment as a reference model to investigate the impact of uncertainties of hydrogeological and geochemical data on the RN transport in a defined model system. Secondly, GRS developed a tool for probabilistic uncertainty and sensitivity analysis by means of flow and transport simulations in a fractured medium realized by an interface between the transport code d³f++ and RepoSTAR (a tool box for controlling and evaluating probabilistic runs). Thirdly, HZDR was working on establishing a workflow from raw data (rock measurements) to model which enables to include different types of host rocks with natural small scale (sub-cm) to meso scale (sub-m) heterogeneities to characterise the geochemical parameter along fluid migration pathways. *Figure 1-1* illustrates how these tasks are interrelated within the DONUT work of task 4. The methods and approaches developed here applicable to geochemical and transport modelling in fractured media but also transferable to other geological systems

Table of content

List of figures

List of Tables

1. General objectives

Simulations of THMC process couplings of radioactive waste disposal involves many input models and data whose level of couplings vary in space and time and could be weak and strong. Effects of uncertainties of input data and models on final indicators are difficult to take into account as there are strong non-linearities, and classical methods have thus some limitations. Relevant methods and tools should be defined and set up to assess (i) effect of all kinds of uncertainties on uncertainty of results indicators (uncertainty analysis) and (ii) relevant input data and models which manage the result (sensitivity analysis).

Task 4 focused on:

- Developments of innovating numerical methods to treat uncertainty and sensitivity analysis on complex coupled representative cases (big system, many media and input data) with strong nonlinearities: global/local sensitivity analysis with specific tools (surrogate/metamodels such Chaos polynomia, neural network, adjoin state methods, …)
- Developments of mathematical representations of uncertain parameters applied to coupled processes (correlations, …)

About developments of mathematical representations of uncertain parameters applied for coupled processes, a joint programme is outlined, since SURAO, GRS, and HZDR collaborated to carry out a complex uncertainty treatment and sensitivity analysis effort based on a coordinated data collection, uncertainty quantification and testing sensitivity analysis methods, applied to geochemical and transport application.

The work for including hydrogeological and geochemical uncertainties into the flow and transport modelling in co-operation between the partners UJV, HZDR and GRS consisted of three different parts. Firstly, it was decided to use the Czech block experiment as a reference model to investigate the impact of uncertainties of hydrogeological and geochemical data on the RN transport in a defined model system. Secondly, GRS developed a tool for probabilistic uncertainty and sensitivity analysis by means of flow and transport simulations in a fractured medium realized by an interface between the transport code d³f++ (“distributed density-driven flow”; Fein et al, 1999, Schneider et al, 2020) and RepoSTAR (a toolbox for controlling and evaluating probabilistic runs, Becker 2016, Reiche 2016). Thirdly, HZDR was working on establishing a workflow from data to model which enables to include different types of host rocks with natural small scale (sub-cm) to meso scale (sub-m) heterogeneities to characterise the geochemical parameter along fluid migration pathways. *Figure 1-1* illustrates how these tasks are interrelated within the DONUT work of task 4.

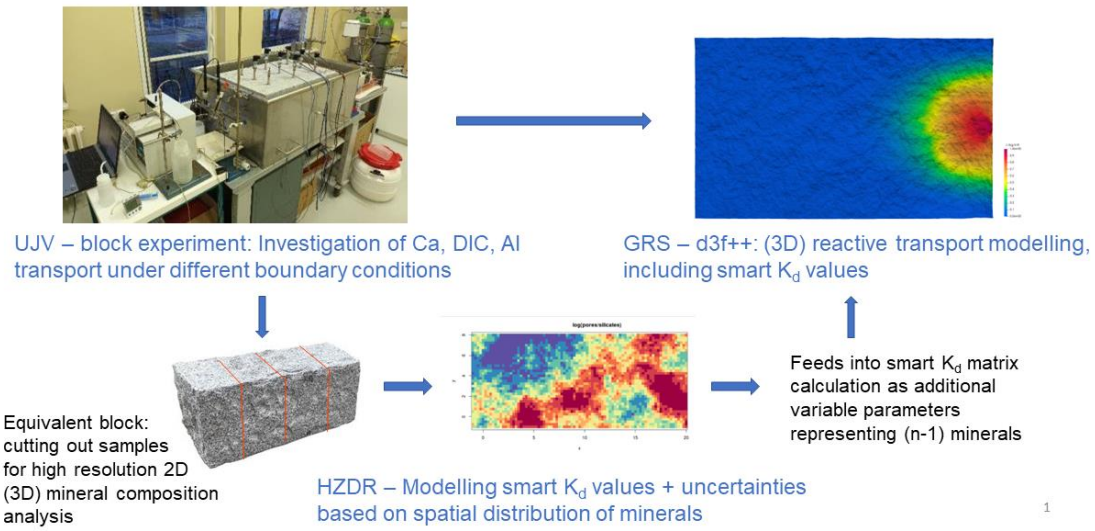


Figure 1-1 Interrelation of the key tasks performed by UJV, HZDR and GRS within task 4 of DONUT

2. Block experiment as an application case

2.1 Objectives

One objective of the work in task 4 was to investigate the impact of uncertainties of hydrogeological and geochemical data on the RN transport in a defined model system. As a basis for this it was decided by the partners from UJV, HZDR and GRS to use as one component a real 2D(3D)-experiment in a fractured medium (see *Figure 1-1*), namely the block constructed at UJV. The original idea was to perform two experiments at different chemical conditions to identify the impact of geochemical conditions on the transport of selected radionuclides (elements) and to test the Smart Kd-concept (Noseck et al, 2018), which is implemented in the transport code d^{3}f++ by accompanying model simulations. However, due to the high effort and the unexpected outcomes, particularly the challenge to establish sorption equilibria, only one experiment was conducted within DONUT. It is foreseen to continue the work in future national cooperations.

The model development and simulations of the block experiment are also the basis for the development and application of the probabilistic approach. The block experiment serves as a reference for these calculations and the calibration of the model performed for the tracer test are used as starting point for the variation of uncertain parameters.

2.2 Preparation stage prior to reactive migration experiment

2.2.1 Preparation of the granite block

The experimental granite block (TH02030543-V10) was constructed within the TACR project TH02030543 (https://starfos.tacr.cz/cs/project/TH02030543). The source bulk rock has been recovered from the quarry in Mrákotín (Czech Republic) and has been divided into 6 identical blocks with dimensions 80 x 50 x 40 cm.

General data provided to the partners (data has been gained within the TACR project TH02030543) are as follows:

- Geometry of the block, including the geometrical information about the fracture
- Permeability of the rock
- Porosity of the rock
- Mineralogical composition
- Aperture, roughness of the fracture
- Location of inflow and outflow openings
- Location of sensors within the fracture

Geometry of the block was derived from a raw data file comprising the cloud of points of roughly 0.1 mm resolution measured by means of 3D laser scanning with Romer Absolute Arm. Fracture aperture is derived from the scan of the reassembled block (*Figure 2-1*).

Figure 2-1 3D scanning of fracture surfaces and reassembled block (TACR TH02030543)

The block was after reassembling equipped with 8 opening which would serve as an inlet/outlet locations (A – H) during flowthrough experiments. Provided pressure data were acquired from hydraulic tests on the experimental block MS2 which was instrumented with 9 pressure sensors in the regular pattern. The final experimental setup is composed by the rock block, on-line detection system (pressure sensors, electrochemical sondes, ion-selective sondes), automatic fraction collector and precise HPLC pump (*Figure 2-2*).

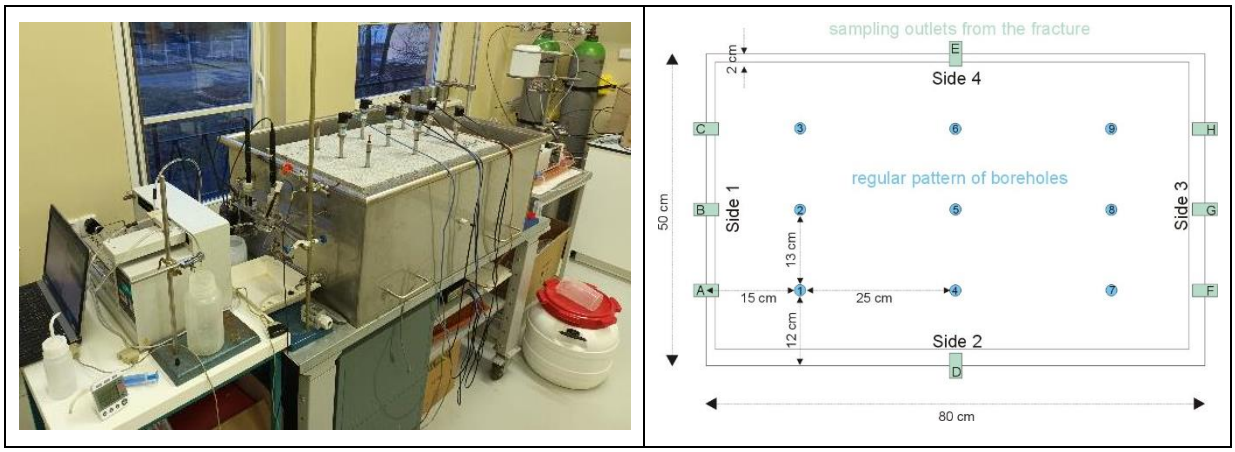


Figure 2-2 Experimental setup with the online detection system and fraction collector (left), plan view of the experimental block with inlet/outlet positions and monitoring boreholes (right)

2.2.2 Selection of radionuclides and their speciation

The following radionuclides (or inactive analogues) were selected to cover a wide range of sorption patterns on one hand and to meet the radiation safety criteria at UJV Rez on the other hand, for example, to avoid that after the experiment the whole block would be classified as contaminated, which would have resulted in an excessive amount of radioactive waste.

U, Ni, Cs, Sr, Ba

U stock solution is usually applied in the form of nitrate ($UO_2(NO_3)_2 \cdot 6H_2O$) by UJV in order to keep it as UO_2^{2+}. However, in SGW2 we can expect U to be in the form of Ca-U(VI)-carbonate complexes which is not sorbing a lot (**Erreur ! Source du renvoi introuvable.***Figure 2-3*).

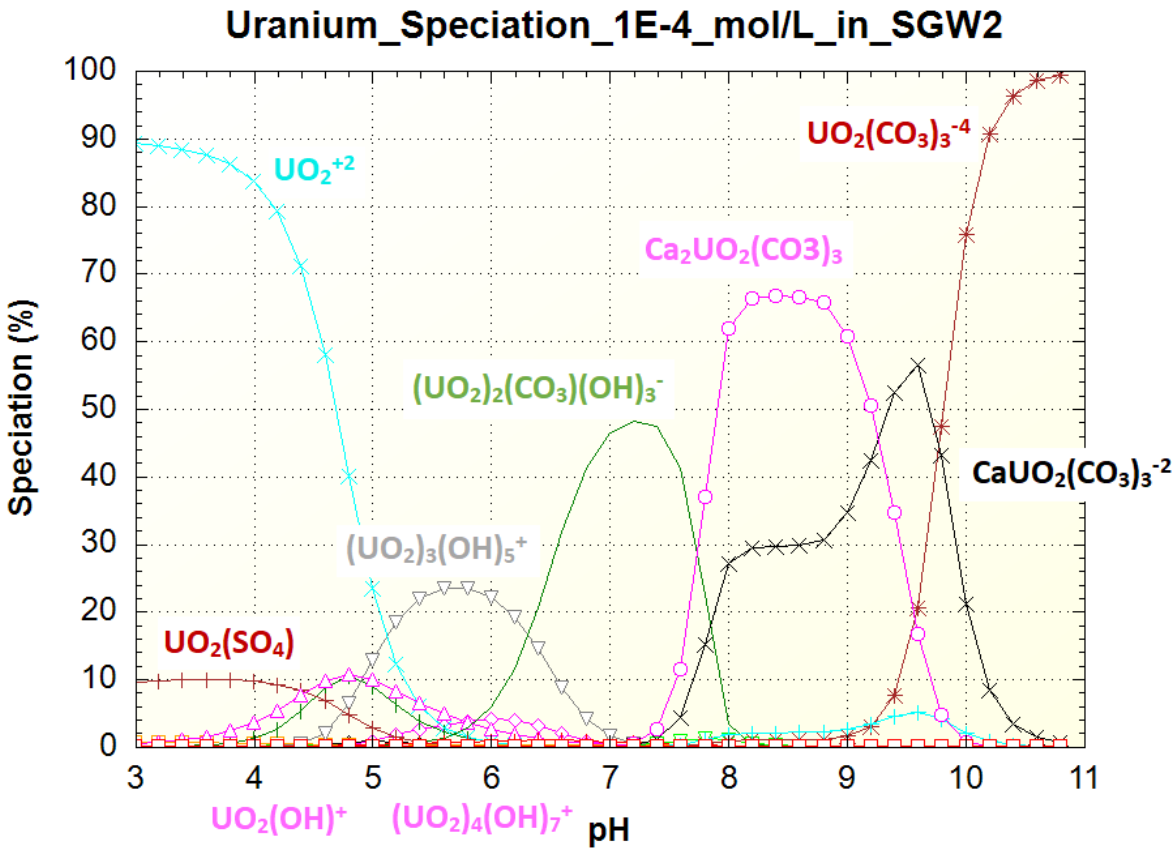
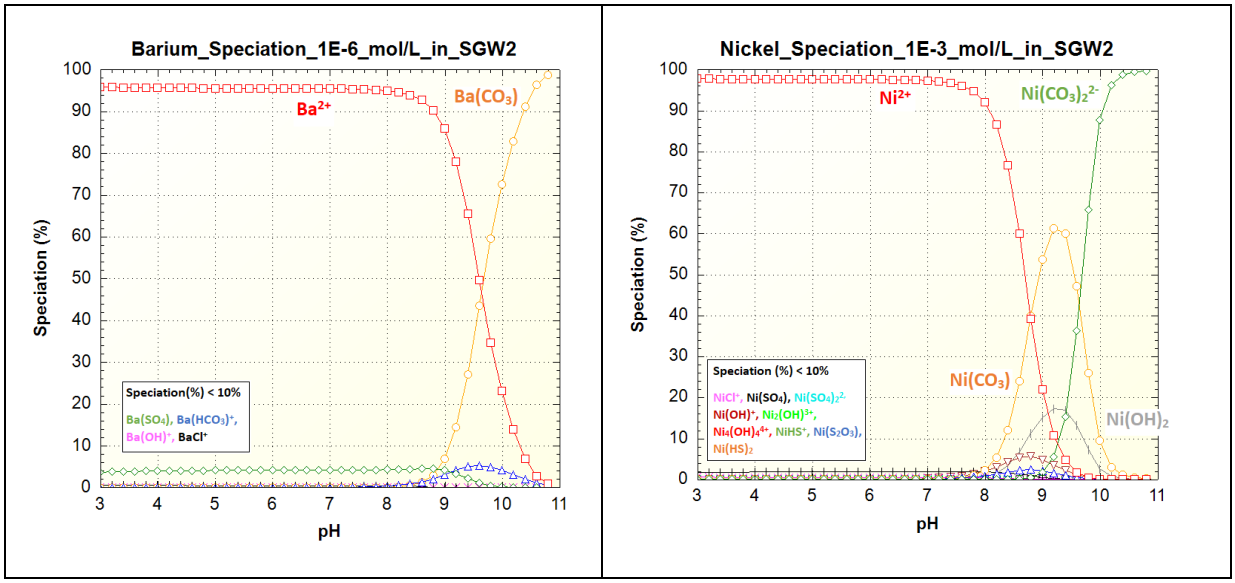


Figure 2-3 Speciation of U in SGW2

The other four elements are applied as inactive analogues in the form of chlorides (CsCl, $NiCl_2$, $SrCl_2$, $BaCl_2$). The speciation diagrams based on the database ThermoChimie (version 11a, Giffault et al., 2014) were constructed for all concerned elements (*Figure 2-4*). In the presence of SGW2 (pH 8.2) we can expect Ba, Ni, Sr and Cs to be predominantly in the cationic form (Ba^{2+}, Ni^{2+}, Sr^{2+} and Cs^{+}).



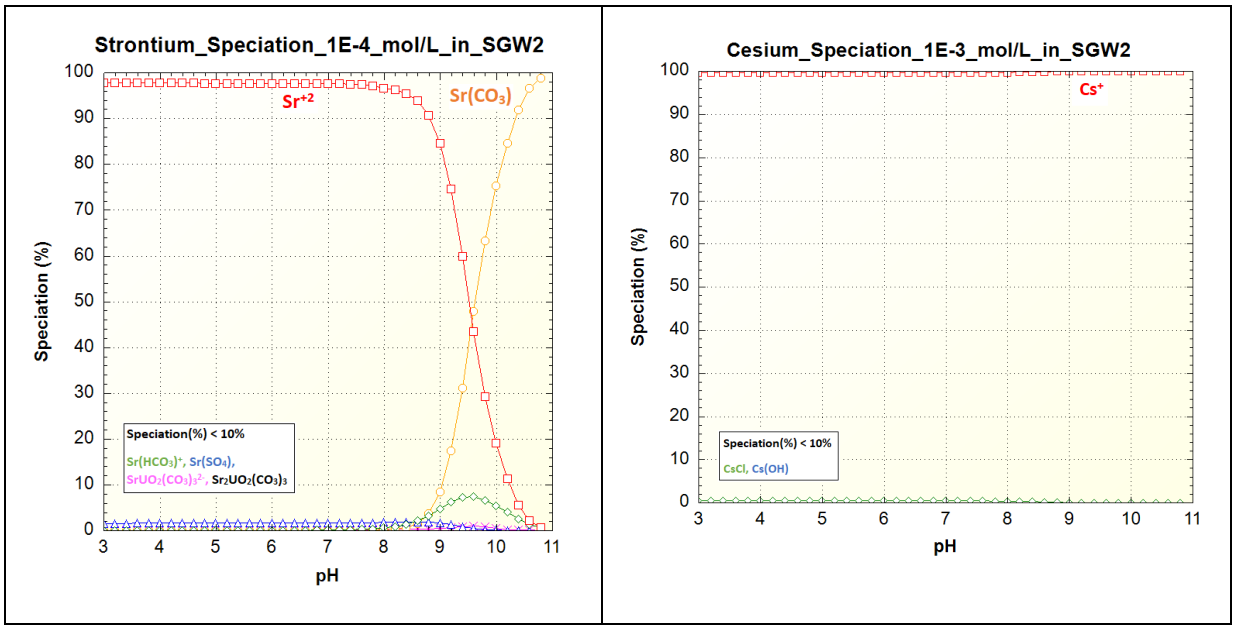


Figure 2-4 Speciation of Ba, Ni, Sr and Cs in SGW2

Speciation calculations with PhreeqC (Parkhurst et al. 2013) and the PSI database (Hummel & Thoenen, 2023) with respect to solubility limits at pH 8.2 revealed, that there are no solubility restrictions for Sr and Cs. The values for both Ni and U can be increased significantly, also avoiding subsequent problems with the ICP-MS detection limits (see next section). Ba astonishingly turned out to be more problematic, here the initial concentration must not exceed 10^{-6} mol/l. The scoping calculation yielded the following limiting solid phases, solubility limits, and recommended start concentrations:

[Ni]	$NiCO_3{:}5.5H_2O$(s)	4.389 10^{-3} mol/L	10^{-3} mol/L
[U]	Becquerelite $CaU_6O_{19}{:}11H_2O$	4.179 10^{-4} mol/L	10^{-4} mol/L
[Ba]	Barite $BaSO_4$	3.084 10^{-6} mol/L	10^{-6} mol/L
[Sr]	no problem expected with 10^{-4} mol/L		
[Cs]	no problem expected with 10^{-3} mol/L		

In addition to aqueous speciation, also the surface speciation is to be forecasted as good as possible to optimize experimental conditions. Here, a combination of models for surface complexation and for ion exchange is used, implementing the so-called smart-Kd approach (Stockmann et al., 2017, www.smartkd-concept.de). To do so, first a respective set of relevant solid phases must be defined. Here, the mineral composition given for the whole rock (sample MS6) based on XRD characterization was used for sorption modelling, i.e. 42.5% quartz, 25 % K-feldspar, 30 % plagioclase (corresponding to Ca-albite or oligoclase) and 2.5 % mica (most likely biotite). Based on this mineral set, the mineral-specific sorption data base RES³T was scanned for respective surface complexation model (SCM) data, as well as additional original literature to capture conventional Kd values to close data gaps. The latter was especially necessary for U and Ni sorption onto biotite as there RES³T did not provide any reliable SCM data and diverting to illite as proxy was too ambitious an extrapolation. Concerning Ni sorption onto biotite (Ticknor, 1994), the Kd [m3/kg] was reported to range between 2.610 and 2.900, with an estimation of 2.750 for pH 8.2. With respect to U sorption onto biotite, Idemitsu et al, 1994, reported that around pH 7.5 there is a strong decrease in Kd values [m3/kg], going down to 0.6-0.7 at pH 8.2. Pan et al, 2022 reported a decrease not that strong, leading to Kd values [m3/kg] of around 1 at pH 8.2.

2.2.3 Boundary conditions, flow and sampling regime

The granite block was during the experiment designed with one inlet (location G) and one outlet (location B) (*Figure 2-5*). Before the start of the actual experiment, it was conditioned with 2 days with SGW, so

that an equilibrium could be established. The solution spiked with radionuclides was added by means of an injection through three-way valve over a period of 20 min. The flow rate of the SGW was 5 ml/min. In addition to the spiked elements, iodides were monitored as inert tracers, as well as important elements typical for geological questions: K, Ca, Al, Rb, Co and Fe.

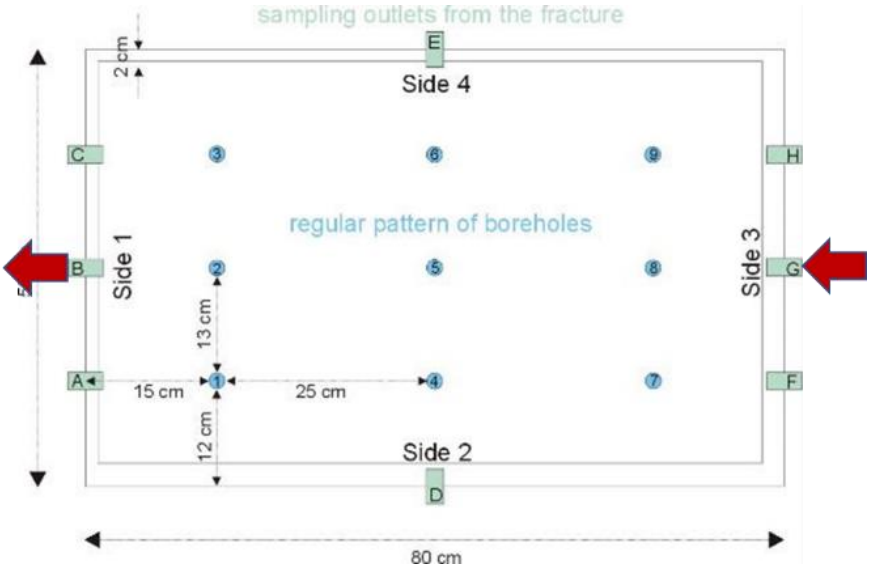


Figure 2-5 Plan view of the experimental block with the arrow marks of inlet (G) and outlet (B) during the experiment

As a flushing medium two different waters were discussed, (1) synthetic granitic water (SGW2) of pH 8.2 and (2) lower pH (6-7) water with reduced mineralization. Only one type of water (SGW2) was selected for our experiment. The theoretical chemical composition is depicted in the *Table 2-1*. The electric conductivity is 318 µS/cm and Eh 157,3 mV.

Table 2-1 Chemical composition of SGW2

Cations / anions	**c_theor (mg/l)**
Na	16.5
K	2.1
Mg	8.3
Ca	37.3
Cl-	3.5
NO_3-	0
SO_4^{2-}	21.9
HCO_3	168.7
CO_3^{2-}	0
OH-	0

The flow velocity impacts the detection type: In case of significantly lower flow velocities detection/sampling rates could be lower. Conventional fraction samplers would then allow a “broader“ elemental analysis than ISEs.

Pump position is at inlet side of the granite block. All experiments was designed with one inlet and one outlet to keep the system as simple as possible.

At the early-stage various flow rates were considered (ranging from 0,017 to 5 ml/min). The selected flow rate strongly impacts the number and volume of collected samples which are the crucial parameters

determining the financial and logistics requirements. Based on these characteristics and calculated K_d values the flow rate of 5 ml/min and sampling volume 10-20 ml was selected. Sampling frequency based on flow rate and suspected breakthrough curves of the tracers was designed accordingly: 0 – 6 hours (every 10 min), 6 – 12 hours (every 20 min), 12 – 24 hours (every 40 min), 24 – 90 hours (every 60 min) and 90 – end (every 120 min).

2.2.4 Time regime

The time regime was deeply discussed in advance of the experiment. There are particularly two conflicting "boundary conditions". On the one hand it is seen necessary to establish low flow conditions to reach equilibrium in the sorption processes during transport through the fracture. On the other hand the duration of the experiment is limited by practical issues. Although the whole setup is rather automatized and experiments can be run overnight and over the weekend after proper announcement, the duration should not increase three weeks. With a longer duration of the experiment the number of samples and – depending on the flow – also the amount of artificial groundwater, will increase and therewith the expenses. Furthermore, the probability of power breakdowns would increase, which can disturb the experiment.

A further requirement is that within the experimental duration the peaks of the breakthrough curves of all tracers should be visible, i.e. even sorbing tracers should have been transported through the block within the duration of the experiment. In order to optimize the system, particularly the flow conditions, predictive transport calculations have been performed (see Section 2.2.5). The final conditions are then described in Section 2.6.1

2.2.5 Analytics

Analytical procedures in UJV

On-line measurement of the concentration of iodide at the outlet was performed by combined ion selected electrode (Mettler Toledo, perfection comb I) in the flowthrough cell. At the same time samples had been collected by automatic fraction collector at specified time periods. Concentration of iodide was subsequently analysed by capillary electrophoresis (Agilent 7100 CE) to confirm the break-through curve obtained on-line. On-line detection of electric conductivity, pH and redox were also measured in the flowthrough cells during the experiment with WTW sondes.

Alkalinity (HCO_3-, CO_3^{2-}) was analysed by potentiometric titration.

Determination of U and Ni concentration in the solutions was realized by two methods, both of which were using complexation reaction of U and Ni ions altogether with arsenazo III agent. The different scale of the reaction due to different concentration of uranium and nickel ions was being observed via UV-VIS and adsorptive stripping voltammetry. In both cases, the solution was adjusted to pH 1,85 using the addition of Britton-Robinson buffer in the volume ratio of 100 µl of AZ III / 1 ml of Britton-Robinson buffer and 1 ml of the sample.

For the electrochemical analysis, printed electrochemical cells provided by Metrohm were used. As a working electrode, these cells use glassy carbon surface as well as for the auxiliary electrode, Ag is used as a quasi-reference electrode. The solution was applied to the cell as a 100 µl drop. Within this setup, adsorptive stripping voltammetry method was used with following parameters: Accumulation time 120 s, scan rate 20 mV/s, pulse 50 mV, initial potential -500 mV, final potential +1.1 V.

Analytical procedures in HZDR

The chemical composition of the aqueous solution at the outlet was carried out by ICP-MS analysis at the HZDR. All samples to HZDR were acidified by 100 µL of conc. HNO_3 and shipped all together. The

transport of samples had to fulfil all restrictions and regulations on the Czech and German side. The list of parameters to be analysed by ICP-MS contained: Co, Rb, K, Ca, Ba, Sr, Cs, Ni, U, Fe and Al.

The samples had been analysed at the HZDR laboratories with an iCAP RQ (Thermo Scientific). The analysis of elements comprised Al, K, Ca, Fe, Co, Ni, Rb, Sr, Cs, Ba and U. The calibration solution was adapted to the expected concentration ranges. The standards for the solutions had been purchased from Merck VI and Berndt Kraft (only K and Cs). The elements Sc and Rh had been used as internal standards. The first hundred samples had been measured as diluted samples (1:100 with 1% HNO3) and undiluted samples. After these samples dilution proved unnecessary.

2.3 Set-up of geometric model in d³f++ and predictive transport calculations

The data set for the geometry of the model as well as for tracer experiments in the fracture, looking at different fracture variables and boundary conditions was provided by UJV. It is based on the characterization of the fracture as described in Section 2.2. The geometric description was made available in the form of grid data for fracture bottom, fracture top and fracture aperture with a resolution of 1 mm. Because of data inconsistencies, in addition to fracture bottom the fracture aperture information was used instead of the fracture top data to develop the model. Furthermore, based on the rock matrix properties and the estimated transport velocity the height of the model was reduced from 400 mm to 60 mm to save computing capacity (diffusive transport deep into the rock matrix is not of relevance for the experiments). The model geometry is shown in *Figure 2-6*. The coarse grid is refined three times to an element length of 2 mm during the solving process (geometric multigrid method). Finally, a special multigrid construction was chosen to optimize the relation between grid resolution and computational effort.

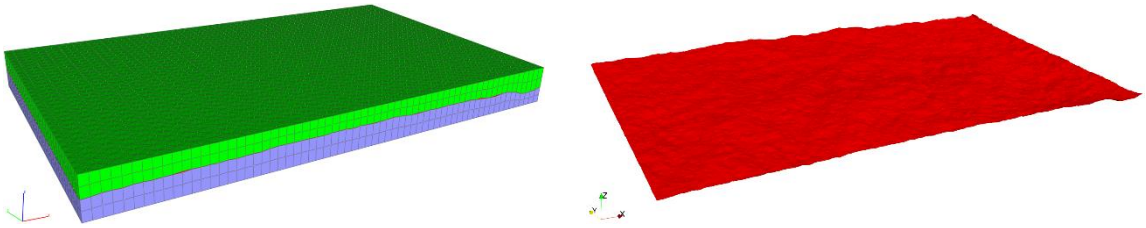


Figure 2-6 Complete 3d geometric model (left; coarse grid) and fracture (right; fine grid)

Based on this model geometry, consisting of 7.9 million prism elements, the code d³f++ was used to simulate RN transport in the block experiment. In the fracture, the porosity was set to 1.0 and a cubic law approach, based on the point-wise aperture measures, was chosen for the permeability. The matrix porosity was set to 0.007 and the permeability to 10^{-18} m². The model was calibrated by a conservative tracer experiment realized by UJV on 10.10.2019, see *Figure 2-7*. Over a period of 60 s 20 ml/min of a NaCl tracer (358 kg/m³) were injected into the fracture in one central point, and outflow concentration was measured in one of the two outlets at the other side of the block.

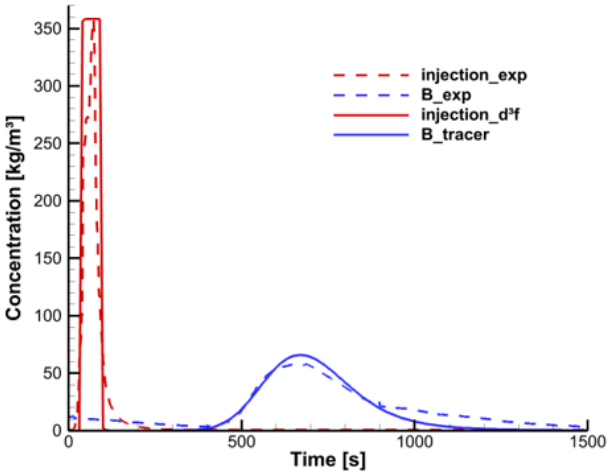


Figure 2-7 Model calibration, based on the conservative tracer experiment of 2019

Figure 2-8 shows the simulated tracer concentration in fracture and matrix after a model time of 600 s. The concentration distribution in the fracture after 400 s is shown in *Figure 2-9*. It illustrates the influence of the fracture roughness and the permeability distribution to the tracer transport. The measured concentrations at outlet B are met very well by simulation.

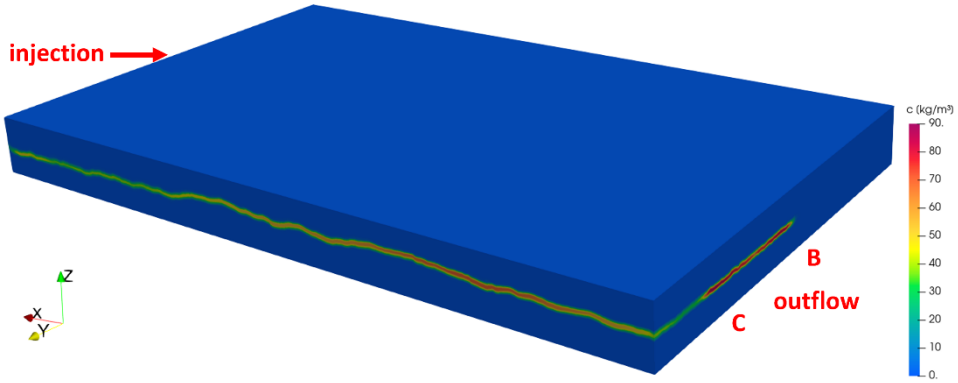


Figure 2-8 Simulated tracer concentration in fracture and matrix after 600 s

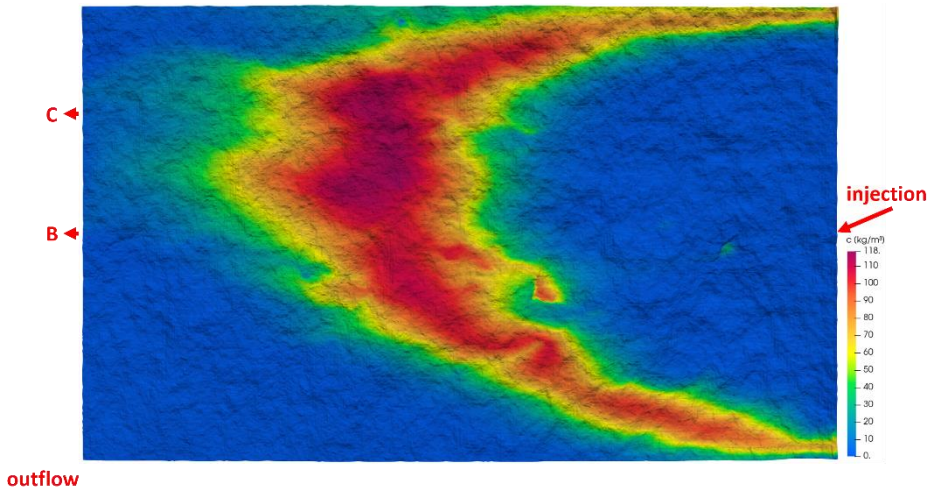


Figure 2-9 Simulated concentration distribution of a non-sorbing tracer under the conditions of the block experiment performed by UJV on 10 October 2019

With regard to the block experiment the model pursues two objectives. Firstly, the calibrated model is used to predict concentrations and travel times of generic pollutants with different Kd values and different flow velocities in the fracture of the block experiment. This was done in order to plan the layout of the experiment. The main aim was to identify the duration of the experiment and the appropriate design to obtain tracer concentrations at the outlet significantly above the detection limits. Secondly, the model is used to accompany the experiment by simulations with d^{3}f++ and the implemented smart Kd-concept, see Section 2.7.

One important aim was to provide equilibrium conditions for sorption processes in the experiment, i.e., to reduce the flow velocity to as low as feasible values. Therefore, the computations were repeated with different inflow rates down to 2 ml/min and 0.2 ml/min and an outflow only at point B (as later realized in the experiment). *Figure 2-10* shows the results for 0.2 ml/min and varying distribution coefficients.

The maximum concentrations at outlet B and the travel times (arrival of peak maximum) for all predictive calculations are summarized in *Table 2-2*.

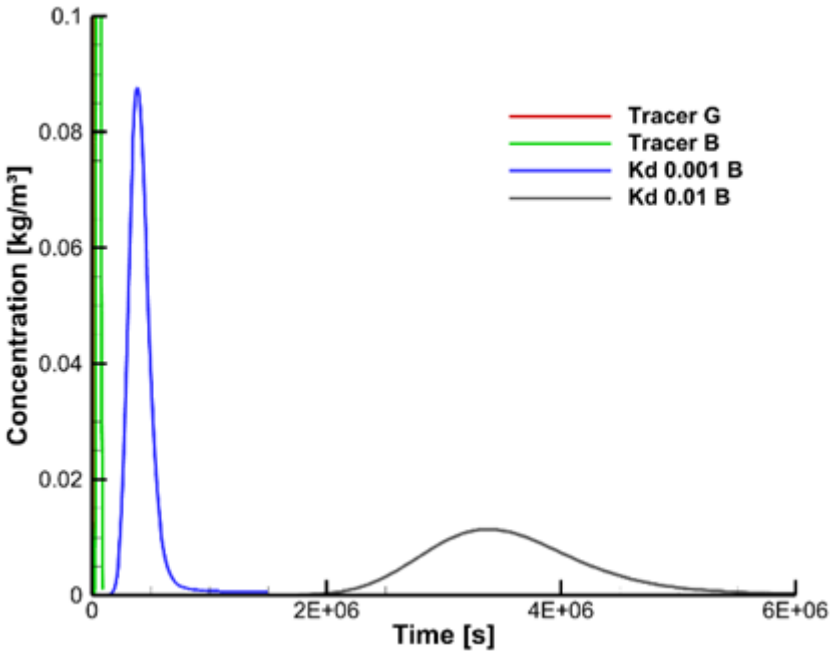


Figure 2-10 Travel time predictions with an inflow rate of 0.2 ml/min and varying distribution coefficients

Table 2-2 Key results from predictive calculations of the break-through curves for the layout of the block experiment.

Kd [m³/kg]	time c_max in B [s]	time c_max in B	c_max in B [kg/m³][1]	Reduction factor c_max
Inflow 20 ml/min				
0	590	9.833 min	65.789	5.4
0.001	4899	82 min	9.618	37.2
0.01	42890	12 h	1.017	352
0.1	422419	4.875 d	0.103	3476
Inflow 1 ml/min				
0	9 900	2.75 h	88.6	4
0.001	78 000	21.70 h	9.92	36.1
0.01	680 000	7.78 d	1.28	278
Inflow 0.2 ml/min				
0	50 000	13.89 h	83	4.3
0.001	385 500	4.46 d	8.8	40.7
0.01	3 355 000	38.83 d	1.14	314

[1] For all calculations c_max was related to the same input mass of the tracer (realized by an increased inflow interval with decreasing flow velocity)

Speciation calculations for all tracers to be injected into the experiment indicated that for Ni distribution coefficients in the range of 0.1 m^3/kg are possible. As the calculations show, to observe the peak of a tracer with Kd-value of 0.1 m^3/kg, the flow must not be lower than 5 ml/min. Therefore, it was finally decided to perform the experiment at this flow rate of 5 ml/min. It was discussed in advance that this is a compromise, since equilibrium sorption conditions might not fully be met. As also shown the concentration of the tracers can be adapted by the injection time. If the same amount of tracer is injected the maximum concentrations are mainly impacted by the Kd-value and rather similar for the different flow conditions.

2.4 Pre-tests: Leaching tests at different flowrates

Based on the discussion within the team the complementary flowthrough leaching tests on the experimental block were conducted. The first set of samples were collected at the flowrate 0,017ml/min over the period of 21 days during flushing of SGW2. The background concentration of U, Cs, Ba, Sr and Ni were analysed in the samples collected 2, 7, 14 and 21 days after the pumping was commenced (**Erreur ! Source du renvoi introuvable.**). The concentration of selected elements were also analysed in the granitic rock itself (Table 2-3).

Table 2-3 Concentration of selected elements in the granitic rock

	U	**Cs**	**Ba**	**Sr**	**Ni**
sample	ppm = mg/kg	ppm = mg/kg	ppm = mg/kg	ppm = mg/kg	ppm = mg/kg
granite	8.0	17.3	220.6	58.5	113.1

Additional leaching test was performed also at the flowrate 5 ml/min. This time only two samples were collected once the equilibration of the system was reached (on-line detection of el. conductivity and pH) (*Figure 2-11*, *Table 2-4*). Concentration of all elements are significantly lower than those collected at the flowrate 0,017 ml/min.

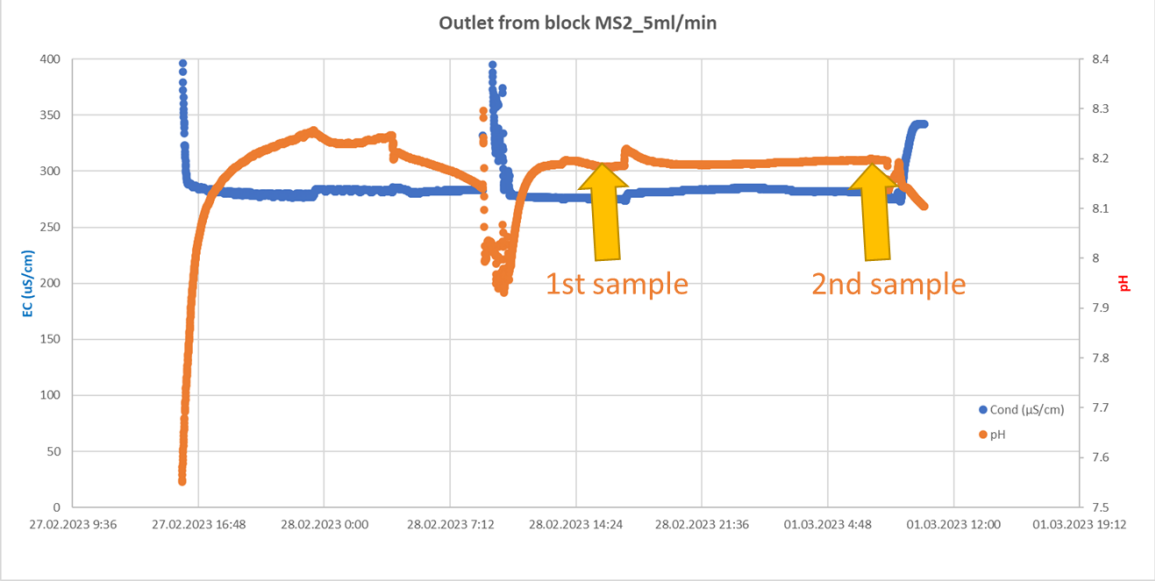


Figure 2-11 On-line monitoring of system equilibrium (conductivity and pH)

Table 2-4 Concentration of elements in the leachates at the flowrate 5 ml/min

parameter	**unit**	**SGW2**	**1. sample (28/2/23)**	**2. sample (1/3/23)**
U	mg/l	3.07E-05	1.14E-03	1.19E-03

Cs	mg/l	2.23E-05	1.35E-03	1.37E-03
Ba	mg/l	2.50E-02	3.26E-02	3.17E-02
Sr	mg/l	2.80E-03	1.97E-02	1.89E-02
Ni	mg/l	1.14E-03	2.77E-02	2.76E-02

2.5 Reference tracer test with iodide

Tracer tests with inactive iodide were performed at two specific flowrates (0,5 and 5 ml/min). The experimental setup was identical in both cases. The injected tracer 0,01M KI was detected both online (ISE) and offline (sampler collector). At the same time pressure gradient was monitored on the inlet/outlet and also within the block in monitoring boreholes.

- Tracer test with iodide at 0,5 ml/min

Tracer test performed at the flowrate 0,5 ml/min served as an initial approach to test the final setup and all detection and monitoring systems. The amount of injected tracer was 19,31 g and the flushing medium was SGW2. The experiment was finished 21 hours after injection. The first arrival of iodide in the outlet was reached in 188 min and the maximum concentration in 278 min. During the sampling period altogether 22 samples were collected (5 ml/sample). The detected BTCs from both online/offline records exhibit good compliance. The mass recovery reached up to ca 90%.

- Tracer test with iodide at 5 ml/min

After the computation the decision to increase the flowrate from 0,5 to 5 ml/min was made. Tracer test performed at this flowrate served as a reference to the planned reactive experiment. The amount of injected tracer was 98,01 g and the flushing medium was SGW2. The experiment was finished 20 hours after injection. The first arrival of iodide in the outlet was reached in 17,5 min and the maximum concentration in 40 min. During the sampling period altogether 24 samples were collected (7,5 ml/sample). The detected BTCs from both online/offline records exhibit good compliance (*Figure 2-12*). The mass recovery reached also ca 90%.

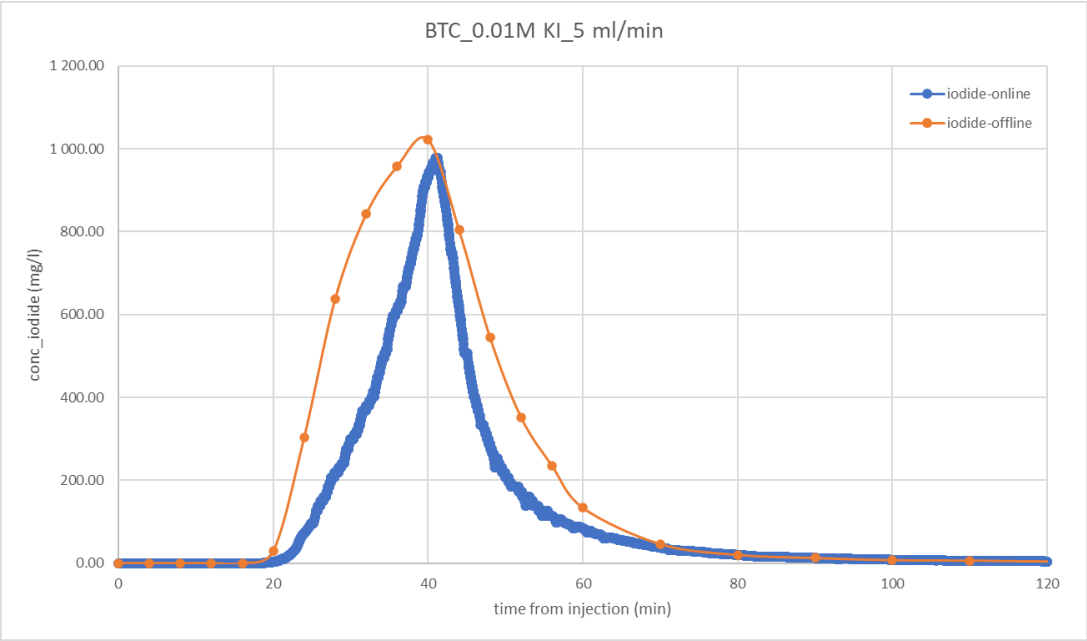


Figure 2-12 Breakthrough curves of iodide detected on-line (ISE) and off-line (fraction collector) at the flowrate 5 ml/min

2.6 Reactive migration experiment on the granitic block

2.6.1 Final experimental setup

The final reactive migration experiment was conducted after several team meetings which were focusing on preparation all necessary steps in detail. The experimental setup was already tested during the previous inactive tracer tests with iodide, this time only complemented with laboratory scale (KERN) for on-line monitoring of waste balance (*Figure 2-13*).

Figure 2-13 Final setup for reactive experiment

In total, 5 batches of SGW2 (40 l each) were prepared continuously to cover the needs of the experiment planned for 3 weeks. Fraction collector (ECOM) was programmed to collect 15 ml, resp. 18 ml of sample for subsequent analysis (UJV, HZDR). The stabile flow was ensured by precise HPLC pump (INGOS). On-line concentration of iodide was detected by combined ISE (Mettler Toledo) connected to the data logger (GRYF).

Samples had been treated and shipped to HZDR as described in section 2.2.5.

2.6.2 Results

The reactive experiment was carried out in the period 20.4. – 12.5.2023. A cocktail with iodide as non-sorbing tracer and the elements U, Cs, Ni, Sr and Ba with concentrations as described in Section 2.2.2 was injected and in total 355 samples were collected. Each sample was divided into two vials for analysis in UJV and HZDR. In order to monitor on-line the progress of the advective fraction of the tracer cocktail it was decided to label the cocktail with 0,01M NaI. The comparison of iodide BTC with the one carried out during the reference tracer test on 1.3.2023 are in a good agreement (*Figure 2-14*).

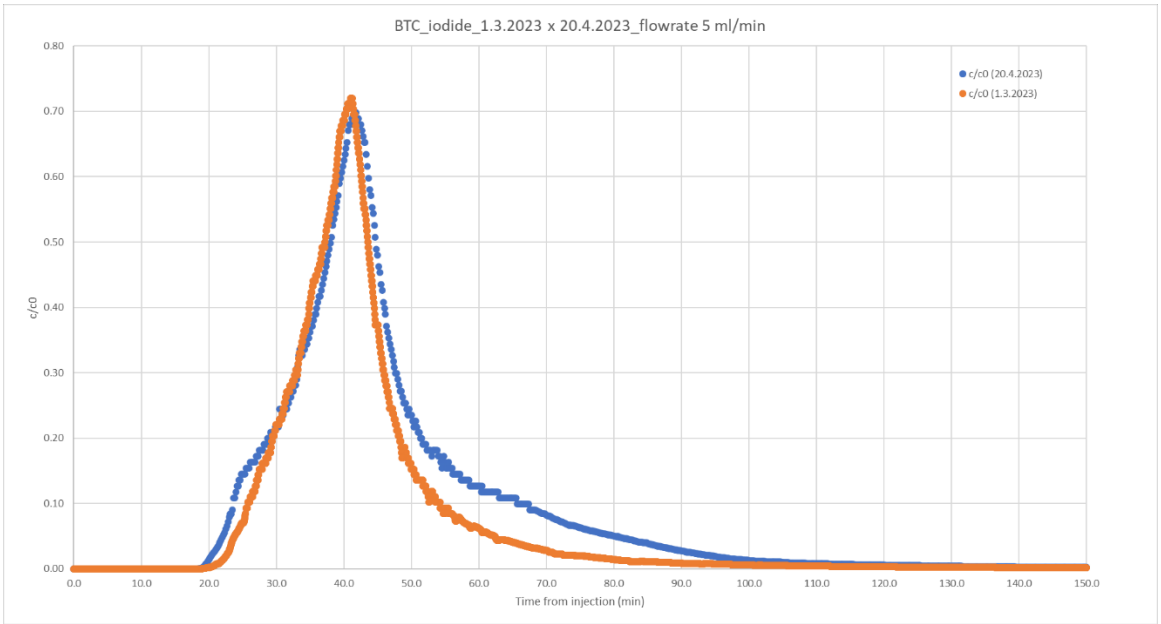


Figure 2-14 Comparison of iodide BTCs from two experiments (1.3.2023 and 20.4.2023)

For the UV-VIS measurements, comparison of results for samples No. 1, 3, 5, 7, 9, 15, 20, 25 a 30 is given in the following figures (*Figure 2-15*) showing that only for samples 5-9, the significant amount of uranium was present.

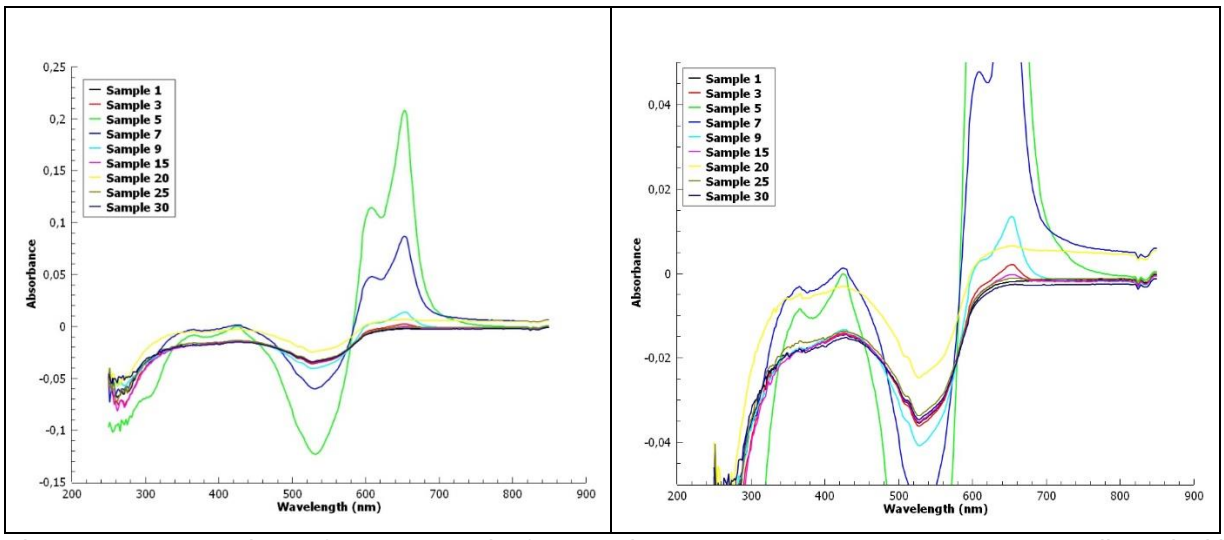


Figure 2-15 Comparison of UV-VIS results for samples No. 1, 3, 5, 7, 9, 15, 20, 25 a 30 adjusted with Britton-Robinson buffer and AZIII

From the electrochemical analysis it can be seen in the following figure that in the row of samples 5, 7, 30 (*Figure 2-16*) the decrease is in agreement with UV-VIS measurement.

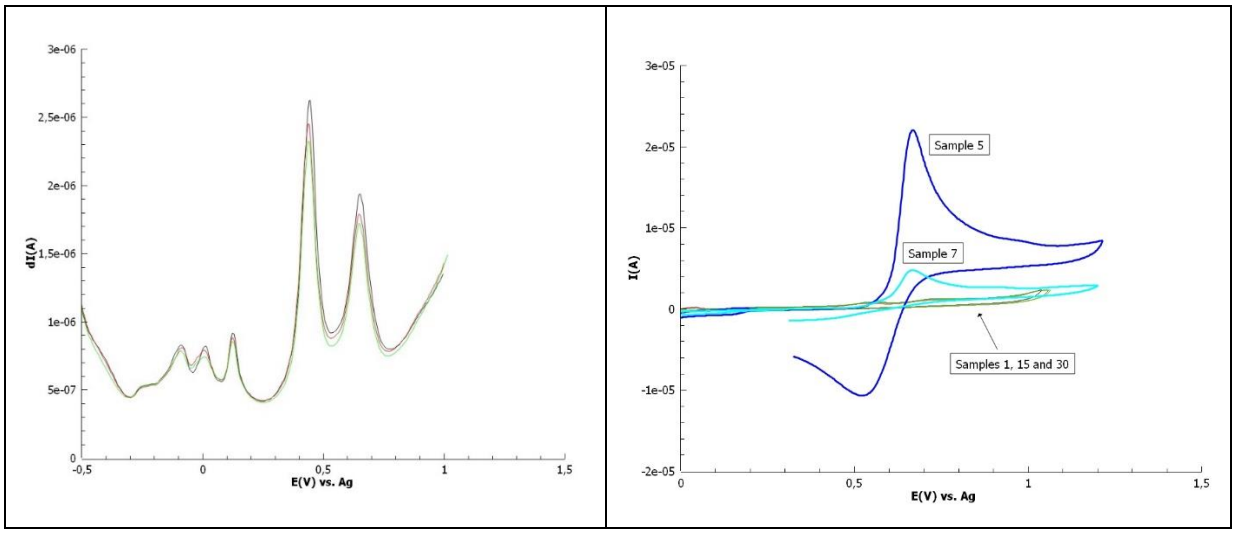


Figure 2-16 Electrochemical analysis - printed electrochemical cells (left), graphene-paste electrode for direct electrochemical measurement (right)

Figure 2-17 shows the breakthrough curves of all injected tracers normalized to its injection concentration. All curves except for Cs show a peak at the arrival time of the non-sorbing tracer iodide. The curve for uranium is very similar to the iodide curve (*Figure 2-12*). The curves for Ni, Sr and Ba show a very strong tailing while for Cs the breakthrough curve is very broad with its peak maximum occurring after more than 100 minutes.

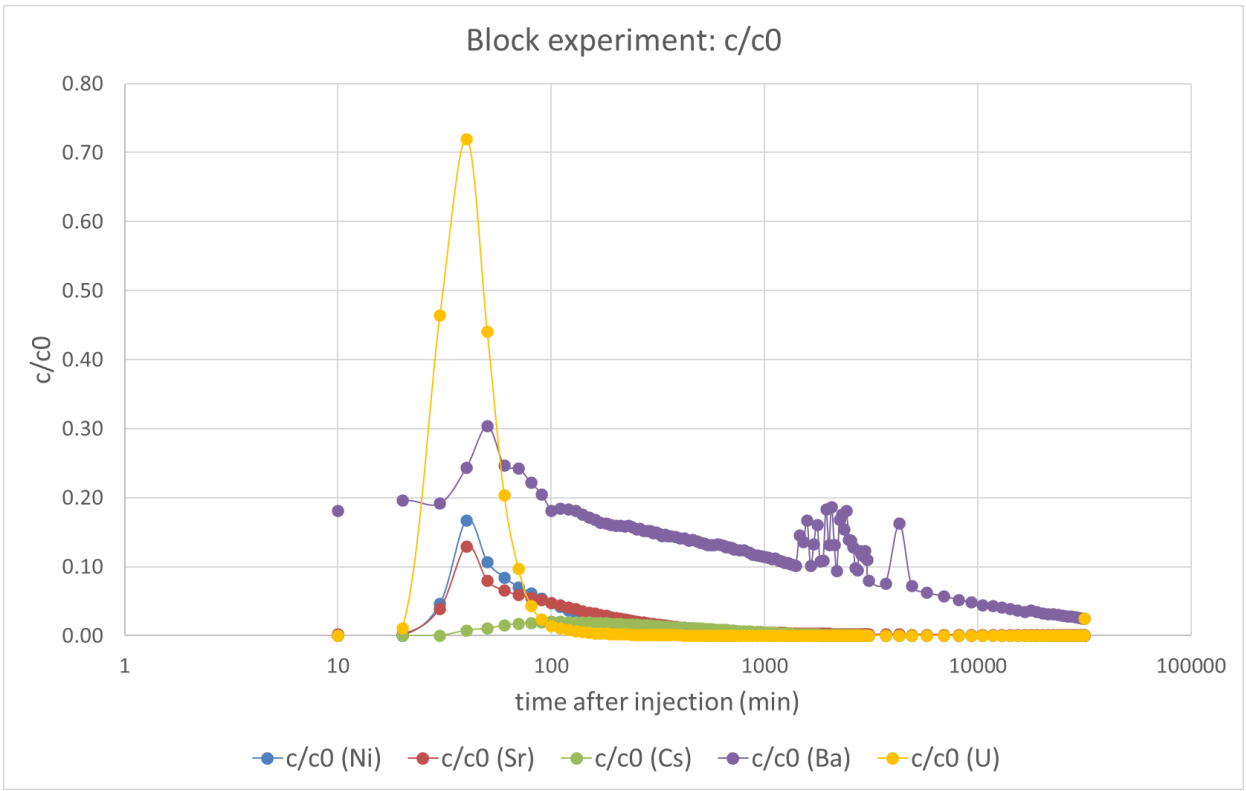


Figure 2-17 Breakthrough curves of the cocktail elements normalized to the initial source concentration

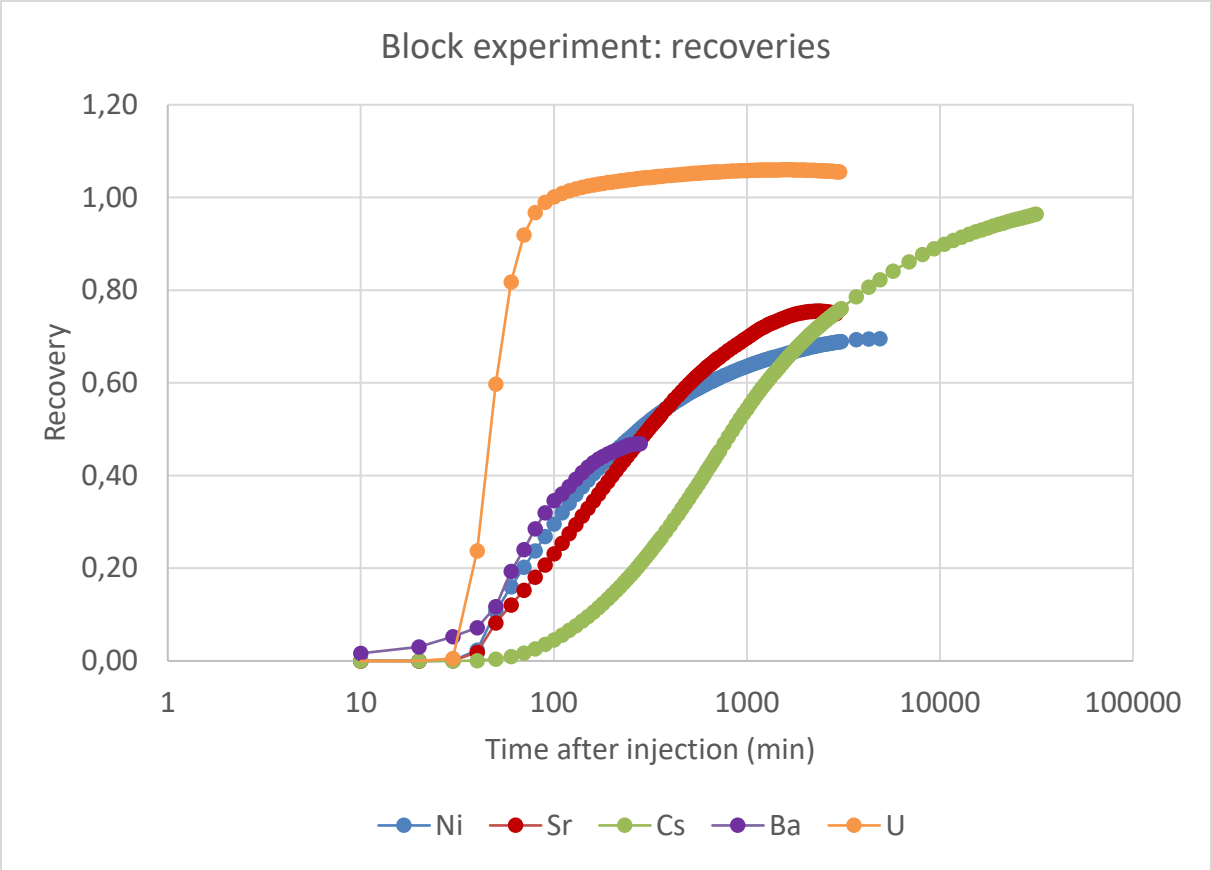


Figure 2-18 Calculated recoveries of the breakthrough curves

Figure 2-18 shows the recoveries for the injected tracers, indicating that most of the tracers is recovered after 10 000 minutes. The recoveries of U and Cs are reaching about 100 %. For Sr and Ni a recovery of 70 to 80 % is observed after a few thousand minutes. Only the curve for Ba seems to provide a lower recovery of about 50 %. However, the calculated recovery strongly depends on the assumed background concentration before tracer injection. This particularly concerns the curve for Ba, where the background concentration is decreasing significantly below the concentration before breakthrough of the injected solution (*Figure 2-17*).

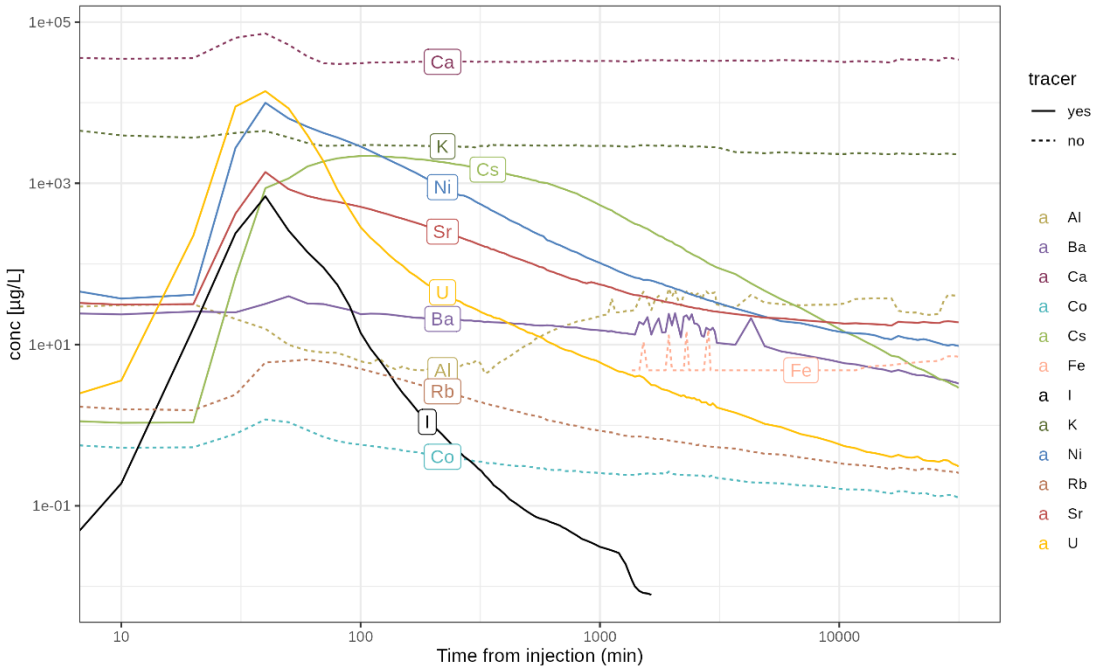


Figure 2-19 Breakthrough curves of all monitored elements

Not only tracer elements are showing a peak, also all monitored elements, except for Al, which even decreases in its concentration during the first 4hs (*Figure 2-19*).

The ratios of the elements vary over time. Roughly three periods can be distinguished in terms of composition of element concentrations: the initial phase from 0 – 110 min after injection, the middle phase from 111- 800 min, and the final phase from 800 min onwards.

Log-ratio variation heatmaps are used to show the log-ratio variation of each element against each other: If the variance of an element's log-ratio within a period is small, it means that the elements behave similarly in terms of retention in the block. If the variance is large, it indicates that the elements respond differently to the conditions within the chosen period. The advantage of using log-ratios is that the difference between major and trace elements is unimportant as the log-ratio measures the distance between elements.

In *Figure 2-20*, the heatmap of the log ratio variation of the samples over the whole period ("Total" sub-figure) shows that log(U/I) has the lowest variance. Also log(Ni/I) has a low variance, followed by log(Ni/U), confirming the general observation that U and Ni are not very well retained and behave most similarly to I among the monitored elements. Ni could also be grouped together with Rb, Cs and Sr in one group of elements, and Ba, Ca, Co and K in a second group of elements, suggesting that the elements within each group behave similarly over the whole period. Al shows a markedly different behaviour.

This overall behaviour is very different from the three periods considered separately. During the initial phase (sub-picture "Minutes 0 - 110") the tracers form a group, except for Ba, which is clearly associated with the other monitored elements. Among the tracers, U and I are most similar, then Ni and Cs, and U, I, Ni and Cs react similarly in general. Sr has a slightly different behaviour. Looking at the monitored elements, Rb seems to be significantly affected by the injection of the tracers, with a much greater variance in the log ratio compared to all the other elements. In the middle phase (sub-picture "Minutes 111 - 800") the U concentrations begin to decouple from the I concentrations, leaving U and Ni as the most similarly behaving elements. Among the tracers, Cs is discharged very differently in this phase compared to the other tracers, suggesting that Cs reacts with the surfaces and is retained. On the other hand, Rb is comparatively more reactive than the tracers and shows even better (relative) run-off than Co. Among the monitored elements and Ba, Al still shows the largest log-ratio variance with the other elements, except for Co. In the final phase, among the tracers, U and Cs show the smallest log-ratio

variance, followed by log(Cs/Ni) and then log(U/Ni). Compared to all other elements, Cs shows a rather large log-ratio variance. The main elements K, Ca and Al form a second group of elements with similar behaviour within the group. The concentrations shown in *Figure 2-19* suggest that the concentrations are back to equilibration. The third group would be Co, Ba, Rb and Sr with comparably stable log-ratios over this time.

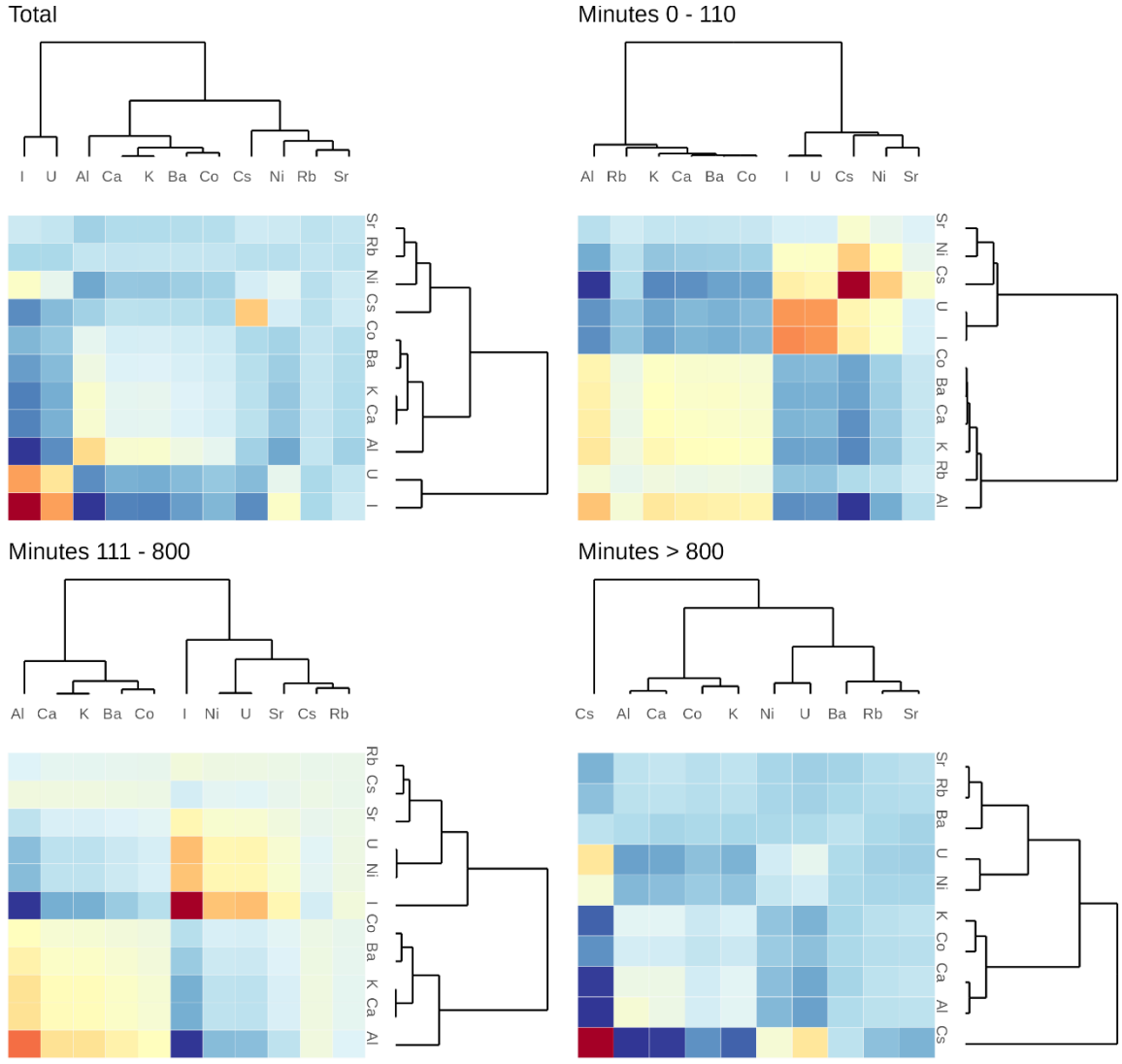


Figure 2-20 The heatmap of the log ratio variation of the samples

In summary, U initially behaves like the inert tracer I, but over time becomes more similar to Ni and Cs. Ni, together with the tracers Sr and Cs, was poorly retained, although Cs had good retention in the middle phase. Ba was generally very well retained or added at concentrations too low to produce a significant breakthrough curve. However, after the breakthrough peak of Ba, the concentration decreases continuously over the monitored time, similar to the monitored elements Rb and Co. Also, the recovery curves indicate that Ba is retained over the period of the experiment.

It should be noted that due to the high flow rates, the fluids were most likely not in equilibrium with the solid phases and there may also be some kinetic component to the reaction.

2.7 Transport modelling

When the experimental results became available, further transport simulations adapted to the results have been performed. The flow aperture and the dispersion lengths was slightly adapted to describe the breakthrough curves of the non-sorbing iodide and the uranium tracer, since the measured curves

indicated that uranium has not been retarded at all. The results for simulated and measured curves are shown in *Figure 2-21*.

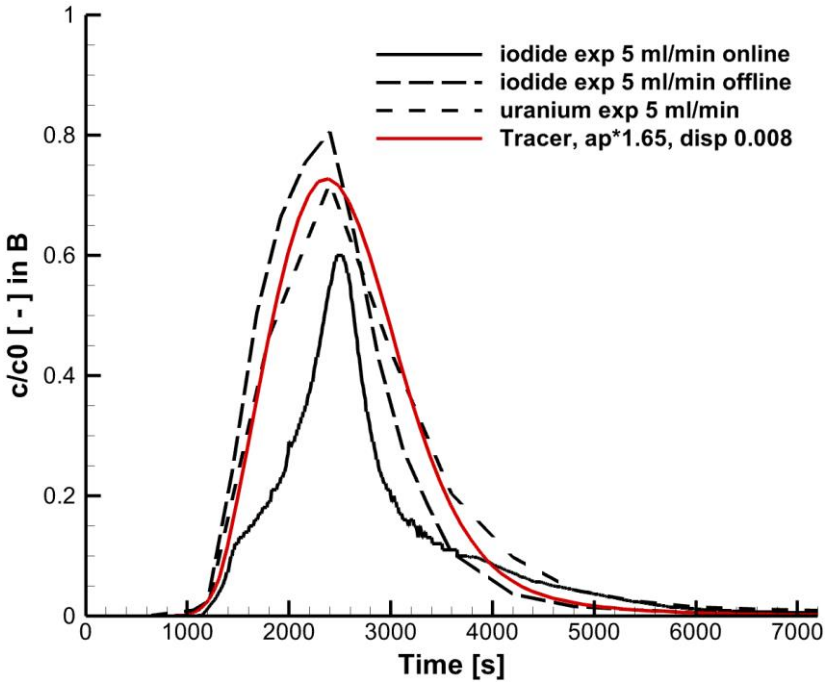


Figure 2-21 Simulated breakthrough curves of a non sorbing tracer and measured breakthrough curves of iodide and uranium

Figure 2-22 shows the same curves but also the measured breakthrough curves of the other elements of the injection cocktail. In addition, a simulation curve for a low sorbing element with a K_d value of 0.001 m^3/kg is added. It is evident that the maxima of the breakthrough curves for Ni, Sr and Ba are similar to the non-sorbing tracer. However, these curves show a strong tailing. This could be due to matrix diffusion, but a more likely explanation is that the sorption processes are not in equilibrium, since the majority of the injected mass is transported through the fracture of the block in less than one hour. The breakthrough curve for Cs does not show a peak at ideal tracer arrival but a very broad, retarded curve, appearing much earlier but being broader than the curve obtained for tracer with equilibrium sorption and a Kd-value of 0.001 m^3/kg. Sorption experiments with Sr and Cs in a system with SGW2 water and granite reveal Kd-values in the range of 0.001 and 0.01. Speciation calculations for Ni indicate Kd-values above 0.1 m^3/kg. The differences between sorption values obtained in batch systems or from calculations on the one hand and the observations from the block experiment on the other hand give evidence that the volume flow was too high to allow complete adjustment of equilibrium conditions.

In order to check this hypothesis, calculations applying approaches fo kinetically-controlled sorption processes are on the way. Some programm modifications are necessary. As soon as available these calculations will be included into this report

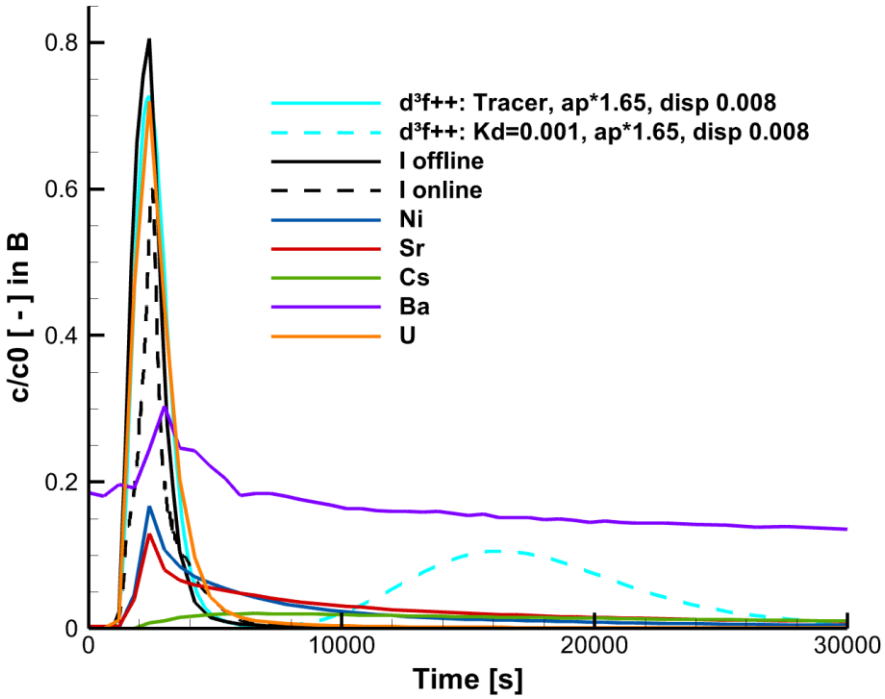


Figure 2-22 Simulated breakthrough curves for a non sorbing tracer and a tracer with Kd=0.001 m³/kg and measured breakthrough curves of all injected tracers

3. Probabilistic approach

The second aspect of the model development was to test and demonstrate the newly developed tool for probabilistic uncertainty and sensitivity analysis by means of flow and transport simulations in a fractured medium. This is related to the work of investigating sophisticated methods of higher-order variance-based sensitivity analysis in order to allow identification of parameter interactions, which GRS has been conducting for several years. Techniques like Sobol-HDMR, which are based on surrogate models, are investigated by application to model systems of different degrees of complexity and nonlinearity and the significance of the results for its application in performance assessment and the safety case.

In parallel, the interface between d³f++ and RepoSTAR (a toolbox for controlling and evaluating probabilistic runs) has been defined and developed. A simple heterogeneous 2D- test model has been defined and used for probabilistic test calculations. The results of a set of 1000 runs have been analysed in view of uncertainty and sensitivity using RepoSTAR.

The d³f++ code is appropriate for modelling the transport of pollutants through a groundwater-conducting geological layer, considering the influences of porosity, permeability, sorption, temperature, fluid density, salinity and viscosity. As these input parameters are more or less uncertain, one might be interested in how the uncertainties affect the total uncertainty of the model output. This is called sensitivity analysis. Global sensitivity analysis considers the total parameter space in a single step, which can be performed by applying a probabilistic approach. This means that a higher number of sets of input parameter values is drawn using, e. g., a random sampling scheme, and the model is calculated with each of them.

GRS has developed the numerical tool RepoSTAR for controlling probabilistic investigations. It supports the user in performing the parameter sampling, assigns the drawn values to the input parameters for each individual run, starts the runs automatically and collects the output. After completion of all runs, the user can apply different mathematical methods to analyse the overall uncertainty and the sensitivities.

GRS has been investigating various methods of probabilistic sensitivity analysis, applied to performance assessment models, during the recent years. There is still little experience, however, with groundwater models. Moreover, RepoSTAR was originally conceptualized as a component of the performance assessment code package RepoTREND, so it is a technical demand to use it with the more complex flow

and transport code d³f++. To test this process and demonstrate the outcome and interpretation of probabilistic analysis, a simple 2D test model was created, first.

Figure 3-1 shows the model. It represents a rectangular region of 10 m x 20 m, consisting of a homogeneous matrix material with two built-in barriers with different permeabilities and sorption coefficients for the tracer. The model grid consists of 67584 elements.

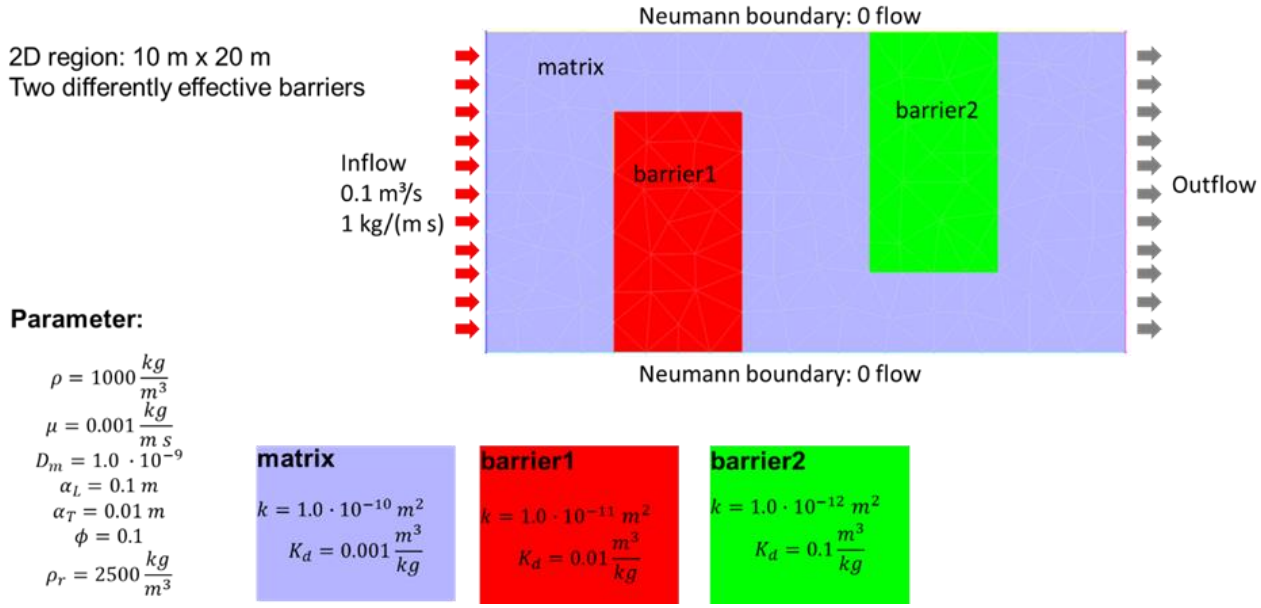


Figure 3-1 Test model and reference data

The results of a deterministic test model run with the reference data are shown in *Figure 3-2*. The fluid flow direction is indicated by arrows, the contaminant (tracer) concentration by colour. It is well visible how the fluid flows around the barriers, which remain at low contaminant concentration.

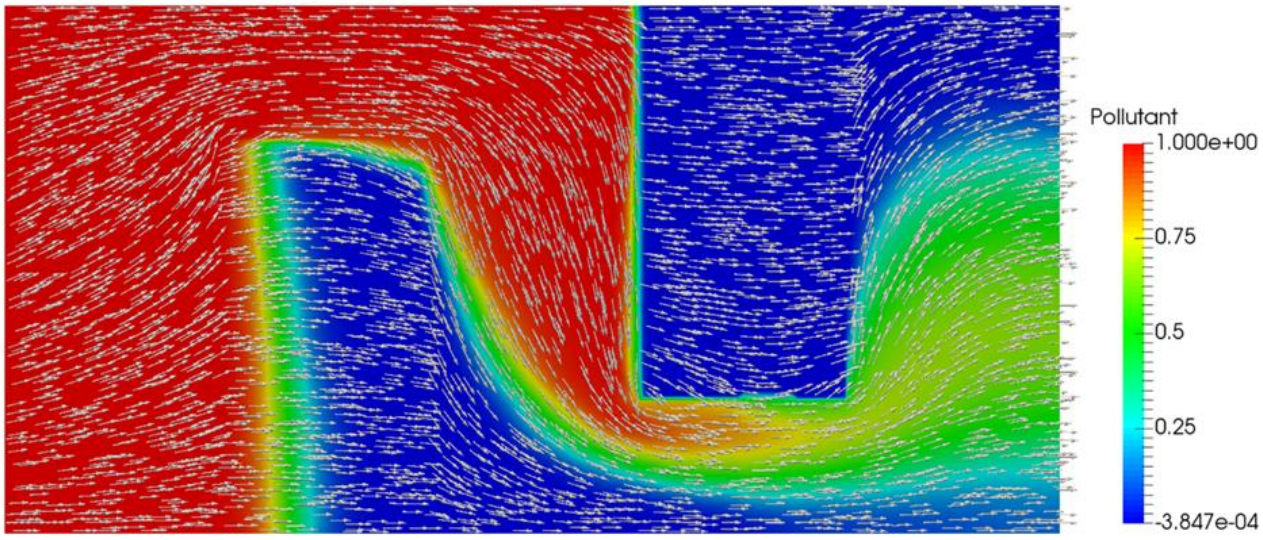


Figure 3-2 Flow direction and tracer concentration after 7 days of model time

Figure 3-3 presents the results of two deterministic parameter variations in comparison with the reference case. This investigation shows the specific effects of some targeted variations of input parameters but does not allow assessment of the overall uncertainty or the global influences of the individual parameters.

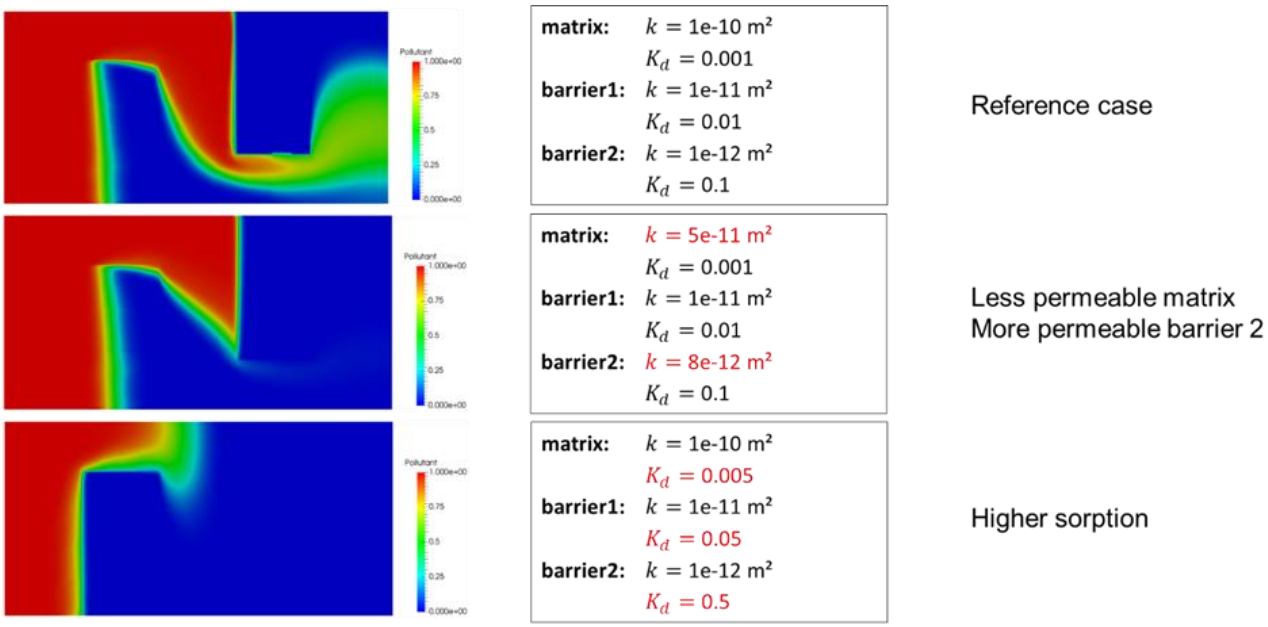


Figure 3-3 Deterministic parameter variations

In order to perform global probabilistic uncertainty and sensitivity analysis, a numerical interface was developed that allows coupling of d³f++ with RepoSTAR. It cannot be done, however, for the total two-dimensional output but it is necessary to define a limited number of scalar quantities of interest. The concentrations at three observation points were chosen, which are all located on the middle line of the rectangular region: in the middle of each of the barriers and at the outflow boundary. The observation points are named after their location on the x-axis: M6, M14 and M20 (see *Figure 3-4*).

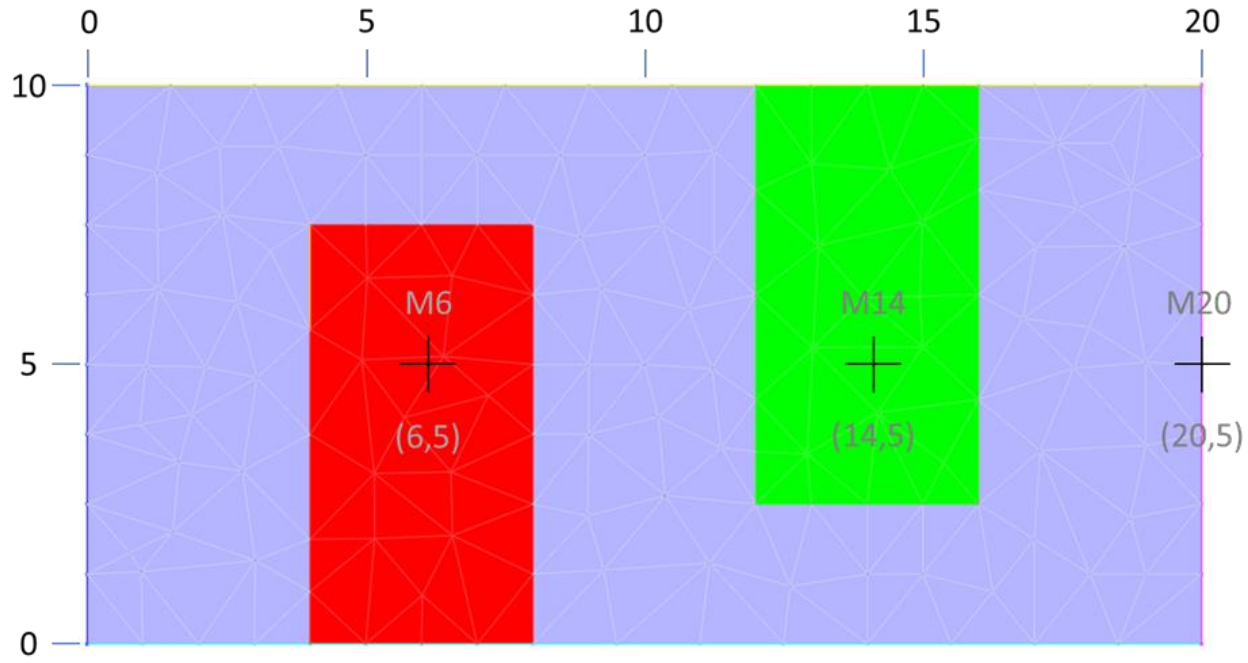


Figure 3-4 Locations of the observation points for probabilistic evaluation

To keep the study simple, the number of input parameters was limited to 6, namely the permeabilities and the Kd-values in the three different parts of the model system. For all of them, log-uniform distributions were assumed see *Figure 3-5*.

Parameter	Unit	Distribution	Lower limit	Upper limit
matrix_k	m^2	log-uniform	1E-10	1E-9
barrier1_k	m^2	log-uniform	1E-12	1E-10
barrier2_k	m^2	log-uniform	1E-13	1E-11
matrix_Kd	m^3/kg	log-uniform	1E-5	1E-3
barrier1_Kd	m^3/kg	log-uniform	1E-3	1E-1
barrier2_Kd	m^3/kg	log-uniform	1E-2	1E0

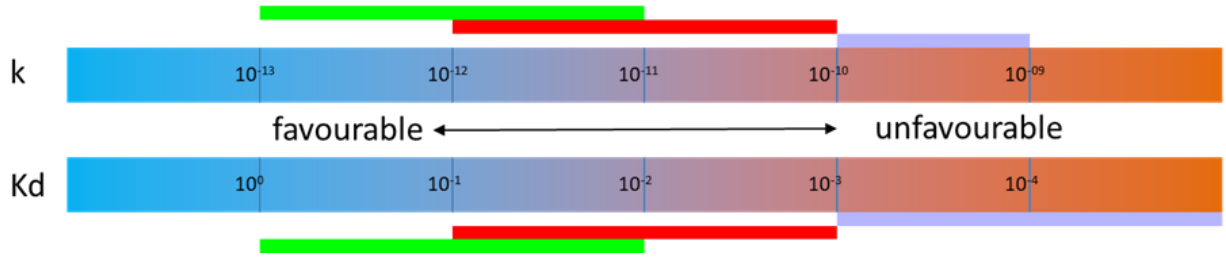


Figure 3-5 Input parameters for the probabilistic analysis. The lower part illustrates the intervals graphically

For uncertainty analysis, a random sample of parameter values with a sample size of 100 was drawn, and accordingly, the model was calculated 100 times. From the calculated time-dependent concentrations at the three observation points the following statistical measures were calculated: mean, maximum, minimum, standard deviation, median, and variance. The time developments are depicted in Figure 3-6.

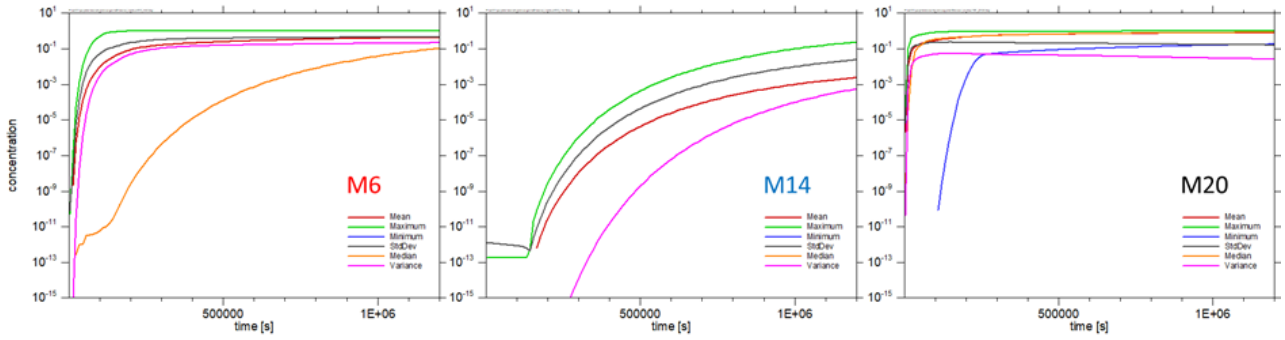


Figure 3-6 Uncertainty analysis of the calculated concentrations at the three observation points

One can draw the following conclusions:

M6 (middle of barrier 1): The median is far below the maximum, especially at earlier times. There are just a few rather high concentration values, becoming more numerous but not higher with time.

M14 (middle of barrier 2): The median is not visible in this presentation, which means that at all times, most output values remain very low. The maximum at late times, however, is just moderately lower than in case of M6. A combination of unfavourable values can lead to higher concentration even in the more effective barrier.

M20 (outflow boundary): From relatively early times, the median is close to the maximum, the minimum is less than 1 order of magnitude below the maximum. That means that the output spreads over a narrow range; the matrix and barrier properties have only limited influence on the concentration at the outflow boundary.

For probabilistic sensitivity analysis, 100 model runs are not sufficient. Therefore, the model was calculated another 1000 times, based on a new random sample, over a model time of 14 days. This took about 50 hours on 10 processor kernels. For the evaluation, four methods were selected:

- CUSUNORO is a graphical method that illustrates the model sensitivity over the quantile of re-ordered parameter values. The x-axis of these plots represents this quantile, for instance, 0.4 means “the 40 % lowest parameter values”. The curves start and end at 0, but the higher the deviation from the x-axis, the more sensitive is the respective parameter. For details, see Plischke 2012, Swiler et al. 2021.
- The Standardized Regression Coefficients (SRC) are calculated from multilinear regression of the model input and output and normalized to the range between -1 and 1. These coefficients work well for quantifying the sensitivities of models that behave close to linear. The “degree of linearity” is quantified by r^2 [Seber et al. 2012, Swiler et al. 2021].
- By replacing each input and output value by its rank in an ordered list, one often gets a relationship closer to linear than the original one. The Standardized Rank Regression Coefficients (SRRC) make use of this and often yield more unique results than SRC. Normally, r^2 is increased by rank transformation. Compared to SRC, SRRC enhances influences on low values and reduces influences on high values [Seber et al. 2012, Swiler et al. 1021].
- EASI is well-approved method to calculate the first-order variance-based sensitivity coefficients (SI1). These coefficients quantify the sensitivity – regardless of the direction of influence – between 0 and 1 on the basis of variance, which does not require linearity [Plischke 2010, Swiler et al. 2021].

RepoSTAR provides, among many others, calculation of these sensitivity measures. The obtained results are presented in the following figures.

Figure 3-7 below shows the CUSUNORO[1] analysis for M6 and M2 at four different points in time (1.1 d, 2.2 d, 8.1 d, 14.0 d). For the concentration in barrier 1 (M6), the properties of barrier 2 as well as the matrix sorption seem to play no role at all, which is no surprise. At all times, the concentration is dominated by barrier 1 permeability and sorption coefficient and, to a lower extent, by matrix permeability. At the output boundary, there is initially a clear dominance of the matrix Kd, which decreases and vanishes soon. Instead, barrier 2 Kd becomes increasingly dominant. The permeabilities of matrix and barrier 1 play minor roles all over the time. Non-symmetric CUSUNORO curves indicate that the parameter influence is not constant over its range. Obviously, at early times, M6 concentration is mainly influenced by the smallest Kd values and the highest permeabilities of barrier 1.

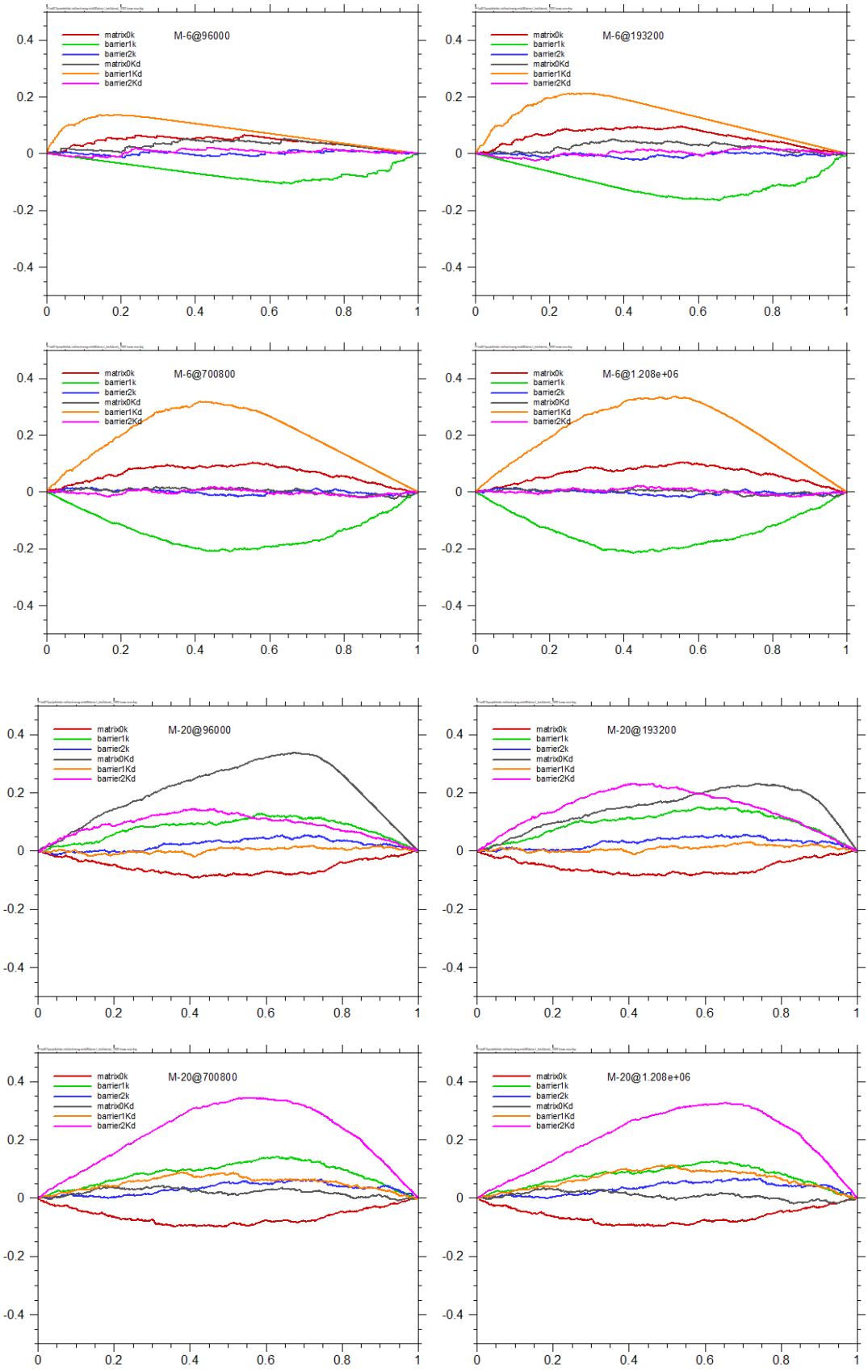


Figure 3-7 CUSUNORO analysis for the observation points M6 (upper) and M20 (lower) at four different times (indicated in seconds)

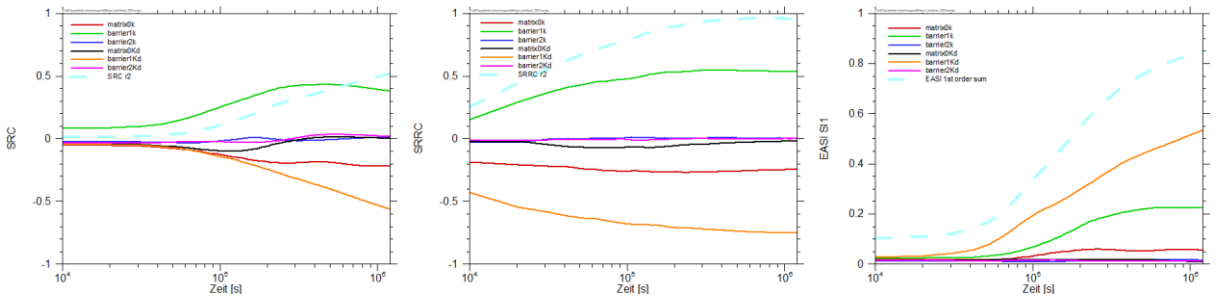


Figure 3-8 SRC, SRRC and EASI analysis for the observation point M6

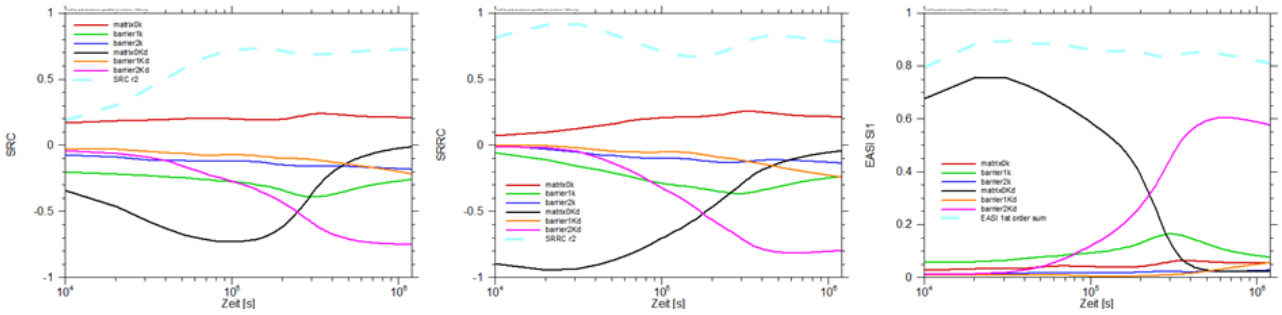


Figure 3-9 SRC, SRRC and EASI analysis for the observation point M20

The time-dependent sensitivity analysis based on SRC, SRRC and EASI SI1 (*Figure 3-8* and *Figure 3-9*) is well in line with the CUSUNORO analysis. One can better see the temporal evolution of the sensitivity but loses the information on the influences of different parameter subranges. It is conspicuous in M6 analysis that for early times, SRC and EASI do not calculate significant sensitivity of matrix permeability and barrier 1 Kd, while SRRC does. Such a difference is a hint that these parameters have an influence only on the lowest model output (concentration in barrier 1) values. As these are often irrelevant, one should be careful when interpreting rank-based sensitivity results.

The test case showed the successful realization of the interface between d^{3}f++ and the statistic tool Repostar to perform probabilistic flow and transport calculations. In a second step probabilistic calculations were performed for a fracture using the block experiment dimensions (see description of the model in Section 2.3). This work will be described in a separate report (Becker et al. 2024).

[1] Note that in the CUSUNORO definition we use here, positive influences (higher parameter value -> higher output) show up by positively curved lines, normally proceeding in the lower half of the diagram, and vice versa. Some implementations define it in the opposite way.

4. Rock as heterogeneous medium: Estimating Uncertainties for distribution coefficients

Conditional distribution coefficients (K_d values) are used to model the partition of a contaminant between solid and aqueous phases. The K_d is defined as the experimentally determined ratio of the sorbed (fixed, immobilized) and unsorbed (free, truly dissolved, mobile) fraction of a component (chemical element) under equilibrium conditions. The K_d values can be calculated based on surface complexation models (SCM). They are quasi-thermodynamic, mechanistic models to describe physi- and chemi-sorption onto mineral surfaces by applying equilibrium reactions between surface binding sites (that can be deprotonated) and aqueous components (most often metal cations), including electrostatic contributions to the binding. The rock material considered in this report are crystalline rocks. The collective term “crystalline” applies to low- to high-grade metamorphic rocks as well as plutonites and is used to distinguish these two rock types from sedimentary rocks (Reinhold 2005, 7). In the context of nuclear waste repositories, the term mostly refers to granitoids or metamorphic rocks of

granitic composition like granites or granodiorites in case of plutonitc rocks and gneisses (one-mica/two-mica) in case of metamorphic rocks.

4.1 Lithological heterogeneity

An important parameter in calculating the distribution coefficient Kd for a given pH and elemental concentrations of the aqueous phase is the mineral surfaces with which potential fluids come into contact. The amount of each mineral interacting with the aqueous phase is obtained by the the volume of the fluid, the rock surface area being in contact with the given volume and a presumed interaction depth of few µm from the surface into the rock. The latter is needed to generate a rock volume, as required by the geochemical modelling software.. Therefore, when calculating Kd values for geological samples, the mean values of mineral abundances determined for e.g., mineralogical composition or pore volume do not necessarily represent the relevant geochemical-mineralogical environment, since the fluid migration pathways transporting the radionuclides may be focused on specific structures and mineral assemblages of the rock. (Andersson et al.,2004). Rock heterogeneity in crystalline rocks can be described by a) structural-geological heterogeneity, in particular the heterogeneous occurrence of faults, fractures and joints, often referred to as discrete fracture network (DFN), and b) lithological heterogeneity, i.e., an irregular spatial variation in modal mineralogy. This lithological heterogeneity depends both on rock type characteristics, e.g., gneisses have a different mineral distribution than granites, and on modal mineralogy along fluid pathways. Therefore, we distinguish here between general spatial variations in modal mineralogy, grain sizes, mineral associations, etc., and differences between the mean composition of modal mineralogy and mineral composition along preferred fluid migration pathways, e.g., fault zones (Andersson et al.,2004).

Of particular interest for K_d uncertainty estimation are potential fluid pathways that are difficult to detect or predict because they have little or no aperture, such as ductile shear zones, fracture zones without offset, weakly developed joints in granites, or veins filled with (readily soluble) secondary minerals. The workflow presented in this report will include these types of fluid migration pathways as they represent an often-overlooked type of "heterogeneity" in the lithology.

So far, modelling of K_d values has mostly assumed that the average composition of the respective geological domain corresponds to the minerals exposed along the fluid migration paths. In this report it will be shown how the estimation of sorption changes when the geostatistical methods are tuned to represent the modal mineral composition along (potential) fluid migration paths or other weak zones (*Figure 4-1*).

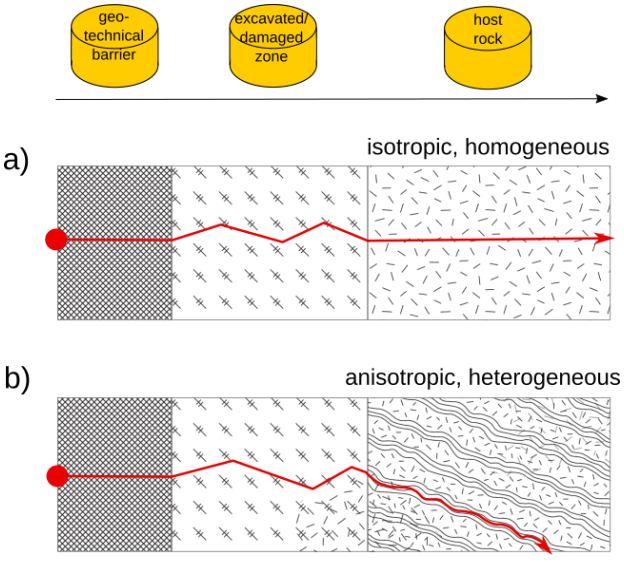


Figure 4-1 Two models for calculating retardation within host rock formations: a) represents the state of the art, the host rock is modelled by mean composition of mineral composition and is assumed to be homogeneous and isotropic. b) shows the model that fluids migrate along specific structures in the host

rock and are therefore exposed to subgroup of modal mineralogy, which can considerably differ to the average composition of the total rock

4.1.1 Sorption models

While the geological aspects dominate the geochemical framework, the intrinsic and globally valid data sets for thermodynamic (and kinetic) characterizations mentioned at the beginning are no less important. Retardation models are based on a selection of data (here mostly sorption and ion exchange) for the respective mineral phases. Accordingly, a careful selection of these data for the overall model is important, since they were usually obtained experimentally and thus can represent the retardation properties only under certain conditions. It is not only important for which mineral phases these data are available at all with respect to which fluid composition, but also for which boundary conditions, such as exact chemical diversity of the minerals or grain size, they are valid. The intrinsically and globally valid data thus consist of the "thermodynamic data", which in addition to the surface-specific data also subsume data for aqueous speciation and for solubility, and the parameters for reaction kinetics. The latter are accessible from databases and literature references.

Furthermore, there are the so-called "field parameters", which describe the totality of physicochemical conditions at a given time and in a given space (to put it generically). Such data are by definition site-specific and consequently highly variable. Well-known examples are pH, redox potential or pore water composition, modal mineral content of the host rock, or temperature. These variable parameters are different for each geologic situation and can vary even at very small scales (sub-cm) due to the natural heterogeneity of the rock. In general, the geological properties beyond the geotechnical barrier determine the geochemical situation for modelling reactive radionuclide transport in the host rock.

4.2 Workflow from rock sample to distribution coefficient

4.2.1 Modelling spatial modal mineralogy

The crystalline rocks discussed in this report are of granitoid composition, i.e., either gneisses, which are metamorphic rocks with potentially very diverse microstructures, or granites, which are plutonic rocks with generally more uniform microstructures, approaching isotropic properties under certain crystallisation conditions. The average modal mineralogy of granitoids consists mainly of quartz, plagioclase, (Na, Ca) feldspar, potassium feldspar, (K, Na) feldspar, biotite and, especially in gneisses, also other phyllosilicates such as muscovite and alteration products such as chlorite, illite, smectite, kaolinite, etc. In addition, other rock-forming minerals may occur, notably amphibole and epidote, in veins and other healed fractures, as well as calcite and other carbonates. There are also iron phases, especially pyrite, but also pyrrothite and iron oxides. Healed veins and fractures are often filled with quartz, calcite, laumontite, chlorite, epidote and prehnite (Nishimoto and Yoshida 2010). Therefore, these are the mineral phases that will be primarily considered.

Predicting the distribution of minerals in a hand specimen, or the distribution of minerals along fluid migration paths, involves two steps: first, simulating or measuring rocks to create a database; second, training and parameterising a model or method based on these data to make predictions or distributions of modal mineralogy. A third step could be to filter the model data and their spatial dependencies with respect to fluid migration paths or areas of the rock that could come into contact with radionuclide-loaded fluids. Several geostatistical methods were tested for their applicability in modelling the modal mineralogy of the exposed surface along (potential) fluid migration paths in combination with the rock volume interacting with the fluid. The methods tested included 3D compositional kriging (Tolosana-Delgado and Mueller 2021), multiple point geostatistics (Mariethoz 2015) and graph-based models. The techniques were tested on simulated rock data as well as measured 2D and 3D rock data.

Graph based models had been finally selected as suitable tool for the requirements for capturing the heterogeneity of rocks, i.e., the discrepancy between mean modal mineralogical composition and the fluid exposed mineral surface composition, for estimating the uncertainty of K_d modelling as further input into follow up reactive transport codes. Graphs have a long history and can be used in many ways to represent

relationships between points (vertices). Being a very versatile tool, graphs can also be used to represent surfaces, respectively volumes, and their interrelationship, for example for models in structural geology (Hillier et al. 2021). Up to the authors knowledge, graphs have not yet been used for microstructure models.

A challenge for using graphs for analysing (and ultimately predicting) the mineral composition along migration paths is the preparation of the input data: The measurement procedure of the rock sample has to result into a data bank, which provides the information about mineral grains, the coordinates of the mineral grain centers, i.e. the spatial position of the grain in a rock sample, the grain size, all adjacent mineral grains and the contact length for each adjacent mineral grain. One possible workflow to gain this type of data is shown in *Figure 4-2*.

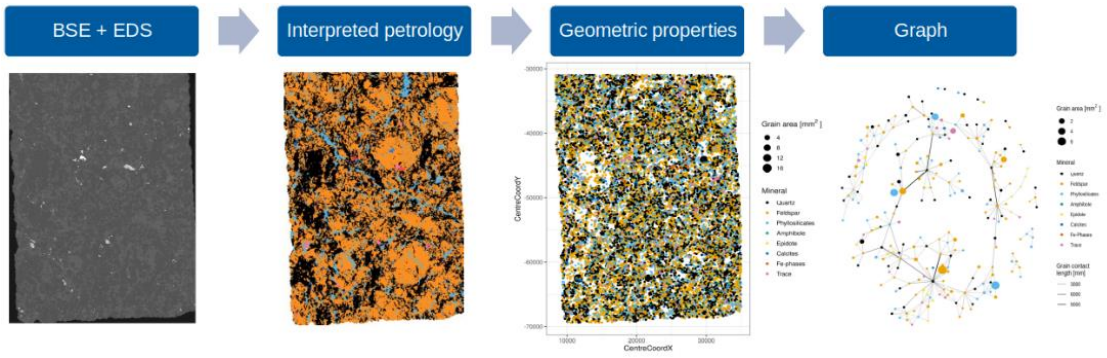


Figure 4-2 Example a workflow from rock sample to graph model: A thin section is made from a rock sample and analysed by back scattered electron (BSE) imaging techniques in combination with energy-dispersive X-ray spectroscopy (EDS). The data of these two measurements are then used to assign to each pixel of the image a mineral phase, which leads to an image with interpreted petrology. Image analysing algorithms can then be used to detect grain boundaries and mineral grains. Furthermore, the software should be able to store the information about grain size, grain neighbours and contact length with each neighbour in a data bank. This information can then be easily transferred into a graph.

Formalised, representing a rock sample by a graph is given by the following: A rock containing N grains is given by an undirected graph, where the vertices represent the grains, and the edges represent the contacts between the grains. Each vertex stores the information about the mineral phase and the grain size. Each edge stores the information about the contact length.

The subgraph is determined by solving the constrained shortest pathway problem with mineral phases as weights for the vertices. Voids (i.e., a phase from the picture which had been determined as ‘void’, this might be technically very challenging) and minerals typically filling weakzones are assigned to high weights. Edges are then used to calculate the exposed mineral surface along the migration pathway. The result is the modal mineralogy of the exposed surface along the fluid migration path. An example of a subgraph is given in *Figure 4-3*.

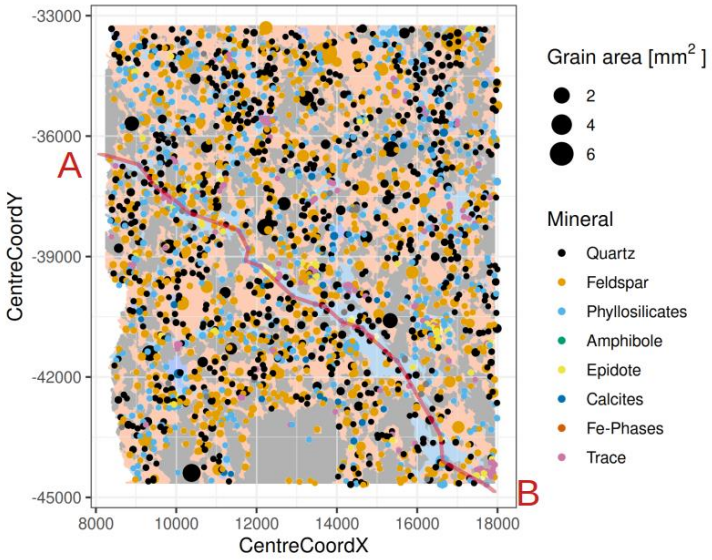


Figure 4-3 Example of a shortest path from A to B, given the condition that a path of the graph should only consist of vertices of the category 'void' or 'phyllosilicate'.

This type of microstructure model requires spatially resolved measurements of mineral phases in sample at least of the size of a thin section, e.g., 2.5x4 cm. The prediction capacity of the method increases with the number of measured samples, but already with one sample, meaningful results can be produced. The focus of this method lies in modelling the "effective mineral surface", i.e., the mineral surfaces exposed along the fluid migration pathways, and as a secondary effect, the offset between the "effective mineral surface" and the mineral composition, the modal mineralogy, of the whole sample. For using this method, it is essential to be aware that the retrieved "effective mineral surface" is controlled by the conditions for the subgraph-producing algorithms. For example, if the condition is strictly, that the vertices of the graph should only be "void" and "phyllosilicate" (for example see *Figure 4-3*), then the effective mineral surface will have a distribution with a very high content of phyllosilicates next to the other phases being present. Thus, conditions for the subgraph production should be chosen very carefully.

4.2.2 Calculating K_d value distributions

The workflow for retrieving K_d values consists of two main tasks: a geological/lithological model and the chemical model, see*Figure 4-4*. The geological/lithological models should provide information about the "effective" mineral (surface) composition along the fluid pathways (see chapter 4.2.1), and through fracture- and void-network modelling the rate of fluid volume being in contact with rock mass. This aspect had not been covered in this study. The rock mass is estimated by the mineral surface in contact with the fluid, and an assumed interaction depth of few µm.

The geological models need as input data analytical or simulated data containing spatially resolved information about the chemical and mineralogical inventory of the rock, as well as structural information like pathways and zones of weakness. These data have to be machine-readable. The models then use this information from measurements for description of spatial dependencies of the mineralogical composition, either for their distribution or their predictions. The result of the geological models should describe for a rock volume pore volume of possible pathways and their mineralogical surfaces, which will be in contact with the fluid.

The chemical models, in turn, require as input a model of the chemistry of the fluids, giving their chemical composition, pH, Eh, temperature, etc., and models of surface complexation for each element of the fluid on each mineral surface that the geological model suggests the fluid will be in contact with.

The geochemical model for estimating the K_d value uses the results from the lithological and chemical models, namely the modal mineralogy of the exposed surface, chemical composition of

the fluid, pore volume, pH, Eh, temperature, SCMs, etc. If for the basic principle of treating (geo)chemical equilibria the concept of Law-of-Mass-Action (LMA) is used, then codes like PhreeqC (Parkhurst et al. 2013), Geochemist's Work Bench (Bethke 2021), Visual MINTEQ (Gustafsson 2011) or several others (Parkhurst et al. 2010) can be used to calculate the K_d value and it's uncertainty for a specific rock volume, respectively geological domain.

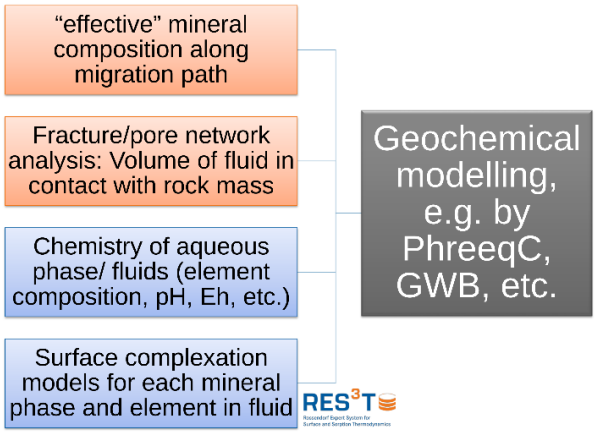


Figure 4-4 Workflow for modelling Kd values from measurements from rock samples. The workflow consists of input information from geological/lithological models (orange, upper two components) and from (geo)chemical models (blue, lower two components). Data from SMC(s) can be retrieved from data repositories, e.g., the RES³T data bank.

4.3 Distribution of K_d values for RN migration by coupling heterogeneities of rocks with thermodynamic modelling

Based on the measurement of several thin slices (workflow see *Figure 4-2*), a distribution of mineral compositions was derived, in total 432 different compositions. Due to the scarcity of data for SCM modelling, areas of predominantly granitoid composition, i.e., dominantly a mixture of quartz, feldspars and phyllosilicates, were selected. Within these areas, the complex mineralogy has been simplified to three mineral phases, quartz, orthoclase, illite and other phases, where 'other phases' should not exceed 5% of the area. Orthoclase was chosen to represent all feldspars and illite to represent all phyllosilicates. It is obvious that this simplification already introduces a bias into the calculations. In particular, the replacement of all phyllosilicates by illite can be viewed very critically, as the larger percentage of phyllosilicates were biotite. Biotite, as an Fe-containing phase, may have significantly different properties in the SCMs than illite or other leucocratic phyllosilicates. Nevertheless, as demonstration for the workflow and resulting distributions of Kd values, Kd values had been calculated by Geochemist's Work Bench with the following input parameter:

- 432 mineral compositions (Quartz, Orthoclase, Illite), given as percentages with constant sum at 100%
- pH range, given as discrete values at 5.5, 6.25, 7, 7.75 and 8,
- UO_2^{++} concentration range, given as discrete values at 5.000e-07 mg/l, 2.875e-06 mg/l, 5.250e-06 mg/l, 7.625e-06 mg/l and 1.000e-05 mg/l.
- Ratio of solid phases to aqueous phase with ratios (g/L) at 1 and 50, reflecting different apertures of fractures.
- Chemical composition of the aqueous phase, as derived from the experimental setup from the block experiment:
    - Eh = 157.3 mV
    - Na = 16.5 mg/l

- K = 2.1 mg/l
- Mg = 8.3 mg/l
- Ca = 37.3 mg/l
- Cl = 3.5 mg/l
- SO42 = 21.9 mg/l
- C = 168.7 mg/l

This resulted into a total of 43700 calculated Kd values for UO_2^{2+}. Results are visualised as density curves in *Figure 4-5*.

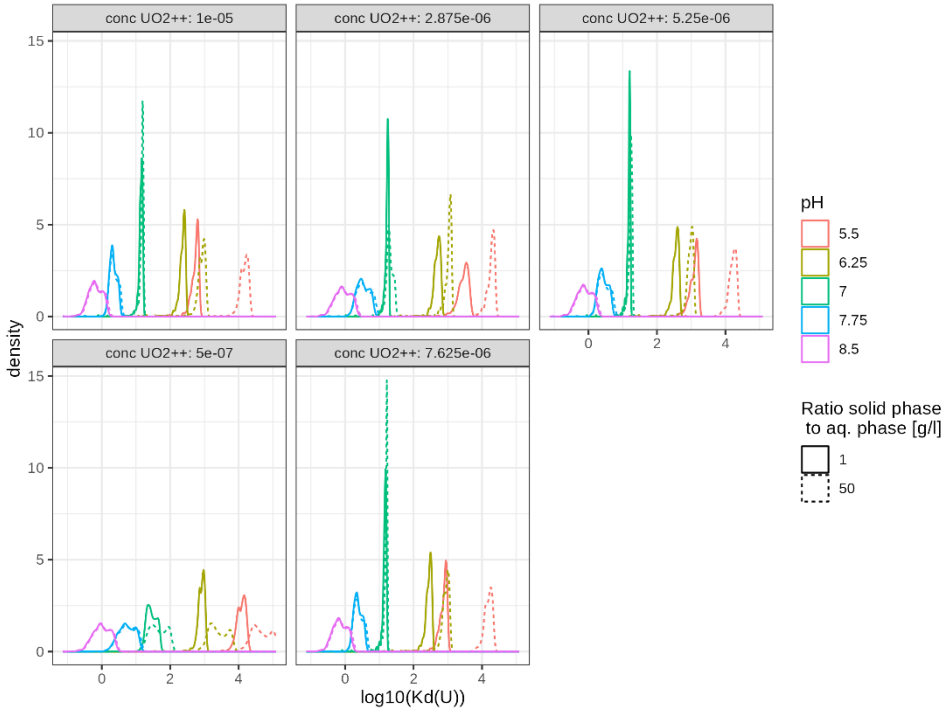


Figure 4-5 Distribution of calculated Kd values for U sorption on granitic rock with respect to different UO_2^{2+} concentrations, a pH range and two ratios of fluid volume being in contact with rock material.

The results show that within the given fluid composition and the variability of the rock composition the main influence variable is the pH value of the aqueous phase, influencing the K_d values by orders of magnitude. The concentration of U in the fluid phase has an influence on the distribution of the K_d values but within the given range of concentration only at a very small part. Concerning the mineral composition, the results show that for this strongly simplified three phase system of the mostly inert reacting phases quartz and feldspar with a certain percentage of more reactive phyllosilicate, the K_d values can vary up to one order of magnitude, depending on the pH value of the fluid.

It must be emphasized that illite as representative for all phyllosilicates, including biotite, is causing certainly a bias in the estimation of the K_d values, because illite is an iron-free phyllosicate, while other micas include Fe in their mineral structure and hence also provide other characteristics of sorption. Additionally, all trace or alteration minerals, which could notably increase the retention potential had not been considered due to missing surface complexation models for UO_2^{++} in combination with an aqueous phase of more complex chemical composition like the SGW. Accordingly, the here calculated distribution of K_d values can be considered as the minimum spread caused by lithological heterogeneity. With increasing complexity of the included mineral phases the maximum range of K_d values will also increase.

5. Summary

A joint programme initiated to carry out a complex uncertainty treatment and sensitivity analysis effort based on a coordinated data collection, uncertainty quantification and testing sensitivity analysis methods, applied to geochemical and transport application has been performed by SURAO, GRS, and HZDR.

One key element of the work was the block experiment performed at the UJV laboratories in a granite block with an artificial fracture. On the one hand the impact of geochemical conditions on the transport of selected radionuclides has been investigated and on the other hand the block experiment served as a reference model to test and apply the newly developed tool for probabilistic calculations on RN transport in fractured media (see below). Regarding the former aspect a cocktail of iodide (non-sorbing tracer), U, Cs, Ni, Sr, and Ba in synthetic groundwater SGW-2 was injected and breakthrough curves were analysed at the outlet over a time frame of about three weeks. Geochemical speciation calculations determining maximum concentrations in the injection cocktail (avoiding precipitation), transport calculations identifying the time regime and the expected element concentrations at the outlet of the experiment, as well as results from surface complexation modelling and sorption experiments on relevant minerals were compiled in order to carefully prepare this experiment and to estimate the expected sorption behaviour of the injected elements.

The results of the block experiment show that the curve for uranium and iodide behave similar, indicating no sorption of uranium. A reason for that may be the high stability of the ternary neutral and negatively charged Ca-uranium-carbonate complexes under the conditions of the SGW-2 water (ph 8.2 and relatively high Ca and inorganic carbon concentrations). Ni, Sr and Ba also arrive with the non-sorbing tracer, but the curves show a very strong tailing, whereas Cs is arriving later than iodide tracer with a very broad breakthrough curve. These results indicate that equilibrium conditions for the sorption process were not established, and kinetic effects occur resulting in broadening of the curves. Further investigations with an experiment under reduced flow conditions and transport simulations including kinetically controlled sorption processes are planned or on the way, respectively.

One aspect of the model development was to test and demonstrate the application of probabilistic uncertainty and sensitivity analysis to flow and transport simulations in a fractured medium. This is related to the work of investigating sophisticated methods of higher-order variance-based sensitivity analysis in order to allow identification of parameter interactions, which GRS has been conducting for several years. To enable the 3D transport code d^{3}f++ to perform probabilistic calculations, an interface between d^{3}f++ and RepoSTAR (a toolbox for controlling and evaluating probabilistic runs) was defined and developed. A simple heterogeneous 2D- test model was defined and used for probabilistic test calculations. The results of a set of 1000 runs were analysed in view of uncertainty and sensitivity using RepoSTAR. Four sensitivity analysis methods were applied; the results are well understandable and more or less in line with what one would expect, but show some interesting details: initially, the contaminant outflow is dominated by the matrix Kd, but its influence decreases with time, while the sorption in the barrier near the outflow boundary becomes more important. Permeabilities in all parts of the system play negligible roles. It is planned to apply uncertainty and sensitivity analysis to the d^{3}f++ model for the granitic block experiment.

A workflow was developed which allows the relevant information for the calculation of K_d values to be determined from measurements of a rock sample with the aid of modern, high-resolution methods. For this purpose, methods from graph theory are used, which allow the computationally highly efficient calculation of the "effective" composition of the minerals, namely the modal mineralogy along pathways or potential pathways, even for very complex crystalline rocks. The combination of the high-resolution, but already industrially used measurement techniques makes it possible to detect, by means of graphs, exactly the type of inconspicuous pathways or migration zones, which, due to their fine ramifications and preferred mineral linings, can represent a large uncertainty factor. For the calculated K_d values show that even in the simplest possible setting for mineral composition, the three phase system quartz, felspar and one mica, the variability in the abundance of the mineral phases caused by rock heterogeneities may well affect the K_d values by an order of magnitude. The distribution of the K_d value is significantly more

complex than the often assumed normal distribution around a mean value. Hence, uncertainties of mineral composition should be taken into account when implementing code for assessing safety of deep geological repositories. It has also been shown that especially for Fe-bearing minerals the data situation for Surface Complexation Models deserves more attention, so that in the future one can reduce the bias of K_d value caused by the choice or replacement of the SCMs.

References

Andersson, P., Byegård, J., Tullborg, E.-L., Doe, T., Hermanson, J., Winberg, A., 2004. In situ tracer tests to determine retention properties of a block scale fracture network in granitic rock at the äspö hard rock laboratory, sweden. Journal of Contaminant Hydrology 70, 271–297. https://doi.org/10.1016/j.jconhyd.2003.09.009

Becker, D.-A., 2016: RepoSTAR - Ein Codepaket zur Steuerung und Auswertung statistischer Rechenläufe mit dem Programmpaket RepoTREND. Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) gGmbH, GRS-411.

Bethke, C.M., 2021. Geochemical and biogeochemical reaction modeling. Cambridge University Press. https://doi.org/10.1017/9781108807005

Fein, E., Schneider, A. (eds.), 1999: d³f - Ein Programmpaket zur Modellierung von Dichteströmungen. Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) gGmbH, GRS-139.

Giffaut, E., Grivé, M., Blanc, P., Vieillard, P., Colàs, E., Gailhanou, H., Gaboreau, S., Marty, N., Madé, B., Duro, L. (2014). Andra thermodynamic database for performance assessment: ThermoChimie. *Appl. Geochem.* 49, 225-236. DOI: 10.1016/j.apgeochem.2014.05.007.

Gustafsson, J.P., 2011. Visual MINTEQ 3.0 user guide. KTH, Department of Land and Water Recources, Stockholm, Sweden 550.

Hillier, M., Wellmann, F., Brodaric, B., Kemp, E. de, Schetselaar, E., 2021. Three-dimensional structural geological modeling using graph neural networks. Mathematical Geosciences 53, 1725–1749. https://doi.org/10.1007/s11004-021-09945-x

Hummel, W., Thoenen, T. (2023). The PSI Chemical Thermodynamic Database 2020. Nagra Technical Report 21-03, Wettingen, Switzerland, 1411pp. https://nagra.ch/wp-content/uploads/2023/05/NTB-21-03.pdf.

Mariethoz, G., 2015. Multiple-point geostatistics stochastic modeling with training images: Stochastic modeling with training images. John Wiley & Sons Inc.

Nishimoto, S., Yoshida, H., 2010. Hydrothermal alteration of deep fractured granite: Effects of dissolution and precipitation. Lithos 115, 153–162. https://doi.org/10.1016/j.lithos.2009.11.015

Noseck, U., Brendler, V., Britz, S., Stockmann, M., Gehrke, A; Richter, C.; Flügge, J.; Lampe, M. 2018: A.: Smart Kd-concept for long term safety assessments. Extension towards more complex applications. Gesellschaft für Anlagen- und Reaktorsicherheit (mbH), GRS-500.

Parkhurst, D.L., Appelo, C., et al., 2013. Description of input and examples for PHREEQC version 3—a computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations. US geological survey techniques and methods 6, 497.

Parkhurst, D.L., Kipp, K.L., Charlton, S.R., 2010. PHAST version 2—a program for simulating groundwater flow, solute transport, and multicomponent geochemical reactions. US Geological Survey Techniques and Methods 6, A35.

Plischke, E.: An effective algorithm for computing global sensitivity indices (EASI). RESS 95 (2010) 354–360.

Plischke, E.: An adaptive correlation ratio method using the cumulative sum of the reordered output. RESS 107 (2012), 149–156.

Reiche, T., 2016: RepoTREND - Das Programmpaket zur integrierten Langzeitsicherheitsanalyse von Endlagersystemen. Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) gGmbH, GRS-411.

Reinhold, K., 2005. Tiefenlage der kristallin-oberfläche in deutschland (research report). Bundesanstalt für Geowissenschaften und Rohstoffe.

Schneider, A., Gehrke, A., Kröhn, K.-P., Zhao, H., 2020: Groundwater Flow and Transport in Complex Real Systems. Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) gGmbH, GRS-566.

Seber, G.A.F., and A.J. Lee. 2012. Linear Regression Analysis 2nd edition. John Wiley & Sons, Hoboken, NJ.

Swiler, L. et al: Sensitivity Analysis Comparisons on Geologic Case Studies: An International Collaboration. SAND2021-11053, Sandia Nat. Labs 2021

Tolosana-Delgado, R., Mueller, U., 2021. Geostatistics for compositional data with r. Springer International Publishing. https://doi.org/10.1007/978-3-030-82568-3

Emulation of computationally demanding reactive transport models to speed up computations

[1]Laloy, E., [1]Jacques, D.

[1]SCK CEN, Belgium

Abstract

Computational tasks that involve running a reactive transport model (RTM) maybe be prohibitively costly with respect to computational time. This is especially true if one needs to do repeated runs with different model parameter values, such as for sensitivity analysis (SA), uncertainty analysis (UA) and model calibration, or in case of long-term reactive transport simulations which are typical to radioactive waste disposal studies. The SCK CEN contribution to DONUT WP4 is all about accelerating RTM simulations by using machine learning (ML). RTMs combine a flow model with a geochemical model. Based on the observation that geochemical calculations classically consume most of the computational time associated with RTM simulations, we investigated to what extent the geochemical solver of a RTM can be replaced by a trained ML method for nonlinear regression. We show that for cement systems of moderate complexity, replacing the geochemical solver by either a deep neural network (DNN) or k-nearest neighbour (kNN) regression both lead to a speedup of about 25-30 against single-threaded RTM calculations while offering a great accuracy. The situation appears to be less favourable for complex cement systems and we thus discuss potential solutions. Furthermore, we also show that for SA, UA and model calibration, the full RTM can be efficiently replaced by a trained ML method. Here we compared DNNs with the other popular ML methods of Gaussian processes (GPs) and polynomial chaos expansion (PCE). We found that GPs offer the best trade-off between efficiency and accuracy. Lastly, we also describe our collaboration to two other, still ongoing projects that were initiated within WP4 DONUT and are respectively led by two other WP4 DONUT partners. Both initiatives concern accelerating numerical simulations by ML. The first project is about speeding up multi-level Monte Carlo (MLMC) uncertainty analysis of flow simulations in fractured rocks by ML. The second project is a multi-partner benchmarking exercise between different ML methods for emulating geochemical solvers.

Significance Statement

The safe disposal of radioactive waste requires computer-based numerical simulations of the relevant physical processes. Moreover, such numerical modelling will form the basis of the so-called digital twins of disposal sites, which upon completion are expected to greatly help site management and maintenance by allowing real-time modelling of the full disposal site. Among the relevant processes to model, reactive transport of radionuclides through different porous media, such as cement and clay, occupies a large part. Unfortunately, computational tasks that involve running a reactive transport model (RTM) maybe be prohibitively costly with respect to computational time. This is especially true if one needs to do repeated runs with different model parameter values, such as for a model uncertainty analysis, or in case of long-term simulations which are typical to radioactive waste disposal studies. In parallel, data-driven prediction methods that rely on machine learning (ML) are extremely fast and have become much easier to train with the relatively recent boom of artificial intelligence and deep learning. The SCK CEN contribution to the DONUT WP4 is about using machine learning (ML) methods, such as deep neural networks (DNNs) and Gaussian processes (GPs), to accelerate reactive transport simulations in various settings relevant to waste disposal. A RTM model combines a flow model with a geochemical model. Since most of the computational time cost of a RTM simulation is typically incurred by the geochemical model component, a large part of our work was to investigate, for the first time, to what extent this geochemical model component can be replaced by a much faster, but potentially less accurate, trained ML method. We showed that for cement systems of moderate complexity, doing so leads to a speedup of about 25-30 against single-threaded RTM simulation and provides quite accurate results. We also highlighted the current limits of the approach for more complex systems and discussed potential way forwards. Furthermore, for model uncertainty analysis under a fixed scenario, which is not computationally feasible with the original RTM, we showed how ML can be used to make calculations tractable and studied the resulting trade-off between affordable computational power and accuracy. Overall, this work paved to the way for ML-based acceleration of reactive transport simulation in complex systems, which as mentioned above is crucial for the future elaboration of digital twins of disposal sites.

Table of content

List of figures

1. Introduction and scope

Within the WP4 DONUT, SCK CEN mainly investigated how computationally-demanding reactive transport model (RTM) simulations can be sped up by machine learning (ML). More specifically, two acceleration strategies were considered. The first and most innovative strategy, referred to as strategy A herein, is to replace the geochemical solver of a reactive transport code by a trained nonlinear regression ML algorithm, such as a deep neural network (DNN). This because geochemical calculations typically incur the largest fraction of the total computational cost of a reactive transport simulation. Our second acceleration strategy is more classical and concerns situations where many runs of the same RTM must be performed, with changing model parameters (for instance, sensitivity analysis, uncertainty assessment and model calibration). In this case, we investigated whether the full RTM can be replaced by a trained ML algorithm, for fixed initial and boundary conditions. In this document, we refer to this acceleration strategy as strategy B.

Besides speeding up reactive transport model (RTM) simulations by ML, SCK CEN also collaborated with TUL, another partner of the WP4 DONUT, on their project of accelerating uncertainty analysis of subsurface flow in fractured rocks by combining a discrete fracture-matrix network model (DFM) with multi-level Monte Carlo (MLMC), using again ML for the acceleration. This project will continue beyond DONUT. The SCK CEN contribution to the work consists of helping to research if the coarse level simulations of the MLMC can efficiently and accurately be replaced by a trained DNN.

Lastly, SCK CEN is actively contributing to the ML benchmark exercise led by PSI, which started within WP4 DONUT. Here the idea is to compare different ML methods in their accuracy and efficiency for emulating a geochemical solver for different cement and Uranium-based multi-component geochemical test cases.

2. ML for accelerating reactive transport simulations – acceleration strategy A

We coupled the C++ written HPx reactive transport code (Jacques et al., 2018) with Python, in order to replace the PHREEQC geochemical solver (Parkhurst and Appelo, 2013) that is embedded within HPx by a trained ML method (surrogate geochemical model) in order to speedup reactive transport simulation in cement systems. We considered two ML algorithms: DNN and kNN. The DNN method is a fully-connected feed forward neural network that needs to be trained beforehand. In contrast, kNN is a non-parametric method that does not require training per se, but still requires a training base. The kNN method basically looks for the k-nearest neighbours of the point to predict within the training base and then performs a weighted average among them. All details about our implementation of these two methods can be found in our associated publication (Laloy and Jacques, 2022). We used two cement systems of increasing complexity to demonstrate the approach. A simpler 4-chemical component system (Ca, Si, O, H) and a more complex 7-chemical component system (Ca, Si, O, H, Al, C, S). For the first test case, both kNN and DNN were found to provide quite accurate results together with a speedup of 5 to 7 (20 to 30) compared to 4-core (single-threaded) HPx. For the second test case, the speedup remained globally similar but the ML-based simulation accuracy decreased. Therefore, some potential solutions to improve accuracy while maintaining speedup were proposed. See Laloy and Jacques (2022) for details. Figure 6 displays some of the obtained results for the first 4-component cement system. It is seen that the ML-based concentration profiles are very similar to those associated with the original model.

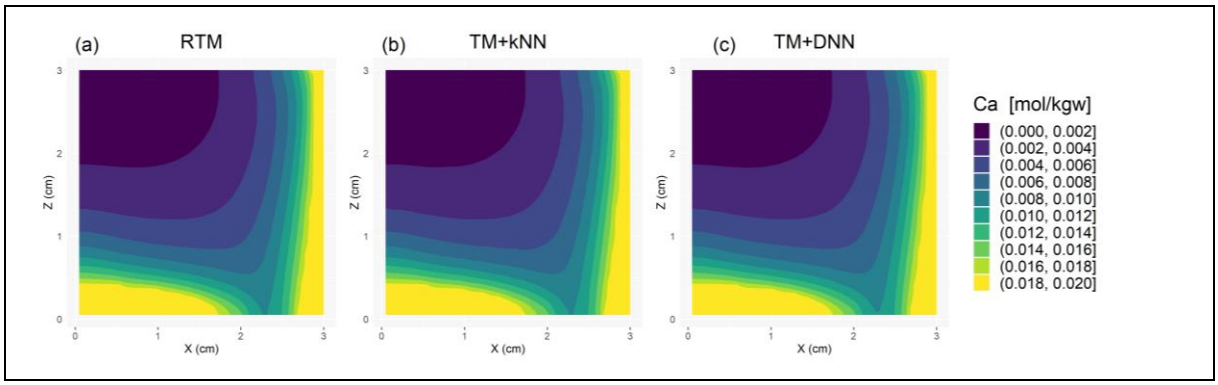


Figure 6 – 2D Ca concentration profile obtained at a given time step for the 4-component cement system simulation of interest. RTM means the original HPx reactive transport model(RTM), TM+kNN denotes the Hydrus (Simunek et al., 2013) transport model (TM) coupled with kNN and TM+DNN signifies the Hydrus transport model (TM) coupled with a deep neural network (DNN). Here the RTM simulation speedups offered by kNN and DNN are about 25 and 7, compared to single-threaded HPx and 4-core parallelized HPx, respectively. Picture from Laloy and Jacques (2022).

3. ML for global emulation of a given RTM – acceleration strategy B

Within the DONUT WP4, SCK CEN also developed non-intrusive statistical emulators (also called metamodels or surrogate models) of a full RTM under given initial and boundary conditions. This is useful to speed up tasks involving repeated runs of the RTM with different parameter values, such as sensitivity analysis (SA), uncertainty analysis (UA) or model calibration (aka inverse modelling). Three ML techniques were compared for the emulation of a well-known reactive transport simulation benchmark involving the 1D reactive transport of Uranium in a soil column (Greskowiak et al., 2015). The considered ML techniques were DNNs, Gaussian processes (GPs) and polynomial chaos expansion (PCE). The three computationally-demanding tasks of SA, UA and model calibration were considered (see Laloy and Jacques, 2019, for details about these ML methods and computational tasks). Given the runtime of the original model, which was about one hour but could increase to several hours for some parameter sets, these tasks can only be efficiently done using an emulator. Indeed, once trained the later takes less than a second per run. A key issue is that the available training set to train the emulator is limited. In our study, training sets of 25 to 500 examples were investigated. Overall, GPs and DNNs were found to perform similarly well for SA and UA with a small advantage for DNNs. In contrast, PCE was found to be less accurate than GPs and DNNs, for the same training set size. Moreover, with respect to model calibration, GPs appeared to provide a much less biased solution than DNNs and PCE. These findings are detailed in Laloy and Jacques (2019). *Figure 7* shows the emulation results of the trained DNN based on either 75 or 500 training samples, for the considered 8-dimensional variant of the original RTM model. It is observed when 500 training samples can be used, DNNs very accurately reproduce the original RTM behaviour. *Figure 8* presents one of the results of the ML-based SA. It appears that when using only 25 runs of the considered 13-dimensional variant of the original RTM to create the training base, GPs and DNNs can provide reliable first-order Sobol (Sobol, 1967) sensitivity indices that resemble relatively well their true counterparts obtained from running the original RTM 500 times.

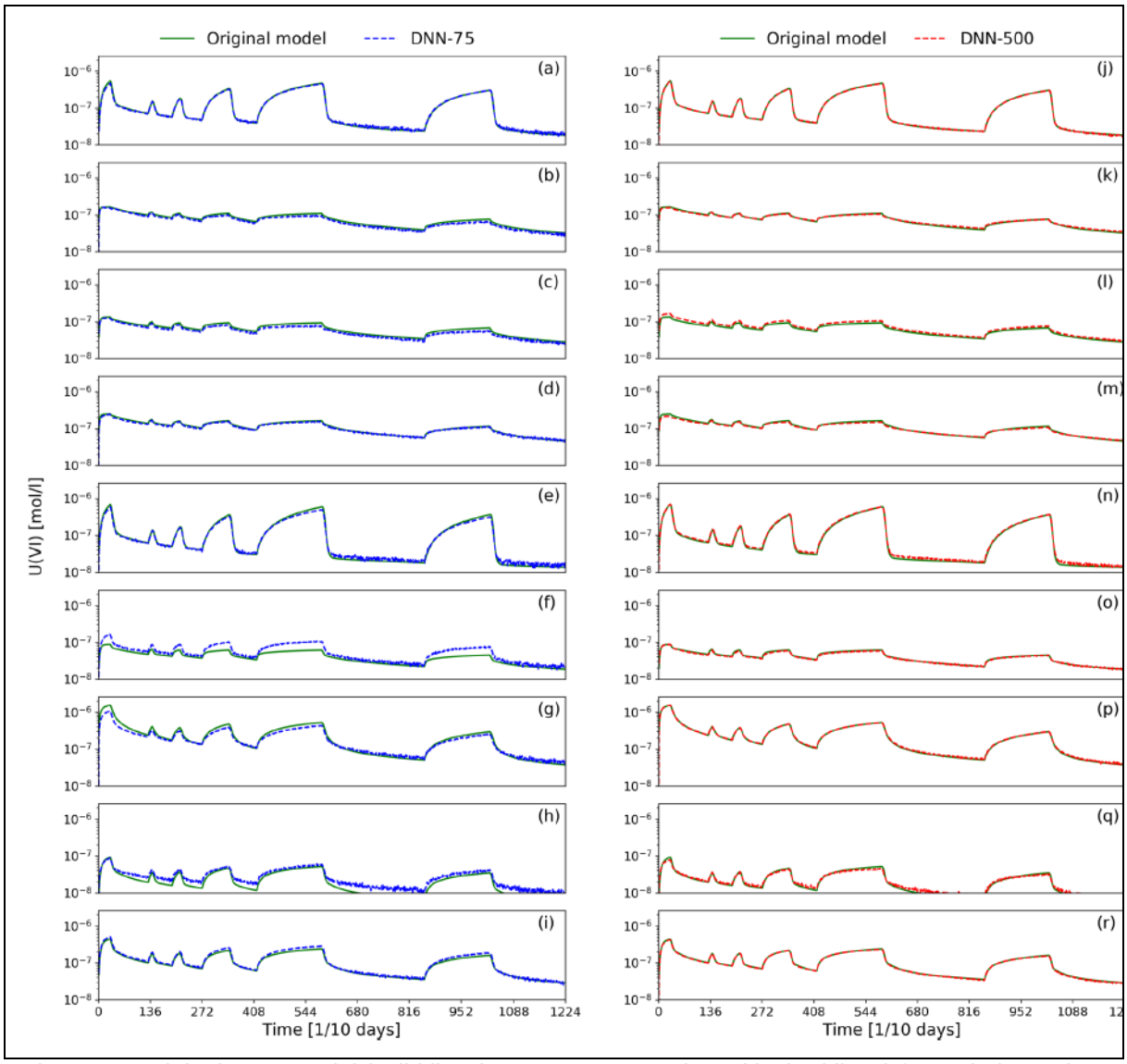


Figure 7 – Original RTM model (solid lines) versus DNN emulator (dashed lines) for an independent evaluation set of simulations of the considered model output: U(VI) concentration in the last node of the modelled soil column. Two training sets sizes are considered to train the DNN, 75 training samples (DNN-75, left plot) and 500 training samples (DNN-500) right plot. Picture from Laloy and Jacques (2019).

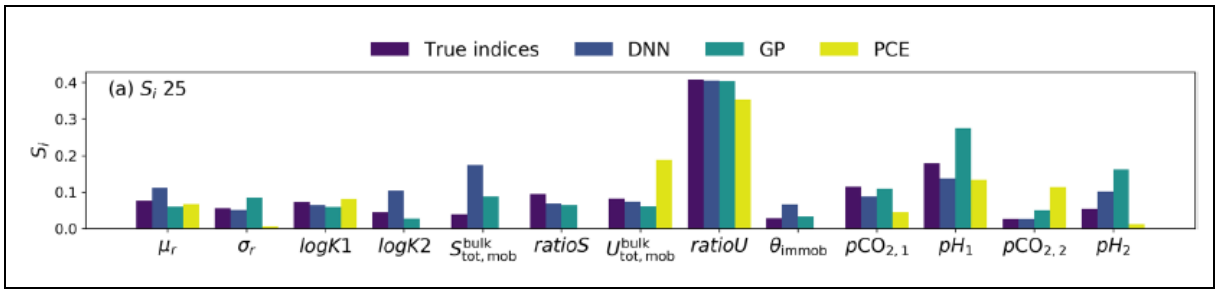


Figure 8 – True and emulated first order Sobol sensitivity indices of each of the 13 parameters of the considered RTM. The emulated indices produced by the DNN, GP and PCE methods are derived using a training set of 25 model simulations only. The true RTM indices are obtained from a set of 500 model simulations. Picture from Laloy and Jacques (2019).

4. Speeding up MLMC uncertainty analysis of subsurface flow in fractured rocks with ML

Under DONUT WP4, SCK CEN also collaborated with TUL within the context of a mobility grant. A PhD student from TUL visited SCK CEN for a 6-month internship, for the period February 1, 2023 – July 31, 2023. In DONUT WP4, TUL is working on subsurface flow simulation in fractured rocks using discrete fracture-matrix models (DFMs). To assess the uncertainty in the simulated flow at an affordable computational cost, TUL is examining if using multi-level Monte Carlo (MLMC) is advantageous over plain Monte Carlo (MC) simulations. To speed up the MLMC, replacing the required coarse level simulations by a trained ML method is under investigation. The main scientific goal of the TUL - SCK CEN collaboration (through the aforementioned 6-month mobility) was to find a sufficiently accurate and computationally efficient ML-based metamodel of the numerical homogenization process taking place in MLMC. This because replacing the numerical homogenization procedure by a trained metamodel, here a deep convolutional neural network (CNN), can significantly accelerate the MLMC analysis. Specifically, we designed a metamodel to approximate the effective conductivity tensor from a DFM model. The metamodels were trained on data from DFM models with non-constant conductivity of fractures (conductivity calculated from aperture). This collaboration will continue outside of DONUT and publication of two journal articles is foreseen.

5. Contribution to ML benchmark exercise – emulation of geochemical solvers

SCK CEN is currently contributing to the ongoing benchmark exercise on ML-based emulation of geochemical solvers, which started within DONUT, involves many partners from DONUT WP4 together with several external partners, and is led by PSI. SCK CEN contributed at different levels. First, SCK CEN co-devised the considered geochemical simulation cases: (i) multi-component geochemical reactions in 5 cement systems of increasing complexity and (ii) Uranium-based geochemical reactions in clay. Second, as three geochemical solvers are considered, PHREEQC (Parkhurst and Appelo, 2013), ORCHESTRA (Meeussen, 2003) and GEMS (Kulik, 2013), SCK – CEN created the reference simulations (and associated training sets) for all test cases, for the PHREEQC solver. Third, SCK – CEN trained and evaluated DNNs for the 5 cement test cases. Lastly, SCK CEN is contributing to the writing of the forthcoming, resulting journal paper. For illustrative purposes, *Figure 9* presents the generalization performance of the trained DNN for a cement test case of large complexity.

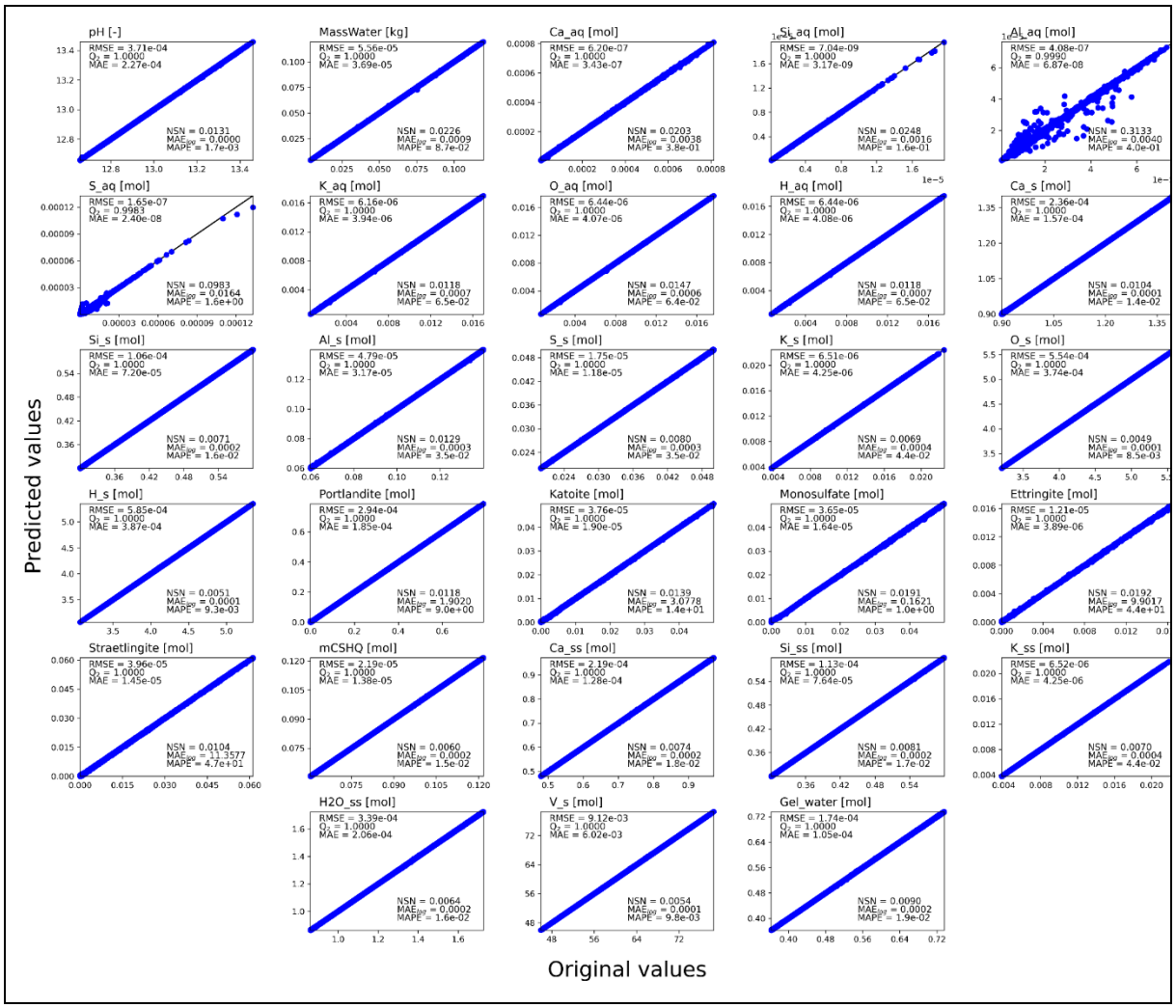


Figure 9 – True (or original) versus DNN-predicted model output values for an independent dataset of a complex cement test case. The “original” values are obtained by running the PHREEQC geochemical solver while the “predicted” values are the computationally way much cheaper DNN predictions.

Code source

An example code for the ML-accelerated HPx reactive transport simulation developed in Laloy and Jacques (2022) will be available on github at elaloy@github.com, pending the latest HPx update. This code and all other developed codes within WP4 DONUT are also available from Eric Laloy (eric.laloy@sckcen.be) or Diederik Jacques (diederik.jacques@sckcen.be) upon email request.

References

Greskowiak, J., Gwo, J., Jacques, D., Yin, J., Mayer, K.U. 2015. A benchmark for multi-rate surface complexation and 1D dual-domain multi-component reactive transport of U(VI). Computational Geosciences. 19, 585–597. https://doi:10.1007/s10596-014-9457-4.

Jacques, D., Simunek, J. Mallants, D., van Genuchten., M.T. 2018. The HPx software for multicomponent reactive transport during variably-saturated flow: Recent developments and applications, Journal of Hydrology and Hydromechanics. 66(2), 211–226. https://doi.org/10.1515/johh2017-0049.

Kulik, D.A., Wagner, T., Dmytrieva, S.V., Kosakowski, G., Hingerl, F.F., Chudnenko, K.V., Berner, U.R. 2013. GEM-Selektor geochemical modeling package: Revised algorithm and GEMS3K numerical kernel for coupled simulation codes. Computational Geosciences. 17(1), 1–24. https://doi.org/10.1007/s10596-012-9310-6.

Laloy, E., Jacques, D., 2019. Emulation of CPU-demanding reactive transport models: a comparison of Gaussian processes, polynomial chaos expansion, and deep neural networks. Computational Geosciences. 23 (5), 1193–1215. https://doi.org/10.1007/s10596-019-09875-y.

Laloy, E., Jacques, D. 2022. Speeding up reactive transport simulations in cement systems by surrogate geochemical modeling: deep neural networks and k-nearest neighbors. Transport in Porous Media.143 (2), 433–462. https://doi.org/10.1007/s11242-022-01779-3.

Meeussen, J.C.L. 2003. ORCHESTRA: An Object-Oriented Framework for Implementing Chemical Equilibrium Models, Environmental Science & Technology. 37 (6), 1175–1182. https://doi.org/10.1021/es025597s.

Parkhurst, D.L., Appelo, C.A.J. 2013. Description of input and examples for PHREEQC version 3: a computer program for speciation, batch-reaction, one664 dimensional transport, and inverse geochemical calculations U.S. Geol. Surv. Techniques and Methods 6-A43. Chapter 43 of Section A: Groundwater in Book 6 Modeling Techniques. https://pubs.er.usgs.gov/publication/tm6A43.

Simunek, J., Sejna, M., Saito, H., Sakai, K., van Genuchten, M. T. 2013. The Hydrus-1D Software Package for Simulating the Movement of Water, Heat, and Multiple Solutes in Variably Saturated Media, Version 4.17. Department of Environmental Sciences, University of California Riverside, Riverside, California, USA.

Sobol, I. M. 1967. The distribution of points in a cube and the accurate evaluation of integrals. Zh. Vych. Mat. Mat. Fiz. 7: 784–802 (in Russian); U.S.S.R Comput. Maths. Math. Phys. 7: 86–112 (in English).

New modelling approach for TSPA coupled modelling

[1]Bothi, Z., [1]Olasz, T., [2]Baksay, A.

[1]Golder Associates (Magyarország) Zrt., Hungary

[2]TS-Enercon Kft., Hungary

Abstract

A new modelling approach was developed to make coupled physical-chemical modelling an integral part of long-term, post-closure, Total System Performance Assessment (TSPA). It aims at improving understanding how coupled processes impact the near-field evolution of a repository. It provides tools for carrying out sensitivity analysis and uncertainty analysis of the TSPA model, and to optimize the disposal system. It also supports QA/QC measures by providing traceable data processing via databases and transparency during the development and modification of the model over different phases of disposal programs.

The model was developed in a way which makes it possible to be applied for TSPA purposes of geological disposal systems regardless of their waste type, host rock formation, or disposal concept. It focuses on describing coupled processes in the Engineered Barrier System (EBS) of the near-field, but also includes migration of radionuclides through the far-field, and the calculation of radiological risks in the biosphere.

It is the first model in which all physical-chemical (Thermal, Hydraulic, geoMechanical, geoChemical, Gas-related, and microbiological - THMCGB) processes considered to be coupled to one another, and mass, heat and radionuclide TRansport (+Tr) are all included in a transparent fashion. It applies a time shifting algorithm in which current processes are used to calculate changes in state variables, such as temperature, pressure and chemical concentration. The values of these state variables from the previous timestep are then applied to derive material properties and rates of processes to update the calculations.

The near-field of the model is spatially discretized by using the localized container and cloning features of the GoldSim environment in which it is implemented. These provide the basis for applying a Finite Element Modelling (FEM) approach, where each discretized volume has the same internal structure, general logic and mathematical description. This approach enables us to consider heterogeneity by defining different material properties and boundary conditions, makes the model easy to modify, upgrade and extend without having to make changes to the underlying structure as new information and knowledge becomes available, and the spatial and temporal discretization of the model can be optimized to meet current requirements.

Significance Statement

Evaluating the long-term, post-closure radiological risks of geological disposal systems for radioactive wastes is essential to demonstrate the safety of such disposal concepts and actual repositories. The Total System Performance Assessment (TSPA) methodology is applied to carry out such evaluations.

Golder developed a new TSPA modelling methodology within the GoldSim environment which focuses on the near-field evolution of repositories by considering the coupled physical and chemical processes, such as thermal, hydraulic, mechanical, chemical, gas-related and microbiological processes together with the mass, heat, and radionuclide transport processes. The new approach aims at providing a transparent and traceable modelling environment within which even such complex systems may be relatively easily implemented, and their adequacy may be checked. These are top priorities for authorities and agencies which deal with radioactive waste disposal.

The model was developed by utilizing some new methodology which had not yet been applied previously in GoldSim. Some of the processes required new mathematical models to be developed in GoldSim, extending and enhancing its built-in capabilities. The model currently is capable of calculating the following processes, and their coupling in the near-field:

- waste package failure and waste form degradation;
- radionuclide release;
- radioactive decay and ingrowth;
- decay heat generation;
- heat conduction, convection and radiation;
- two-phase, variably saturated flow;
- saturation and effective porosity related permeability;
- advective-dispersive-diffusive transport of radionuclides and aqueous species in pore water;
- partitioning among fluids and solid media;
- hybrid equilibrium-kinetic chemical reactions;
- bentonite swelling, chemical alteration and ion-exchange;
- degradation of cementitious materials and consequent changes in porosity and permeability;
- corrosion and microbiological activity induced gas generation.

The newly developed modelling approaches include

- the direct calculation of decay heat from radioactive decay;
- calculating complex, unsaturated hydraulics;
- deriving chemical reactions based on complex stoichiometry and kinetics;
- the hydration process of the bentonite and its impact on bound and effective porosity;
- microbiological degradation of organic materials and calculation of the microbial mass.

Table of content

List of Figures

Glossary and Abbreviations

PA	Performance Assessment evaluates the radiological risks of waste repositories
SA	Safety Assessment compares radiological risks to legislative limits
TSPA	Total System Performance Assessment evaluates the whole disposal system within a single framework
QA	Quality Assurance determines how the system meets requirements
QC	Quality Control is a set of procedures aiming to ensure that quality criteria are met
EBS	Engineered Barrier System is the part of the multi-barrier system emplaced by human technological processes
HLW	High Level radioactive Waste is such waste for which its heat generation must be considered
ILW–LL	Long-Lived Intermediate Level radioactive Waste is such waste in which longlived radionuclides have major contribution
SNF	Spent Nuclear Fuel is waste originated from nuclear power plant reactor
RIP	Repository Integration Program is one of the predecessors of GoldSim
STRIP	Strategy Integration Program is one of the predecessors of GoldSim
CTM	Contaminant Transport Module is a GoldSim module with which transport calculations may be carried out
RTM	Radionuclide Transport Module is a special version of CTM enabling the calculation of radioactive decay and ingrowth
PDF	Probability Distribution Function is the mathematical description of probability density for continuous distributions
CDF	Cumulative Distribution Function is the integral of PDF
Monte Carlo Simulation	stochastic simulation using repeated random sampling
Top-down Modelling	modelling approach where complexity and detail level of the model increases with consecutive phase
Lumped Parameter Modelling	modelling technique in which larger spatial volumes are characterized by average parameter values

Process Level Modelling	modelling technique aiming to model all relevant processes at a very detailed level
Abstraction Level Modelling	modelling technique focusing on general behavior of the system
KBS–3V	disposal concept for HLW / SNF developed by SKB and POSIVA with granitic host rock
aleatoric uncertainty	natural variability of the system which cannot be lowered
epistemic uncertainty	uncertainty related to lack of knowledge and understanding which may be lowered by obtaining more information

1. Goals of model development

The goal is to develop a modelling approach which integrates coupled processes into the Total System Performance Assessment (TSPA). The approach should support the top-down modelling approach and the possibility to carry out sensitivity and uncertainty analyses, provide adequate transparency, traceability, and QA/QC (Quality Assurance / Quality Control) measures, as required by radioactive waste disposal programs.

2. Basic assumptions

To achieve the goals defined above first and foremost some definitions and assumptions must be clarified:

- In a TSPA the purpose of the assessment is to investigate the whole disposal system evolution after closure and to derive its radiological consequences. Therefore, the TSPA must include all relevant parts of the disposal system (the wastes, the Engineered Barrier System (EBS), the host rock and other relevant geological formations and features (together the geosphere), and the biosphere) and must be able to predict processes and events and their interactions which may take place during the timeframe of the investigation.
- The timeframe considered for such systems is generally in the thousands and millions of years, so it extends to much longer time periods for which normal industrial standards and approaches could be applied. That requires special predicting modelling approaches to be developed and applied. The prediction of the models and consequently the validity of the models is difficult to verify since no laboratory experiment or site investigation can cover such long time periods. So, model results must be compared to data covering limited time intervals and must be extrapolated from then on.
- The TSPA generally must use abstraction level simplification for the description of the disposal system and its possible evolution. The level of information and understanding relating to different parts of the system could be quite different. It is generally beneficial if the TSPA (in general) deals at a similar level with all elements of the system regardless of whether available knowledge enables to model certain processes in a much higher detail. The only exemption to this general approach is that previous sensitivity analyses might have identified parts of the system as having dominant consequences and these parts may be modelled at a higher detail in consequent TSPAs.
- TSPAs generally apply the principle of Lumped Parameter Modelling. Since the Abstraction Level Modelling used in these applications enforces higher level simplification of the modelled system than Process Level Modelling (generally applied to derive the closest possible description of the real system), it also requires a coarser (both spatial and temporal) discretization of the system. In Lumped Parameter Modelling a certain part of the system is characterized by a set of average (or integral) parameters and state variables. This is an upscaling technique which is based on the fact that in a TSPA the major factors we are interested in are the consequences and the general behavior of the system, rather than understanding in detail the underlying physical-chemical-mechanical processes.
- The available information on different parts of the system and their potential evolution is extremely varied. At least in theory, the wastes and the EBS (in their primary state) can be characterized quite accurately, whereas information on the geological environment will always be inevitably burdened with uncertainty and current biosphere characteristics (including human behavior) may change considerably in a very short time. The uncertainty originating from the possible processes and events significantly increases with time. Therefore, one of the most important requirements for a TSPA is to properly characterize and implement both aleatoric and epistemic uncertainties.

- Performance Assessment (PA) or Safety Assessment (SA) (in this document they are used interchangeably) could be applied to both operational and post-closure (or long-term) radiological consequence assessments. Here we are focusing solely on post-closure PAs. Nonetheless, the starting point in time for a post-closure PA is when the repository is closed and left to evolve in a passive manner, without any further human intervention. This follows a (relatively) long period in which human activities (like the construction of tunnels and galleries, water extraction and ventilation) have altered primary conditions significantly. Consequently, the conditions are unstable (transient), and then they are further perturbated by processes taking place as the repository is closed. This means that even the initial conditions of a post-closure PA are burdened with uncertainty and the evolution of the system may differ according to what initial conditions are estimated or assumed.

- Although TSPA may be applied to any kind of repository, in the current task we focus on High Level Wastes (HLW) including Spent Nuclear Fuels (SNF) and such Intermediate Level Long-Lived Wastes (ILW–LL) as ion-exchange resins as these types of wastes require the most sophisticated modelling approaches to be applied. The layout, the materials used in the engineered barrier system and the host rocks in which such disposal systems are constructed (or planned to be constructed) are varied and therefore provide different critical (coupled) processes to be investigated. There are some processes which must be evaluated for all disposal systems, such as their re-saturation, gas generation, advective-dispersive transport of contaminants, but there are also some specific processes like swelling of clays, degradation of cementitious materials, corrosion of different metals, plastic deformation of rock salt, transport in unsaturated media which may prove to be the critical processes in certain repositories, and which may be missing altogether from others. In the current sub-task, we try to provide a general framework in which all such processes and their potential interactions could be evaluated and implemented.

- In the current phase, our goal is to develop a generic model in which all relevant processes are considered, so that it may be modified relatively easily to apply for any given disposal system. On the other hand, to present this generic model, an existing disposal concept should be selected as a basis on which the applicability of the modelling method can be demonstrated. Calculated results should be compared to published literature data as a benchmarking exercise.

- It is further assumed that such PA models are to be used in different phases of a radioactive waste disposal program during which the available information and the requirements may be significantly different. Consequently, the modelling approach should be flexible enough to be applied with restricted modifications to all those phases and to produce results which can be evaluated at the detail levels required by the current phase. It is also assumed that the model should provide possibilities to focus on different parts of the disposal system as the current assessment requires it.

3. The conceptual model

For the purposes of demonstrating the efficiency of the modelling approach that we developed, a generic model was developed which follows in its main structure and assumptions the KBS-3V disposal concept of SKB and POSIVA without aiming to reproduce it in detail.

The model describes the physical-chemical evolution after final closure of a repository for spent nuclear fuel and high-level radioactive waste. It assumes that the closure has already taken place instantaneously in the whole repository, so initial conditions are what they have developed during the operation phase. It means that within the EBS unsaturated conditions prevail, and when the repository is closed ingress of water starts from the adjacent rock matrix re-saturating the system. The other major process which takes place from the moment of closure is the thermal pulse which is the consequence of decay heat generated by the radioactive decay of the waste. It is also possible that some chemical processes also start acting from the start, due to the assumed chemical disequilibrium existing in the system. It is assumed that originally oxidative conditions progressively become reductive as the oxygen

in the system is consumed. One of the chemical processes which may have impact on other processes and conditions is the hydration of the bentonite buffer around the waste containing canister. Since bentonite is made up of some swelling clay minerals, such as montmorillonite, as it hydrates it also swells, and a swelling pressure develops. The characteristics of the bentonite (such as effective porosity, permeability) are dependent on its hydration state. Bentonite may also be chemically altered which also affects its properties. This transient phase may take hundreds to thousands of years, but eventually these processes reach some equilibrium and a tertiary, stationary condition is achieved within which the velocities of processes are very slow until some external event disturbs the system. Depending on local conditions the waste package fails some time, and it starts releasing radionuclides from the waste matrix. Released isotopes start migrating from the source through the near-field and then through the far-field. It is assumed that the repository is located in a sparsely fractured, low permeability crystalline host rock, such as granite. It is further presumed that the location will be selected where downwards gradients prevail in primary conditions and generally flow rates are extremely low, so diffusive processes dominate, at least within the near-field. Radionuclides released from the repository migrate through the far-field part of the granite through the fracture network. It is assumed that downgradient receptors (humans) will use wells to tap contaminated groundwater which has gone through some dilution.

3.1 Thermal model

Radioactive decay generates decay heat. In the model SNF and HLW waste types are assumed, and each is defined as having such high activity as their heat generation must be considered. The produced heat must be dissipated for which theoretically there are three different processes: radiation, convection and conduction. Radiation is the emittance of electromagnetic waves due to the temperature of the material. Heat conduction is the transfer of heat energy by the collision of microscopic particles, similar to diffusion. Convection happens when heat is transferred by the flow of some fluid, generally either water or air.

The ambient temperature of the host rock of the repository is mainly dependent on its depth and the local thermal gradient. When a repository is constructed at depth, human activity related to ventilation and groundwater drainage may change it a bit, but not significantly. On the contrary, when high activity radioactive waste is disposed, it elevates local temperatures significantly. It will create a heat pulse due to the decay heat it produces. The temperature will increase first at the waste (within the canister or container) and from there it will dissipate outside elevating neighboring temperatures, a temperature front will be formed moving outward from the waste. Since, the activity of the waste will decrease with time, this heat pulse will have a limited longevity which could be in the hundreds to few thousand years depending on the initial radionuclide inventory of the waste, the thermal properties of the EBS and the host rock. After the heat pulse is gone, temperatures will slowly get back to normal ambient temperatures.

Since temperature plays a major role in determining both material properties and the rates of some processes (*Figure* 1), the local elevated temperature during the heat pulse phase will impact a lot of other processes taking place in and around the repository.

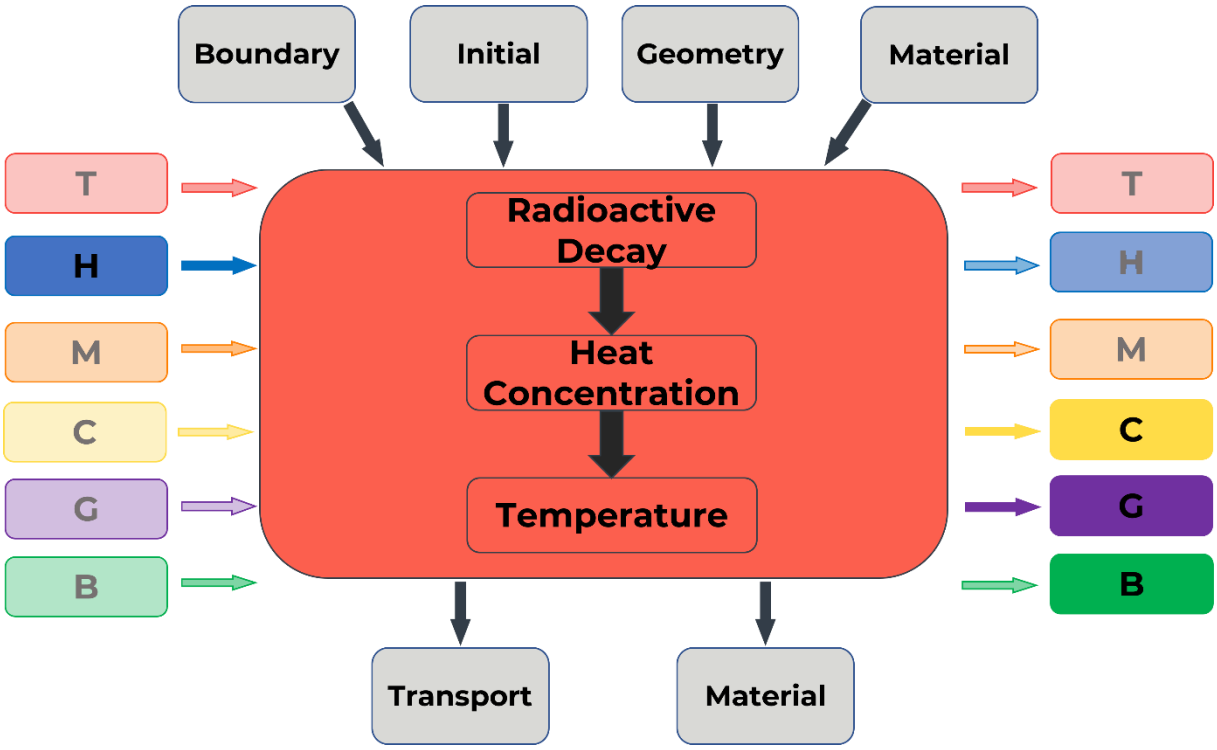


Figure 1 – The thermal model and its coupling to other processes

3.2 Hydraulic model

After the repository is constructed, during its operation, continuous water drainage will take place which will create unsaturated hydraulic conditions within the repository and in the near-field part of the host rock. After the wastes are disposed of, cavities will be backfilled, and no longer active water drainage will be carried out. This will result in the start of ingress of groundwater into these parts of the repository, gradually re-saturating it.

Effective permeability of solid materials is a function of their water saturation, with the higher their water saturation the closer their effective permeability (hydraulic conductivity) will be to their intrinsic permeability. So, during the re-saturation of the repository, generally permeabilities of materials will increase, providing means for higher flow rates to occur.

Hydraulic conditions are intrinsically coupled to both gas-related processes and thermal conditions (*Figure*), so these processes must be considered as interacting.

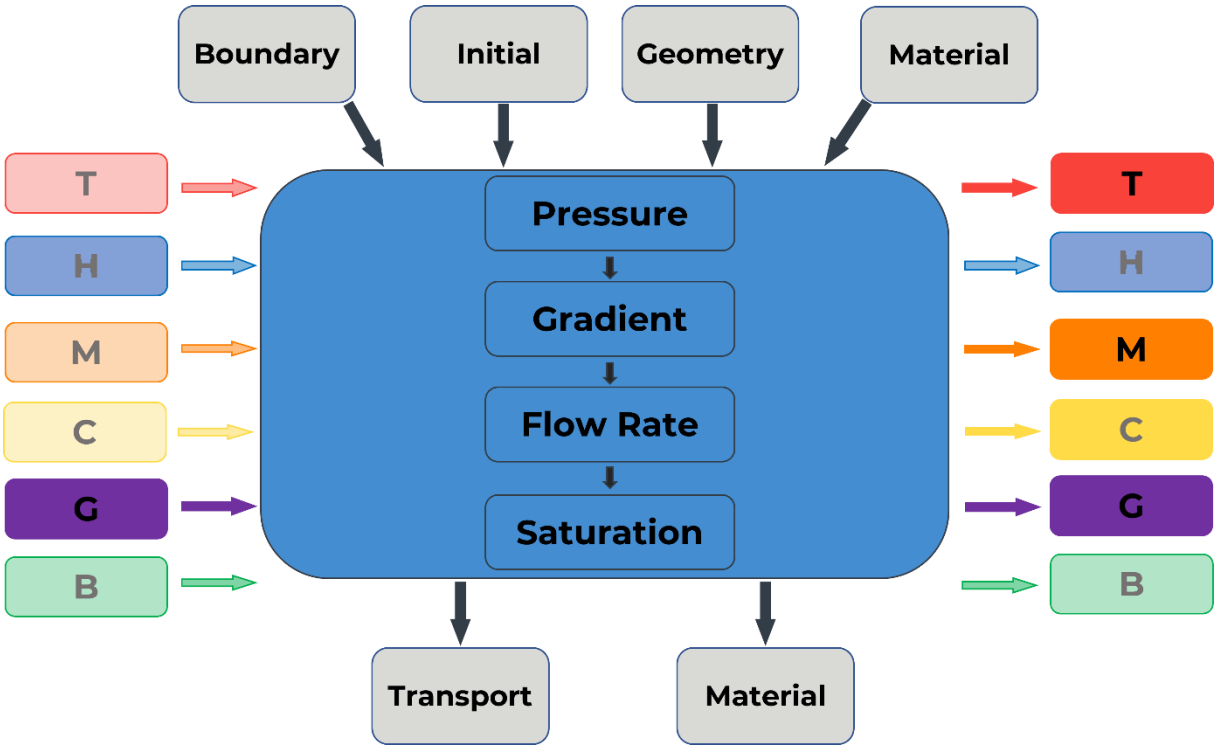


Figure 2 – The hydraulic model and its coupling to other processes

3.3 Geomechanical model

Although there are several geomechanical processes which might be considered, the most significant one is the swelling pressure build-up of the smectite minerals within the buffer (*Figure* 3).

Smectites (montmorillonites and the likes) are swelling clay minerals, which means that as they are hydrated, they take up water into their crystal lattice and their volume increases. As they swell, they exert pressure on the wall of the canister and on the wall of the disposal hole.

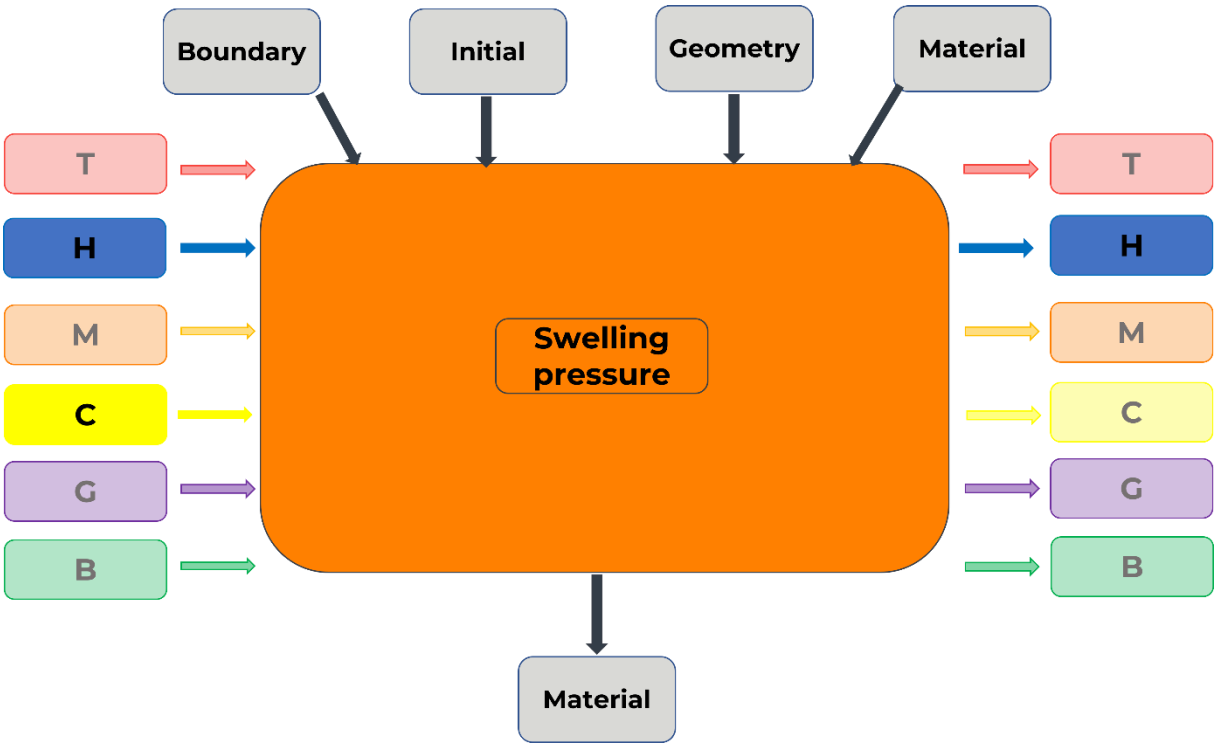


Figure 3 – The geomechanical model and its coupling to other processes

3.4 Geochemical model

It can be assumed that the host rock and its pore water (the groundwater) are in chemical equilibrium, but this equilibrium cannot be presumed to prevail when EBS materials – mainly the cementitious materials, the bentonite and the metals – are emplaced within the repository. This results in chemical alteration, or degradation of solid materials. This process is further enhanced by the varying geochemical conditions which are partly due to the re-saturation of the system and the consumption of oxygen.

The major geochemical processes which must be considered are (*Figure* 4):

- the alteration of cementitious materials,
- the corrosion of metals,
- the alteration of smectite materials,
- biodegradation of organic materials, and
- the consumption of oxygen.

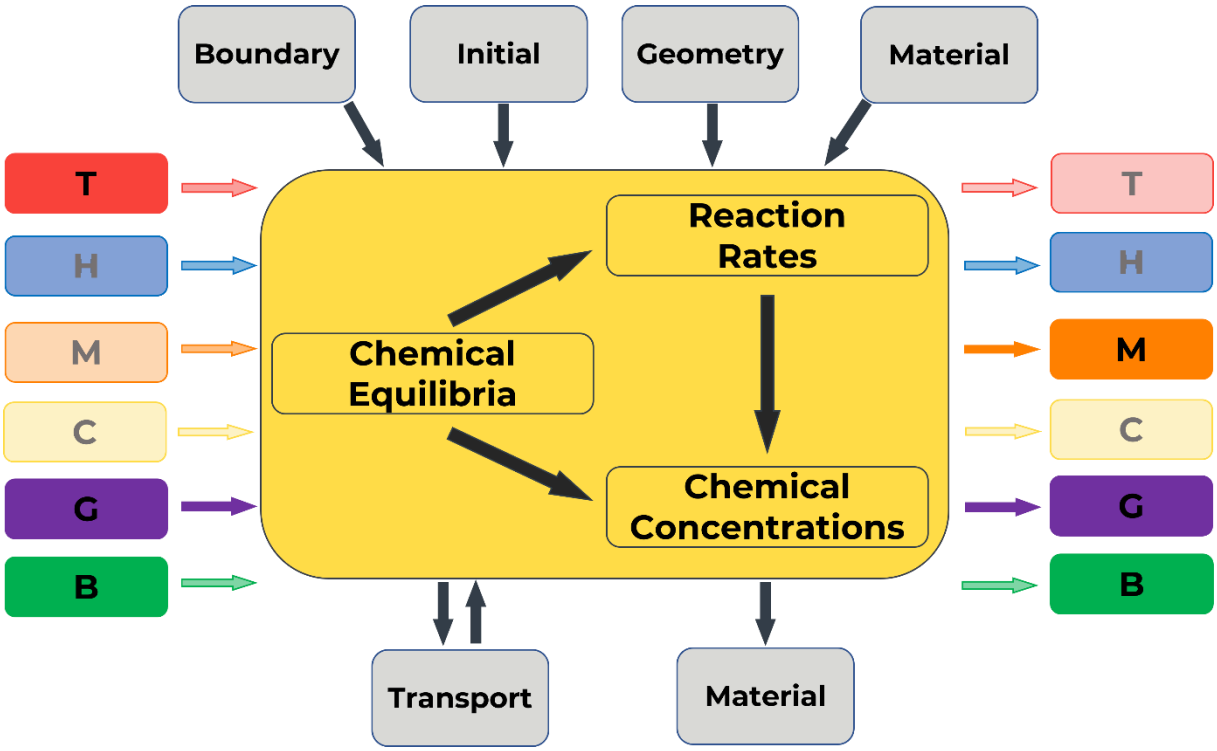


Figure 4 – The geochemical model and its coupling to other processes

Cementitious materials go through hydration and setting processes as they are emplaced. As they are contacted with the pore waters of different geochemical composition, they start their slow alteration process which goes through several main phases. During these processes alkalic ions are released from the cementitious materials elevating the pH of the porewater.

Metals, mainly different types of steel materials and copper for the canister, emplaced in the repository go through corrosion processes as they are contacted with air or porewater. The corrosion processes should be divided up into two major groups whether they take place in aerobic conditions (when oxygen is available) or in anaerobic conditions. Generally, the first type of corrosion has much higher rate but it is very limited in time as oxygen is depleted from the system in a relatively short period. Anaerobic corrosion is a significant process as it may lead to hydrogen gas generation (which is further detailed in Section 3.5).

As the buffer is one of the main EBS components with several safety functions, the alteration of smectite minerals it is composed of may impact its performance and consequently the long-term evolution and safety of the whole disposal system. There are two main smectite alteration processes which must be investigated: the illitization and its change of cation composition. Illitization occurs when there is enough potassium in the porewater and the temperature is high enough. Since illite (and the mixed minerals which are intermediates in this process) possess no or very limited swelling capacity compared to the original montmorillonite, some of the beneficial properties of the bentonite buffer may get lost due to the illitization process. The other process takes place as montmorillonite takes up cations from the porewater. A typical version of this process is when Ca^{++} replaces Na^{+} in the crystal lattice of the montmorillonite. The structure of the mineral and consequently its properties – like its swelling capacity, permeability and sorption capacity – change as this alteration takes place resulting in less favorable retardation capabilities.

Geochemical alterations are driven by current local geochemical conditions, from which the pH and the redox conditions are the most significant. pH may be changed due to chemical alterations (like the release of cations from the cementitious materials). As the repository is re-saturated and oxygen is consumed by chemical processes, it is getting depleted from the system. As this happens aerobic conditions turn into anaerobic conditions.

The chemical processes take place in two timescales: fast chemical processes prevail within the aqueous phase, where changes happen within short time periods; and precipitation and dissolution of mineral phases, where a transfer between the aqueous and the solid phase takes place happen in a much longer time period. The first ones could be assumed to be always in local equilibrium, so an equilibrium-based approach may be used for the description of these processes. The latter ones require considering kinetic rates as these are limited by some factors.

3.5 Gas-related model

Initially the system is unsaturated, which means that some air is trapped within the pores as a separate phase. During re-saturation air is replaced by water.

Gases may be generated due to chemical and microbiological (biogeochemical) processes. As gas is produced, it dissolves in the porewater until it becomes saturated. If gas cannot be released from a location it starts to accumulate as a separate free gas phase. This results in gas pressure build-up (*Figure* 5) until it exceeds a critical pressure, the gas entry pressure, when it enters the solid phase and expels some of the porewater.

One of the major potential gas generation processes is the corrosion of metals (mainly steel and copper) under anaerobic conditions which may produce huge amounts of hydrogen gas over extended time period. Although metal corrosion also takes place under aerobic conditions also, it does not lead to gas generation and accumulation. Another potential gas generation process is the microbial degradation of organic materials. This may lead to the production of either methane or carbon-dioxide depending on redox conditions.

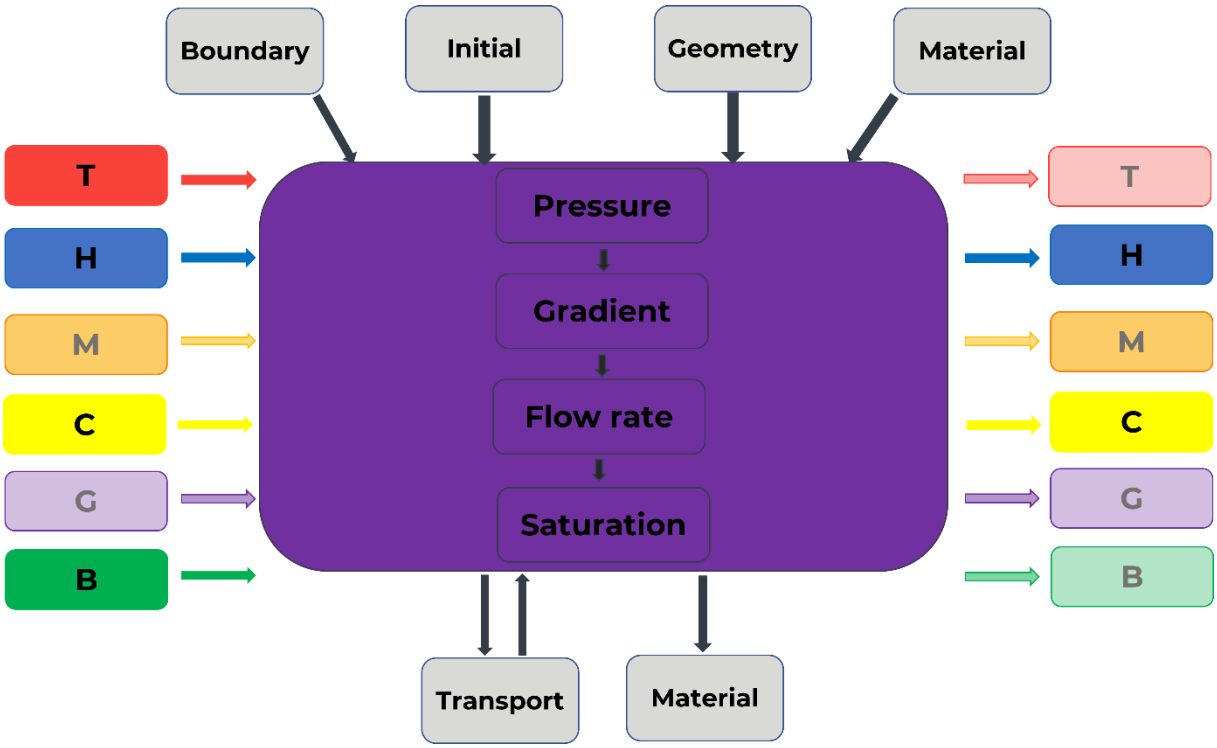


Figure 5 – The gas model and its coupling to other processes

3.6 Microbiological model

Microbiological processes are very limited in a HLW/SNF repository environment. Organic materials can be consumed by microorganisms (*Figure* 6). These processes may lead to the chemical alteration or consumption of some chemical components, and they may produce some gases. The bacterial mass may either increase or decrease due to the availability of these chemical components.

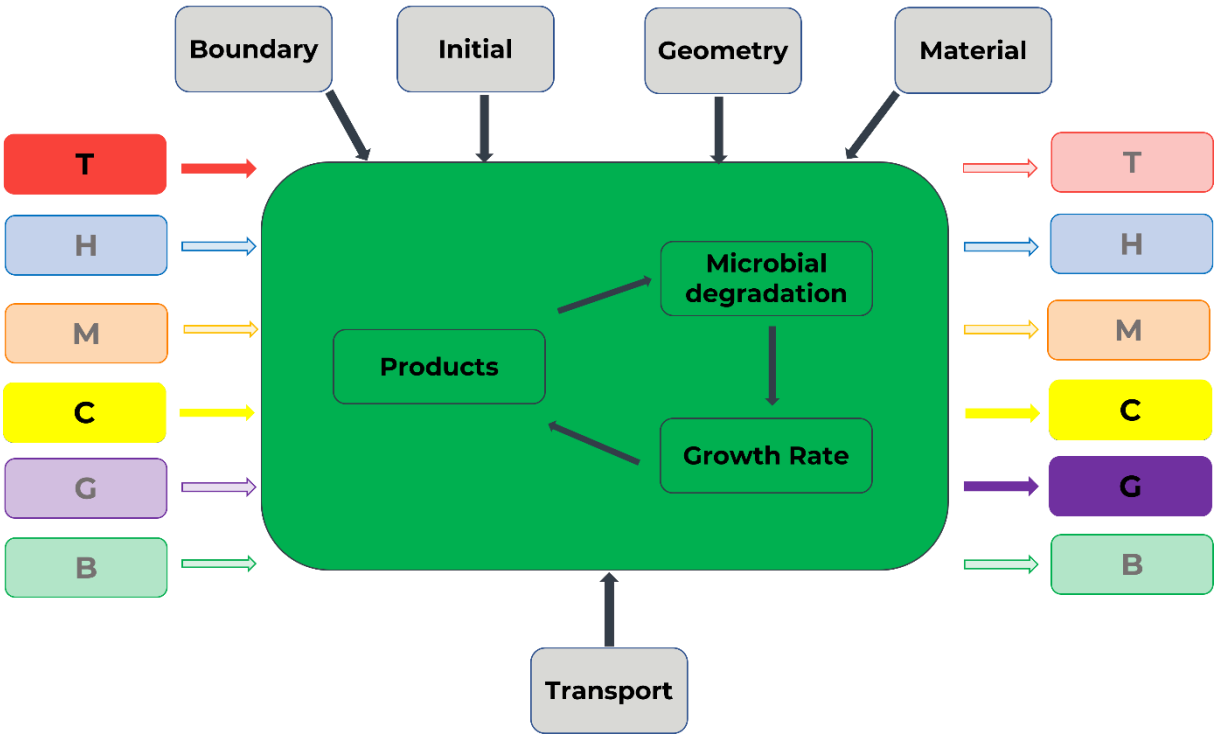


Figure 6 – The microbiological model and its coupling to other processes

3.7 Transport model

In the long-term PA we are interested in the radiological effects of radionuclides. To assess this, their migration must be investigated with a transport model.

The transport processes of radionuclides start with their release from the waste form when the waste container, like the canister fails (a hole is formed through which porewater can enter it and radionuclides can be released). Most of the transport processes are mediated by porewater, but in some cases they can be transported in the gas phase.

Radionuclides dissolve in the porewater up to their elemental solubility limit. The remaining part is precipitated and remains as precipitate until the dissolved concentration decreases below the solubility limit. Dissolved radionuclides may be transported by either diffusion or advection. Diffusion is driven by chemical concentration differences. Advection is the transport of dissolved components by the flowing water. Since low permeability materials and low hydraulic gradients are dominant in a disposal system, diffusive transport processes are presumed to prevail resulting in a halo of contaminated porewater being formed slowly around the waste packages. Radionuclides may sorbed onto solid materials further retarding their migration away from the waste package and from the repository.

Radioactive decay may also be considered a transport phenomenon. Decay occurs naturally for instable isotopes and this process may be characterized by the decay rate or half-life of the isotopes. In case there is a daughter product which has much shorter half-life than its parent and both remain in place (they are both immobile elements) then the ingrowth results in secular equilibrium in which the activity of the daughter product equals its parent isotope's activity.

4. New approaches and methods developed for the model

Development of the model required some new methods and approaches to be developed and applied which had not yet been previously utilized within the GoldSim modelling environment. Listed below the major new developments:

- Decay heat directly influences the temperature of the waste form.

- Process variables directly influence all relevant material properties.
- Hybrid kinetic-equilibrium approach for chemistry in GoldSim.
- FEM-like approach in GoldSim using Cloning and spatial volumes.
- Swelling pressure calculation using a multi-factor empirical equation system.
- Water volume considers chemical production and loss, biological production, vapor condensation and evaporation, and hydraulic fluxes within the same timestep.
- Gas mixture properties are directly derived from the current and updating composition together with pressure and temperature.
- All equations and connections can be changed by the user at any stage of the modelling process without having to alter the source code.

5. Preliminary model results and their interpretation

Results of the thermal evolution show expected trends. In *Figure* 7 we can see the temperature of the core over 200 years. In the beginning, and upwards trend is visible. The temperatures quickly rise to a high, 190 °C value. After this peak temperature, cooling down is observable. The cool-down phase lasts longer then the warming up period.

We can see the effect of the Spent Nuclear Fuel core and its remaining radioactive isotopes here. They cause the warming up of the spatial volume. The generated heat also transports into neighboring volumes, cooling down the core. This general process is successfully captured here.

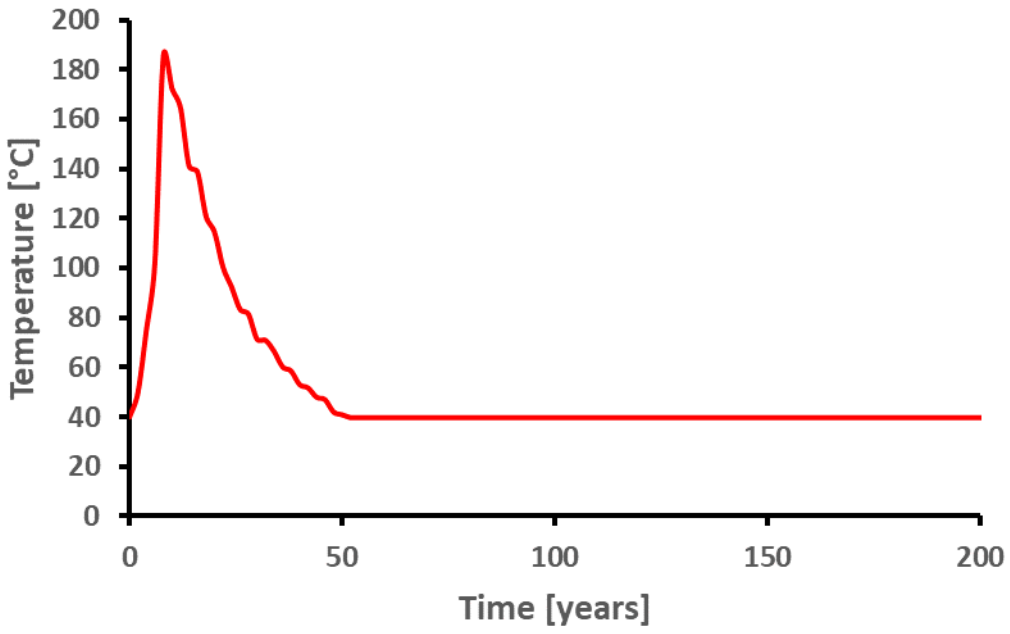


Figure 7 – Thermal evolution of the Spent Nuclear Fuel core

Looking at a spatial volume containing the bentonite buffer, we can observe a similar trend (*Figure* 8), but with different values. The temperature peak occurs later, and it only reaches around half of the core's peak value. The cool-down phase here is an even longer process.

What we see here is the effect of heat transport by the different means included in our model (conduction, convection and radiation). The heat generated in the core transfers into nearby volumes, but since it is a transient process, a delay can be observed. The transport of the heat energy between different spatial volumes also results in a lower peak temperature.

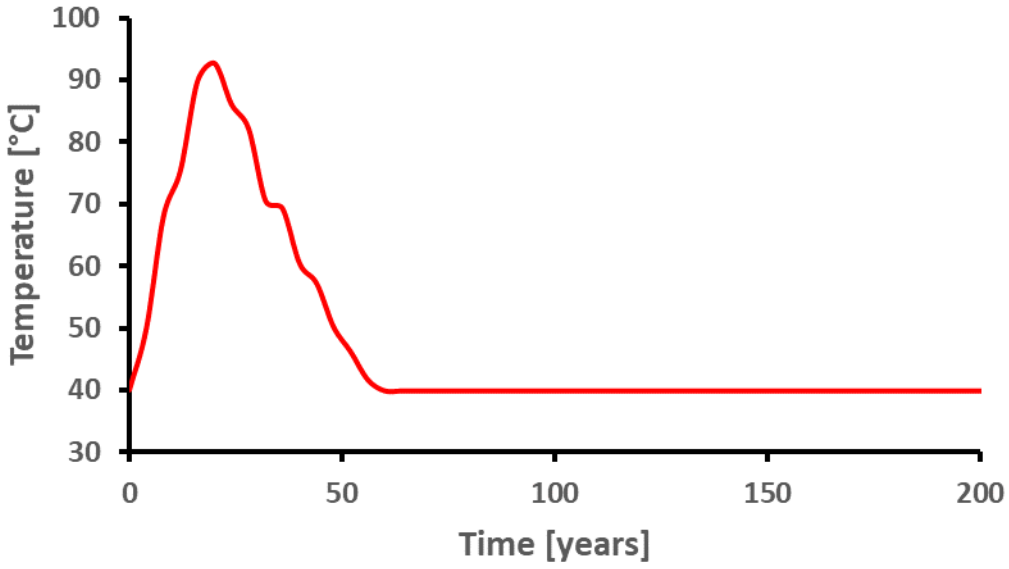


Figure 8 – Thermal evolution of a spatial volume containing bentonite material

Looking at the patterns of inflow, outflow and the saturation (*Figure* 9 and *Figure 10*0) of the spatial volumes, we can observe different patterns. It is likely that the spatial discretization of the model is too sparse to accurately represent the expected processes.

It can be seen that in one of the buffer volumes, after a large amount of inflow, the bentonite saturates fully. Inflow after this period do not saturate the volume, but flow through it. A higher peak inflow occurs around 90 years.

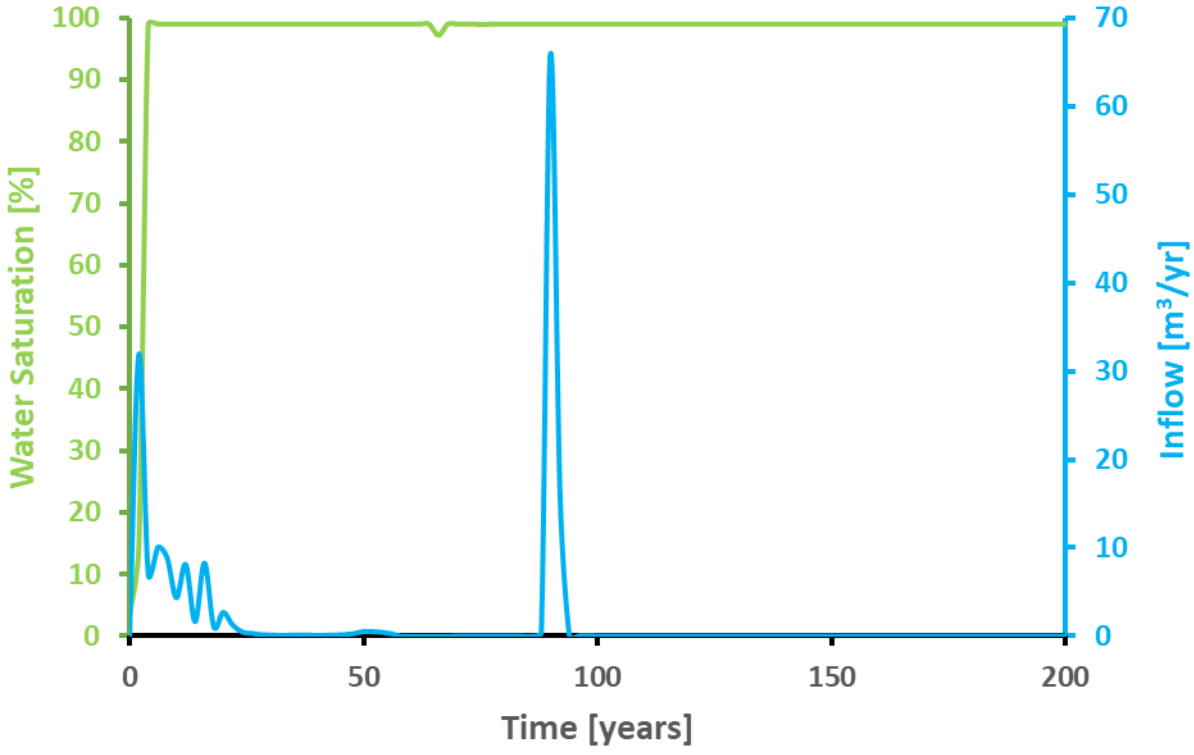


Figure 9 – Water saturation and inflow rate of a spatial volume containing bentonite material

In a different bentonite spatial volume, the material becomes less saturated. It can be seen that a large outflow rate is present in the first few years, which causes the saturation to lower slowly to around 30%. After this, the outflow rate is low, and the saturation fluctuates only slightly.

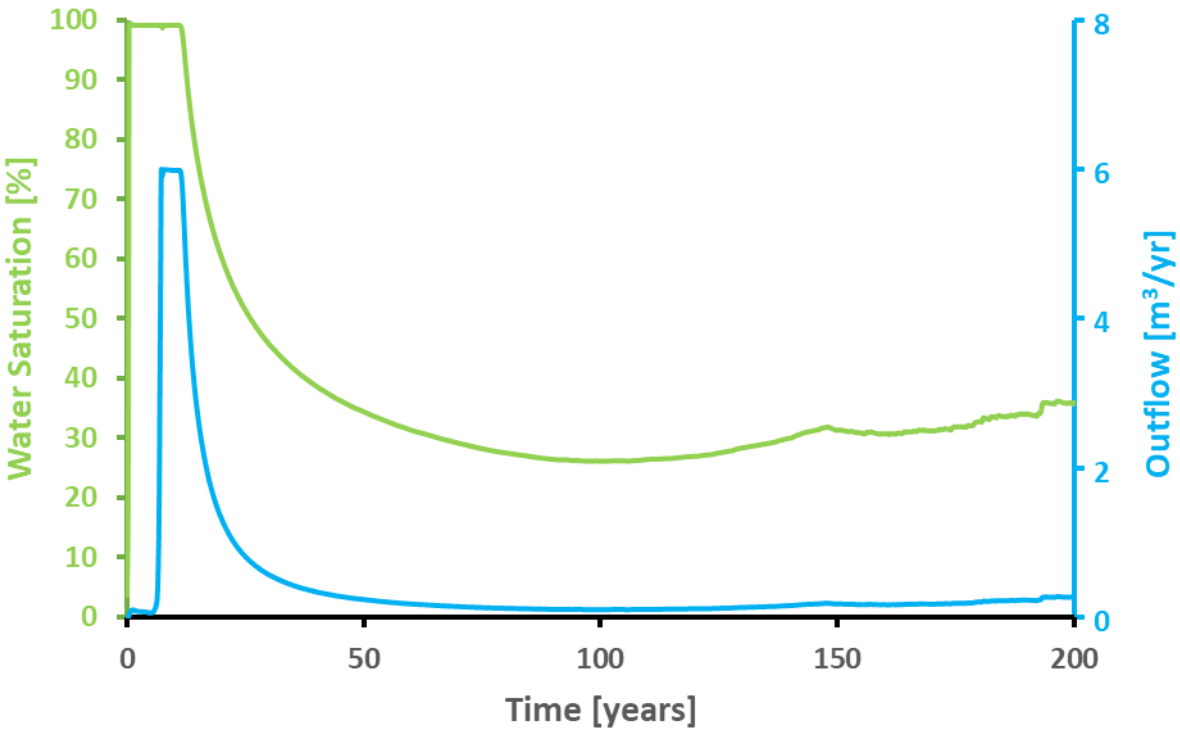


Figure 100 – Water saturation and outflow rate of a spatial volume containing bentonite material

Chemical alterations are one of the most complicated parts of the model and are the most prone to numerical errors. This can be seen on the decay rate of carbonate and bicarbonate ions (*Figure* 11), which shows how much is reacting from these ions. They show correlation, mostly due to them being able to transform into each other. Rates vary from very high starting values to lower ones for the rest of the 200-year period. The rate of bicarbonate decay stabilizes towards the end of the period, but the decay rate of carbonate continues to fluctuate.

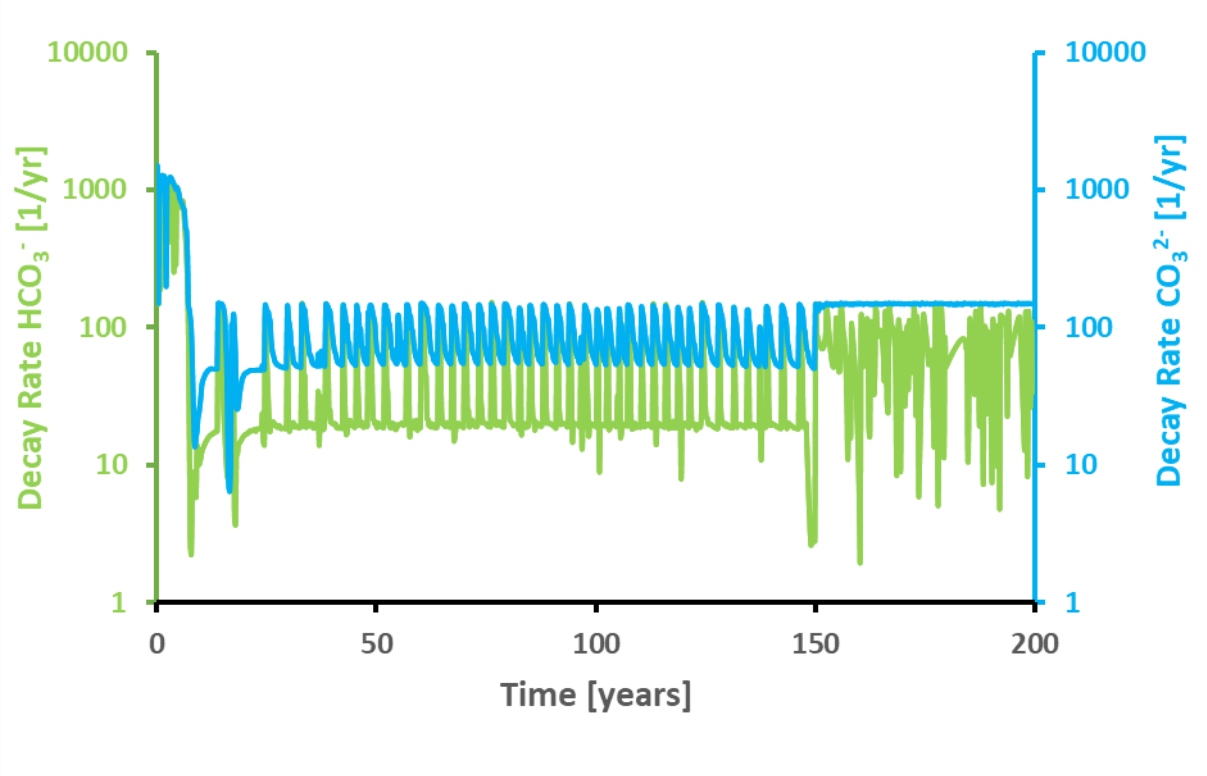


Figure 11 – Decay rate of carbonate and bicarbonate ions over time

In the canister, a slowly increasing corrosion rate can be observed (*Figure* 12). The process occurring is the transformation of iron into magnetite and hydrogen gas. It can be seen that hydrogen gas is present in the system. Gas flow is most likely responsible for the peaks in the amount of hydrogen, the generated gas is transferred from the canister spatial volumes to nearby ones like the bentonite, after they are generated by the means of corrosion.

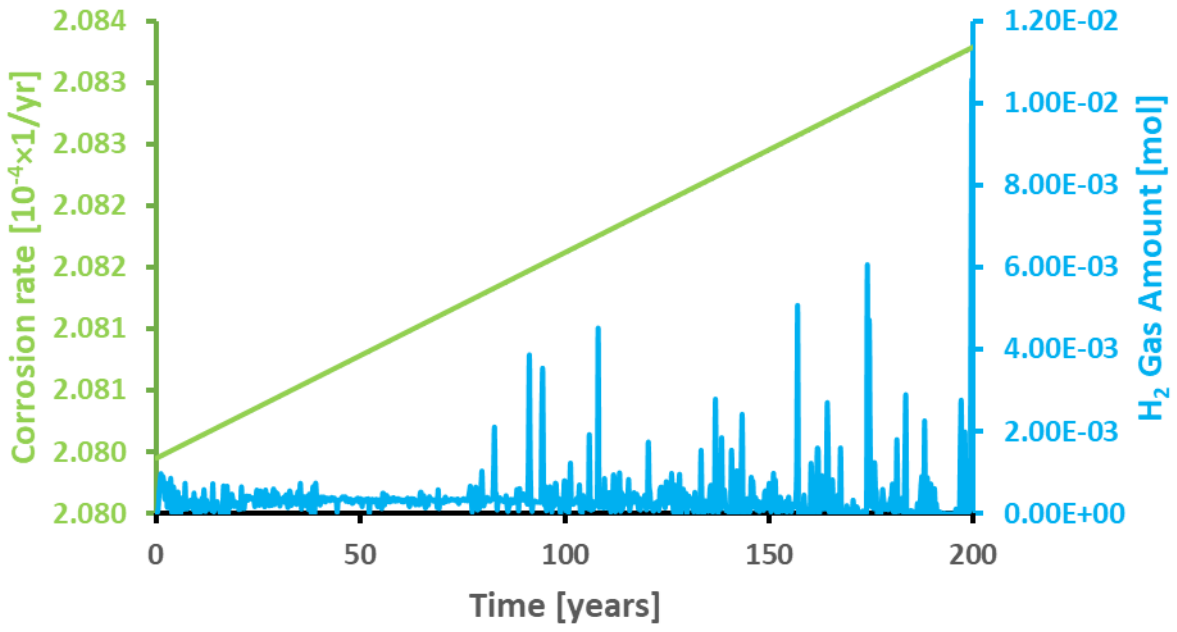


Figure 12 – Corrosion rate of the iron in the canister and the amount of hydrogen gas in the spatial

Code source

The model was developed within the GoldSim (GoldSim Technology Group, 2021d) proprietary modelling software tool using its Contaminant (Radionuclide) Transport Module (GoldSim Technology Group, 2021a). The model was presented as a poster (Bothi *et al.*, 2022) at IDG-TP's (Implementing Geological Disposal for radioactive waste Technical Platform) Symposium and Webinar: The role of optimisation in radioactive waste disposal programmes held in Zurich, Switzerland on 20–22 September and as a webinar (Bőthi and Olasz, 2024) for the GoldSim users in February 2024. There is no source code which may shared.

References

Bothi, Z., Olasz, T., Ditroi, A., Katzer, A., Wieser, M., Baksay, A., 2022. Novel Coupled Modelling Approach for Total System Performance Assessment in GoldSim, IGD-TP Symposium and Webinar: The role of optimisation in radioactive waste geological disposal programmes, Zurich, Switzerland.

Bőthi, Z., Olasz, T., 2024. An Approach for Representing Coupled Processes in Complex Contaminant Transport Models.

GoldSim Technology Group, 2021a. GoldSim Contaminant Transport Module. User's Guide., 14.0 ed. GoldSim Technology Group, Seattle, Washington, USA.

GoldSim Technology Group, 2021b. GoldSim Dashboard Authoring Module. User's Guide., 14.0 ed. GoldSim Technology Group, Seattle, Washington, USA.

GoldSim Technology Group, 2021c. GoldSim Distributed Processing Module. User's Guide., 14.0 ed. GoldSim Technology Group, Seattle, Washington, USA.

GoldSim Technology Group, 2021d. GoldSim. User's Guide. Probabilistic Simulation Environment., 14.0 ed. GoldSim Technology Group, Seattle, Washington, USA.

Innovative numerical methods for adjoint-state methods and sensitivity analyses and uncertainty analysis

[1]J. Samper, [1]A.C. Samper, [1]J. Samper II, [1]A. Mon, [1]B. Pisani, [1]L. Montenegro

[1]UDC(ENRESA), Spain

Abstract

This report presents the work done so far by UDC on task 4 of DONUT WP, which deals with innovative numerical methods for adjoint-state methods and sensitivity analyses and uncertainty analysis.

Adjoint state methods

The adjoint state (AS) method has been widely used for calculating local derivatives of objective functions in the solution of the inverse problem of parameter estimation and computing model sensitivities. The adjoint state method can be applied: (a) to the continuous version of the original problem (continuous AS method) or (b) to the discretized form of the original problem (discrete AS method). The continuous AS method consists of deriving the AS equations from the partial differential equations (PDEs) of the original problem and then solving numerically the resulting PDEs of the AS. The discrete adjoint state method consists of obtaining the finite element equations of the AS directly from the discretized equations of the original problem.

The work done on adjoint-state methods has included: 1) Development of the equations for the continuous and discretized flow and transport equations for conservative species; 2) Development of the continuous and discrete adjoint state equations; 3) Development of the equations for the continuous and discretized transport equations for weakly-reactive species; 4) Development of the adjoint state equations for the continuous reactive transport equations; 5) Application of AS to Case 1 of reactive system with mineral phases at equilibrium or kinetically controlled; 6) Application of AS to Case 2 of reactive system with multicomponent cation exchange reactions; 7) Development of the adjoint state equations for the discretized reactive transport equations and development of the equations for its implementation into CORE^{2D} V5; and 8) Application of AS to a test case of 1D solute transport through a column.

Sensitivity analyses and uncertainty quantification

The work on sensitivity analysis and uncertainty quantification methods includes: 1) The compilation of global sensitivity methods applications in radioactive waste disposal; 2) The use and the application of global sensitivity methods such as Sobol, Morris and VARS (“Variogram Analysis of Response Surfaces”) methods. Sophisticated sensitivity analysis methods have been developed for numerical models based on the analysis of model results. VARS is a comprehensive and efficient tool to analyze the structure of a response function Y. Here we present the application of the VARS method to the calculation of the global sensitivities of the predictions of a reactive transport model for a HLW repository

in granite. The variables Yi considered in the analysis include pH, redox potential and the volume of corrosion products at 10^4, $2.5 \cdot 10^4$, $5 \cdot 10^4$ years. The parameters considered include: the corrosion rate of the metallic canister, the diffusion coefficient of the solutes in the bentonite barrier, the flow of groundwater through the granitic host rock, the cation exchange selectivity of Fe and the equilibrium constant of magnetite. The results of the global sensitivity analysis lead to the conclusion that the most influential parameters are the corrosion rate, the flow in the granite and the cationic selectivity of Fe.

Significance Statement

The evaluation of the long-term safety of radioactive waste repositories requires using complex non-linear reactive transport models. Model predictions have uncertainties which often come from uncertainties in model parameters. Reducing such uncertainty requires the identification of the parameters causing the largest model output uncertainties and the ranking of the sensitivity indexes. UDC contribution is important because combines both local (adjoint state) and global (VARS, Variogram Analysis of Response Surfaces) sensitivity methods. The adjoint state method provides CPU-efficient estimates of the local sensitivities of model outputs to model parameters. VARS is a comprehensive and efficient tool to analyze the global structure of a model output. UDC has performed an innovative application of VARS to the output of a reactive transport model of a HLW radioactive repository in granite. The selected output variables include pH, redox potential (Eh) and volume of corrosion products at t = 10^4, $2 \cdot 10^4$, $5 \cdot 10^4$ years. The sensitivity analysis was performed for: the metallic canister corrosion rate, the diffusion coefficient of the solutes in the bentonite barrier, the groundwater flux through the granitic host rock, the cation exchange selectivity of Fe and the log-equilibrium constant of magnetite (log K). The results of the global sensitivity analysis indicate that the most influential parameters on the selected outputs are the corrosion rate, the flow in the granite and the cationic selectivity of Fe. The sensitivity of the computed pH with respect to corrosion rate and Fe selectivity increase with time while that of the log K of magnetite decreases. The sensitivity index of the bentonite diffusion coefficient, however, remains constant in time. Like pH, the most influential parameters for the computed Eh are the corrosion rate and the groundwater flow through the granite. The sensitivities of the corrosion rate and the log K of magnetite decrease with time while that of groundwater flow increases with time. The most influential parameter for the volume of corrosion products is the corrosion rate while the second is the diffusion coefficient of the bentonite. The parameter ranking of VARS is similar to those of Morris and Sobol methods VARS (IVARS50) show some differences. These differences are largest for the Morris indexes.

Table of content

List of figures

List of Tables

1. ADJOINT STATE METHODS

1.1 Introduction

The adjoint state method has been widely used for calculating the gradient of the objective function in the solution of the inverse problem of parameter estimation (Townley and Wilson, 1985; Carrera and Neuman, 1986, Samper et al., 1989). Adjoint state (AS) methods have been widely used in optimization and inverse problems in groundwater flow models (Townley and Wilson, 1985; Sun, 1994) and conservative solute transport models (Samper, 1986; Samper et al., 1986, Ackerer et al., 2014). Evaluating the adjoint state requires a single calculation like that of the original problem, regardless the number of parameters. The AS method was used for the estimation of variograms and spatial covariance structures in combination with maximum likelihood cross-validation (Samper and Neuman, 1989). The adjoint state method can be applied to the continuous version of the original problem or to the discretized form of the problem. Most of the reported applications rely upon discrete approaches. A comparison of the continuous and discrete adjoint states for solving the inverse problem of groundwater flow in heterogeneous dual porosity systems is presented by Delay et al (2017). The use of a coarse discretization grid for solving the continuous AS equations may not result in a significant gain of computational costs Delay et al (2017). Both discrete and continuous AS with sufficiently refined grids lead to similar results. They concluded that: 1) The content of data and the class of sought solutions have more practical incidence in the inverse problem than the numerical techniques used in solving the adjoint states; 2) The increase of the time step does not enhance the overall efficiency of the inverse procedure; and 3) The continuous adjoint state is a smart choice, given its non-intrusive character and the subsequent ease of implementation on non-proprietary codes. The AS method has been applied recently to model the steady-state sequentially coupled radionuclide transport in porous media (Hayek et al., 2019, 2020). To the best of our knowledge, the AS method has not been applied yet for modelling reactive solute transport.

The work done on adjoint-state methods has included: 1) Development of the equations for the continuous and discretized flow and transport equations for conservative species; 2) Development of the continuous and discrete adjoint state equations; 3) Development of the equations for the continuous and discretized transport equations for weakly-reactive species; 4) Development of the adjoint state equations for the continuous reactive transport equations; 5) Application of AS to Case 1 of reactive system with mineral phases at equilibrium or kinetically controlled; 6) Application of AS to Case 2 of reactive system with multicomponent cation exchange reactions; 7) Development of the adjoint state equations for the discretized reactive transport equations and development of the equations for its implementation into CORE^{2D} V5; and 8) Application of AS to a test case of 1D solute transport through a column.

1.2 AS methods for single-species conservative transport

1.2.1 Single-species conservative solute transport

The partial differential equation (PDE) of conservative solute transport is given by:

$$\boldsymbol{\nabla} \cdot (\emptyset \boldsymbol{D} \nabla c - \boldsymbol{q} c) + q_r c_r - q_d c - \lambda_d \emptyset c + \emptyset g = \frac{\partial}{\partial t} \emptyset c \quad \text{in RxT} \tag{1.1}$$

where $\emptyset$ is porosity, $\boldsymbol{D}$ is the dispersion/diffusion tensor, c is the solute concentration, $\boldsymbol{\nabla c}$ is the solute concentration gradient, $\boldsymbol{q}$ is the Darcy velocity, q_r and q_d are water source/sink terms, c_r is the concentration of the water inflow, λ_d is the decay constant, g is the solute mass source term, R is the space domain, t is time and T is the time domain. The Darcy velocity is given by Darcy's law:

$$\boldsymbol{q} = -\boldsymbol{K}\boldsymbol{\nabla}\boldsymbol{h} \tag{1.2}$$

where $\boldsymbol{K}$ is the hydraulic conductivity tensor and $\boldsymbol{\nabla h}$ is the gradient of the hydraulic head.

The initial and boundary conditions of the solute transport in Eq.(1.1) are given by:

$$(\emptyset \boldsymbol{D}\boldsymbol{\nabla}\boldsymbol{c} - \boldsymbol{q}c)\,\boldsymbol{n} = -\beta_{bc}(c - C) + G - \boldsymbol{q}\boldsymbol{n}C \quad \text{in } \Gamma_1 \text{ x T} \tag{1.3}$$

$$(\emptyset \boldsymbol{D}\boldsymbol{\nabla}\boldsymbol{c} - \boldsymbol{q}c)\,\boldsymbol{n} = -\beta_{bc}(c - C) + G - \boldsymbol{q}\boldsymbol{n}c \quad \text{in } \Gamma_2 \text{ x T} \tag{1.4}$$

$$c = c_0 \quad for \;\; t = t_0 \quad in\; R \tag{1.5}$$

where $\Gamma 1$ denotes the inflow and no-flow boundaries, $\Gamma 2$ denotes the outflow boundaries, $\boldsymbol{n}$ is a unit vector normal to the boundary pointing outwards, C is the external concentration along the boundary, G is the prescribed solute mass flux along the boundary, β_{bc} is a parameter controlling the type of solute transport boundary and c_0 is the steady-state concentration.

1.2.2 Objective function

Solving the inverse problem of solute transport requires to minimize an objective function, Ω. The integral form of the objective function has the following expression:

$$\Omega = \int_R J_1(c_0)\, dR + \iint_{R,T} [J_2(c) + J_3(\bar{\boldsymbol{\gamma}})]\, dRdt \tag{1.6}$$

where J_1, J_2, J_3 are known functions defined by the user and $\bar{\boldsymbol{\gamma}}$ is a vector of model parameters. In conservative solute transport, these parameters are $\emptyset, \boldsymbol{D}, q_r, q_d, c_r, g, \boldsymbol{K}, \beta_{bc}, C, G$.

Similarly to Hayek et al. (2019), the following objective functions can be defined:

$$\Omega_1(t^l) = \iint_{R,T} \emptyset\, c(x,t)\, \delta(t^l)\, dRdt \tag{1.7}$$

$$\Omega_2(x_j, t^l) = \iint_{R,T} c(x,t)\, \delta(x_j)\, \delta(t^l)\, dRdt \tag{1.8}$$

In order to minimize Ω, its gradient can be obtained in terms of the adjoint state variables, which in turn are obtained from the original problem. The AS equations can be derived from the continuous and discretized versions of the original problem.

1.2.3 Continuous adjoint state

The continuous adjoint state method consists in deriving the AS equations from the PDEs of the original problem and then solving numerically the resulting PDEs of the AS. The continuous adjoint state method involves the following steps:

- Define the adjoint state variable for solute concentrations, $\boldsymbol{\tau}$. The model parameters γ_s are collected in a vector $\bar{\boldsymbol{\gamma}}$.
- Define the objective function in terms of the variables and model parameters, such as that in Eq. (1.6)
- Take derivatives of the solute transport PDEs and associated initial and boundary conditions in Eqs. (1.1) - (1.5) with respect to a generic parameter γ_s. The derivatives are denoted with a superscript $'$, i.e., $\boldsymbol{r}'$ denotes the derivative of $\boldsymbol{r}$ with respect to γ_s.
- Multiply the resulting derivative equations by the adjoint state τ and integrate in RxT.
- Apply Green's first identity and divergence theorem to substitute the divergence terms by lower-order derivatives. The resulting boundary integrals are evaluated by their respective equations

obtained after steps 3 and 4.

- Similarly, eliminate the terms involving $\nabla c'$ by applying Green's first identity and divergence theorem.
- Lump together all the terms containing the derivatives of the solute concentrations with respect to the parameters in ${F_c}^1$. The resulting expression is given by:

$$F^1 + {F_c}^1 = 0 \tag{1.9}$$

- Take derivatives of the objective function Ω with respect to γ_s and add it to Eq. (1.9). The gradient of the objective function becomes:

$$\frac{\mathrm{d}\Omega}{\mathrm{d}\gamma_s} = F + F_c \tag{1.10}$$

where F and F_c have the following expressions:

$$F = F^1 + \iint_{R,T} \frac{\partial J_3}{\partial \gamma_s} dRdt \tag{1.11}$$

$$F_c = {F_c}^1 + \iint_{R,T} \frac{\partial J_2}{\partial c} c' dRdt \tag{1.12}$$

- Select $\boldsymbol{\tau}$ so that F_c cancels out. This leads to a system of adjoint PDEs for the adjoint state variable, which include its initial and boundary conditions.
- Calculate the derivative of the objective function with respect to the selected model parameters, which involves the original and the adjoint state variables.

The PDE of the continuous adjoint state of transient solute concentration and its associated "final" and boundary conditions are given by:

$$\boldsymbol{\nabla} \cdot (\emptyset \boldsymbol{D} \boldsymbol{\nabla} \boldsymbol{\tau} + \boldsymbol{q} \boldsymbol{\tau}) - \boldsymbol{\tau} \cdot (q_d + \lambda_d \emptyset) + \frac{\partial J_2}{\partial c} = -\frac{\partial \tau \emptyset}{\partial t} \text{ in RxT} \tag{1.13}$$

$$\emptyset \boldsymbol{D} \nabla \tau \cdot \boldsymbol{n} = -\beta_{bc} \tau \text{ in } \Gamma 1 \text{ x T} \tag{1.14}$$

$$(\emptyset \boldsymbol{D} \nabla \tau + \boldsymbol{q} \tau) \cdot \boldsymbol{n} = -\beta_{bc} \tau \text{ in } \Gamma 2 \text{ x T} \tag{1.15}$$

$$\tau(x, t_f) = 0 \text{ in R} \tag{1.16}$$

where t_f is the final time.

Numerical methods are used to solve the adjoint PDE. By applying the same discretization method as in the original problem (Galerkin finite elements for spatial discretization and finite differences for time discretization), the finite element equations of the adjoint states are given by:

$$\left(\bar{\bar{\boldsymbol{E}}}\theta_1 - \frac{\bar{\bar{\boldsymbol{F}}}}{\Delta t^k}\right) \bar{\boldsymbol{\tau}}^{\boldsymbol{k-1}} + \left[\bar{\bar{\boldsymbol{E}}}(1-\theta_1) + \bar{\bar{\boldsymbol{F}}}/\Delta t^k\right] \bar{\boldsymbol{\tau}}^{\boldsymbol{k}} = \bar{\boldsymbol{r}}_{\boldsymbol{\tau}}^{\boldsymbol{k}} \quad k = 1, 2, \ldots, K \tag{1.17}$$

where $\bar{\bar{\boldsymbol{F}}}$ is a square N x N matrix, $\bar{\bar{\boldsymbol{E}}}$ is a N x N symmetric square matrix, θ_1 is a time discretization parameter ($0 < \theta_1 < 1$), $\bar{\boldsymbol{\tau}}^{\boldsymbol{k}}$ is the vector of nodal adjoint state concentrations at time t^k and K is the total number of time intervals where concentrations are calculated. Matrixes $\bar{\bar{\boldsymbol{E}}}$, $\bar{\bar{\boldsymbol{F}}}$ are the same matrixes defined in the original problem in Eqs. 1.17 and vector $\bar{\boldsymbol{r}}_{\boldsymbol{\tau}}^{\boldsymbol{k}}$ is given by:

$${r_\tau}^k{}_i = \int_R \frac{\partial J_2}{\partial c} \xi_i \, dR \Big|_k \tag{1.18}$$

The analysis of the continuous version of the adjoint state equations indicates that the adjoint state τ satisfies a PDE similar to that of c. However, the "adjoint solute" corresponding to τ moves upstream, unlike the solute with concentration c. Therefore, τ must be solved backwards in time from the final to the initial time. Furthermore, the PDE of the adjoint state are intrinsically linear, unlike the equations of the original problem which may be non-linear for non-linear chemical sinks/sources.

1.2.4 Comparison of the continuous and discrete adjoint methods for solute transport

Both discrete and continuous adjoint state methods lead to linear equations. Moreover, in both cases the adjoint state equations must be solved backwards in time, starting at the final conditions where $\tau = 0$.

Although the discrete and continuous adjoint state methods lead in general to different algebraic equations, their solutions are consistent because both methods reflect different ways of discretizing the same partial differential equations. The continuous method is versatile because it allows using a numerical method and a discretization scheme for the adjoint states independent of those of the forward problem.

1.3 AS methods for multicomponent conservative transport

This chapter presents the continuous AS equations for a conservative transport system of N species.

1.3.1 Multicomponent conservative solute transport

The mathematical formulation of the original problem is given by the following PDEs, boundary and initial conditions:

$$\boldsymbol{L}(\bar{\boldsymbol{c}}) + \phi\bar{\boldsymbol{g}} = \frac{\partial\phi\bar{\boldsymbol{c}}}{\partial t} \tag{1.19}$$

$$(\phi\boldsymbol{D}\boldsymbol{\nabla}\bar{\boldsymbol{c}} - \mathrm{q}\bar{\boldsymbol{c}}) \cdot \boldsymbol{n}|_{\Gamma} = -\beta(\bar{\boldsymbol{c}} - \bar{\boldsymbol{C}}) + \bar{\boldsymbol{G}} \tag{1.20}$$

$$\bar{\boldsymbol{c}}(\boldsymbol{x}, 0) = \bar{\boldsymbol{c_0}} \tag{1.21}$$

where $\bar{\boldsymbol{c}}$ is the vector of N primary species concentrations, $\boldsymbol{L}(\bar{\boldsymbol{c}}) = \boldsymbol{\nabla}(\phi\boldsymbol{D}\boldsymbol{\nabla}\bar{\boldsymbol{c}} - \mathrm{q}\bar{\boldsymbol{c}})$ is the solute transport operator, ϕ is the porosity, q is Darcy's velocity, $\boldsymbol{D}$ is the diffusion/dispersion coefficient, $\bar{\boldsymbol{g}}$ are solute mass sources, $\boldsymbol{x}$ is space and t is time. The boundary is defined by Γ, where β is a parameter controlling the type of transport boundary, $\bar{\boldsymbol{C}}$ are the external concentrations along the boundary and $\bar{\boldsymbol{G}}$ are the prescribed solute mass flux along the boundary.

The objective function in this case is given in terms of the N solute concentrations $\bar{\boldsymbol{c}}$ and the model parameters $\bar{\boldsymbol{\gamma}}$:

$$\Omega = \iint_{R,T} J_2(\bar{\boldsymbol{c}})\, dRdt + \iint_{R,T} J_3(\bar{\boldsymbol{\gamma}})\, dRdt \tag{1.22}$$

where RxT are the space and time domains, respectively.

The selected model parameters are the porosity ϕ, the dispersion/diffusion coefficient $\boldsymbol{D}$, the external concentrations along the boundary $\bar{\boldsymbol{C}}$ and the prescribed solute mass flux along the boundary $\bar{\boldsymbol{G}}$.

1.3.2 Adjoint state formulation

The adjoint states associated to the original problem are collected in a column vector of N components, $\bar{\boldsymbol{\tau}}$. These adjoint states satisfy the following equations:

$$\boldsymbol{L_\tau}(\bar{\tau}) + \frac{\partial J_2}{\partial \bar{c}} = -\phi \frac{\partial \bar{\tau}}{\partial t} \tag{1.23}$$

$$(\phi \boldsymbol{D} \boldsymbol{\nabla} \bar{\tau} + \mathrm{q}\bar{\tau}) \cdot \boldsymbol{n}|_\Gamma = -\overline{\tau \beta} \tag{1.24}$$

$$\bar{\tau}(\boldsymbol{x}, t_f) = \bar{\mathbf{0}} \tag{1.25}$$

where $\boldsymbol{L_\tau}(\bar{\tau})$ is the adjoint state solute transport operator:

$$\boldsymbol{L_\tau}(\bar{\tau}) = \boldsymbol{\nabla}(\phi \boldsymbol{D} \boldsymbol{\nabla} \bar{\tau} + \mathrm{q}\bar{\tau}) \tag{1.26}$$

Eqs.(1.23) and (1.25) show that adjoint states $\bar{\tau}$ satisfy similar equations as in the original problem, with an extra source term which depends on the derivatives of the objective function. Moreover, the AS equations must be solved backwards in time.

1.4 AS methods for a reactive system with n kinetically-controlled minerals

This chapter presents the continuous AS equations for reactive transport involving N primary aqueous species and N_p kinetically controlled mineral phases with precipitation/dissolution rates $\bar{\boldsymbol{r}}_{\boldsymbol{p}}$.

1.4.1 Reactive transport problem

The mathematical formulation of the original problem is given by:

$$\boldsymbol{L}(\bar{c}) + \bar{\bar{\boldsymbol{S}}}_{\boldsymbol{p}} \bar{\boldsymbol{r}}_{\boldsymbol{p}} = \frac{\partial \phi \bar{\boldsymbol{c}}}{\partial t} \tag{1.27}$$

$$(\phi \boldsymbol{D} \boldsymbol{\nabla} \bar{\boldsymbol{c}} - \mathrm{q}\bar{\boldsymbol{c}}) \cdot \boldsymbol{n}|_\Gamma = -\beta(\bar{c} - \bar{\boldsymbol{C}}) + \bar{\boldsymbol{G}} \tag{1.28}$$

$$\bar{\boldsymbol{c}}(\boldsymbol{x}, 0) = \overline{\boldsymbol{c_0}} \tag{1.29}$$

The matrix $\bar{\bar{\boldsymbol{S}}}_{\boldsymbol{p}}$ is a (NxNp) rectangular matrix of stoichiometric coefficients given by:

$$\bar{\bar{\boldsymbol{S}}}_{\boldsymbol{p}} = \begin{pmatrix} \upsilon_{c1}^{p1} & \upsilon_{c1}^{p2} & \dots & \upsilon_{c1}^{Np} \\ \upsilon_{c2}^{p1} & \upsilon_{c2}^{p2} & \dots & \upsilon_{c2}^{Np} \\ \vdots & \vdots & \ddots & \vdots \\ \upsilon_{N}^{p1} & \upsilon_{N}^{p2} & \dots & \upsilon_{N}^{Np} \end{pmatrix}_{NxNp} \tag{1.30}$$

where υ_{cj}^{pi} is the stoichiometric coefficient of the j-th primary species involved in the chemical reaction of the i-th mineral phase.

The column vector $\bar{\boldsymbol{r}}_{\boldsymbol{p}}$ of N_p precipitation/dissolution kinetic rates given by:

$$\bar{\boldsymbol{r}}_p = \begin{pmatrix} r_{p1}(c_1, \dots, c_N) \\ \vdots \\ r_{Np}(c_1, \dots, c_N) \end{pmatrix}_{Npx1} \tag{1.31}$$

where the term r_{pi} refers to the kinetic rate of the i-th mineral phase, which in general depends on the concentrations of the primary species involved in the reaction.

The term $\overline{\overline{\boldsymbol{S}}}_{\boldsymbol{p}}\bar{\boldsymbol{r}}_{\boldsymbol{p}}$ in Eq. (1.27) is a sink/source term of the concentrations of primary species associated to the kinetic precipitation/dissolution of the mineral phases with concentrations $\bar{\boldsymbol{P}}$:

$$\overline{\overline{\boldsymbol{S}}}_{\boldsymbol{p}}\bar{\boldsymbol{r}}_{\boldsymbol{p}} = \pm\frac{\partial\phi\bar{\boldsymbol{P}}}{\partial t} \tag{1.32}$$

The objective function is given in terms of the N primary species concentrations $\bar{\boldsymbol{c}}$ and the model parameters $\bar{\boldsymbol{\gamma}}$:

$$\Omega = \iint_{R,T} J_2(\bar{\boldsymbol{c}})\, dRdt + \iint_{R,T} J_3(\bar{\boldsymbol{\gamma}})\, dRdt \tag{1.33}$$

where RxT are the space and time domain.

The selected model parameters are the porosity ϕ, the dispersion/diffusion coefficient $\boldsymbol{D}$, the external concentrations along the boundary $\bar{\boldsymbol{C}}$ and the prescribed solute mass flux along the boundary $\bar{\boldsymbol{G}}$.

1.5 Adjoint state formulation

The adjoint states associated to the original problem are defined by the column vector of N components, $\bar{\boldsymbol{\tau}}$. The adjoint state equations for a reactive system with N_p kinetically controlled mineral phases are given by:

$$\boldsymbol{L}_{\boldsymbol{\tau}}(\bar{\boldsymbol{\tau}}) + \overline{\overline{\boldsymbol{R}}}^{\boldsymbol{T}}\,\overline{\overline{\boldsymbol{S}}}_{\boldsymbol{p}}^{\boldsymbol{T}}\,\bar{\boldsymbol{\tau}} + \frac{\partial J_2}{\partial \bar{\boldsymbol{c}}} = -\phi\frac{\partial\bar{\boldsymbol{\tau}}}{\partial t} \tag{1.34}$$

$$(\phi\boldsymbol{D}\boldsymbol{\nabla}\bar{\boldsymbol{\tau}} + \mathrm{q}\bar{\boldsymbol{\tau}})\cdot\boldsymbol{n}|_{\Gamma} = -\overline{\boldsymbol{\tau}\boldsymbol{\beta}} \tag{1.35}$$

$$\bar{\boldsymbol{\tau}}(\boldsymbol{x}, t_f) = \bar{\boldsymbol{0}} \tag{1.36}$$

where $\boldsymbol{L}_{\boldsymbol{\tau}}(\bar{\boldsymbol{\tau}})$ is the adjoint state solute transport operator given by Eq. (1.26).

The matrix $\overline{\overline{\boldsymbol{S}}}_{\boldsymbol{p}}$ is given by Eq. (1.30) and $\overline{\overline{\boldsymbol{R}}}$ is a (NpxN) matrix associated to the first-order derivatives of the precipitation/dissolution kinetic rates with respect to the primary species concentrations:

$$\overline{\overline{\boldsymbol{R}}} = \begin{pmatrix} \frac{\partial r_{p1}}{\partial c_1} & \frac{\partial r_{p1}}{\partial c_2} & \cdots & \frac{\partial r_{p1}}{\partial c_N} \\ \frac{\partial r_{p2}}{\partial c_1} & \frac{\partial r_{p2}}{\partial c_2} & \cdots & \frac{\partial r_{p2}}{\partial c_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial r_{Np}}{\partial c_1} & \frac{\partial r_{Np}}{\partial c_2} & \cdots & \frac{\partial r_{Np}}{\partial c_N} \end{pmatrix}_{NpxN} \tag{1.37}$$

Eqs.(1.34) - (1.36) show that the adjoint states of a reactive system with kinetically controlled mineral phases satisfy similar equations as those of the original problem, with additional sink/source terms which depend on the derivatives of the objective function. Moreover, there is a source term in $\bar{\boldsymbol{\tau}}$ which is associated to the chemical reactions. It should be noticed that contrary to the PDEs of the original problem, which in general are nonlinear, the PDEs of the adjoint state variables are linear.

1.6 AS methods for a system with minerals in chemical equilibrium

This chapter presents the continuous AS equations for reactive transport involving N primary aqueous species and N_p minerals with precipitation/dissolution reactions in chemical equilibrium.

1.6.1 Reactive transport problem

The mathematical formulation of the original problem is given by:

$$\boldsymbol{L}(\bar{\boldsymbol{c}}) = \frac{\partial \phi \bar{\boldsymbol{c}}}{\partial t} + \bar{\bar{\boldsymbol{S}}}_{\boldsymbol{p}} \frac{\partial \phi \bar{\boldsymbol{p}}}{\partial t} \tag{1.38}$$

$$(\phi \boldsymbol{D} \boldsymbol{\nabla} \bar{\boldsymbol{c}} \; - \; \mathrm{q}\bar{\boldsymbol{c}}\,) \cdot \boldsymbol{n}|_{\Gamma} = -\beta(\bar{\boldsymbol{c}} - \bar{\boldsymbol{C}}) + \bar{\boldsymbol{G}} \tag{1.39}$$

$$\bar{\boldsymbol{c}}(\boldsymbol{x}, 0) = \bar{\boldsymbol{c_0}} \tag{1.40}$$

where the matrix $\bar{\bar{\boldsymbol{S}}}_{\boldsymbol{p}}$ of stoichiometric coefficients is given by Eq.(1.30) and $\bar{\boldsymbol{p}}$ is a column vector (Npx1) of mineral phases concentrations, which are related to the precipitated concentrations, $\bar{\boldsymbol{P}}$, through the stoichiometric coefficients:

$$\bar{\boldsymbol{P}} = \bar{\bar{\boldsymbol{S}}}_{\boldsymbol{p}} \bar{\boldsymbol{p}} \tag{1.41}$$

The chemical equilibrium is given by the following equation:

$$\boldsymbol{ln}\bar{\boldsymbol{K}} = \bar{\bar{\boldsymbol{S}}}_{\boldsymbol{p}}^{\boldsymbol{t}} \, \boldsymbol{ln}\bar{\boldsymbol{c}} \tag{1.42}$$

where $\bar{\boldsymbol{K}}$ is the column vector (N_px1) of chemical equilibrium constants associated to each precipitation/dissolution reaction.

The original problem formulation includes N transport PDEs given by Eq.(1.38) and N_p algebraic equations associated to the chemical reactions in Eq. (1.42), so that the dependent variables are the primary species concentrations, $\bar{\boldsymbol{c}}$, and the mineral phases concentrations, $\bar{\boldsymbol{p}}$.

The objective function in this case is given in terms of the N primary species concentrations $\bar{\boldsymbol{c}}$, the N_p mineral concentrations $\bar{\boldsymbol{p}}$ and the model parameters $\bar{\boldsymbol{\gamma}}$:

$$\Omega = \iint_{R,\,T} J_2(\bar{\boldsymbol{c}})\, dRdt + \iint_{R,\,T} J_3(\bar{\boldsymbol{\gamma}})\, dRdt + \iint_{R,\,T} J_4(\bar{\boldsymbol{p}})\, dRdt \tag{1.43}$$

The selected model parameters are the porosity ϕ, the dispersion/diffusion coefficient $\boldsymbol{D}$, the chemical equilibrium constants $\bar{\boldsymbol{K}}$**,** the external concentrations along the boundary $\bar{\boldsymbol{C}}$ and the prescribed solute mass flux along the boundary $\bar{\boldsymbol{G}}$.

1.6.2 Adjoint state formulation

Given that the original formulation includes N + N_p equations, the adjoint state method for a reactive system of mineral phases in equilibrium involves: a) N adjoint state variables $\bar{\boldsymbol{\tau}}$ for the transport PDEs in Eq. (1.38), and b) N_p additional adjoint states $\bar{\boldsymbol{\varepsilon}}$ for the algebraic equilibrium conditions in Eq. (1.42).

As a result, the AS equations for a reactive system with mineral phases in equilibrium are given by:

$$\boldsymbol{L_{\tau}}(\bar{\boldsymbol{\tau}}) - \bar{\bar{\boldsymbol{S}}}_{\boldsymbol{p}}\, \bar{\boldsymbol{\varepsilon}} \; + \; \frac{\partial J_2}{\partial \bar{\boldsymbol{c}}} = -\phi \frac{\partial \bar{\boldsymbol{\tau}}}{\partial t} \tag{1.44}$$

$$(\phi \boldsymbol{D} \boldsymbol{\nabla} \bar{\boldsymbol{\tau}} \; + \; \mathrm{q}\bar{\boldsymbol{\tau}}) \cdot \boldsymbol{n}|_{\Gamma} = -\,\overline{\boldsymbol{\tau \beta}} \tag{1.45}$$

$$\bar{\boldsymbol{\tau}}\big(\boldsymbol{x}, t_f\big) = \bar{\boldsymbol{0}} \tag{1.46}$$

$$\overline{\overline{\boldsymbol{S}}}_{\boldsymbol{p}}^{\,T}\frac{\partial\overline{\boldsymbol{\tau}}}{\partial t}=-\frac{\partial J_4}{\partial\overline{\boldsymbol{p}}} \tag{1.47}$$

where $\boldsymbol{L}_{\boldsymbol{\tau}}(\overline{\boldsymbol{\tau}})$ is the adjoint state solute transport operator given by Eq.(1.26) and the matrix $\overline{\overline{\boldsymbol{S}}}_{\boldsymbol{p}}$ is given by Eq. (1.30).

Pre-multiplying Eq.(1.44) by $\overline{\overline{\boldsymbol{S}}}_{\boldsymbol{p}}^{\,T}$ and substituting the right-hand side by Eq. (1.47) leads to the following expression:

$$\overline{\overline{\boldsymbol{S}}}_{\boldsymbol{p}}^{\,T}\boldsymbol{L}(\overline{\boldsymbol{\tau}})-\overline{\overline{\boldsymbol{S}}}_{\boldsymbol{p}}^{\,T}\overline{\overline{\boldsymbol{S}}}_{\boldsymbol{p}}\,\overline{\boldsymbol{\varepsilon}}+\overline{\overline{\boldsymbol{S}}}_{\boldsymbol{p}}^{\,T}\frac{\partial J_2}{\partial\overline{\boldsymbol{c}}}=\phi\frac{\partial J_4}{\partial\overline{\boldsymbol{p}}} \tag{1.48}$$

1.7 AS methods for a system with cation exchange reactions

This chapter presents the continuous AS equations for multicomponent cation exchange reactive transport in groundwater.

1.7.1 Reactive transport problem

The transport PDEs for a set of exchanging cations in a 3D domain is given by:

$$L(\overline{\boldsymbol{c}})=\phi\frac{\partial\overline{\boldsymbol{c}}}{\partial t}+\lambda\overline{\overline{\boldsymbol{D}}}_{\boldsymbol{z}}\frac{\partial\overline{\boldsymbol{\beta}}}{\partial t} \tag{1.49}$$

$$L(c_N)=\phi\frac{\partial c_N}{\partial t}+\lambda\frac{\partial\beta_N}{\partial t} \tag{1.50}$$

where $\overline{\boldsymbol{c}}$ is the column vector of N-1 dissolved cations concentrations, $\overline{\boldsymbol{\beta}}$ is the column vector of (N-1) equivalent fractions of the exchange cations, $\overline{\overline{\boldsymbol{D}}}_{\boldsymbol{z}}$ is a (N-1)x(N-1) diagonal matrix of elements ${z_i}^{-1}$, where z_i are the cation charges. c_N is the concentration of the reference cation, β_N is the equivalent fraction of the reference cation and λ is a constant given by:

$$\lambda=\frac{\mathrm{CEC}\,\rho_d}{100} \tag{1.51}$$

where CEC is cation exchange capacity (meq/100g) and ρ_d is bulk density. Initial and boundary conditions are given in general by:

$$\phi\boldsymbol{D}\boldsymbol{\nabla}\overline{\boldsymbol{c}}\cdot\boldsymbol{n}|_{\Gamma}=-\beta_{bc}(\overline{\boldsymbol{c}}-\overline{\boldsymbol{C}})+\overline{\boldsymbol{G}} \tag{1.52}$$

$$\overline{\boldsymbol{c}}(\boldsymbol{x},t_0)=\overline{\boldsymbol{c_0}} \tag{1.53}$$

Cation exchange takes place when dissolved cations exchange with cations at interlayer positions, which can be described as an equilibrium reaction. According to the Gaines-Thomas convention, the mass-action-law of cation exchange is given by the following expression in terms of equivalent fractions β_i:

$$K_{ij}=\frac{{\beta_i}^{1/z_i}\cdot{c_j}^{1/z_j}}{{\beta_j}^{1/z_j}\cdot{c_i}^{1/z_i}} \tag{1.54}$$

where K_{ij} is the exchange coefficient or selectivity coefficient.

Equivalent fractions add to one by definition:

$$\sum_{i=1}^{N} \beta_i = 1 \tag{1.55}$$

Exchange reactions are usually expressed in terms of a reference cation. Considering N as the reference cation, Eq. (1.54) becomes:

$$K_{Nj} = \frac{{\beta_N}^{1/z_N} \cdot {c_j}^{1/z_j}}{{\beta_j}^{1/z_j} \cdot {c_N}^{1/z_N}} \quad j = 1, \dots, N-1 \tag{1.56}$$

Using matrix notation and taking logarithms in Eq. (1.56) leads to:

$$ln\overline{\boldsymbol{K}} = \overline{\overline{\boldsymbol{D}}}_{\boldsymbol{z}}\left(ln\overline{\boldsymbol{c}} - ln\overline{\boldsymbol{\beta}}\right) + \overline{\mathbf{1}}(ln\beta_N - lnc_N) \tag{1.57}$$

$$\overline{\mathbf{1}}^T\overline{\boldsymbol{\beta}} + \beta_N = 1 \tag{1.58}$$

The previous equations define the mathematical formulation of cation exchange reactive transport.

1.7.2 Adjoint state formulation

By following the steps of the continuous AS method, the adjoint states associated to the original problem are defined in *Table* 1.

The objective function in terms of the variables and model parameters is given by:

$$\Omega = \iint_{R,T} \left[J_2(\overline{\boldsymbol{c}}, c_N) + J_3(\overline{\boldsymbol{\gamma}}) + J_4(\overline{\boldsymbol{\beta}}, \beta_N)\right] dRdt \tag{1.59}$$

Table 1 - Adjoint states of reactive system with cation exchange.

Original problem equations	Adjoint state associated	Dimensions
Eq. (1.45)	$\overline{\boldsymbol{\tau}}$	(N-1) x 1 column vector
Eq. (1.46)	τ_N	Scalar
Eq. (1.53)	$\overline{\boldsymbol{\varepsilon}}$	(N-1) x 1 column vector
Eq. (1.54)	σ	Scalar

Once the AS method is applied, the adjoint state equations for a reactive system with cation exchange involving N cations are given by:

$$L(\overline{\boldsymbol{\tau}}) - \overline{\overline{\boldsymbol{D}}}_{\boldsymbol{c}}\overline{\overline{\boldsymbol{D}}}_{\boldsymbol{z}}\overline{\boldsymbol{\varepsilon}} + \frac{\partial J_2}{\partial \overline{\boldsymbol{c}}} = -\phi\frac{\partial \overline{\boldsymbol{\tau}}}{\partial t} \tag{1.60}$$

$$\overline{\overline{\boldsymbol{D}}}_{\boldsymbol{\beta}}\overline{\overline{\boldsymbol{D}}}_{\boldsymbol{z}}\overline{\boldsymbol{\varepsilon}} + \frac{\partial J_4}{\partial \overline{\boldsymbol{\beta}}} + \sigma\,\overline{\mathbf{1}} = -\lambda\overline{\overline{\boldsymbol{D}}}_{\boldsymbol{z}}\frac{\partial \overline{\boldsymbol{\tau}}}{\partial t} \tag{1.61}$$

$$L(\tau_N) + \frac{1}{c_N}\,\bar{\mathbf{1}}^T\bar{\boldsymbol{\varepsilon}} + \frac{\partial J_2}{\partial c_N} = -\phi\frac{\partial \tau_N}{\partial t} \tag{1.62}$$

$$-\frac{1}{\beta_N}\,\bar{\mathbf{1}}^T\bar{\boldsymbol{\varepsilon}} + \frac{\partial J_4}{\partial \beta_N} + \sigma = -\lambda\frac{\partial \tau_N}{\partial t} \tag{1.63}$$

where $\bar{\bar{\boldsymbol{D}}}_{\boldsymbol{c}}, \bar{\bar{\boldsymbol{D}}}_{\boldsymbol{z}}, \bar{\bar{\boldsymbol{D}}}_{\boldsymbol{\beta}}$ are (N-1)x(N-1) diagonal matrixes with entries $D_{c,ii} = c_i^{-1}$, $D_{z,ii} = z_i^{-1}$ and $D_{\beta,ii} = \beta_i^{-1}$.

The system of equations can be simplified by elimination. The resulting adjoint state equations are given by:

$$\sigma = -\frac{c_N}{\beta_N}\left[L(\tau_N) + \left(\phi + \lambda\frac{\beta_N}{c_N}\right)\frac{\partial \tau_N}{\partial t} + \frac{\beta_N}{c_N}\frac{\partial J_4}{\partial \beta_N} + \frac{\partial J_2}{\partial c_N}\right] \tag{1.64}$$

$$L(\bar{\boldsymbol{\tau}}^*) + \bar{\bar{\boldsymbol{A}}}^{-1}\bar{\boldsymbol{s}}^* = \bar{\bar{\boldsymbol{A}}}^{-1}\bar{\bar{\boldsymbol{B}}}\frac{\partial \bar{\boldsymbol{\tau}}^*}{\partial t} \tag{1.65}$$

where the superscript * indicates that vectors and matrixes are N dimensional.

The matrixes $\bar{\bar{\boldsymbol{A}}}^{-1}$, $\bar{\bar{\boldsymbol{B}}}$ and the vector $\bar{\boldsymbol{s}}^*$ are given by:

$$\bar{\bar{\boldsymbol{A}}}^{-1} = \begin{pmatrix} & \bar{\bar{\boldsymbol{I}}} & \\ -\frac{c_1 z_1}{c_N} & \cdots & -\frac{c_{N-1} z_{N-1}}{c_N} \end{pmatrix} \begin{matrix} w_1 \\ \vdots \\ w_{N-1} \\ w_N \end{matrix} \tag{1.66}$$

$$\bar{\boldsymbol{s}}^* = \begin{pmatrix} s_1 \\ \vdots \\ s_{N-1} \\ \sum_{i=1}^{N} c_i z_i \frac{\partial J_2}{\partial c_i} \end{pmatrix} \tag{1.67}$$

$$\bar{\bar{\boldsymbol{B}}} = \begin{pmatrix} & -\bar{\bar{\boldsymbol{R}}} & \\ -\phi c_1 z_1 & \cdots & -\phi c_{N-1} z_{N-1} \end{pmatrix} \begin{matrix} -\frac{c_N \beta_1}{\beta_N c_1} R_{NN} \\ \vdots \\ -\frac{c_N \beta_{N-1}}{\beta_N c_{N-1}} R_{NN} \\ -\phi c_N \end{matrix} \tag{1.68}$$

where $\bar{\bar{\boldsymbol{I}}}$ is the (N-1)x(N-1) identity matrix, $\bar{\bar{\boldsymbol{R}}}$ is a diagonal matrix with entries $R_{ii} = \phi + \lambda\frac{\beta_i}{c_i}$ and w_i, s_j are given by:

$$w_i = \frac{\beta_j}{c_j \sum_{i=1}^{N} \beta_i z_i} \;;\; i = 1, \dots, N \tag{1.69}$$

$$s_j = \frac{\partial J_2}{\partial c_j} + \frac{\beta_j}{c_j}\cdot\frac{\partial J_4}{\partial \beta_j} + \frac{\beta_j}{c_j}\frac{c_N}{\beta_N}\cdot\left(\frac{\beta_N}{c_N}\frac{\partial J_4}{\partial \beta_N} + \frac{\partial J_2}{\partial c_N}\right) \;;\; j = 1, \dots, N-1 \tag{1.70}$$

1.8 AS methods for a system with linearized cation exchange reactions

This chapter presents the continuous AS equations for multicomponent cation exchange reactive transport.

1.8.1 Reactive transport problem

The reactive transport PDEs for a set of N exchanging cations are given by:

$$L(\bar{\boldsymbol{c}}) = \frac{\partial \bar{\boldsymbol{c}}}{\partial t} + \frac{\partial \bar{\boldsymbol{w}}}{\partial t} \tag{1.71}$$

$$(\boldsymbol{D}\boldsymbol{\nabla}\bar{\boldsymbol{c}} - \mathrm{v}\bar{\boldsymbol{c}}) \cdot \boldsymbol{n}|_{\Gamma} = -\beta(\bar{\boldsymbol{c}} - \bar{\boldsymbol{C}}) + \bar{\boldsymbol{G}} \tag{1.72}$$

$$\bar{\boldsymbol{c}}(\boldsymbol{x}, 0) = \bar{\boldsymbol{c_0}} \tag{1.73}$$

where $\bar{\boldsymbol{c}}$ and $\bar{\boldsymbol{w}}$ are the dissolved cations and exchanged cations concentrations, respectively, and the solute transport operator is given by $L(\bar{\boldsymbol{c}}) = \boldsymbol{\nabla}(\boldsymbol{D}\boldsymbol{\nabla}\bar{\boldsymbol{c}} - \mathrm{v}\bar{\boldsymbol{c}})$. The exchanged cations concentration can be expressed in terms of the equivalent fractions β_i as follows:

$$w_i = \beta_i \frac{CEC\rho_d}{100\phi z_i} \tag{1.74}$$

where CEC is the cation exchange capacity, ρ_d is bulk density, ϕ is porosity and z_i is the cation charge. The equivalent fractions β_i add to one by definition:

$$\sum_{i=1}^{N} \beta_i = 1 \tag{1.75}$$

The mass-action-law is given by:

$$K_{1j} = \frac{{\beta_1}^{1/z_1} \cdot {c_j}^{1/z_j}}{{\beta_j}^{1/z_j} \cdot {c_1}^{1/z_1}} \tag{1.76}$$

where the subscript "1" denotes the reference cation and "j" is the j-th exchange cation.

According to Samper and Yang (2007), to derive semi-analytical solutions the mass-action-law in Eq. (1.76) is linearized by taking logarithms and performing 1st order Taylor expansion around selected reference values ${\beta_i}^0$ and ${c_i}^0$ [0].

As a result, the linear reactive transport equations of the original problem are given by:

$$L(\bar{\boldsymbol{c}}) = \bar{\bar{\boldsymbol{R}}} \frac{\partial \bar{\boldsymbol{c}}}{\partial t} \tag{1.77}$$

where $\bar{\bar{\boldsymbol{R}}}$ is a square matrix (NxN) of retardation coefficients given by:

$$\bar{\bar{\boldsymbol{R}}} = \bar{\bar{\boldsymbol{I}}} + CEC^* \bar{\bar{\boldsymbol{Z}}}^{-1} \bar{\bar{\boldsymbol{A}}}^{-1} \bar{\bar{\boldsymbol{B}}} \tag{1.78}$$

where $\bar{\bar{\boldsymbol{I}}}$ is the identity matrix, $\bar{\bar{\boldsymbol{Z}}}$ is a diagonal matrix of cation charges, $CEC^* = CEC\ \rho_d/100\phi$ and $\bar{\bar{\boldsymbol{A}}}, \bar{\bar{\boldsymbol{B}}}$ are matrixes given by:

$$\bar{\bar{\boldsymbol{A}}} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \frac{1}{z_1\beta_1^0} & \frac{-1}{z_2\beta_2^0} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{z_1\beta_1^0} & 0 & \dots & \frac{-1}{z_N\beta_N^0} \end{pmatrix} \tag{1.79}$$

$$\bar{\bar{B}} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ \frac{1}{z_1 c_1^0} & \frac{-1}{z_2 c_2^0} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{z_1 c_1^0} & 0 & \dots & \frac{-1}{z_N c_N^0} \end{pmatrix} \tag{1.80}$$

The objective function is given in terms of the N exchange cations concentrations and the model parameters:

$$\Omega = \iint_{R,T} J_2(\bar{\boldsymbol{c}})\, dRdt + \iint_{R,T} J_3(\bar{\boldsymbol{\gamma}})\, dRdt \tag{1.81}$$

where RxT are the space and time domain.

The selected model parameters are porosity ϕ and the dispersion/diffusion coefficient $\boldsymbol{D}$.

1.8.2 Adjoint state formulation

The adjoint state PDEs for a reactive system with a set of N exchange cations is given by:

$$\boldsymbol{L_\tau}(\bar{\tau}) + \frac{\boldsymbol{\partial J_2}}{\boldsymbol{\partial \bar{c}}} = -\bar{\bar{\boldsymbol{R}}}^{\boldsymbol{T}} \frac{\boldsymbol{\partial \bar{\tau}}}{\boldsymbol{\partial t}} \tag{1.82}$$

$$(\boldsymbol{D\nabla\bar{\tau}} + \mathrm{v}\bar{\tau}) \cdot \boldsymbol{n}|_\Gamma = -\,\overline{\boldsymbol{\tau\beta}} \tag{1.83}$$

$$\bar{\boldsymbol{\tau}}(\boldsymbol{x}, t_f) = \bar{\boldsymbol{0}} \tag{1.84}$$

where $\boldsymbol{L_\tau}(\bar{\tau})$ is the adjoint state solute transport operator given by Eq. (1.26).

Eqs.(1.82) - (1.84) show that the adjoint states of a reactive system with multicomponent cation exchange transport satisfy similar equations as those of the original problem, with additional sink/source terms which depend on the derivatives of the objective function.

1.9 Conclusions

The adjoint state method can be applied: (a) to the continuous version of the original problem (continuous AS method) or (b) to the discretized form of the original problem (discrete AS method). The continuous AS method consists of deriving the AS equations from the partial differential equations (PDEs) of the original problem and then solving numerically the resulting PDEs of the AS. The discrete adjoint state method consists of obtaining the finite element equations of the AS directly from the discretized equations of the original problem. We have presented the formulation of both the continuous and discrete AS for conservative solute transport in porous media. The methods have been presented and the properties of both continuous and discrete AS formulations have been analysed and compared. The AS equations for multicomponent reactive systems with mineral dissolution/precipitation and cation exchange reactions have been also presented and discussed. The work done on adjoint-state methods has included: 1) Development of the equations for the continuous and discretized flow and transport equations for conservative species; 2) Development of the continuous and discrete adjoint state equations; 3) Development of the equations for the continuous and discretized transport equations for weakly-reactive species; 4) Development of the adjoint state equations for the continuous reactive transport equations; 5) Application of AS to Case 1 of reactive system with mineral phases at equilibrium or kinetically controlled; 6) Application of AS to Case 2 of reactive system with multicomponent cation exchange reactions; 7) Development of the adjoint state equations for the discretized reactive transport

equations and development of the equations for its implementation into CORE^{2D} V5; and 8) Application of AS to a test case of 1D solute transport through a column.

2. SENSITIVITY ANALYSES AND UNCERTAINTY QUANTIFICATION

This chapter presents the work done on sensitivity analysis and uncertainty quantification methods, which includes: 1) Compilation of applications; 2) Use of global sensitivity methods such as Sobol, Morris and VARS ("Variogram Analysis of Response Surfaces") methods; 3) Select and develop metamodels for the long-term geochemical evolution of a disposal cell in granite; 4) Apply all the methods to the long-term geochemical evolution of a disposal cell in granite.

2.1 Compilation of global sensitivity and uncertainty quantification methods

2.1.1 Morris and Sobol methods

The Morris elementary effects quantify the mean absolute value of the changes in the model output, Y, due to the perturbations in the model input parameters, $\boldsymbol{x_i}$. Morris effects are useful to identify the relevant and irrelevant input parameters. The Morris method provides global sensitivity measures because the mean changes in Y are computed at a large number of parameter combinations evenly distributed in the parameter domain.

The Sobol indexes quantify the contributions of parameters $\boldsymbol{x_i}$ to the variance of the output Y. The total Sobol index, S_{T_i}, quantifies the total contribution of $\boldsymbol{x_i}$ to the variance of Y:

$$\boldsymbol{S_{Ti} = S_i + S_{ij} + \cdots + S_{ij\ldots n}} \quad (2.1)$$

Similar to the Morris elementary effects, the Sobol indexes allow the identification of relevant and irrelevant parameters. The Sobol method is more powerful than the Morris method because the Sobol index quantifies the relative importance of each parameter and the parameter interdependencies. Morris and Sobol methods typically require thousands of simulations to compute meaningful results.

2.1.2 The VARS method

The variogram analysis of response surfaces (VARS) is a method which provides a more comprehensive and efficient tool to analyze the structure of a response function Y (see https://www.youtube.com/watch?v=YxeMdA8QIfU). Similar to the Morris and Sobol methods, the VARS method requires performing Monte Carlo simulations of the output, Y. The VARS method estimates the directional variograms for each one of the input parameters (Razavi and Gupta, 2016). The Morris and the Sobol methods are particular cases of the VARS method. The key advantage of VARS is its efficiency: the number of required Monte Carlo simulations can be up to two orders of magnitude smaller than those required by the Morris or Sobol methods. The results of a benchmark of the performance of eight different Monte Carlo estimators of the total Sobol index show that the Razavi and Gupta's VARS method is the preferred estimator, especially when there are stringent limitations on the total number of model runs (Puy et al., 2020).

Input parameters x_i are treated as random variables which are normalized into uniform distributions in the interval [0, 1]. Then, each axis is discretized with constant parameter increments Δh (resolution). Latin hypercube sampling (LHS) is performed at the so-called "star points". The response function Y is evaluated at all the discrete points along lines parallel to the axes (see *Figure* 1). Directional variograms of Y are calculated along each of the lines from the values of (Y, x_i). VARS interpolates the experimental variogram with smooth functions.

The variograms γ_i along the lines hold all the required answers. The limit of the variogram, γ_i, divided by h^2 when h tends to 0 is similar to the Morris effects (Razavi and Gupta, 2016):

$$\lim_{h\to 0}\frac{\gamma_i(h)}{h^2} = E\left(\frac{\partial y}{\partial h}\right)^2 = \frac{1}{2}\frac{d^2\gamma_i}{dh^2}\bigg|_{h=0} \tag{2.2}$$

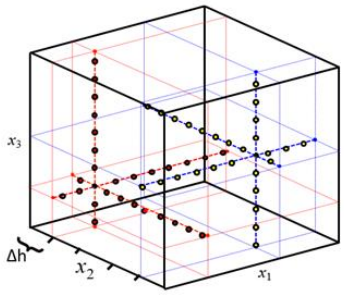


Figure 1 – Scheme of the sampling scheme of the VARS method.

On the other hand, the limit of the variogram γ_i for large h is the total variance which is linked to the Sobol sensitivity index, S_{T_i}, through (Razavi and Gupta, 2016):

$$\lim_{h\to 0}\gamma_i(h) = \boldsymbol{S_{Ti}} - E\left[C_{\chi_{\sim i}}(h)\right] \tag{2.3}$$

The VARS method estimates the directional variograms by interpolating the experimental variogram with smooth functions. This method, however, does not ensure non-negativity of the computed variances. To produce a valid variogram, we propose to use non-parametric variograms estimates such as those reported by Hayek et al. (2019) and implemented in R by Fernández-Casal (2016). The non-parametric variograms provide an extremely large range of valid isotropic covariance functions c(h). Any covariance c(h) can be expressed as:

$$c(h) = \sum_{j=1}^{\infty} p_j \Omega_d(t_j h) \tag{2.4}$$

where d is the number of parameters, p_j are unknown coefficients, t_j are preselected coefficients, and $\Omega_d(u)$ is given by:

$$\Omega_d(u) = \Gamma\left(d/2\right) J_{\frac{d-2}{2}}(u)\left(\frac{2}{u}\right)^{\frac{d-2}{2}} \tag{2.5}$$

where Γ is the Gamma function and J_v is a Bessel function of the first kind.

Figure 11 shows the variogram of the response surface and the links with the Morris elementary effects and the total-order Sobol index.

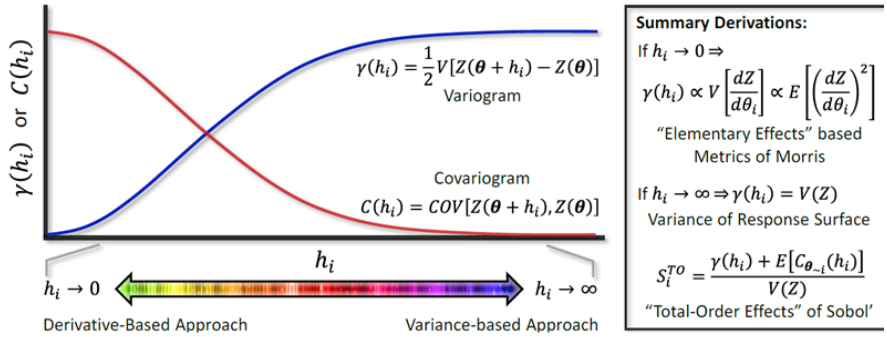


Figure 11 – Illustration of the variogram of the response surface and the links with the Morris elementary effects and the total-order Sobol index. Here the upper limit of hi is 1 because parameters are normalized into the [0.1] interval.

2.2 Application of global sensitivity methods to the long-term geochemical evolution in a HLW repository in granite

2.2.1 Numerical model of canister corrosion and bentonite interactions

The global sensitivity methods have been applied to 1D axi-symmetric reactive transport models of the long-term geochemical evolution of a HLW repository in granite.

The HLW disposal cell concept in granitic host rock includes: (1) the vitrified waste (40 cm in diameter) which is explicitly considered in the reactive transport model and is coupled with the rest of the engineered barrier components; (2) the canister, which is made of carbon-steel of 5 cm thick; (3) the bentonite buffer (75 cm thick) which is composed of water-saturated FEBEX bentonite with a dry density of 1600 kg/m3; and (4) the Spanish Reference Granitic (SRG) host rock (25 m) (*Figure* 3). This HLW disposal cell concept in granite is not specific of any country but aims at representing several national concepts. This case is inspired on the Spanish spent fuel reference concept in granite, known as ENRESA 2000 (ENRESA, 2000). The main differences between these two concepts are the type of waste (vitrified waste vs. spent fuel) and the dimensions of the carbon-steel canister (40 cm in diameter for vitrified waste and 90 cm for spent fuel).

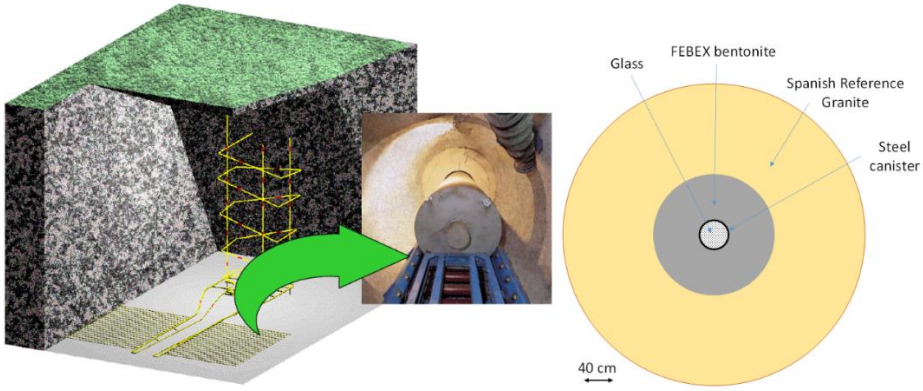


Figure 3 – Underground installations of a radioactive waste repository in granite according to the Spanish Reference Concept for spent fuel (ENRESA, 2000). Layout of the representative HLW disposal cell concept in a granitic host rock selected in the ACED Work Package of the EURAD Project (Samper et al., 2022; Montenegro et al., 2023).

2.2.2 Conceptual THC model

The non-isothermal multicomponent reactive transport model of the HLW disposal cell in granite involves the canister and the bentonite. The reactive transport model assumes that the bentonite is initially water

saturated. The conceptual geochemical model includes: 1) Carbon-steel canister corrosion, 2) Aqueous complexation; 3) Acid/base; 4) Redox; 5) Mineral dissolution/precipitation; 6) Cation exchange of Ca^{2+}, Mg^{2+}, Fe^{2+}, Na^{+} and K^{+}; and 7) Surface complexation of H^{+} and Fe^{2+} on strong sites, $S^{s}OH$, weak #1 sites, $S^{w1}OH$ and weak #2 sites, $S^{w2}OH$. Cation exchange and surface complexation reactions occur only in the bentonite (Montenegro et al., 2023).

The geochemical system is defined in terms of H_2O, H^{+}, $O_2(aq)$, Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Fe^{2+}, Al^{3+}, Cl^{-}, SO_4^{2}, HCO_3^{-}, 39 secondary aqueous species, 7 minerals, 5 exchanged cations and 13 surface complexes. The Gaines-Thomas convention is used for cation exchange reactions (Appelo and Postma, 1993). Surface complexation reactions in the bentonite were modelled with the non-electrostatic triple-site sorption model of Bradbury and Baeyens (1997, 2005).

All the reactions are assumed to be at chemical equilibrium, except for carbon-steel corrosion. The available oxygen in the HLW disposal cell in granite is consumed soon after its closure, and anaerobic conditions are expected to prevail in the long term. The canister is treated as a porous material made of metallic iron, Fe(s). Under anaerobic conditions, H_2O is the oxidizing agent of Fe(s) (Lu et al., 2011; Samper et al., 2016; Mon et al., 2017; Montenegro et al., 2023) according to:

$$Fe(s) + 2H^{+} + 0.5O_2(aq) \Leftrightarrow Fe^{2+} + H_2O \qquad (2.6)$$

Carbon-steel corrosion is kinetically controlled. The base run assumes a constant corrosion rate of 1.41 μm/year (Samper et al., 2016; Mon et al., 2017).

2.2.3 Numerical model

The model was performed with a non-uniform 1-D axisymmetric finite element mesh with 152 nodes (*Figure* 4). Water flow through the granite is simulated with a constant water flux Q parallel to the axis of the gallery. The initial accessory mineral volume fractions in the bentonite include calcite (1%) and smectite (57%). The initial mineral volume fraction of calcite in the granite is 5%. The initial volume fraction of Fe(s) in the canister is 0.99%. The cation exchange capacity (CEC) of the bentonite is 102 meq/100 g (Fernández et al., 2004). Cation selectivity coefficients for exchanged Ca^{2+}, Mg^{2+}, K^{+}, Na^{+} and Fe^{2+} were taken from Samper et al. (2008) and Tournassat et al. (2008). Surface complexation reactions in the bentonite are modelled with the triple sorption site model of Bradbury and Baeyens (1997, 2005). The total concentration of sorption sites is 0.322 mol/L. Model simulations were performed with the finite element reactive transport code CORE^{2D} V5 (Molinero et al., 2004; Samper et al., 2009; Fernández, 2017; Águila et al., 2020).

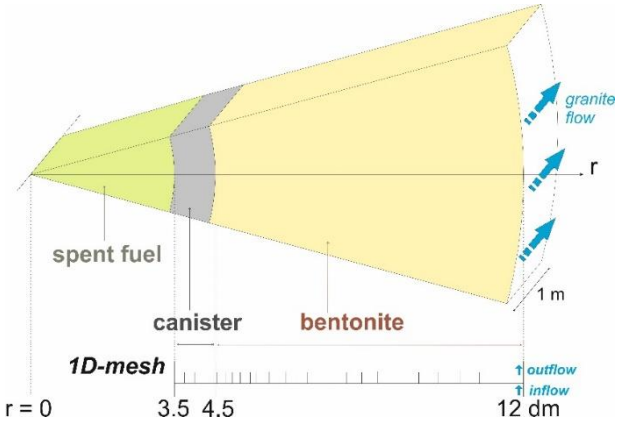


Figure 4 – 1D axi-symmetric section of a HLW repository in granite (Samper et al. 2016).

2.2.4 Model results

Canister corrosion causes an increase in pH in the canister and the bentonite near the bentonite interface. The time evolution of the computed pH and the concentration of dissolved Fe are controlled by: 1) The dissolution-precipitation of Fe minerals in the canister and in the bentonite near the canister interface, 2) Fe exchange and 3) Fe and proton surface complexation reactions.

Magnetite precipitation takes place mostly in the canister and proceeded as long as the canister is corroding. Siderite precipitates at both sides of the canister/bentonite interface. Goethite does not precipitate. Calcite precipitates in the canister near the bentonite interface, in the bentonite and in the granite. Gypsum does not precipitate.

The concentration of Cl^- decreases with time in the bentonite due solute diffusion from the bentonite into the granite because the initial Cl^- concentration in the granite porewater ($3.95 \cdot 10^{-4}$ M) is much smaller than the initial concentration in the bentonite ($2.75 \cdot 10^{-1}$ M). The profiles of the concentration of Cl^- show very small gradients because the solute transfer from the bentonite into the granite is not controlled by solute diffusion through the bentonite, but by solute advection and diffusion in the granite. The computed concentrations of dissolved Ca, Mg, Na and K decrease with time due to solute diffusion into the granite. See numerical model details in Montenegro et al (2023).

2.3 VARS simulations for a HLW repository in granite

2.3.1 Implementation of VARS

The variables considered in the analysis include the pH, the redox potential and the volume of corrosion products at the interface of the container with the bentonite for different times (10,000, 25,000 and 50,000 years). The parameters considered include: the corrosion rate, the diffusion coefficient of the solutes in the bentonite, the flow of groundwater through the granite, the cation exchange selectivity of Fe and the stability constant of the magnetite. Log-uniform probability density functions were used for the parameters, except for the corrosion rate, which was uniform (Table 1). The number of simulations required for d = 5 parameters and Δh = 0.1 is equal to 46 per star point. The number of star points which are needed to get reliable estimates of the variograms is the matter of on-going work.

The selected parameters are listed in *Table* . The following options were considered:

1) pH evolution at the canister/bentonite and bentonite/granite interfaces.
2) Redox potential at the canister/bentonite and bentonite/granite interfaces.
3) Volume of corrosion products at the canister/bentonite and bentonite/granite interfaces.

Table 2 – Ranges and statistical distributions of the selected parameters.

Parameter	**Minimum**	**Maximum**	Unit	**Distribution**
Corrosion rate (CR)	0.1	10	µm/y	uniform
Effective solute diffusion in the bentonite (D_e)	0.0631	0.631	dm^2/year	log-uniform
Groundwater flow through granite (Q_{gra})	0.01	0.1	L/y	log-uniform
Fe exchange selectivity (K_{Fe})	0.001	1	-	log-uniform
Log magnetite solubility constant (log K)	-8.56	-4.56	-	uniform

More than 27600 sensitivity runs were performed with the numerical code CORE2D V5 in the Galician Supercomputing Center [CESGA, Centro de Supercomputación de Galicia (in Spanish)]. The selection of the computing center was done on the basis of making thousands of simultaneous/parallel model runs, together with a very large data storage capacity.

The simulation runs took 6 months of computing time in CESGA. The effort would have taken more than 50 years running in a personal computer.

2.4 VARS results

2.4.1 Scatterplots

Figure and *Figure* show the scatterplots of the computed pH values versus the 5 parameters (CR, D_e, Q_{gra}, K_{Fe}, logK). Each plot also contains linear regression results (red dotted line and corresponding values of R^2 determination coefficient).

The determination coefficients, R^2, are rather small, except for the corrosion rate (Figure 5). The pH seems to decrease with increasing corrosion rate. The scatterplot of pH versus groundwater flow (Q_{gra}) shows that there are two branches with a bifurcation.

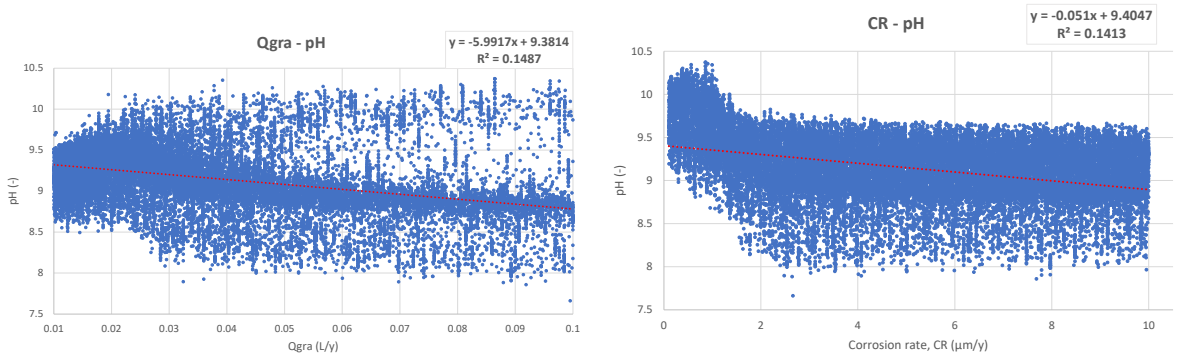


Figure 5 – Scatterplot of the computed pH values versus groundwater flow (Qgra) ((left) and corrosion rate (CR) (right) in the granite for large time increments. Also shown the regression line and the determination coefficient R^2.

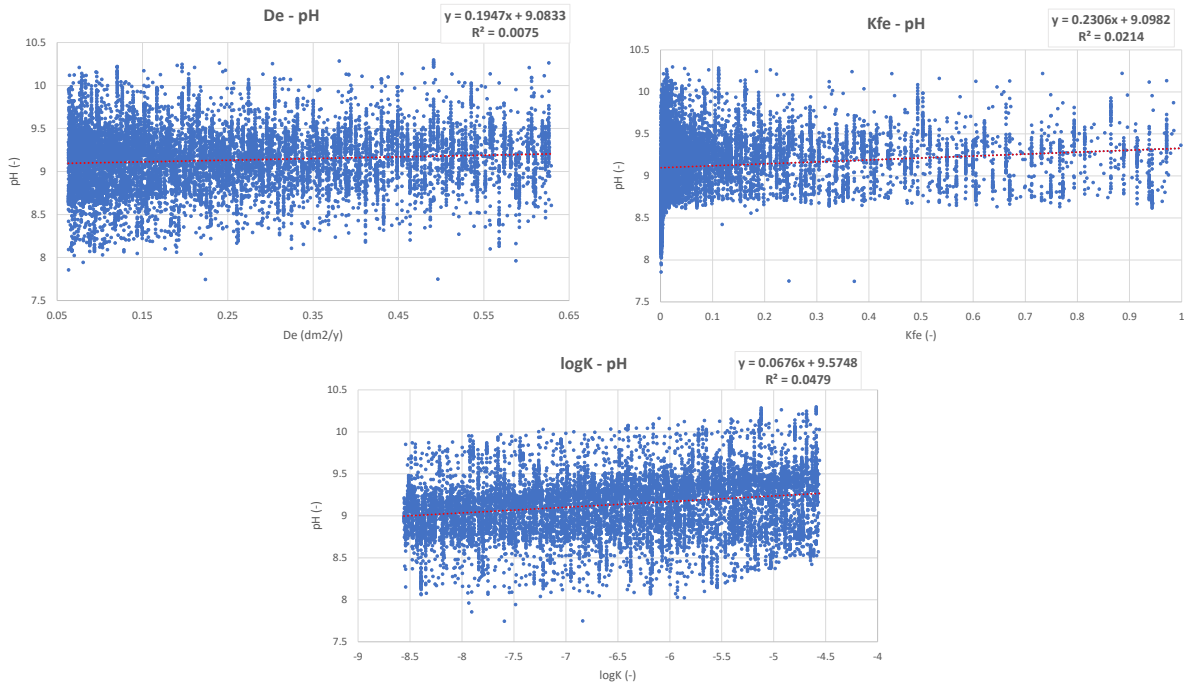


Figure 6 – Scatterplot of the computed pH values versus the effective diffusion coefficient (left), Fe exchange selectivity of bentonite (right) and magnetite log-solubility constant (down). Also shown the regression line and the determination coefficient R^2

2.4.2 VARS indexes

Figure 7 shows the IVARS50 indexes for the selected parameters as a function of the number of stars for the computed pH at 50,000 years. It should be noticed that the indexes achieve stable values after 100 stars. The largest index is associated with the corrosion rate. The second largest index corresponds to the flow in the granite. The thirst largest index is associated with the selectivity of Fe. The other two parameters, the bentonite diffusion coefficient, and the Log K of magnetite, have smaller indexes.

Figure 8 shows the robustness of rankings as a function of the number of stars, N. It should be noticed that the robustness measure is greater than 90% after 100star centers. This means that for N > 100 stars, with a 90% confidence value, the ranking of the parameters is the ranking indicated by VARS results.

Figure 9 shows the sample directional variograms of pH at 50,000 years for the five parameters. There are three main parameters which exhibit larger variograms. These parameters are the corrosion rate, the flow in the granite and selectivity of Fe. These sample variograms hold all the required answers. The IVARS-0, IVARS-50 and IVARS-100 are computed from these variograms.

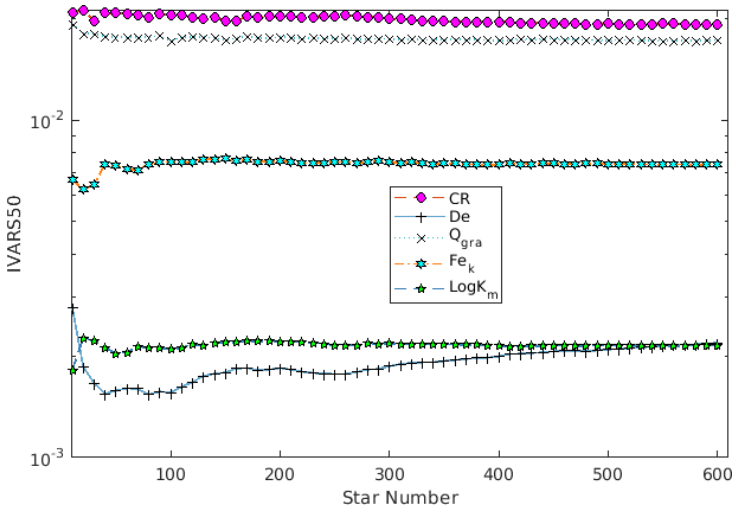


Figure 7 – IVARS50 indexes for the selected parameters as a function of the number of stars for the computed pH at 50,000 years. It should be noticed that the indexes achieve stable values after 100 stars.

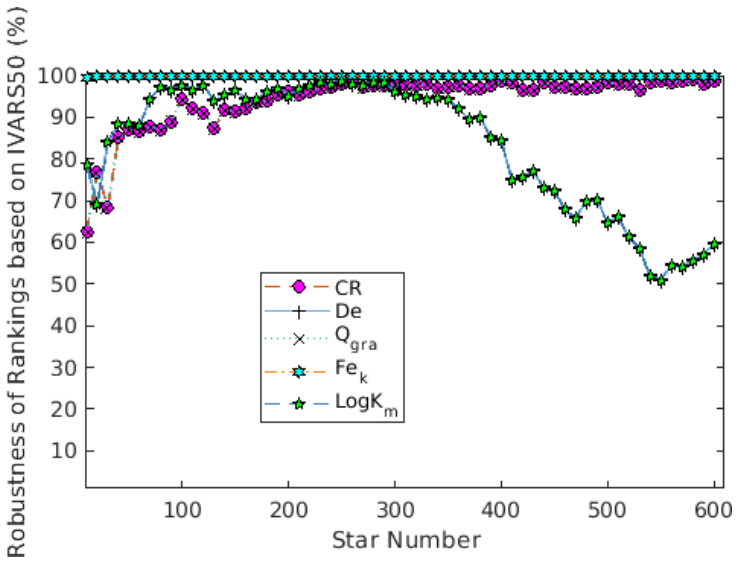


Figure 8 – Robustness of rankings as a function of the number of stars. It should be noticed that the robustness measure is greater than 90% after 100 star centers.

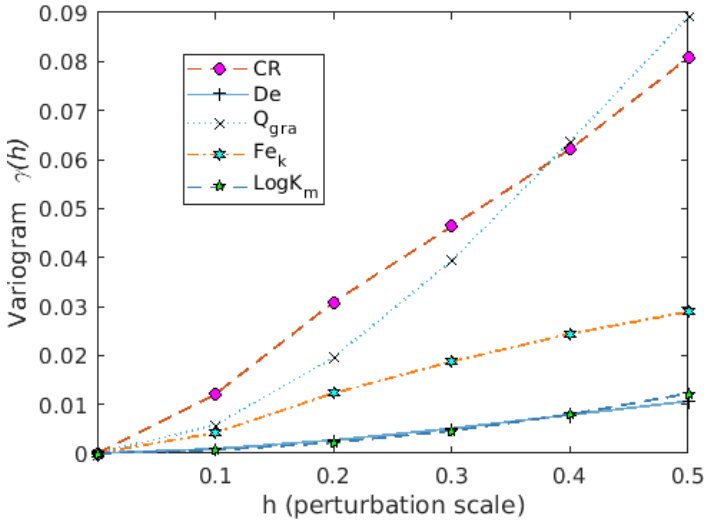


Figure 9 – Sample variograms of the output (pH) for the five parameters.

Figure 10 shows the values of the IVARS-0 (IVARS-ABE or Morris), IVARS-50 and IVARS-100 (Sobol) indexes for pH at t = 10^4, $2{\cdot}10^4$ y $5{\cdot}10^4$ years. The sensitivity indexes of the corrosion rate and the Fe selectivity for pH increase with time while that of the log K magnetite decreases. The sensitivity index of the bentonite diffusion coefficient, however, remains constant in time.

Figure shows the values of the IVARS-0 (IVARS-ABE or Morris), IVARS-50 and IVARS-100 (Sobol) indexes for Eh at t = 10^4, $2{\cdot}10^4$ y $5{\cdot}10^4$ years. Similar to pH, the most influential parameters for Eh are the corrosion rate and the flow in the granite. The sensitivity indexes of the corrosion rate and the log K of magnetite for Eh decrease with time while that of groundwater flow increases with time.

Figure 12 shows the values of the IVARS-0 (IVARS-ABE or Morris), IVARS-50 and IVARS-100 (Sobol) indexes for the volume of corrosion products at t = 10^4, $2{\cdot}10^4$ y $5{\cdot}10^4$ years. Similar to pH and Eh, the most influential parameter for the volume of corrosion products is the corrosion rate while the second is the diffusion coefficient of the bentonite. All the sensitivity indexes for the corrosion products increase with time except for the corrosion rate.

The indexes corresponding to Morris, Sobol and VARS (IVARS-50) show some differences. These differences are largest for the Morris indexes. However, the three methods lead to similar results in terms of rankings and significance. Morris index (IVARS-ABE) tends to reach larger values than those of the IVARS-50 and Sobol indexes for the less influential parameters.

The VARS analysis has been performed for a time discretization scheme with larger time steps and large tolerances for the convergence criterion of the sequential iteration scheme. VARS results are slightly sensitive to the temporal discretization scheme (not shown here).

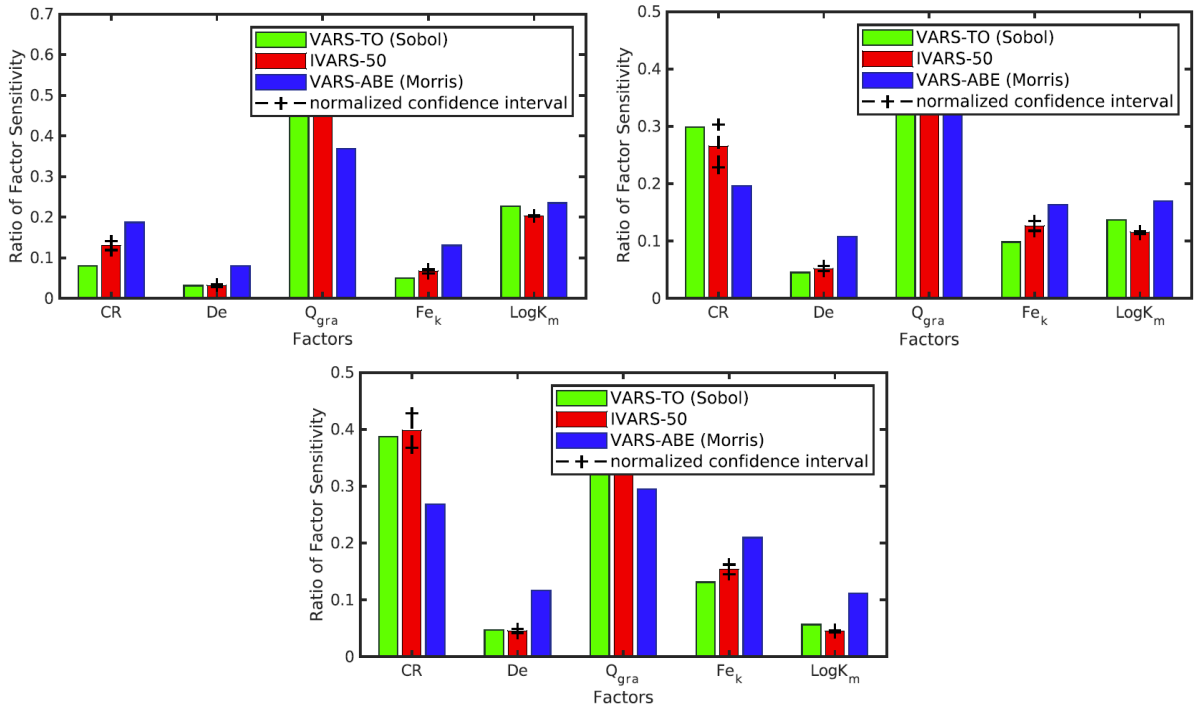


Figure 10 – Values of the IVARS-0, IVARS-50 and IVARS-100 indexes for the computed pH at 10,000, 25,000 and 50,000 years.

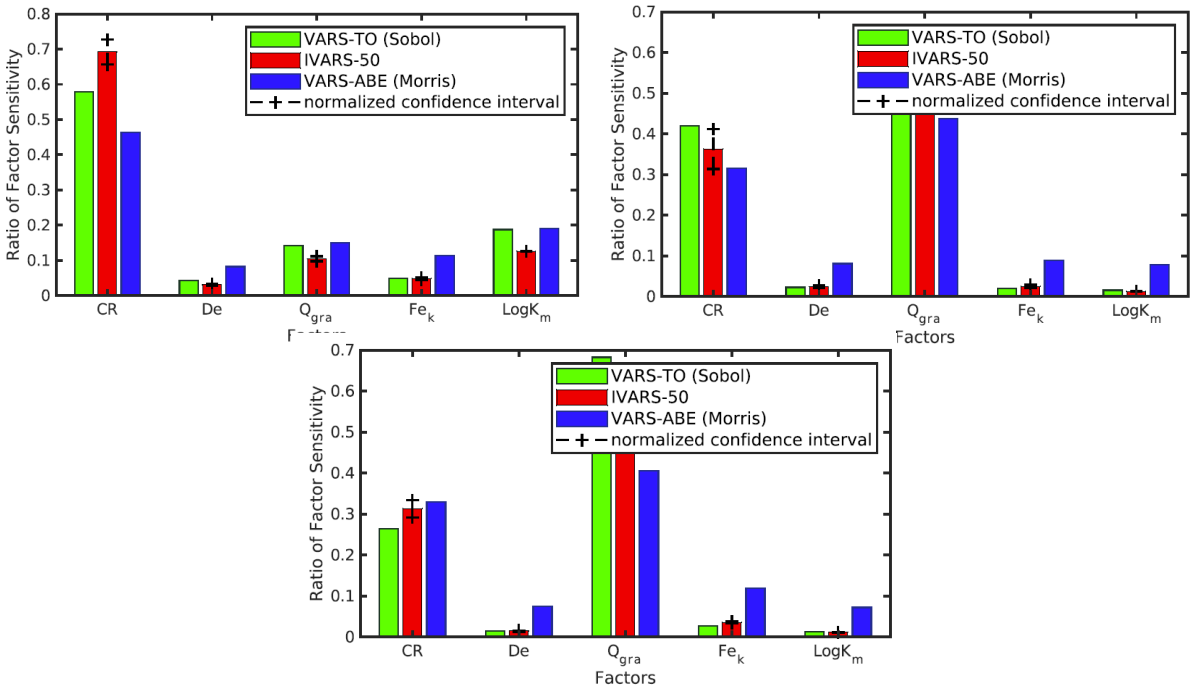


Figure 11 – Values of the IVARS-0, IVARS-50 and IVARS-100 indexes for the computed Eh at 10,000, 25,000 and 50,000 years.

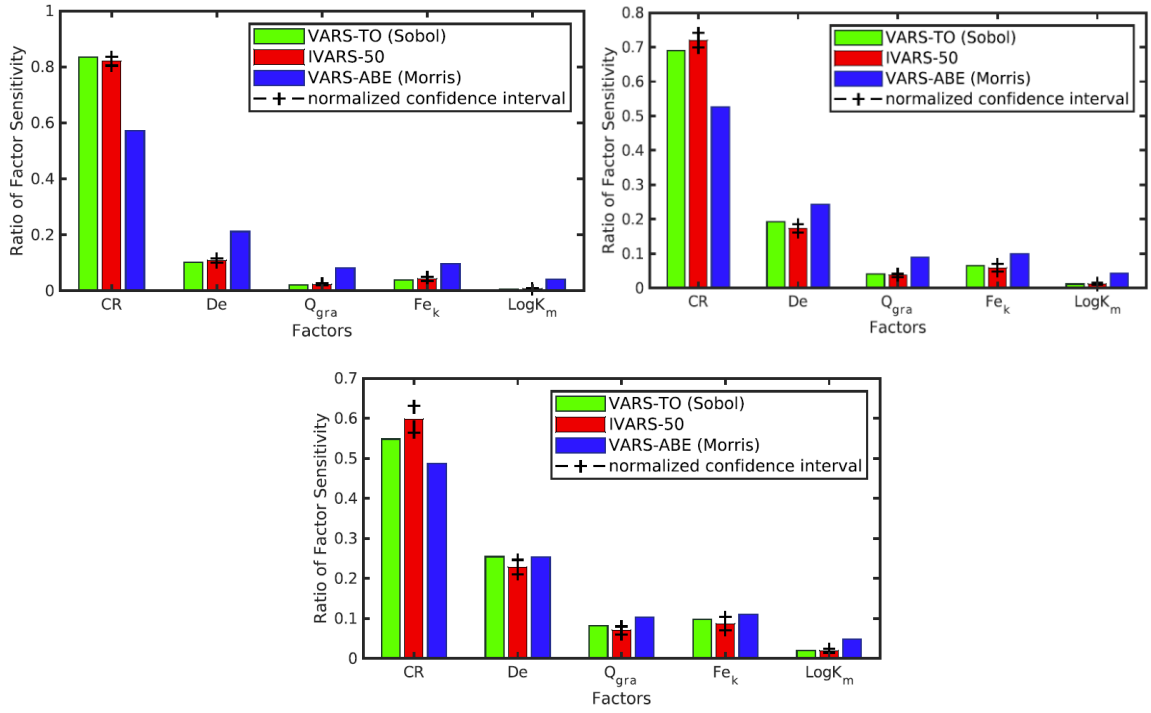


Figure 12 – Values of the IVARS-0, IVARS-50 and IVARS-100 indexes for the computed corrosion products at 10,000, 25,000 and 50,000 years.

2.5 Conclusions

The results of the global sensitivity analysis lead to the conclusion that the most influential parameters for pH are the corrosion rate, the flow in the granite and the cationic selectivity of Fe. The main interactions between parameters for pH occur between the corrosion rate and the groundwater flow through granite. The sensitivity indexes of the corrosion rate and the Fe selectivity for pH increase with time while that of the log K magnetite decreases. The sensitivity index of the bentonite diffusion coefficient, however, remains constant in time.

Similar to pH, the most influential parameters for Eh are the corrosion rate and the flow in the granite. The sensitivity indexes of the corrosion rate and the log K of magnetite for Eh decrease with time while that of groundwater flow increases with time.

Similar to pH and Eh, the most influential parameter for the volume of corrosion products is the corrosion rate while the second is the diffusion coefficient of the bentonite. All the sensitivity indexes for the corrosion products increase with time except for the corrosion rate.

The indexes corresponding to Morris, Sobol and VARS (IVARS-50) show some differences. These differences are largest for the Morris indexes. However, the three methods lead to similar results in terms of rankings and significance. Morris index (IVARS-ABE) tends to reach larger values than those of the IVARS-50 and Sobol indexes for the less influential parameters.

The VARS method estimates the directional variograms from the experimental variograms. This method, however, does not ensure non-negativity of the computed variances. Therefore, to improve the sensitivity analysis we propose to use non-parametric variograms estimates such as those reported by Gorsich and Genton (2004) and implemented in R by Fernández-Casal (2016) The non-parametric variograms provide the full range of valid isotropic covariance functions c(h).

3. CONCLUSIONS

This report has presented the work done so far by UDC on Task 4 of DONUT WP, which deals with innovative numerical methods for adjoint-state methods and sensitivity analyses and uncertainty analysis.

3.1 Adjoint state methods

The adjoint state method can be applied: (a) to the continuous version of the original problem (continuous AS method) or (b) to the discretized form of the original problem (discrete AS method). The continuous AS method consists of deriving the AS equations from the partial differential equations (PDEs) of the original problem and then solving numerically the resulting PDEs of the AS. The discrete adjoint state method consists of obtaining the finite element equations of the AS directly from the discretized equations of the original problem.

We have presented the formulation of both the continuous and discrete AS for conservative solute transport in porous media. The methods have been presented and the properties of both continuous and discrete AS formulations have been analysed and compared. The AS equations for multicomponent reactive systems with mineral dissolution/precipitation and cation exchange reactions have been also presented and discussed.

The detailed work on adjoint-state methods has included: 1) The development of the equations for the continuous and discretized transport equations for conservative species; 2) The development of the continuous and discrete adjoint state equations; 3) The development of the equations for the continuous and discretized transport equations for weakly-reactive species; 4) The development of the adjoint state equations for the continuous reactive transport equations; 5) The application of AS to a reactive system with mineral phases at equilibrium or kinetically controlled; 6) The application of AS to a reactive system with multicomponent cation exchange reactions; 7) The application of AS to a reactive system with linearized cation exchange reactions; 8) The development of the adjoint state equations for the discretized reactive transport equations; 9) The implementation of AS equations into CORE^{2D} V5; and 10) The application of AS to test cases.

3.2 Sensitivity analyses and uncertainty quantification

The results of the global sensitivity analysis lead to the conclusion that the most influential parameters for pH are the corrosion rate, the flow in the granite and the cationic selectivity of Fe. The main interactions between parameters for pH occur between the corrosion rate and the groundwater flow through granite. The sensitivity indexes of the corrosion rate and the Fe selectivity for pH increase with time while that of the log K magnetite decreases. The sensitivity index of the bentonite diffusion coefficient, however, remains constant in time.

Similar to pH, the most influential parameters for Eh are the corrosion rate and the flow in the granite. The sensitivity indexes of the corrosion rate and the log K of magnetite for Eh decrease with time while that of groundwater flow increases with time.

Similar to pH and Eh, the most influential parameter for the volume of corrosion products is the corrosion rate while the second is the diffusion coefficient of the bentonite. All the sensitivity indexes for the corrosion products increase with time except for the corrosion rate.

The indexes corresponding to Morris, Sobol and VARS (IVARS-50) show some differences. These differences are largest for the Morris indexes. However, the three methods lead to similar results in terms of rankings and significance. Morris index (IVARS-ABE) tends to reach larger values than those of the IVARS-50 and Sobol indexes for the less influential parameters.

The VARS method estimates the directional variograms from the experimental variograms. This method, however, does not ensure non-negativity of the computed variances. Therefore, to improve the sensitivity analysis we propose to use non-parametric variograms estimates such as those reported by

Gorsich and Genton (2004) and implemented in R by Fernández-Casal (2016) The non-parametric variograms provide the full range of valid isotropic covariance functions c(h).

Code source

For DONUT Task 4: UDC/ENRESA has performed local and global sensitivity calculations with the following codes: CORE2D V5.1 and INVERSE-FADES-CORE V2. CORE2D V5.1 is a code for transient saturated and unsaturated water flow, heat transport and multicomponent reactive solute transport under both local chemical equilibrium and kinetic conditions in heterogeneous and anisotropic media. INVERSE-FADES-CORE V2 is a finite element code for modelling non-isothermal multiphase flow, heat transport and multicomponent reactive solute transport under both chemical equilibrium and kinetic conditions in deformable media. VARS simulations have been performed in the super-computers of the Galician Super-computing Center (CESGA). VARS analyses have been performed in the MATLAB version of VARS.

References

Águila, J.F., Samper, J., Mon, A., Montenegro, L., (2020). Dynamic update of flow and transport parameters in reactive transport simulations of radioactive waste repositories. Applied Geochemistry, Vol 117: 104585 https://doi.org/10.1016/j.apgeochem.2020.104585.

Appelo, C.A.J., Postma, D., (1993). Geochemistry, Groundwater and Pollution, A.A. Balkema, Brookfield, Vt.

Ackerer, P., Trottier, N., Delay, F., 2014. Flow in double porosity aquifers: parameter estimations using an adaptive multiscale method. Adv. Water Resour. 73, 108–122.

Bradbury, M.H., Baeyens, B., (1997). A mechanistic description of Ni and Zn sorption on Na-montmorillonite. Part II: Modelling. J. Contaminant Hydrology, 27, 223-248.

Bradbury, M.H., Baeyens, B., (2005). Modelling the sorption of Mn(II), Co(II), Ni(II), Zn(II), Cd(II), Eu(III), Am(III), Sn(IV), Th(IV), Np(V) and U(VI) on montmorillonite: linear free energy relationships and estimates of surface binding constants for some selected heavy metals and actinides. Geochim. Cosmochim. Acta 69, 875-892.

Carrera, J., S.P Neuman (1986). Estimation of aquifer parameters under transient and steady-state conditions. 2. Uniqueness, stability and solution algorithms. Water Resour. Res. 22 (2), 211–227.

Delay, F, H Badri, M Fahs, P Ackerer (2017), A comparison of discrete versus continuous adjoint states to invert groundwater flow in heterogeneous dual porosity systems, Ad. Water Resour. 110: 1-18.

ENRESA (2000) FEBEX Project. Full-scale engineered barriers experiment for a deep geological repository for high-level radioactive waste in crystalline host rock. Technical Report. ENRESA 1/2000, Madrid, 354 pp.

Fernández, A.M., Baeyens, B., Bradbury, M., Rivas, P., (2004). Analysis of the pore water chemical composition of a Spanish compacted bentonite used in an engineered barrier. Physics and Chemistry of the Earth 29(1), 105-118.

Fernández-Casal, R. (2016) npsp: nonparametric spatial (geo)statistics. R package version 0.5-3 https://rubenfcasal.github.io/npsp/index.html

Fernández, J., (2017). Reactive Transport Models of low Permeability Structured Porous and Fractured Media. Ph.D. Dissertation. Universidad de A Coruña, Spain.

Gorsich, D. J. and Genton, M. G. (2004). On the discretization of nonparametric isotropic covariogram estimators. Statistics and Computing 14: 99–108.

Hayek, M., RamaRao, B. S., & Lavenue, M. (2019). An adjoint sensitivity model for steady-state sequentially coupled radionuclide transport in porous media. Water Resources Research, 55, 8800– 8820. https://doi.org/10.1029/2019WR025686

Hayek, M., RamaRao, B.S., and Lavenue, M. (2020). An adjoint sensitivity model for transient sequentially coupled radionuclide transport in porous media. Water Resources Research, 56 (7), e2020WR027274. https://doi.org/10.1029/2020WR027274

Lu, C., Samper, J., Fritz, B., Clement, A., Montenegro, L., (2011). Interactions of corrosion products and bentonite: An extended multicomponent reactive transport model, Physics and Chemistry of the Earth 36, 1661-1668, doi: 10.1016/j.pce.2011.07.013.

Molinero, J. J. Samper, G. Zhang & C. Yang (2004). Biogeochemical reactive transport model of the redox zone experiment of the Äspö hard rock laboratory in Sweden, Nuclear Technology, 148. 151-165.

Mon, A., Samper, J., Montenegro, L., Naves, A., Fernández, J., (2017). Reactive transport model of compacted bentonite, concrete and corrosion products in a HLW repository in clay. Journal of Contaminant Hydrology 197, 1-16.

Montenegro, L, J Samper, A Mon, L De Windt, AC Samper & E García 2023, Reactive transport models of the geochemical evolution in a disposal cell at a HLW repository in granite, Applied Clay Science, https://doi.org/10.1016/j.clay.2023.107018.

Puy, A.; Becker, W.; Lo Piano, S. and Saltelli, A. (2020). The battle of total-order sensitivity estimators. arXiv:2009.01147

Razavi, S., Gupta, H. V., 2016. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. Water Resour. Res., 52, 440 – 455.

Samper, F. J., Statistical methods of analyzing hydrological, hydrochemical and isotopic data from aquifers, Ph.D. dissertation, Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson, 1986

Samper, F.J., S.P. Neuman (1986) Adjoint state equations for advective-dispersive transport, In: VI International Conference on Finite Elements in Water Resources, Lisbon, 423-437.

Samper F.J., S.P. Neuman (1989) Estimation of Spatial Covariance Structures by Adjoint State Maximum Likelihood Cross-validation: 1. Theory, Wat Resour Res,25(3):351-362.

Samper, J., R. Juncosa, J Delgado, L Montenegro, 2000. CORE-2D. A code for non-isothermal water flow and reactive solute transport. Users manual version 2. Publicación Técnica de ENRESA 6/2000. Madrid. 131 pp.

Samper-Calvete, J., Yang, C. (2007). A semi-analytical solution for linearized multicomponent cation exchange reactive transport in groundwater. Transp Porous Med 69, 67–88. https://doi.org/10.1007/s11242-006-9065-4

Samper, J., Lu, C., Montenegro, L., (2008). Coupled hydrogeochemical calculations of the interactions of corrosion products and bentonite. Physics and Chemistry of the Earth 33, S306-S316, doi:10.1016/j.pce.2008.10.009.

Samper, J., Xu, T., Yang, C., (2009). A sequential partly iterative approach for multicomponent reactive transport CORE2D. Comput. Geosci. 13, 301-316.

Samper J, A Naves, L Montenegro, A Mon (2016) Reactive transport modelling of the long-term interactions of corrosion products and compacted bentonite in a HLW repository in granite: Uncertainties and relevance for performance assessment, Applied Geochemistry. Vol 67: 42-51.

Samper, J., L. Montenegro, L. De Windt, V. Montoya, J. Garibay-Rodríguez, D. Grigaliuniene, A. Narkuniene, Poskas, P., (2022): Conceptual model formulation for a mechanistic based model implementing the initial SOTA knowledge (models and parameters) in existing numerical tools. Deliverable D2.16 of the EURAD Joint Project. EC Grant Agreement no: 847593.

Sun, N-Z., 1994. Inverse problems in groundwater modeling. Kluwer Acad. Publishers, Dordrecht, The Netherlands.

Tournassat, C, C Lerouge, P Blanc, J Brendlé, JM Greneche, S Touzelet, E C. Gaucher (2008). Cation exchanged Fe(II) and Sr compared to other divalent cations (Ca, Mg) in the Bure Callovian–Oxfordian formation: Implications for porewater composition modelling, Applied Geochemistry, 23 (4): 641-654.

Townley, L.R., Wilson, 1985. Computationally efficient algorithms for parameter estimation and uncertainty propagation in numerical models of groundwater flow. Water Resour. Res. 21 (12), 1851–1860.

Feasibility of Neural Network Surrogate Models for Sensitivity Analysis of Radionuclide Transport and Geochemical Processes

Lipping, T., Turunen, J., Pohjola. J.

Tampere University, Finland

Abstract

In this report we first present a framework of using sensitivity analysis to investigate and validate the models used in planning and safety assessment of geological repositories of spent nuclear fuel. Variance based methods such as Sobol and RBDFast have been selected. Also, the feasibility of using surrogate models for sensitivity analysis is considered. Case studies involving three different models – a lake-farm-forest compartmental model of the biosphere at a geological repository, a differential equation model of the propagation of radioactive substance in sand, and a geochemical model implemented using the Orchestra solver – are then considered in more detail. The significance of the parameters of the compartmental model differs for different radionuclides, however, in general, the properties of the soil, the properties of the lake sediment and the concentration ratio from soil to plants were found to affect the dose conversion factor most. The relative importance of the lake, agricultural and forest ecosystems depend on the use of the lake water for irrigation. In the case study involving radionuclide propagation, the focus was on the validation of Convolutional Neural Network based surrogate model while using the advection-dispersion equation model as a reference. It was found that the sensitivity analysis results largely depend on the range of input parameter values selected for sampling the parameter space as well as if the first or total order analysis was used. The third case study focused on the sensitivity analysis of geochemical models and validation of Neural Network based surrogate models for sensitivity analysis. Two models of different complexity were chosen. The study shows that the surrogate model gives similar sensitivity index values compared to the geochemical solver when trained using a data set of 5000 realizations. However, the accuracy of the surrogate model and the consistence of the surrogate model with the solver is further improved when the training sample size is increased to 50 000. The Sobol first and total order sensitivity indices behave in a coherent way as predicted by their theoretical background, assuring the reliability of the results.

Significance Statement

In complex problems such as the assessment of the safety of geological repositories of spent nuclear fuel, only very limited phenomena can be studied experimentally and, therefore, various models are essential. Models are just an approximation of the reality and the inputs as well as internal parameters of the models always involve uncertainties. These uncertainties may come from inaccuracies in measurements, limited knowledge about the modelled processes and/or the changes in the environment the modelled processes take place in. Therefore, investigating the accuracy of the results given by the models is not enough but the sensitivity of model outputs on model parameters should also be considered. Among the various techniques introduced for sensitivity analysis, those based on the propagation of variance from inputs / model parameters to the outputs have proven useful in many fields of science. The methods are data-driven, i.e., the analysis is based on a set of input-output data calculated using the model. In the case studies presented in this report, the behaviour of three different types of models and their surrogate counterparts have been investigated from the perspective of sensitivity analysis. The case studies indicate the usefulness but also point out some limitations of using sensitivity analysis in model validation.

Table of content

List of figures

List of Tables

1. Sensitivity analysis: basic principles and methods

Sensitivity analysis investigates the effect of the uncertainty or variability in the values of the model parameters on the model output. Swiler et al., (2021) recognize the following classes of methods that can be used for sensitivity analysis:

- correlation analysis,
- regression,
- variance-based measures,
- moment-independent sensitivity measures,
- graphical analysis,
- derivative-based measures.

More detailed discussion of the listed methods is out of the scope of this report, however, when defined in terms of variability, probably the most natural way to perform sensitivity analysis is to study the relation between the variance of the input parameters and the model output. Except in trivial cases, no tools exist to express the relation between variances in an analytical form. Therefore, the parameter space is usually sampled in a certain way to form a set of input vectors and the respective model outputs are calculated. These input-output pairs then form the basis for variance-based sensitivity analysis.

When calculating model outputs, the parameter space formed by the parameter probability distributions has to be sampled. Various issues should be considered in forming and sampling the parameter space. If the parameter values can be considered uniformly distributed, simple grid-based sampling can be used. Alternatively, random sampling, Latin hypercube sampling or Sobol sampling have been proposed, to name a few sampling schemes (Renardy et al., 2021). Another issue to consider is the ranges of the parameter values and interdependencies between the parameters. Whereas the ranges of individual parameters may well be physically plausible, certain areas of the parameter space (i.e., certain combinations of parameter values) might never exist in nature. The sensitivity of the model output can be considered for one parameter at a time (first order sensitivity analysis), for a group of parameters, or for all the model parameters at once (total order sensitivity analysis).

To form the input-output pairs for variance-based sensitivity analysis, the model has to be evaluated repeatedly. The number of realizations required depends on the number of input parameters and the ranges of their values but usually thousands or tens of thousands of input-output pairs need to be calculated. In case of complex models, this is time consuming. Especially due to the rapid development in the field of Machine Learning (ML), using surrogate models have become an attractive approach as these models can be trained to learn the input-output relationship of the physical models with good accuracy and their implementation is computationally efficient using tailored hardware. The drawback of using ML models is that their behaviour is difficult to interpret in physical terms and therefore, they are often considered as black boxes. Recently, a lot of research has been conducted to improve the interpretability of the data-driven ML models resulting in solutions such as Physics Informed Neural Networks (PINNs). However, developing and training of PINNs is often limited to certain types of physical and ML models. Therefore, when using ML surrogate models to speed up computations and to perform sensitivity analysis, in addition to the accuracy also the models' behaviour in sensitivity analysis should be validated.

1.1 The FAST/EFAST/RBD-FAST and Sobol sensitivity analysis methods

In the case studies presented below, the variance-based Sobol, FAST (Fourier Amplitude Sensitivity Test), EFAST (Extended FAST), and RBD-FAST (Random Balanced Design - FAST) sensitivity analysis methods are used. All these methods are based on the decomposition of the model function $f(\boldsymbol{X})$ into orthogonal components of 1, 2, …, *d* variables, $f_i(X_i), f_{i,j}(X_i, X_j), \dots, f_{i,\dots,d}(X_i, \dots, X_d)$, respectively, where *d* is the number of input parameters. These component functions can then be used to obtain conditional estimates of the model output with respect to 1, 2, …, *d* input parameters together with corresponding variances. The first order sensitivity indices are calculated as the one-parameter variances V_i scaled by

the total variance of the model output whereas the total order sensitivity indices for the individual parameters are calculated as one minus the variance due to all the other parameters, scaled by the variance of the model output.

The Sobol, FAST and RBD methods differ in the way the decomposition and variance calculation is implemented. In the Sobol method, the so called pick'n'freeze scheme is used where the columns of a sample matrix are replaced by the columns of the reference sample matrix one by one and corresponding correlations are obtained. Several variants of the method have been suggested, some of which require certain sampling scheme of the parameter space, also referred to as Saltelli sampling (Saltelli et al., 2010). The method is efficient and provides first and total order sensitivity indices in the same run.

The FAST and EFAST methods use the Fourier' transform to calculate the functional decomposition of the model function. The input factors are assigned corresponding frequencies and the contribution of these frequencies to the output is estimated. First order sensitivity can be obtained using the FAST method while EFAST, assigning high frequency to the factor of interest and lower frequencies to all the other factors, can be used to obtain total order effects. The RBD method is a simplification of the FAST using one basic frequency and permutations of the model output for the input factors.

1.2 Sensitivity analysis framework used in the case studies

In this report three cases of sensitivity analysis, related to geological repositories for spent nuclear fuel, are considered. In the first case (marked by blue boxes in *Figure 1*), a compartmental model of radionuclide transport in biosphere is considered with the target parameter being the dose conversion factor of a certain radionuclide from the repository to a human living in the area. In this case, no surrogate model was developed as the calculations required for implementing the compartmental model are of relatively low complexity. The mode involves over 100 input parameters and therefore, the Morris method was used first to screen for the most important parameters. Two methods, EFAST and Sobol were then compared for sensitivity analysis.

The second case (presented using red boxes in *Figure 1*) involves surrogate model based on the Convolutional Neural Network (CNN) architecture. The aim was to study if the surrogate model gives similar results when used for sensitivity analysis compared to the physical model expressed by the advection-dispersion equation (ADE). The case study is based on the experiment and the ADE model described in (Flury et al., 2004). ^{137}Cs concentration values along the time-distance plane of the sand tube experiment are used to train the surrogate model and the sensitivity of the model output to four ADE parameters is studied.

The third case (presented using green boxes in *Figure 1*) involves geochemical models where the physical model is represented by the Orchestra solver and a fully connected Neural Network (NN) surrogate model is trained to follow the input-output map of the solver as accurately as possible. Two models, one with 3 inputs and 16 outputs, and another with 6 inputs and 23 outputs are considered. The models belong to the benchmark dataset described in more detail in other parts of this volume.

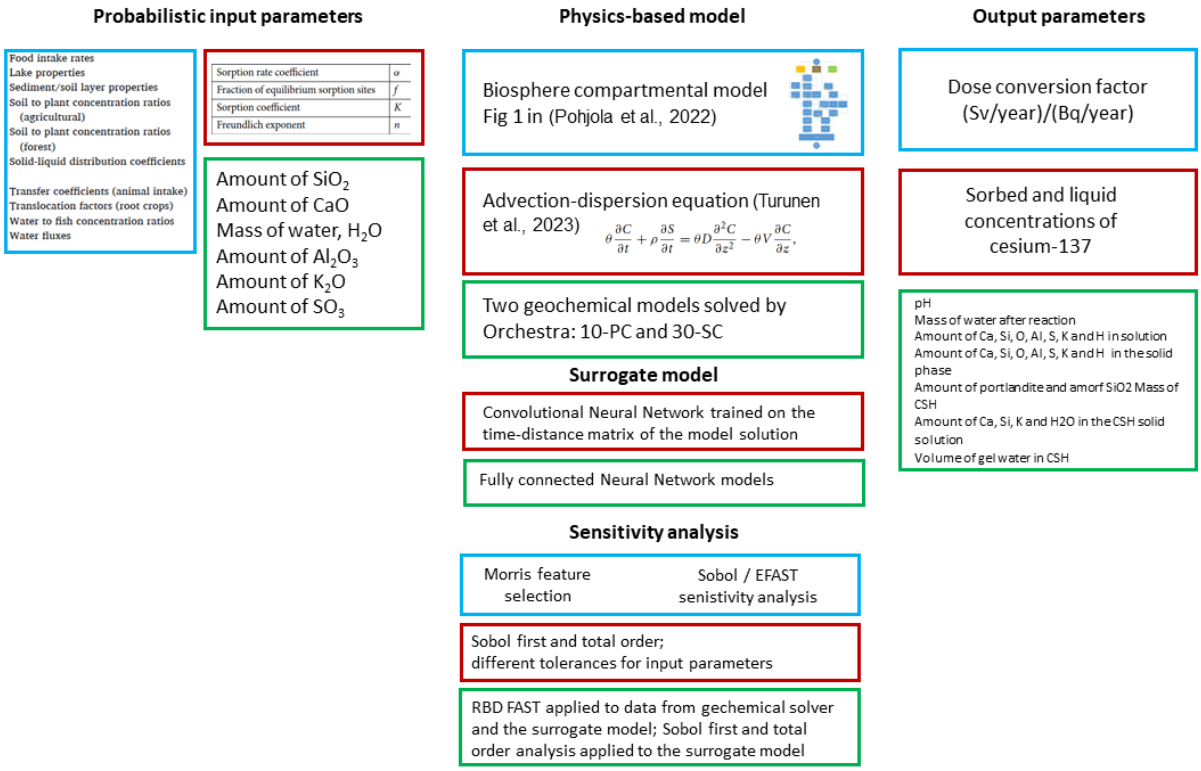


Figure 1 – Framework of the feasibility study of using surrogate models for sensitivity analysis.

2. Sensitivity analysis of radionuclide transport in biosphere

To study the effect of a possible release of radionuclides from a geological disposal site, compartmental models are commonly used. These models are described by the concentration ratios (CRs) between compartments and diffusion coefficients within compartments. The model is used to calculate the dose conversion factor from the model input (a unit activity of the considered radionuclide is applied) to a human, living in the vicinity of the disposal site. Among the various scenarios describing the pathways of radionuclides from the repository to humans, those involving a self-sustaining farmer, consuming mostly the food products produced in the region is commonly considered as the most critical.

2.1 A lake-farm-forest model

In our modelling efforts, the context of the Olkiluoto spent nuclear fuel disposal site in Finland is considered. Geological repository of radioactive waste in Olkiluoto, built by Posiva Oy, has progressed to the phase were the company's application for the permission to start the actual deposition of the waste is currently pending, being thus the furthest advanced geological repository in the World. In Olkiluoto, the effect of postglacial land uplift is still observed, causing the coastline to gradually shift westwards with new land areas raising up from the Baltic Sea. The land is relatively flat in the area, and it is expected that new lakes will be formed with surrounding farming areas and patches of forest. This presumption is based on the development of the coastal areas in Western Finland – the areas that were still under seawater some centuries ago.

We have used a lake-farm compartmental model to study the contribution of the various factors on the dose received by a human inhabitant of the area in case of a leakage from the repository (Pohjola et al., 2016; J. K. Pohjola et al., 2019). The model takes into account the postglacial land uplift with the time horizon of 10 000 years from the present. The land uplift model used in the radionuclide transport model is described in (Pohjola et al., 2014) with a more broader historical context given in (J. Pohjola et al., 2019) and employs the data published in (Pohjola et al., 2020). In the EURAD project we have further

elaborated the model to consider the relative effect of forest areas to the dose conversion factor (Pohjola et al., 2022). The model is presented in *Figure 1* using blue boxes. There were over 100 probabilistic input parameters including, in addition to concentration ratios and distribution coefficients, parameters such as food intake rates for various food products and the properties of the lake.

2.2 Sensitivity analysis results

The lake-farm-forest model was implemented and the sensitivity analysis performed using the Ecolego software suite (AFRY AB, version 8.0.6; Avila et al., 2003). The Morris method was used to determine the most significant parameters. Then, 20 most influential parameters were selected for further sensitivity analysis using the Sobol and EFAST methods. Altogether, seven radionuclides were considered: ^{36}Cl, ^{135}Cs, ^{129}I, ^{237}Np, ^{90}Sr, ^{99}Tc and ^{238}U. The following issues were studied:

- *Accumulation of the radionuclides in various compartments in the next 2000 – 10000 years in 2000-year steps*. Whereas ^{36}Cl had mostly exited the model already after the first 2000 years, ^{135}Cs, ^{129}I, ^{99}Tc and ^{238}U were mostly accumulated in the deep sediment or soil. Most of ^{90}Sr had decayed and ^{237}Np was mostly accumulated in the lake sediment.
- *Sensitivity of dose conversion factors to model parameters*. It was found that, in general, the most influential parameters were the physical properties of the agricultural soil, the physical properties of the deep soil/sediment and the concentration ratios from soil to plants. The results depend on the radionuclide with deep soil/sediment properties dominating for ^{135}Cs and concentration ratios from soil to plant dominating for ^{99}Tc, for example. The proportions of sensitivity among input parameters differed slightly for the two analysis methods (i.e., EFAST and Sobol).
- *The importance of different pathways of the radionuclides from the repository to humans*. It was found that the relative importance of the pathways involving forest vs. agricultural land depended significantly on the level of irrigation. If irrigation of the agricultural land using the lake water was taken into account, the relative contribution of the forest pathways remained significantly lower compared to the case with no irrigation. Especially the relative contribution of mushroom consumption was significant for ^{135}Cs if agricultural land was not irrigated.

The detailed results are presented in (Pohjola et al., 2022).

2.3 Discussion on sensitivity analysis of the biosphere model

The assessment of the sensitivity of the dose conversion factors on the parameters of a compartmental biosphere model relies highly on various assumption underlying the model (such as the food intake profile or the way of life of the people living in the area). These results should not be interpreted in the absolute scale; they rather indicate relative importance of different factors and scenarios. However, the study clearly indicates the feasibility of using sensitivity analysis in the assessment of the importance of the various factors.

The reliability of the results depends on the reliability of the parameter values used in the model. For example, the values of concentration ratios and diffusion coefficients for the different radionuclides have been looked up from various reports and publications and it is an open question how, for example, animal and plant species should be grouped or what is the effect of the environmental factors on these parameters.

3. Surrogate models in sensitivity analysis: the Hanford case

This case is based on the experiment carried out to investigate the risks of radioactive contamination in the area of the Hanford plutonium production site in the US State of Washington, where leakage of radioactive material has occurred. The original experiment is described in (Flury et al., 2004); sand from the Hanford site was packed into glass cylinders of 7.2 cm length and NaCl water solution was pumped through the cylinders, after which ^{133}Cs was injected. The propagation of the ^{133}Cs was observed for

both liquid and sorbed concentrations. The phenomenon was modelled by one-dimensional advection-dispersion equation (ADE) extended with linear and nonlinear Freundlich isotherms (see Eqs. 1…4 in Flury et al., 2004). The model involves 10 parameters, 6 of which were fixed and 4 were selected for sensitivity analysis (see *Figure 1*). We implemented the ADE model using the Python FiPy package (Guyer et al., 2009); the model implementation yielded very similar results to those obtained in the actual experiment of (Flury et al., 2004) for the ratio of liquid vs sorbed concentrations.

3.1 Convolutional Neural Network based surrogate model

In the EURAD project, our aim was to study if a Convolutional Neural Network (CNN) based surrogate model gives similar results compared to physical model implementation in sensitivity analysis (Turunen and Lipping, 2023). The model results were arranged along the time-distance plane to visualize the sorbed and liquid concentrations of ^{133}Cs at any distance from the injection point in the sand tube at any time up to 12 min after the injection. We then trained CNNs to learn these propagation surfaces when the four model parameters were randomly varied within prespecified ranges. Altogether, up to 163 840 realizations of the probabilistic parameter values were used in training. The error between the surrogate model output and the FiPy package realization of the physical model was at least two orders of magnitude less than the corresponding concentrations.

3.2 Sensitivity analysis of the Flury model

Sensitivity analysis was performed using the Python SAlib package. Sobol first and total order analysis was applied. In addition, sensitivity analysis was performed using two tolerance ranges for the four input parameters. We considered the tolerances suggested in the original work by Flury et al. as narrow tolerances and used tolerances ranging half of the mean value of each parameter to both sides as wide tolerances (see Tables 2 and 3 in (Turunen and Lipping, 2023)). The main lessons learned are summarized in *Figure 2*. When the first order Sobol sensitivity analysis was applied and the four input parameters were varied within relatively narrow ranges, both the ADE model and the surrogate model indicated similar sensitivities of the liquid concentration to the four parameters along the whole time-distance plane. When wider tolerances of the input parameters were used, the sensitivities obtained with the surrogate model (NN in the figure) remained similar but those obtained with the ADE model (DE in the figure) changed. When using the narrow ranges but switching from first order to total order Sobol analysis, the sensitivity indices changed for both models. Indeed, this indicates that the four parameters are not independent which can also be seen from the ADE equations.

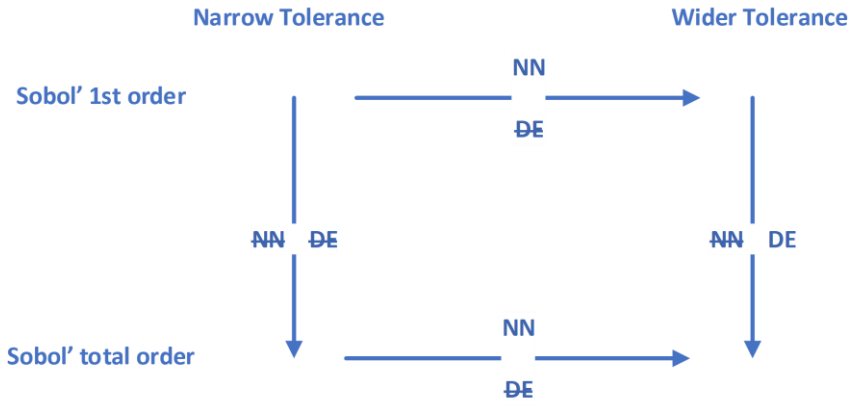


Figure 2 – Lessons learned from the sensitivity analysis using surrogate model in the Hanford case.

3.3 Discussion on sensitivity analysis of the Hanford case

The study described in this section is, indeed, a brief snapshot of the possibility to use surrogate models for sensitivity analysis. There are various ways to develop surrogate models such as polynomial chaos expansion, Gaussian processes or neural networks. Even when using neural networks, different

networks architectures might behave in a different manner. On the other hand, there are also a large variety of ways to perform sensitivity analysis. The study indicated, however, that even if in a certain case (in terms of the input parameter ranges and/or sensitivity analysis methodology) the surrogate model yields similar sensitivities of the model output to its inputs as the physical model, when changing the methodology or analysis framework slightly, different results can be obtained. The study thus points out the need for careful specification of the analysis framework.

4. Surrogate models in sensitivity analysis: geochemical processes

The third case study is based on benchmark data from a geochemical model generated in the EURAD DONUT work package. Cementitious systems were modelled using three solvers: the GEM Software, Orchestra and Phreeqc. Surrogate models were developed to emulate the behaviour of the solvers. The dataset and the surrogate models are described in more detail in the reports of other groups involved in the DONUT work package. In this case study we focused on the following research questions:

- How do the Neural Network based surrogate models compare with the geochemical solvers in terms of sensitivity analysis?
- How does the model complexity affect the results of the sensitivity analysis?
- How does the size of the data set used to train the NN-based surrogate models affect its feasibility in sensitivity analysis?

For this preliminary study two models of different complexity (referred in the following as models ’10-PC’ and ’30-SC’) were selected and the NN-based surrogate models were trained using 5000 / 50000 realizations of the geochemical solver outputs. The model parameters are given in *Table 1*. The input parameters were sampled in the ranges given in the table using the Latin Hypercube sampling (LHS) scheme.

Table 1 – Model parameters and their tolerances

Model 10-PC						
Inputs				**Outputs**		
Description	**Symbol**	**Unit**	**Range**	**Description**	**Symbol**	**Unit**
Amount of SiO_2	SiO_2	mole	0.2…0.7	pH	pH	
Amount of CaO_2	CaO_2	mole	0.1…1.8	Mass of water after reaction	MassWater	kg
Mass of water	H2O	kg	0.05…0.15	Amount of Ca in solution	Ca_aq	mole
				Amount of Si in solution	Si_aq	mole
				Amount of O in solution	O_aq	mole
				Amount of H in solution	H_aq	mole
				Amount of Ca in the solid phase	Ca_s	mole
				Amount of Si in the solid phase	Si_s	mole
				Amount of O in the solid phase	O_s	mole
				Amount of H in the solid phase	H_s	mole
				Mass of CSH	mCSHQ	kg
				Amount of Ca in the CSH solid solution	Ca_ss	mole
				Amount of Si in the CSH solid solution	Si_ss	mole
				Amount of Si in the CSH solid solution	H2O_ss	mole
				Volume of all solids	V_s	cm^3
				Volume of gel water in CSH	GelWater	cm^3
Model 30-SC						

Inputs				Outputs (in addition to the 10-PC model)		
Description	**Symbol**	**Unit**	**Range**	**Description**	**Symbol**	**Unit**
Amount of SiO_2	SiO_2	mole	0.3…0.6	Amount of Al in solution	Al_aq	mole
Amount of CaO_2	CaO_2	mole	0.9…1.4	Amount of Al in solution	S_aq	mole
Mass of water	H2O	kg	0.05…0.15	Amount of Al in solution	K_aq	mole
Amount of Al_2O_3	Al_2O_3	mole	0.03…0.07	Amount of Al in the solid phase	Al_s	mole
Amount of K_2O	K_2O	mole	6…12 E^{-3}	Amount of S in the solid phase	S_s	mole
Amount of SO_3	SO_3	mole	0.02…0.05	Amount of K in the solid phase	K_s	mole
				Amount of K in the CSH solid solution	K_ss	mole

4.1 Surrogate models

Fully connected neural networks were used for surrogate models. The model architecture was optimized separately for the two models using the Optuna Python package. Based on the optimization results, a network with 3 and 6 hidden layers were used for the models 10-PC and 30-SC, respectively. The models were trained using 5000 / 50000 Orchestra realizations for training and a separate set of 500 / 5000 realizations for validation, respectively. After training, a separate set of 5000 / 50000 realizations were used to test the surrogate models. The mean absolute percentage errors over the test set for both models and all output parameters are given in *Table 2*.

Table 2 – Estimation errors of the surrogate models on test data

Model 10-PC							
Surrogate model trained with 5 000 realizations				**Surrogate model trained with 50 000 realizations**			
Output	**MAPE %**	**Output**	**MAPE %**	**Output**	**MAPE %**	**Output**	**MAPE %**
pH	*0.082*	O_s	*0.062*	pH	*0.012*	O_s	*0.013*
MassWater	*1.299*	H_s	*0.124*	MassWater	*0.601*	H_s	*0.024*
Ca_aq	*0.093*	mCSHQ	*0.682*	Ca_aq	*0.027*	mCSHQ	*0.126*
Si_aq	*0.092*	Ca_ss	*0.361*	Si_aq	*0.034*	Ca_ss	*0.080*
O_aq	*0.099*	Si_ss	*0.459*	O_aq	*0.028*	Si_ss	*0.075*
H_aq	*0.085*	H2O_ss	*0.239*	H_aq	*0.027*	H2O_ss	*0.070*
Ca_s	*0.199*	V_s	*0.036*	Ca_s	*0.068*	V_s	*0.005*
Si_s	*0.198*	GelWater	*0.450*	Si_s	*0.073*	GelWater	*0.122*
Model 30-SC							
Surrogate model trained with 5 000 realizations				**Surrogate model trained with 50 000 realizations**			
Output	**MAPE %**	**Output**	**MAPE %**	**Output**	**MAPE %**	**Output**	**MAPE %**
pH	*0.035*	S_s	*2.661*	pH	*0.007*	S_s	*0.366*
MassWater	*4.488*	K_s	*6.695*	MassWater	*0.545*	K_s	*3.931*
Ca_aq	*0.159*	O_s	*0.050*	Ca_aq	*0.037*	O_s	*0.006*
Si_aq	*0.138*	H_s	*0.057*	Si_aq	*0.048*	H_s	*0.007*
Al_aq	*0.340*	mCSHQ	*0.651*	Al_aq	*0.039*	mCSHQ	*0.079*
S_aq	*0.576*	Ca_ss	*0.258*	S_aq	*0.048*	Ca_ss	*0.031*
K_aq	*0.108*	Si_ss	*0.269*	K_aq	*0.033*	Si_ss	*0.035*
O_aq	*0.133*	K_ss	*7.224*	O_aq	*0.025*	K_ss	*3.906*
H_aq	*0.123*	H2O_ss	*0.230*	H_aq	*0.026*	H2O_ss	*0.024*

Ca_s	*0.153*	V_s	*0.032*	Ca_s	*0.022*	V_s	*0.002*
Si_s	*0.267*	GelWater	*0.292*	Si_s	*0.043*	GelWater	*0.050*
Al_s	*1.334*			Al_s	*0.175*		

4.2 Sensitivity analysis of the geochemical models and the surrogates

The following cases of sensitivity analysis were considered:

1. using the RBD Fast sensitivity analysis method and the input-output data obtained with the Orchestra solver
2. using the RBD Fast sensitivity analysis method and the input-output data obtained with the NN surrogate
3. using the Sobol first order sensitivity analysis and the input-output data obtained with the NN surrogate
4. using the Sobol total order sensitivity analysis and the input-output data obtained with the NN surrogate.

The SALib Python package was used to perform the sensitivity analysis. To use the Sobol method, the input parameter space has to be sampled using the Saltelli sampling scheme (see section 1.1). In this case study, precalculated Orchestra realizations based on the LHS sampling scheme were used and therefore the Sobol method could only be applied to the NN surrogate model.

When the NN surrogate model was trained using 50 000 realizations, the sensitivity index values for the Orchestra data and the surrogate model were virtually the same. In Figure 3 this can be seen as nearly equal heights of the two leftmost bars at each input parameter. The figure also shows that for the particular output, the Sobol method gives somewhat different sensitivity values to the input parameters compared to the RBDFast method. Furthermore, the Sobol total order sensitivities also differ from the first order analysis results. Interestingly, the sensitivity of H_{aq} to CaO is highest among the inputs in the 10-PC model whereas for the 30-SC model the effect of CaO is negligible. Discussion of the results for all models, outputs and sensitivity analysis methods is out of the scope of this report, however, for most of the outputs the four analyses methods gave similar input sensitivities. Naturally, some of the outputs were sensitive only to a certain input parameter (such as Ca_s to CaO and Si_s to SiO_2 in the 10-PC model case as well as Mass water to H_2O and Al_s to Al_2O_3 in the 30-SC model case).

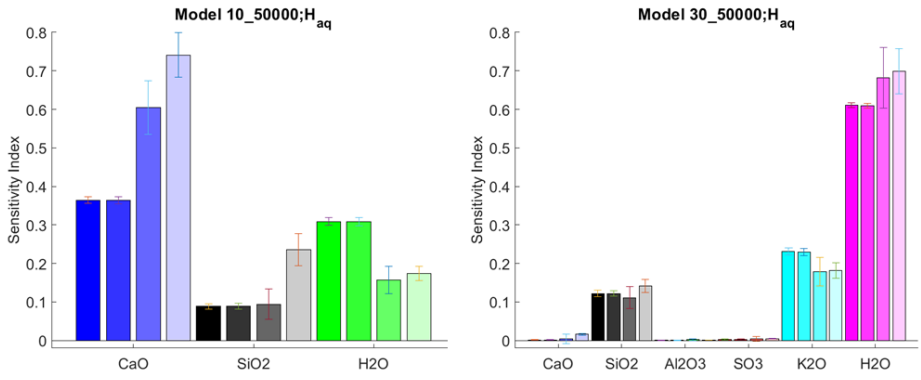


Figure 3 – *Example of sensitivity analysis results for the output Haq for the 10-PC (left panel) and 30-SC (right panel) models. Sensitivity to the different inputs is marked by different colours for clarity. The darkness of the bar denotes the analysis method according to the order of the list at the beginning of section 4.2 with the darkest (leftmost) bar presenting the sensitivity index obtained with the RBD Fast method applied to the Orchestra data and the lightest (rightmost) bar presenting the sensitivity index obtained with the Sobol total order analysis applied to the NN surrogate model data.*

When comparing the sensitivity index values obtained using 5 000 vs. 50 000 realizations, the results were similar indicating that 5 000 realizations is mostly sufficient for sensitivity analysis. An example of

a discrepancy between the two cases is shown in *Figure 4* for the Al_{aq} output parameter where the Sobol total order sensitivity to some inputs increases significantly when sample size is increased.

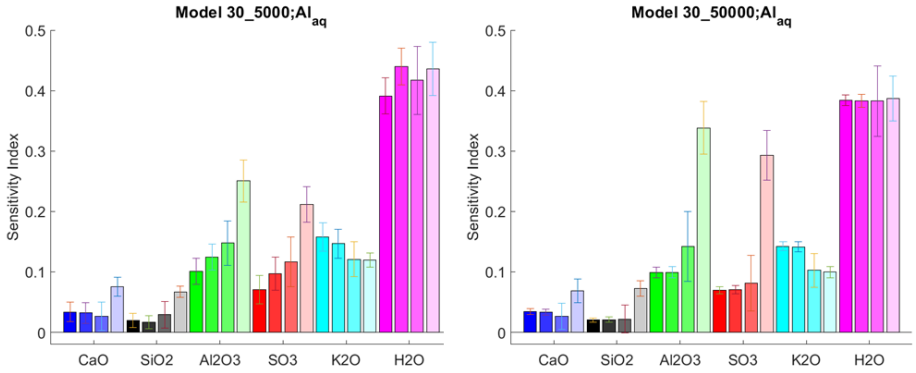


Figure 4 – Comparison of sensitivity analysis results for Al_{aq} in case of using 5 000 (left panel) vs. 50 000 (right panel) realizations.

Interesting insight can be obtained when looking at the sum of the sensitivity indices over all inputs for the Orchestra vs. surrogate model case as well as the first order vs. total order analysis case. These results are presented in *Figure 5*. If 50 000 realizations are used, the total sensitivities for the Orchestra model and the surrogate model (red and green curves, respectively) almost coincide. Sobol analysis, applied only to the surrogate models, shows that the total order sensitivities sum up to one or more while the first order sensitivities sum up to one or less. This comes from the fact that variance-based total order indices calculate the higher-order terms twice whereas the first order indices do not take higher-order terms into account. We can conclude from the figure that the output parameters for which the first and total order curves diverge are affected by the interactions between the input parameters whereas the output parameters for which all the curves stay close to one, are affected either by a single input or the inputs affect them more-or-less independently.

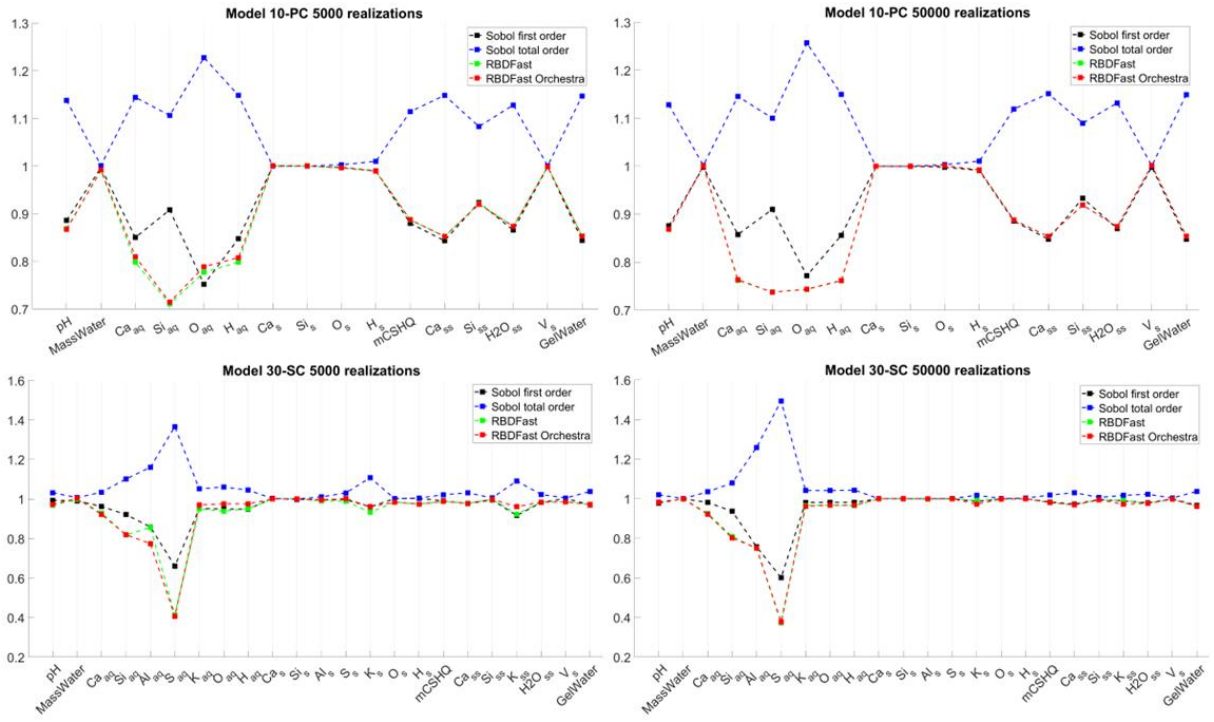


Figure 5 - Sensitivity indices summed over the input parameters.

4.3 Discussion on sensitivity analysis of geochemical models

Sensitivity analysis is an important tool when validating models of geochemical processes. Various tools are available and choosing between them is sometimes more an art than science. Visualizations and interpretation of the results plays crucial role. Surrogate models serve as proxies for the complex physical models, however, as the computers become more powerful, obtaining thousands of realizations with geochemical solvers becomes more feasible. Replacing certain time-consuming calculations by surrogates is, however, still a useful option.

The results presented in this section indicate that Neural Network surrogate models can well emulate the behaviour of geochemical solvers and the approximation error decreases when the number of realizations is increased from 5 000 to 50 000 (see *Table 2*). In many cases, however, 5 000 realizations is enough for analysing the relative sensitivity of model outputs to the input parameters. More complex models require deeper network architectures to achieve similar approximation error and confidence in sensitivity analysis. A useful way to validate the analysis is to look at the sum of sensitivities over input parameters. *Figure 5* shows that the RBDFast method gives almost identical results for the data obtained using the geochemical solver and the surrogate model. The results obtained using the Sobol analysis differ somewhat but point to the same direction. Due to using presampled data, Sobol analysis could only be performed on the surrogate model. The results for first and total order Sobol analysis are coherent based on the theory underlying the methods.

5. Conclusions

In this report three cases of sensitivity analysis of models related to geological disposal of radionuclide waste have been presented. In two of the cases surrogate models have also been considered. Sensitivity analysis has various roles when evaluating models; in addition to revealing the sensitivity of the model output to the input parameters (and thus indicating which input parameters should be paid more attention when designing models or performing measurements), it is also a tool for validating the models and studying the interactions between model inputs. As indicated by our results, different models provide somewhat different results, and the sensitivity indices depend also on the tolerances used for model inputs and the sampling schemes of the input parameter space. The study on the sensitivity analysis of the geochemical models is still ongoing and the results presented in this report are preliminary.

Code source

The following codes were used:

Python packages:

- FiPy for solving systems of differential equations
- SALib for sensitivity analysis
- Scikit-learn for implementation of machine learning algorithms
- Optuna for optimizing neural network architectures.

Ecolego software suite (Avila et al., 2003) was used for implementation and sensitivity analysis of compartmental models. The software is available from https://afry.com/en/service/afry-ecolego-modelling-and-simulation-software

References

Avila, R., Broed, R., Pereira, A., 2003. Ecolego - A toolbox for radioecological risk assessment. International Atomic Energy Agency (IAEA).

Flury, M., Czigány, S., Chen, G., Harsh, J.B., 2004. Cesium migration in saturated silica sand and Hanford sediments as impacted by ionic strength. J. Contam. Hydrol. 71, 111–126. https://doi.org/10.1016/j.jconhyd.2003.09.005

Guyer, J.E., Wheeler, D., Warren, J.A., 2009. FiPy: Partial Differential Equations with Python. Comput. Sci. Eng. 11, 6–15. https://doi.org/10.1109/MCSE.2009.52

Pohjola, J., Turunen, J., Lipping, T., 2020. Lake and mire isolation data set for the estimation of post-glacial land uplift in Fennoscandia. Earth Syst. Sci. Data 12, 869–873. https://doi.org/10.5194/essd-12-869-2020

Pohjola, J., Turunen, J., Lipping, T., Ikonen, A.T.K., 2022. On the inclusion of forest exposure pathways into a stylized lake-farm scenario in a geological repository safety analysis. J. Environ. Radioact. 255, 107019. https://doi.org/10.1016/j.jenvrad.2022.107019

Pohjola, J., Turunen, J., Lipping, T., Ikonen, A.T.K., 2016. Probabilistic assessment of the influence of lake properties in long-term radiation doses to humans. J. Environ. Radioact. 164, 258–267. https://doi.org/10.1016/j.jenvrad.2016.08.001

Pohjola, J., Turunen, J., Lipping, T., Ikonen, A.T.K., 2014. Landscape development modeling based on statistical framework. Comput. Geosci. 62, 43–52. https://doi.org/10.1016/j.cageo.2013.09.013

Pohjola, J., Turunen, J., Lipping, T., Sivula, A., Marila, M., 2019. Historical Perspectives to Postglacial Uplift: Case Studies from the Lower Satakunta Region, SpringerBriefs in Geography. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-00970-0

Pohjola, J.K., Turunen, J.J., Lipping, T., Ikonen, A.T.K., 2019. Probabilistic assessment of the impact of bottom sediment on doses to humans from a groundwater-mediated radionuclide release in a farm-lake scenario. J. Radiol. Prot. 39, 564–578. https://doi.org/10.1088/1361-6498/ab134d

Renardy, M., Joslyn, L.R., Millar, J.A., Kirschner, D.E., 2021. To Sobol or not to Sobol? The effects of sampling schemes in systems biology applications. Math. Biosci. 337, 108593. https://doi.org/10.1016/j.mbs.2021.108593

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. 181, 259–270. https://doi.org/10.1016/j.cpc.2009.09.018

Swiler, L., Becker, D.-A., Brooks, D., Govaerts, J., Koskinen, L., Kupiainen, P., Plischke, E., Röhlig, K.-J., Saveleva, E., Spiessl, S., Stein, E., Svitelman, V., 2021. Sensitivity Analysis Comparisons on Geologic Case Studies: An International Collaboration (No. SAND2021-11053, 1822591, 700264). https://doi.org/10.2172/1822591

Turunen, J., Lipping, T., 2023. Feasibility of neural network metamodels for emulation and sensitivity analysis of radionuclide transport models. Sci. Rep. 13, 6985. https://doi.org/10.1038/s41598-023-34089-9

Sparse Polynomial chaos expansion for uncertainty quantification in both pore water composition calculation and model sorption

[1,2,*]Sochala, P., [2]Chiaberge, C., [2]Claret, F. [2,3,4]Tournassat, C.

[1]CEA, DAM, DIF, F-91297 Arpajon, France

[2]BRGM, 3 avenue Claude Guillemin, 45060 Orléans, France

[3]ISTO,Université d'Orléans-CNRS-BRGM, France

[4]Lawrence Berkeley National Laboratory, Berkeley, CA, USA

[*]Pierre Sochala was employed by BRGM the first two years of the project before moving to CEA

This work was published in:

- Sochala, P., Chiaberge, C., Claret, F and Tournassat. C. Uncertainty propagation in pore water chemical composition calculation using surrogate models. Sci Rep 12, 15077 (2022). https://doi.org/10.1038/s41598-022-18411-5
- Sochala, P., Chiaberge, C., Claret, F and Tournassat. C. Dimension reduction for uncertainty propagation and global sensitivity analyses of a cesium adsorption model, Journal of Computational Science, 2023. https://doi.org/10.1016/j.jocs.2023.102197.

The present report is an extended summary of the study. For a complete description of the work, please refer to the two above (gold open access) publications.

Abstract

Performance assessment in deep geological nuclear waste repository systems necessitates an extended knowledge of the pore water chemical conditions prevailing in host-rock formations. Indeed many clay based materials properties such as sorption, diffusion and swelling are influenced by porewater composition. In the last two decades, important progress has been made in the experimental characterization and thermodynamic modelling of pore water speciation, but the influence of experimental artifacts and uncertainties of thermodynamic input parameters are seldom evaluated. In this respect, we conducted an uncertainty propagation study in a reference geochemical model describing the pore water chemistry of the Callovian-Oxfordian clay formation. Nineteen model input parameters were perturbed, including those associated to experimental characterization (exchanged cations, cation exchange selectivity coefficients) and those associated to generic thermodynamic databases (solubilities). A set of 13 quantities of interest were studied by the use of polynomial chaos expansions built non-intrusively with a least-squares forward stepwise regression approach. Training and validation sets of simulations were carried out using the geochemical speciation code PHREEQC. The statistical results explored the marginal distribution of each quantity of interest, their bivariate correlations as well as their global sensitivity indices. The influence of the assumed distributions for input parameters uncertainties was evaluated by considering two parametric domain size. Then a caesium adsorption model was upstream chained with the pore water composition model as the number of uncertain input parameters is about twenty for each of the two models, a dimension reduction technique is implemented to build a polynomial approximation of the caesium distribution coefficient in a reduced subspace. The interpretation of the linear transformations projecting the original inputs to the reduced

coordinates is broadly consistent with the geochemical features of the model. Global sensitivity analysis highlights that the variance of the cesium distribution coefficient is overwhelmingly governed by the adsorption model. Still, this conclusion is nuanced when considering the whole cumulative distribution function for which the interaction effects between the two models account for a fifth.

Significance Statement

Radionuclides speciation and sorption depends on porewater composition. Therefore deciphering uncertainties of pore water model composition and its influence on sorption processes if of paramount importance. To tackle this issue surrogate modelling that aims to provide a simpler, and hence faster, model which emulates the specified output of a more complex model in function of its inputs and parameters can be used. Here, polynomial chaos expansion, method for representing a random variable in terms of a polynomial function of other random variables has been used to elaborate the surrogate model. Doing so it was possible not only to quantify the uncertainties of the quantities of interest for future performance evaluation calculations, but also to identify the main influential input parameters in the pore water models. This latter plays a minor role into the caesium adsorption process.

Table of content

List of figures

1. Context and main conclusions

Clay-based materials are foreseen as buffer and backfill materials in almost all repository concepts. Many properties of the clay barrier is depending on the porewater chemistry as illustrated on the Figure 12.

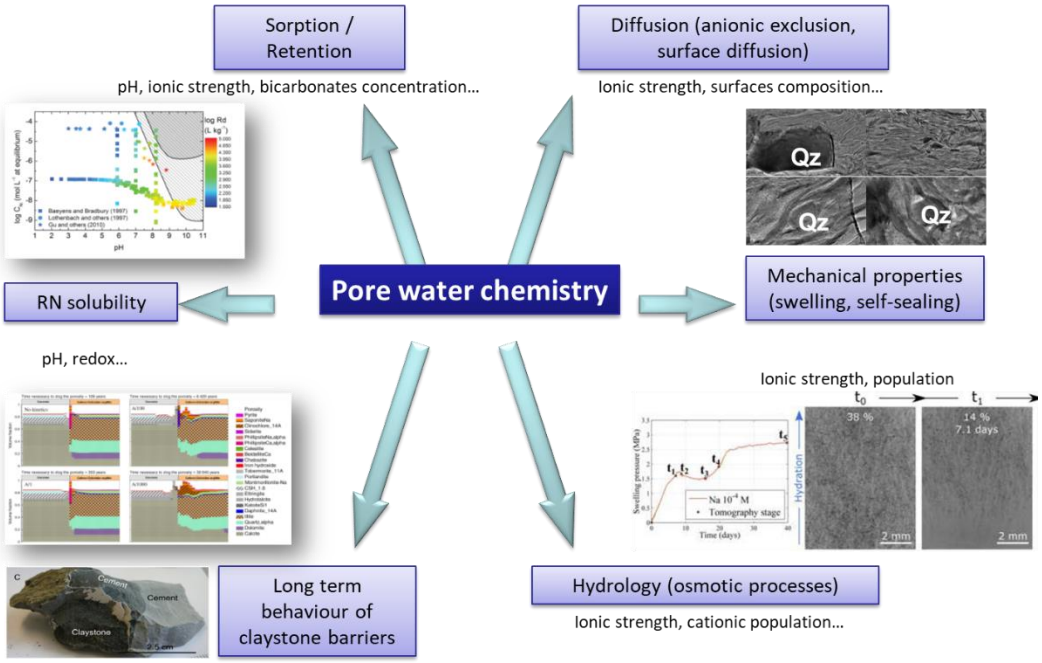


Figure 12: Pore water chemistry influences many interesting properties of clay barrier.

Pore water composition models at equilibrium that rely on the Gibbs' phase rule, and that have been validated by comparison with in situ seepage water composition, are available for various clay-rocks, providing reliable models for far-field initial conditions. Those models relies on the use of thermodynamic database for mineral solubility or cationic exchange database and some parameters can suffer of significant uncertainties. In this context been able to quantify such uncertainties is of paramount importance. Uncertainty propagation in a pore water composition model that mimics the composition of pore water in the Callovian-Oxfordian formation at Bure, France was developed using polynomial chaos expansion to elaborate surrogate models (Sochala *et al.*, 2022). Surrogate modelling allow amongst others to make sensitivity analysis. For example we were able to show that pH, pe, Fe, Si, Al, S(-II), P_{CO2} are only governed by the solubilities of mineral phases used to build the pore water model while K depends on the selectivity coefficient Na/K and Si only depends on quartz solubility. To which extend uncertainties on this pore water models influence caesium adsorption was examined by chaining a caesium adsorption model with the pore water composition model (Sochala *et al.*, 2023). Two approaches were used: a monolithic one or a fragmented one. In the first case, the computational software chain as a single block while in the second one each solver of the computational chain is considered separately. Due to the high dimension problem (37 in the monolithic approach against 19+23 in the fragmented one) dimension-reduction techniques have been applied. A linear supervised dimension reduction approach combined with a functional approximation has proved effective for constructing a surrogate model of the caesium concentration. Assessment of the surrogate models revealed that both monolithic and fragmented approaches have comparable mean squared relative errors. Direct exploitation of the surrogate models through global sensitivity analysis emphasized that the upstream porewater composition model has no effect on the caesium K_D variance. Such an approach could be used for other radionuclides to investigate further the role of the pore water composition model into adsorption processes.

Code source

All the used algorithm are given in the appendix of the published paper, which is in the next section.

References

Sochala, P., Chiaberge, C., Claret, F., Tournassat, C., 2022. Uncertainty propagation in pore water chemical composition calculation using surrogate models. Scientific Reports 12, 15077.

Sochala, P., Chiaberge, C., Claret, F., Tournassat, C., 2023. Dimension reduction for uncertainty propagation and global sensitivity analyses of a cesium adsorption model. Journal of Computational Science, 102197.

Seras coller ici papier 1 puis papier 2