Advances in Uncertainty and Sensitivity Analysis Methods and Applications in GDSA Framework

Fuel Cycle Research and Development

Prepared for
U.S. Department of Energy
Spent Fuel and Waste Science and Technology Campaign

L.P. Swiler, E. Basurto, D.M. Brooks, A.C. Eckert, P.E. Mariner, T. Portone, and E.R. Stein

Sandia National Laboratories

September 25, 2020

M3SF-20SN010304032
SAND2020-10802 R

DISCLAIMER

This information was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness, of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. References herein to any specific commercial product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

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

APPENDIX E

NFCSC DOCUMENT COVER SHEET[1]

Name/Title of Deliverable/Milestone/Revision No.	*Advances in Uncertainty and Sensitivity Analysis Methods and Applications in GDSA Framework* M3SF-20SN010304032
Work Package Title and Number	GDSA-Uncertainty and Sensitivity Analysis Methods-SNL SF-20SN01030403
Work Package WBS Number	1.08.01.03.04
Responsible Work Package Manager	Laura Swiler (Name/Signature)
Date Submitted	September 25, 2020

Quality Rigor Level for Deliverable/Milestone[2]: ☐ QRL-1 ☐ Nuclear Data ☐ QRL-2 ☒ QRL-3 ☐ QRL-4 Lab QA Program[3]

This deliverable was prepared in accordance with Sandia National Laboratories *(Participant/National Laboratory Name)*

QA program which meets the requirements of

☒ DOE Order 414.1 ☐ NQA-1 ☐ Other

This Deliverable was subjected to:

☒ Technical Review ☐ Peer Review

Technical Review (TR)
Review Documentation Provided

☐ Signed TR Report or,
☐ Signed TR Concurrence Sheet or,
☒ Signature of TR Reviewer(s) below

Peer Review (PR)
Review Documentation Provided

☐ Signed PR Report or,
☐ Signed PR Concurrence Sheet or,
☐ Signature of PR Reviewer(s) below

Name and Signature of Reviewers

Jon Helton

Jon Helton

NOTE 1: *Appendix E should be filled out and submitted with the deliverable. Or, if the PICS:NE system permits, completely enter all applicable information in the PICS:NE Deliverable Form. The requirement is to ensure that all applicable information is entered either in the PICS:NE system or by using the NFCSC Document Cover Sheet.*

- *In some cases there may be a milestone where an item is being fabricated, maintenance is being performed on a facility, or a document is being issued through a formal document control process where it specifically calls out a formal review of the document. In these cases, documentation (e.g., inspection report, maintenance request, work planning package documentation or the documented review of the issued document through the document control process) of the completion of the activity, along with the Document Cover Sheet, is sufficient to demonstrate achieving the milestone.*

NOTE 2: *If QRL 1, 2, or 3 is not assigned, then the QRL 4 box must be checked, and the work is understood to be performed using laboratory QA requirements. This includes any deliverable developed in conformance with the respective National Laboratory / Participant, DOE or NNSA-approved QA Program.*

NOTE 3: *If the lab has an NQA-1 program and the work to be conducted requires an NQA-1 program, then the QRL-1 box must be checked in the work Package and on the Appendix E cover sheet and the work must be performed in accordance with the Lab's NQA-1 program. The QRL-4 box should not be checked.*

ACKNOWLEDGEMENTS

This report has benefited from the insights and assistance of many. The authors thankfully acknowledge Michael Nole for his contributions to PFLOTRAN debugging many advanced features for the crystalline reference case. Participants in the international uncertainty and sensitivity analysis collaboration, especially Klaus Röhlig (TU Clausthal) and Dirk Becker (GRS), provided valuable discussion and perspective. The authors gratefully acknowledge the significant contributions of S. David Sevougian (retired) in repository science and of Jon Helton in performance assessment methodology. The authors appreciate and thank the technical reviewers for comments and suggestions that improved the quality of the manuscript. Finally, the authors give a special thanks to Prasad Nair and Tim Gunter of the Department of Energy's Spent Fuel and Waste Science and Technology campaign for their on-going support and guidance.

EXECUTIVE SUMMARY

The Spent Fuel and Waste Science and Technology (SFWST) Campaign of the U.S. Department of Energy (DOE) Office of Nuclear Energy (NE), Office of Fuel Cycle Technology (FCT) is conducting research and development (R&D) on geologic disposal of spent nuclear fuel (SNF) and high-level nuclear waste (HLW). Two high priorities for SFWST disposal R&D are design concept development and disposal system modeling. These priorities are directly addressed in the SFWST *Geologic Disposal Safety Assessment* (GDSA) control account, which is charged with developing a geologic repository system modeling and analysis capability, and the associated software, *GDSA Framework*, for evaluating disposal system performance for nuclear waste in geologic media. *GDSA Framework* is supported by SFWST Campaign and its predecessor the Used Fuel Disposition (UFD) campaign.

This report fulfills the GDSA Uncertainty and Sensitivity Analysis Methods work package (SF-20SN01030403) level 3 milestone – *Advances in Uncertainty and Sensitivity Analysis Methods and Applications in GDSA Framework* (M3SF-20SN010304032). It presents high level objectives and strategy for development of uncertainty and sensitivity analysis tools, demonstrates uncertainty quantification (UQ) and sensitivity analysis (SA) tools in *GDSA Framework* in FY20, and describes additional UQ/SA tools whose future implementation would enhance the UQ/SA capability of *GDSA Framework*. This work was closely coordinated with the other Sandia National Laboratory GDSA work packages: the GDSA Framework Development work package (SF-20SN01030404), the GDSA Repository Systems Analysis work package (SF-20SN01030405), and the GDSA PFLOTRAN Development work package (SF-20SN01030406). This report builds on developments reported in previous *GDSA Framework* milestones, particularly M2SF-19SN01030403.

Contents

Table of Figures

Table of Tables

1. INTRODUCTION

This report presents high level objectives and strategy for development of uncertainty and sensitivity analysis tools in *Geologic Disposal Safety Assessment (GDSA) Framework*, a software toolkit for probabilistic post-closure performance assessment (PA) of systems for deep geologic disposal of nuclear waste. *GDSA Framework* is supported by the Spent Fuel and Waste Science and Technology (SFWST) Campaign of the U.S. Department of Energy (DOE) Office of Nuclear Energy (NE) and its predecessor the Used Fuel Disposition (UFD) campaign.

This report fulfills the GDSA Uncertainty and Sensitivity Analysis Methods work package (SF-20SN01030403) level 3 milestone – *Advances in Uncertainty and Sensitivity Analysis Methods and Applications in GDSA Framework* (M3SF-20SN010304032). It presents high level objectives and strategy for development of uncertainty and sensitivity analysis tools, demonstrates uncertainty quantification (UQ) and sensitivity analysis (SA) tools in *GDSA Framework* in FY20, and describes additional UQ/SA tools whose future implementation would enhance the UQ/SA capability of *GDSA Framework*.

This work was closely coordinated with the other Sandia National Laboratory GDSA work packages: the GDSA Framework Development work package (SF-20SN01030404), the GDSA Repository Systems Analysis work package (SF-20SN01030405), and the GDSA PFLOTRAN Development work package (SF-20SN01030406). This report builds on developments reported in previous *GDSA Framework* milestones, particularly M2SF-19SN01030403 [1].

The repository community must maintain leadership in UQ/SA methods. Computational codes are becoming increasingly complex and require high performance computers to run, resulting in costly sample evaluations. Geologic repository performance assessment in the U.S. involves a code base that includes coupled, multiphysics modeling at high resolution. Having relatively few simulation samples highlights the need to consider surrogate models to sample and explore the input parameter space more extensively. However, this must be done in a careful way so that surrogate accuracy can be tracked and understood in the context of UQ/SA results. Variance-based sensitivity indices are now a standard practice in the sensitivity analysis community but require a large number of function evaluations of the predictive model. Much research has focused on calculating variance-based sensitivity indices while keeping the computational cost reasonable. We note that other sensitivity analysis methods [1] may be better than variance-based methods at identifying patterns of behavior or trends. Another recent approach is to employ “multifidelity” UQ in which many low-fidelity simulation runs (e.g., coarser mesh, simpler physics) augment a small number of high-fidelity runs. Keeping abreast of improvements to existing UQ/SA methods as well as employing new methods is critical to performing sensitivity and uncertainty analysis of new repository systems which will involve large parameter spaces and computationally expensive simulations.

1.1 Overview of this Report

This report provides documentation of the UQ/SA work performed in 2020. The outline of this report is as follows:

- Chapter 2 provides an overview of an exciting new uncertainty quantification method involving models at multiple levels of fidelity, hence the name "Multifidelity Uncertainty Quantification." The main idea in multifidelity UQ is to extract information from a limited number of high-fidelity model evaluations and complement them with a much larger number of a set of lower fidelity evaluations. The final result is an estimator of the mean response with a lower variance: a more accurate and reliable estimator can be obtained at lower computational cost. Chapter 2 demonstrates the use of multifidelity UQ on a PFLOTRAN test problem involving a tank containing chemical and radioactive waste meant to be illustrative of a tank at the Hanford site.
- Chapter 3 discusses the verification exercises performed on the discrete fracture networks (DFN). In FY19, we demonstrated that the DFNs were a main source of uncertainty for the crystalline reference case. This year, we examined the DFNs closely to ensure consistency of the number of fractures and fracture area per unit volume with respect to estimates from the SKB Forsmark site. We also upscaled the DFNs to obtain equivalent continuous porous medium (ECPM) results in terms of the hydraulic conductivity tensor values K. We compared the ECPM results for the different K tensor directions that were obtained from upscaling the DFNs with those reported by SKB and found good agreement. These results are documented in Chapter 3.
- In Chapter 4, we extend the sensitivity analysis of the crystalline reference case performed in FY19. Last year, we identified the need for better indicators of fracture connectivity as well as information about the flow between the repository and the aquifer. These metrics can then be used in SA to help identify why higher peak concentrations happen for certain realizations. To facilitate this, PFLOTRAN added several new quantities of interest, including peak concentrations (of ^{129}I and tracers) in the aquifer at each time step, along with the location of the peaks, mean travel time from the repository to various locations, residence time within the repository, and total water flux in various directions. The results of the sensitivity analysis of the crystalline reference case based on a new set of 800 PFLOTRAN runs (20 realizations of spatial uncertainty x 40 epistemic samples) are documented in Chapter 4.
- Chapter 5 documents the *GDSA Workflow* for a UQ study. Uncertainty quantification workflows are not trivial to define and get running, even when using tools such as Dakota to generate nested studies involving sampling loops over both aleatory and epistemic samples. The analyst has to spend a significant amount of time writing scripts to interface the sample values to PFLOTRAN, extract the results, and put the entire workflow on a high-performance computing platform. Further, as the quantities of interest increase and many vectors of results are generated for each simulation, plotting and aggregating the results in a variety of ways (e.g. averaging over epistemic or aleatory slices as a function of time) becomes very involved. To address this, the *GDSA Workflow* was developed this year. This workflow couples Dakota, PFLOTRAN, and SAW (the Sandia Analysis Workflow software) to present the user with a unified interface where the actual workflow can be dictated in an easy-to-use graphical format. This workflow also allows greater reproducibility and traceability of the actual files and scripts

used for a particular study. This year, we demonstrated the UQ workflow for a GDSA study involving spatial, aleatory, and epistemic uncertainty.

- Chapter 6 documents the sensitivity analysis results we performed on two international case studies as part of an international working group focused on UQ/SA for repository analysis.
- Chapter 7 provides a summary.

1.2 GDSA Framework

GDSA Framework (Figure 1-1) capabilities include multi-physics simulation of coupled processes affecting deep geologic repository performance, uncertainty and sensitivity analysis, pre- and post-processing, and visualization. For a given performance assessment, these tools will be linked to a version-controlled parameter database and an automated run-control system. The overall objectives of *GDSA Framework* development are to:

- create a framework that is flexible enough to take advantage of future advances in hardware, software, and simulation and analysis methods;
- leverage existing high-performance computing capabilities (e.g., meshing, simulation, analysis, and visualization);
- enable increasingly coupled, mechanistic multi-physics modeling;
- provide analysis methods for prioritization of SFWST Disposal Research R&D activities;
- provide transparent implementation of simulation and analysis methods;
- develop and distribute in an open-source environment so that software is freely available to stakeholders ([1-5]).

Objectives specific to the uncertainty and sensitivity analysis capability in *GDSA Framework* are to make available standard sampling-based methods of uncertainty propagation, sensitivity analysis, and uncertainty quantification typically used within U.S. nuclear waste disposal programs (e.g., DOE 2008 [6], DOE 2014 [7], RESS2000[13], RESS2014[14]); and to enable future adoption of new methods consistent with the current standard of practice in the UQ/SA community and appropriate for high-dimensional, highly coupled, nonlinear problems resulting from the implementation of mechanistic multi-physics simulations. Having a consistent, common framework which enables a user to perform a range of sensitivity analysis and UQ approaches for a particular problem or set of simulations allows for reproducibility, comparative analyses, use of verified algorithms, and documentation of best practices. These are important goals for performance assessments now and in the future.

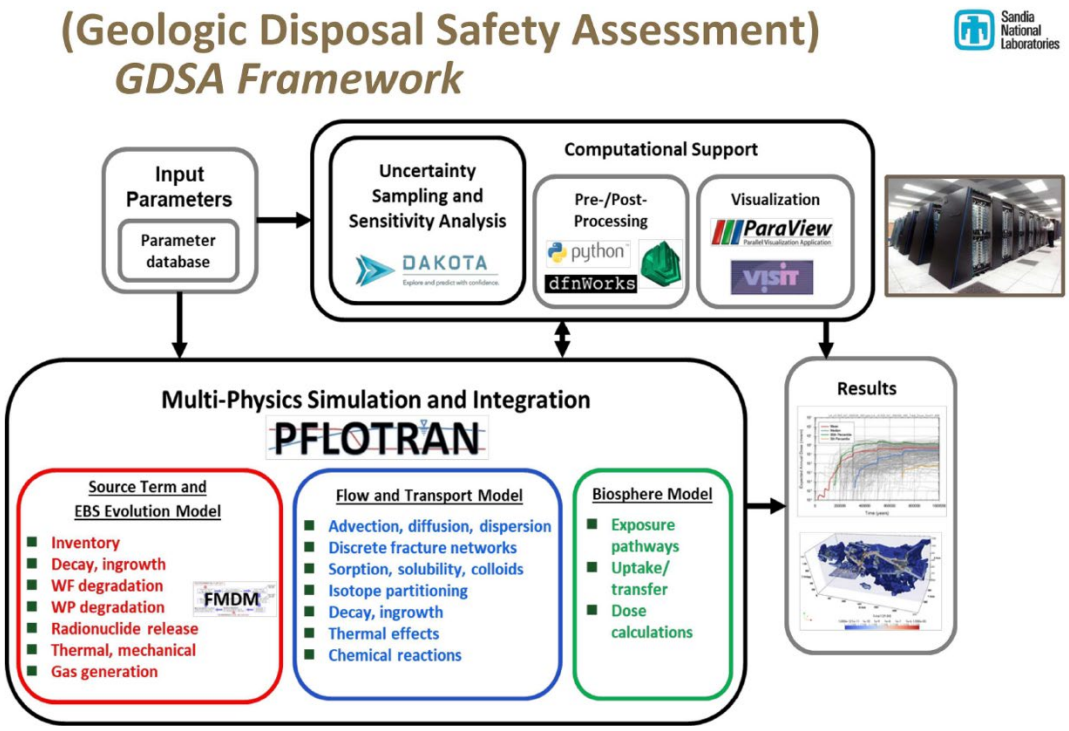


Figure 1-1. The GDSA Framework

The following sections highlight the key components of *GDSA Framework*. More information about each can be found by following the links at https://pa.sandia.gov.

1.2.1 PFLOTRAN

PFLOTRAN is an open source, state-of-the-art, massively parallel subsurface flow and reactive transport simulator ([8-10]). Written in object-oriented Fortran 2003, PFLOTRAN is a porous medium continuum code for modeling multicomponent, multiphase flow and transport, heat conduction and convection, biogeochemical reactions, geomechanics, and isotope decay and ingrowth. The code is developed under a GNU LGPL license allowing for third parties to interface proprietary software with the code. The availability and continuing development of PFLOTRAN are due to an ongoing collaborative effort of several DOE laboratories led by Sandia. PFLOTRAN development for *GDSA Framework* is described by Mariner et al. ([2-4]) and Sevougian et al. 2018 [11]. PFLOTRAN installation instructions are available at: https://www.pflotran.org.

1.2.2 Dakota

Dakota is an open-source toolkit of algorithms that contains both state-of-the-art research and robust, usable software for optimization and uncertainty quantification (UQ). It is available at: https://dakota.sandia.gov [12]. The algorithms allow a user to explore a computational simulation to answer questions such as:

- what is the best design?
- how safe is this design?
- what are the most important parameters?
- what effects do uncertainties have on my system?

The Dakota software has a rich set of parametric analysis methods that enable design exploration, model calibration, risk analysis, and quantification of margins and uncertainty with computational models. Dakota provides a flexible, extensible interface between simulation codes and these iterative analysis methods, which include:

- optimization with gradient and nongradient-based methods;
- uncertainty quantification with sampling, reliability, stochastic expansion, and epistemic methods;
- parameter estimation using nonlinear least squares (deterministic) or Bayesian inference (stochastic); and
- sensitivity/variance analysis with design of experiments and parameter study methods.

These capabilities may be used on their own or as components within advanced strategies such as hybrid optimization, surrogate-based optimization, mixed-integer nonlinear programming, or optimization under uncertainty. Dakota is a C++ code and has been under development at Sandia since 1994 and has been primarily sponsored by DOE's Advanced Simulation and Computing (ASC) program. As such, it has a focus on interfacing to and running simulations which are computationally expensive, require high performance computing and parallel execution, and exhibit nonlinearities, non-monotonic and/or discontinuous responses, and often involve noisy responses and high-dimensional inputs. Thus, a focus of the algorithm development in Dakota has been on methods that are as efficient as possible and minimize the number of runs required of a high-fidelity simulation model. Such algorithms include surrogate or emulator models, adaptive sampling approaches, and multifidelity UQ methods which augment a small number of high-fidelity runs with many low-fidelity runs to obtain comparable accuracy in statistical estimators.

Dakota contains the uncertainty quantification and sensitivity analysis methods typically used in the U.S. repository program. Dakota implements Latin Hypercube Sampling (LHS) with correlation control on input parameters. It calculates moments on responses of interest as well as correlation matrices (simple, partial, and rank correlations) between inputs and outputs. Dakota also contains an algorithm for performing incremental LHS which allows one to double an initial LHS study such that the second LHS study is a Latin design and the combined initial and second LHS studies together form a Latin hypercube design. Dakota allows nested studies to perform an "outer loop" epistemic sampling and an "inner loop" aleatory sampling to generate ensembles of distributions. Dakota returns tables of input and output amenable to further processing and visualization with additional tools developed within *GDSA Framework* or by an individual user. Additional methods that have been implemented in Python for use in *GDSA Framework* include calculation of standardized regression coefficients via stepwise linear regression and calculation of partial correlation coefficients based on raw data or rank-transformed data.

A graphical depiction of Dakota interfacing with a computational model such as a repository simulation in PFLOTRAN is shown in Figure 1-2. Based on the type of study being performed (optimization, uncertainty quantification, etc.), Dakota chooses the next set of parameters at which to evaluate the simulator and runs the simulator, which returns the performance metrics of interest back to Dakota. Dakota then generates the next set of parameters according to the algorithm being used for the study and keeps iterating until the specified number of samples is reached.

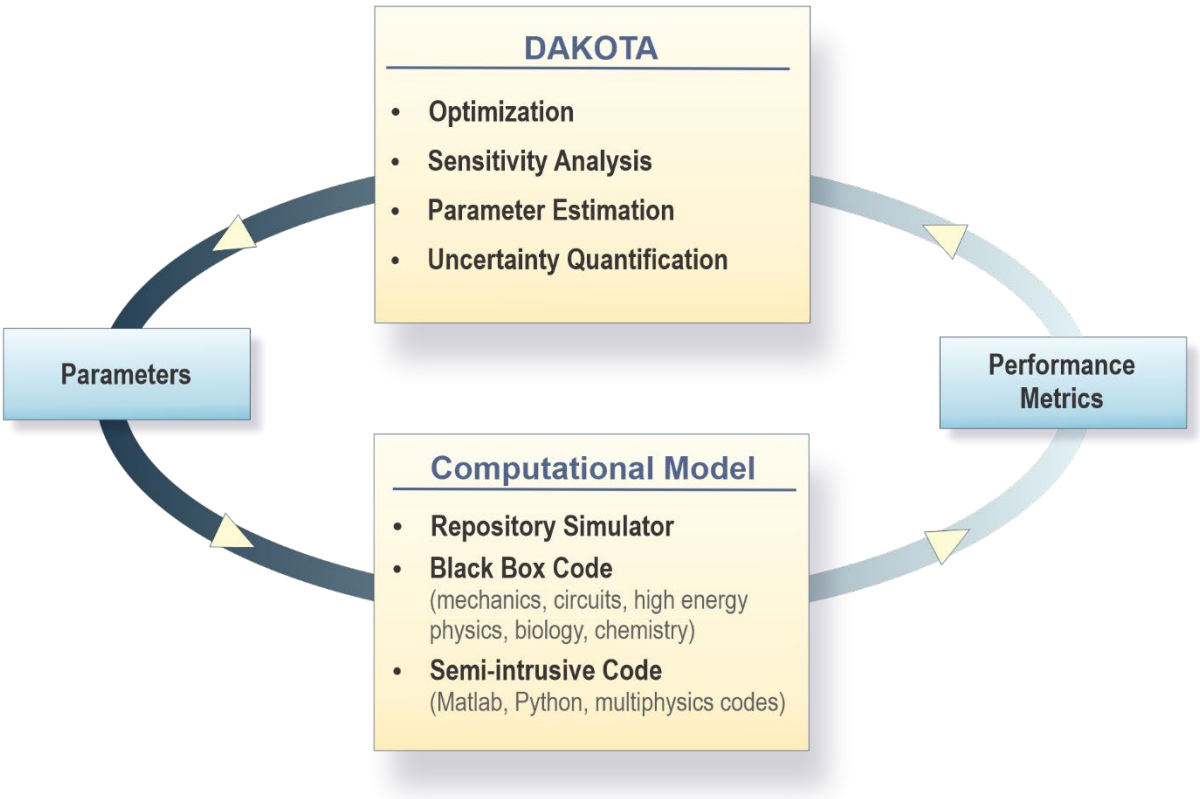


Figure 1-2. Dakota interfacing to a computational model such as a repository simulator

The UQ/SA methods in Dakota have evolved as the standard of practice evolves. Over the past ten years, the Dakota team has invested in methods which calculate the Sobol' variance-based sensitivity indices in an efficient manner. Currently, a Dakota user can calculate these by extensive sampling of the simulation code, by using surrogate methods such as regression or Gaussian process models, and by the use of polynomial chaos expansions. These advanced methods are presented in more detail later as used in *GDSA Framework*. Dakota is an actively maintained and developed code with formal releases issued twice per year. Dakota uses formal software quality development processes including advanced version control, unit and regression testing, agile programming practices, and software quality assessment.

1.2.3 Visualization Tools

GDSA Framework employs ParaView and/or VisIT for visualization of results. ParaView is an open-source, multi-platform data analysis and visualization application developed by Sandia National Laboratories. As stated on the ParaView website (https://www.paraview.org/): "ParaView users can quickly build visualizations to analyze their data using qualitative and quantitative techniques. The data exploration can be done interactively in 3D or or programmatically using ParaView's batch processing capabilities. ParaView was developed to analyze extremely large datasets using distributed memory computing resources. It can be run on

supercomputers to analyze datasets of petascale size as well as on laptops for smaller data, has become an integral tool in many national laboratories, universities and industry, and has won several awards related to high performance computation."

VisIT has been developed at Livermore National Laboratory. As stated on the VisIT website (https://wci.llnl.gov/simulation/computer-codes/visit): "VisIT is an open source, interactive, scalable, visualization, animation, and analysis tool. From Unix, Windows, or Mac workstations, users can interactively visualize and analyze data ranging in scale from small ($<10^1$ cores) desktop-sized projects to large ($>10^5$ cores) leadership-class computing facility simulation campaigns. Users can quickly generate visualizations, animate them through time, manipulate them with a variety of operators and mathematical expressions, and save the resulting images and animations for presentations. VisIt contains a rich set of visualization features to enable users to view a wide variety of data including scalar and vector fields defined on two- and three-dimensional (2D and 3D) structured, adaptive and unstructured meshes. Owing to its customizeable plugin design, VisIt is capabable of visualizing data from over 120 different scientific data formats."

1.2.4 GDSA Workflow

Uncertainty Quantification workflows are not trivial to define and get running, even when using tools such as Dakota to generate nested studies involving sampling loops over both aleatory and epistemic samples. The analyst has to spend a significant amount of time writing scripts to interface the sample values to PFLOTRAN, extract the results, and put the entire workflow on a high-performance computing platform. Further, as the quantities of interest increase and many vectors of results are generated for each simulation, plotting and aggregating the results in a variety of ways (e.g. averaging over epistemic or aleatory slices as a function of time) becomes very involved. To address this, the GDSA workflow was developed this year. This workflow couples Dakota, PFLOTRAN, and SAW (the Sandia Analysis Workflow software) to present the user with a unified GUI where the actual workflow can be dictated in an easy-to-use graphical format. This workflow also allows greater reproducibility and traceability of the actual files and scripts used for a particular study. This year, we demonstrated the UQ workflow for a GDSA study involving both aleatory and epistemic sampling. This is further discussed in Chapter 5.

1.2.5 GDSA Software Strategy

The software strategy for GDSA is to leverage and use open source software that is actively maintained and developed, whenever possible. That is why the *GDSA Framework* utilizes PFLOTRAN, Dakota, Paraview, and VisIT. Another goal is to support HPC computing, which is a primary focus for all of the software tools listed above. In addition, the *GDSA Framework* should have the flexibility to develop and adopt new capabilities as state-of-the-art hardware, software, and methodology evolves. Again, the codes chosen for GDSA exhibit this flexibility and are constantly evolving and adopting to utilize new software and hardware capabilities. Much of the interfacing between the codes is currently performed with Python scripts and other scripting tools, and there will be tighter integration as *GDSA Framework* and the GDSA workflow progresses.

1.3 References: Chapter 1

1. Swiler, L.P., J.C. Helton, E. Basurto, D.M. Brooks, P. Mariner, L.M. Moore, S. Mohanty, S.D. Sevougian, and E.R. Stein. "Status Report on Uncertainty Quantification and Sensitivity Analysis Tools in the Geologic Disposal Safety Assessment (GDSA) Framework." SAND2019-13835 R. Sandia National Laboratories, Albuquerque, NM.
2. Mariner, P. E., E. R. Stein, S. D. Sevougian, L. J. Cunningham, J. M. Frederick, G. E. Hammond, T. S. Lowry, S. Jordan, and E. Basurto (2018). Advances in Geologic Disposal Safety Assessment and an Unsaturated Alluvium Reference Case. SFWD-SFWST-2018-000509; SAND2018-11858R. Sandia National Laboratories, Albuquerque, NM.
3. Mariner, P. E., E. R. Stein, J. M. Frederick, S. D. Sevougian, and G. E. Hammond (2017). Advances in Geologic Disposal System Modeling and Shale Reference Cases. SFWD-SFWST-2017-000044 / SAND2017-10304R. Sandia National Laboratories, Albuquerque, NM.
4. Mariner, P. E., E. R. Stein, J. M. Frederick, S. D. Sevougian, G. E. Hammond, and D. G. Fascitelli (2016). Advances in Geologic Disposal System Modeling and Application to Crystalline Rock. FCRD-UFD-2016-000440 / SAND2016-9610R. Sandia National Laboratories, Albuquerque, NM.
5. Mariner, P. E., W. P. Gardner, G. E. Hammond, S. D. Sevougian, and E. R. Stein (2015). Application of Generic Disposal System Models. FCRD-UFD-2015-000126 / SAND2015-10037R. Sandia National Laboratories, Albuquerque, NM.
6. DOE (2008). Yucca Mountain Repository License Application Safety Analysis Report. DOE/RW-0573, Revision 1, http://www.nrc.gov/waste/hlw-disposal/yucca-lic-app/yucca-lic-app-safety-report.html#1. US Department of Energy, Washington, DC.
7. DOE (2014). Title 40 CFR Part 191 Subparts B and C Compliance Recertification Application 2014 for the Waste Isolation Pilot Plant. DOE/WIPP 15-3503. U.S. Department of Energy, Carlsbad, New Mexico.
8. Lichtner, P. C. and G. E. Hammond (2012). Quick Reference Guide: PFLOTRAN 2.0 (LA-CC-09-047) Multiphase-Multicomponent-Multiscale Massively Parallel Reactive Transport Code. LA-UR-06-7048. December 8, 2012. Los Alamos National Laboratory, Los Alamos, New Mexico.
9. Hammond, G. E., P. C. Lichtner, R. T. Mills, and C. Lu. (2008). Toward petascale computing in geosciences: application to the Hanford 300 Area. In R. L. Stevens (Ed.), Scidac 2008: Scientific Discovery through Advanced Computing (Vol. 125).
10. Hammond, G. E., P. C. Lichtner, and R. T. Mills (2014). "Evaluating the performance of parallel subsurface simulators: An illustrative example with PFLOTRAN". Water Resources Research, 50(1), 208-228. doi: 10.1002/2012wr013483.
11. Sevougian, S. D., G. E. Hammond, P. E. Mariner, E. R. Stein, J. M. Frederick, and R. J. MacKinnon (2018). "GDSA Framework: High-Performance Safety Assessment Software to Support the Safety Case." in Proceedings of the IGSC Safety Case Symposium 2018, Rotterdam, The Netherlands, October 10-11, 2018, SAND2018-9975C, Sandia National Laboratories. Albuquerque, NM.

12. Adams, B.M., Bohnhoff, W.J., Dalbey, K.R., Ebeida, M.S., Eddy, J.P., Eldred, M.S., Hooper, R.W., Hough, P.D., Hu, K.T., Jakeman, J.D., Khalil, M., Maupin, K.A., Monschke, J.A., Ridgway, E.M., Rushdi, A.A., Seidl, D.T., Stephens, J.A., Swiler, L.P., and Winokur, J.G., "Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.12 User's Manual," Sandia Technical Report SAND2020-5001, May 2020.
13. Helton J.C. and M.G. Marietta (Editors). 2000. Special Issue: The 1996 Performance Assessment for the Waste Isolation Pilot Plant. *Reliability Engineering and System Safety* 69:1-451.
14. Helton J.C., C.W. Hansen and P.N. Swift (Editors). 2014. Special Issue: Performance Assessment for the Proposed High-Level Radioactive Waste Repository at Yucca Mountain, Nevada. *Reliability Engineering and System Safety* 122:1-456.

2. NEW UQ METHODOLOGY: MULTILEVEL/MULTIFIDELITY METHODS

Repository performance assessment requires the repeated evaluation of complex, high-fidelity models that simulate million-year time horizons. The high computational cost of these models limits the number of simulations that can be performed in an uncertainty analysis, which in turn limits the accuracy of summary statistics and statistical analyses using such few samples. The emerging UQ methodology of multifidelity/multilevel methods can help to mitigate the effects of these limitations from computational cost by deploying lower-fidelity, less-resolved models which are still well-correlated with the high-fidelity model used for performance assessment.

The remainder of this chapter proceeds as follows. In Section 2.1, multilevel methods are briefly introduced. In Section 2.2, multifidelity polynomial chaos expansions are briefly introduced. Finally, in Section 2.3 these two methods are demonstrated on a benchmark problem in PFLOTRAN that mimics some of the properties of models for performance assessment to show the potential utility.

2.1 Multilevel Methods

Key statistics of interest such as the mean concentration of a radionuclide or its .95 quantile involve the numerical evaluation of an integral. The main cost is typically in calculating a numerical solution to a system of ordinary or partial differential equations which is used to obtain the quantities of interest. In the presence of uncertainty, this integration is typically performed using Monte Carlo integration, which is unbiased and has a rate of convergence that is independent of number of uncertain random variables. However, this rate of convergence is slow; the variance of an MC estimator decays as the square root of the number of samples used in the estimation. The aim of multilevel methods is to reduce the variance of the estimator without incurring additional high-fidelity model evaluations. This variance reduction is achieved by exploiting coarser discretizations of the model.

The key concept behind multilevel UQ methods is introduced here in the context of two model discretizations, which are denoted *f* for fine and *c* for coarse. The quantity of interest (the integrand) is denoted Q. Here the desired statistic is $\mathbb{E}[Q_f]$, which is the expectation of the quantity of interest at the finest level. This expectation can be approximated using Monte Carlo integration and evaluations of the finest model discretization as

$$\hat{Q}_f = \frac{1}{N}\sum_{i=1}^{N} Q_f^{(i)}. \tag{2.1}$$

The expected value of $\hat{Q}_f$ is $\mathbb{E}[Q_f]$, and its variance is $\mathbb{V}(Q_f)/N$. Note that in practice $\mathbb{V}(Q_f)$ is unknown and must be additionally approximated using a sample variance. When N is small, this variance can be quite large, meaning the estimated mean may be inaccurate. Instead, one can consider the multilevel (ML) estimator, $\hat{Q}_f^{ML}$, where the multilevel estimator is written as the mean of the coarse estimator plus the mean of the difference between the coarse and fine estimators:

$$\hat{Q}_f^{ML} = \hat{Q}_c + \widehat{Q_f - Q_c} = \frac{1}{N_c}\sum_{i=1}^{N_c} Q_c^{(i)} + \frac{1}{N_{ML}}\sum_{j=1}^{N_{ML}}\left(Q_f^{(j)} - Q_c^{(j)}\right). \tag{2.2}$$

Note that in the evaluation of Equation 2.2., we assume that the number of samples of the coarse model used to estimate the first term is much higher than the number of multilevel samples used to estimate the second term, which is governed by the cost to run the fine model: $N_c \gg N_{ML}$.

The expected value of $\hat{Q}_f^{ML}$ is

$$\mathbb{E}\left[\hat{Q}_f^{ML}\right] = \mathbb{E}\left[\hat{Q}_c\right] + \mathbb{E}\left[\widehat{Q_f - Q_c}\right] = \mathbb{E}[Q_c] + \mathbb{E}\left[Q_f\right] - \mathbb{E}[Q_c] = \mathbb{E}\left[Q_f\right], \tag{2.3}$$

which is the target quantity to be estimated. However, the variance of this multilevel estimator is

$$\mathbb{V}\left[\hat{Q}_f^{ML}\right] = \frac{\mathbb{V}[Q_c]}{N_c} + \frac{\mathbb{V}[Q_f - Q_c]}{N_{ML}}. \tag{2.4}$$

The benefit of the multilevel estimator is that the variance of $\mathbb{V}[Q_f - Q_c]$ can be much smaller than $\mathbb{V}\left[Q_f\right]$ alone, if Q_f and Q_c vary similarly with respect to changes in the random inputs. This means that the contribution of the term $\mathbb{V}\left[Q_f - Q_c\right]$ to the variance of the estimator can be much smaller for the same number of simulations at the fine level. There is an additional contribution to the multilevel estimator's variance from the coarse level evaluations. The contribution of the first term in Equation 2.3 can be made small by driving the variance down with a large number of samples N_c at the coarse level, which is presumably possible because the coarse model is cheaper to run.

Only two discretization levels were considered to describe multilevel methods above, but it is common to employ several discretization levels in a multilevel estimator. These discretization levels are typically indexed by ℓ, with $\ell = 0$ being the coarsest level and $\ell = L$ being the finest level. The same concept above applies, except now there is a telescoping difference between a discretization level ℓ and the next finer level $\ell + 1$. The general multilevel estimator of the quantity of interest Q at the finest level L can be thus written as

$$\begin{aligned} \hat{Q}_L^{ML} &= \frac{1}{N_0}\sum_{i=1}^{N_0} Q_0^{(i)} + \frac{1}{N_1}\sum_{i=1}^{N_1}\left(Q_1^{(i)} - Q_0^{(i)}\right) + \frac{1}{N_2}\sum_{i=1}^{N_2}\left(Q_2^{(i)} - Q_1^{(i)}\right) + \cdots \\ &= \frac{1}{N_0}\sum_{i=1}^{N_0} Q_0^{(i)} + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{i=1}^{N_\ell}\left(Q_\ell^{(i)} - Q_{\ell-1}^{(i)}\right). \end{aligned} \tag{2.5}$$

This approach can be especially powerful in the regime where mesh convergence is observed, as the variance of the difference between respective discretizations decays rapidly. A demonstration of multilevel Monte Carlo (MLMC) for the mean of a tracer concentration at an outflow boundary is presented in Subsection 2.3.2. Further reading on the theory and application of multilevel Monte Carlo methods includes [1,2,3].

2.2 Multifidelity Polynomial Chaos Expansion

Polynomial chaos expansions (PCEs) have gained popularity in uncertainty analyses in recent years and have been documented in detail in [4]. PCE is a stochastic expansion method whereby the output response is expanded in a series of polynomials that are orthogonal with respect to the densities of the random inputs. One advantage of stochastic expansion methods such as PCE is

that the moments of the expansion (means, variances, etc.), as well as Sobol' indices for global sensitivity analysis, can be written analytically.

Coefficients for the PCE are determined either by projecting model responses onto the basis, or by solving a regression problem. In either case, many model evaluations are required. The number of evaluations needed increases with the complexity of the model response, as more terms are needed in the PCE to resolve it. In scenarios where multiple fidelities of model exist, for instance models with multiple discretizations, or models with simplified or highly-resolved physics, the number of evaluations of the highest-fidelity model required to construct a PCE representation of its response can be reduced. This is achieved using a so-called "multifidelity PCE," first introduced in a paper by Ng and Eldred in 2012 [5].

As in Section 2.1, consider two levels of model discretization, indexed by f for fine and c for coarse. To simplify discussion, the coarse and fine models are assumed to depend on a single uncertain parameter ξ. The goal is to derive a PCE, denoted χ_f, for the response of the high-fidelity fine model, denoted R_f. In a single-fidelity framework, the high-fidelity model is sampled multiple times and a PCE is constructed using those evaluations:

$$R_f(\xi) \approx \chi_f(\xi) = \beta_0 + \sum_{i=1}^{p} \beta_i \Psi_i(\xi). \tag{2.6}$$

On the other hand, in a multifidelity PCE using two model discretizations, a PCE is constructed for the coarse model response R_c and another PCE is constructed for the discrepancy between the coarse and fine model, $\Delta_{f,c}$. This discrepancy may be defined as the difference or the ratio between the two discretizations, or a combination of the two.

In this work, only the difference between the two levels is considered. In this setting, a multifidelity PCE is derived by constructing a PCE for the coarse model, then for the discrepancy between coarse and fine, $\Delta_{f,c} \equiv R_f - R_c$:

$$R_f(\xi) \approx \chi_c(\xi) + \chi_\Delta(\xi) = \left[\beta_{0,c} + \sum_{i=1}^{p_c} \beta_{i,c}\Psi_i(\xi)\right] + \left[\beta_{0,\Delta} + \sum_{i=1}^{p_\Delta} \beta_{i,\Delta}\Psi_i(\xi)\right]. \tag{2.7}$$

The hope in employing this method is that R_c is similar enough to R_f that the difference between the two responses is simpler than R_f alone. In cases where this is true, fewer terms are required to resolve the PCE of the discrepancy $\Delta_{f,c}$ than to resolve the PCE of the high-fidelity model directly. The reduction in the number of required terms may manifest in the form of a rapid decay of coefficients for χ_Δ or in the form of sparsity in the number of terms that are excited.

When this occurs, the coarse PCE χ_c will likely have many more terms than the discrepancy PCE χ_Δ, requiring more evaluations at the coarse level. It is presumed that this is feasible because the coarse model is cheap to run. The coefficients of the PCEs can be easily recombined into a single PCE by combining like terms. The combined PCE is written

$$R_f(\xi) \approx (\beta_{0,c} + \beta_{0,\Delta}) + \sum_{i=1}^{p_\Delta}(\beta_{i,c} + \beta_{i,\Delta})\Psi_i(\xi) + \sum_{i=p_\Delta+1}^{p_c} \beta_{i,c}\,\Psi_i(\xi), \tag{2.8}$$

where it is assumed that $p_c > p_\Delta$. This combined PCE can be used in any of the ways a PCE constructed in a single-fidelity fashion would be used, as a surrogate or to compute response

statistics or Sobol' indices. Only two models were considered in the discussion above, but a multifidelity PCE can be constructed using an arbitrary number of models. The multifidelity PCE can be constructed using a series of discrepancies between discretization levels, indexed from $\ell = 0$ at the coarsest level to $\ell = L$ at the finest level, as

$$R_L \approx \chi_0 + \sum_{l=1}^{L} \chi_{\Delta_{\ell,\ell-1}}. \tag{2.9}$$

A demonstration of the construction and use of a multifidelity PCE for global sensitivity analysis (GSA) is presented in Subsection 2.3.3.

2.3 SX-115 Hanford Tank Farm Example

In this section the multilevel and multifidelity techniques for uncertainty quantification described in Sections 2.1 and 2.2 are applied to one of the benchmark problems in PFLOTRAN, the SX-115 Hanford Tank Farm. The physical problem is described in Subsection 2.3.1, multilevel Monte Carlo is used for forward uncertainty propagation in Subsection 2.3.2, and a multifidelity PCE is constructed and used for global sensitivity analysis in Subsection 2.3.3.

2.3.1 Physical problem description

The Hanford Tank Farm contains nuclear waste storage tanks and is located in south-central Washington. A number of leaks have occurred at the site, prompting the study by Lichtner et al. in 2004 [6] on which a benchmark problem for PFLOTRAN is based. The benchmark problem simulates a leak from one of the tanks (the SX-115 tank) in a 1D vertical column of layered porous medium with properties chosen to be representative of the earth at the Hanford site. The column extends from ground level at 68 m to the water table at 0 m. A steady-state infiltration rate is imposed at the top of the column. Fluid is injected for two weeks during the simulation at a depth of 16m to represent a leak occurring at the bottom of the tank. The leak then travels downward through the column to the water table at the bottom of the domain.

A detailed description of the benchmark problem and its parametrization is available in the PFLOTRAN user manual [7] and in the online documentation at the time of writing: https://documentation.pflotran.org/. The problem setup described in the user manual was used with almost no modification for the uncertainty analyses in this chapter. Any changes made to the problem setup are detailed here. For this work the leak is represented by a passive tracer. A typical quantity of interest (QoI) for performance assessment is the concentration as a function of time of a radionuclide at a point on the outflow boundary. To mimic this type of QoI, an observation point was placed at the bottom of the domain, the water table, to track the time-varying concentration of the tracer. The infiltration rate was modified to 5.0e-10 m/sec from 2.53678e-10 m/sec (corresponding to 8mm/yr) to shorten the timescales of the problem, allowing each simulation to run for less time.

The original problem setup specified the leak as a point source from the REGION 'well.' This was specified in the following input blocks from the original problem specification:

```
REGION well
 COORDINATES
  1.d0 1.d0 52.d0
  1.d0 1.d0 52.d0
 /
```

```
FLOW_CONDITION source
 TYPE
  RATE mass_rate
 /
 …
```

This specification is dependent on the discretization of the problem. This is because fluid is injected into each cell in the region at the specified mass rate. As the grid is refined in the *z* direction, more cells occupy the region compared to the initial discretization and cause more fluid to be injected with each refinement. To account for the different discretizations that will be used in the multilevel/multifidelity studies to come, the 'well' region was modified to be a small volume and the source was modified to inject the fluid at a scaled rate so that the same volume of fluid was injected into the volume regardless of how many cells were in the region. These modified input blocks are below.

```
REGION well
 COORDINATES
  1.d0 1.d0 52.d0
  1.d0 1.d0 52.5d0
 /
```

```
FLOW_CONDITION source
 TYPE
  RATE scaled_mass_rate volume
 /
 …
```

All other parameters in the problem are fixed at the values specified for the benchmark problem, except for the permeabilities of each layer of the porous medium, which are assumed uncertain for the purposes of this study. Their distributions will be defined later in the discussion. Variations in permeability affect the steady-state pressure and saturation profiles before the leak is introduced. To account for this, the steady-state and leak runs have been combined into one simulation. The times in the problem have been modified so that the simulation runs for 300 years without any leak to allow the pressure and saturation to come to steady state. At 300 years, the leak occurs for two weeks, then the simulation runs to 600 years, until the tracer reaches a significant concentration at the outflow boundary

A convergence study was performed for the maximum timestep used in the solver for the original discretization of 136 elements in the z-direction ($N_z = 136$) and the nominal values for the permeability in each layer. The following timestep specification was required to see convergence in the dynamics of the problem.

```
TIME
# 300.0 y to run to steady state,
# another 300 y for the leak.
FINAL_TIME 600.d0 y
INITIAL_TIMESTEP_SIZE 1.d-6 y
MAXIMUM_TIMESTEP_SIZE 1.d-2 y
MAXIMUM_TIMESTEP_SIZE 1.d-1 y at 10 y
# Starting over for transience after the leak
MAXIMUM_TIMESTEP_SIZE 1.d-6 y at 300 y
MAXIMUM_TIMESTEP_SIZE 1.d-2 y at 300.000001 y
MAXIMUM_TIMESTEP_SIZE 1.d-1 y at 310 y
END
```

A mesh convergence study was performed in terms of the discretization in the z direction, parameterized in terms of the number of discretization points in z, denoted N_z. As the mesh resolution increases, mesh convergence was observed for N_z = 1088 and N_z = 2176, as shown in Figure 2-1. The Newton solver used in these calculations for each implicit timestep terminates when one of three convergence tolerances are met. These are in terms of the 2-norm of the residual; the 2-norm of the residual at the current iteration divided by the initial residual; and the 2-norm of the difference between the current solution minus the solution at the previous iteration, divided by the solution at the previous iteration. A nonphysical discontinuous step occurred in the tracer concentration for the finer meshes using the tolerances for the Newton solver specified in the benchmark problem.

To mitigate this issue, the tolerances for the flow Newton solver were lowered to 1e-50 and were set to the default tolerances for the transport Newton solver. Only the concentrations after 550 years were used for the following studies; concentrations before 550 years include (nearly) zero concentrations, which do not vary as a function of the model parameters and therefore are not of interest for UQ. Furthermore, the following uncertainty analyses were performed in terms of the natural logarithm of the concentration and of the uncertain permeabilities. This was motivated by previous observed challenges in [4] with using untransformed concentrations for GSA with PCEs.

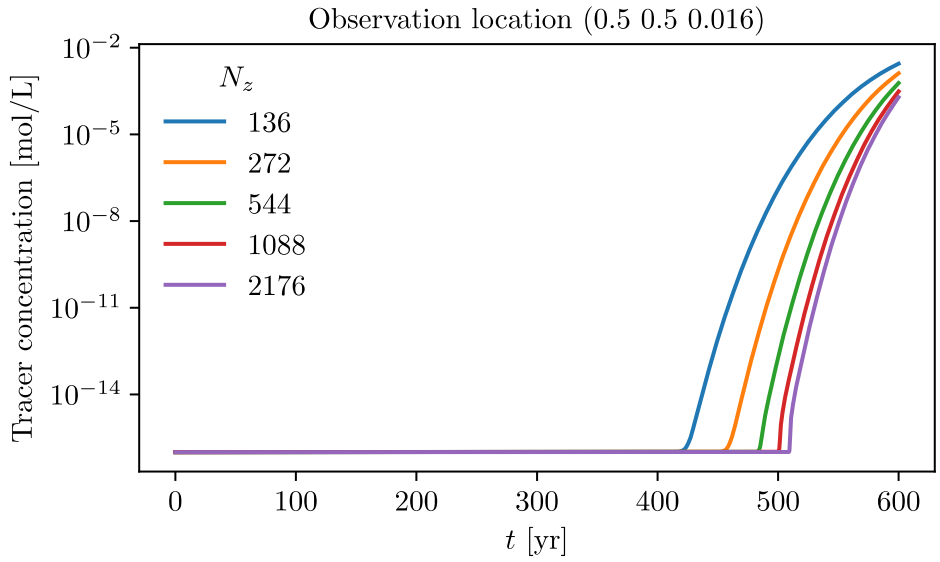


Figure 2-1. Time history of tracer concentration at the bottom of the domain for different discretizations.

The permeability of each layer of the porous medium is considered uncertain for this work. The permeabilities are assumed to be log-normally distributed. Their distributions were specified in terms of the underlying normal distribution of their log values. The means of the normal distribution are set to the natural logarithm of the nominal values for permeabilities specified in the benchmark problem. The standard deviations are set to 1% of the absolute value of the means.

2.3.1.1 Computational cost by discretization

Multifidelity (MF) and multilevel (ML) methods balance the computational cost of running the model at each discretization level with the relative contribution to estimator variance to determine how many samples are needed (see, e.g., [8,9] for different methods of distributing cost across discretizations/model fidelities). Because of this, it is necessary to estimate the relative cost of the models at each level. The absolute cost (in terms of floating point operations or linear solves in PFLOTRAN) is not necessary. The only thing needed is an estimate of how many times more expensive it is to evaluate finer discretizations of the model, relative to the coarsest level. For instance, in the case of a linear 1D problem that is solved directly, the relative cost of solving a problem of size $h/2$ compared to h is 2.

For this problem, PFLOTRAN employs a fully-implicit adaptive timestepper. Inverse solves at each timestep are performed using the iterative biconjugate gradient stabilized (BCGS) method, which runs until a desired tolerance is reached. The timestep changes throughout the simulation, but it is generally the largest timestep allowed by the user and for which the Newton solver can converge. Compared to direct solvers with fixed timestep, it is more challenging to assess the cost of running the model at different discretization levels. The cost metric used in this work is the number of linear solves required for simulation at each discretization level. The number of linear solves for the transport part of the solver was greatest, so it was the basis of comparison here. The relative cost of the simulations in the mesh convergence study were used to approximate the cost

of the model at each level, which are reported in Table 2-1. These relative costs are rounded up when they are used to evaluate equivalent costs for various ML/MF methods.

N_z	Linear transport solves	Cost relative to $N_z = 136$
136	990466	1
272	5097567	5.15
544	16729384	16.9
1088	28276950	28.55
2176	48003520	48.47

Table 2-1. Linear solves and relative cost as a function of mesh size.

Note that for this simple, 1D problem the cost rises approximately linearly with respect to discretization level, as shown in Figure 2-2. However, for practical applications it is more common to see superlinear increase in cost across discretization levels, potentially leading to more dramatic increase in computational efficiency through use of these methods.

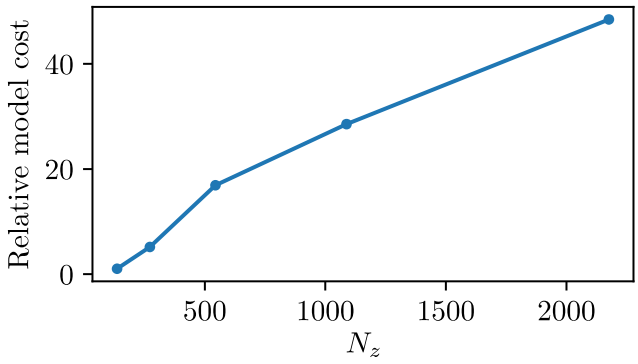


Figure 2-2. Relative model costs as a function of N_z.

2.3.1.2 Equivalent number of high-fidelity evaluations

A common metric to compare the cost of a multilevel/multifidelity scheme to a single-fidelity scheme using the high-fidelity model only is the equivalent number of high-fidelity evaluations a method incurs. This equivalent number of evaluations is determined by considering the number of samples taken at each discretization level and multiplying them by the fraction of their cost relative to the high-fidelity model. For instance, revisiting the two-discretization multilevel scheme, recall that the mean estimator is obtained as

$$\hat{Q}_f^{ML} = \hat{Q}_c + \widehat{Q_f - Q_c} = \frac{1}{N_c}\sum_{i=1}^{N_c} Q_c^{(i)} + \frac{1}{N_{ML}}\sum_{j=1}^{N_{ML}} \left(Q_f^{(j)} - Q_c^{(j)}\right). \quad (2.10)$$

In this case, the coarse model would be evaluated $N_c + N_{ML}$ times for the two different terms, and the fine model would be evaluated N_{ML} times. Another way to think about the cost is that N_c samples are obtained at a cost of C_c, the cost of the coarse model evaluation, and N_{ML} samples were obtained at the cost of $C_c + C_f$, the cost of a coarse and fine model solve. Then the equivalent high-fidelity cost of the total samples required to calculate Equation 2.9 is equal to

$$N_{equiv} = \frac{C_c}{C_f}(N_c + N_{ML}) + N_{ML} = \frac{C_c}{C_f}N_c + \frac{(C_c+C_f)}{C_f}N_{ML}. \quad (2.11)$$

This can be expanded to an arbitrary number of levels $\ell = 0, \dots, L$, with L being the finest level, as

$$N_{equiv} = \sum_{\ell=0}^{L-1}\frac{C_\ell}{C_L}(N_\ell + N_{\ell+1}) + N_L = \frac{1}{C_L}[N_0 C_0 + \sum_{\ell=1}^{L} N_\ell(C_\ell + C_{\ell-1})]. \quad (2.12)$$

2.3.2 Multilevel Monte Carlo for Forward Propagation

As described in Section 2.1, multilevel Monte Carlo was used to obtain an estimate of the mean (log) concentration as predicted by the high-fidelity ($N_z = 2176$) model. A convergence study was performed using the UQ software Dakota [10] with increasing reduction in the sampling variance relative to an initial sampling variance, generated from 20 pilot samples. These pilot samples are used to estimate the variance at each level and optimally allocate samples by level to minimize estimator variance with minimal cost. The algorithm for this allocation scheme is described in [11]. The algorithm reports how many samples were evaluated at each level, as well as the equivalent number of high-fidelity evaluations the method performed in terms of cost.

To compare performance of this estimator with respect to a standard, single-level Monte Carlo mean estimator, the sample mean and variance using the equivalent number of evaluations of the high-fidelity model were computed to correspond to each MLMC study. A plot of the sampling variance in the multilevel mean estimator compared to the single-level mean estimator as a function of equivalent high-fidelity evaluations is presented in Figure 2-3. Sample variance of the mean estimator for multilevel Monte Carlo compared to single level Monte Carlo, for increasing number of equivalent high-fidelity model evaluations. The values of the sampling variance are reported in Table 2-2. Because the model output is time-varying between 550 and 600 years, the sample mean and variance are time-varying as well. To report a scalar value for the convergence study the statistics discussed here are averaged over time. Note that, as expected, the sample variance using MLMC is less than the single-level MC for the same computational effort.

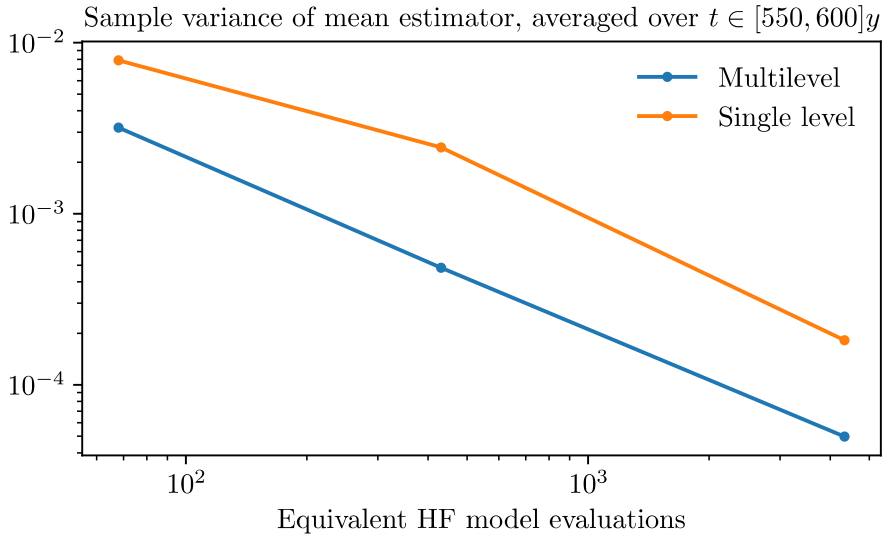


Figure 2-3. Sample variance of the mean log-concentration estimator for multilevel Monte Carlo compared to single level Monte Carlo, for increasing number of equivalent high-fidelity model evaluations.

Equivalent HF evaluations	MLMC sampling variance	MC sampling variance
68	3.19×10^{-3}	7.88×10^{-3}
431	4.83×10^{-4}	2.44×10^{-3}
4341	4.98×10^{-5}	1.82×10^{-4}

Table 2-2. The variance on the mean log-concentration estimator computed using MLMC and single-level MC for the same equivalent number of high-fidelity evaluations.

To compare the efficiency of the MLMC vs. standard MC estimator, the ratio of the time-averaged sampling variance of the respective mean estimators was computed and is reported in Table 2-3. Additionally, a percent variance reduction, computed as $100\% \times (1 -$ ratio of variances), is reported.

Equivalent HF evaluations	$\frac{\langle Estimator\ var, MLMC \rangle_t}{\langle Estimator\ var, SLMC \rangle_t}$	% Variance reduction
68	0.40	60%
431	0.20	80%
4341	0.27	73%

Table 2-3. The ratio of sample variance for the MLMC estimator compared to the single level MC (SLMC) estimator. The percent variance reduction is computed as 100% (1 – ratio of variances).

For this problem the reduction in variance is significant, with at least a 60% reduction and as much as 80% reduction over the range of equivalent high-fidelity evaluations that were used in sampling. This means that the estimation of the mean is significantly more accurate than using samples of the high-fidelity model alone, for the same amount of computational effort. The amount of variance reduction that is possible for a given problem is dependent on a combination of factors, namely the relative costs of the models at each level and the correlation between model levels.

Recall that

$$\mathbb{V}[Q_\ell - Q_{\ell-1}] = \mathbb{V}[Q_\ell] + V[Q_{\ell-1}] - 2\mathbb{C}ov[Q_\ell, Q_{\ell-1}]. \quad (2.13)$$

The more correlated two levels are, the smaller the variance of their difference will be. On the other hand, if the cost of obtaining a sample of their difference is not a significant reduction from the high-fidelity model, the equivalent number of high-fidelity model evaluations per sample increases, reducing the ability to drive the sampling variance down with more samples.

An alternative way of looking at these results is to consider the relative cost, in terms of equivalent high-fidelity evaluations, to achieve the same accuracy in the estimator. This was done by computing a linear fit to the number of high-fidelity evaluations as a function of the estimator variance (averaged over 550 to 600 years) and computing the ratio of costs for the same estimator variance. Using the results from the convergence study shown in Figure 2-3, it was found that the MLMC estimator could achieve the same accuracy with half the cost of the standard, single-level MC estimator.

Note that in this case the forward propagation studies were specified in terms of a convergence tolerance based on a reduction in the initial sampling variance of the mean. The MLMC algorithm kept adding samples across levels until the desired accuracy was achieved. However, in many practical problems, the number of samples that can be afforded, especially at the highest level, is limited. Recent works such as [1,8] discuss allocation strategies that assume a fixed number of high-fidelity evaluations and derive the sampling scheme that achieves the best accuracy within a given computational budget. This type of approach will be necessary for a practical application of these methods to GDSA and will be employed in future work.

2.3.3 Multifidelity Polynomial Chaos Expansion for Global Sensitivity Analysis

As described in Section 2.2, a multifidelity polynomial chaos expansion (MF PCE) was constructed to demonstrate its utility for global sensitivity analysis (GSA). While it is common to employ quadrature-based approaches for constructing PCEs, regression approaches have typically been employed in the context of GDSA because it allows for reuse of samples from forward propagation studies, and because of the high computational cost of quadrature as number of variables increases. With this in mind, all the results presented herein employ regression methods to construct the MF PCE.

MF PCEs and single-fidelity PCEs using high-fidelity evaluations were constructed. For both MF PCE and single-fidelity PCE, Dakota [10] was used for regression with an adaptive algorithm that increased the order of the expansion uniformly across all uncertain parameters until the L^2 norm of the change in the response covariance matrix fell below a specified convergence tolerance. For MF PCE the initial order of the PCE was 0; for the standard PCE expansion the initial order

was 2. The higher initial order for the standard PCE expansion was necessary because the algorithm diverged without cross validation activated; cross validation requires at least a starting order of 2. Both algorithms employed orthogonal matching pursuit with a collocation ratio of 0.9. These methods and options are detailed in the Dakota User's Manual [10] and the Dakota Theory Manual [11].

Prior to running the regression problems, the amenability of the problem to an MF PCE approach was studied using 64 pilot samples taken across all model levels. An MF PCE is built with a baseline PCE fit of the coarsest model level, then with additional PCE fits to the discrepancy between model levels, up to the finest level. The problem is fit for an MF PCE approach if relatively few coefficients can be used to represent the discrepancies between each of the model levels. This was tested by computing the difference between each of the model levels over the pilot samples and passing them as build points in a regression PCE study. It was found that few coefficients were needed to resolve each of the discrepancies, so the problem was deemed a good candidate for an MF PCE approach.

A table of the model evaluations required at each discretization level and the equivalent high-fidelity evaluations are reported in Table 2-4. Additionally, in Table 2-5, the convergence tolerances and model evaluations required for the single-fidelity PCE built on high-fidelity evaluations are reported. Again, note that the convergence tolerance is based on the L^2 norm of the change in the response covariance matrix, so there is no assumed equivalent accuracy between the MF PCE and the single-fidelity PCE constructed using the same convergence tolerance. Note that the actual number of model evaluations needed for the adaptive MF PCE algorithm ended up being much less than the 64 pilot samples taken at each level for the initial study. Instead, it could have been possible to use the pilot samples from the forward UQ propagation study in 2.3.2, rather than incurring additional model evaluations for a PCE-specific pilot study. This will be the approach taken in future.

Convergence tolerance	**Number of evaluations**					
	Equivalent HF	$N_z = 136$	$N_z = 272$	$N_z = 544$	$N_z = 1088$	$N_z = 2176$
10^{-1}	18.73	69	24	10	10	5
10^{-2}	38.45	69	38	38	24	5
10^{-3}	79.71	100	100	69	38	19

Table 2-4. The number of model evaluations at each level and equivalent number of high-fidelity evaluations to construct the MF PCE.

Convergence tolerance	**Number high-fidelity evaluations**
10^{-1}	50
10^{-2}	50
10^{-3}	713

Table 2-5. The number of high-fidelity model evaluations required for each convergence tolerance for the single-fidelity PCE.

Across all convergence tolerances and between MF PCE and single-fidelity PCE (SF PCE) the predicted mean and variances were strikingly similar, as shown in Figure 2-4. Note that the responses are so similar that they overlap completely. The Sobol' indices are also essentially identical, as shown in Figure 2-5—the indices are so similar that they are indistinguishable. Sensitivity of the response was dominated by *HF ksat*, which is the water saturated permeability for Hanford Fine Sand, the layer of earth just below the location of the leak in the column. This holds across MF/SF PCE and different convergence tolerances, as shown in Figure 2-5. Even the coarsest MF PCE, which incurred only ~19 equivalent high-fidelity evaluations, was consistent with the most-resolved SF PCE's Sobol' indices. The Sobol' indices for the permeabilities of the other layers in the simulation domain were also identical across type and convergence tolerance of PCE.

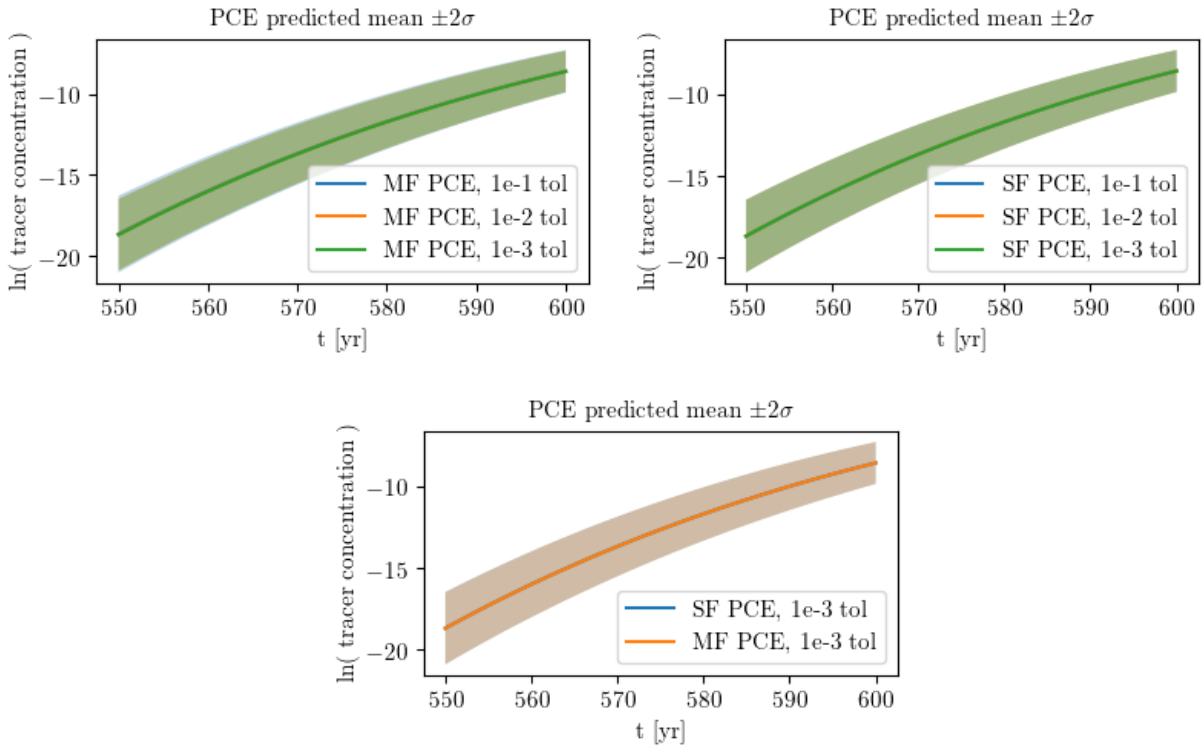


Figure 2-4. Mean log concentration $\pm$ 2σ for different convergence tolerances and MF/SF PCEs. Note that the responses are so similar that they overlap completely.

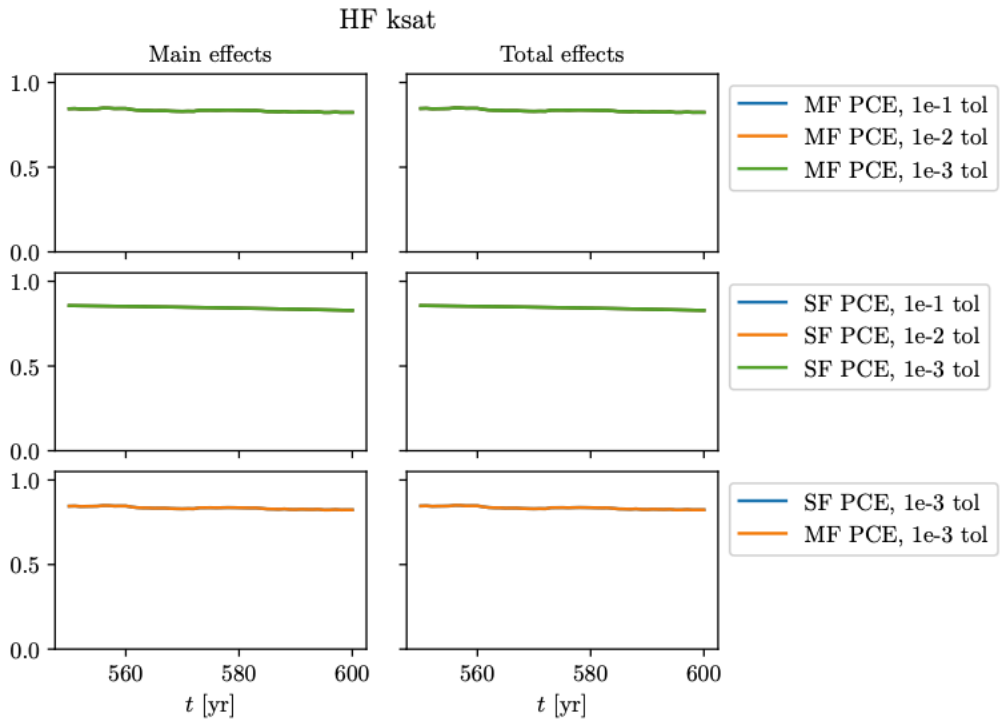


Figure 2-5. The main and total effect Sobol index for *HF ksat* as a function of time for different convergence tolerances of MF PCE, SF PCE, and comparing MF and SF PCE from the tightest tolerance runs. The indices are so similar that they overlap completely.

The relative invariance of these results to the convergence tolerance of the PCE algorithms is likely due to the simplicity in behavior of the model responses. As illustrated in Figure 2-6, only coefficients up to first order are significantly nonzero in all the cases considered, indicating that the variability of the model responses can be captured with only a first-order PCE. Any additional refinement of the PCE by adding more orders does not add any further accuracy.

Figure 2-6 only shows the coefficients for the PCE fit to model outputs at 550 years, but the coefficients over all times exhibited this behavior. Given the simplicity, very few model evaluations are required to adequately capture the model's variation, so even the lowest-order PCEs were sufficient in this case.

Even so, the potential benefits of employing an MF PCE approach are shown in the low number of equivalent high-fidelity evaluations required across all tolerances, and the low number of evaluations of the most refined model, maxing out at 19 evaluations total. Additionally, the computational gains compared to a Saltelli-based sampling method [12] for estimating the indices can't be overstated. Even a rough estimate of the Sobol' indices for the five permeability parameters using 100 samples would require $(100 \text{ samples}) \times (5 \text{ parameters} + 2) = 700$ model evaluations. In a practical performance assessment study, even a rough estimate using such methods could be computationally infeasible using the high-fidelity model only.

A focus of future work will be to employ MF PCE for a problem with a more complex dependence of model outputs on uncertain inputs. This will result in a more challenging fitting problem and enable a more detailed study of the relative benefits and downsides of a MF PCE vs SF PCE approach for problems of interest in NW performance assessment. Additionally, methods by which existing HF samples can be augmented to create a regression problem for an MF PCE will be explored. This is motivated by the typical situation where only a limited number of high-fidelity model simulations can be run due to computational cost, thus precluding adaptive algorithms that terminate based on a convergence tolerance.

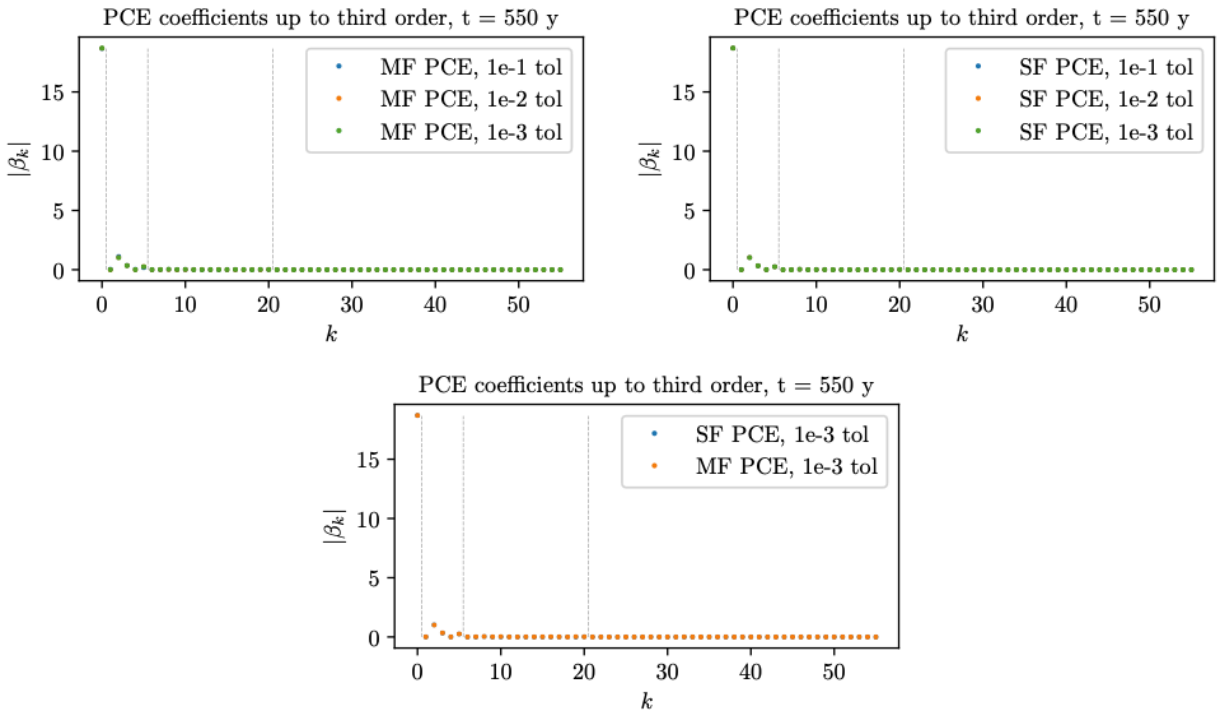


Figure 2-6. Modulus of PCE coefficients β_k (indexed by k) up to third order, compared across convergence tolerance for MF/SF PCEs, and comparing the tightest-tolerance MF and SF PCE coefficients. The vertical lines delineate changes in the order of the term with which each coefficient is associated, with the first vertical line separating the 0th and 1st order coefficients. Note that the coefficients overlap almost completely for the different cases.

2.4 Conclusions and Future Work

In this chapter multilevel Monte Carlo (MLMC) and multifidelity PCEs were introduced and demonstrated on a benchmark tracer transport problem in PFLOTRAN. The MLMC methods exhibited a significant decrease in estimator variance for the mean of model outputs, exhibiting 60-80% decrease in estimator variance compared to a standard MC using an equivalent number of high-fidelity model evaluations. The MF PCE produced Sobol' indices and predicted mean and variance of model outputs that were consistent with those of similarly-constructed single-fidelity PCEs built off of high-fidelity model evaluations alone. The MF PCEs overall required very few equivalent high-fidelity model evaluations, but the simplicity of the behavior of model outputs

with respect to uncertain inputs was very simple for this problem, so the method wasn't challenged very much. Care will be taken to ensure the next application problem has a more complex dependence on uncertain inputs, thereby providing more of a challenge to the MF PCE algorithms.

Future work will focus on identifying and applying these methods to a more challenging problem that exhibits more of the properties and challenges that arise in practical NW performance assessment studies. Properties that will be targeted are: a mixture of aleatory and epistemic uncertainties, no explicit discretization hierarchy to rely on, and additional computational cost. Additionally, best practices for employing multifidelity methods within a fixed computational budget will be determined, as this will typically be the scenario for GDSA studies. Focusing on these aspects will ensure that the methods and infrastructure that are developed can be successfully extended to a large-scale production study without too much modification.

2.5 References: Chapter 2

1. G. Geraci, M. S. Eldred, and G. Iaccarino. A multifidelity multilevel monte carlo method for uncertainty propagation in aerospace applications. In *19th AIAA Non-Deterministic Approaches Conference,* 2017.
2. M. B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.
3. M. B. Giles. Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, 2015.
4. Swiler, L. P., Helton, J. C., Basurto, E., Brooks, D. M., Mariner, P., Moore, L. M., Mohanty, S., Sevougian S.D., & Stein, E. (2019). *Status Report on Uncertainty Quantification and Sensitivity Analysis Tools in the Geologic Disposal Safety Assessment (GDSA) Framework* (No. SAND2019-13835R). Sandia National Lab. (SNL-NM), Albuquerque, NM (United States); Sandia National Laboratories, Las Vegas, NV.
5. Ng, L. W. T., & Eldred, M. (2012). Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA* (p. 1852).
6. Lichtner, P. C., Yabusaki, S., Pruess, K., & Steefel, C. I. (2004). Role of competitive cation exchange on chromatographic displacement of cesium in the vadose zone beneath the Hanford S/SX tank farm. *Vadose Zone Journal*, *3*(1), 203-219.
7. Lichtner, P. C., Hammond, G. E., Lu, C., Karra, S., Bisht, G., Andre, B., ... & Kumar, J. (2015). *PFLOTRAN user manual: A massively parallel reactive flow and transport model for describing surface and subsurface processes* (No. LA-UR-15-20403). Los Alamos National Lab.(LANL), Los Alamos, NM (United States); Sandia National Lab.(SNL-NM), Albuquerque, NM (United States); Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States); Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States); OFM Research, Redmond, WA (United States).

8. A. A. Gorodetsky, G. Geraci, M. Eldred, and J. D. Jakeman. A generalized framework for approximate control variates. *arXiv preprint* arXiv:1811.04988v2, 2018.
9. B. Peherstorfer, K. Willcox, and M. Gunzburger. Optimal model management for multifidelity Monte Carlo estimation. *SIAM Journal on Scientific Computing*, 38(5):A3163–A3194, 2016
10. Adams, B.M., Bohnhoff, W.J., Dalbey, K.R., Ebeida, M.S., Eddy, J.P., Eldred, M.S., Hooper, R.W., Hough, P.D., Hu, K.T., Jakeman, J.D., Khalil, M., Maupin, K.A., Monschke, J.A., Ridgway, E.M., Rushdi, A.A., Seidl, D.T., Stephens, J.A., Swiler, L.P., and Winokur, J.G., "Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.12 User's Manual," Sandia Technical Report SAND2020-5001, May 2020.
11. Dalbey, K.R., Eldred, M.S., Geraci, G., Jakeman, J.D., Maupin, K.A., Monschke, J.A., Seidl, D.T., Swiler, L.P., Tran, A., Menhorn, F., and Zeng, X., "Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.12 Theory Manual," Sandia Technical Report SAND2020-4987, May 2020.
12. Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M. and Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer physics communications,* 181(2), pp.259-270.

3. DISCRETE FRACTURE NETWORKS

Discrete fracture network (DFN) modeling is an alternative to continuum approaches for simulating flow and transport through sparsely fractured rocks in the subsurface [1]. Continuum approaches use effective parameters to include the influence of fractures on the flow. In contrast, a DFN approach involves a network of fractures where the geometry and properties of individual fractures are explicitly represented as lines in two dimensions or planar polygons in three dimensions. These generated networks are then meshed for computation. dfnWorks [1] is a parallelized computational suite developed at Los Alamos National Laboratory to generate three-dimensional discrete fracture networks and simulate flow and transport. The work presented in this report uses the "dfnGen" capability of dfnWorks to generate the networks, with the flow and transport calculations run in PFLOTRAN. This capability was used to generate 20 discrete fracture networks for a crystalline reference case, which is based off properties at the Forsmark site in Sweden [9], but it not a direct replica of Forsmark. They were then converted to equivalent porous media (ECPMs) for flow and transport simulation with PFLOTRAN. These 20 DFNs are used in the uncertainty analysis presented in Chapter 4.

In dfnWorks, one defines various layers in a 3-D domain and one can define various fracture families to be placed in each layer. The fractures in each family may be specified as rectangles or ellipses. This work uses circular plates, defined by radii which vary according to a truncated power-law distribution. The user can define characteristics such as aspect ratio of the ellipses per fracture family, and various angles describing the ellipse fracture orientation such as the angle the normal vector makes with the z-axis. The orientation of the ellipses is characterized by a 3-dimensional Fisher distribution describing the orientation of the fracture poles. The fractures may be placed at specified coordinates but typically are placed randomly throughout the layer. There are many user controls governing the generation of fractures in dfnWorks: the keywords are described in the user documentation: https://dfnworks.lanl.gov/dfngen.html. Finally, we note that dfnWorks uses an underlying graph to represent the fractures. For example, a fracture graph assigns fractures to nodes and intersections to edges. This is a computationally efficient way to represent and store the network.

There are several fracture domains at Forsmark, each containing fracture sets with different statistical properties. For our implementation of fractures for the crystalline reference case, we define three fracture families and three depth intervals, which are a subset of the fractures sets and domains defined at the Forsmark site. The selection of the applicable fracture domains and sets to include and other simplifying assumptions are detailed in [6]. Fracture domains FF01/06 were used because they constituted the majority of the Forsmark site and contained the repository. The fracture sets NS, NE, HZ were selected because they had the largest number of open and flowing fractures for depths [-200 m, -400 m], according to analysis reported in [5]. A random fracture network is generated within these defined zones such that the fracture density decreases with increasing depth (see Table 3-1). Each fracture set has been characterized using a 3-dimensional Fisher distribution describing the orientation of fracture poles and a truncated power-law distribution for the fracture radii. Each set is further characterized by the fracture density, P_{32}, which is defined as the area of fractures per volume of rock (m^2/m^3).

The columns of Table 3-1 are as follows: the layers are defined by depth (meters below sea level) and fracture set type (NS=North/South, NE=North/East, HZ=horizontal). The fracture

orientation is defined by a Fisher distribution which describes the orientation in space (in a sphere) of the poles normal to the fracture planes. Trend and plunge are geospatial terms used to describe the orientation of a pole or lineament. The trend is the degrees from North, meaning the compass direction. The plunge is the degrees from horizontal. Note that trend and plunge define the orientation of normal vector to the plane of the fracture. This means, for example, that the 0.0° plunge for the NS and NE fracture sets means that they are located in a vertical orientation on average and the 90.0° plunge for the HZ fracture set means that it is located horizontally on average. Kappa (κ) is "concentration" (a shape parameter) which describes how tightly the distribution is clustered about the mean: larger values means the normal vectors of the ellipses are more similar.

The parameters governing the truncated power law distribution which generates the radii of the elliptical fractures are as follows: k is a rate parameter, r_0 is the minimum fracture radius, and r_{upper} is the maximum fracture radius used in the truncated power law in Equation 3.4. The fracture density can be described by the number of target fractures per cubic kilometer or by the target P_{32} value (area of fractures per volume of rock in units of m^2/m^3) which is defined by Equation 3.2 and discussed in Section 3.1 below.

After the 20 DFNs for the crystalline reference case were generated with dfnWorks, they were mapped to the equivalent continuous porous medium domain using mapDFN.py [11], a code that approximates hydraulic fracture properties by calculating and assigning permeability and porosity on a cell-by-cell basis. Conversion to equivalent continuous porous media (ECPMs) is described in [6] and further depends on a fracture radius-transmissivity relationship. An example of a particular DFN used for uncertainty analysis is shown in Figure 4-2 in Chapter 4.

Depth (meters below sea level) / Fracture Set Name	**Orientation: Fisher Distribution for Poles**			**Size: Truncated Power Law for Radii**			**Fracture Density (Requested)**	**P_{32} value defined in Ref. [2]**
	Mean Trend	Mean Plunge	κ	Power Law k	Min Radius r_0 (m)	Max Radius r_{upper} (m)	Number of fractures in 1 km^3	$P_{32}, r \in [0.038,564]$ $[m^2/m^3]$
0-200 / NS	90°	0.0°	22	2.5	30	500	184	0.073
0-200 / NE	180°	0.0°	22	2.7	30	500	274	0.319
0-200 / HZ	360°	90.0°	10	2.4	30	500	2217	0.543
200-400/ NS	90°	0.0°	22	2.5	30	500	357	0.142
200-400/NE	180°	0.0°	22	2.7	30	500	296	0.345
200-400 / HZ	360°	90.0°	10	2.4	30	500	1290	0.316
>400 / NS	90°	0.0°	22	2.5	30	500	236	0.094
>400 / NE	180°	0.0°	22	2.7	30	500	140	0.163
>400 / HZ	360°	90.0°	10	2.4	30	500	576	0.141

Table 3-1. Fracture distribution parameters for Forsmark, FFM01/FFM06 domains. The Fisher distribution parameters are pulled from Table 2-14 in [10], and the power-law *k* parameters and P_{32} values are pulled from Table 2 in [2].

This chapter focuses on verifying the implementation of DFNs and conversion of the discrete fracture networks to ECPMs for comparison, specifically in terms of the properties that are expected to mimic those of the Forsmark site. The DFNs are analyzed in terms of the fracture density and fracture distributions for the different fracture sets, as well as in terms of the bulk properties of the equivalent continuous porous media (ECPMs) generated from the DFNs. In Section 3.1, a verification of the DFNs generated by dfnWorks is performed to ensure the expected number of fractures and P_{32} values are obtained using the specification in Table 3-1. In Section 3.1.2, bulk properties of the ECPMs that are expected to coincide between the Forsmark site and the crystalline reference case are compared for consistency. Note that these DFNs are used in the spatial uncertainty loop as part of the uncertainty analysis of the crystalline reference case described in Chapter 4.

3.1 Verifying consistency of DFN statistics with Forsmark

Fracture distribution parameters from Table 2 of [2] were used to define the distribution of radii of circular fractures across the range of fracture zones at Forsmark. Only a subset of the fracture families was used in the generated DFNs for the reference crystalline case. These distribution parameters are presented in Table 3-1. The interdependence of the fracture-size power-law distribution with the fracture density (P_{32}) is described in [4] and is repeated here. As discussed in [5], the probability density function for a fracture of radius r corresponds to a power-law distribution:

$$p(r) = \frac{k r_0^k}{r^{k+1}}, r_0 \leq r < \infty, \tag{3.1}$$

where r_0 is the smallest possible radius. In practice, as is done in [5], this is set to the radius of the smallest observable fracture, which is limited by the radius of the boreholes used for observation at a particular site. In [5], $r_0 = 0.038$ m. Define n_0 as the average number of fractures per unit volume. Then the P_{32} value (area of fractures per volume of rock in units of m^2/m^3) can be approximated as the average number of fractures per unit volume, multiplied by the average area of a fracture, to obtain an area per unit volume:

$$P_{32} = n_0 \int_{r_0}^{r_{upper}} p(r)\pi r^2 dr = n_0 \int_{r_0}^{r_{upper}} \frac{n_0 k r_0^k}{r^{k+1}} \pi r^2 dr = -\frac{\pi n_0 k r_0^k}{k-2} [r^{2-k}]_{r=r_0}^{r=r_{upper}}, \tag{3.2}$$

where r_{upper} is the upper limit of fracture radii included in the empirical determination of the P_{32} value. A value of r_{upper} = 564 m was used to determine the P_{32} values reported in Table 3-1.

If P_{32}, r_0 and r_{upper} are known, one can approximate the average number of fractures per unit volume by dividing the P_{32} by the right-hand side of (3.2) to solve for n_0. In this case, for a power law defined by k = 2.5, r_0 = 0.038 m, r_{upper} = 564 m, and a P_{32} of 0.073 m^2/m^3, we find that

$$n_0 = P_{32}\left[-\frac{\pi k r_0^k}{k-2}\left(r_{upper}^{2-k} - r_0^{2-k}\right)\right]^{-1} = 3.245 \frac{\text{fractures}}{m^3} = 3.245 \times 10^9 \frac{\text{fractures}}{km^3}. \tag{3.3}$$

This operation was used to determine the number of fractures per unit volume for the difference fracture families in Table 3-1; the average number of fractures per unit volume could then be multiplied by the volume of each subdomain to obtain the expected number of fractures in each one.

Given the site extends several kilometers in each direction, it is not feasible to generate a DFN that contains every fracture. The power-law distributions defined as in Table 3-1 are heavily weighted toward small fractures, as shown in Figure 3-1. Note, for $k = 2.5$, 99% of the probability mass of the distribution falls between radii of 0.038 and 0.24 m. The corresponding P_{32} value of 0.073 in the specified subdomain would be equivalent to having the order of 10^9 fractures with radii ranging from 0.038 m to 564 m, most of them lying in the subgrid scale. However, the largest fractures tend to dominate the flow, so it was decided to only store fractures with radii ≥ 30 m. Since the vast majority of fractures are much smaller than this threshold, a far more reasonable number of fractures, having the order of 10^3, are stored.

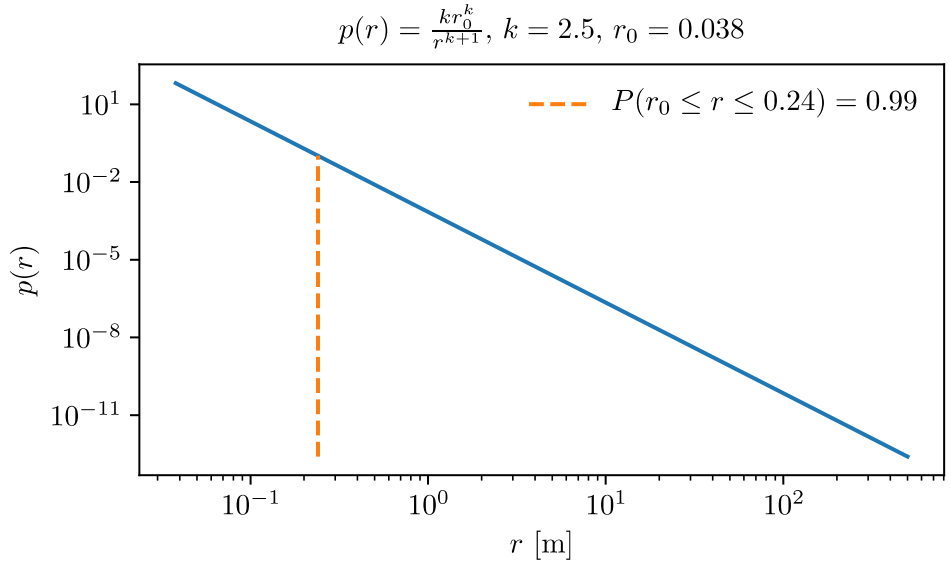


Figure 3-1. The fracture radius distribution. 99% of the probability mass lies in the range of radii between 0.038 and 0.24 m.

The software package dfnWorks has two ways of specifying a DFN. In both cases fracture are generated using a truncated power-law distribution defined over the range of radii that should be stored. The truncated power-law distribution defined over the range $[r_0, r_u]$ is

$$p_{[r_0,r_u]}(r) = \frac{kr_0^k}{r^{k+1}}\left[1 - \left(\frac{r_0}{r_u}\right)^k\right]^{-1}. \tag{3.4}$$

This is the same as (3.1) except with a modified normalization constant to account for the truncated upper bound. For this work the range of radii used was [30, 500] m. To determine when a sufficient number of fractures has been generated, the user can then either specify the number of fractures to be produced and the relative frequency of fractures in each family, or the P_{32} values for each of the

fracture families. The P_{32} values from [2] are for the entire range of measured radii [0.038, 564]. To determine the corresponding P_{32} for a smaller range of radii, denoted $P_{32}[r_{min}, r_{max}]$, one need only compute the integral (3.2) over that range, using the untruncated power law:

$$P_{32}[r_{min}, r_{max}] = \int_{r_{min}}^{r_{max}} n_0 p(r) \pi r^2 dr$$
$$= -\frac{\pi n_0 k r_0^k}{k-2} [r_{min}^{2-k} - r_{max}^{2-k}]. \quad (3.5)$$

In this case the rescaled P_{32} values are listed in Table 3-2.

Fracture elevation [meters above sea level]	Fracture set name	$P_{32}, r \in [30,500]$ $[m^2/m^3]$
> -200	NS	1.98×10^{-3}
	NE	2.58×10^{-3}
	HZ	2.60×10^{-2}
$[-200, -400]$	NS	3.85×10^{-3}
	NE	2.79×10^{-3}
	HZ	1.69×10^{-2}
< -400	NS	2.55×10^{-3}
	NE	1.32×10^{-3}
	HZ	7.53×10^{-3}

Table 3-2. $P_{32}[30, 500]$ computed from P_{32} in Table 3-1 using (3.5).

Alternatively, if one were to specify the number of fractures that dfnWorks should generate to produce these P_{32} values, one would need to specify the total number of fractures to produce, as well as the relative probability of fractures in each family being produced. First, to compute the number of fractures in the range [30, 500] that would need to be generated, one would need to compute

$$\int_{r_{min}}^{r_{max}} n_0 p(r) dr = -n_0 r_0^k [r^{-k}]_{r=r_{min}}^{r=r_{max}} = n_0 r_0^k \left(r_{min}^{-k} - r_{max}^{-k}\right) \quad (3.6)$$

for each of the fracture families. The relative probability of a fracture occurring in each fracture family is then defined as the number of fractures in the family, divided by the total number of fractures across all families. These numbers and probabilities are presented in Table 3-3.

Fracture elevation [meters above sea level]	Fracture set name	Subdomain # fractures, $r \in [30,500]$	Relative probability of fractures
> -200	NS NE HZ	227 335 2767	0.020 0.030 0.245
$[-200, -400]$	NS NE HZ	440 363 1771	0.039 0.032 0.157
< -400	NS NE HZ	1252 737 3398	0.111 0.065 0.301

Table 3-3. The number of fractures in each fracture family.

3.1.1 Consistency of DFN specifications

Initially, the properties of the DFNs generated by specifying target P_{32} values to dfnWorks were inconsistent with the properties of the DFNs generated by specifying the target number of fractures. It was determined that dfnWorks computes P_{32} using both faces of the fracture. Compared to the definition as specified in [4] and repeated in (3.2), this means the P_{32} that dfnWorks computes will be approximately twice the value computed using (3.2) and as reported in [2]. To account for this, care must be taken when specifying target P_{32} values to dfnWorks to instead pass twice the target value. Additionally, it should be noted that the values dfnWorks reports for P_{32} need to be halved to compare to P_{32} values computed as in [2][4]. Taking this change into account, the number of fractures in each family and the P_{32} values for the DFNs generated using either method are reported in Table 3-4.

The agreement is not perfect between the two methods, but the results are in relatively good agreement with each other. It should be noted that the families with the highest P_{32} values exhibited the greatest discrepancy between the resulting number of fractures and P_{32} values. This discrepancy can be explained by the differences in how the algorithms terminate. In the case of specifying the number of fractures, fractures are added according to the underlying of distribution of fractures by family and by radius until the total number of fractures is reached, at which point it terminates. In the case of specifying P_{32}, on the other hand, the algorithm will keep adding fractures until the P_{32} is within a small tolerance of the target. In general, fractures of smaller radius will allow for more refined values of P_{32}, so this method results in more fractures with smaller radius.

A comparison of the distributions of fractures over all families between the two methods of specification indicate the P_{32} specification has almost 2000 more fractures with radius near the lower bound of 30 m, compared to the number of fractures specification. For the DFNs used in the crystalline reference case elsewhere in this work, the method of specifying the number of fractures was used to generate the DFNs. In future work, specifying a target P_{32} value may be favorable because it enables as close a replication to the parametrization in [2] as possible. However, as will be discussed in Section 3.1.2, the difference between the P_{32} values from [2] and those of the

DFNs generated by specifying the number of fractures per family did not significantly impact the bulk properties of the resulting ECPMs.

Fracture elevation [m.a.s.l.]	Fracture set name	Converge on P_{32}			Converge on n fractures		
		Target P_{32}	Final P_{32}	n fractures	Final P_{32}	Target n fractures	Final n fractures
> -200	NS	1.98×10^{-3}	1.98×10^{-3}	271	1.47×10^{-3}	227	209
	NE	2.57×10^{-3}	2.58×10^{-3}	326	2.30×10^{-3}	335	329
	HZ	2.60×10^{-2}	2.60×10^{-2}	3298	2.24×10^{-2}	2767	2697
$[-200, -400]$	NS	3.85×10^{-3}	3.85×10^{-3}	414	3.22×10^{-3}	440	440
	NE	2.79×10^{-3}	2.81×10^{-3}	386	2.00×10^{-3}	363	342
	HZ	1.69×10^{-2}	1.69×10^{-2}	1835	1.41×10^{-2}	1771	1559
< -400	NS	2.55×10^{-3}	2.55×10^{-3}	1475	2.27×10^{-3}	1252	1211
	NE	1.32×10^{-3}	1.32×10^{-3}	822	1.21×10^{-3}	737	745
	HZ	7.53×10^{-3}	7.53×10^{-3}	3663	6.43×10^{-3}	3398	3100

Table 3-4. P_{32} values and number of fractures per family for DFNs generated by either specifying P_{32} or number of fractures per family. The number of fractures reported are in range $r \in [30, 500]$ m.

3.1.2 Comparing crystalline reference case bulk ECPM properties to Forsmark

Having verified the DFN statistics are consistent with the desired values in Table 3-1, attention turns to whether the bulk properties of the derived ECPMs for the current GDSA crystalline reference case are consistent with similarly-derived ECPMs for the Forsmark site in Sweden. Because only a subset of the fracture domains and sets were used to define the crystalline case, only a subset of properties will be expected to agree within a reasonable degree. The comparison will be made in terms of the bulk hydraulic conductivity computed using mapDFN.py [11]. The process by which this is derived from the DFNs is described in [6].

The standard assumption for the relationship between permeability and hydraulic conductivity is

$$K = \frac{k\rho g}{\mu}, \tag{3.8}$$

where K is hydraulic conductivity, k is permeability, ρ is the density of the transported fluid, g is the gravitational constant, and μ is the fluid's viscosity. For the GDSA work the fluid is water.

When dfnWorks computes permeability from fracture transmissibility, it uses $\rho = 997.7\ kg\ m^{-3}$, $\mu = 8.9 \times 10^{-4}\ kg\ m^{-1}s^{-1}$ (from dfnWorks insertShape.cpp source code). These values will be used to convert k to K. Let $\gamma = \frac{\rho g}{\mu} = 1.1 \times 10^{7} m^{-1}s^{-1}$ so that

$$K = \gamma k. \tag{3.9}$$

The computational domain for the GDSA crystalline reference case is one continuous block; it is not subdivided by the depth zones. The ranges of the computational domain for dfnWorks and PFLOTRAN are summarized in Table 3-5 so that the appropriate subdomains can be mapped between dfnWorks, PFLOTRAN, and the Forsmark information. Note that the 15 m grid cells do not evenly subdivide [0,1260], so the bounds for each subdomain were taken to fall on the interior of each. The derived permeability tensor is only nonzero in the axial directions, so the corresponding hydraulic conductivity is also only nonzero in the axial directions.

Fracture elevation [m.a.s.l.]	dfnWorks subdomains	PFLOTRAN subdomains	15 m grid subdomains
> -200	$[430, 630]$	$[1060,1260]$	$[71,83]$
$[-200, -400]$	$[230,430]$	$[860,1060]$	$[58,70]$
< -400	$[-630,230]$	$[0,860]$	$[0,57]$

Table 3-5. The corresponding subdomains for dfnWorks and PFLOTRAN.

Statistics for the hydraulic conductivity for three depth zones were provided by the SKB team through private correspondence [7] and refer to runs described in [3]. These properties were reported in terms of the axial directions of the hydraulic conductivity tensor, K_{xx}, K_{yy}, and K_{zz}, as well the geometric mean of the three directions: $K_{eff} = (K_{xx}K_{yy}K_{zz})^{1/3}$. The goal here is to compute the same statistics for the crystalline reference case and check for agreement between them. As described in [3], when converting to an ECPM from a DFN, SKB set the hydraulic conductivity of any cell without a fully-connected network of fractures to 10^{-11} m/s. The appropriate minimum was derived by calculating the minimum values seen when the DFN was truncated only at very small fractures relative to cell size, so that they were free of effects from truncating the radii used in computing the ECPM. A minimum hydraulic conductivity of 10^{-11} m/s is equivalent to a minimum permeability of 10^{-18} m^2. The minimum permeability for the ECPMs derived from dfnWorks for the GDSA crystalline reference case is 10^{-20} m^2. To enable a comparison with the SKB statistics, the background hydraulic conductivity was replaced with 10^{-11} to compute the means reported here.

The SKB team reported the log base 10 of the arithmetic mean of the hydraulic conductivity denoted $\log_{10}(\overline{K_{xx}})$, as well as the log base 10 of the geometric mean, denoted $\overline{\log_{10}(K_{xx})}$. Any cells containing a deformation zone were not included in the computation of the mean. These values are reported in Table 3-6.

Depth [m]	$\log_{10}(\overline{K_{xx}})$	$\overline{\log_{10}(K_{xx})}$	$\log_{10}(\overline{K_{yy}})$	$\overline{\log_{10}(K_{yy})}$	$\log_{10}(\overline{K_{zz}})$	$\overline{\log_{10}(K_{zz})}$	$\overline{\log_{10}(K_{eff})}$
> -200	-7.1	-8.0	-6.3	-8.1	-6.3	-9.4	-8.5
$[-400, -200]$	-8.1	-9.7	-8.1	-9.6	-9.4	-10.5	-9.9
< -400	-9.8	-10.8	-9.7	-10.8	-9.7	-10.8	-10.8

Table 3-6. The statistics for the hydraulic conductivity of the ECPM representing FFM01/06, provided by SKB in [6].

The SKB ECPMs were computed using a finite element discretization with 20 m grid size, compared to the 15 m grid size used for the GDSA crystalline reference case. Smaller grid sizes tend to result in lower mean hydraulic conductivity and higher variance, so this will be considered when comparing results. The range of fracture radii used in the upscaling was from 5.6 m to 564 m, compared to the range of 30 m to 500 m used in in the reference case.

As described in [2], equation 2, SKB assumed the following relationship between fracture size and hydraulic aperture/transmissivity for their hydrogeological base case in [3]:

$$\log_{10} T = \log_{10}(ar^b) + \sigma\mathcal{N}(0,1) \ , \tag{3.10}$$

where T [m^2/s] is transmissivity, r [m] is fracture radius, a and b are coefficients in a power-law relationship between r and $\log_{10} T$, $\sigma$is the standard deviation of $\log_{10} T$, and $\mathcal{N}(0,1)$ denotes the normal distribution of a random deviate of zero mean and unit variance. This relation was specified for each depth zone individually and affects the hydraulic conductivity of the ECPM. As of dfnWorks version 2.1, it is only possible to specify a single relationship between transmissivity and fracture size that will apply to the entire domain, and the relationship is of the form

$$T = ar^b. \tag{3.11}$$

Because only one relationship could be defined, the parameters for the fully-correlated relationship of FFM01/06 and depth zone 200-400 meters below sea level (mbsl) from Table 2 in [2] were used in the DFN specification for this work. This means that only the statistics for depth zone 200-400 mbsl can be expected to agree to some degree to those reported by SKB.

SKB did not include the properties of any element cut by a deformation zone in the computation of their statistics. Correspondingly, the deformation zones, specified as user-defined faults to dfnWorks, were filtered out for computing the statistics reported here. To reiterate, the background hydraulic conductivity for the ECPMs was increased to 10^{-11}, compared to the 1.1×10^{-13}, to make it consistent with the background value for SKB. SKB only computed the statistics for the single reference hydrogeological base case, while 30 identically distributed ECPMs were computed for this work. To compare to the SKB statistics, the log-base-10 algebraic and geometric mean for each hydraulic conductivity field was computed, then a sample mean was computed over the 30 algebraic/geometric means from each ECPM. To compute the means over the three depth subdomains >-200 m, [-400 m, -200 m], and <-400 m, the computational domain was partitioned according to the z indices specified in Table 3-5. The arithmetic and geometric means were computed on each of these subdomains. The resulting (overall) means are reported in Table 3-7.

Depth [m]	$\log_{10}(\overline{K_{xx}})$	$\overline{\log_{10}(K_{xx})}$	$\log_{10}(\overline{K_{yy}})$	$\overline{\log_{10}(K_{yy})}$	$\log_{10}(\overline{K_{zz}})$	$\overline{\log_{10}(K_{zz})}$	$\overline{\log_{10}(K_{eff})}$
> -200	-8.7	-10.0	-8.7	-10.0	-9.2	-10.3	-10.1
$[-400, -200]$	-8.8	-10.2	-8.8	-10.2	-9.2	-10.4	-10.3
< -400	-9.2	-10.7	-9.2	-10.7	-9.5	-10.7	-10.7

Table 3-7. The log base 10 of the mean values for the ECPMs computed for the GDSA crystalline reference case with the background hydraulic conductivity set to 10^{-11} m/s for consistency with SKB.

As expected, because of the constant transmissivity-radius relationship as a function of depth, the statistics in Table 3-7 do not vary as significantly as those for SKB. This makes sense, since the only change as a function of depth in Table 3-7 is induced by the decrease in fracture frequency and radius as a function of depth. Given the difference in correlation model, minimum fracture radius, and cell size, the values for the depth zone 200-400 mbsl agree reasonably well. The SKB and GDSA values for this depth zone are compared in Figure 3-2. As mentioned previously, because the SKB team used a larger grid size and a larger ranges of fracture sizes, it is expected that their hydraulic conductivities would be slightly higher than for the crystalline reference case. This holds for all directions except the z (depthwise) direction, and these values still agree well with those of SKB. These results verify the specification and implementation of the DFN and conversion to ECPM in this work, insofar as the ECPM is consistent with that of SKB to the extent it is expected to be.

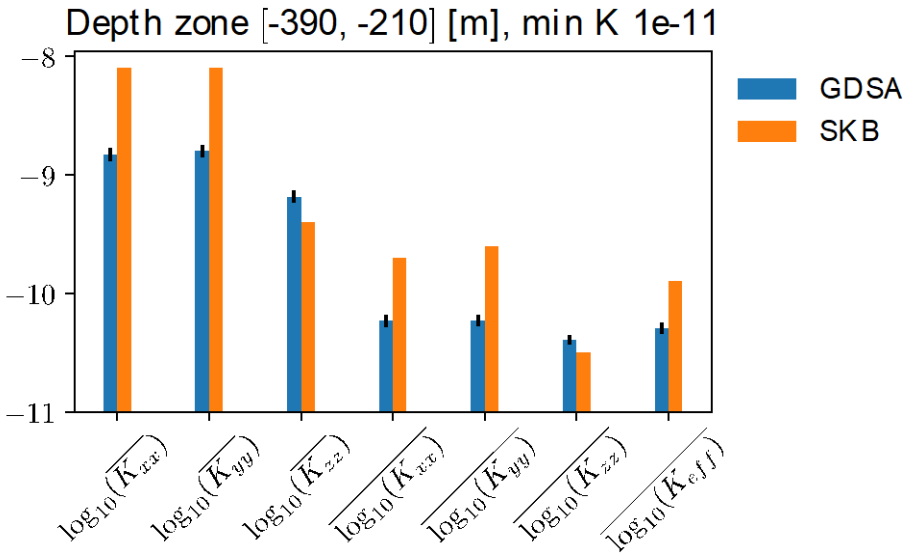


Figure 3-2. A bar chart comparing the mean hydraulic conductivities for the second depth zone between [-400, -200] m. The vertical black lines are one standard deviation on the averages, taken over the 30 DFNs.

References: Chapter 3

[1] Hyman, J. D., S. Karra, N. Makedonska, C. W. Gable, S. L. Painter and H. S. Viswanathan (2015). "dfnWorks: A discrete fracture network framework for modeling subsurface flow and transport," *Computers & Geoscience*, **84**:10-19.

[2] Steven Joyce, Lee Hartley, David Applegate, Jaap Hoek, and Peter Jackson. Multi-scale groundwater flow modeling during temperate climate conditions for the safety assessment of the proposed high-level nuclear waste repository site at Forsmark, Sweden. *Hydrogeology Journal*, 22(6):1233–1249, 2014.

[3] Steven Joyce, Trevor Simpson, Lee Hartley, David Applegate, Jaap Hoek, Peter Jackson, David Swan, Niko Marsic, and Sven Follin. Groundwater flow modelling of periods with temperate climate conditions–Forsmark. Technical report, Swedish Nuclear Fuel and Waste Management Co., 2010.

[4] Allan Hedin. Semi-analytic stereological analysis of waste package/fracture intersections in a granitic rock nuclear waste repository. *Mathematical geosciences*, 40(6):619–637, 2008.

[5] Sven Follin, Lee Hartley, Ingvar Rh´en, Peter Jackson, Steven Joyce, David Roberts, and Ben Swift. A methodology to constrain the parameters of a hydrogeological discrete fracture network model for sparsely fractured crystalline rock, exemplified by data from the proposed high-level nuclear waste repository site at Forsmark, Sweden. *Hydrogeology Journal*, 22(2):313–331, 2014.

[6] Paul E Mariner, Emily R Stein, Jennifer M Frederick, S David Sevougian, Glenn Edward Hammond, and DG Fascitelli. Advances in geologic disposal system modeling and application to crystalline rock. Technical report, Sandia National Laboratories.(SNL-NM), 2016.

[7] Lee Hartley and Steven Joyce. Responses to SSM on Hydrogeology, SKBdoc 1396325 ver 1.0, Svensk Kärnbränslehantering AB., 2013. SKBdoc documents will be submitted upon request to document@skb.se.

[8] J. D. Hyman, C. W. Gable, S. L. Painter, and N. Makedonska. Conforming Delaunay triangulation of stochastically generated three dimensional discrete fracture networks: A feature rejection algorithm for meshing strategy. *SIAM J. Sci. Comput.*, 36(4):A1871–A1894, 2014.

[9] SKB, Long-Term Safety for the Final Repository for Spent Nuclear Fuel at Forsmark. Main Report of the SR-Site Project. Technical Report SKB TR-11-01, Swedish Nuclear Fuel and Waste Management Co., Stockholm, Sweden, 2011

[10] Wang, Y., E. Matteo, J. Rutqvist, J. Davis, L. Zheng, J. Houseworth, J. Birkholzer, T. Dittrich, C. W. Gable, S. Karra, N. Makedonska, S. Chu, D. Harp, S. L. Painter, P. Reimus, F. V. Perry, P. Zhao, J. B. Begg, M. Zavarin, S. J. Tumey, Z. R. Dai, A. B. Kersting, J. Jerden, K. E. Frey, J. M. Copple and W. Ebert (2014). Used Fuel Disposal in Crystalline Rocks: Status

and FY14 Progress. FCRD-UFD-2014-000060, SAND2014-17992 R. Sandia National Laboratories, Albuquerque, New Mexico.

[11] Stein, E. R., J. M. Frederick, G. E. Hammond, K. L. Kuhlman, P. E. Mariner and S. D. Sevougian (2017). "Modeling Coupled Reactive Flow Processes in Fractured Crystalline Rock," *Proceedings of the 16th International High-Level Radioactive Waste Management (IHLRWM 2017) Conference*, Charlotte, North Carolina, April 9-13, American Nuclear Society.

4. CRYSTALLINE REFERENCE CASE: UPDATES

For a nuclear waste repository located in crystalline rock, a major source of uncertainty in performance assessment is the spatial heterogeneity of potential fracture flow paths through the host rock. Conceptually, a long-lived radionuclide released from a waste package will initially migrate through the buffer material and into the surrounding damaged rock zone (DRZ). From there it will migrate along the DRZ until it enters a fracture that takes it farther into the host rock, where connected fractures can provide a path to a nearby fracture zone. It might then migrate along this fracture zone and through connected fracture zones to the biosphere. Along the flow path, the radionuclide will undergo radioactive decay and ingrowth and diffuse into and out of dead-end pores and fractures. Additionally, depending on its properties and the environmental conditions along the flow path, it will adsorb and desorb from colloids and immobile mineral surfaces, chemically react with aqueous species, possibly change oxidation state, and, if solubility-limited, precipitate and dissolve.

This chapter presents a new uncertainty analysis of a reference case repository for commercial spent nuclear fuel in fractured crystalline rock. The reference case is identical to that in Stein et al. [1], Mariner et al. [4], and Sevougian et al. [9] except for improvements to the fracture network implementation (discussed in Chapter 3) and the design of the uncertainty analyses. Note that this analysis builds extensively on the crystalline reference case uncertainty analysis performed in 2019 and documented in Chapter 8 of [14].

4.1 Uncertainty Analysis (UA)

The uncertainty analysis includes multiple types of uncertainty: spatial uncertainty, parameter epistemic uncertainty, and aleatory uncertainty. Uncertainties described as spatial uncertainty relate to the unknown structure of the repository and surrounding rock. The spatial uncertainty is not strictly categorized as epistemic or aleatory, in part because it incorporates some of both types of uncertainties. Uncertainties described as parameter epistemic uncertainty relate to inputs with fixed but unknown scalar values.

Section 4.1.1 discusses different perspectives on the categorization of spatial uncertainty and Section 4.1.2 summarizes the structure and implementation of the UA.

4.1.1 Uncertainty Categorization

The categorization of spatial uncertainty in this analysis was an issue of debate. Last year, the discrete fracture networks were treated as aleatory [14], and that study motivated further separation of the spatial uncertainty from epistemic parameter uncertainty for the current study. In this report, spatial heterogeneity in the discrete fracture networks was separated from most of the parameter uncertainty because one of the goals of the analysis was to understand the relative contributions of the spatial heterogeneity and parameter uncertainties to total uncertainty. Though this separation was illustrative, there was not agreement on whether it represents a classical separation between aleatory and epistemic uncertainty.

The categorization of the spatial uncertainty depends on the definitions of aleatory and epistemic uncertainty, the intent of the crystalline reference case, the difference between theoretically and practically reducible uncertainty, and historical usage. With respect to

definitions, epistemic uncertainty is lack-of-knowledge uncertainty, also called reducible uncertainty, and aleatory uncertainty is stochastic or irreducible uncertainty. Aleatory uncertainty can also be described with respect to time; it is uncertainty about future occurrences, not the present state [20]. However, outside of PA, this connection to time is not always included in the definition [21].

One perspective on spatial uncertainty is to treat it as an epistemic uncertainty. This assumes that the underlying fracture network at the crystalline reference case site is fixed, but we do not know what it is. Our knowledge of the fracture network is imperfect: we cannot perfectly characterize the fractures due to limitations of geophysical imaging capabilities. As discussed in Chapter 3, we have used the dfnWorks software [3] to generate twenty realizations that are consistent with field data and with what we understand about crystalline formations and their properties. We cannot say that any one of the twenty DFNs we have generated is the exact real fracture network, but they are possible realizations each intended to represent a realistic possibility. This follows from earlier work at WIPP and Yucca Mountain where geologic field data was treated as epistemic. For example, the WIPP analysis involved the generation of realizations of the transmissivity field in the Culebra dolomite formation (LaVenue and RamaRao, 1992) and the Yucca Mountain analysis involved realizations of flow fields for both the unsaturated and saturated zones for flow and transport modeling (Rechard et al. 2014).

An alternative perspective on spatial uncertainty is to treat it as an aleatory uncertainty. This argument is based on two concepts: 1) predicting the distribution of random fractures through space is analogous to predicting the occurrence of random events through time, and 2) uncertainties that may be reducible philosophically are not necessarily practically reducible. The dfnWorks software produces random fracture networks that are consistent with a set of fixed characteristics specified for the site (this is done by generating random fractures from user specified distributions on fracture size, orientation, etc. and placing them in the fracture network, see Chapter 3 for details). Analogous to future random events, there are infinitely many fracture networks that can occur in space which are consistent with these fixed characteristics. This can be perceived either as inherent randomness or model uncertainty: the characteristics are insufficient to define a fixed fracture network.

The perspective that spatial uncertainty should be treated as aleatory also relates to the practical motivation for separating uncertainty. Epistemic uncertainty can be reduced, so separating epistemic uncertainty from aleatory uncertainty identifies where uncertainty could be reduced, as well as how much uncertainty cannot be eliminated. Even with the most advanced measurement technologies, it is not feasible to measure and model the true fracture network at a potential repository site. Though the fracture network is fixed, and so the uncertainty is hypothetically reducible, it cannot be reduced practically. If the goal behind separating out epistemic uncertainty is to identify uncertainty that can actually be reduced, it makes more sense to include uncertainty in the fracture network as an aleatory uncertainty.

Finally, the categorization of spatial uncertainty is further complicated by the introduction of another type spatial heterogeneity that clearly mixes epistemic and aleatory components: uncertainty in the order in which waste packages breach in the repository (e.g., this is a spatial variability in the sense that waste package 37 may breach first, followed by waste package 483, etc. These waste packages are at different locations in the repository). Uncertainty in waste

package corrosion rate is introduced by sampling from a truncated log normal distribution for the base corrosion rate in a temperature-dependent rate expression for general corrosion [4]; this log normal distribution for the (spatially distributed) waste package corrosion rate is sampled within PFLOTRAN [6, 7] and is considered an aleatory uncertainty (random, stochastic, or irreducible uncertainty). Both the mean (*meanWPrate*) and standard deviation (*stdWPrate*) of this log normal distribution, however, are treated as epistemically uncertain inputs.

This analysis also contains epistemic uncertainty in parameter values, which were categorized without debate. Epistemic uncertainties include porosity and permeability of the bentonite buffer (*pBuffer* and *pBuffer*, respectively), permeability of the DRZ (*kDRZ*) and of the overlying sedimentary unit (*kGlacial*), and rate of spent (used) nuclear fuel dissolution (*rateUNF*). Note that the epistemic parameter uncertainty is simpler to treat than the fracture network uncertainty. For epistemic parameter uncertainty, we assume that the parameters have a fixed but unknown value for this particular case study. That is, there is one value for porosity of the bentonite buffer but we do not know what it is, thus we generate possible realizations of a number for this porosity.

Because the debate on how to categorize the spatial uncertainty is ongoing, we refrain from making an explicit statement about the categorization in this report. The implementation of spatial heterogeneity includes epistemic and aleatory uncertainties, and there are philosophical arguments for either categorization. Regardless of how it is characterized, the spatial uncertainty was separated from the parameter uncertainty in a two-loop structure for this UA. To avoid implying a categorization, the sampling loop with spatial heterogeneity is referred to simply as the spatial loop.

4.1.2 Uncertainty Implementation

The UA comprises a spatial loop of sample size 20 fracture network realizations, and a parameter loop of sample size 40 for a total of 800 simulations. Note that both of these loops involve epistemic uncertainties, but the spatial uncertainty loop involves the spatial heterogeneity exhibited by the 3-D variability in the fracture networks and the epistemic parameter loop involves the parametric uncertainty associated with the epistemic parameters. Because the aleatory uncertainty is also spatial, it is included in the spatial loop. For each DFN realization in the spatial loop, a different random seed is used when sampling the waste package corrosion rate distribution, so that the order of waste package breach associated with each DFN is different. The sampling loops are shown in Figure 4-1.

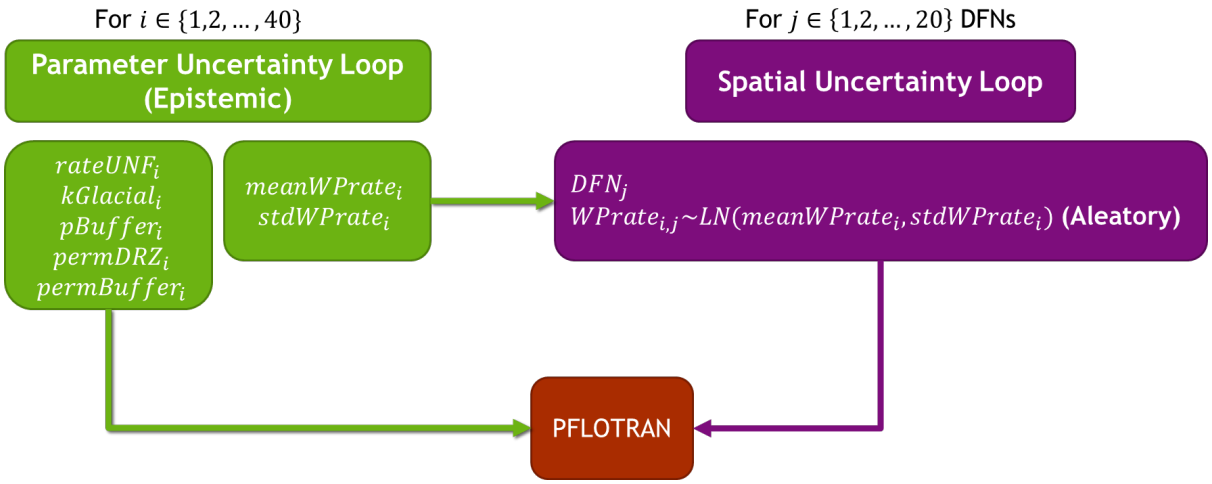


Figure 4-1. Sampling Loops used in the Uncertainty Analysis of the Crystalline Reference Case presented in this Chapter.

Latin hypercube sampling of epistemically uncertain parameters is performed using Dakota [8]. The uncertainty analysis includes *pBuffer, kBuffer, kDRZ, kGlacial*, and *rateUNF* as epistemically uncertain inputs, and both *meanWPrate* and *stdWPrate* are also included as epistemically uncertain inputs in the parameter loop. The uncertain parameters are listed in Table 4-1. The outputs include maximum ^{129}I concentration over time and location in the aquifer as well as quantities such as mean travel time, mean residence time, and flow rates as defined in Table 4-2. These new Quantities of Interest (QoIs) that have been implemented and are the subject of this study (see Table 4-2 and the discussion in Section 4.2.7 below).

When structuring the UA, the same set of epistemic samples (e.g., same 40 sample vectors for the epistemic variables) were run for each of the 20 spatial realizations. Repeating the epistemic samples allows explicit separation of effects from the epistemic uncertainty and effects from the spatial uncertainty (the DFN and the spatial distribution of base normalized general corrosion rate) but comes at the cost of covering less of the epistemic sample space. The epistemically uncertain parameters and their distributions are listed in Table 4-1.

Table 4-1. Uncertainty distributions propagated in crystalline reference case UA.

Input	Description	Range	Units	Distribution	Sampling Loop
DFN	Particular realization of a discrete fracture network	1-20			Spatial loop
Aleatory_1	Aleatory sampling to generate a fixed ordering of waste package degradation for each DFN	0-1		Uniform	Spatial loop
rateUNF	Fractional dissolution rate of spent (used) nuclear fuel	10^{-8} – 10^{-6}	yr^{-1}	log uniform	Parameter loop
kGlacial	Glacial till permeability	10^{-15} – 10^{-13}	m^2	log uniform	Parameter loop
pBuffer	Buffer porosity	0.3 – 0.5	-	uniform	Parameter loop
permDRZ	DRZ permeability	10^{-19} – 10^{-16}	m^2	log uniform	Parameter loop
permBuffer	Buffer permeability	10^{-20} – 10^{-17}	m^2	log uniform	Parameter loop
meanWPrate	Mean of the truncated log normal distribution on base normalized general corrosion rate (*R*)	-5.5 – (-4.5)	$log(yr^{-1})$	Uniform	Parameter loop
stdWPrate	Standard deviation of the truncated log normal distribution	0.15 – 0.4	$log(yr^{-1})$	Uniform	Parameter loop

4.2 Model Set-up

4.2.1 Model Domain

The model domain [1] is 3015-m in length, 2025-m in width, and 1260-m in height, partially depicted in Figure 4-2. Overlying the host rock is a 15-m thick overburden of glacial sediments (not shown). The repository is located at a depth of 585 m. Forty-two disposal drifts contain 80 12-PWR waste packages each (3360 12-PWR waste packages in total). Drifts are backfilled with bentonite buffer and are surrounded by a 1.67-m thick DRZ. Within the repository, grid cells are as small as 1.67-m on a side; elsewhere grid cells are 15-m on a side. The model domain contains 4,848,260 cells; of these, approximately 2.5 million are the smaller cells in and around the repository that allow representation of individual waste packages with surrounding buffer materials. Additional information on the grid and dimensions may be found [1] available for download at https://pa.sandia.gov.

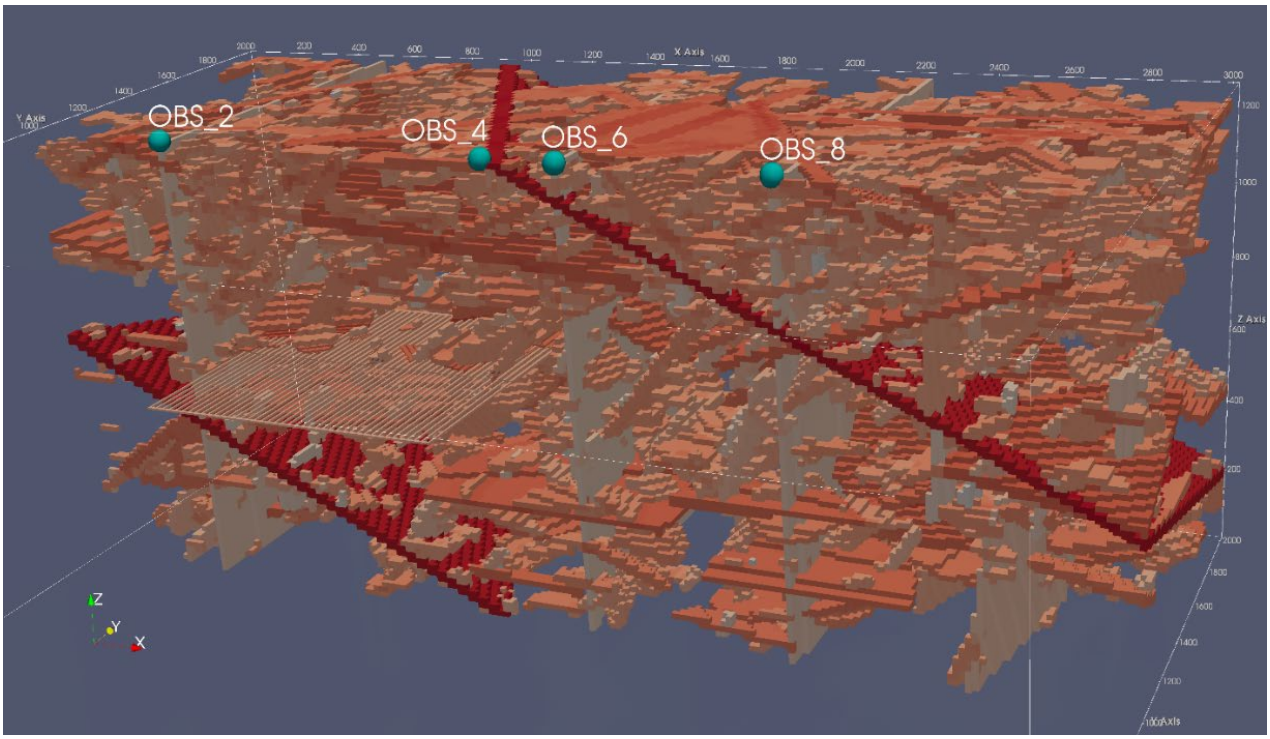


Figure 4-2. Cut-away of DFN 1 realization mapped to porous medium grid, showing the full repository and the far half of the model domain.

4.2.2 Discrete Fracture Networks

As described by Mariner et al. [4], the representation of fractured crystalline rock in the GDSA reference case is based on the well-characterized, sparsely fractured metagranite at Forsmark, Sweden [2]. At Forsmark, volumes of relatively undeformed rock (termed a "fracture domain") are bounded by large-scale mappable features of concentrated brittle and/or ductile deformation (termed "deformation zones"). The fracture domains are sparsely fractured, and the fractures within each domain can be described in terms of a number of "fracture sets", which are distinguished from each other on the basis of fracture orientation. Six fracture domains are defined at Forsmark, each containing five fracture sets. As appropriate, three depth zones are defined (<200 m below sea level (mbsl), 200-400 mbsl, and >400 mbsl) to account for the decrease in fracture density and fracture transmissivity with depth. The parameters describing the DFNs for this case study are described in Chapter 3 of this report.

The crystalline host-rock reference case analyzed here [4], based on the Forsmark data set, is modeled using some fixed features and an uncertain fracture network. The case contains 5 fixed fracture zones and three depth intervals. Twenty random networks were generated with dfnWorks, one for each realization of the spatial uncertainty loop. The deterministic, user-defined fracture zones represent large, mappable features, such as faults, and are common to all realizations. There are three subvertical fracture zones (Figure 4-1, in gray) and two fracture zones with a dip of approximately 30 degrees (in red). The discrete fracture networks, two-dimensional planes distributed in the three-dimensional model domain, are generated using dfnWorks [3], and mapped to the equivalent continuous porous medium domain using mapDFN.py, a code that approximates

hydraulic fracture properties by calculating and assigning permeability and porosity on a cell-by-cell basis [1].

The fracture set properties and deterministic fracture zones employed in this study provided sufficient fracture connectedness such that each DFN realization resulted in direct fracture pathways from the repository to the top boundary of the fractured crystalline host-rock. The existence of connected fracture pathways was determined by dfnWorks.

4.2.3 Waste Package Corrosion Model

The waste package corrosion model implemented in PFLOTRAN (Mariner et al. [4], Section 3.2.1) calculates normalized thickness of the waste package wall at each time step as a function of a base waste package corrosion rate, a canister material constant, and temperature. Waste package breach occurs when the normalized thickness reaches zero. The normalized thickness is initialized to 1, and is reduced at each time step as a function of the effective waste package corrosion rate R_{eff} defined by

$$R_{eff} = R \cdot e^{C\left(\frac{1}{333.15} - \frac{1}{T(t,\vec{x})}\right)} \tag{4-1}$$

where R is the base corrosion rate at 60ºC in units of yr^{-1}, T is the local temperature (in Kelvin), and C is the canister material constant. This equation assumes that reaction rates are a function of temperature as described by the Arrhenius equation. It is also assumed that waste package corrosion occurs via general corrosion. R represents the normalized general corrosion rate at 60ºC in units of 1/T (time) (i.e., units of L/T normalized by the thickness (in units of length L) of the waste package wall). In PFLOTRAN, R is in units of yr^{-1} and is entered as a lognormal distribution across waste packages.

PFLOTRAN assigns a base normalized general corrosion rate (R) for each waste package by sampling on a truncated log normal distribution whose mean, standard deviation, and upper truncation limit may be treated as epistemic uncertain inputs sampled by Dakota [8]. In UA the mean of the distribution is sampled over the range $10^{-5.5}$ to $10^{-4.5}$ (in units of yr^{-1}). The standard deviation is sampled over the range $10^{0.15}$ to $10^{0.4}$ (also in units of yr^{-1}). The uncertainty analysis uses a constant upper truncation limit of $10^{-3.5}$ yr^{-1}. A mean of 10^{-5} for R results in a mean waste package breach time of 100,000 years if waste packages are held at a constant temperature of 60 °C.

Sampling on both the mean and standard deviation implements uncertainty in the location and spread of the normalized general corrosion rate distribution. This high level of uncertainty is appropriate because the state of knowledge is low and, since the distribution is sampled for each waste package, it makes sense to have a different corrosion rate distribution for each realization. Unlike the other uncertainty distributions included in the UA, the normalized general corrosion rate distribution is sampled many times within each realization. First, the mean and standard deviation are sampled once per realization to establish the corrosion rate distribution for that realization. Then this distribution is sampled independently for each waste package within that realization. Sampling of the distribution parameters implements the epistemic uncertainty in the corrosion rate and sampling each waste package independently from that distribution implements spatial aleatory uncertainty.

4.2.4 Initial and Boundary Conditions

Initial conditions specified are pressure and temperature. Nominal nonzero radionuclide concentrations are also specified as initial conditions but this is for numerical necessity, not to represent reality. Initial pressures and temperatures throughout the model domain are calculated by applying a liquid flux of 0 m/s and an energy flux of 60 mW/m^2 to the base of the domain and holding temperature (10°C) and pressure (approximately atmospheric) constant at the top of the domain and allowing the simulation to run to 10^6 years. Pressure at the top of the domain decreases from west (left) to east (right) with a head gradient of –0.0013 (m/m). This technique results in initial conditions that represent a geothermal temperature gradient and hydrostatic pressure gradient in the vertical direction, and a horizontal pressure gradient that drives flow from west to east.

The initial concentration of ^{129}I in all cells is 10^{-22} mol/L. A non-zero value is necessary, because PFLOTRAN transport equations are formulated in terms of the log of concentration. A concentration of 10^{-22} mol/L is approximately 60 atoms of ^{129}I per liter of water.

At all six faces of the model domain, pressures and temperatures are held constant at initial values. Concentration of ^{129}I is held at the initial concentration at inflow boundaries. At outflow boundaries, the concentration gradient is set to zero.

4.2.5 Observation Points

Previous analysis of the crystalline reference case analyzed uncertainty in ^{129}I concentrations at specific observation points [14]. Rather than focusing on each spatial location individually as in that analysis, this UA uses the maximum ^{129}I concentration calculated over all observation points. Because the peak ^{129}I concentration varies spatially over time, the maximum over all of the tracked observation points is a better measure of overall repository performance; this measure will indicate poor performance if the ^{129}I concentration at any observation point exceeds requirements.

4.2.6 Timestep Size

During previous analysis of the crystalline reference case [14], oscillations in ^{129}I concentrations which could impact estimates of ^{129}I breakthrough times were found at observation points. An analysis was conducted to identify the cause of the oscillations and determine how the effect could be minimized in future studies. In [14], we document a study showing the tradeoff between maximum timestep size and computational time: Oscillations in the ^{129}I concentrations become less frequent with smaller timesteps but this results in an increased computational time. From the study, we determined that a maximum timestep of 5,000 years was acceptable for our quantities of interest.

For the PFLOTRAN runs presented in this report, a maximum timestep size of 5,000 years was also used. We examined the peak ^{129}I concentration over all time as well as the maximum ^{129}I as a function of time. These quantities did not exhibit significant discontinuities or oscillations, therefore we continued to use the 5,000 year maximum timestep.

4.2.7 New Quantities of Interest (QoIs)

The new QoIs implemented in the UA are listed in Table 4-2. They include a new performance metric capability and several new DFN-related bulk characteristic calculations potentially relevant to performance.

The new performance metric capability allows for monitoring the entire aquifer at each time step for ^{129}I concentration, recording the highest ^{129}I concentration in the aquifer at that time, and recording its location. The peak ^{129}I concentration in the aquifer in the simulation can then be determined, which is important because it is the ultimate measure of repository performance in this current model.

The DFN-related bulk characteristics that are potentially relevant to performance include the mean travel time (MTT) of a conservative tracer from the repository to the aquifer, the mean residence time (MRT) of an initial conservative tracer within the repository, net mass flux (NMF) of a conservative tracer out of the repository, and rock boundary water mass flow rates.

A more detailed description of these new QoIs may be found in Mariner et al. (2020) [15]. The mean residence time of a conservative tracer initially present in the repository (MRT) at each time step in the simulation is determined by tracking the total mass of a tracer remaining in the repository over time. Because the initial tracer slowly flushes from the repository and never completely flushes from the repository, the MRT has an associated time series of residence times. The MRT is given by Equations 4.1 and 4.2 below; these are equations 29-30 in [15] and discussed in more detail in that report. The cumulative MRT of an initial tracer of a model region is calculated by tracking the total mass of the tracer in the region over the time interval (τ) using the following equation

$$\mathrm{MRT}(\tau) = \int_{t=0}^{\tau} F(t)\mathrm{d}t \tag{4.1}$$

where (i) $F(t)$ is the fraction of initial pulse remaining, calculated from

$$F(t) = \frac{N(t)-N_b}{N_o-N_b}, \tag{4.2}$$

(ii) $N(t)$ is the total mass of the tracer in the region at a given time, (iii) N_o is the initial total mass of the tracer in the region, and (iv) N_b is the total background mass of the tracer in the region.

The MTT measures the mean travel time of a conservative (non-decaying) tracer from the repository to an observation point beyond the repository. For the equations defining MTT, see equations 32-35 in [15]. These are reproduced here as Equations 4.3-4.6. Mean travel time, like mean residence time, can be directly measured using tracers. Identical concentrations of two conservative tracers are artificially and continuously injected at a constant rate at the starting point. The only difference between the two tracers is that one of them decays or ingrows exponentially over time since injection. Because the movement of these tracers within the domain is identical, the difference in concentration at a distant location is solely due to the mean time since tracer injection. This measure of time is considered the mean travel time (MTT) of a conservative tracer. We note a difference between MRT and MTT: for MRT, there is one initial concentration of the

conservative tracer at the beginning of the simulation and no more of it is injected into the region, then MRT is obtained by tracking the tracer mass remaining in the repository over time. In contrast, for MTT, two tracers are continuously injected and the differences in their concentrations at locations beyond the injection region can be used to calculate the MTT of injected tracer as defined below.

Exponential decay (or ingrowth) is described by the equation:

$$C(t) = C_o e^{-rt} \tag{4.3}$$

where C is the concentration at time t, C_o is the initial concentration, and r is the rate of the reaction. The rate can be calculated from the half-life $t_{1/2}$ using the equation:

$$r = \frac{\ln(2)}{t_{1/2}} \tag{4.4}$$

Solving for t gives:

$$t = \frac{-\ln\left({}^{C(t)}/_{C_o}\right)}{r} \tag{4.5}$$

In terms of MTT, C_o is analogous to the concentration of stable tracer C_s, and $C(t)$ is analogous to the concentration of unstable (decaying or ingrowing) tracer C_u. Thus, MTT is estimated from the equation:

$$\text{MTT} = \frac{-\ln\left({}^{C_u}/_{C_s}\right)}{r} \tag{4.6}$$

This formulation assumes that initial background concentrations of these tracers, the ingrowing concentrations of an unstable background tracer, and trace concentrations of these tracers flowing into the domain from flux boundaries are negligible compared to the concentrations observed at the monitor location. In the analysis below, the MTT is calculated at the location of the maximum ^{129}I concentrations in the aquifer at each time step of interest. It is considered a characteristic of the system and is not the mean travel time of the ^{129}I concentration observed at that location at that time.

The MRT and MTT evaluated at 10^6 years (1 Ma) were of interest because these represent the total flow for the simulation; values at 10^2, 10^3, 10^4, and 10^5 years were also analyzed. The repository MRT and MTT as calculated from the tracers were also recorded at times corresponding to peak ^{129}I. These values are also of interest because they reflect the flow associated with the highest ^{129}I concentrations.

Table 4-2. New QoIs for crystalline reference case sensitivity analysis

#	Quantities of Interest (QoIs)	Scalar or Vector
1	Maximum ^{129}I concentration in aquifer in mol/L (M)	Max concentration vector that is a function of time; vector includes (X,Y,Z) location of concentration
2	Peak ^{129}I concentration in aquifer in mol/L (M)	Scalar with (X,Y,Z) location of peak
3	Repository MRT (mean residence time in repository) in years	Vector that is a function of time τ. Each value in the vector is the MRT evaluated over the interval $[0,\tau]$.
4	Repository MRT (1 Ma) in years	Scalar, MRT evaluated over the interval [0, 1 Ma]
5	MTT (mean travel time from repository to aquifer location of maximum ^{129}I concentration) in years	Vector that is a function of time. Each value of MTT is calculated from the tracer concentration ratio, evaluated at time τ.
6	MTT (1 Ma) in years	Scalar, MTT as calculated from the tracer concentration ratio, evaluated at 1 Ma
7	MTT for Peak in years	Scalar, MTT as calculated from the tracer concentration ratio, evaluated at the time of Peak ^{129}I concentration
8	Water flow rate rock to aquifer (kg/y)	Vector that is a function of time
9	Water flow rate east boundary (kg/y)	Vector that is a function of time

As an example of the spatial and temporal resolution of these QoIs, the first QoI listed in Table 4-2 is plotted in Figure 4-3 for two of the 800 PFLOTRAN simulations. These plots show maximum ^{129}I concentration in the aquifer plotted as a function of (x,y) location and also as a function of time (z-axis). The left column of Figure 4-3 shows a 3-D view and the right column shows a front view in the x-direction. The top row is DFN1, epistemic realization 1 and the bottom row is DFN1, epistemic realization 5. The plots in Figure 4-3 are colored by the magnitude of the maximum ^{129}I concentration so that one can see (for example) that early in the simulation, the maximum values tend to be in one corner of the domain and are low: in later years, the maximum ^{129}I concentration is more centrally located and is on the order of 1.E-8M.

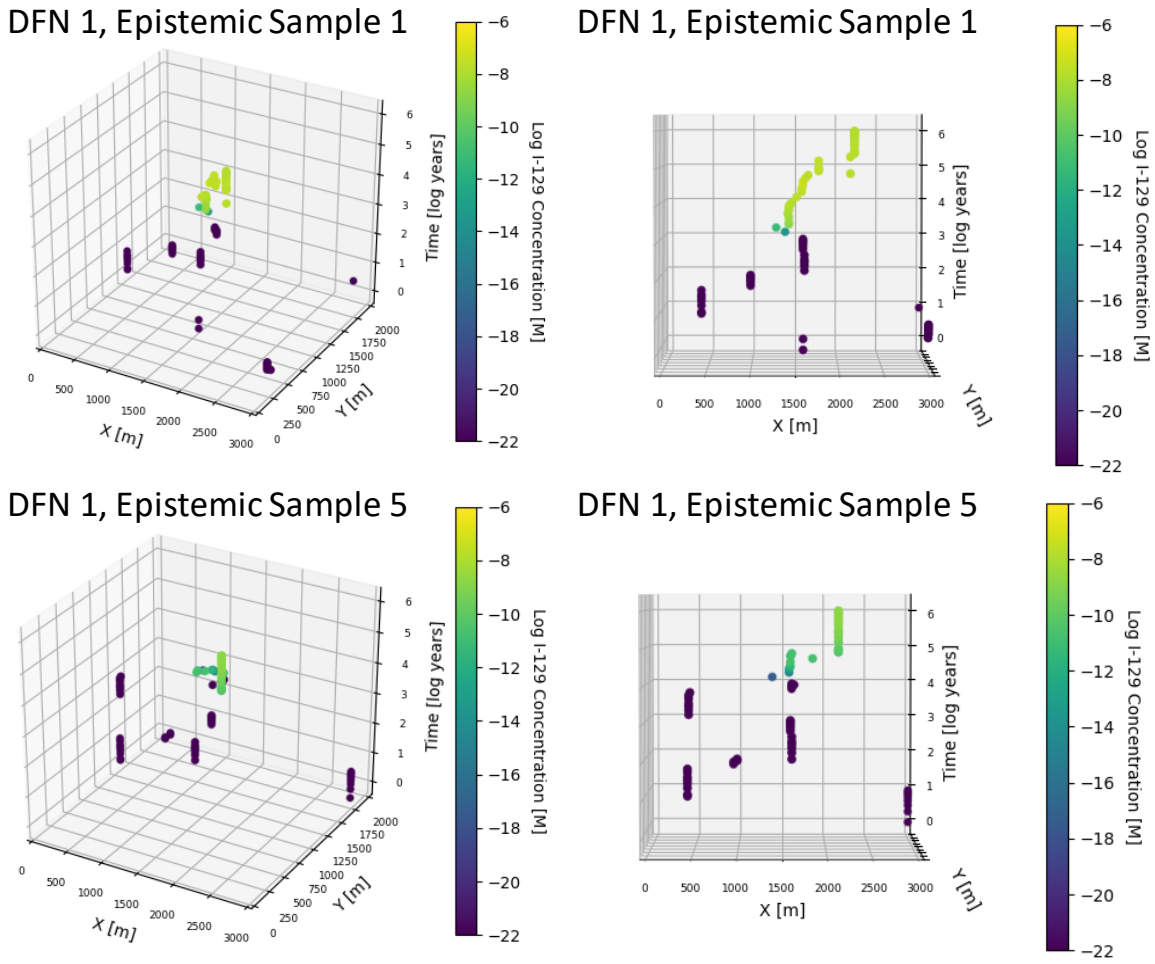


Figure 4-3. Example 3-D plots of maximum ^{129}I concentration in the aquifer shown for two of the 800 PFLOTRAN simulations. The upper row shows the results from DFN1, epistemic sample 1; the bottom row shows DFN1, epistemic sample 5. There are two views on each row: a 3-D view looking down (left column) and a frontal view (right column). The spread and location of the maximum ^{129}I concentration varies for each of the 800 runs.

4.3 Results

The uncertainty analysis described above was extensive, with spatial and epistemic parameter sampling loops generating 800 PFLOTRAN runs on a high-performance computing system. The average run time was 1.54 hours on 32 nodes or 512 processors. The total number of processor hours for the full set of 800 PFLOTRAN runs was approximately 631,000 processor hours.

The uncertainty and sensitivity analyses involved a number of time-dependent quantities of interest, and responses for which the quantity of interest was determined as the maximum (e.g.,

maximum ^{129}I concentration) over a spatial domain (i.e., the aquifer) at a particular time or for which the quantity was the peak concentration both spatially and temporally: the maximum over all time points. Below, we provide a variety of summary plots and statistical analysis of results, including cumulative distribution functions of results, scatterplots and interaction plots, and sensitivity analysis.

4.3.1 Distribution of output quantities

In this section, we first show the overall distribution of several output quantities of interest using cumulative distribution function (CDF) plots. At each time step, we calculate the maximum value of ^{129}I concentration over all locations in the aquifer for each of the 800 sampling-based calculations. The CDFs of these maximum values for selected time steps are shown in Figure 4-4.

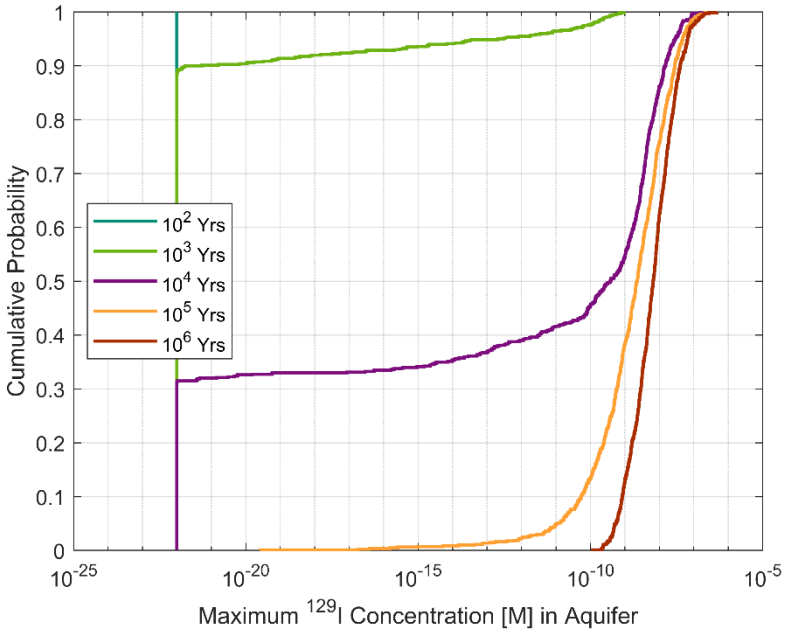


Figure 4-4. CDF of Maximum ^{129}I concentration in Aquifer at various times with step size 1/800.

Figure 4-4 shows the CDFs for maximum ^{129}I values at particular time points. Figure 4-5 shows the CDF of the peak ^{129}I concentration in the aquifer, where the peak for each sample is simply the maximum concentration for that sample over all time points, regardless of location. Note that this peak ^{129}I concentration varies by over three orders of magnitude.

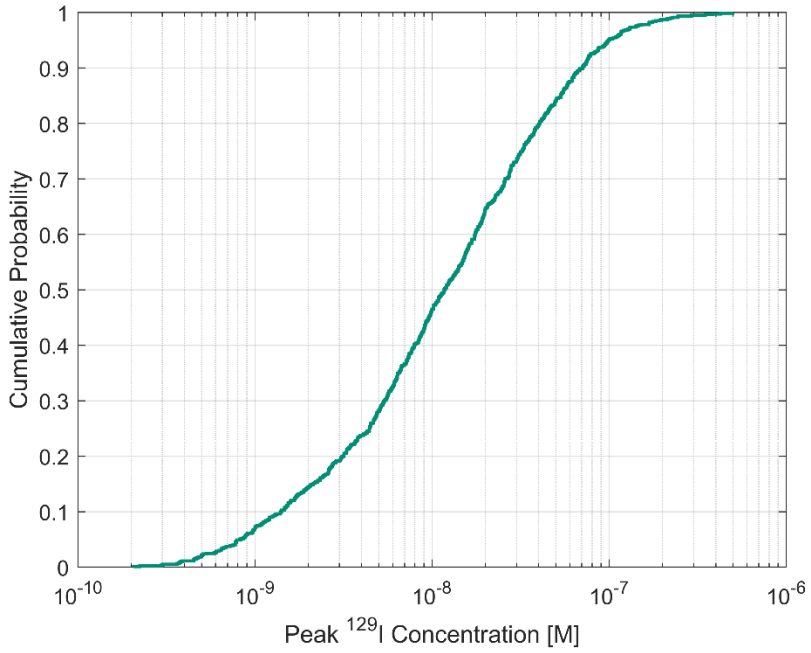


Figure 4-5. CDF of Peak ^{129}I concentration in aquifer with step size 1/800.

Because this Peak ^{129}I concentration is a significant quantity of interest, we also show two ways of slicing the data: CDFs for each spatial realization showing the variation in parameter uncertainty per CDF and CDFs per each parameter realization showing the variation in spatial uncertainty. These are shown below in Figure 4-6 and Figure 4-7.

Each curve in Figure 4-6 shows the CDF of the Peak ^{129}I concentration for a spatial realization across all parameter realizations; hence, 40 points on each (spatial) curve. The same data are plotted in Figure 4-7 but instead each curve plots the CDF of a parameter realization across all spatial realizations; hence, 20 points on each (parameter) curve in Figure 4-7.

The CDFs in black in Figure 4-6 and Figure 4-7 are the CDFs of the means of the peak ^{129}I concentrations in each spatial realization across all parameter realizations (Figure 4-6) and of each parameter realization across all spatial realizations (Figure 4-7). The CDFs of the means provide the best measure of repository performance of all the parameters calculated in these simulations. They provide the calculated spatial and parameter uncertainty distributions about the mean Peak ^{129}I concentration in the aquifer. A comparison of these figures clearly shows that the overall parameter uncertainty in the mean Peak ^{129}I concentration in the aquifer exceeds the overall spatial uncertainty. This observation is best illustrated by the ranges of the CDFs of the mean peak ^{129}I concentrations (black curves). In Figure 4-6 the range across all parameter realizations (40 points) is approximately 2 orders of magnitude, while in Figure 4-7 the range across all spatial realizations (20 points) is slightly greater than 1 order of magnitude. This relationship is opposite of what is observed at a fixed monitor location (Mariner et al. 2019 [16]).

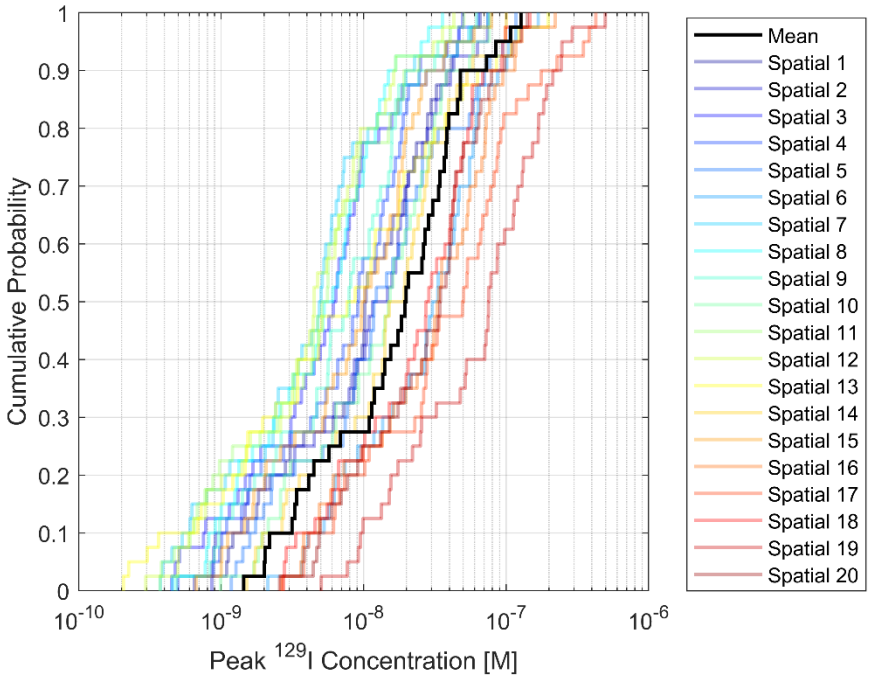


Figure 4-6. CDFs of Peak ^{129}I concentration for each spatial loop realization with step size 1/40 (20 CDFs).

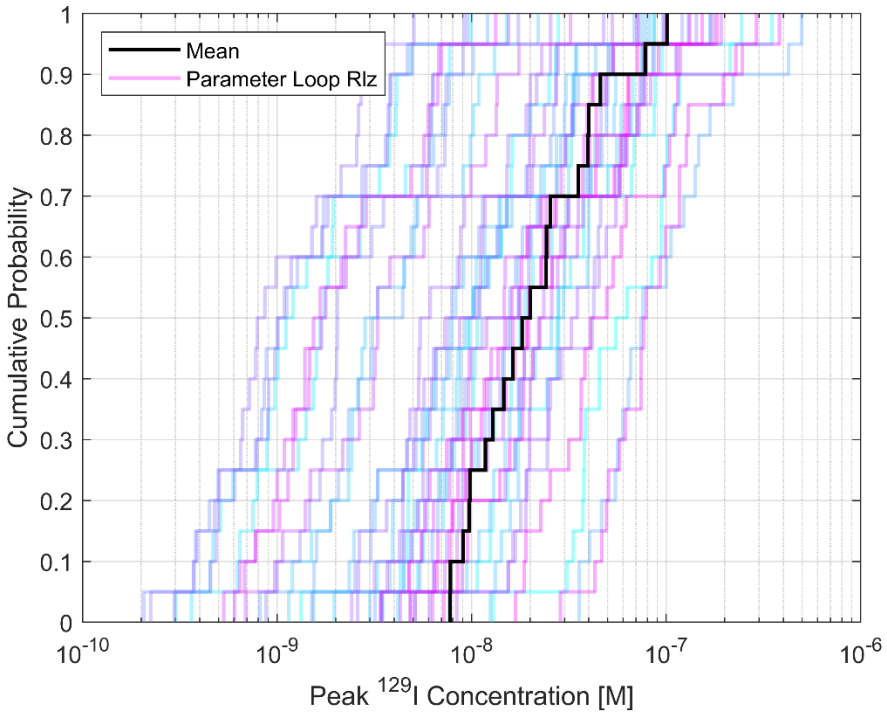


Figure 4-7. CDFs of Peak ^{129}I concentration for each each epistemic parameter realization (Rlz) with step size 1/20 (40 CDFs).

The next set of figures shows the CDFs for the MTT of a conservative tracer from the repository to the aquifer location of the maximum ^{129}I concentration (Figure 4-8) and the MRT of an initial conservative tracer in the repository (Figure 4-9). Figure 4-8 shows the CDF of the MTT evaluated at one million years, also referred to as Mega-annum (Ma). Figure 4-9 shows the CDF for the MRT in the repository region on [0, 1 Ma]. Note that the CDF of MTT is less smooth than that for the MRT. Residence time depends only on a tracer concentration in the repository region while travel time calculations depend on tracer transport through the rock and aquifer domains.

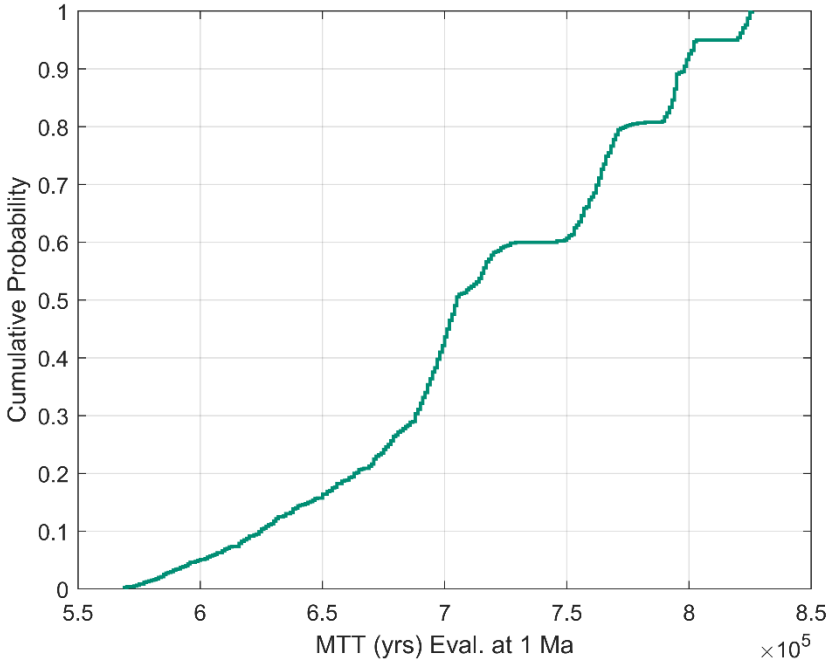


Figure 4-8. CDF of MTT of a conservative tracer from the repository to the location of the maximum ^{129}I concentration at 1Ma with step size 1/800.

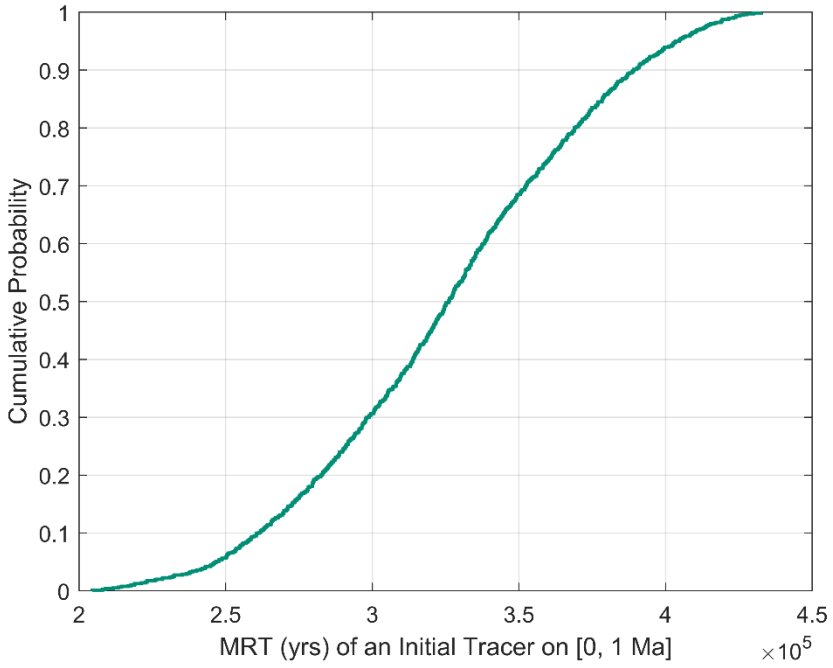


Figure 4-9. CDF of MRT of an initial conservative tracer within the repository on [0, 1 Ma] with step size 1/800.

Figure 4-10 and Figure 4-11 show the CDFs of the maximum value of the rock to aquifer flow rate and the flow rate at the east boundary of the domain.

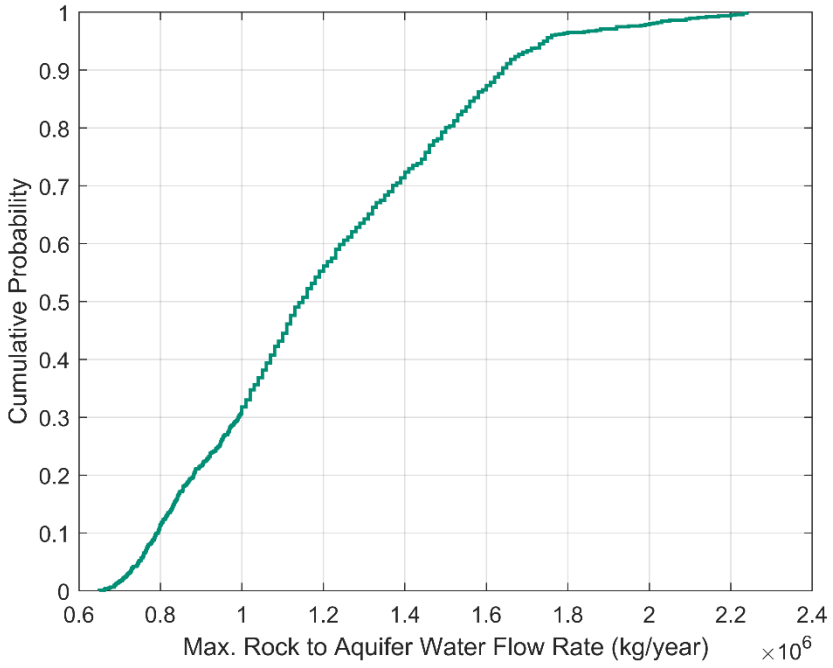


Figure 4-10. CDF of the maximum water mass flow rate from the rock to the aquifer (kg/year) with step size 1/800.

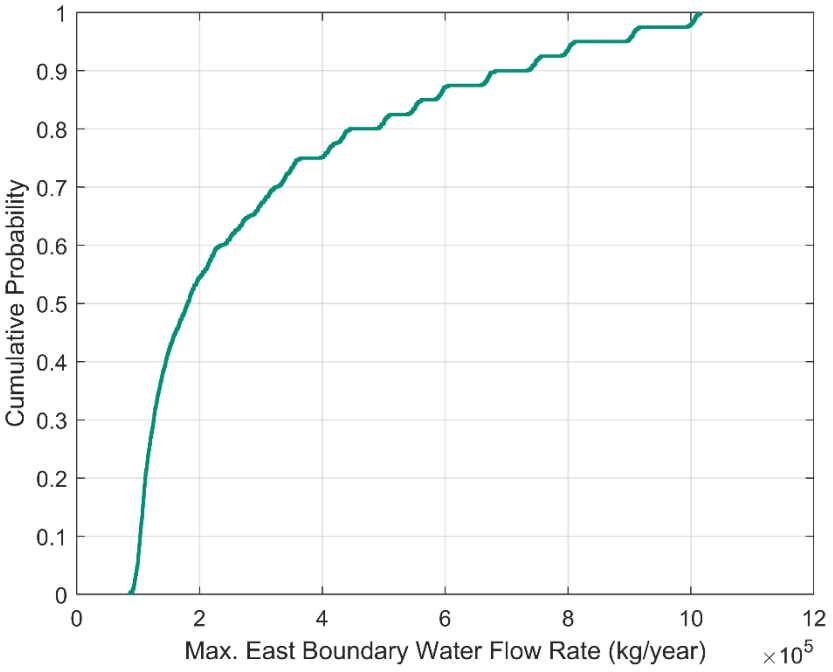


Figure 4-11. CDF of the maximum water mass flow rate at the east boundary of the domain (kg/year) with step size 1/800.

4.3.2 Scatterplots: Inputs to Outputs

The following set of plots show scatterplots of input quantities (on x-axis) to various output quantities (on y-axis). The first output quantity plotted is peak ^{129}I concentration. Figure 4-12 shows the peak ^{129}I values with respect to the samples of the epistemic parameters. Note that the many values for a particular input parameter setting (long vertical lines) is due to the spatial sample realization: a given epistemic parameter is paired with 20 spatial realizations. There are 40 parameter realizations and the variability in peak ^{129}I concentration at each sample value is due to the 20 spatial realizations run with that parameter sample. Figure 4-12 shows the importance of *kGlacial* and *meanWPrate* on the peak ^{129}I values.

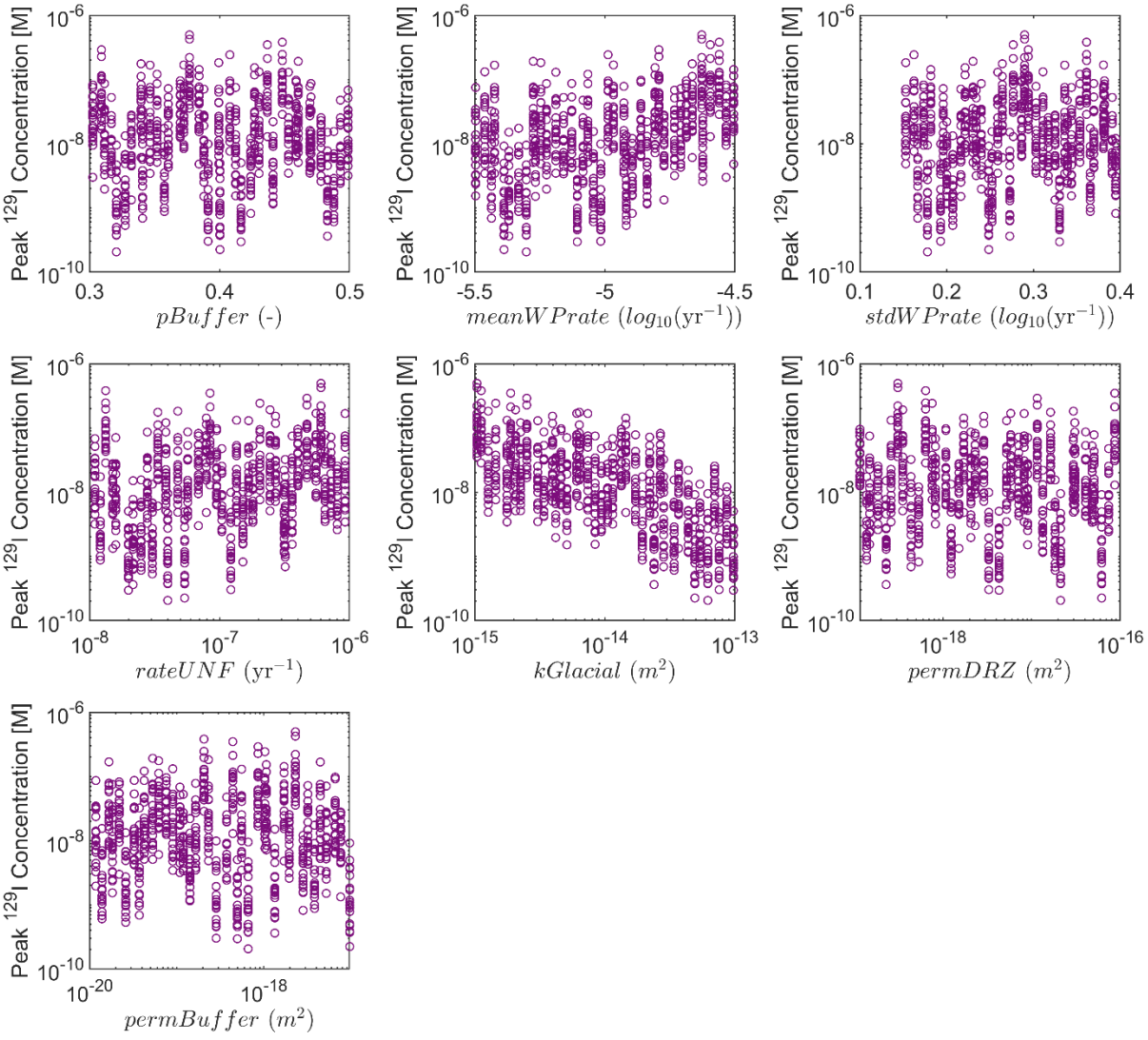


Figure 4-12. Scatterplots of Peak ^{129}I concentration vs. epistemic input parameters defined in Table 4-1.

Figure 4-13 through Figure 4-15 show the output quantity vs. the same set of input parameters, but the responses are also color coded with respect to the peak ^{129}I concentration so that one can see the interaction of ^{129}I with respect to some of these outputs.

Figure 4-13 shows the scatterplots of MTT (at 1Ma) with respect to the input parameters. We note that *meanWPrate*, *stdWPrate*, and *rateUNF* cannot have any effect on travel time due to the model construction. When *pBuffer* increases, the effective diffusion coefficient in the buffer increases, which would tend to decrease travel time (diffusive behavior). Also, when *pBuffer* increases, the average linear porewater velocity decreases which would increase travel time (advective behavior). There are some interesting observations from these plots: it is clear that there are two groupings of results: the one on the top with high mean travel times appears to be from 7 spatial realizations and the bottom from the remaining 13 spatial realizations. It is likely that there are certain characteristics of the 7 DFNs that lead to larger travel times: we are still investigating these properties, but analysis of additional DFNs (beyond the 20 generated for this analysis) suggests that this relationship is due to random chance and the pattern between the arbitrary spatial realization number and QoIs would not persist with a higher DFN sample size. Also, one can see input parameter effects. For example, high values of *kGlacial* tend to be correlated with low values of ^{129}I and low values of *kGlacial* with high values of ^{129}I.

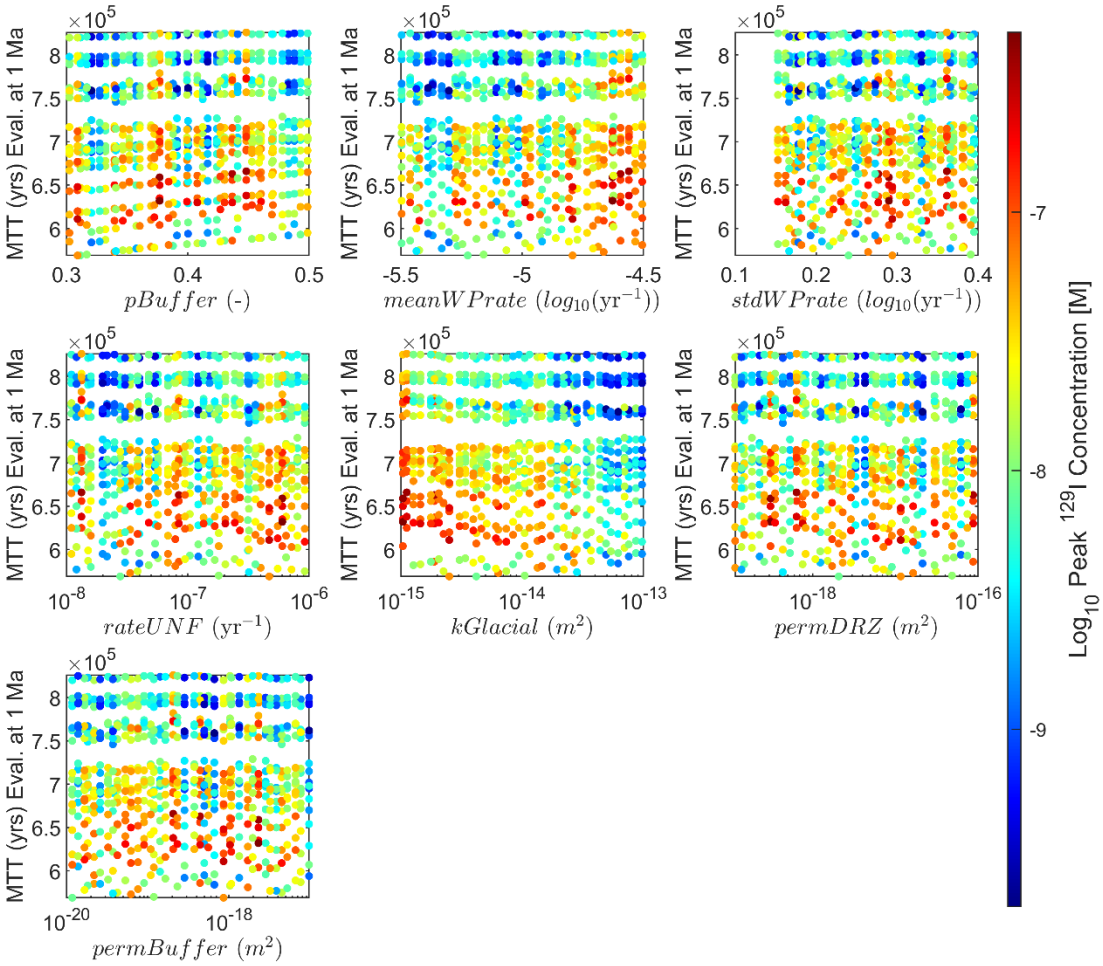


Figure 4-13. Scatterplots of MTT of a conservative tracer from the repository to the aquifer evaluated at 1Ma vs. epistemic input parameters defined in Table 4-1, colored by Peak ^{129}I.

Figure 4-14 shows the same type of scatterplot for the repository MRT evaluated from zero to one million years. Again, there is a clear effect of DFNs. There is a strong correlation between mean residence time and pBuffer, for example. This is likely due to the average linear porewater velocity behavior, where the porewater velocity decreases as pBuffer increases which results in an increase in residence time. That is, it takes longer to flush a region of higher volume given the same flow rate. However, we also note that higher values of pBuffer do not correlate with peak ^{129}I concentrations, as lower peak ^{129}I concentrations are observed across the pBuffer spectrum. This implies that while pBuffer increases the repository MRT for a given DFN, the major differences in MRT between DFNs are not due to buffer porosity but instead due to the DFNs themselves. Clearly, the cooler colors of lower peak ^{129}I concentration are observed in the pBuffer plot for DFNs that result in greater repository MRT. Again, we see that the high values of kGlacial are associated with low values of ^{129}I.

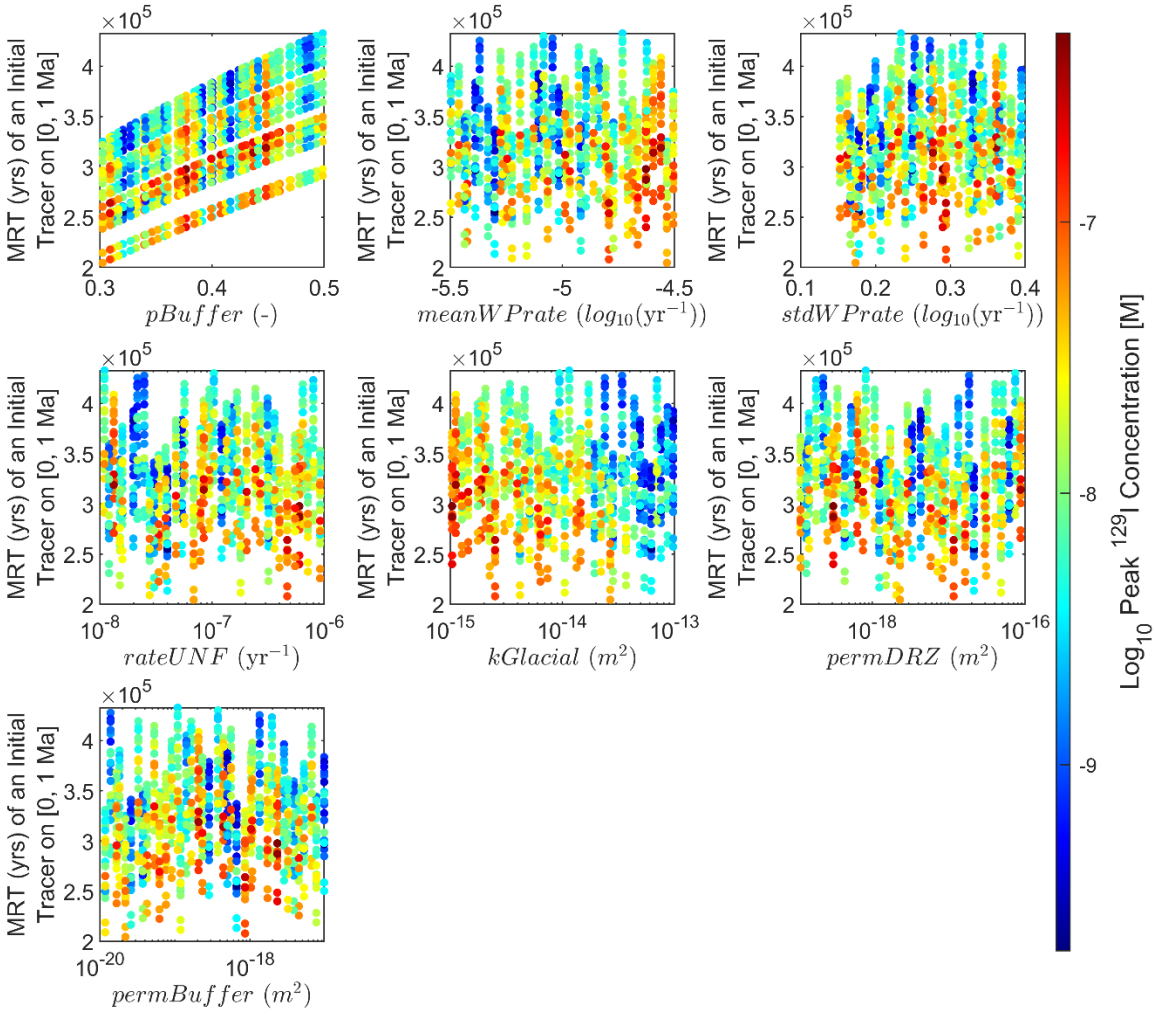


Figure 4-14. Scatterplots of MRT of an initial conservative tracer evaluated on [0, 1Ma] vs. epistemic input parameters defined in Table 4-1, colored by Peak ^{129}I.

Figure 4-15 shows the scatterplots with respect to the maximum flow rate from the rock to the aquifer. The maximum rock to aquifer flow rates observed during the simulations occur at the beginning of the simulations around 1 year when the newly emplaced waste packages heat up the water in the repository and cause it to expand. Thus, this causes a somewhat artificial pulse of water into the aquifer from the underlying rock that can be used to help characterize the bulk connection between the repository and aquifer. This response quantity does not correlate as strongly with the peak ^{129}I concentration. pBuffer is the most important variable affecting maximum water mass flow rate from rock to aquifer because with increased porosity there is more water to expand.

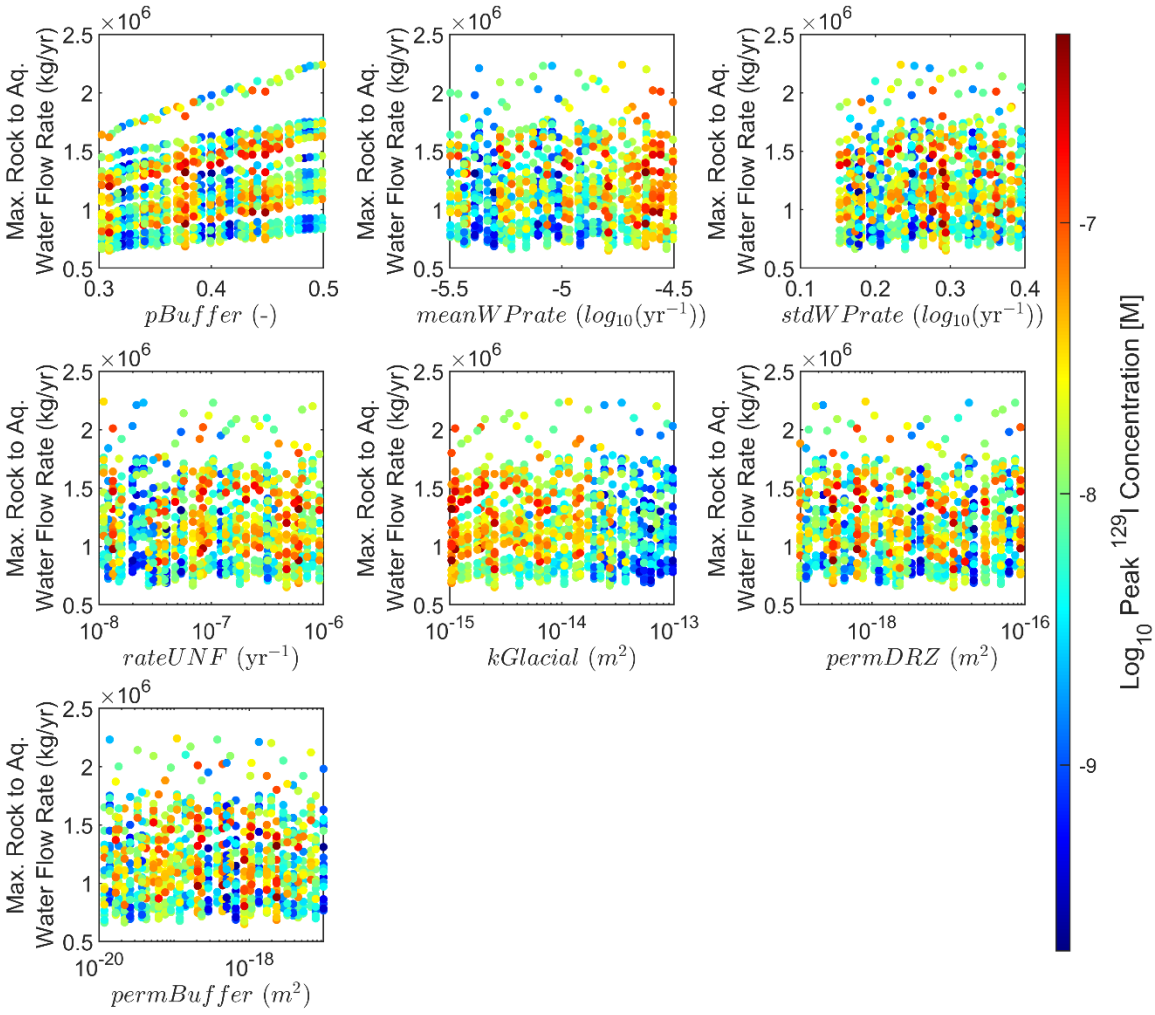


Figure 4-15. Scatterplots of maximum water mass flow rate from rock to aquifer vs. epistemic input parameters defined in Table 4-1, colored by Peak ^{129}I.

4.3.3 Scatterplots: Outputs to Outputs

The analysis of the crystalline reference case presented here has many new output quantities that we have added to the PFLOTRAN code base specifically to gain a better understanding of host rock characteristics, the influence of discrete fracture networks, and an understanding of the fluxes throughout the system. Our ultimate goal is to understand which DFNs are "leakier" and how this affects travel time, residence time, and peak concentrations over time. Thus, it is helpful to plot the peak ^{129}I concentration against various outputs and also to plot output quantities as a function of time against each other.

First, we plot peak ^{129}I vs. the other integral quantities at the final time, as shown in Figure 4-16 below. Note that Peak ^{129}I is negatively correlated with MTT and MRT, as well as with the peak flow rate at the east boundary. The Mean Travel Time for the Peak ^{129}I concentration has two distinct groupings of results.

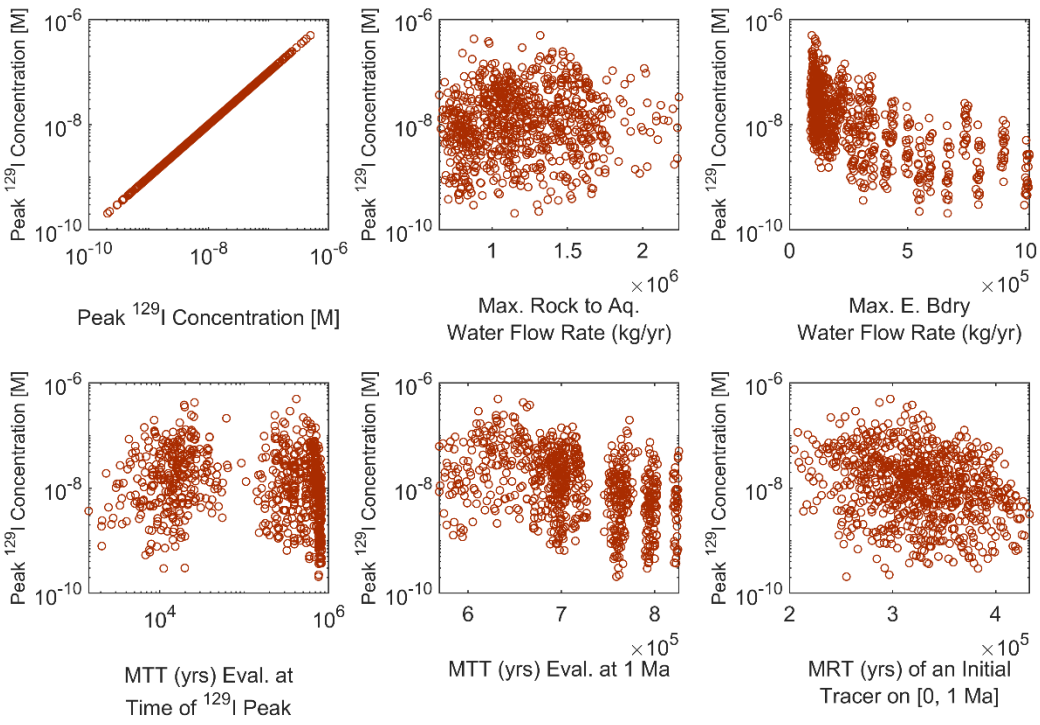


Figure 4-16. Scatterplots of peak ^{129}I concentration vs. other output quantities defined in Table 4-2 for all 800 simulations.

We have a rich set of time-dependent data. Figure 4-17 shows the max ^{129}I concentration as a function of the other responses at 5 time points, from 100 years to 1M years. Note that max ^{129}I concentration values on the y-axes are the maximum ^{129}I value over the spatial domain at that time point, whereas peak ^{129}I (x-axis, first column) represents the peak ^{129}I value over all times. This plot shows some trends especially at later times but not as much correlation as we expected.

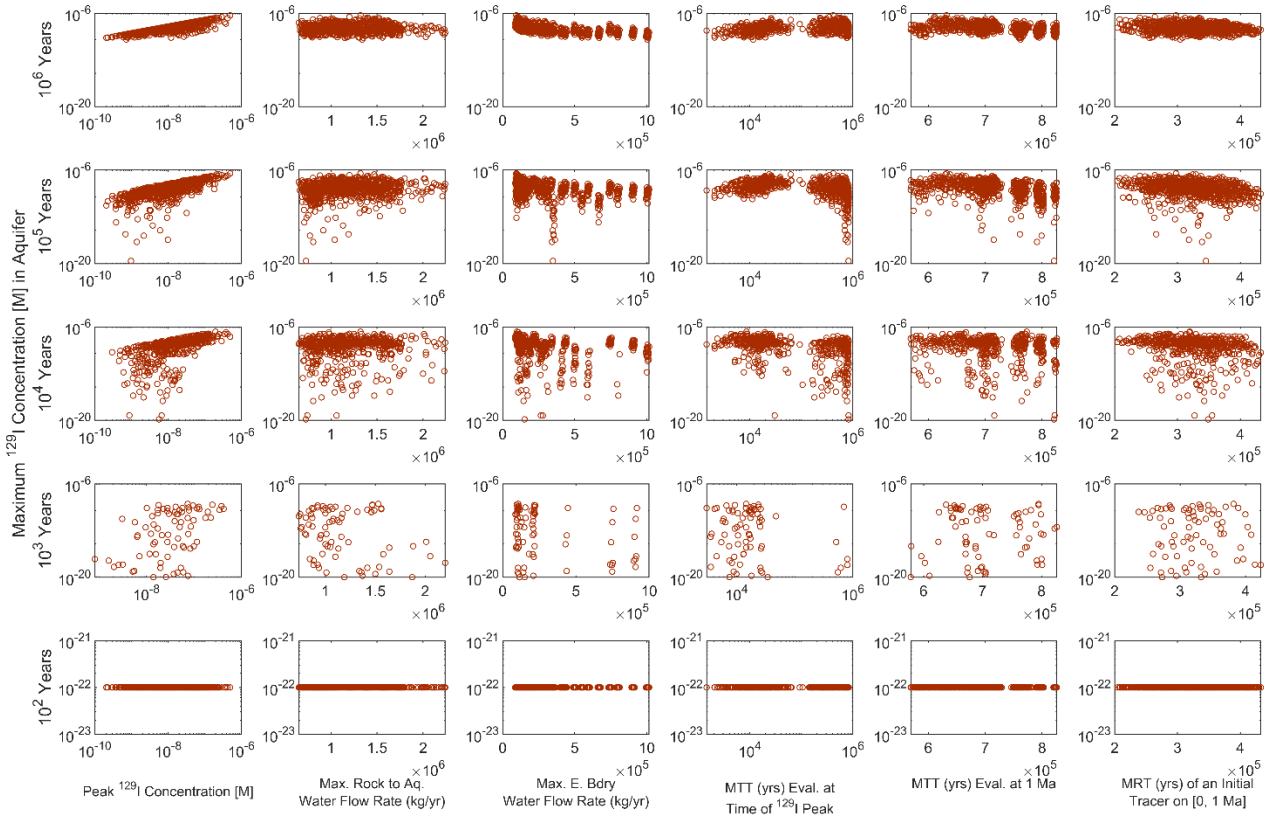


Figure 4-17. Scatterplots of maximum ^{129}I concentration at various time points vs. other output quantities for all 800 simulations.

Figure 4-18 shows a similar plot with MTT at different time points on the y-axes vs. the other quantities on the x-axes. Again, we can see some time dependence, especially at 1M years where mean travel time is negatively correlated with peak ^{129}I concentration and peak flow rate from the rock to aquifer, and positively correlated with mean residence time. However, the distinctive striations or clusters of points in the plots are the result of the spatial uncertainty. Within each spatial realization, the relationships can be different. For example, there does not seem to be a strong relationship between mean travel time and peak ^{129}I within a spatial "striation" at 1 million years as shown by nearly horizontal lines on the top row of Figure 4-18. For the relationship between MTT and rock to aquifer flow rate, there appears to be a small positive correlation within each spatial striation.

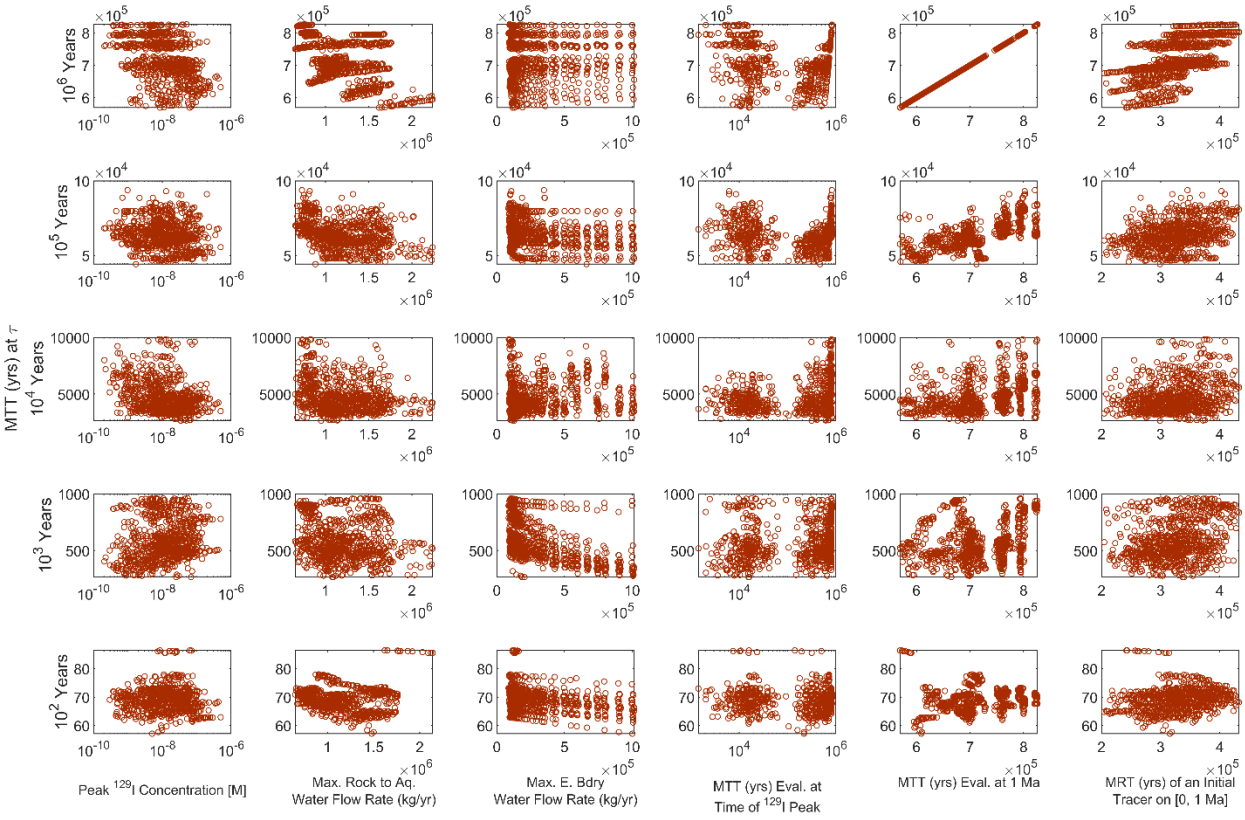


Figure 4-18. Scatterplots of MTT at various time points vs. other output quantities.

Figure 4-19 shows a plot with MRT at different time points on the y-axes vs. the other responses on the x-axes. Here, we see strong correlation patterns of MRT with respect to peak MTT and flow rate from rock to aquifer. This correlation is consistent across DFNs, which is why the scatter plots for these quantities against MRT have distinct clusters of points. The rock to aquifer flow rate is positively correlated with the MRT for a fixed spatial realization, as was the case for MTT.

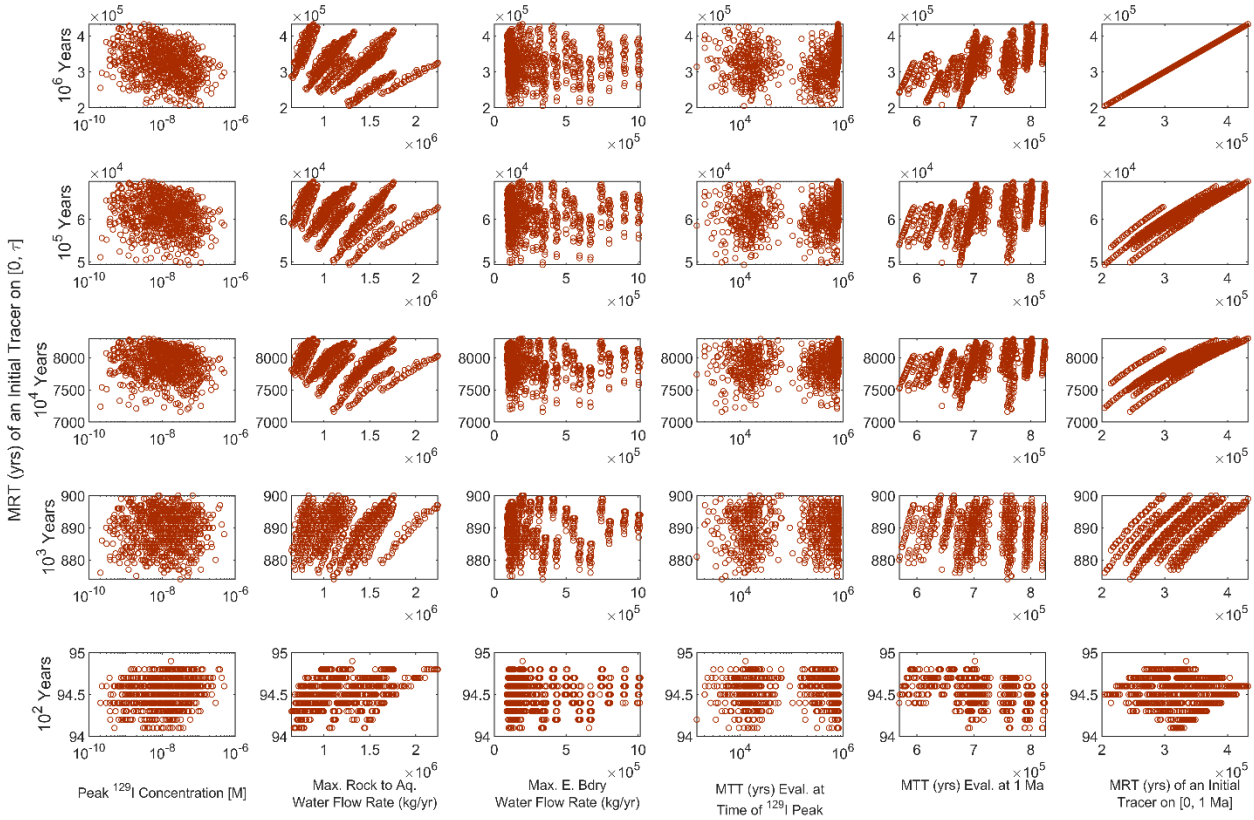


Figure 4-19. Scatterplots of MRT at various time points vs. other output quantities.

The scatterplots of the flow rate from the rock to the aquifer is shown vs. other response quantities in Figure 4-20 below. Unlike the MTT and MRT, the correlation between on the Rock to Aquifer Flow Rate and other output quantities is more pronounced earlier in the simulation.

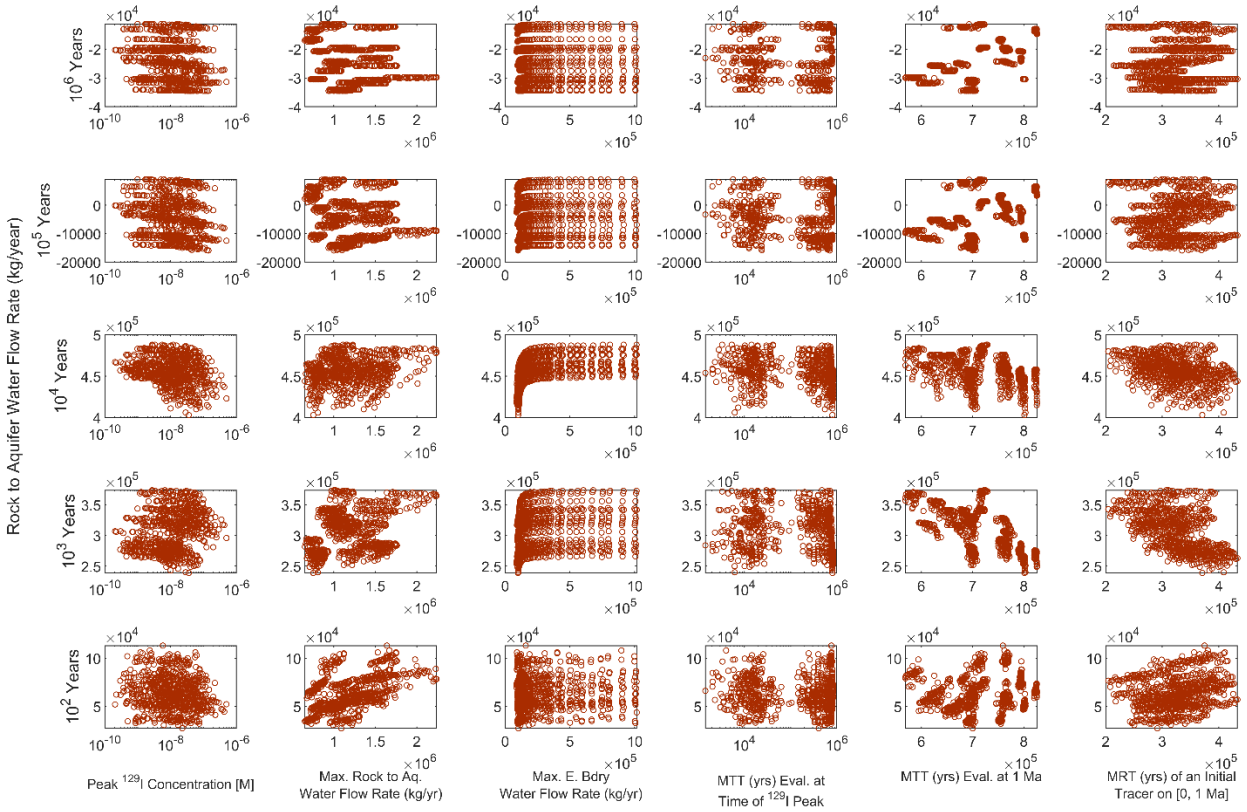


Figure 4-20. Scatterplots Rock to Aquifer flow rate vs. other output quantities.

4.3.4 Graph Analysis

This year, a number of graph-based analyses on the DFNs were performed to obtain more information about the nature of the DFNs, with the goal of correlating various graph attributes with the output quantities of interest. Graphs are a powerful tool for representing DFNs because they are able to capture the inherent network topology of the DFN without the high computational cost of meshing thousands of fractures ranging in size from millimeter to kilometer. The difference in computational cost is striking—derived quantities from a graph representation of a DFN can be computed in a matter of seconds or minutes, while a fully-resolved, 3D DFN flow and transport simulation would require several hours to run. Graphs have been used with some success as reduced-order models (with corrections) for predicting flow and transport through the DFN, and as surrogates in UQ propagation studies in this context [11,12].

These promising applications of graphs as surrogates of DFN connectivity motivated this work. The goal was to identify correlations between graph QoIs and metrics of repository performance derived from the PFLOTRAN simulations defined in Table 4-2. The metrics of repository performance we sought to correlate with graph QoIs were the Peak ^{129}I concentration, the mean travel time, the mean residence time of a tracer in the repository, and the mass flow rate from the rock to the aquifer. These QoIs are defined and detailed in Table 4-2. Graph QoIs for which such correlations are identified can prove invaluable by providing a quantitative metric to characterize the DFN for use in a sensitivity analysis, thereby providing a more accurate attribution of uncertainty.

The graphs were constructed using dfnWork's dfnGraph utility [11, 12], and postprocessing to obtain QoIs from the graphs were obtained using dfnWorks and NetworkX, a python network analysis package [13]. There are two types of graphs that can be constructed from a DFN: a fracture graph, wherein fractures are assigned to nodes and intersections to edges, and an intersection graph, wherein intersections are assigned to nodes and fractures to edges. Unless otherwise stated, the graph QoIs discussed herein are in terms of the fracture graph.

The graph attributes examined included number of nodes (fractures), number of edges (intersections), average degree (average number of intersections a fracture is part of), relative shortest travel time (as calculated by the dfnFlow algorithm in dfnWorks), and number of intersections with the repository. Note that the graph attributes are defined per DFN. However, within the UA, there are 40 parameter realizations run for each DFN. These means that there are 40 output QoIs associated with each single value of a graph attribute. Thus, the output QoIs (e.g., Peak ^{129}I concentration) were averaged over the parameter realizations. These averaged QoIs were correlated with the graph attributes and examined using scatterplots. Several correlations were identified, but they were relatively weak, with the strongest correlations still only explaining about 20-40% of the variance in the output QoIs. This may be in part because one DFN involves significant stochastic variability of the particular fracture network instantiated. Additionally, the variability of the epistemic samples of the performance metrics for each DFN is significant. It can be expected that the variability across epistemic variabilities is not fully reflected when using a mean value to summarize the performance metric then correlate it with DFN attributes. There are some correlations, however, and we can use the graph metrics both to verify the behavior of the DFNs and to perform sensitivities.

As one would expect, the graph attributes themselves (average degree, number of nodes, and number of edges) are highly correlated—see, for example, Figure 4-21a. Each of these QoIs was positively correlated with the water mass flow rate between rock and aquifer, denoted "Rock to Aquifer Flow Rate" in the scatterplot Figure 4-21b. The fact that these quantities are correlated makes intuitive sense because the graph metrics are measures of the number of pathways in the DFN—the more pathways between rock and aquifer, the greater the mass flow rate will be. However, there is significant variation, especially for larger average degree. This may be because an increased average degree corresponds to more pathways, possibly introducing greater variability in the flow rate. A more targeted analysis would be necessary to confirm or deny this postulation, however.

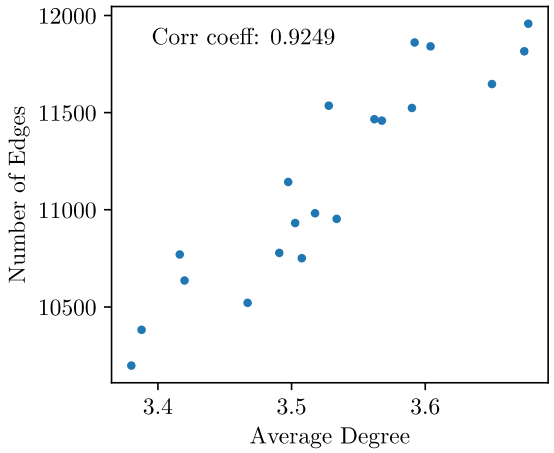


Figure 4-21a A scatterplot of the average degree vs. number of edges, along with the Pearson correlation coefficient.

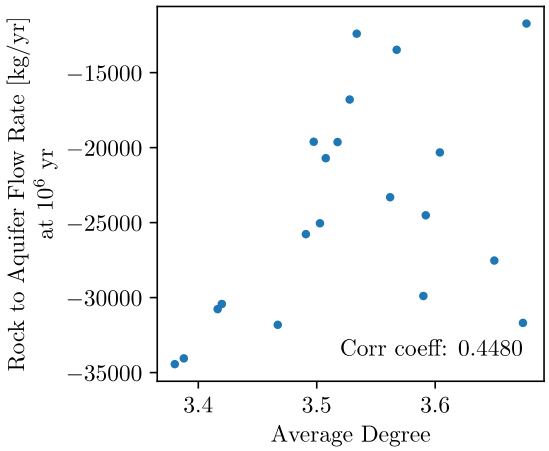


Figure 4-21b A scatterplot of the average degree vs. mass flow rate from rock to aquifer at 1 My, along with the Pearson correlation coefficient.

As is shown in Figure 4-22, the number of intersections with the repository exhibited a negative correlation with the mean residence times in the repository (as defined in 4.2.7) starting at observation times of 1,000 y and later. It makes intuitive sense that the mean residence time in the repository is negatively correlated with with the number of fractures intersecting it, since these intersections increase the pathways out of the repository, making it easier for fluid to flow out.

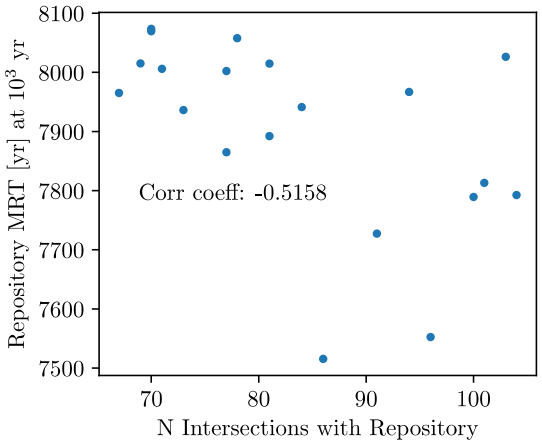


Figure 4-22. A scatterplot of the number of intersections with the repository and the mean residence time (MRT) in the repository at 10 kyr, along with the Pearson correlation coefficient.

The graph QoI that yielded the most correlations to repository performance metrics was the relative shortest travel time (RSTT). The relative shortest travel time is defined by normalizing the graph-based shortest travel time by the minimum graph-based shortest travel time taken over all DFNs:

$$\mathrm{RSTT_i} = \frac{(\text{graph shortest travel time})_\mathrm{i}}{\min\limits_{j}(\text{graph shortest travel time})_\mathrm{j}}, \qquad i,j = 1,2,\ldots,20. \tag{4.7}$$

The procedure for computing the graph-based shortest travel time for each DFN is described herein. For each DFN, a shortest travel time was computed using dfnGraph's flow functionality in dfnWorks, which uses an intersection graph representation of the DFN, with intersections as nodes and fractures as edges, and assigns weights to the edges of the graph based on the area and permeability of each fracture. These are used to estimate a flow time between input and output nodes in the graph, instantiated with inflow and outflow pressures and fluid viscosity specified by the user. The default values were used for these quantities, set to 2×10^6 Pa, 1×10^6 Pa and 8.9×10^{-4} Pa-s respectively. The RSTT was then obtained by scaling all the flow times by the minimum over all DFNs, as is defined in (4.7). Clearly there is no reason to assign any physical relevance to the absolute shortest travel time derived from this procedure, since the pressures and viscosity were assigned in an arbitrary fashion. The meaning of this quantity should be seen as a relative ranking of the speed with which fluid can move from the repository to the aquifer across DFNs.

In accordance with intuition, it was found that the RSTT was positively correlated with the mean travel time of the peak ^{129}I concentration (as defined in 4.2.7) for observation times after 10,000 y, as well as the mean travel time of the maximum ^{129}I concentration taken over all times (see Figure 4-23a and b). Both QoIs are measures of how rapidly fluid flows through the DFN, so it is expected that an increase in one would correspond to an increase in the other.

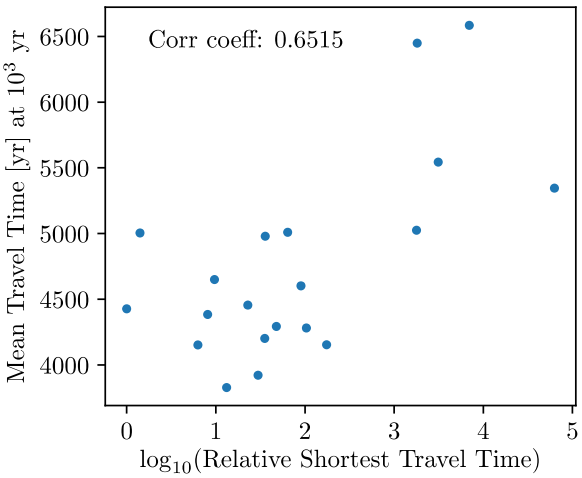


Figure 4-23a A scatterplot of the $\log_{10}$ of the relative shortest travel time vs. the mean travel time of the peak ^{129}I concentration at 10 ky, along with the Pearson correlation coefficient.

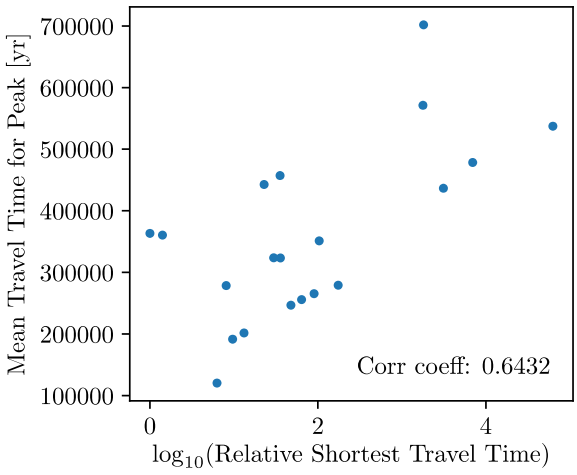


Fig 4-23b A scatterplot of the $\log_{10}$ of the relative shortest travel time vs. the mean travel time of the maximum (peak) ^{129}I concentration taken over all times, along with the Pearson correlation coefficient.

Finally, the RSTT was found to be positively correlated with the rock-to-aquifer mass flow at late times (100,000 and 1 million years), but negatively correlated with the maximum (over time) rock-to-aquifer mass flow, as shown in Figure 4-24. This difference in the sign of the correlation may be explained by the change in sign of the rock-to-aquifer mass flow from early to late time. In early time the mass flow is positive, while it is negative for later times, as can be seen Figure 4-24a. For positive mass flows, as for the peak rock-to-aquifer mass flow, one would expect the negative correlation, since shorter travel times would correspond to DFNs where fluid could flow more rapidly. On the other hand, for the negative mass flows at later time, it is less clear why there would be a positive correlation.

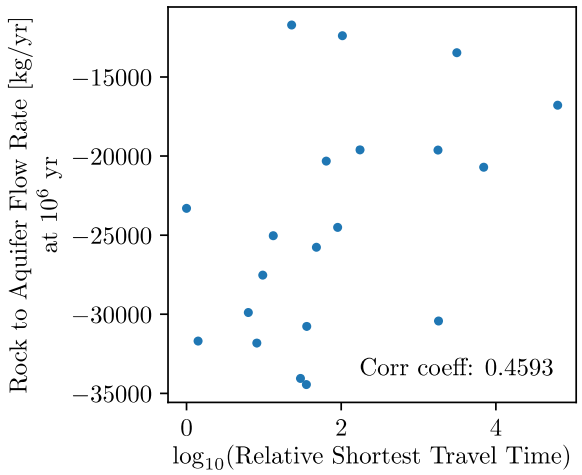


Figure 4-24a A scatterplot of the $\log_{10}$ of the relative shortest travel time vs. the rock-to-aquifer mass flow at 1 million years, along with the Pearson correlation coefficient.

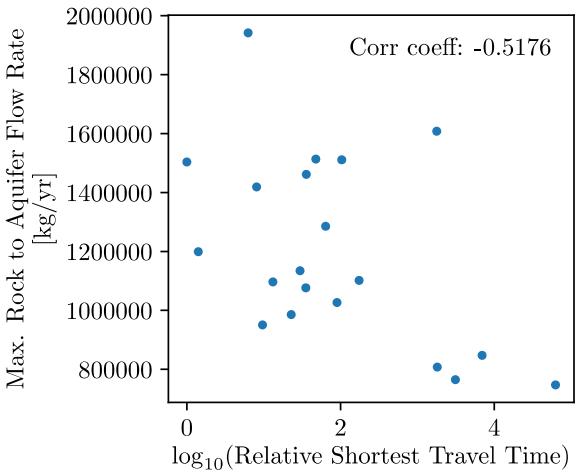


Figure 4-24b A scatterplot of the $\log_{10}$ of the relative shortest travel time vs. the max-over-time rock-to-aquifer mass flow, along with the Pearson correlation coefficient.

Figure 4-25 shows the correlation of RSTT with the mean peak ^{129}I concentration, showing that shorter travel times correlate with higher mean peak ^{129}I concentrations.

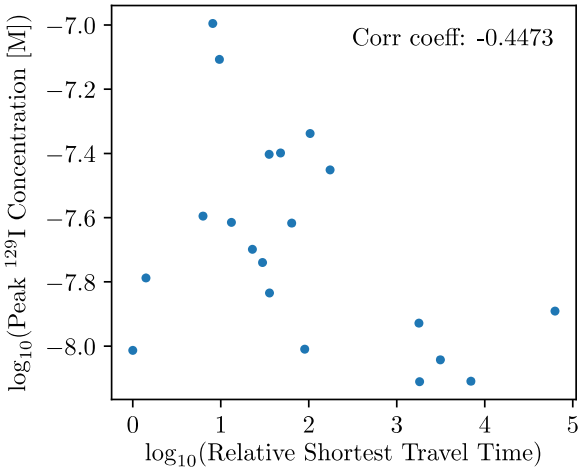


Figure 4-25. Mean of Peak ^{129}I Concentration (averaged over epistemic parameter realizations) vs. RSTT.

Figure 4-26 does not average over the epistemic parameter realizations but shows the graph attribute of relative shortest travel time vs. the peaks of all the QoIs. Note that relative shortest travel time is most strongly correlated with mean travel time from the PFLOTRAN simulations. This is a good sanity check because these metrics quantify the same thing. Note that graphs with larger values of "relative shortest travel time" have the smallest maximum flow rates to the aquifer (third graph from left). These larger values of relative shortest travel time also correspond to longer

residence times in the respository and lower peak ^{129}I concentrations. Striations in the plot are the result of spatial uncertainty; there are 20 vertical clusters, one for each DFN.

These results show promise that correlations between graph QoIs and repository performance metrics of interest can be identified and used for analysis downstream.

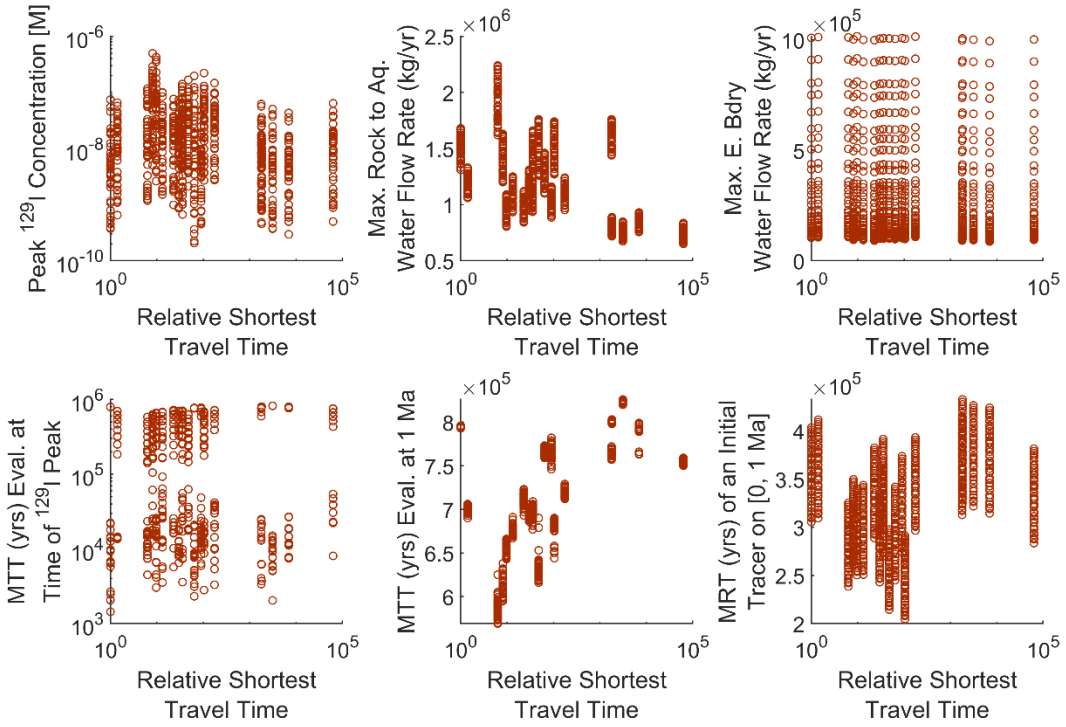


Figure 4-26. Scatter plots of RSTT vs. repository performance QoIs.

4.3.5 Interaction plots

As mentioned, one key focus for this analysis was to identify peak ^{129}I concentration and its relationship to the DFN and associated characteristics. This next series of plots highlights a particular QoI on the y-axis as a function of inputs, but the responses are colored by DFNs. These plots highlight clear patterns with respect to DFN, which is an artifact of the small spatial sample size. Offline, we performed a graph analysis study with the 20 DFNs as well as with a larger set of 100 DFNs. We saw significant correlations of the graph QoIs and DFN number with 20 DFNs but very little correlation with 100 DFNs.

Figure 4-27 is similar to Figure 4-12 but colored by the spatial realization. As discussed, the predominance of the higher DFN numbers associated with larger ^{129}I in this figure is an artifact of having only 20 DFNs. The number associated with a spatial realization is the order in which that DFN was generated by dfnWorks; it has no meaningful relationship to features of the DFN. There are 40 vertical clusters (one for each parameter loop sample) and 20 points within each cluster (one for each spatial loop sample).

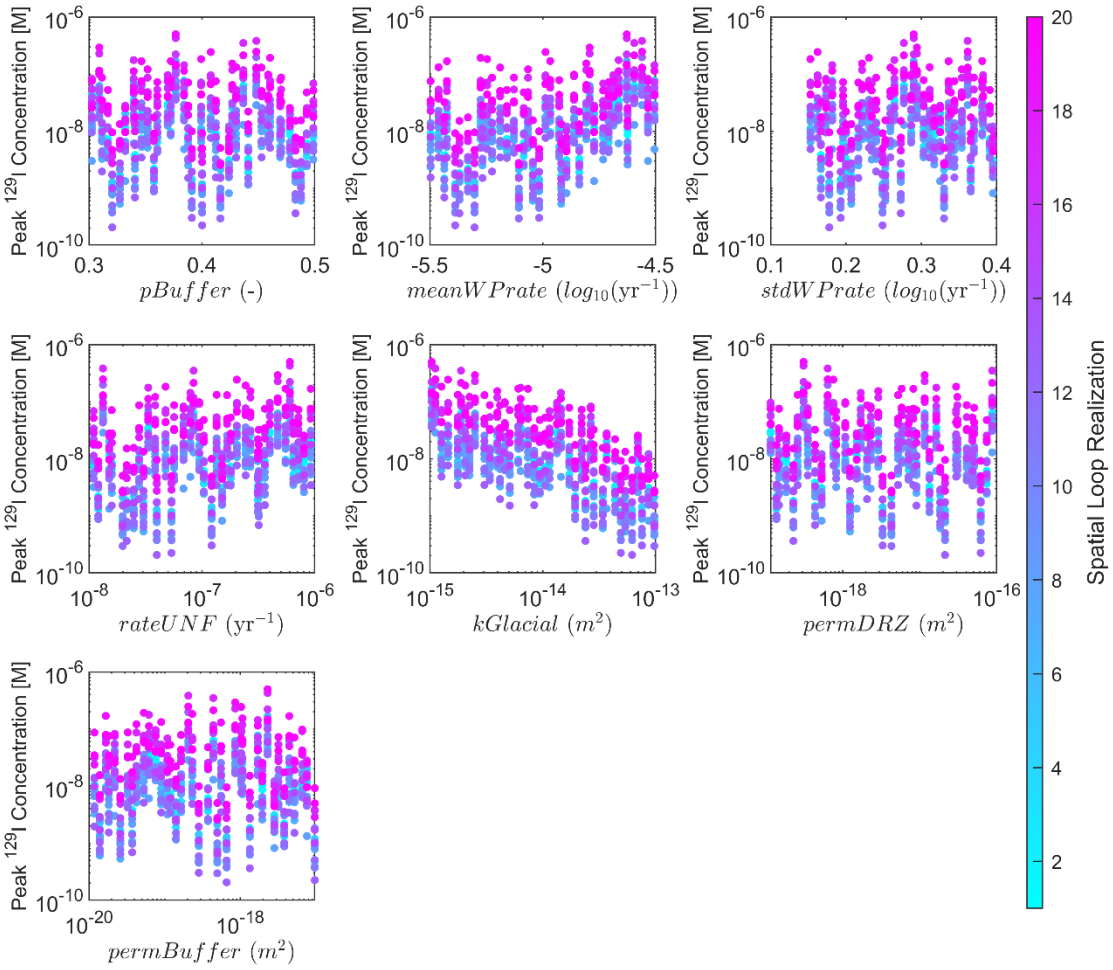


Figure 4-27. Peak ^{129}I Concentration vs. epistemic inputs defined in Table 4-1, colored by DFN.

Figure 4-28 shows a strong dominance of DFN on mean travel time. Note that this figure indicates the mean travel time is almost totally dominated by DFN: there is not a strong influence of the epistemic parameters. The dominance of the DFN uncertainty is why the clusters in the plot (which appear like horizontal striations at different QoI magnitudes) are each a uniform color.

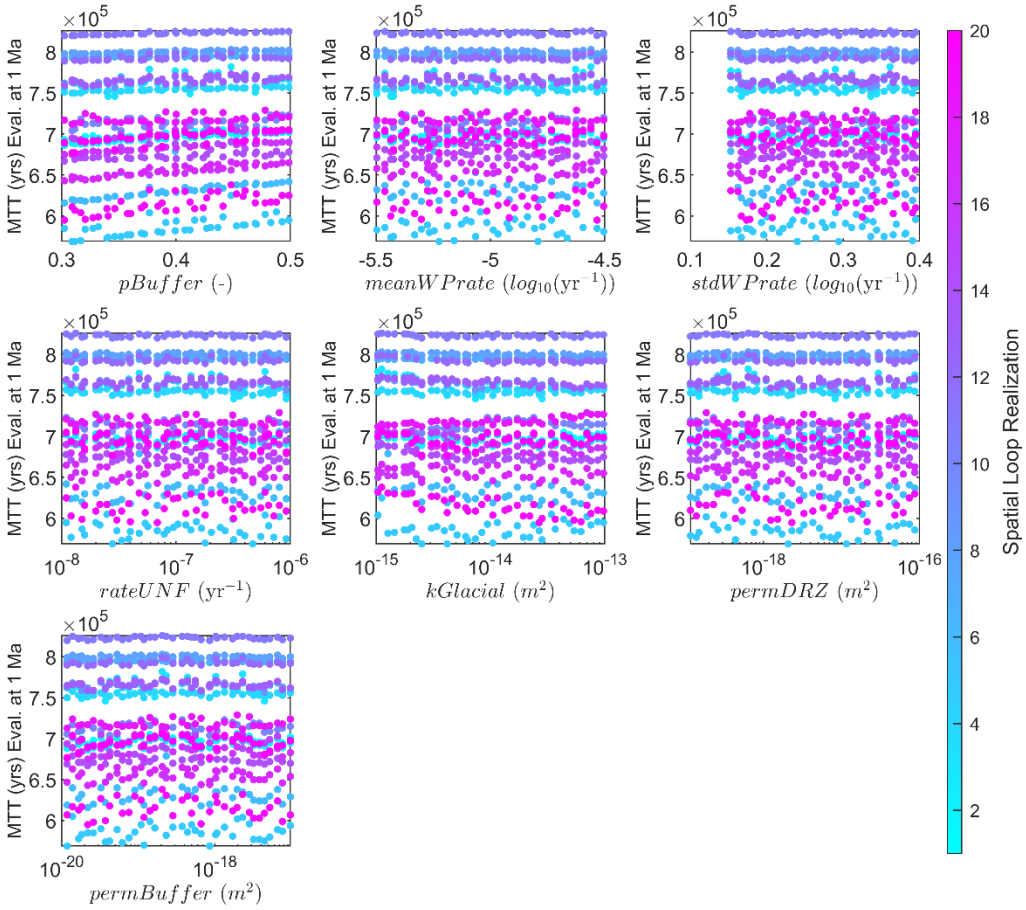


Figure 4-28. MTT at 1Ma vs. epistemic inputs defined in Table 4-1, colored by DFN.

Figure 4-29 shows a similar plot for mean residence time. The correlation between *pBuffer* and the mean residence time appears to be a significant trend, with the DFN effect being somewhat less obvious for other parameters.

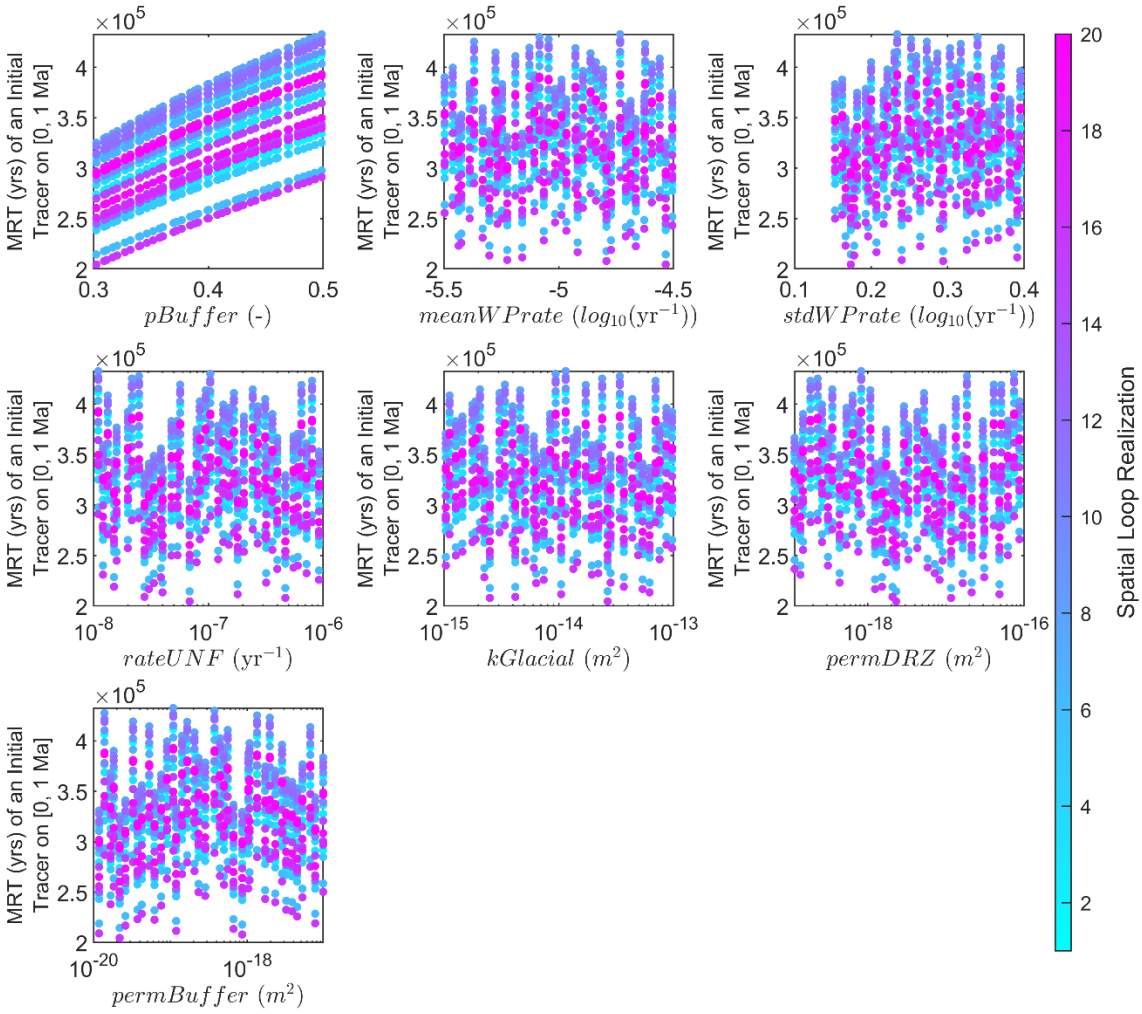


Figure 4-29. MRT at 1Ma vs. epistemic inputs defined in Table 4-1, colored by DFN.

Figure 4-30 shows the interaction between the time of the peak ^{129}I concentration and its value, colored by mean travel time of a conservative tracer from the repository to the aquifer location of the peak ^{129}I concentration. There are some interesting observations from Figure 4-30. The time of the peak ^{129}I concentration is divided into two groups. For the group with the earlier time of peak, the earliest peaks occurred with shorter (blue) travel times. The second grouping showed increasing travel times as the time of the peak concentration increased.

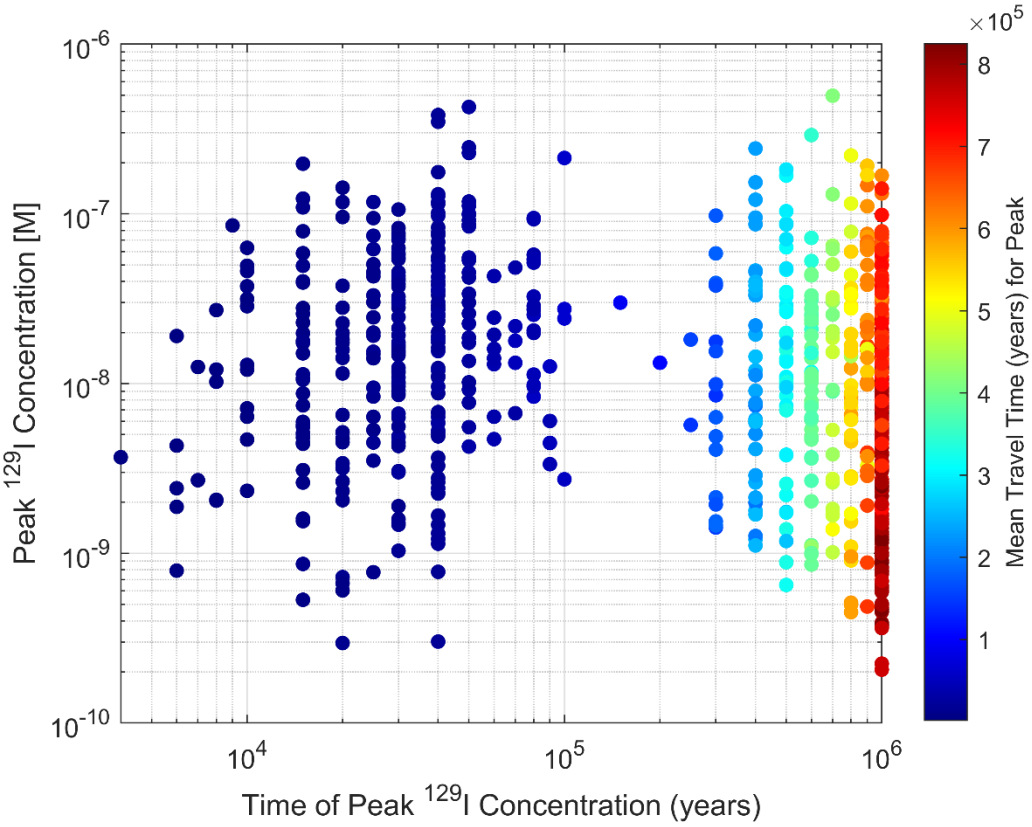


Figure 4-30. Peak ^{129}I concentration vs. Time of Peak, colored by Mean Travel Time of a conservative tracer from the repository to the aquifer location of the peak ^{129}I concentration.

Another example of a time-dependent relationship between parameters is shown in Figure 4-31. This figure shows again the time of peak ^{129}I concentration vs. the peak value, but it is colored by the mean waste package degradation rate. Here, higher values of mean WP rate correspond to early peak ^{129}I times and lower values correspond to later peak ^{129}I times. This correlation is driven by the waste package degradation: the faster the waste package degrades, the earlier it fails, which turns on the radionuclide source term.

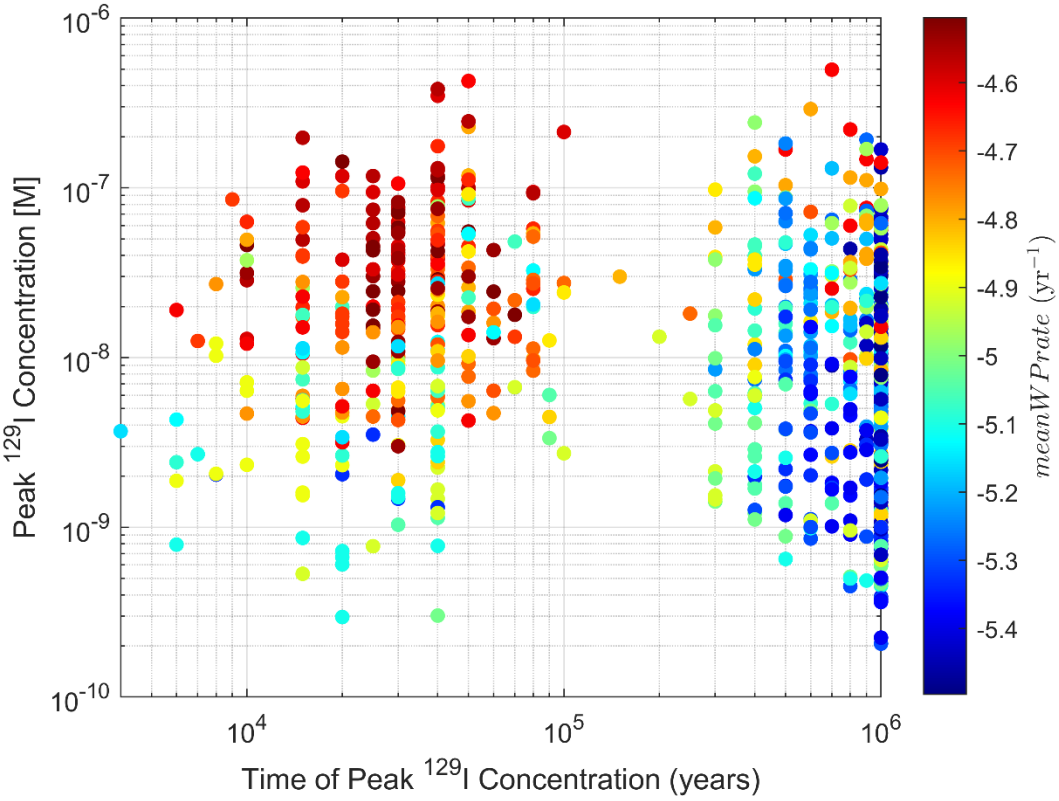


Figure 4-31. Peak ^{129}I concentration vs. Time of Peak, colored by Mean Waste Package Degradation Rate.

4.3.6 Sensitivity Analysis

The next series of plots show the sensitivity analysis of various outputs with respect to the 7 epistemic parameters that were sampled as well as a graph characteristic that was used as a proxy for the DFNs. Note that the inclusion of graph metrics (see Section 4.3.4) in the sensitivity analysis is a new development. It was only attempted with one graph metric (the relative shortest travel time) for this analysis, but this demonstrated its utility; we anticipate being able to use more graph characteristics as proxies for a parametric representation of the DFNs in the future.

The plots below show the Sobol' main and total effect indices. Recall that the Sobol' indices are variance-based indices. The main effect index indicates the fraction of the total response variance that can be attributed to a particular parameter alone. For example, if a parameter has a main effect index of 0.3 with respect to a response, that means that 30% of the variance in the response can be attributed to the effect of that parameter's variance. The total effect index indicates the combined or total effect of a variable and its interaction with other variables on the variance of the response. The main effect indices should sum to one if there are no interaction effects; in the presence of interaction effects, their sum will be less than one. The total effect indices are bounded below by one because they include interactions in addition to the main effects. If the main and total effect indices are the same value for a given parameter, this indicates the parameter does not have significant interaction effects on the response with respect to other parameters.

The calculation of these sensitivity indices is a costly computation, requiring tens of thousands of model evaluations. Because this is not feasible, the 800 sample realizations were used to generate two types of surrogate models that were then used to perform the Sobol' variance-based decomposition. The two surrogates used were polynomial regression (a second-order quadratic regression) and polynomial chaos expansion. Both surrogates were applied without including the graph metric as an input and a second quadratic regression was applied with the relative shortest travel time (RSTT) included. The sensitivity results from the polynomial chaos expansion surrogate are denoted by PCE2 (for polynomial chaos expansion, order 2) in the plots and the results from the quadratic surrogate are denoted by Quad. Thus, the main effect Sobol index estimated using the PCE surrogate is denoted PCE2 Main in the plots. The total effect Sobol index estimated using the PCE surrogate is similarly denoted PCE2 Total, and the indices from the quadratic surrogate are denoted analogously. The quadratic surrogate results are also labeled in this fashion.

These surrogate model methods and the calculation of the Sobol' indices were discussed extensively in last year's GDSA UQ/SA report [14]. Note that this year's dataset posed some challenges to surrogate construction because there were 20 spatial replicates for each parameter sample. That is, there was a 1 input → 20 output values mapping. Many surrogate types perform better with a 1 input → 1 output mapping and this limited the order of the surrogate models that could be used. Higher order polynomials or polynomial chaos expansions as well as Gaussian process models require more unique input samples.

The data from this crystalline reference case poses an additional challenge because the variability over the 20 replicates at each input parameter value is often quite large and dominated by spatial uncertainty. When the RSTT is not included in the surrogate model, the spatial uncertainty cannot be attributed to any of the parameters, so the surrogate model fails to explain much of the variance or may spuriously attribute it to parameter interactions. This can be seen in

some of the sensitivity analysis plots in this section. In the top plot of Figure 4-32, both the quadratic and PCE models suggest interactions between many of the input parameters and no single parameter stands out as substantially more important. When the RSTT is included, however, the bottom plot shows that most of the variance is accounted for by the spatial uncertainty.

The plots below corroborate the trends shown in the scatterplots in Section 4.3.2 and provide a quantitative ranking of parameter importance. Because this ranking is dependent on the quality of the surrogate model, multiple models are provided. Inconsistencies between models highlight the difficulty of fitting accurate surrogate models to this type of data. When the models are consistent with each other, this suggests a relatively strong trend.

Figure 4-32 shows the sensitivity indices for the Peak ^{129}I concentrations when the RSTT is not included to account for spatial uncertainty (top) and when the RSTT is included (bottom). The surrogate models for sensitivity analysis on the peak ^{129}I concentration were fit using the $\log_{10}$ transformed peak concentrations. This technique can reduce the effects of different input/output scales and improved the quality of the surrogate model results for peak ^{129}I concentration. The only parameter with a clearly significant main effect according to both surrogates is *kGlacial*. The mean waste package degradation rate and the permeability parameters all appear to have some significance. Though the relationship with permeability parameters is less clear, the scatterplots in Figure 4-27 support a negative correlation between *kGlacial* and peak ^{129}I concentration and a positive correlation between *meanWPrate* and peak ^{129}I concentration. The total sensitivity indices for the permeability parameters are at least double the value of the main sensitivity indices, indicating that much of the permeability effects are interaction effects, which would explain why the trends are not obvious in two-dimensional scatterplots.

Inclusion of the RSTT (bottom) did not improve the quality of the polynomial surrogate model for Peak ^{129}I concentration. The model attributed very little variance to this graph metric, so the Sobol indices do not suggest a different conclusion regarding variable importance. Results from the quadratic model including RSTT identify the same dominant parameters as the polynomial chaos expansion surrogate without the RSTT. For the remaining figures, the analysis including the RSTT is only included when it improved the quality of the surrogate model.

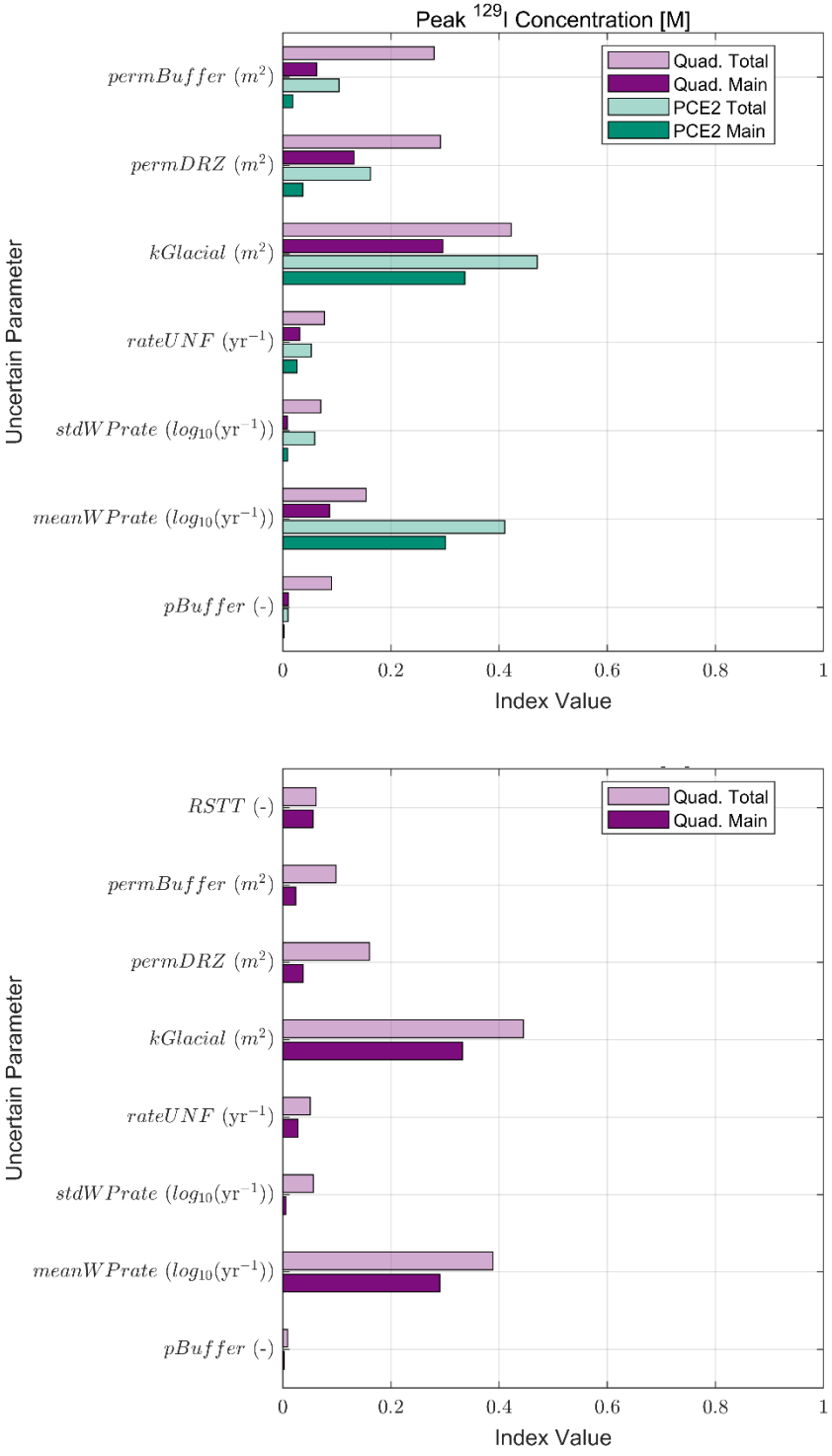


Figure 4-32. Sobol indices for Peak ^{129}I concentration without including the graph metric (top) and with the RSTT (bottom).

Figure 4-33 shows the sensitivity indices for the mean travel time at 1Ma. Again, we see several parameters have some effect, with *pBuffer* ranking the highest followed by the permeability of the buffer, the permeability of the DRZ, and *kGlacial*. As mentioned previously, the *meanWPrate*, *stdWPrate*, and *rateUNF* cannot affect travel time or residence time. We suspect that the small main effect indices are an artifact of using sampling methods to estimate the Sobol' indices and also due to the fact we are using surrogate models to calculate these indices. The surrogate models are constructed over replicates with significant replicate variability: that is, for one parameter setting, there are 20 DFN realizations that yield significantly different outputs. The surrogate estimation is thus challenging in the face of such noise, and we see that noise reflected in part with some spurious main effects indices which should be zero but are small (<0.1). There is also some disagreement between the models regarding the top ranked parameter. The quadratic surrogate attributes more variance to the permeabilities of the buffer and DRZ than to *pBuffer*, whereas the polynomial chaos expansion model clearly attributes most of the variance to *pBuffer*.

We added one of the graph attributes, the RSTT, to the set of inputs and re-ran the calculations to see if the graph attribute (representing DFN variability) would account for the variance in the output. Figure 4-32 shows that adding the graph attribute does not significantly change the ranking of parameter significance with respect to peak ^{129}I concentration. However, it does significantly change the parameter ranking for mean travel time and in fact now is the dominant factor. Figure 4-33 shows that all the variance in peak mean travel time may be attributed to the RSTT, which is a proxy for the DFN. The RSTT also appears to have a significant effect on mean residence time (accounting for 35% of the variance, Figure 4-34) and on the flow rate from the rock to aquifer (accounting for 60% of the variance, Figure 4-35) but not on the east boundary flow rate which is still dominated by *kGlacial* (Figure 4-36). We are currently investigating more graph attributes. The fact that the RSTT shows a significant main and total effect on many of the QoIs is consistent with the DFN effect observed in scatter plots and demonstrates that the importance of the DFN on these results can be captured by a summary statistic.

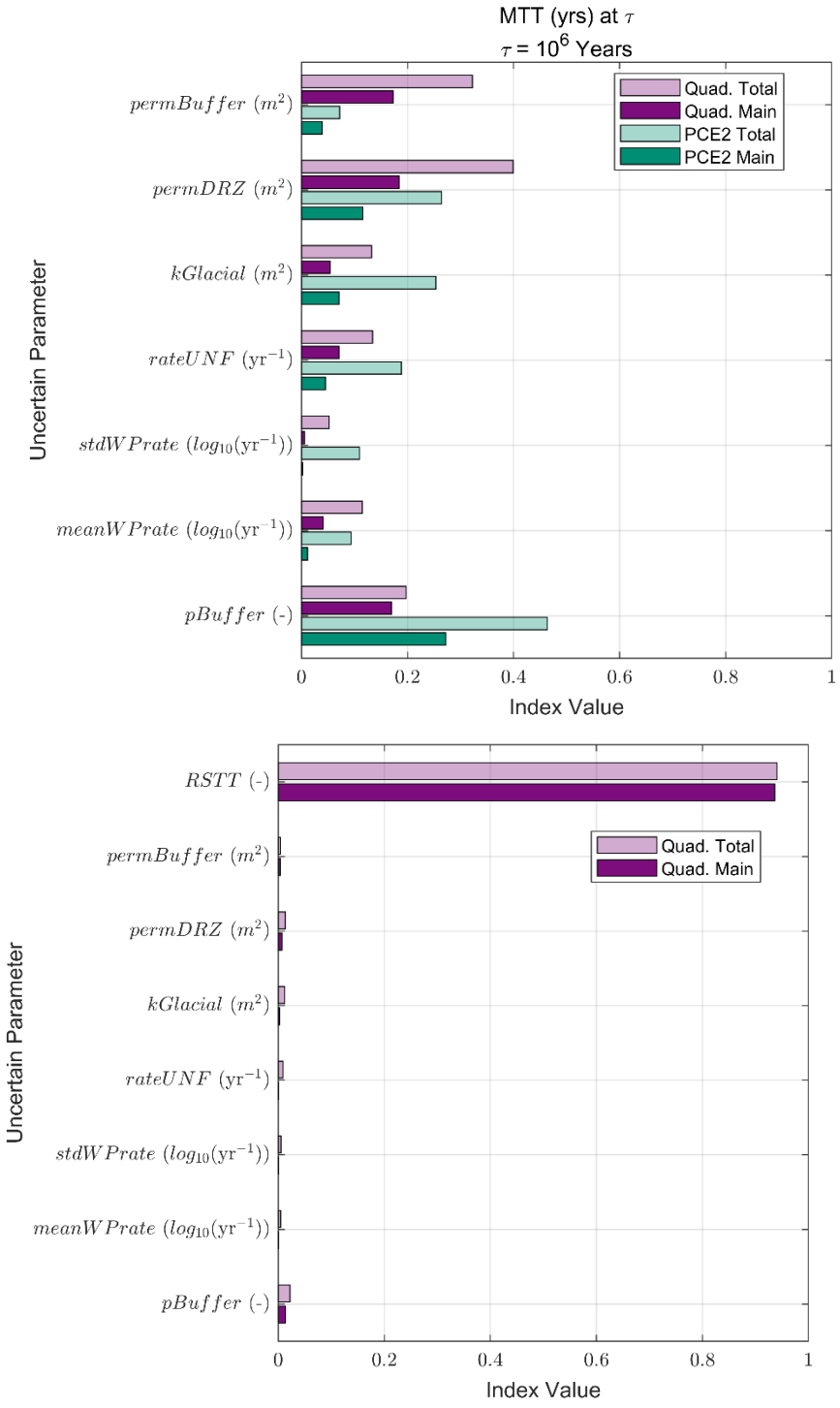
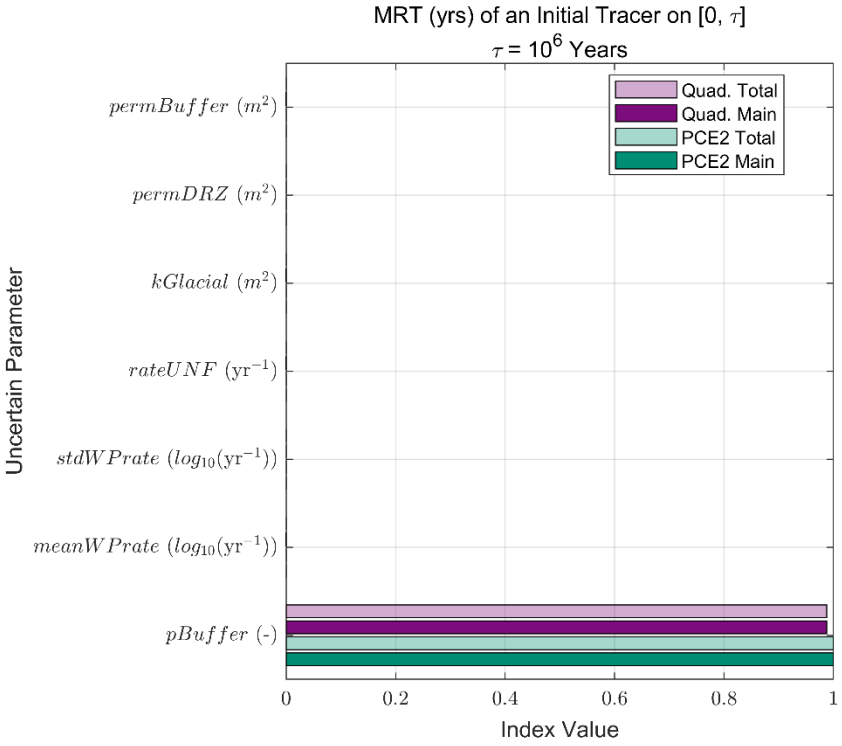


Figure 4-33. Sobol indices for mean travel time of a conservative tracer from the repository to the aquifer at 1Ma without including the RSTT (top) and with the RSTT (bottom).

The peak mean residence time shows a different result: the variance it is totally dominated by *pBuffer* as shown in Figure 4-34, where *pBuffer* accounts for nearly 100% of the variance. However, when the RSTT is added, the effect of the DFN influence is seen. This is consistent with the trends observed in scatterplots; the mean residence time is positively correlated with the RSTT in Figure 4-26 and positively correlated with *pBuffer* in Figure 4-29. Both trends are distinct but the *pBuffer* trend is stronger and more monotonic, which is consistent with its higher index values.



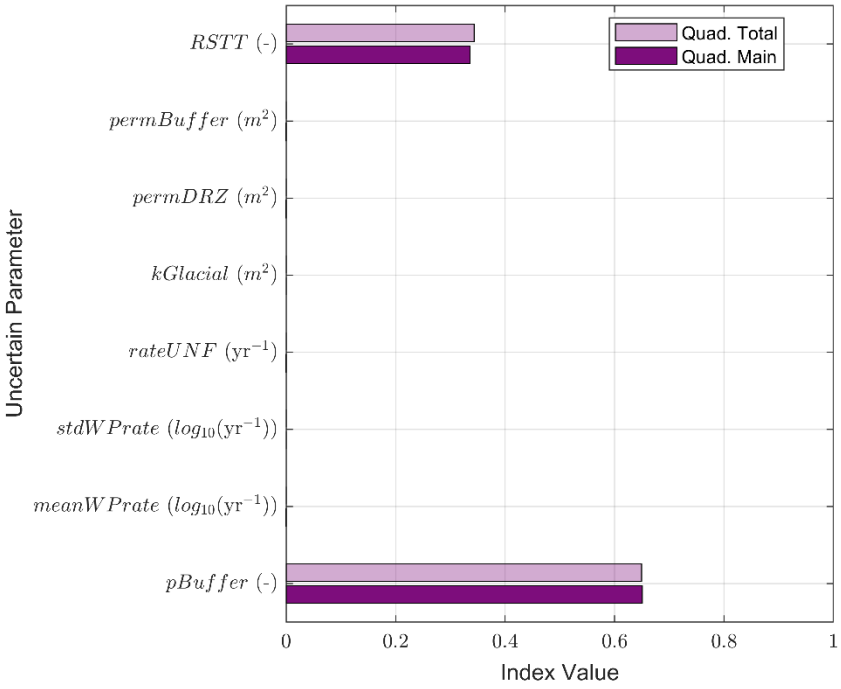


Figure 4-34. Sobol indices for mean residence time of an initial conservative tracer within the repository on [0, 1 Ma] without including the RSTT (top) and with the RSTT (bottom).

For the rock to aquifer flow rate (Figure 4-35), we see *pBuffer* having a strong effect on the peak flow rate from the rock to the aquifer, contributing nearly 50% of the variance with some other parameters having a small main or total effect. Adding the graph metric changes the sensitivity analysis results, with RSTT accounting for 60% of the variance. As with the mean residence time, the sensitivity results including the RSTT are consistent with trends observed in the scatterplots. Figure 4-15 shows a weak positive trend between *pBuffer* and the maximum rock to aquifer flow rate and Figure 4-26 shows a stronger negative trend between the RSTT and the maximum rock to aquifer flow rate.

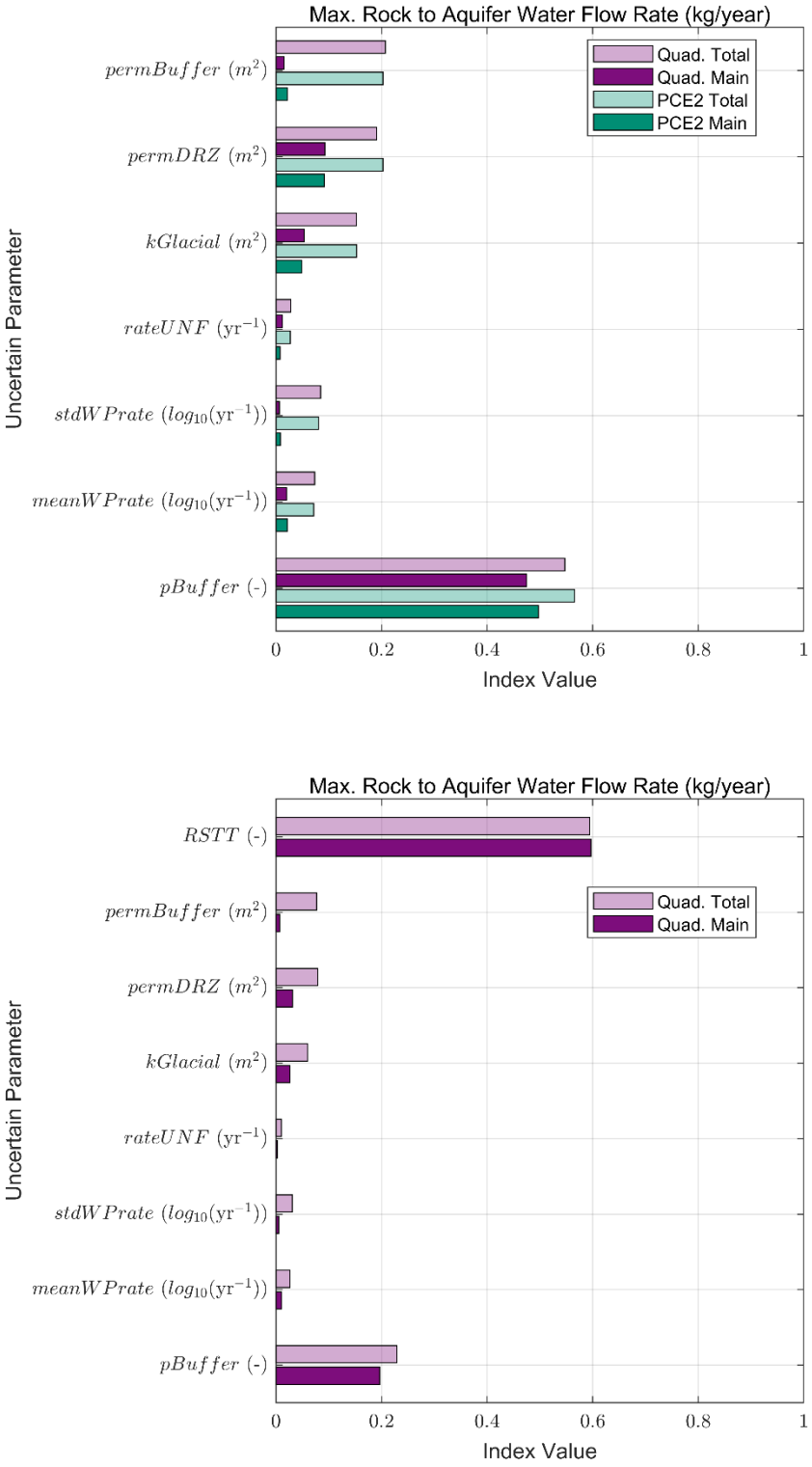


Figure 4-35. Sobol indices for rock to aquifer maximum water flow rate without including the RSTT (top) and with the RSTT (bottom).

Finally, *kGlacial* is the most significant parameter affecting the flow rate from the rock to the east boundary, as shown in Figure 4-36, accounting for all the variance. This result also make sense. The maximum east boundary rate is plotted against the epistemic input parameters in Figure 4-37. Note that there are 40 clusters in the plots, one for each realization of the parameter loop, and 20 points within each cluster, one for each spatial realization. The positive trend with *kGlacial* is clearly the dominant trend. Unlike the other QoIs, there is comparatively very little effect from the spatial uncertainty, so it makes sense that the RSTT did not change the sensitivity index results.

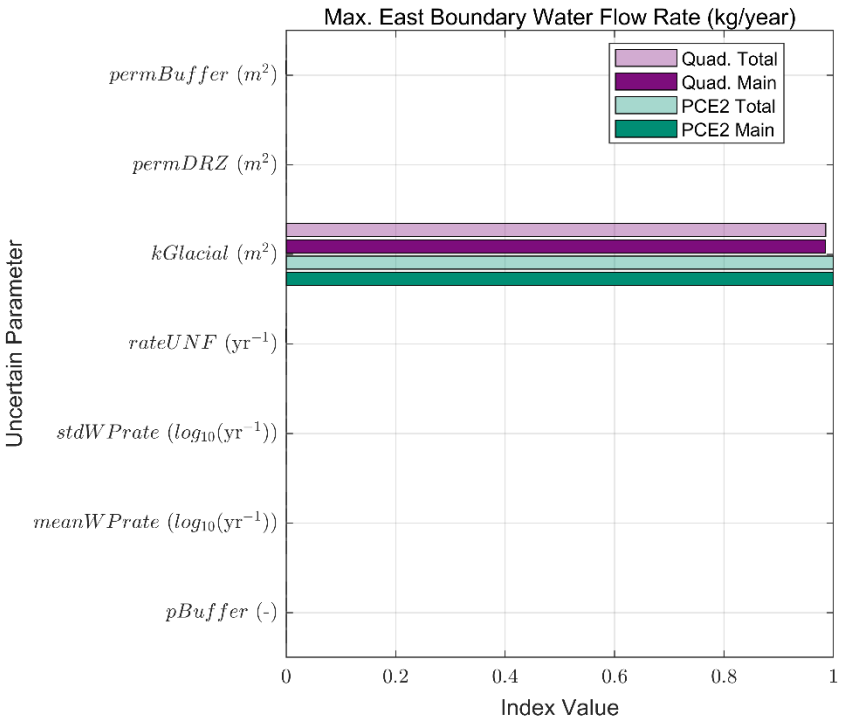


Figure 4-36. Sobol indices for maximum east boundary water flow rate.

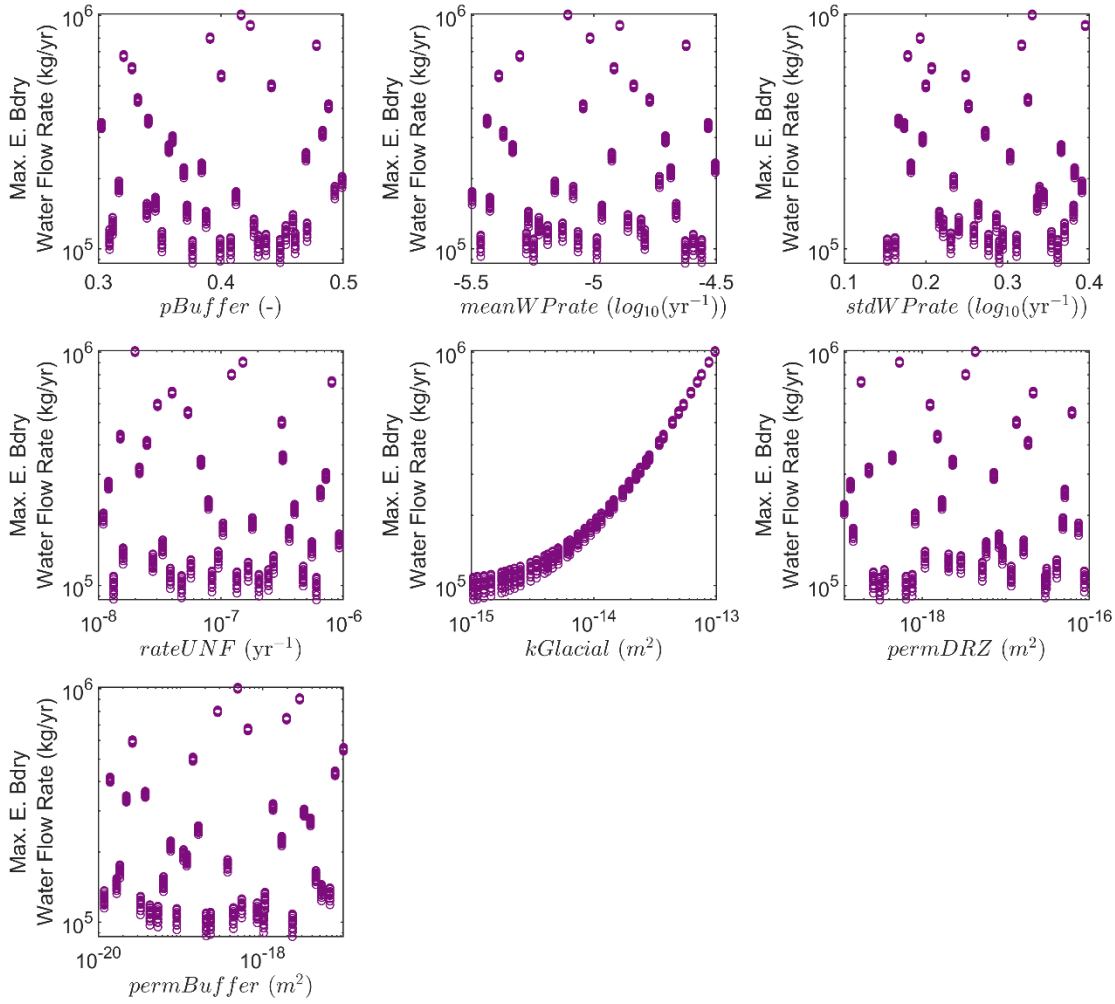


Figure 4-37. Scatterplots of epistemic input parameter values vs. the maximum east boundary water flow rate, showing the strong dependence on *kGlacial*.

4.4 Conclusions for Crystalline Reference Case UA

This analysis was performed for a reference case repository for commercial spent nuclear fuel in fractured crystalline host rock. The uncertainty and sensitivity analyses were designed using lessons learned to build upon the analyses performed on this reference case in [14]. The same epistemic uncertainties were included as in the previous uncertainty analysis, but the sampling structure was changed such that each epistemic parameter sample was repeated for each spatial loop. This sampling strategy was chosen to distinguish the effects of epistemic parameter uncertainties from the effects of spatial heterogeneity) uncertainties on key QoIs. In addition to changing the sampling strategy, we defined new DFN-related bulk characteristic QoIs to add insights into repository performance beyond those gained from analysis of peak ^{129}I concentrations.

The peak ^{129}I concentration QoI was also changed to be the maximum taken over all observation points in the aquifer, rather than analyzing the ^{129}I concentrations only at a few observation points. The work described in Section 4.3 emphasizes the spatial heterogeneity of the problem and metrics designed to tease out such effects.

The effects of spatial uncertainty were emphasized as a result of the new sampling strategy. This interfered with the ability of surrogate models to adequately characterize sensitivity of the QoIs to the epistemic parameter uncertainties, due to the small overall number of unique paramter samples and because of the high variance in the 20 different estimates of the QoI at each of the 40 epistemic samples introduced by spatial uncertainties. We were able to investigate sensitivity to the epistemic parameter uncertainties by including a graph metric (the RSTT) in the surrogate model that acted as a proxy for the DFN. This improved performance of the surrogate for QoIs that are sensitive to the DFN and led to sensitivity results less dominated by (potentially spurious) interaction effects.

In summary, some of the key findings from the sensitivity analyses are:

1) Relative shortest travel time, mean travel time, mean residence time, and even peak flow rate from rock to aquifer all provide similar measures of connectivity/transport in the rock.

2) There is a significant scatter in peak ^{129}I. We did not find one parameter that clearly drives uncertainty in the peak ^{129}I concentration. *kGlacial* and the permeability parameters have some effect, and there are small interaction effects. We had hypothesized that the measures of connectivity would be strongly influential factors, but they were not. In some ways, we are divorcing the fracture connectivity relationship from the iodine concentration by taking the peak ^{129}I over the entire aquifer compared to a maximum ^{129}I at a location, because the location of the peak could be due both to different fracture paths and different waste package breaches. That is, even if we don't see an explicit effect of connectivity, there still could be some interaction between fracture path, waste package breach order, and location of waste package that accounts for most of the remaining variance in the ^{129}I concentration not accounted for by *kGlacial* and the permeability parameters. This would be very hard to tease out.

3) The mean travel time, mean residence time, and maximum flow rate from the rock to aquifer were all heavily influenced by the relative shortest travel time which is a proxy for DFN: these quantities were significantly influenced by DFNs. Additionally, the relationships between buffer porosity and measures of connectivity can be explained by the physics of the problem and may be of secondary importance compared to the dominant influence of the DFN.

4) Repeating the epistemic parameter samples helped us identify how much variability comes from the spatial uncertainties but also limited the utility we got out of the sensitivity analysis.

5) The inclusion of a graph attribute for sensitivity analysis showed promise and helped improve the sensitivity analyses given our small epistemic sample size. In particular, the graph attributes allowed us to quantify the effect of spatial uncertainty in the sensitivity analysis of the crystalline reference case, something that could not be done previously. We will continue investigation of such graph characteristics in future work.

4.5 References: Chapter 4

1. Stein, E. R., J. M. Frederick, G. E. Hammond, K. L. Kuhlman, P. E. Mariner and S. D. Sevougian (2017). "Modeling Coupled Reactive Flow Processes in Fractured Crystalline Rock," *Proceedings of the 16th International High-Level Radioactive Waste Management (IHLRWM 2017) Conference*, Charlotte, North Carolina, April 9-13, American Nuclear Society.
2. Joyce, S., L. Hartley, D. Applegate, J. Hoek and P. Jackson (2014). "Multi-scale groundwater flow modeling during temperate climate conditions for the safety assessment of the proposed high-level nuclear waste repository site at Forsmark, Sweden," *Hydrogeology Journal*, **22**(6):1233-1249
3. Hyman, J. D., S. Karra, N. Makedonska, C. W. Gable, S. L. Painter and H. S. Viswanathan (2015). "dfnWorks: A discrete fracture network framework for modeling subsurface flow and transport," *Computers & Geoscience*, **84**:10-19.
4. Mariner, P.E., E.R. Stein, J.M. Frederick, S.D. Sevougian, G.E. Hammond, and D.G. Fascitelli (2016), *Advances in Geologic Disposal System Modeling and Application to Crystalline Rock*, FCRD-UFD-2016-000440, SAND2016-9610 R, Sandia National Laboratories, Albuquerque, NM.
5. SKB 2006. *Long-term safety for KBS-3 repositories at Forsmark and Laxemar – a first evaluation*. SKB TR-06-09. Svensk Kärnbränslehantering AB, Stockholm, Sweden.
6. Hammond, G.E., P.C. Lichtner, C. Lu and R.T. Mills (2011). "PFLOTRAN: Reactive Flow and Transport Code for Use on Laptops to Leadership-Class Supercomputers." *Groundwater Reactive Transport Models*. F. Zhang, G. T. Yeh and J. Parker. Bentham Science Publishers.
7. Lichtner. P.C. and G.E. Hammond (2012). *Quick Reference Guide: PFLOTRAN 2.0 (LA-CC-09-047) Multiphase-Multicomponent-Multiscale Massively Parallel Reactive Transport Code*. LA-UR-06-7048. December 8, 2012. Los Alamos National Laboratory, Los Alamos, New Mexico.
8. Adams, B.M., Bauman, L.E., Bohnhoff, W.J., Dalbey, K.R., Ebeida, M.S., Eddy, J.P., Eldred, M.S., Hough, P.D., Hu, K.T., Jakeman, J.D., Stephens, J.A., Swiler, L.P., Vigil, D.M., and Wildey, T.M., *Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.0 User's Manual*, SAND2014-4633, July 2014. Other versions released December 2009 (Version 5.0), December 2010 (Version 5.1), November 2011 (Version 5.2), February 2013 (Version 5.3), May 2013 (Version 5.3.1), November 2013 (Version 5.4), November 2014 (Version 6.1), May 2015 (Version 6.2), November 2015 (Version 6.3), May 2016 (Version 6.4), November 2016 (Version 6.5), May 2017 (Version 6.6), November 2017(Version 6.7), May 2018 (Version 6.8), November 2018 (Version 6.9), May 2019 (Version 6.10).
9. Sevougian, S. D., E. R. Stein, G. E. Hammond, P. E. Mariner, J. M. Frederick, and E. Basurto 2018. "Simulating the Effect of Fracture Connectivity on Repository Performance with GDSA Framework – 18589," in *Proceedings of WM2018 Conference*, March 18-22, 2018, Phoenix, AZ, USA, SAND2017-12677C.

10. Mariner, P.E., Stein, E., Frederick, J.M, Sevougian, S.D., and Hammond, G.E. *Advances in Geologic Disposal System Modeling and Shale Reference Cases*. SFWD-SFWST-2017-000044 / SAND2017-10304R. Sandia National Laboratories, Albuquerque, NM (2017).
11. Viswanathan, H. S., Hyman, J. D., Karra, S., O'Malley, D., Srinivasan, S., Hagberg, A., & Srinivasan, G. (2018). "Advancing graph-based algorithms for predicting flow and transport in fractured rock." Water Resources Research, 54(9), 6085-6099.
12. Srinivasan, G., Hyman, J. D., Osthus, D. A., Moore, B. A., O'Malley, D., Karra, S., ... & Viswanathan, H. S. (2018). "Quantifying topological uncertainty in fractured systems using graph theory and machine learning." Scientific reports, 8(1), 1-11.
13. Aric A. Hagberg, Daniel A. Schult and Pieter J. Swart, "Exploring network structure, dynamics, and function using NetworkX", in Proceedings of the 7th Python in Science Conference (SciPy2008), Gäel Varoquaux, Travis Vaught, and Jarrod Millman (Eds), (Pasadena, CA USA), pp. 11–15, Aug 2008.
14. L.P. Swiler, J.C. Helton, E. Basurto, D.M. Brooks P.E. Mariner, L.M. Moore, S. Mohanty, S.D. Sevougian, and E.R. Stein. "Status Report on Uncertainty Quantification and Sensitivity Analysis Tools in the Geologic Disposal Safety Assessment (GDSA) Framework." SAND2019-13835R. Sandia National Laboratories, Albuquerque, NM (2019).
15. Mariner, P. E., M.A. Nole, E. Basurto, T.M. Berg, K.W. Chang, B.J. Debusschere, A.C. Eckert, M.S. Ebeida, M. Gross, G.E. Hammond, J. Harvey, S.H. Jordan, K.L. Kuhlman, T.C. LaForce, R.C. Leone, W.C. McLendon III, M.M. Mills, H.D. Park, F.V. Perry, A. Salazar III, D.T. Seidl, S.D. Sevougian, E.R. Stein and L.P. Swiler. "Advances in GDSA Framework Development and Process Model Integration." Sandia National Laboratories, Albuquerque, NM (2020).
16. Mariner, P. E., E. Basurto, D. M. Brooks, E. R. Stein and L. P. Swiler. "A Study of Epistemic and Spatial Aleatory Uncertainties on Crystalline Repository Performance." SAND2019-13399 PE. Albuquerque, New Mexico, Sandia National Laboratories.
17. LaVenue, A.M. and RamaRao, B.S. "A Modeling approach to address spatial variability within the Culebra Dolomite transmissivity field." SAND92-7306.
18. Rechard, R.P, B. W. Arnold, B. A. Robinson, J. E. Houseworth, "Transport modeling in performance assessments for the Yucca Mountain disposal system for spent nuclear fuel and high-level radioactive waste," *Reliability Engineering & System Safety*, Volume 122, 2014.
19. Helton J.C. 2011. Quantification of Margins and Uncertainties: Conceptual and Computational Basis. *Reliability Engineering and System Safety* 96:976-1013.
20. Helton J.C. and C.J. Sallaberry. 2012. Uncertainty and Sensitivity Analysis: From Regulatory Requirements to Conceptual Structure and Computational Implementation. *IFIP Advances in Information and Communication Technology* 377 AICT:60-76.

21. Hacking I. 2006. The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction and Statistical Inference. 2nd edition. London; New York: Cambridge University Press.

5. UQ WORKFLOW

The GDSA computational framework is largely comprised of two primary computational capabilities: PFLOTRAN and Dakota, as described in Section 1.2. An analysis supporting the performance assessment of a geologic repository necessarily requires the development and use of many additional connective computational components beyond these two primary components, including the development of input files, scripts that connect the capabilities of PFLOTRAN and Dakota, scripts to submit calculations to computational resources, capabilities to gather and post-process results, and much more. The collection of these computational components that comprise the complete body of work required to produce results of interest from a computational simulation capability, such as the GDSA framework, is hereafter referred to as an *analysis workflow*.

The development of GDSA analysis workflows often occurs on an analysis-by-analysis basis and requires the expertise of highly experienced modelers/analysts who are very familiar with the PFLOTRAN code base. Often, these complex analysis workflows involve many manual steps and continuous monitoring of simulations. This can make it difficult to replicate previous analyses, hand-off analyses between analysts, and/or train new analysts to produce analysis workflows as the reproducibility of these analyses depends on how well the modeler/analyst organized and documented what was done.

To improve analysis workflow automation, development, reproducibility, and traceability for repository PA simulations, the GDSA team began developing automated analysis workflows using the Next-Generation Workflow (NGW) capability in FY20. The Next-Generation Workflow (NGW) capability is an open source engine that was developed at Sandia National Laboratories to provide analysts with a capability to construct, execute, and communicate end-to-end computational simulation analysis workflows [1]. This capability is a graphical, node-based interface that includes many pre-programed support functions which are often utilized within computational simulation analysis workflows. NGW is available within the Dakota Graphical User Interface (GUI) [2] and is thus available to the GDSA analysis community.

Development of NGW analysis workflows for the GDSA Framework is focused providing an automated workflow capability with the following objectives:

- Reduce the learning curve for new users to set up and run simulations and analysis workflows
- Speed up analysis workflow execution time
- Eliminate/reduce the need for manual intervention and allow for automated monitoring
- Reduce the potential for the introduction of human errors
- Increase traceability and reproducibility

A series of NGW analysis workflows for GDSA analysis exemplars, including a short course exercise and the Crystalline Reference Case, were developed in FY20 to pilot the use of this capability for the GDSA program [3]. Section 3.2.1 of the GDSA Framework report [3] details the

development of these workflows, provides examples of their use, and proposes a path forward for additional development. Appendix C of [3] provides a complete graphical workflow, demonstrating an uncertainty quantification study. This workflow sets parameters of interest, inserts these parameters into the PFLOTRAN input file, collects additional input files, runs PFLOTRAN, and develops and runs a post-processing script. The eventual goal of this work is to develop an analysis workflow library that can be made available to the GDSA analysis community.

5.1 References: Chapter 5

1. Orient, G., Clay, R., Friedman-Hill, E., & Hoffman, E., 2020. Next Generation Workflow – an Open Source Software System for Computational Modeling to Support Agile VVUQ. Presented at the ASME V&V Symposium, May 20-22, 2020.

2. Ridgway, E., 2020. Dakota. GUI Version 6.12 User Manual – Next-Gen Workflow. https://dakota.sandia.gov/content/next-gen-workflow. Accessed August 5, 2020.

3. Mariner, P. E., M.A. Nole, E. Basurto, T.M. Berg, K.W. Chang, B.J. Debusschere, A.C. Eckert, M.S. Ebeida, M. Gross, G.E. Hammond, J. Harvey, S.H. Jordan, K.L. Kuhlman, T.C. LaForce, R.C. Leone, W.C. McLendon III, M.M. Mills, H.D. Park, F.V. Perry, A. Salazar III, D.T. Seidl, S.D. Sevougian, E.R. Stein and L.P. Swiler. "Advances in GDSA Framework Development and Process Model Integration." Sandia National Laboratories, Albuquerque, NM (2020).

6. INTERNATIONAL COLLABORATIONS

6.1 Sensitivity Analysis of Reference Cases

This year, two cases studies were performed as part of an international collaboration on sensitivity analysis. This activity is part of a working group focused on sensitivity analysis methods that falls under the umbrella of the Integration Group for the Safety Case (IGSC) Symposium on the Safety Case of of OECD/NEA [1]. Participating organizations or countries include GRS (Germany), Clausthal University of Technology (TUC) (Germany), Sandia National Laboratories (SNL) (USA), Posiva (Finland), SCK-CEN (Belgium), and the Nuclear Safety Institute of the Russian Academy of Sciences (Russia). The cases are briefly described below. For a more complete description, see [2]. Note that a variety of sensitivity analysis approaches were used by the participating organizations, including simple and partial correlation coefficients on the raw data and on rank-transformed data, regression approaches, CUSUNORO curves, and scatterplots. Below we present only the variance-based results because this was of particular interest to this working group.

6.1.1 GRS: LILW repository in salt

The case study provided by GRS involved a Low and Intermediate Level Waste (LILW) in a salt repository [3]. There were four separate data sets provided by GRS involving 6 uncertain inputs or 11 uncertain inputs, and 4096 samples or 8192 samples. The analyses performed involved creating surrogate models (linear regression, quadratic regression, multivariate adaptive regression splines (MARS), and polynomial chaos expansion (PCE)). The surrogates were then used to estimate the Sobol' variance-decomposition sensitivity indices. Correlations and scatterplots were also used.

Figure 6-1 shows the main effects indices as a function of time, based on different surrogate models for the 11 variable case and 8192 samples. These show similar patterns of variable importance over time. *IniPermSeal*, *AEBConv*, and *GasEntryP* are the most important parameters in terms of significant main effects.

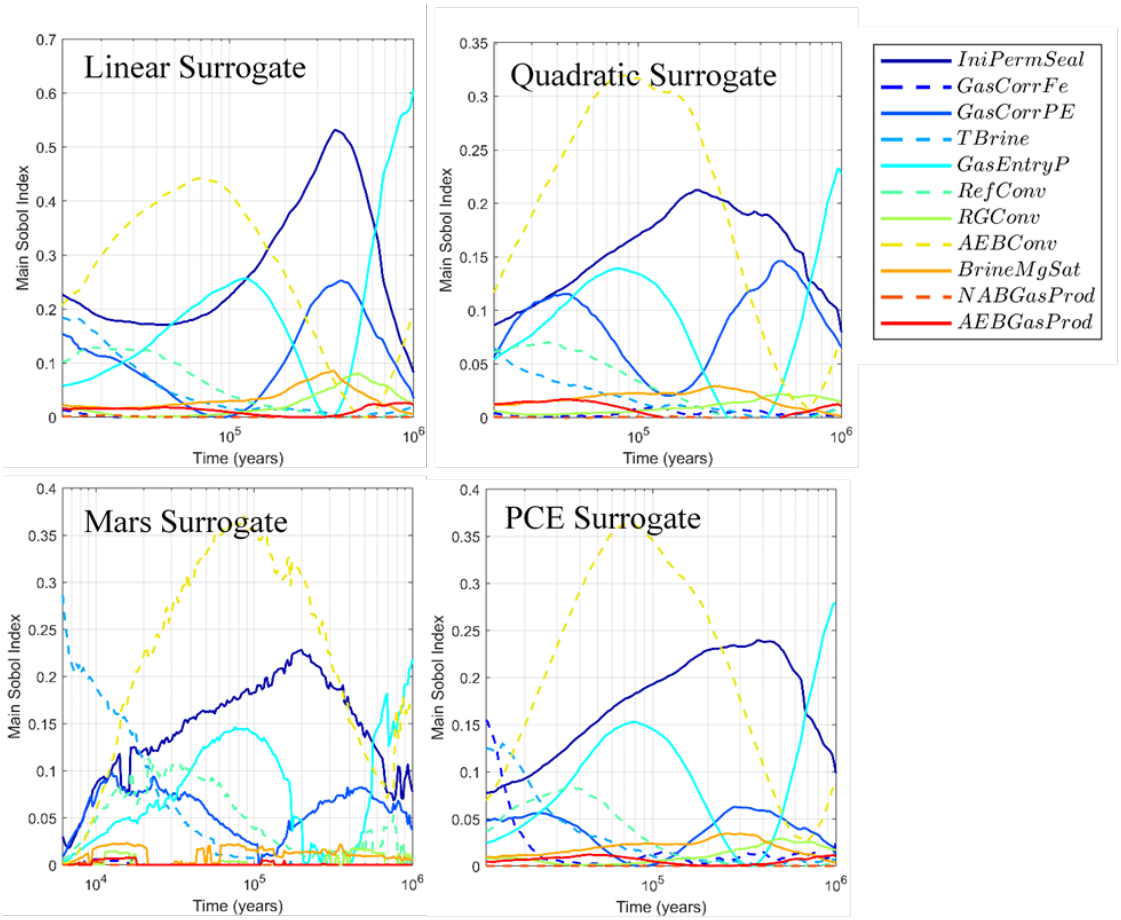


Figure 6-1. Main Sobol' index calculated for the peak ^{129}I concentration [M] for the GRS Salt case.

Figure 6-2 shows the main effects with the total effects obtained using the PCE model. The total effects indices account for the variable itself as well as interactions with other variables. The higher total sensitivity indices in Figure 6-2 suggest significant interaction effects (note scale differences compared with Figure 6-1). Including second-order interactions, *RefConv* gains some importance but the rankings of the top three parameters are similar. *GasEntryP* appears to have more significant interaction effects than main effects later in time. This sharp decrease in the main index is due to a threshold effect [3].

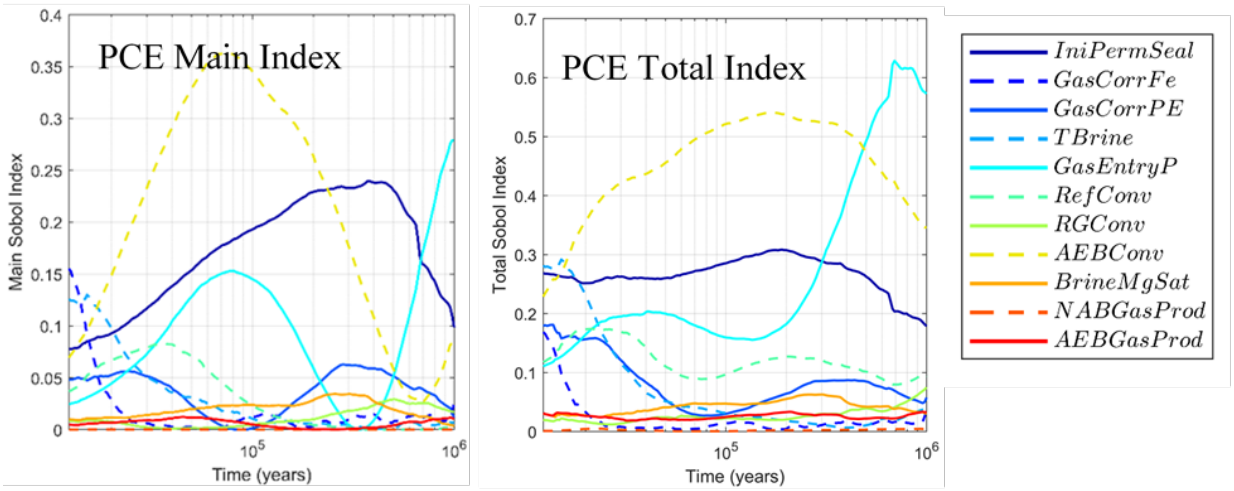


Figure 6-2. Main and total Sobol' indices calculated for the peak ^{129}I concentration [M] for the GRS Salt case.

To summarize the sensitivity analysis results for the GRS salt case: most of the surrogate models performed well and gave similar results, though Mars did not perform as well with small sample sizes. Some Sobol' index estimates are not precise (e.g. values greater than 1 or less than 0), which should be resolved by increasing the sample size applied to the surrogate. Nonetheless, trends and rankings are consistent.

Variance-based methods with surrogates that can detect non-linear and non-monotonic relationships show that there are interactions that may not be captured using only linear methods. *AEBConv* importance remains high over time when measured with the total index, even though the main index decreases. This is consistent with the GRS analysis which showed that linear methods did not adequately detect *AEBConv* importance for the full (relevant) duration of the simulation.

6.1.2 Ibrae

The Ibrae case study was provided by the Nuclear Safety Institute of Russia. It is a case involving a cross-sectional groundwater flow model in heterogeneous geological media. There are 12 inputs: seven representing hydraulic conductivities of the different structural elements in the media and five boundary conditions. There are 37 outputs, representing 37 values of hydraulic heads at different observation points in the domain.

We used the same approach as for the GRS salt case: surrogate model construction and Sobol' index calculation (main and total effects). The *BC2* input was identified as the most influential parameter by far, regardless of which surrogate was used. There was also essentially no difference between main and total indices, indicating the Ibrae model does not include variable interaction. The Sobol' indices estimates were repeated for multiple sample sizes (number of simulations). Results were essentially the same for 140 samples as for 28,000. These results were consistent with the sensitivity analyses performed by Ibrae on this case.

6.2 Survey of Sampling-Based Sensitivity Analysis Methods in the International Community

Sitakanta Mohanty, in collaboration with Laura Swiler and Emily Stein, has written a report titled "Survey of Sampling-Based Sensitivity Analysis Methods Used in Geologic Disposal Safety Assessments." (SAND2020-XXXX) [4]. This report presents a review of UQ/SA methods used by selected non-US countries (seven countries) and international consortia engaged in repository PAs over the past two decades. The report describes sampling-base sensitivity analysis methods formally used in performance assessment calculations to meet regulatory requirements and describes methods used for supplemental, experimental, or developmental purposes. It lays out (i) the current practices by various countries, (ii) international collaborations, and (iii) some recommendations on future developments.

For each country, a brief description is provided of the PA model, which is the UQ/SA performance function model. The review also suggest that future efforts should focus on handling parameter interactions in the PA model with a large number of sampled parameters (>>10), elegantly handling correlated parameters, dealing with discontinuity in the model output, developing a protocol for selecting the appropriate UQ/SA methods based on input-output relationships, and continued development of UQ/SA toolboxes with a full suite of traditional and advanced methods.

6.3 References: Chapter 6

1. Röhlig, K-J., E. Plischke, D-A. Becker, E.R. Stein, J. Govaerts, B.J. Debusschere, L. Koskinen, P. Kupiainen, C.D. Leigh, P. Mariner, O. Nummi, B. Pastina, D. Sevougian, S. Spießl, L.P. Swiler, E. Weetjens, and T. Zeitler. "Sensitivity Analysis in Performance Assessment: Towards a Joint Approach," in the IGSC Symposium 2018, held in Rotterdam, The Netherlands, 10-11 October 2018. Nuclear Energy Agency (NEA), Issy-les-Moulineaux, France.

2. Brooks, D. "SNL Analyses of the GRS Salt and IBRAE Reference Cases." SAND2020-6398PE.

3. Spiessl, S.M., S. Kucherenko, and D.A. Becker. "Application of the RS-HDMR Meta-Modeling Approach in the SobolGSA Software to LILW Models." Presented at the International SA-UQ Workshop 2018, Brussels, Belgium, October 4-5, 2018.

4. S. Mohanty, E. R. Stein, and L.P. Swiler. "Survey of Sampling-Based Sensitivity Analysis Methods Used in Geologic Disposal Safety Assessments." SAND2020-XXXX.

7. CONCLUSIONS

This report covers the main topics addressed in the GDSA UQ/SA work performed in 2020.

The report provides an overview of an exciting new uncertainty quantification method involving models at multiple levels of fidelity. We presented results demonstrating the use of these multifidelity uncertainty quantification approaches to calculate a mean estimate of a quantity of interest more cheaply and with lower variance than an equivalent number of high-fidelity simulations. We also demonstrated the use of multifidelity methods in calculating Sobol' sensitivity indices.

This year, we performed several verification exercises on the discrete fracture networks because the FY19 results demonstrated that DFNs were a main source of uncertainty for the crystalline reference case. We examined the DFNs closely to ensure consistency of the number of fractures and fracture area per unit volume with respect to estimates from the SKB Forsmark site. We also upscaled the DFNs to obtain equivalent continuous porous medium (ECPM) results in terms of the hydraulic conductivity tensor values. We compared the ECPM obtained from upscaling the DFNs with those reported by SKB and found good agreement.

This year, we extended the sensitivity analysis of the crystalline reference case performed in FY19, focusing on indicators of fracture connectivity as well as information about the flow between the repository and the aquifer. To facilitate this, PFLOTRAN added several new quantities of interest, including maximum concentrations (of ^{129}I and tracers) in the aquifer at each time step, along with the location of the peak concentrations, mean travel time from the repository to various locations, residence time within the repository, and total water flow rate in various directions. The results of the sensitivity analysis of the crystalline reference case based on a new set of 800 PFLOTRAN runs (20 realizations of spatial uncertainty x 40 epistemic samples) showed that these new quantities of interest gave us several new insights. The flow rates, mean travel time, and mean residence time were somewhat correlated. There was not one parameter that clearly drives the uncertainty in peak ^{129}I concentration. The most important parameters for several of the QoIs were *kGlacial* and *pBuffer*.

A new capability we developed this year was to calculate graph attributes from the DFNs, such as number of nodes and edges, average degree and relative shortest travel time. When we included a graph attribute in the sensitivity analysis, it had a significant effect on responses such as mean travel time, mean residence time, and flow rates, suggesting that the variability in DFNs drives the uncertainty in these responses. The inclusion of these graph attributes in sensitivity analysis shows promise and helped improve sensitivity results with our small epistemic sample size.

Another activity started this year was the *GDSA Workflow*. This workflow couples Dakota, PFLOTRAN, and SAW (the Sandia Analysis Workflow software) to present the user with a unified interface where the actual workflow can be dictated in an easy-to-use graphical format. This workflow also allows greater reproducibility and traceability of the actual files and scripts used for a particular study. This is increasingly important as we scale up analyses and extract more quantities of interest that must be tracked both as function of time and location (e.g. peak ^{129}I concentration). We demonstrated the UQ workflow for a GDSA study involving two sampling loops over spatial and epistemic uncertainties.

Finally, we continue to engage with the international community on sensitivity analysis topics. This year, we presented results for two international case studies as part of an international working group focused on UQ/SA for repository analysis.

There is a rich legacy of UQ/SA being performed in repository assessment. Development, implementation, and demonstration of new tools and methods for uncertainty and sensitivity analysis in *GDSA Framework* will maintain leadership of the repository science community in UQ/SA methods, while also maintaining an infrastructure of proven tools. Geologic repository performance assessment in the U.S. involves coupled, multi-physics modeling at high resolution, large parameter spaces, and greater use of random (stochastic) field modeling. It requires high-performance computing and costly sample evaluations. UQ/SA methods discussed in this report, including surrogate modeling to reduce computational expense, variance-based sensitivity analysis to quantify importance of parameter interactions in a multi-physics system, and new multi-fidelity methods will enable analysis methods to keep pace with the increasing sophistication of the physics models. *GDSA Framework* development seeks to keep abreast of improvements to existing UQ/SA methods, employ new methods, and maintain an infrastructure of proven tools that can be extended to support computationally expensive analyses.

There are several possible avenues of future UQ/SA development, including:

- Time-dependent sensitivity analysis methods
- Dimension reduction
- Importance sampling
- Surrogate models
- Bayesian calibration
- Estimating/generating distributions/ranges for uncertain inputs
- Multifidelity UQ approaches
- Model form uncertainty
- Density-based SA methods

Future work will involve further development of the items above as they align with the *GDSA Framework* and the repository performance assessment needed.