Uncertainty and Sensitivity Analysis Methods and Applications in the GDSA Framework (FY2022)

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, R. Leone, P.E. Mariner, T. Portone, M. L. Smith

Sandia National Laboratories

August 19, 2022

M3SF- 22SN01030408
SAND2022-XXXXX

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 *Uncertainty and Sensitivity Analysis Methods and Applications in the GDSA Framework (FY2022)*

Deliverable/Milestone/Revision No. M3SF-22SN010304082

Work Package Title and Number GDSA-Uncertainty and Sensitivity Analysis Methods-SNL SF-22SN01030408

Work Package WBS Number 1.08.01.03.04

Responsible Work Package Manager Laura Swiler (Name/Signature)

Date Submitted August 17, 2022

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

Laura Price — Laura Price

Nathan Porter — Nathan Porter

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 gratefully acknowledge the life and legacy of Jon Helton (deceased) for his foundational contributions in performance assessment methodology. The authors thankfully acknowledge Michael Nole for his contributions to PFLOTRAN and supporting 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 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-22SN01030408) level 3 milestone, *Uncertainty and Sensitivity Analysis Methods and Applications in GDSA Framework (FY2022)* (M3SF-22SN010304082). 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 FY22, 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-22SN01030409), the GDSA Repository Systems Analysis work package (SF-22SN01030410), and the GDSA PFLOTRAN Development work package (SF-22SN01030411). This report builds on developments reported in previous *GDSA Framework* milestones, particularly M3SF-21SN010304042.

Contents

Table of Figures

Table of Tables

ACRONYMS and DEFINITIONS

Abbreviation	Definition
ANN	Artificial Neural Network
CDF	cumulative distribution function
CRC	crystalline reference case
DOE	U.S. Department of Energy
DFN	discrete fracture network
DRZ	disturbed rock zone
ECPM	equivalent continuous porous medium
FDR	Fractional Dissolution Rate
FMD	Fuel Matrix Degradation
GDSA	Geologic Disposal Safety Assessment
GUI	Graphical User Interface
HLW	High-level waste
LHS	Latin Hypercube sampling
Ma	mega annum (one million years)
MF	multifidelity
MFMC	Multifidelity Monte-Carlo: A particular MF UQ technique
NGW	Next Generation Workflow
PA	Performance Assessment
PCE	Polynomial Chaos Expansion
QoI	Quantity of Interest
R&D	Research and development
SA	sensitivity analysis
SFWST	Spent Fuel Waste Science and Technology
SNL	Sandia National Laboratories
UA	uncertainty analysis
UFD	Used fuel disposition
UQ	uncertainty quantification
WP	waste package

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-22SN01030408) level 3 milestone, *Uncertainty and Sensitivity Analysis Methods and Applications in GDSA Framework* (FY2022) (M3SF-22SN010304082). 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 FY22, 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-22SN01030409), the GDSA Repository Systems Analysis work package (SF-22SN01030410), and the GDSA PFLOTRAN Development work package (SF-22SN01030411). This report builds on developments reported in previous GDSA Framework milestones, particularly M3SF-21SN010304042 [1], M3SF-20SN010304032 [2], and M3SF-19SN010304032 [3].

Geologic repository performance assessment in the U.S. involves a code base that includes coupled, multiphysics modeling at high resolution. Due to the high cost of these models which require high performance computers to run, relatively few simulation samples are available for analysis. This 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 many evaluations of the predictive model. Much research has focused on accurately calculating variance-based sensitivity indices while keeping the computational cost reasonable. We note that other sensitivity analysis methods [3] 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 [1, 2]. 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. The repository community must maintain awareness of and leadership in UQ/SA methods to best inform our assessment of costly computational models.

1.1 Overview of this Report

This report provides documentation of the UQ/SA work performed in FY 2022. The outline of this report is as follows:

- Chapter 2 provides results using a 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 lower fidelity evaluations. The final result is an estimator of the mean or statistic of the response (e.g. percentile) with a lower variance: a more accurate and reliable estimator can be obtained at lower computational cost. In 2020, we demonstrated the use of multifidelity UQ on a simplified 1-D test problem [2]. This year, we provide a more realistic example using a simplified crystalline reference case. We demonstrate various options for estimation of a mean quantity, along with the relative costs and variance.
- Chapter 3 focuses on identifying biases that may arise in our sensitivity analysis results (specifically, the Sobol' variance-based indices) due to the use of surrogate models in the sensitivity index calculations. This chapter shows that much of the variance of certain quantities of interest (QoIs) is due to the spatial heterogeneity and not due to variance of epistemic input parameters. However, it is quite difficult to capture that spatial heterogeneity (coming from various discrete fracture network realizations) using graph metrics or other proxies. This chapter dives deeply into the surrogate accuracy and resulting influence on the sensitivity calculations.
- Chapter 4 discusses the discrete fracture networks (DFN) that were generated this year. The focus of this chapter is a comparison of the correlated-constant vs. correlated depth-dependent transmissivity relationship for the crystalline reference case DFNs. One hundred DFNs were generated for each transmissivity relationship and the resulting QoIs are compared. These results are documented in Chapter 4.
- In Chapter 5, we extend the sensitivity analysis of the crystalline reference case performed in FY21. We investigated the effects of DFNs (spatial heterogeneity) as well as epistemic parameters on an updated version of the crystalline reference case. Many quantities of interest are extracted from the PFLOTRAN results, including peak concentrations of ^{129}I in the aquifer at each time step along with the location of the peaks, fraction of tracers remaining at various time points, mean travel time from the repository to various locations, median residence time of a tracer within the repository, and total water fluxes in various directions.

  This year, we added another type of uncertainty to the crystalline reference case analysis: model form uncertainty. We consider two alternatives to the Fuel Matrix Degradation (FMD) model. One is a fractional dissolution rate (FDR) model and the other involves an Artificial Neural Network (ANN) surrogate approximating the FMD model. An entire set of 1000 PFLOTRAN simulations (25 DFNs x 40 epistemic parameters) were run using the FDR model and then repeated for the ANN FMD model. The results and comparisons are discussed in Chapter 5.
- Chapter 6 presents an overview of an initial sensitivity analysis study using the DECOVALEX crystalline case. These results are very preliminary but are provided to show a different case study as well as the generalizability of the *GDSA Workflow* and sensitivity analysis framework.
- Chapter 7 provides a summary.
- Chapter 8 lists the references.

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, 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, 2, 3, 4, 5, 6, 7]).

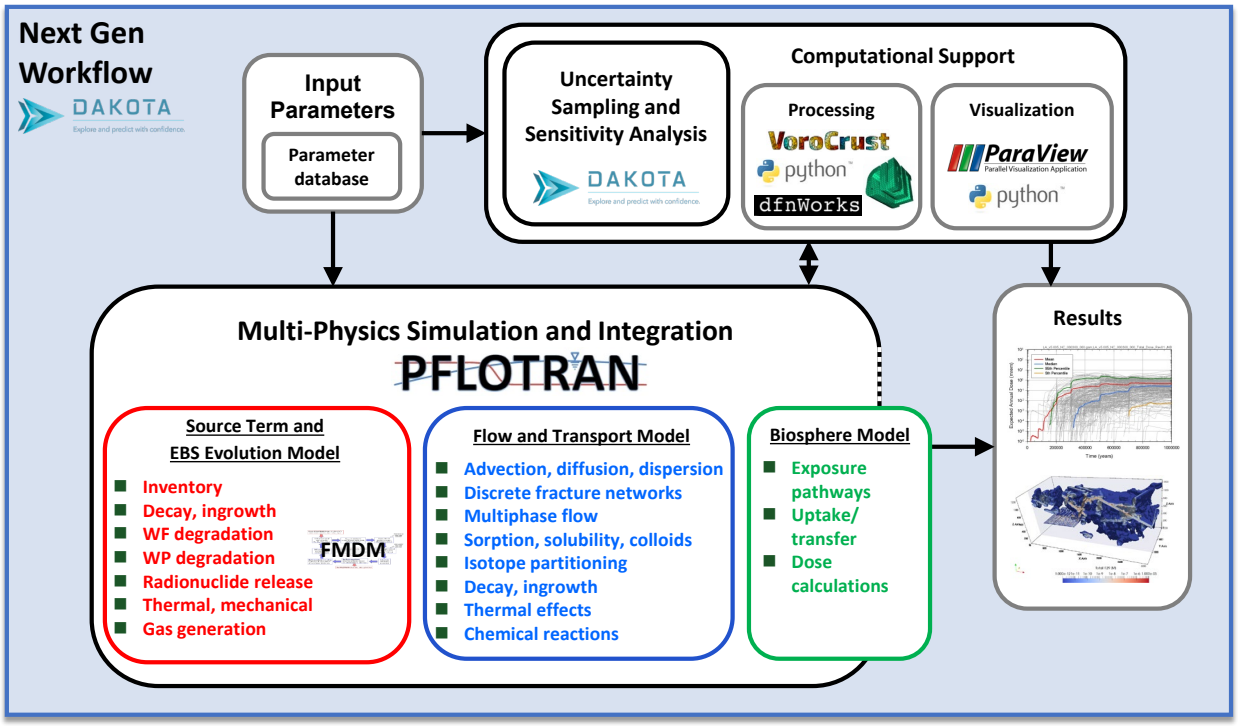


Figure 1-1. The GDSA Framework

One objective of the UQ/SA capability in GDSA Framework is to make 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 [8], DOE 2014 [9],

RESS2000 [10], RESS2014 [11]) available. Another objective is to enable future adoption of new methods consistent with the current standard of practice in the UQ/SA community which are 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.

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 ([12, 13, 14]) written in object-oriented Fortran 2003. PFLOTRAN models subsurface flow using a porous medium continuum approach, which includes capabilities for multicomponent systems, multiphase flow and transport, heat conduction and convection, biogeochemical reactions, geomechanics, and radionuclide decay and ingrowth. The code is developed under a GNU Lesser General Public License, which allows third parties to interface proprietary software with the code. The availability and continuing development of PFLOTRAN for GDSA 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. ([4, 5, 6]) and Sevougian et al. 2018 [15]. 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 [16]. The Dakota software has parametric analysis methods that enable design exploration, model calibration, optimization, uncertainty quantification, and sensitivity analysis with computational models. Dakota is a C++ code which has been under development at Sandia since 1994. It has been primarily sponsored by DOE's Advanced Simulation and Computing (ASC) program. Dakota supports computationally expensive simulations which require high performance computing and parallel execution. 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.

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 allows nested studies to perform an "outer loop" epistemic sampling and an "inner loop" aleatory sampling to generate ensembles of distributions. Dakota includes additional capabilities, such as the use of surrogate models, adaptive sampling approaches, and multifidelity UQ methods. 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.

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.

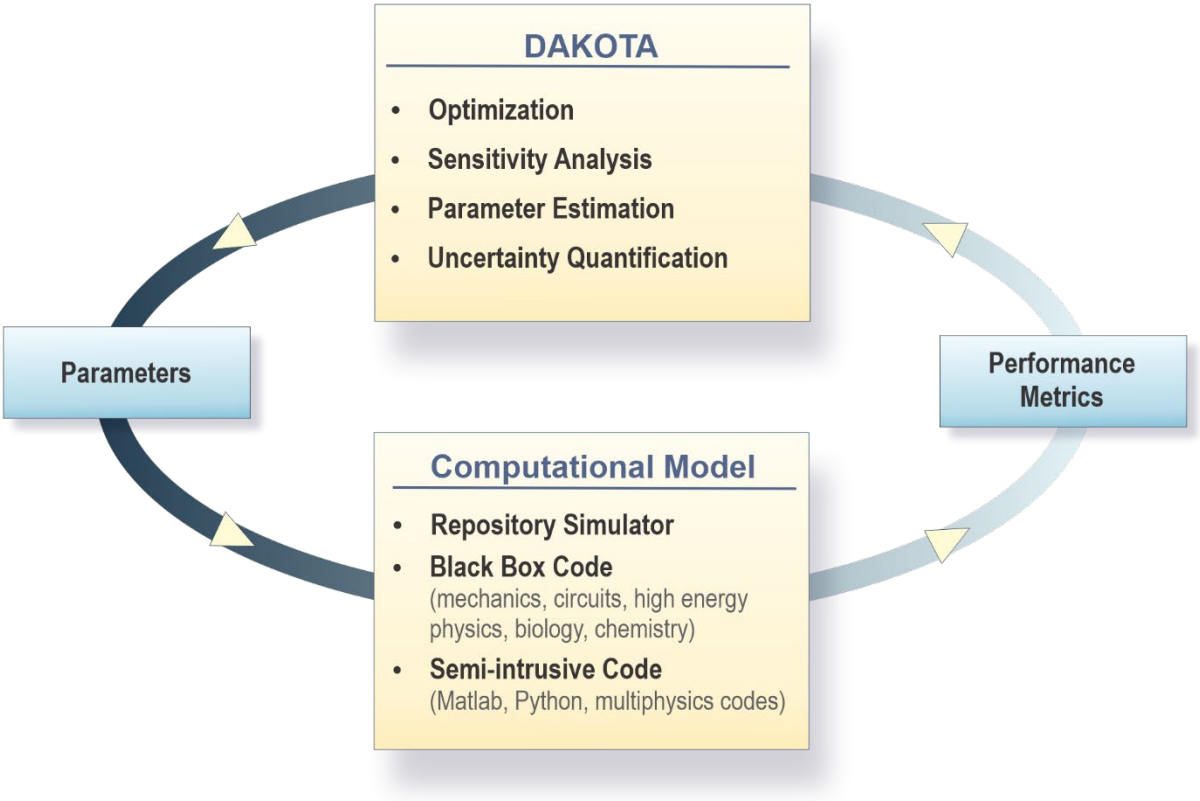


Figure 1-2. Dakota interfacing to a computational model such as a repository simulator

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. These indices estimate the proportion of variance in a quantity of interest that can be attributed to variance in each uncertain input parameter. 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, by the use of polynomial chaos expansions, and by using multifidelity methods. 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 Automated Analysis Workflow Development for GDSA

Uncertainty Quantification analysis 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 number of quantities of interest increases 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 *Crystalline Reference Case Uncertainty Analysis (UA) Nested Workflow* was developed. This workflow couples Dakota, PFLOTRAN, and NGW (the Next-Generation Workflow software) to present the user with a unified graphical user interface (GUI) where the actual workflow can be dictated and automated 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.

The *Crystalline Reference Case UA Nested Workflow* that was developed in FY20 and expanded and used in FY21 was leveraged for the studies presented in this report. Details regarding the development and use of the workflow for past analyses are given in [1, 2, 32]. The portability and automation provided by the workflow significantly speed up the overall analysis. A journal article [17] is currently in preparation to summarize the development of the *Crystalline Reference Case UA Nested Workflow* and the successful adaptation and application of this workflow through time to performance assessment problems. This journal article will describe the use of the Next-Generation Workflow (NGW) software within the Dakota GUI to support performance assessment analyses with complex treatments of UQ/SA.

Modifying the workflow from the crystalline reference case (presented in Chapter 5) to the DECOVALEX case (presented in Chapter 6) required editing of various input files defining the parameters and associated input files to PFLOTRAN. It also required updating run directories and job submission scripts, extracting the outputs for DECOVALEX, and including these outputs in the workflow. We are developing guidelines about what changes are required to use the workflow with a new case study. Having both case studies (the crystalline reference case and the DECOVALEX case) was helpful to generalize the workflow.

1.2.4 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, dfnWorks, and visualization tools such as Paraview. 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 and other scripting tools. There will be tighter integration as *GDSA Framework* and the GDSA workflow progresses.

2. UQ METHODOLOGY: MULTIFIDELITY METHODS

Surrogate-based and sample-based multifidelity methods have the potential to increase the fidelity and efficiency of probabilistic post-closure performance assessment of geologic repositories for nuclear waste. These methods reduce computational cost while still producing high quality estimates of repository performance. With single-fidelity modeling, reducing computation cost typically results in lower quality estimates. Multifidelity methods avoid most or all of the reduction in quality by fusing estimates from cheaper, lower-fidelity models with more computationally costly, higher-fidelity models. Intuitively, these methods use the lower-fidelity, cheap models to increase the precision of statistical estimates (e.g. mean value of a quantity of interest), while the high-fidelity model is used to increase the accuracy of statistical estimates.

Last year, the potential utility of surrogate-based multifidelity methods was explored in the context of a simplified model of a repository in sparsely fractured crystalline rock. The simplified model exhibits many of the features of the Crystalline Reference Case (CRC) but is much less costly to evaluate (~3 minutes on 36 cores vs ~30 minutes on 512 cores, ~1/100th the cost). Full details of this investigation can be found in [1]. Surrogate-based methods are attractive for uncertainty analyses where statistics other than moments of a quantity of interest (e.g. mean, variance) must be computed, e.g. a percentile or a probability distribution. However, the multifidelity variants of commonly used surrogate models such as polynomial chaos expansions and Gaussian processes suffer some of the same challenges as their traditional single-fidelity counterparts: namely, loss of accuracy from discontinuities in model outputs and the need to explicitly represent each source of uncertainty in a model as an input to the surrogate. These challenges were identified and discussed in detail in [1].

Sampling-based approaches are robust to discontinuity and do not require sources of uncertainty to be explicitly represented as long as they can be sampled. However, their applicability is currently limited to single statistics, e.g. moments of the quantity of interest [18, 19, 20], Sobol' indices [21], or percentile estimates [22]—there is not currently any practical multifidelity sampling-based method to compute a probability distribution. Typically, sampling-based multifidelity methods achieve increased efficiency by reducing the variance of statistical estimators, such as a sample mean. This reduction in the variance of the estimator means that the error in the sample estimator will, on average, be lower relative to an estimator computed using only the high-fidelity model. Because mean estimates are of interest for PA, this year the potential utility of sampling-based multifidelity methods was explored in the context of the simplified crystalline repository model. A brief summary of this model is provided here; complete details of its specification are provided in [1].

2.1 Simplified repository model in crystalline rock

For the purposes of this preliminary feasibility study, a simplified model of a repository in crystalline rock was developed which mimics the CRC as closely as possible while significantly reducing the computational cost of the model (by about 100x). This enabled the assessment of sampling-based methods for PA problems in crystalline rock without the computational overhead of running the CRC. The model is a simplification of the CRC, which is detailed extensively in [1, 2, 3]. A detailed description of the simplified model is provided in [1], but relevant features of the problem are summarized here.

The model represents a repository embedded in a sparsely fractured granite; the spatial heterogeneity induced by fractures is represented using a discrete fracture network (DFN), which is generated by randomly sampling fractures of varying size, orientation, and location. The parameterization of the probability distributions for fracture size, orientation, and location are derived from the well-characterized Forsmark site in Sweden [23]. The model simplifies from the CRC by simulating a single waste package in the repository (rather the hundreds), ignores thermal effects, is a smaller computational domain, and runs to 100,000 years rather than 1,000,000 years. However, it retains the strong spatial heterogeneity characteristic of the CRC, and most of the uncertain parameters are shared between the two models (the way waste package breach time uncertainty is parameterized is simplified here because there is only one). One additional difference between the simplified model and the CRC is sampling structure; here a new DFN is generated for each sample evaluation, while the CRC employs a nested sampling structure, where a DFN is fixed for a small number (typically 25) of epistemic samples.

A diagram of the computational domain is shown in Figure 2-1. The waste package is located in the buffer region; a leak occurs at a random time during the simulation, which is sampled from an uncertainty distribution. The repository region constitutes the buffer region and the disturbed rock zone. Surrounding the repository is the spatially heterogeneous sparsely fractured crystalline rock (spatial heterogeneity ignored for the diagram). Finally, a glacial aquifer sits at the top of the computational domain. The maximum concentration of the radionuclide ^{129}I, released from the waste package, is tracked in the aquifer region.

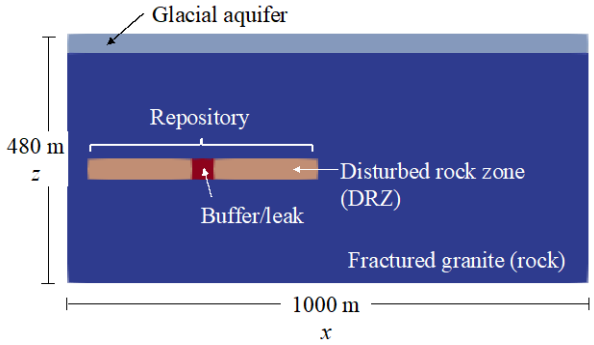


Figure 2-1. A vertical slice of the computational domain, taken at y=500 m

2.2 Sampling-based multifidelity methods

Sampling-based multifidelity methods such as multilevel Monte Carlo (MLMC) [19], multifidelity Monte Carlo (MFMC) [18], and approximate control variate Monte Carlo (ACVMC) [20] leverage evaluations from multiple models of varying accuracy and cost to achieve increased efficiency of statistical estimates. For example, the standard Monte Carlo estimator of the mean of a quantity of interest Q would be computed as

$$\hat{Q} = \frac{1}{N}\sum_{i=1}^{N} Q^{(i)}. \quad (1)$$

This estimator has variance $\mathbb{V}(Q)/N$; this means the standard error of the estimator decays slowly, at a rate of $\sqrt{N}$. For two models (Q high fidelity, costly; Q_1 low fidelity, cheaper), the MLMC mean estimator is defined as

$$\hat{Q}^{ML} = \frac{1}{N_1}\sum_{i=1}^{N_1} Q_1^{(i)} + \frac{1}{N_\Delta}\sum_{j=1}^{N_\Delta} Q^{(j)} - Q_1^{(j)} = \hat{Q}_1 + \widehat{Q - Q_1}. \tag{2}$$

This estimator is unbiased, but its variance is

$$\mathbb{V}\left(\hat{Q}^{ML}\right) = \frac{\mathbb{V}(Q_1)}{N_1} + \frac{\mathbb{V}(Q - Q_1)}{N_\Delta}. \tag{3}$$

The first part of this variance can be driven down by evaluating Q_1 many times—this is possible because it is cheap to evaluate. The second part of the variance will be small if the variance of the discrepancy between the models is small; this can occur if the variations in Q and Q_1 are very similar as functions of the uncertain inputs, e.g. as model predictions converge with mesh refinement. If these conditions hold, the MLMC estimator can achieve orders-of-magnitude smaller variance relative to MC for the same computational cost. It can be extended to more than two models by adding a series of telescoping discrepancies between models to (2). MFMC and approximate control variate (ACV) methods operate similarly to MLMC; they exploit correlations between models to achieve a similar variance reduction. MFMC exploits recursive correlation between models in a hierarchy, while ACV relaxes the assumption of a hierarchy and exploits the correlation between all models in an ensemble. See [20] for a detailed definition of each of these methods, how they are related, and how they differ. Each of these estimators is implemented in the most recent release (6.16) of Dakota [16].

For each of these estimators it is possible to derive an optimization problem for the optimal number of samples to allocate to each model in the ensemble. The objective function is different for each estimator, but they all contain variances and/or correlations which must be estimated numerically. This is typically done using a small number of samples shared between all models, called a pilot study. See [20] for further discussion on allocation strategies.

A typical sampling-based multifidelity procedure entails an initial pilot study, wherein a small number of samples are evaluated for all models in the ensemble, and these are used to compute sample correlations, which feed into the optimal sample allocation and projected estimator variance for a given budget. Following the pilot study, the sample allocation obtained from the pilot study can either be used directly for a UA simulation campaign, or an iterative procedure can be followed wherein additional sample evaluations from all models are obtained to improve the accuracy of the sample correlations before embarking on the simulation campaign. Both options are available in Dakota. The second option was taken here.

Multifidelity sampling-based methods are generally more effective when there is a significant difference in cost and high correlation between models. However, each sampling-based multifidelity method exploits the relationships between models in an ensemble and combines samples from the models in a different way. Because of this, any one of the methods can outperform the others (in terms of variance reduction), depending on the nature of the model ensemble. The pilot study can be used to project the performance of each of the sampling-based

multifidelity estimators for a given budget, so it is recommended to compare the projected performance of the estimators before embarking on a costly simulation campaign. This approach was taken here.

2.3 Results

The potential utility of sampling-based multifidelity methods such as multilevel Monte Carlo (MLMC) [19], multifidelity Monte Carlo (MFMC) [18], and approximate control variate Monte Carlo (ACVMC) [20] was explored here for the mean log-base-10 peak (maximum over time and space) ^{129}I concentration in the aquifer. A model ensemble was defined in terms of the spatial discretization applied to the computational domain. A uniform hexahedral mesh of 10, 20, and 40 m edge length was applied.

A 25-sample pilot study was performed; the sample correlations between models for this study are reported in Table 2-1. This study was also used to compute the average computational cost, measured in core time, for each of the mesh sizes, reported in Table 2-2. The correlations and relative costs for this model ensemble are good, but additional modifications could be made to improve them even further; e.g. modifying timestep algorithm settings to make the coarser spatial discretizations even cheaper. Because this is an exploratory study, we did not explore such modifications; however, it will be a key area of investigation when these methods are deployed to the CRC.

Table 2-1. Correlations between mesh sizes for the $\log_{10}$ peak ^{129}I concentration in the aquifer

Mesh size [m]	10	20	40
10	1.00	0.98	0.91
20	0.98	1.00	0.95
40	0.91	0.95	1.00

Table 2-2. Absolute and relative model costs

Mesh size [m]	Core time [s]	Relative cost
10	83498	1.0
20	917	1.1e-2
40	83	9.9e-4

The performance of MFMC and ACVMC were considered here relative to MC. MLMC was not considered here despite the fact that the model ensemble arises from varying numerical discretizations of the same equation, a common use case for the method, because the previous year's analysis of the model ensemble indicated there was a lack of convergence as the mesh was refined for the peak ^{129}I, which is expected to lead to suboptimal performance for the method. (This lack of convergence is hypothesized to arise from the false connections that create spurious flow pathways between the repository and the aquifer region. Adaptive meshing could mitigate this issue and will be considered in future studies.)

The pilot study was used to obtain and compare the projected estimator variances of MFMC and ACV MFMC (a multifidelity variant of ACV) for a computational budget equivalent to 500 high-fidelity simulations. These are reported in Table 2-3.

Table 2-3. Projected estimator variances for single-fidelity and multifidelity sampling-based methods

Method	Projected estimator variance	MC variance / Projected variance
MC	2.00e-2	1.0
MFMC	2.11e-3	9.5
ACV MFMC	2.41e-3	8.3

The greatest projected variance reduction is achieved with MFMC, with a variance reduction of almost 10x. An iterated study with a computational budget of 500 equivalent high-fidelity evaluations was performed using MFMC, where the sample correlations between models are updated with additional shared evaluations between models. This iteration resulted in 409 shared samples between models to compute sample correlations and projected MFMC estimator variance. This updated projected variance reduction relative to MC dropped to 7.6x, which is lower than was projected with the smaller pilot sample size, but it is still a nonnegligible reduction in variance. A summary of the number of evaluations for each model and a comparison of the approximate variances for MC and MFMC are presented in Table 2-4 (variances are approximate since sample variances and correlations must be substituted in analytical expressions for estimator variance).

Table 2-4. The number of evaluations per model, computational cost, and estimator variance for MC and MFMC methods

Method	Number of evaluations	Computational cost	Estimator variance
MC	Q_{10}: 500	500	3.63e-3
MFMC	Q_{10}: 409 Q_{20}: 4780 Q_{40}: 38507	499.8	4.89e-4

This is a positive result; however, with a good model ensemble, multifidelity sampling-based methods can achieve a much greater reduction in variance for the same computational cost—often multiple orders of magnitude. Given the limited exploration of possible ways to derive lower-fidelity models for this problem, it is likely that even better variance reduction could be achieved by exploring other methods besides simply coarsening the spatial discretization to make the lower-fidelity models cheaper and/or more correlated with the highest fidelity. For instance, time step algorithm settings could be modified to make the coarser spatial discretizations even cheaper to run, though this will also impact accuracy. For more complex PA cases, models with simplified physics could be used as lower-fidelity models in the ensemble. Since this problem is meant only to inform the feasibility of multifidelity methods for the repository PA application area, further exploration of the potential lower-fidelity models was not considered here.

2.4 Conclusions

Multifidelity surrogate-based and sampling-based methods were explored in the context of a simplified crystalline repository model. These methods show promise to improve the efficiency of uncertainty analyses for PA, but additional tuning of model ensembles and surrogate construction algorithms will be needed to realize their full potential. Additional numerical studies for sampling-based methods and a detailed discussion of the surrogate-based method's results are documented in a forthcoming journal article, which is in preparation [24]. A draft may be available from the authors upon request. Future research directions will focus on deploying multifidelity sampling-based methods to the CRC and other reference cases.

As shown in Table 2-4, the reduction in the variance of the mean estimator was shown with a MFMC approach using samples from multiple fidelity runs with a total computational cost that was the same as the cost of the original high fidelity samples. Note that another way the increased precision in the multifidelity sampling-based estimators can be exploited is to maintain the same accuracy in the estimator, but at a reduced cost. In this situation, the goal would not be to reduce the estimator variance but to achieve it with a number of samples across fidelities that have a total computational cost which is less than the cost of the high fidelity samples (e.g. 500 high fidelity samples in the case study in Section 2.3). This cost savings can thus enable even higher fidelity models to be developed and incorporated in probabilistic PAs. This will be considered in future investigations.

3. SENSITIVITY ANALYSIS METHODOLOGY: ASSESSMENT OF SURROGATES

Previous sensitivity analyses with the GDSA Framework have utilized linear methods and variance decomposition [1, 2, 3, 4, 5]. Linear methods are appealing because they are relatively easy to apply and interpret whereas variance decomposition is appealing because it offers more flexibility in detecting significant nonlinear relationships. However, variance decomposition is also more complicated to apply and interpret than linear methods.

As described in [3], there are multiple ways to perform variance decomposition. Application in the GDSA Framework has focused on estimating Sobol' indices [25]. These indices describe the proportion of variance in a quantity of interest that can be attributed to variance in each input parameter. It can be calculated using a large number of simulations with a specific sampling structure [26]. This process is feasible for computationally inexpensive models but is not always feasible for computationally expensive models.

Previous analyses with the GDSA Framework have overcome this challenge using surrogate models [1, 2, 3]. Surrogate models are computationally inexpensive models that can be trained to estimate the behavior of a computationally expensive model. Sobol' indices can then be estimated by exercising the efficient surrogate model many times. The accuracy of the Sobol' indices estimated using a surrogate model depends on the quality of this surrogate.

Surrogate models have previously proven successful at identifying the variables driving uncertainty in QoIs for GDSA Framework simulations [1, 2, 3]. These results were consistent with our phenomenological understanding of the modeled systems and with patterns identified in scatter plots. However, we have also noted that further study on surrogate quality is required [1].

This year's work on sensitivity analysis aimed to further investigate surrogate model quality and the factors that interfere with surrogate quality. For the crystalline reference case, in particular, the separation of spatial and parameter uncertainties presents a challenge for surrogate models for two reasons: 1) the spatial uncertainty is not defined parametrically, so spatial uncertainty cannot be directly included in the surrogate model construction, and 2) the nested sampling structure (Section 5.1.1) results in repeated samples within the spatial loop, which may bias surrogates.

To investigate these issues, we structured an analysis to compare Sobol' indices calculated by directly sampling the model (Section 3.1) and Sobol' indices estimated using surrogate modeling (Section 3.1). Comparison between the calculated indices and the estimated indices indicates the quality of the surrogate. This would not be possible with the crystalline reference case (CRC); it is costly, requiring ~256 core hours per simulation in its current form, so a numerical study resolving the Sobol' indices to adequate accuracy with several thousand samples would not be computationally feasible. Instead, we performed this analysis with the coarsest mesh for the simplified crystalline model which is described briefly in Section 2.1 and in detail in [1]. This model requires ~30 seconds on one core to run, so is it computationally feasible to obtain the computed Sobol' indices for this model, which provides a unique opportunity to make this comparison. However, it still retains the features in the CRC which are challenging to surrogates, namely the spatial uncertainty which is not defined parametrically but instead manifests through a sampled DFN. This year we focused on the first challenge of unparameterized spatial

heterogeneity and its impact on surrogate accuracy. Understanding the effect of sampling structure and sample size on surrogate accuracy will be the subject of future work.

Another advantage of using the simplified model is that the DFNs are also relatively inexpensive to generate and use. The procedure for calculating Sobol' indices requires an initial sample set with unique spatial and variable samples for each realization. This sample is then modified to construct the remaining samples for the calculation, as described in Section 3.1, which introduces repeated values to each dimension in the set. As a result of this procedure, there is a smaller sample set with all unique samples, and a larger sample set with repeated values. Constructing the surrogate on the smaller set allows it to be constructed with unique DFN samples for each parametric sample, removing the issue of repeated sampling in the spatial loop which biases the surrogate. This allows for an isolated study of the impact unparameterized spatial uncertainty has on surrogate quality.

Section 2.1 describes the simplified case. Section 3.1 describes how it was used to calculate Sobol' indices, Section 3.1 describes the sampling, and Section 3.3 details each of the analyses and their results. This chapter ends with a brief summary discussion in Section 3.4.

3.1 Uncertain random variables

The uncertain random variables are similar to the full-scale crystalline reference case, except for the way uncertainty in waste package breach is modeled. Since there is only one waste package in the simplified model, a single uncertain breach time is simulated, rather than a distribution over the degradation of the canister, which is sampled for each waste package in the full-scale crystalline reference case. The range of values for the breach time was selected so that the waste package would breach in the first part of the simulation; it is not based on any physical information. The uncertain random variables and their distributions are reported in Table 3-1. The $\mathcal{U}$ notation describes a uniform distribution with upper and lower bounds. The $\log \mathcal{U}$ notation describes a log-uniform distribution with upper and lower bounds. A log-uniform distribution is uniform in log space.

These parameters are explained in more detail in [6], where Chapter 4 of [6] provides literature references for each particular aspect (e.g. buffer, waste package, etc.) Most of these parameters have a nominal value obtained from the literature. The uncertainties around these estimates were best estimates provided by the authors of [6]. These are meant to be reasonable uncertainty estimates given limited information and should only be taken as illustrative for the purposes of demonstrating the UQ/SA analyses in this report. For an actual performance assessment, these parameters would need to be updated and re-examined with site-specific information.

Table 3-1. Uncertain parameters in the simplified crystalline problem

Parameter	Description	Distribution
rateUNF	Waste form bulk dissolution rate [yr^{-1}]	$\log \mathcal{U}[10^{-8}, 10^{-6}]$
kGlacial	Glacial aquifer permeability [m^2]	$\log \mathcal{U}[10^{-15}, 10^{-13}]$
permDRZ	DRZ permeability [m^2]	$\log \mathcal{U}[10^{-19}, 10^{-16}]$
permBuffer	Buffer permeability [m^2]	$\log \mathcal{U}[10^{-20}, 10^{-17}]$
pBuffer	Buffer porosity [-]	$\mathcal{U}[0.3,0.5]$
wpBreachTime	Waste package breach time [yr]	$\mathcal{U}[2500,10000]$

Additionally, for the direct sampling-based sensitivity analysis, the *float_seed* random variable distributed as $\mathcal{U}[0,1]$ is included in the set of random variables. The *float_seed* provides a seed value to the DFN random number generator.

For some of the studies reported below, spatial uncertainties represented by graph metrics were also included. Their distributions were determined using the procedure reported in [1]. They were included because of their nonnegligible correlation with performance quantities of interest. The graph metrics investigated are [1]: nIntersections (the number of intersections in the DFN), NIwR (the number of fractures intersecting the repository region), and STT (the shortest travel time between repository and aquifer along one of the flow pathways between them; it is computed relative to the median over all DFNs and is log-base-10 transformed).

3.2 Sobol' Index Calculation

The Sobol' indices incorporating spatial heterogeneity were computed using a Saltelli sampling-based method [26] implemented in Dakota [16]. The benefit of this approach relative to a surrogate-based approach is that the effect of spatial heterogeneity does not have to be parameterized as an input as it would for a surrogate; it is sufficient that it can be sampled (by generating a DFN) and fixed (by fixing the seed for the DFN random number generator). To this end, the set of uncertain parameters for the simplified model was augmented with a uniform random variable with range [0,1]. This variable, called the *float_seed*, was multiplied by 10^9 and rounded to the nearest integer to provide an integer seed to the DFN generation algorithm.

The Saltelli sampling-based method for computing Sobol' indices employs a so-called "pick and freeze" procedure. This procedure is briefly explained here, but a more detailed description of the approach is available in [26]. First, two sets of independent, identically distributed input sample sets of size N are generated. Let these be denoted $\boldsymbol{X}, \boldsymbol{X}' \in \mathbb{R}^{N \times d}$, where d is the number of random inputs. Sobol' indices associated with the k^{th} random input are computed using model evaluations $f(\boldsymbol{X}), f(\boldsymbol{X}')$, and $f(\boldsymbol{X}'_{\boldsymbol{k}})$, where $\boldsymbol{X}'_{\boldsymbol{k}} = [\boldsymbol{x}'_{\boldsymbol{1}}, \boldsymbol{x}'_{\boldsymbol{2}}, \dots, \boldsymbol{x}_{\boldsymbol{k}}, \dots, \boldsymbol{x}'_{\boldsymbol{d}}]$ (note that the k^{th} column is taken from $\boldsymbol{X}$ while all other columns are taken from $\boldsymbol{X}'$). The structured sampling associated with $\boldsymbol{X}'_{\boldsymbol{k}}$ enables the computation of the conditional expectations necessary for the Sobol' indices. A full set of Sobol' indices for all random variables thus requires $2N$ model evaluations (for $f(\boldsymbol{X})$ and

$f(\boldsymbol{X}'))$ and dN evaluations (for $f(\boldsymbol{X}'_{\boldsymbol{k}}), k = 1, \ldots, d$). Thus, for a user-specified N, $(d+2)N$ model evaluations are required to compute the Sobol' indices. Note that $2N$ independent samples are generated as a result of this sampling procedure.

An adequately converged set of sampling-based Sobol' indices were computed to serve as a reference for comparison for the surrogate-based indices. These computed indices are referred to as the "calculated" values in the Section 3.3 analyses because they are inherently more accurate than surrogate-based indices and are treated as the true values that the surrogate models are used to estimate. The convergence of the indices was assessed by generating five replicate sets of indices for a given sample size and inspecting the variation in the indices using box plots. First, a sample size of 1,000 was considered. For each replicate study, this resulted in (7+2)1,000=9,000 model evaluations. This brought the total number of model evaluations for the five replicate studies to 45,000; there are 5 replicates, each with 9,000 model evaluations. A scatterplot of the five replicate sets of Sobol' indices are presented in Figure 3-1 for the peak ^{129}I concentration in the aquifer. Note that the bounds on the total effect indices overlap significantly for *kGlacial* and *float_seed*, which makes it challenging to confidently rank the variables by importance.

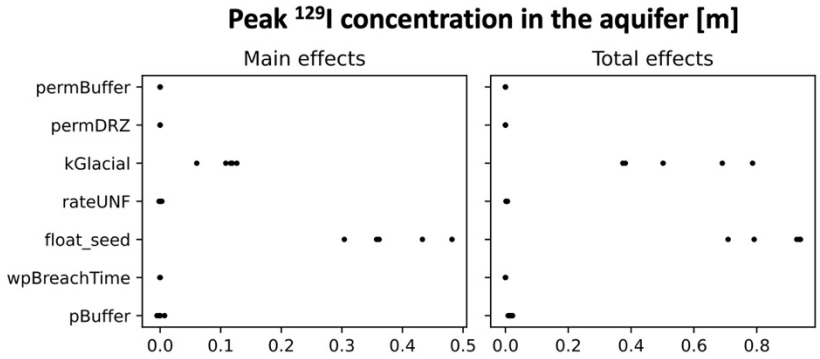


Figure 3-1. Scatterplots for the five replicate sets of Sobol' indices computed for sample size 1,000

Based on these results, Sobol' indices were computed using five additional replicates each with a sample size of 5,000 to drive down the variation in the Sobol' indices. A scatterplot of the five replicate studies is shown in Figure 3-2 for the peak ^{129}I concentration in the aquifer. There is still some spread in the Sobol' indices for *kGlacial* and *float_seed*, but there is no longer an overlap in their values, so ranking of the importance of the variables is possible. For this reason, the 5,000 sample size case was considered adequately converged for use as a reference for comparison with the surrogate-based indices. A similar check was performed for the other performance QoIs, but discussion of the process focused on a single QoI where there was overlap for 1000 samples for simplicity.

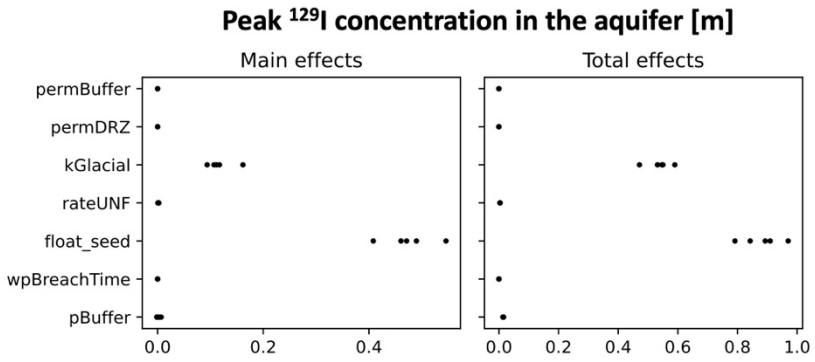


Figure 3-2. Scatterplots for the five replicate sets of Sobol' indices computed for sample size 5,000

It should be noted that the computational costs to obtain these Sobol' indices were significant. A sample size of 5,000 corresponds to (7+2)5,000=45,000 model evaluations for each replicate study; a set of five replicate studies corresponds to 225,000 model evaluations. The low computational cost of the coarsest simplified crystalline repository model (only ~30 seconds on one core) made this computationally feasible (though impractical), but such a numerical campaign would not be possible for production level models like the crystalline reference case. Because this sampling method uses $2N$ unique samples, each of the five replicates produced 10,000 unique samples to be used in the surrogate model construction.

3.2.1 Extracting Samples for Surrogate Analysis

This year, analysis focused on the challenges to surrogate accuracy associated with unparameterized spatial heterogeneity. To isolate the issues associated with this and mitigate its conflation with accuracy of the surrogate from too few samples, the surrogates were constructed with a large number of samples extracted from the Saltelli-based sampling used for the computed Sobol' indices. The independent samples produced as a part of the Saltelli-based sampling were extracted as build points for the subsequent surrogate analyses. These samples were pulled both from the 1,000 and 5,000 sample size cases. For a single replicate study, recall that there are $2N$ independent input/output samples. However, each of the replicate studies constitutes an independent set of samples, so these can be concatenated together to create a larger independent sample set. The number of independent samples extracted from each of the sample sizes used in the Saltelli-based sampling are presented in Table 3-2.

Table 3-2. The number of independent input/output samples extracted from each sample size for the Saltelli-based sampling procedures

Saltelli-based sample size	Number of independent input/output samples (single replicate study)	Number input/output samples (combining 5 replicates)
1,000	2,000	5(2,000)=10,000
5,000	10,000	5(10,000)=50,000

3.3 Surrogate Analysis

Multiple small studies were performed with simulations from the simplified crystalline reference case to prioritize the factors that affect surrogate quality. We have noted previously [1, 2] that the inclusion of spatial heterogeneity creates challenges for surrogate modeling with the full-scale crystalline reference case (see Section 5). However, there are additional characteristics of data sets for the crystalline reference case or surrogate models that may impact surrogate quality, such as non-monotonic behavior, differing scales between variables and QoIs, surrogate modeling form, etc. The studies described in this section aim to investigate these factors individually to qualitatively inform our future approaches to improve surrogate quality for the full-scale model.

The types of analyses performed with the simplified case are described in Table 3-3. The sample sizes described in this table are the samples used in the surrogate model construction and Sobol' index estimation; a larger sample was used to construct the sample required for Sobol index calculation.

Results in this section are provided as plots showing the Sobol' indices estimated using the surrogates and those that were calculated as described in Section 3.1. The calculated indices are treated as the true values that the surrogate models are intended to predict. Main effect indices are plotted on the left frame in the figures. These indices describe the proportion of variance in the QoI that the surrogate model attributes to each individual input variable on its own. The total effect indices are plotted in the right frame in the figures. These indices describe the proportion of variance in the QoI that the surrogate model attributes to each individual input variable on its own and through its interactions with other variables. Total indices that are significantly larger than main effects indices indicate significant interaction effects.

Table 3-3. Analysis performed to investigate surrogate model quality for sensitivity analysis

Analysis	Description	Analysis Question(s)
Surrogate Choice Analysis Section 3.3.1	Samples*: 10,000	Does our choice of surrogate model have a significant effect on sensitivity analysis results?
	Spatial Variable: Not Included	
	Surrogate Models: Polynomial Chaos Expansion (PCE), Polynomial, Multiple Adaptive Regression Splines (MARS), Neural Network (NN)	
Float Seed Analysis Section 3.3.3	Samples*: 10,000	Can surrogate models attribute spatial variability to a random seed?
	Spatial Variable: Float Seed	
	Surrogate Models: Mars	
Y-Location Analysis Section 3.3.2	Samples*: 10,000	Can y-location of peak ^{129}I be used as a proxy variable to summarize spatial heterogeneity in sensitivity analysis?
	Spatial Variable: Y-Location of Peak ^{129}I	
	Surrogate Model: Mars	
Graph Metric Analysis Section 3.3.4	Samples*: 10,000	Do graph metrics improve surrogate quality? Are they sufficient for summarizing spatial heterogeneity?
	Spatial Variables: nIntersections, NIwR, STT	
	Surrogate Model: Mars	
	Samples*: 10,000	

Transformation Analysis Section 3.3.5	Spatial Variables: nIntersections, NIwR, STT	Can transformations improve surrogate model performance?
	Surrogate Model: Mars	
Replicate Analysis Section 3.3.6	Samples*: 10,000 x 5 replicates	Are sensitivity analysis results consistent across replicates? Should they be?
	Spatial Variables: nIntersections, NIwR, STT	
	Surrogate Model: Mars	
* These sample sizes refer to the number of realizations used for surrogate model construction, $2N$ (see Section 3.1)		

3.3.1 Surrogate Choice Analysis

The surrogate choice analysis compares four different surrogate models that were trained to the same data set. The models include quadratic polynomial (Poly2), second order polynomial chaos expansion (PCE2), multiple adaptive regression splines (MARS), and neural network (NN) models, all applied within Dakota [16]. This analysis was performed on simulations that include spatial heterogeneity, but no variables characterizing the spatial heterogeneity were included in the surrogate model construction. This means that the data includes uncertainty that cannot be fully described by the surrogate models. Comparing the surrogates in this context is interesting because different models may have different propensities to over-fit. In other words, some models may be more likely than others to misattribute variation from spatial heterogeneity to epistemic variables or combinations of epistemic variables. Overfitting could lead to misleading sensitivity analysis results with our current imperfect characterization of spatial heterogeneity; other potential differences between surrogate model options were not addressed in this analysis.

The sensitivity analysis results from the surrogate choice analysis for the ratio between the aquifer to east boundary flux and the rock to east boundary flux at 100,000 years are plotted in Figure 3-3 and the corresponding scatter plots of input variables versus this QoI are plotted in Figure 3-4. This is an example of a quantity of interest with a fairly straightforward sensitivity analysis. As we can see in the scatter plots, there is a very clear trend between the glacial till permeability (*kGlacial*) and this flux ratio. However, there is some variation in this linear trend, which is due to spatial heterogeneity. In the absence of any variable characterizing this heterogeneity, all four models attribute the variation solely to *kGlacial*, estimating main and total effect indices near 1.0. In context, we would conclude that the sensitivity analysis identifies the most important epistemic variable, but all models over-estimate the sensitivity indices because they estimate that epistemic variables describe 100% of the variance even though it is not possible to explain 100% of the variance without attributing any to spatial heterogeneity.

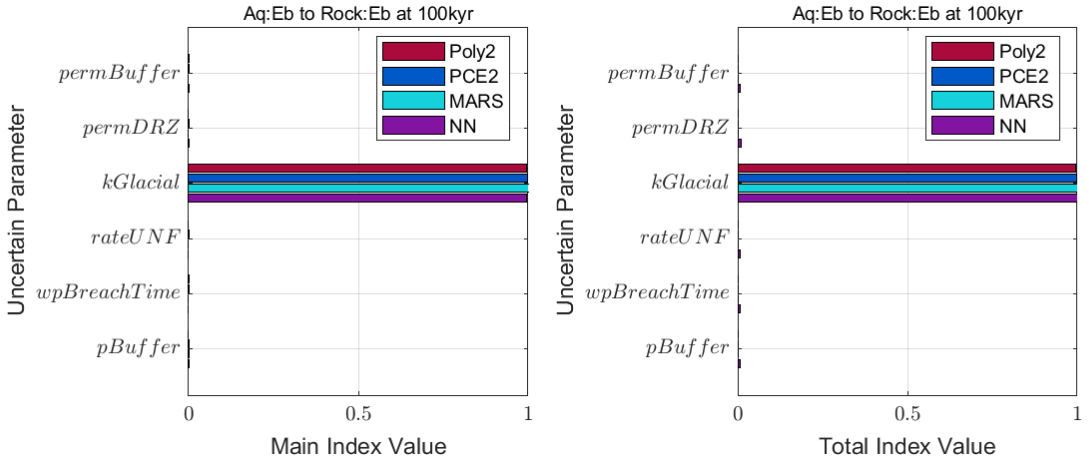


Figure 3-3. Sensitivity analysis results for the ratio of aquifer to east boundary flux and rock to east boundary flux at 100,000 years from the surrogate choice analysis

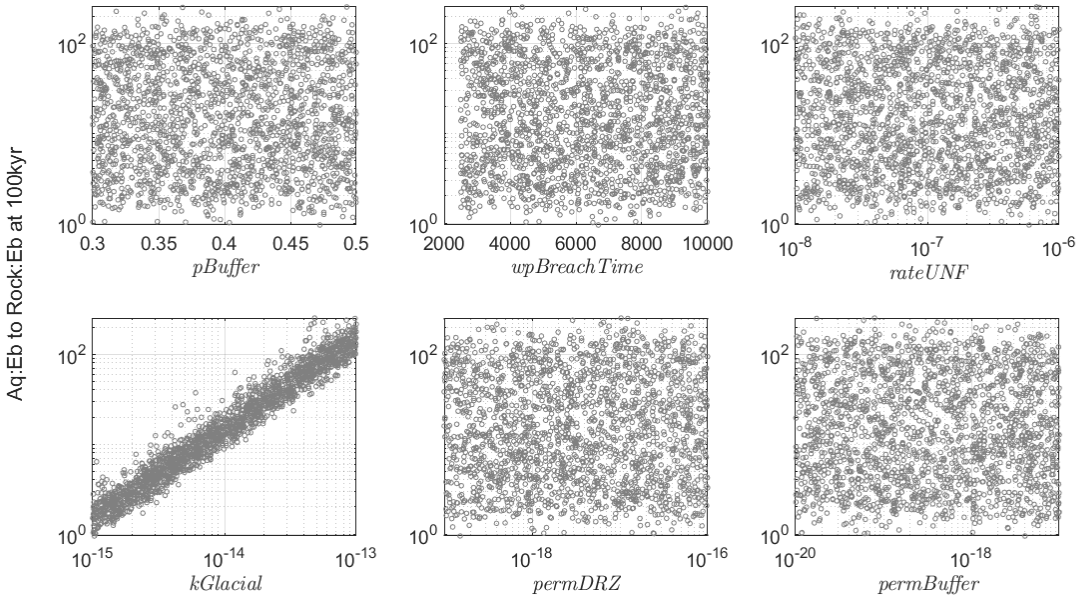


Figure 3-4. Scatterplots of ratio of aquifer to east boundary flux and rock to east boundary flux at 100,000 years versus the uncertain parameters

The sensitivity analysis results from the surrogate choice analysis for the peak ^{129}I concentration are plotted in Figure 3-5 with the corresponding scatterplots in Figure 3-6. This analysis is similar to that for the previous QoI except that more uncertainty in the peak ^{129}I is due to spatial heterogeneity, yet all models still assign almost all of the variance to *kGlacial*; since no variables representing spatial heterogeneity were included in the surrogate construction, this means the surrogate is overfitting by assigning that variance to *kGlacial*. The neural network model is unique in that it spreads more of the variability over variable interactions than the other models.

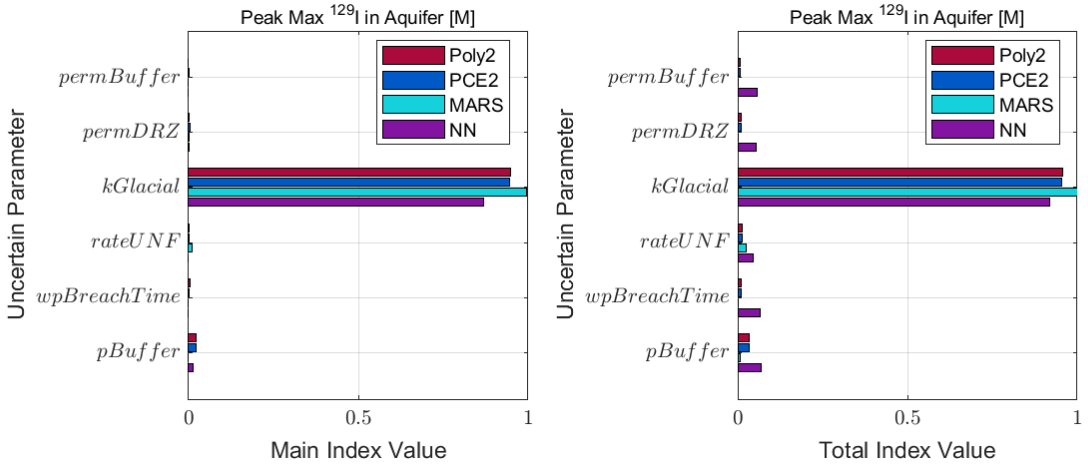


Figure 3-5. Sensitivity analysis results for the peak ^{129}I concentration [M] in the aquifer from the surrogate choice analysis

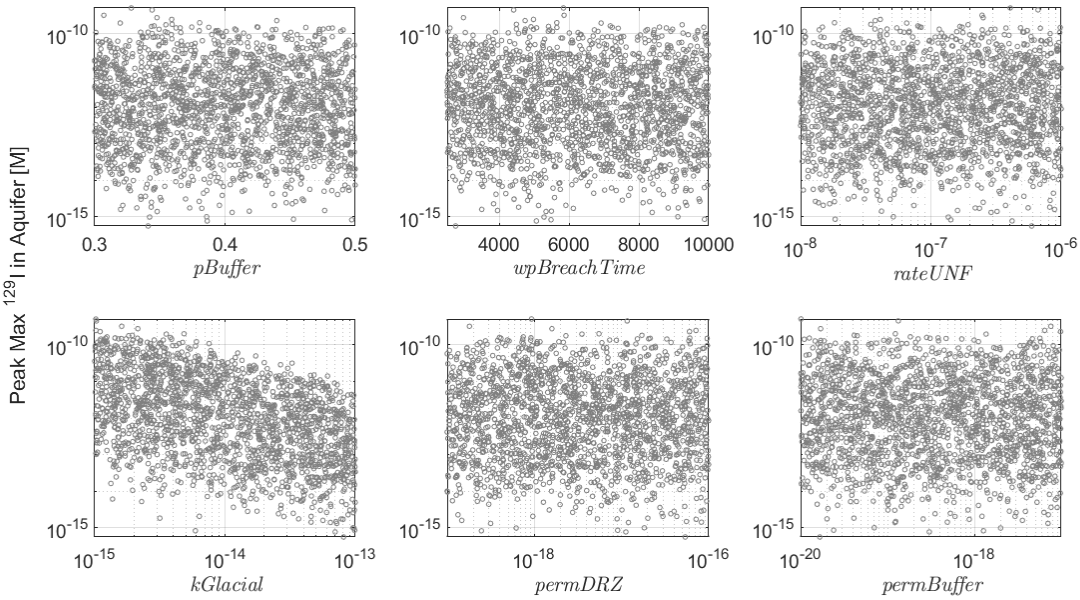


Figure 3-6. Scatterplots of the peak ^{129}I concentration [M] versus the uncertain parameters

The sensitivity analysis results from the surrogate choice analysis for the y-location of the peak ^{129}I concentration and corresponding scatter plots are shown in Figure 3-7 and Figure 3-8, respectively. Note that the scatter plots do not have any discernable trends. In the sensitivity analysis results, the polynomial and neural network models both assign some significant variance to individual variables on their own. Neither the MARS model nor the PCE model attribute any significant variance to the epistemic variables. This may suggest that these models are less prone to overfitting with respect to main effects indices. However, the analysis was also run with the smaller sample size ($2N_{samples} = 2000$). When trained to less data, the PCE model overfit both main and total effects indices, and the MARS model overfit total effects indices. Hence, all of the

models are, to some extent, vulnerable to overfitting and the occurrence overfitting depends on the size of the training data set. This analysis demonstrates the potential for overfitting regardless of model choice with the sample size limitations imposed on a PA by computational costs.

The scatterplots for this the y-location of peak ^{129}I concentration show that it is largely independent of the uncertain input variables. Figure 3-9 shows a grid of scatterplots. Each plot frame is the scatterplot of an input variable versus another input variable, and all plots are colored by the corresponding values for the y-location. The purpose of a plot like this is to identify any two-way interaction effects; if any of the plot frames show a pattern with respect to color, then the two variables in that plot frame have a conjoint influence on the y-location.

The lack of patterns within the color scale for all plots in the grid shows there are no clear two-way interactions between input variables that drive the behavior of the y-location QoI. Because of this independence, we considered the possibility that the y-location may be a good proxy for the input spatial heterogeneity. Phenomenologically, it makes sense that the location of the peak concentration should be impacted by the spatial heterogeneity within the fracture network, as well as the spatial variability in waste package breach times. Even though it is not an uncertain input to the repository model, we considered using y-location as an input in the sensitivity analysis to incorporate spatial heterogeneity in the surrogate model construction (see Section 3.3.2).

Overall, the surrogate models in this study performed similarly. They provided similar results for clearly important variables and tended to overfit when not provided with variables characterizing spatial heterogeneity. The MARS model may be slightly less prone to overfitting with respect to the main index estimation based on the results for y-location. For the sample sizes used in this study, the polynomial, PCE, and MARS models have roughly the same computation cost, and the NN model is less efficient in terms of its training cost. This analysis does not heavily favor one model over the other, but a single model (MARS) was chosen for the remaining studies to simplify those comparisons because it was less prone to overfitting with sufficient sample size in this analysis.

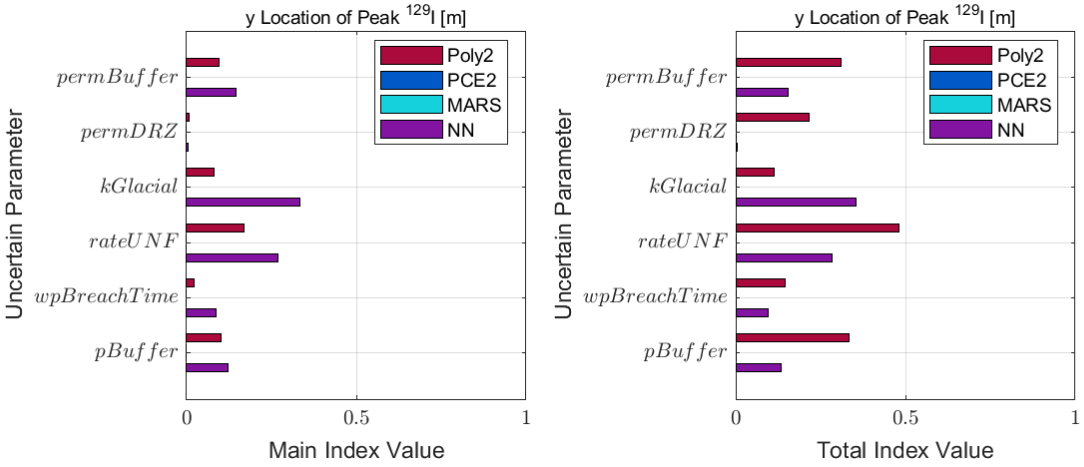


Figure 3-7. Sensitivity analysis results for the y-location [m] of the peak ^{129}I concentration in the aquifer from the surrogate choice analysis

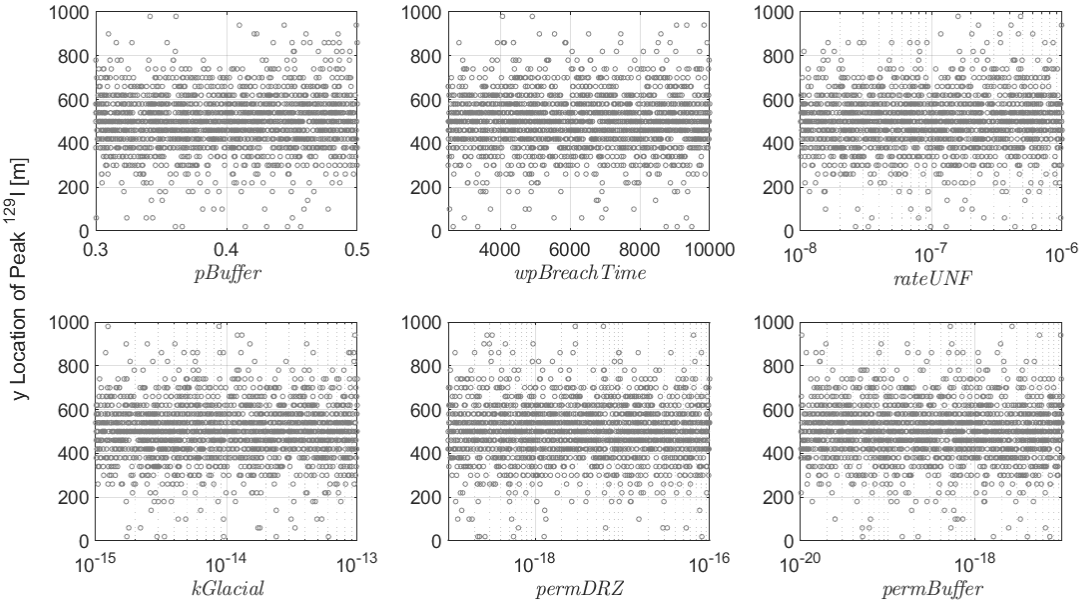


Figure 3-8. Scatterplots of the y-location [m] of the peak ^{129}I concentration versus the uncertain parameters

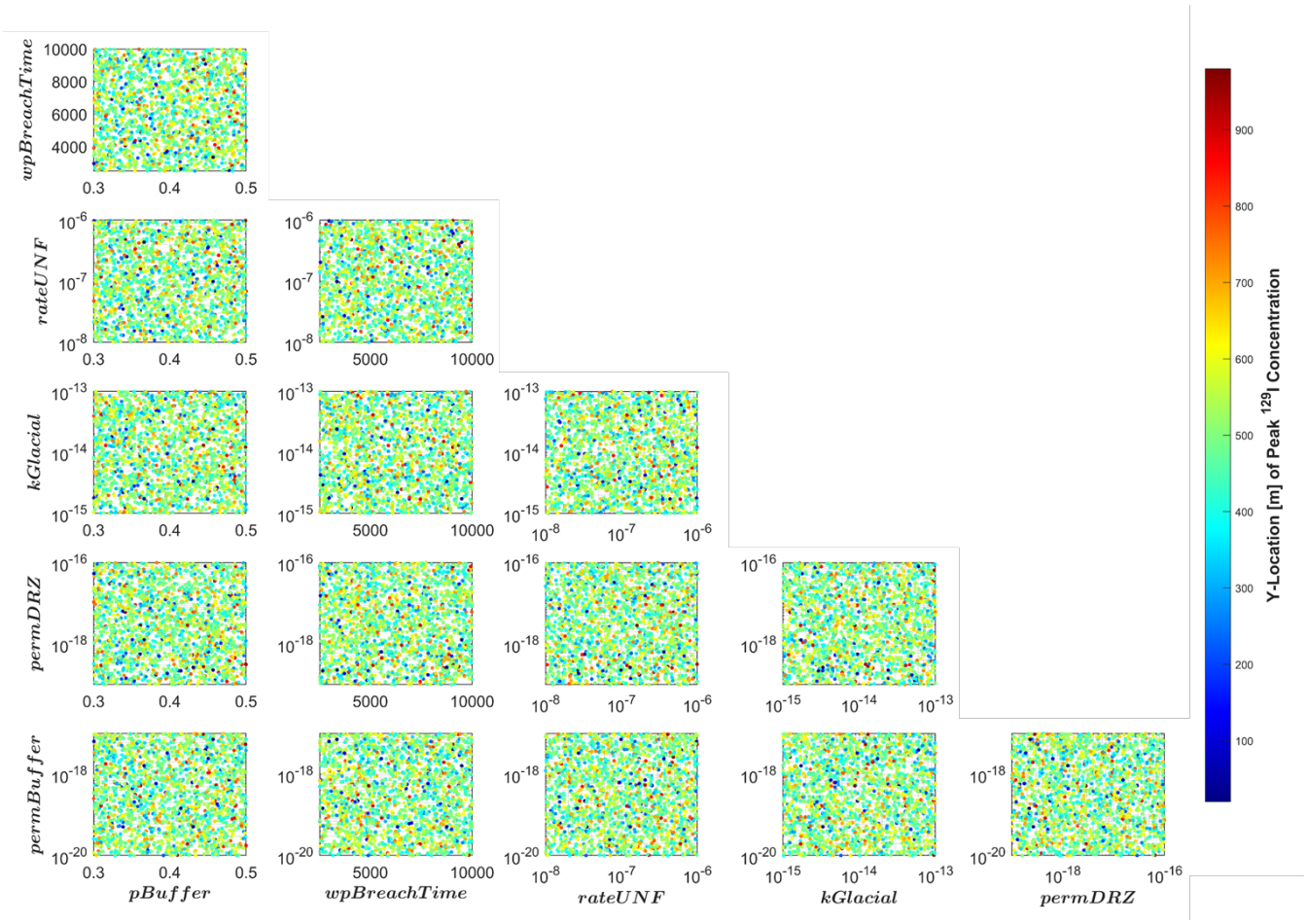


Figure 3-9. Plot matrix of the uncertain parameters plotted versus each other and colored by the y-location [m] of peak ^{129}I

3.3.2 Y-Location Analysis

As discussed in the previous section, the y-location of the peak ^{129}I concentration appears to be independent of the uncertain input variables but should be influenced by the spatial heterogeneity. This led us to consider using the y-location as a proxy variable for spatial heterogeneity in the simplified crystalline reference case. For this study, we included the y-location in the surrogate construction and then plotted the sensitivity analysis results with the calculated Sobol indices (see Section 3.1). The calculated Sobol' indices for the float seed are plotted for the y-location because these are the indices that the analysis aims to estimate using the proxy variable. Recall from Section 3.1 that *float_seed* defines the random seed for DFN generation and the value of *float_seed* is used in the Sobol' "calculated" values but not in the surrogate-based approach. Hence, the Sobol' indices for *float_seed* are the Sobol' indices characterizing the effects of spatial heterogeneity. They are plotted with the y-location Sobol' indices in this analysis so we can evaluate whether using y-location results in a Sobol' surrogate-based index estimate similar to the calculated Sobol' indices for spatial heterogeneity. Calculated values from five replicates are

shown, which enables comparison between the estimates and the level of variation seen in calculated values.

The sensitivity analysis results for the ratio between the aquifer to east boundary flux and the rock to east boundary flux at 100,000 years are plotted in Figure 3-10. The inputs are plotted versus this QoI in Figure 3-4, and y-location is plotted versus this QoI in Figure 3-11. Analysis of the scatterplots in Figure 3-4 suggests that *kGlacial* should be the most important variable driving uncertainty in this QoI, but influence from spatial heterogeneity is also expected. The calculated values in Figure 3-10 show this to be the case. Over 10% of the variance is attributed to *float_seed* from the Sobol' index calculation. However, the surrogate model does not attribute any of this variance to the y-location, and the scatterplot in Figure 3-11 supports this conclusion. This analysis showed that, at least for this QoI, the y-location of peak ^{129}I is not a viable proxy to represent spatial heterogeneity in sensitivity analyses.

Sensitivity analysis results are presented for the median residence time in Figure 3-12. Scatterplots for comparison are plotted in Figure 3-13. and Figure 3-14.. This analysis is similar to that for the previous QoI, except that much more of the variance should be attributed to the spatial heterogeneity. The calculated indices attribute more than 90% of the variance to spatial heterogeneity. However, the surrogate model attributed almost all of the variance to the buffer porosity (*pBuffer*). In this case, the scatterplot versus y-location (Figure 3-14.) also suggests there is no meaningful trend. The y-location is not a useful proxy for spatial heterogeneity for this QoI.

Results from the sensitivity analysis for ^{129}I concentration are plotted in Figure 3-15. (scatterplots in Figure 3-6 and Figure 3-16). We are particularly interested in characterizing the effects of spatial heterogeneity for this QoI because it is our performance metric for the full-scale crystalline reference case at this stage in development. Even for this QoI, however, the y-location does not capture much of the influence from spatial heterogeneity. Overall, the y-location did not prove to be a useful proxy for spatial heterogeneity for any of the QoIs.

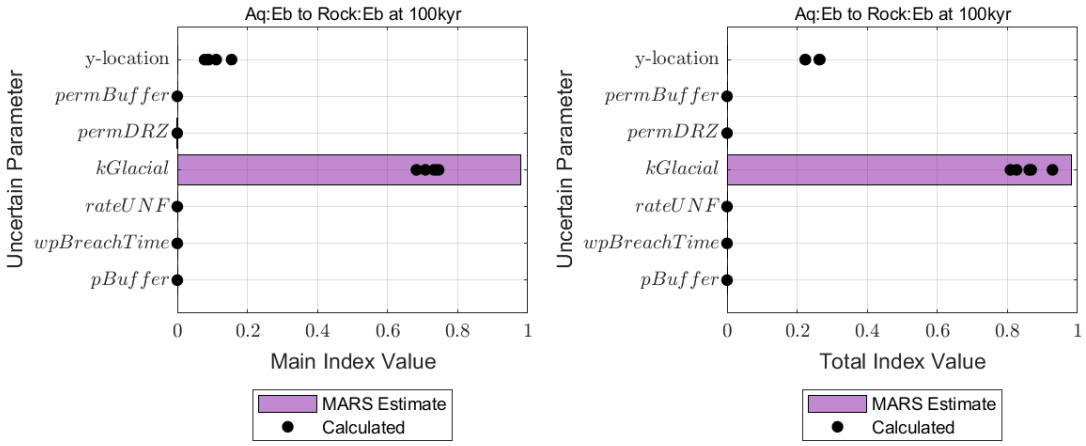


Figure 3-10. Sensitivity analysis results for the ratio between aquifer to east boundary flux and the rock to east boundary flux at 100,000 years from the y-location analysis; calculated values plotted with y-location are the calculated values for the float seed

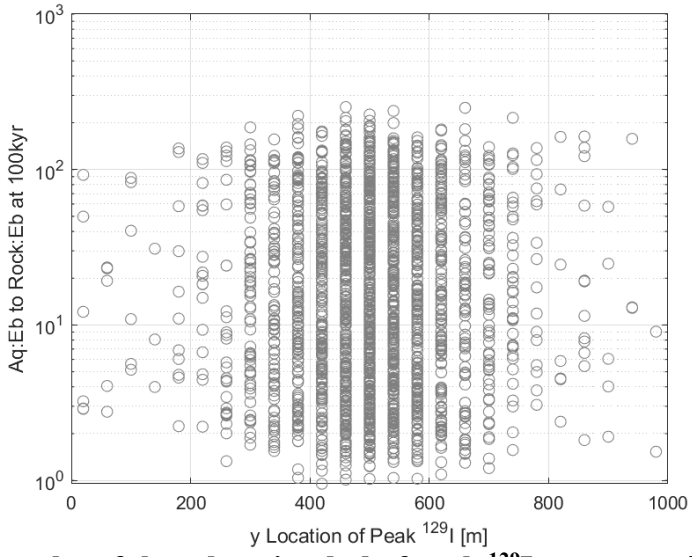


Figure 3-11. Scatterplot of the y-location [m] of peak ^{129}I concentration versus the ratio between aquifer to east boundary flux and the rock to east boundary flux at 100,000 years

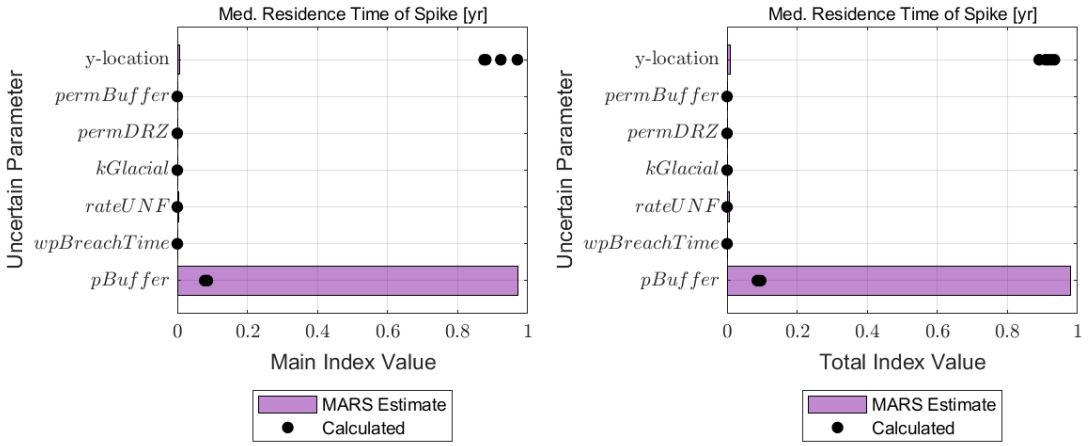


Figure 3-12. Sensitivity analysis results for the median residence time [yr] of the spike in the repository from the y-location analysis; calculated values plotted with y-location are the calculated values for the float seed

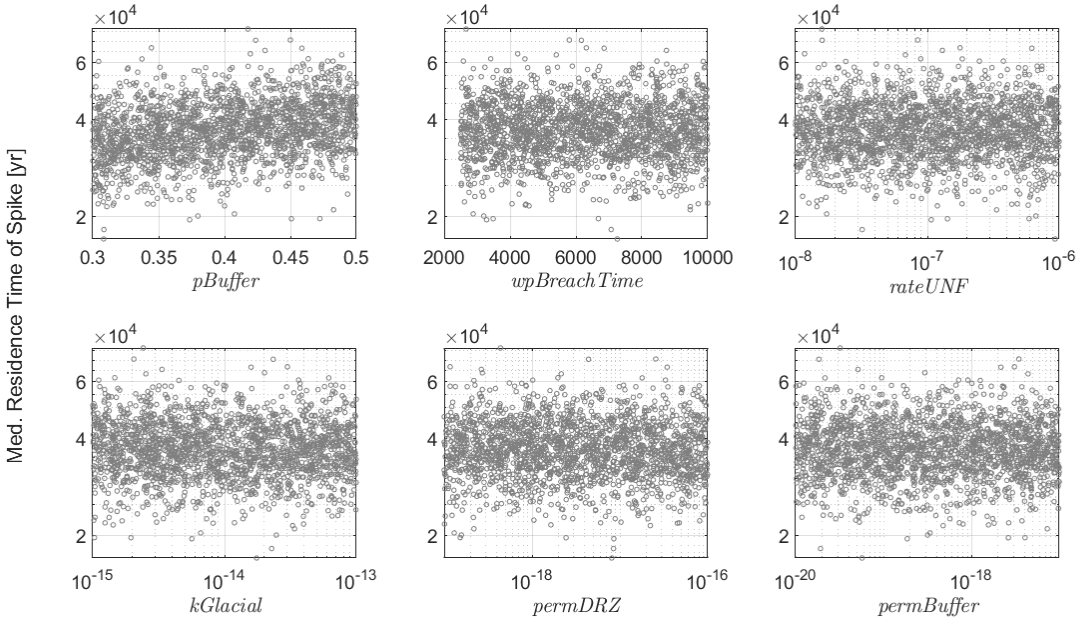


Figure 3-13. Scatterplots of the median residence time [yr] versus the uncertain parameters

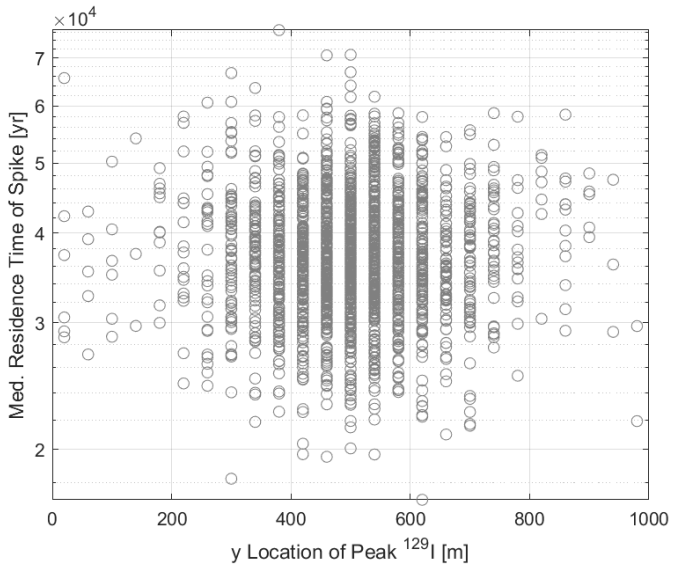


Figure 3-14. Scatterplot of the y-location [m] versus median residence time [yr]

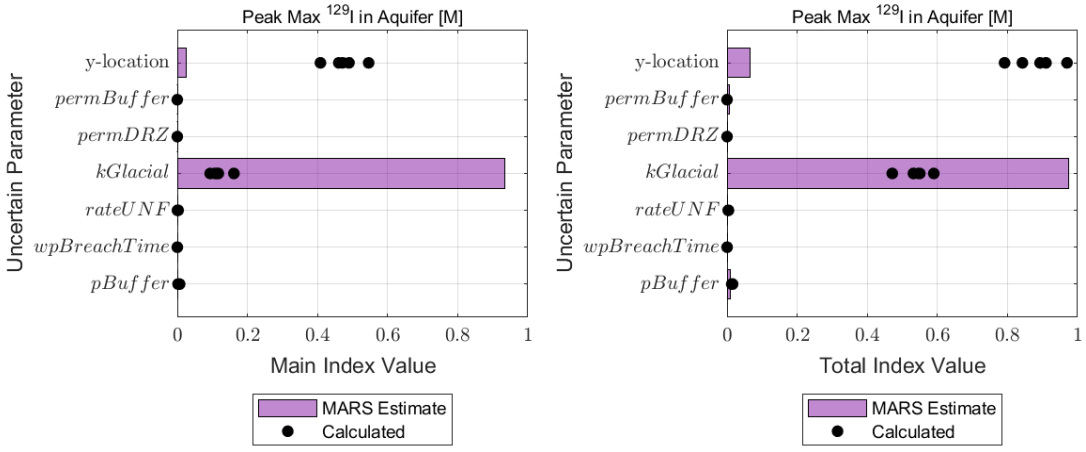


Figure 3-15. Sensitivity analysis results for peak ^{129}I concentration [M] from the y-location analysis; calculated values plotted with y-location are the calculated values for the float seed

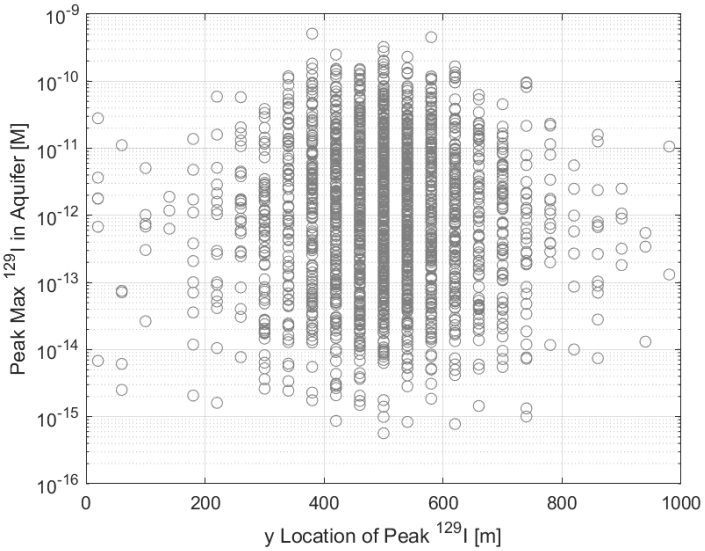


Figure 3-16. Scatterplot of the peak ^{129}I concentration [M] plotted versus the y-location [m] of the peak

3.3.3 Float Seed Analysis

The float seed analysis was performed to test whether the surrogate model could attribute spatial variability to, essentially, a noise variable. In the Sobol' index calculation (Section 3.1), *float_seed* is used just for indexing; the actual values of *float_seed* are not incorporated in the Sobol' index calculation. Because the float seed value is not used directly in the repository simulation, the surrogate models were not expected to attribute any variance to the seed. To perform this analysis, *float_seed* was included as a parameter in the surrogate model construction and subsequent Sobol index estimation.

Sensitivity analysis results from the float seed analysis for the time of peak ^{129}I concentration [M] are plotted in Figure 3-17. The results for the ratio between the rock to aquifer flux and the aquifer to east boundary flux are plotted in Figure 3-18. In both cases, the calculated indices attribute a large proportion of the variance to *float_seed*. The surrogate model analysis attributes none of the variance to the float seed for the time of peak but does attribute some to *float_seed* for the flux ratio QoI; the total effect index for the float seed in this case is relatively high. It is not clear why the surrogate behaves slightly differently in each case; however, the surrogate overestimates the main and total effect indices for the other variables so this may be due to random variation in overfitting behavior.

Sensitivity analysis results from the float seed analysis for the ratio between the rock to aquifer flux and the rock to east boundary flux at 100,000 years is plotted in Figure 3-19. This is an interesting case of overfitting; the total effect index for *float_seed* is quite high, but so are the total effect indices for some of the epistemic variables, such as *permDRZ*, which was calculated to have a total effect index of zero.

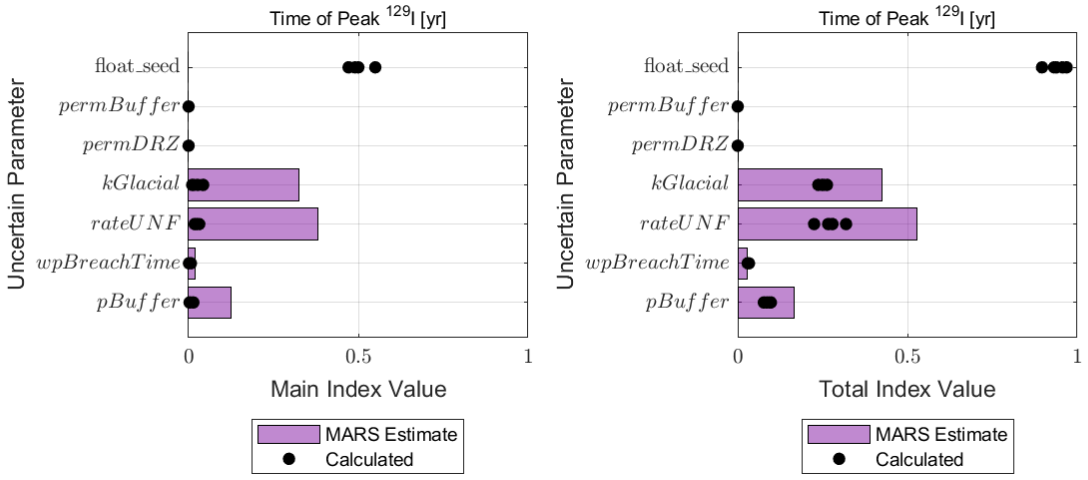


Figure 3-17. Sensitivity analysis results for the time [yr] of peak ^{129}I concentration from the float seed analysis

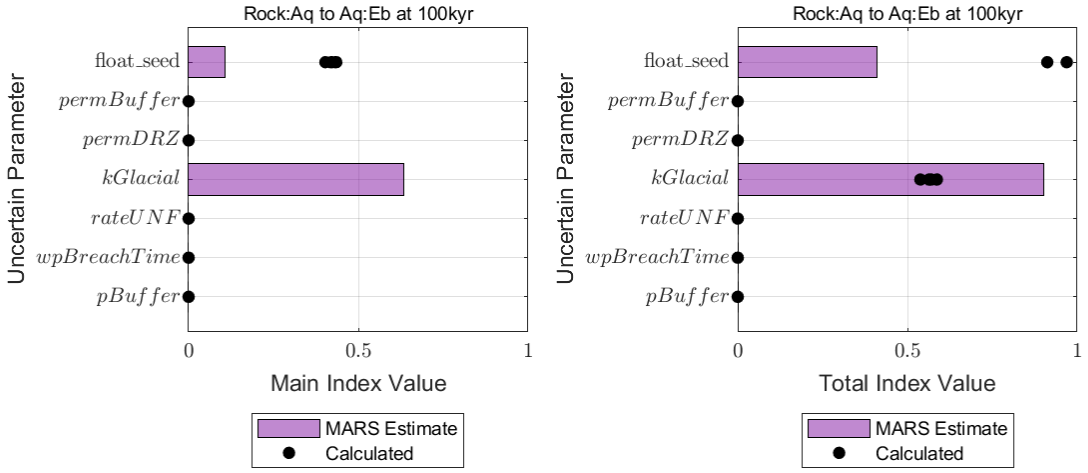


Figure 3-18. Sensitivity analysis results for the ratio between rock to aquifer flux and aquifer to east boundary flux at 100,000 years from the float seed analysis

The float seed analysis was also performed with the smaller initial sample size (see Section 3.1) of $2N_{samples} = 2{,}000$. The sensitivity results from this analysis for the y-location are plotted in Figure 3-20. When 10,000 samples were used, however, all of the main and total effect indices were estimated to be zero. This example shows what overfitting can look like; there may be suspicion of overfitting when most variables have large total effects indices and small main effects indices. However, a larger sample size resolved the overfitting issue in this case.

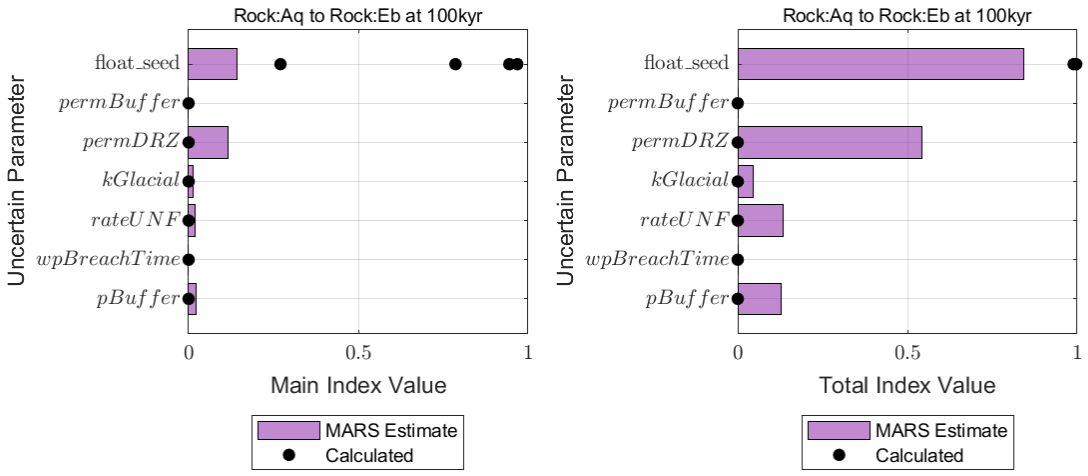


Figure 3-19. Sensitivity analysis results for the ratio between the rock to aquifer flux and the rock to east boundary flux from the float seed analysis

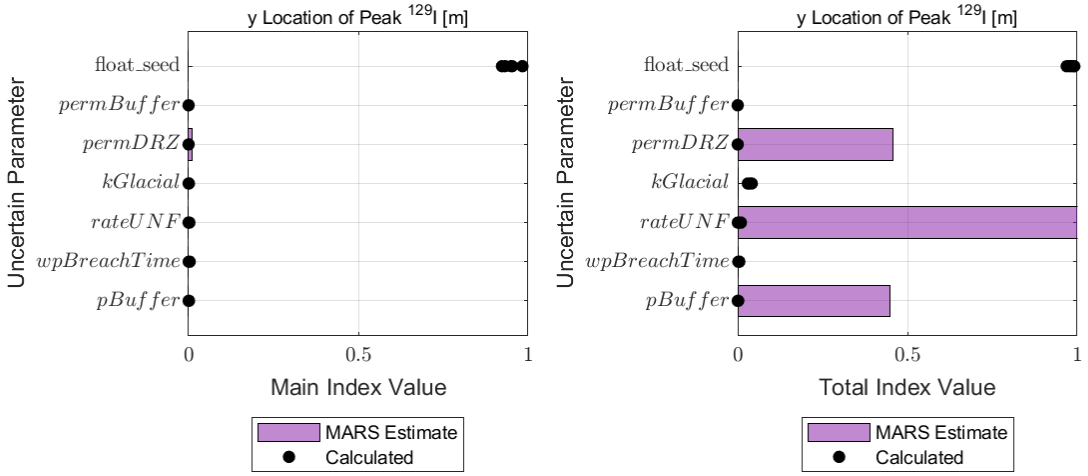


Figure 3-20. Sensitivity analysis results for the y-location [m] of the peak ^{129}I concentration in the aquifer from the float seed analysis with a smaller sample size (2,000)

3.3.4 Graph Metric Analysis

The y-location quantity did not prove to be a useful proxy for spatial heterogeneity (Section 3.3.2) and the float seed does not have a meaningful relationship to QoIs that can be modeled by a surrogate (Section 3.3.3). So far, the only method we have found successful for incorporating spatial heterogeneity in the sensitivity analysis is through the use of graph metrics [1, 2]. These previous analyses have shown that there are meaningful relationships between the graph metrics and many QoIs, and the surrogate models are able to detect this influence. However, it has never been feasible to calculate the Sobol' indices without use of a surrogate for the full-scale crystalline reference case. This comparison is possible for the simplified crystalline reference case.

The graph metric analysis compares the Sobol' indices estimated for the graph metrics to the Sobol' indices calculated for *float_seed* (see Section 3.1). This is not a direct comparison because there are three separate graph metrics and only one float seed. For plots, the float seed Sobol'

indices are repeated so they can be plotted against each of the graph metrics. It would not make sense to plot the float seed Sobol' indices versus the sum of the graph metric indices because the interactions included in the total effects indices would be double-counted. For example, an interaction effect between STT and nIntersections would be counted in the total effect index for both the STT and nIntersections. Thus, the comparison is somewhat qualitative.

Sensitivity analysis results for the y-location of peak ^{129}I are plotted in Figure 3-21. In this case, the surrogate model still misattributes a significant portion of the variance to some of the parameter uncertainties via interactions (between *rateUNF*, *wpBreachTime*, and *pBuffer*). However, it assigns significant variance to the STT and nIntersections graph metrics. Even if we were to add the total effects indices for these graph metrics (which would be an overestimate of their combined effect if they interact with each other), the total effect index estimate would still be significantly lower than the calculated value for spatial heterogeneity, which is repeated in the plot for each graph metric. The scatterplots for this comparison are shown in Figure 3-22; the trend for STT is slightly more pronounced than the trends for NIwR and nIntersections, so the surrogate accurately attributes most of the effect from spatial heterogeneity to the correct graph metric.

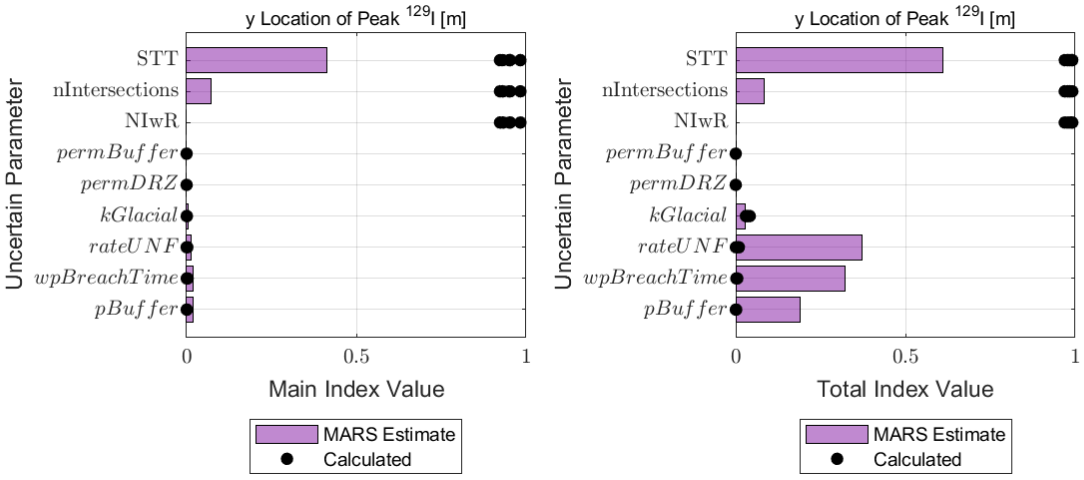


Figure 3-21. Sensitivity results for the y-location [m] of the peak ^{129}I concentration from the graph metric analysis; the calculated indices for *float_seed* are repeated in the plots for each graph metric

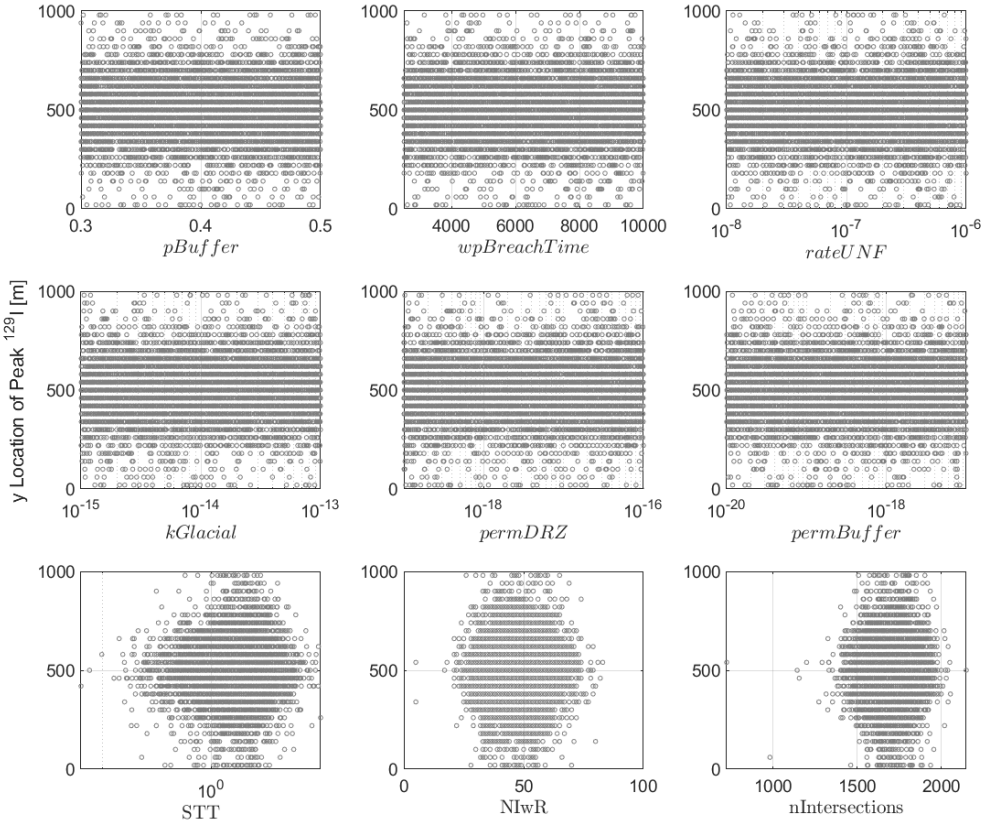


Figure 3-22. Scatter plots of the y-location [m] of peak ^{129}I versus all of the uncertain parameters and the graph metrics

The sensitivity analysis results for the ratio between the aquifer to east boundary flux and the rock to east boundary flux at 100,000 years from the graph metric analysis are plotted in Figure 3-23 with corresponding scatterplots in Figure 3-24. This is an interesting case because there is a

very clear trend with *kGlacial* which the surrogate correctly detects, however the surrogate does not attribute significant variance to the graph metrics. This may be due to the dominance of *kGlacial*, but the scatterplots suggest that the graph metrics do not characterize the effect of spatial heterogeneity on this QoI.

Sensitivity analysis results for the ratio between the rock to aquifer flux and the aquifer to east boundary flux are plotted in Figure 3-25 with scatterplots in Figure 3-26. For this QoI, the calculated Sobol' indices show that spatial heterogeneity is the dominant uncertainty, but that *kGlacial* also has high significance based on its interactions; the main effect index calculated for *kGlacial* is low, but the total effect index is high. The surrogate model in this case matches the relative ranking of the uncertainties from the calculated indices. The surrogate model also ranks the spatial heterogeneity as dominant, but still attributes significant variability to the glacial till permeability (*kGlacial*). The trends identified by these indices can be seen in Figure 3-26; variation in the flux ratio decreases for very low/high values of all of the graph metrics and decreases consistently with *kGlacial*.

Finally, sensitivity analysis results and corresponding scatterplots for the peak ^{129}I concentration in the aquifer are plotted in Figure 3-27 and Figure 3-28, respectively. Surrogate model quality for this QoI is a priority because it is our primary QoI for evaluating repository performance at this stage in our study of the full-scale crystalline reference case. Unfortunately, the surrogate model failed to detect the importance of the graph metrics in this analysis, even though some minor trends are evident in the scatterplots. The trend between peak ^{129}I is clearly linear (in log-log space), but with substantial variation around that linear trend. Therefore, it makes sense that interaction effects would be significant and that there would be some interaction effects with *kGlacial*.

Figure 3-29 shows interaction scatterplots for the graph metrics with *kGlacial*. Each plot frame shows the scatterplot for *kGlacial* versus the peak ^{129}I concentration, colored by one of the graph metrics. Despite the substantial interaction effects suggested by the strong calculated total effect indices for *float_seed* (plotted with the graph metrics) and *kGlacial* in Figure 3-27, the interaction scatter plots do not have any distinctive patterns with respect to color. This suggests that there are underlying interactions between the spatial heterogeneity and the glacial till permeability that are not captured by the graph metrics. The surrogate in this case is not necessarily a poor fit; it is limited by the extent to which the graph metrics reflect the effects of spatial heterogeneity. The poor predictions for the Sobol' indices from the surrogate for this QoI reflect a need to refine the graph metrics of investigate alternative approaches for assessing the effects of spatial heterogeneity.

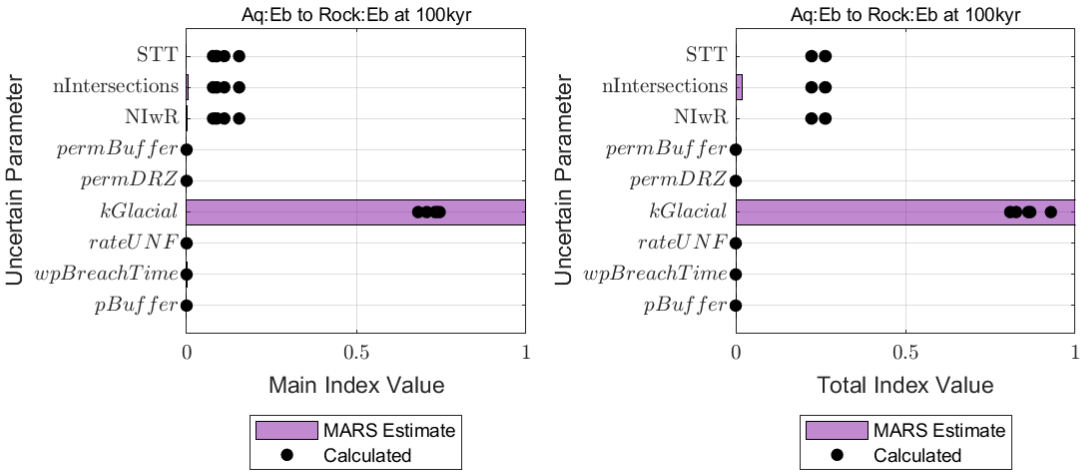


Figure 3-23. Sensitivity results for the ratio between the aquifer to east boundary flux and the rock to east boundary flux at 100,000 years from the graph metric analysis; the calculated indices for *float_seed* are repeated in the plots for each graph metric

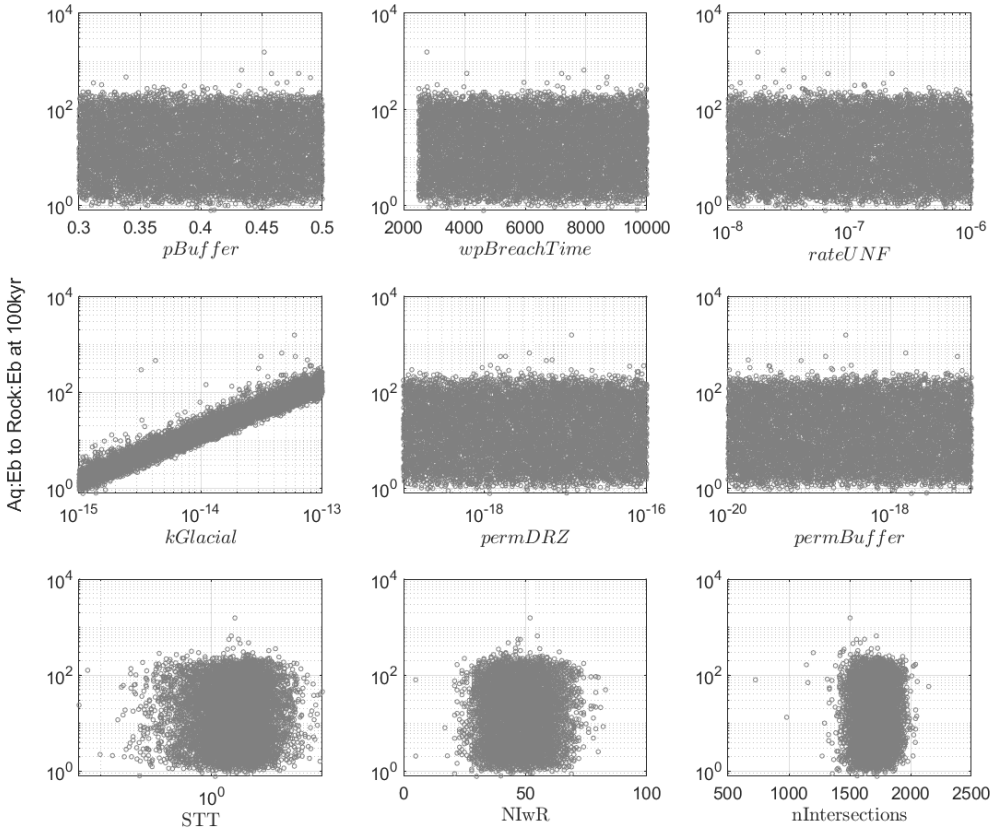


Figure 3-24. Scatter plots of the ratio between the aquifer to east boundary flux and the rock to east boundary flux at 100,000 years versus the uncertain parameters and the graph metrics

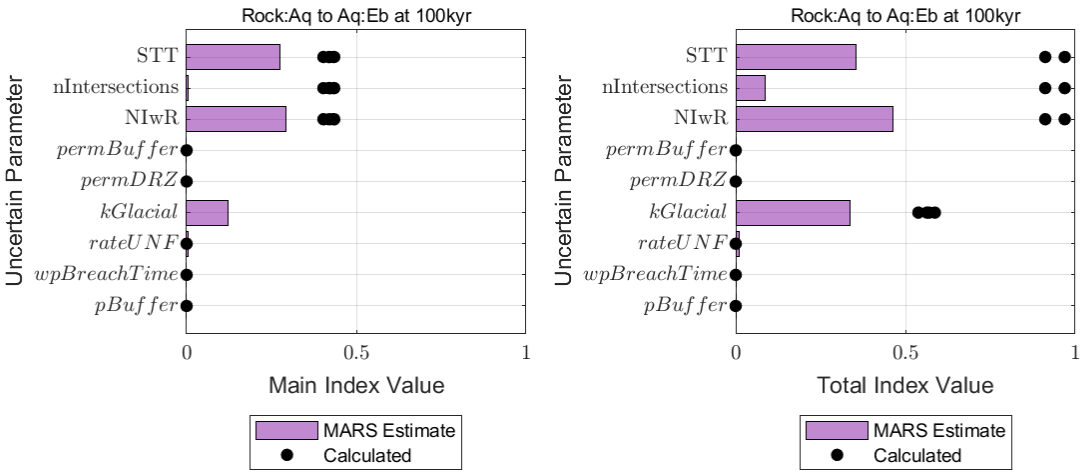


Figure 3-25. Sensitivity results for the ratio between the rock to aquifer flux and the aquifer to east boundary flux at 100,000 years from the graph metric analysis; the calculated indices for *float_seed* are repeated in the plots for each graph metric

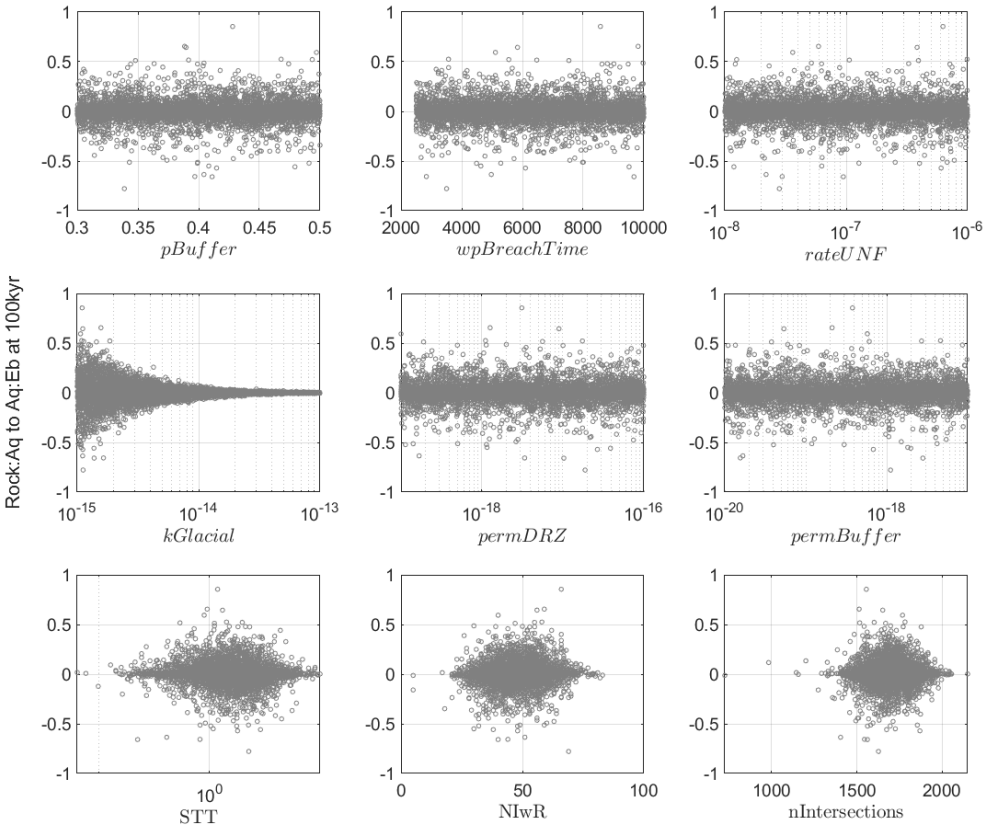


Figure 3-26. Scatterplots of the ratio between the rock to aquifer flux and the aquifer to east boundary flux at 100,000 years versus the uncertain parameters

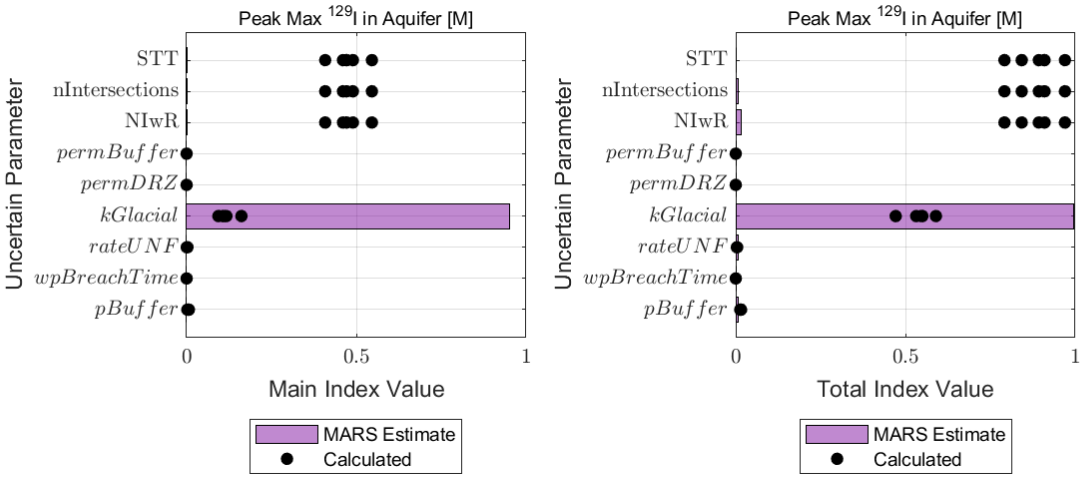


Figure 3-27. Sensitivity results for the peak ^{129}I concentration in the aquifer [M] from the graph metric analysis; the calculated indices for *float_seed* are repeated in the plots for each graph metric

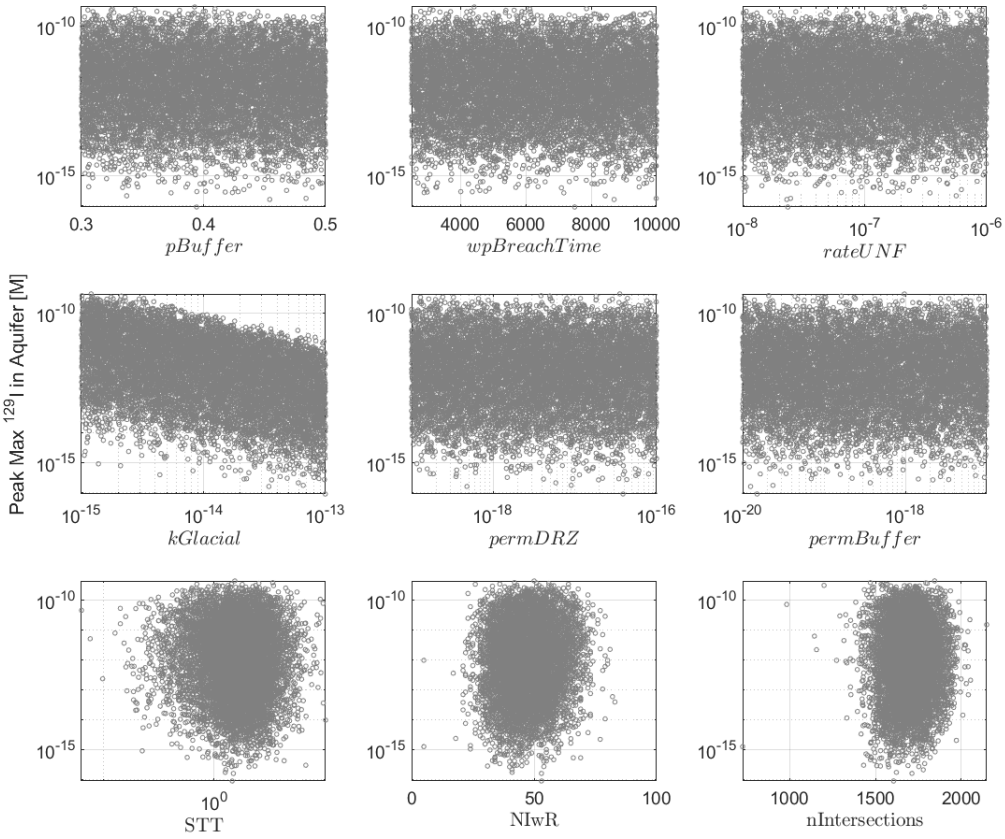


Figure 3-28. Scatterplots of the peak ^{129}I concentration in the aquifer [M] versus the uncertain parameters and graph metrics

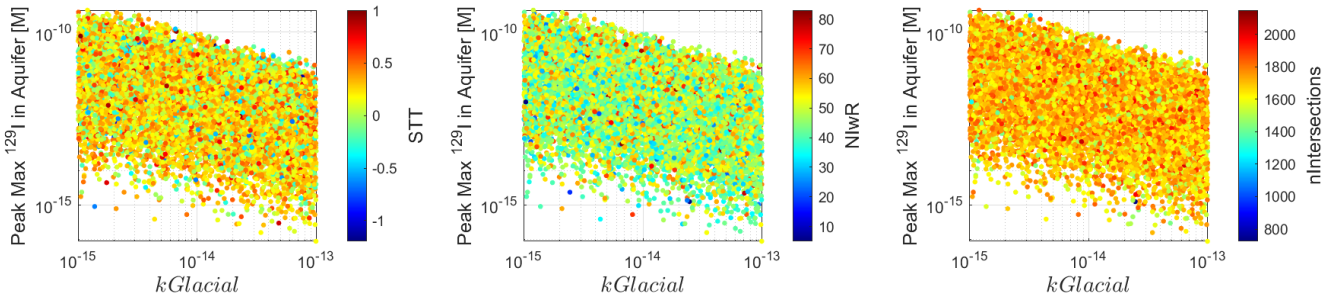


Figure 3-29. Interaction scatterplots for the graph metrics and *kGlacial* for peak ^{129}I concentration in the aquifer [M]

Overall, the graph metrics in this analysis showed utility in characterizing the spatial heterogeneity and its effect on QoIs for some, but not all, of the QoIs. There are important relationships that cannot be reflected in the surrogate because they are not reflected in the graph metrics. This is the case for ^{129}I, which is sensitive to spatial heterogeneity, for which the graph metrics do not reflect a significant interaction between spatial heterogeneity and glacial till permeability. Because of the importance of ^{129}I for understanding repository performance in the full-scale reference case, this motivates further development of the graph metrics or alternative methods to better detail the effects of spatial heterogeneity.

When calculating a graph metric, all of the spatial heterogeneity is reduced to a single value for that realization, resulting in a reduction of information. We use multiple graph metrics to try and include information that is likely important, but even multiple graph metrics cannot characterize all of the important features of the DFNs. For the full-scale crystalline model, even more information is lost because the graph metrics do not describe any of the spatial heterogeneity introduced by the assignment of individual waste package degradation rates to each canister. As a result, we cannot expect graph metrics to ever perfectly summarize spatial heterogeneity, and it may be the case that they cannot sufficiently summarize the effects of spatial heterogeneity to truly understand the relative importance between spatial and parametric uncertainties.

Alternatives to the graph metrics should be considered for sensitivity analysis. The simplified crystalline reference cased used in this study made it feasible to calculate Sobol' indices with a large number of model evaluations, which reduces sample-size uncertainty in the calculated Sobol' indices. However, the same method could still be used with the full-scale model to calculate Sobol' indices for the spatial heterogeneity with a much smaller number of model evaluations. This would result in more accurate estimates of the relative importance of spatial heterogeneity versus parameter uncertainty, but with much lower precision than could be accomplished with the simplified case. If the precision is too low, however, then the increased representativeness of the Sobol' indices for spatial heterogeneity would not be useful; this would essentially just trade the surrogate model uncertainty that currently interferes with interpretation for sample-size uncertainty, which would interfere with interpretation just as much. Whether Sobol' index calculation is a viable alternative for the full-scale crystalline case will depend on what precision can be accomplished with the relatively small number of model evaluations that are feasible for the full-scale model.

3.3.5 Transformation Analysis

Transformations can also be considered as a tool for improving surrogate modeling. Scaling and log transformations have been applied for ^{129}I concentrations in previous GDSA Framework sensitivity analysis with the shale reference case [3, 5] and unsaturated alluvium reference case [4]. Log transformations can be useful for reducing differences in scale between input variances and output variances. Similarly, the scaling applied to the alluvium reference case [4] was necessary because the ^{129}I concentrations were so low that substantial variation still resulted in negligible variance numerically. Scaling the concentrations by a constant factor enabled sensitivity analysis to identify the input variables driving variation in ^{129}I despite the low overall values for the concentration.

This analysis investigates two potential transformations for ^{129}I concentration to see if they improve the ability of the surrogate to estimate the Sobol' indices. Graph metrics were included in this analysis because they are, to date, the best method we have for characterizing the spatial heterogeneity. The first transformation we applied is a rank transformation. In this case, it was applied to all of the input variables, the graph metrics, and the ^{129}I concentration. This kind of transformation can be useful for identifying non-linear and non-monotonic relationships and removes any concerns about the variables being on different scales. The sensitivity analysis results from this transformation are plotted in Figure 3-30. The results are strikingly similar to those in Figure 3-27; the rank transformation did not improve the ability of the surrogate to account for the variance. As discussed in the previous section, the graph metrics may not reflect all of the important impacts of the spatial heterogeneity. This is not a problem we would expect the rank transformation to solve, but the relatively unchanged results support our conclusion that spatial heterogeneity is the dominant challenge for surrogate modeling with this QoI.

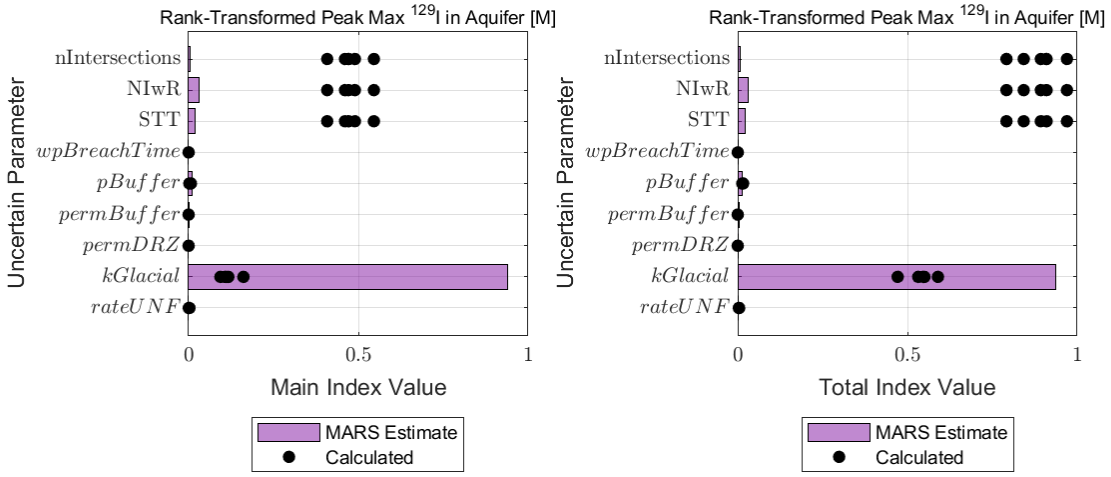


Figure 3-30. Sensitivity analysis results for the peak ^{129}I concentration [M] in the aquifer from the transformation analysis, with rank transformation applied to all inputs and the QoI; the calculated indices for *float_seed* are repeated in the plots for each graph metric

The second transformation we applied to the peak ^{129}I concentration was a $\log_{10}$ transformation. This transformation was only applied to the QoI, and the results are plotted in Figure 3-31. As in the previous case, this transformation resulted in no appreciable change to the

quality of the surrogate model; the sensitivity analysis results are very close and the conclusions unchanged.

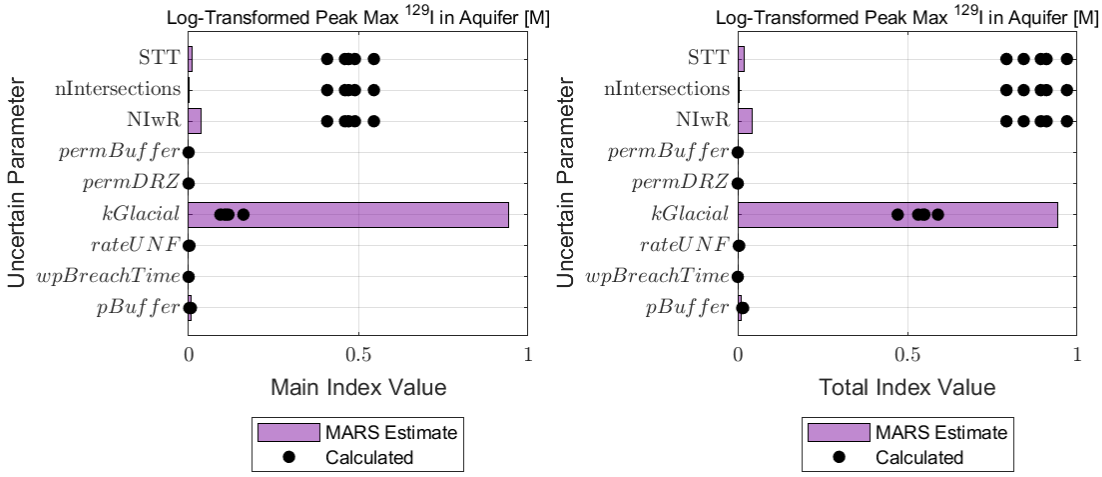


Figure 3-31. Sensitivity analysis results for the peak ^{129}I concentration [M] in the aquifer from the transformation analysis, with a log10 transformation applied to the QoI; the calculated indices for *float_seed* are repeated in the plots for each graph metric

This transformation analysis was limited to the peak ^{129}I concentration because this QoI is a priority in our performance assessment analysis and presents a challenge for sensitivity analysis. Transformations did not provide any benefit for this QoI because the central challenge to estimating Sobol' indices is the non-parametric definition of spatial heterogeneity for the crystalline reference case. This does not mean that transformations could not be useful in the future; transformations may be able to improve the surrogate modeling for the peak ^{129}I concentration once we are able to incorporate all of the effects from spatial heterogeneity.

3.3.6 Replicate Analysis

The consistency of our surrogate modeling is another concern for the full-scale crystalline reference case. Because the model is computationally expensive, we have not previously been able to perform a replicate analysis to understand how much surrogate model results differ between replicates. This is possible, however, with the simplified crystalline case.

For this analysis, we used all five replicates with sample size 10,000 from Section 3.1. The replicates are the same size but have different independent samples from one another. The replicate analysis was performed by fitting the MARS surrogate to each replicate independently and using each surrogate to estimate the Sobol' indices. We do not expect the surrogates for each replicate to give identical estimates of the Sobol' indices because even the calculations are not identical for each replicate. However, a high-quality surrogate should show about the same amount of variation between Sobol' index estimates as we see between the calculated values.

The replicate analysis results for the ratio between the rock to aquifer flux and the aquifer to east boundary flux at 100,000 years are plotted in Figure 3-32. These results show significant variation between the surrogates with respect to the Sobol' index estimates for graph metrics.

There is somewhat less variation and reasonable agreement to the calculated indices for the total effect indices for *kGlacial* for replicates 1,2,3, and 5. However, replicate 4 estimates a substantially larger total effect index for *kGlacial*.

The scatterplots for this QoI (Figure 3-26) may hint at why the surrogate model differs so much between replicates. There is substantial symmetry around the abscissa in the plot against *kGlacial* and this type of symmetry presents a challenge for a surrogate of any form.

The replicate analysis results for the ratio between the rock to aquifer flux and the rock to east boundary flux at 100,000 years are plotted in Figure 3-33. The behavior here is similar to the previous flux ratio QoI, except that we see many more significant total effects indices when the calculated total effects index is zero. This type of result suggests some overfitting, so these results would still be viewed suspiciously in an analysis that did not include replicates [1, 2]. There does appear to be agreement, however, between the surrogates and the calculations in attributing significant variance to spatial heterogeneity. There is also large variation in the calculated main effects indices for the *float_seed* (plotted with the graph metrics), so the variation between replicates in the main effect Sobol' indices for the graph metrics is reasonable.

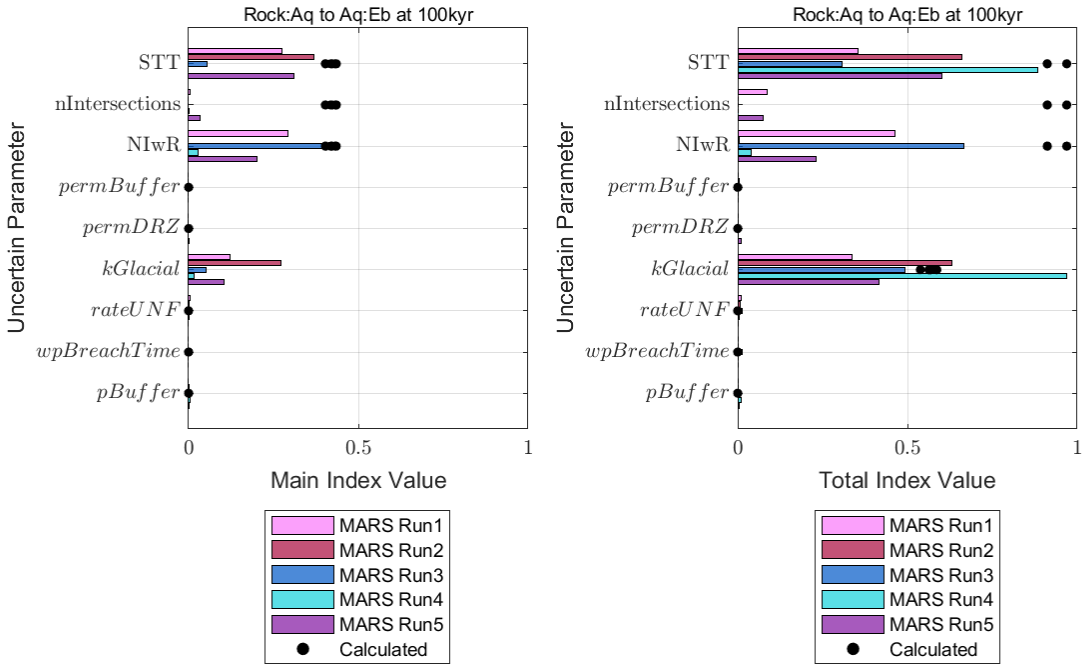


Figure 3-32. Sensitivity analysis results for the ratio between the rock to aquifer flux and the aquifer to east boundary flux from the replicate analysis; the calculated indices for *float_seed* are repeated in the plots for each graph metric

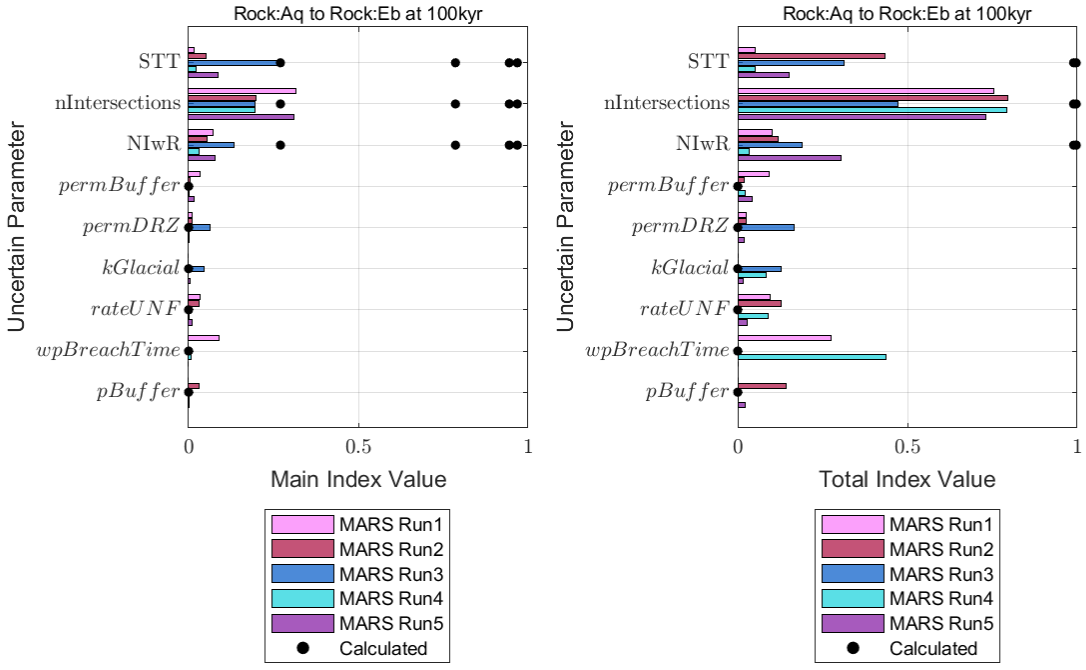


Figure 3-33. Sensitivity analysis results for the ratio between the rock to aquifer flux and the rock to east boundary flux from the replicate analysis; the calculated indices for *float_seed* are repeated in the plots for each graph metric

Results from the replicate analysis for the median residence time are plotted in Figure 3-34. The results from these replicates are very consistent with each other, both between the estimates from the surrogates and between the calculated indices. These results suggest that the graph metrics characterize some of the significant effects from spatial heterogeneity for this QoI, which can be reliably detected by the MARS surrogate. The graph metrics are not sufficient to characterize all of the effects from spatial heterogeneity, however, which is why the indices for the graph metrics are still substantially underestimated and the indices for *kGlacial* are consistently overestimated.

The replicate results for the time of peak ^{129}I concentration are plotted in Figure 3-35. These results are similar to the previous QoI in that they identify some importance of the spatial heterogeneity consistently, but also consistently underestimate its importance relative to the other uncertainties. Here, however, there is very good agreement across all replicates for the total effects indices for *pBuffer* and *wpBreachTime* and good agreement across all replicates for the total effects indices for *kGlacial* and *rateUNF*. The ranking of relative importance between these input variables would be roughly the same, regardless of which calculation was used or which surrogate replicate was used to estimate Sobol' indices. The weakness for the surrogate across all replicates for this QoI is due to the incompleteness in the graph metric characterization of spatial heterogeneity; this is another sensitivity analysis that may be improved in the future by advances to our treatment of the spatial heterogeneity.

Replicate analysis results for the peak ^{129}I concentration analysis are shown in Figure 3-36. The results are consistent with what we saw for this QoI in the other analyses from this chapter. Regardless of which tactic we use, the surrogate model for peak ^{129}I concentration fails to accurately attribute variance to the spatial heterogeneity, and this is consistent across replicates. This highlights the need to advance our treatment of spatial heterogeneity for sensitivity analysis in support of performance assessment.

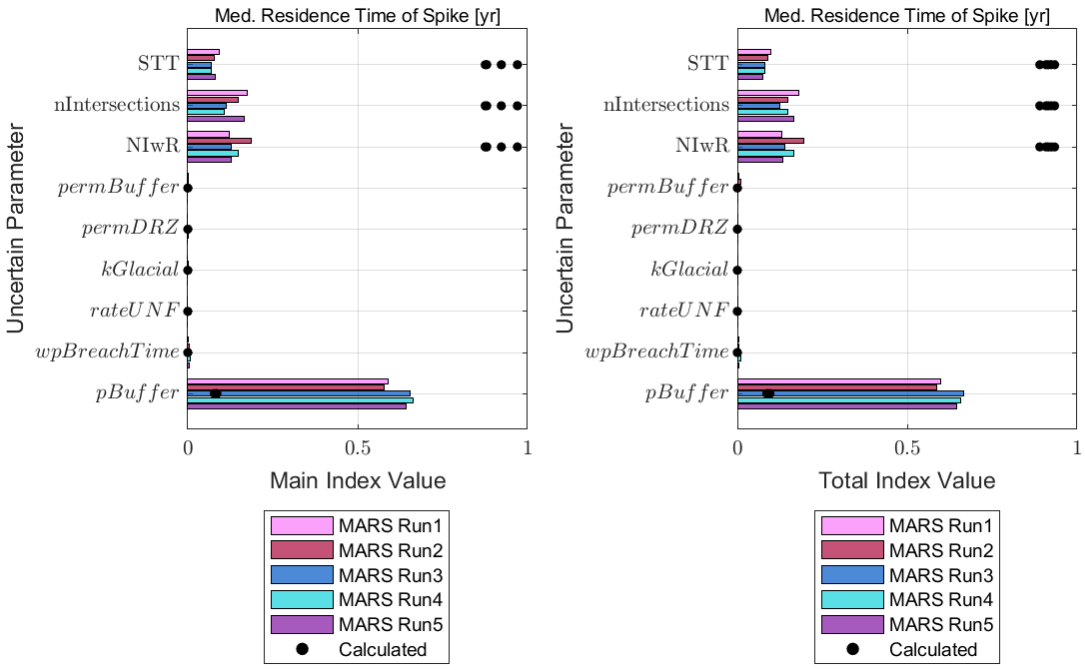


Figure 3-34. Sensitivity analysis results for the median residence time [yr] from the replicate analysis; the calculated indices for *float_seed* are repeated in the plots for each graph metric

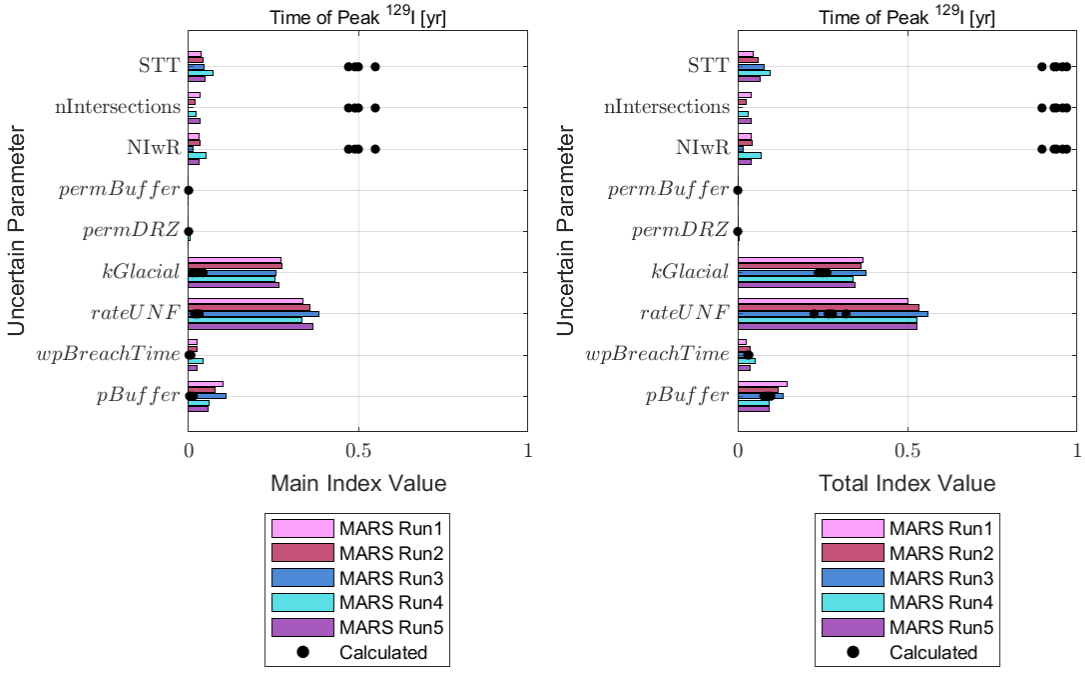


Figure 3-35. Sensitivity analysis results for the time [yr] of peak ^{129}I concentration from the replicate analysis; the calculated indices for *float_seed* are repeated in the plots for each graph metric

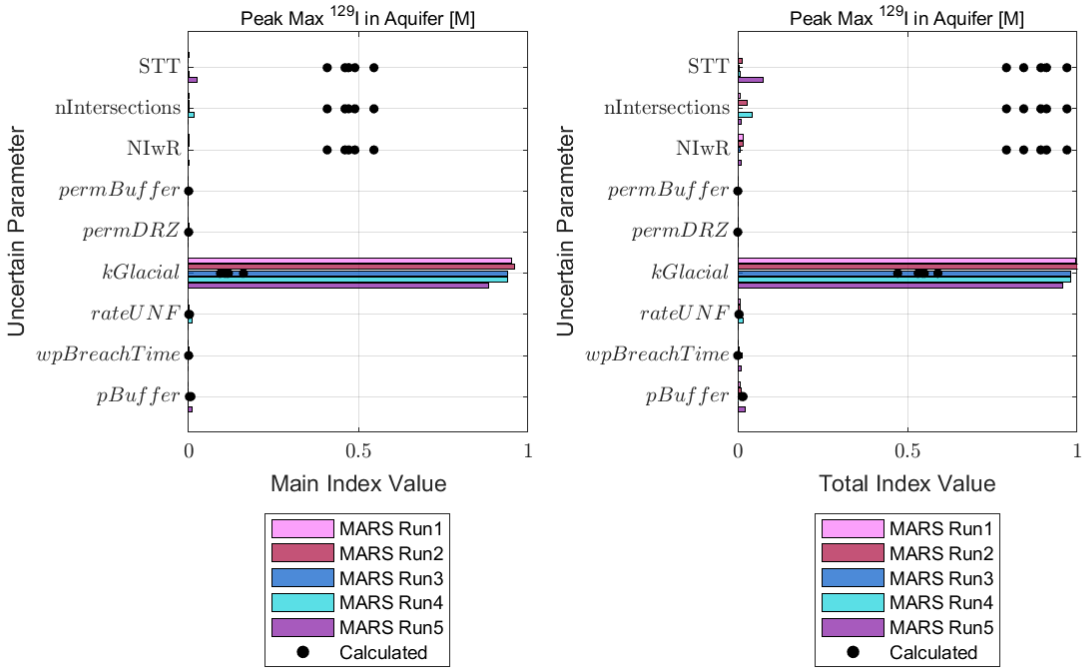


Figure 3-36. Sensitivity analysis results for the peak ^{129}I concentration [M] in the aquifer from the replicate analysis; the calculated indices for *float_seed* are repeated in the plots for each graph metric

3.4 Discussion

The analyses discussed in this chapter showed that the surrogate choice has some effect on sensitivity analysis results, but it is minimal for most QoIs. All surrogate models included in these studies demonstrated overfitting behavior. Our analysis showed that y-location of peak ^{129}I is not a viable proxy for spatial heterogeneity in sensitivity analysis; while y-location does not depend significantly on the uncertain input variables, it still does not reflect the significant impacts from spatial heterogeneity. Graph metrics still proved to be our best tool to date for improving surrogate model performance but are demonstrably incomplete tools. Improvement of our surrogate modeling for peak ^{129}I concentration was evasive, but these studies in combination show that improvement in our treatment of spatial heterogeneity is the most promising avenue for accomplishing better surrogate models.

The biggest gap affecting surrogate quality is our treatment of spatial heterogeneity. We use surrogate modeling to reduce the computational cost of sensitivity analysis, but this requires that the spatial heterogeneity is characterized parametrically. Such parametric characterizations necessarily lose information, and the analyses in this section show that some of the information lost is important for understanding the relative importance of spatial and parameter uncertainties. We may be able to refine our graph metric methodology to include more information about the spatial heterogeneity that influences QoIs, but graph metrics may not be sufficient for this type of analysis.

Another option would be to calculate the Sobol' indices with the full-scale model using a structured sampling approach as discussed in Section 3.1, but with a much smaller sample size due to the high computational cost of full-scale model simulations. This would prevent the loss of information that occurs when converting the spatial heterogeneity to graph metrics but would incorporate substantial uncertainty due to the limited sample size. It is not clear yet if this strategy would be precise enough to improve our understanding of how spatial heterogeneity influences uncertainty in performance measures. Multifidelity sampling-based methods for Sobol' indices such as that in [21] can improve the precision that can be achieved with relatively few full-scale, high-fidelity model evaluations. This will be a subject of future investigation.

This year, analysis focused on the challenges associated with unparameterized spatial heterogeneity; future analyses will focus on the effect that limited samples and nested sampling structure has on surrogate accuracy.

4. DISCRETE FRACTURE NETWORKS

The past two years, we have studied the effect of discrete fracture networks on the uncertainty and sensitivity of key quantities of interest (QoIs) for the crystalline reference case [1, 2]. We continue our investigation this year, focusing on the transmissivity relationship and its effect on permeability fields and QoI. Some of the results in this chapter are taken from a paper submitted to the Intermediate High-Level Radioactive Waste Management conference in Nov. 2022, [27].

Discrete fracture network (DFN) modeling has become the alternative approach to continuum approaches for simulating flow and transport through sparsely fractured rocks in the subsurface [28]. 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. In this work, DFNs are generated using dfnWorks, a parallelized computational suite developed by Los Alamos National Laboratory [28]. The DFNs are then mapped to an equivalent continuous porous medium (ECPM) for flow and transport simulation with PFLOTRAN [12].

The previous two GDSA milestone reports on uncertainty and sensitivity analysis [1, 2] contained extensive studies of the DFN properties for a crystalline repository reference case with host rock properties comparable to the Forsmark site in Sweden [23]. Chapter 3 of the 2020 GDSA SA/UQ report [2] verified the implementation of DFNs to ensure the expected number of fractures and fracture intensity values obtained were consistent with the Forsmark specification. The conversion of the discrete fracture networks to ECPMs then led to a comparison between the ECPM bulk properties from an analysis of Forsmark provided in the literature vs. the ECPM bulk properties for the crystalline reference case calculated from the DFN conversion. The log of the hydraulic conductivity field at each depth zone and for each direction were compared and found consistent.

Chapter 3 of the 2021 GDSA SA/UQ report [1] extended the previous analysis in two ways: it focused on a comparison of the transmissivity relationships defining the DFNs and it included graph metrics in the analysis of the DFNs. The fracture transmissivity is used to determine the continuum permeability field of the ECPM, as described in [29]. The initial implementation in dfnWorks assumed a single “correlated” transmissivity relationship for the entire computational domain. However, new parameterizations were provided for the correlated transmissivity relationship based on depth [23] and new capabilities in dfnWorks enabled the use of the depth-dependent correlated relationship for the crystalline reference case. Chapter 3 of the 2021 report showed that although the correlated constant vs. correlated depth-dependent transmissivity relationships resulted in significantly different mean permeability fields in each depth zone, many of the Quantities of Interest such as peak total ^{129}I concentration were not significantly different across the two transmissivity relationships. This comparison was based on statistical testing of mean differences using 20 DFNs for each type of transmissivity.

This year, we expanded the study, generating and analyzing 100 discrete fracture networks for each transmissivity relationship. The concern was that 20 DFNs were not enough to tease out some of the differences in QoIs such as peak ^{129}I or some of the fluxes. Having a larger sample size results in a tighter estimate of the confidence interval around the mean, resulting in greater ability

to distinguish statistically significant differences. Additionally, this year we tracked many QoIs as a function of time. These time-dependent outputs are presented in the Results section below.

4.1 Approach

As mentioned, in the 2021 study [1], it was found that only one quantity of interest showed a statistically significant difference between the two transmissivity relationships. It was therefore concluded that the type of transmissivity relationship did not have a significant impact on repository performance characteristic for the sample of DFNs used. However, since the original study only sampled 20 DFNs in total, there was interest in if a larger sample set might yield different results. Therefore, the analyses described here utilize a new set of 100 randomly generated DFNs. Note that the transmissivity relationship is the only thing that varies in these data sets: there are 100 DFNs with the correlated constant transmissivity and 100 DFNs with the correlated depth-dependent transmissivity. All other parameters are held constant.

The same modeling and data extraction process used for the previous study was used in this analysis and the same QoIs (listed below) were assessed to observe if there were any changes to the initial conclusions. For more detail about the QoIs, specifically the tracer implementation, see Section 5.2.6 and Table 5-2 of this report. For a figure depicting the repository studied here, see Figure 5-5. The QoIs for this study are:

- Maximum ^{129}I concentration in the aquifer
- Median residence time (MdRT) of a tracer in the repository in years
- Fraction of tracer still in the repository at 1 million years
- Fractional mass flux of tracer from the repository at 3 thousand years
- Ratio between the mass flow rate of water from the aquifer to east boundary and the mass flow rate from the rock to east boundary
- Ratio between the mass flow rate of water from the rock to aquifer and the mass flow rate of water from the rock to east boundary

This study also compares QoIs to graph metrics. Graph metrics reflect the topology of the network using a set of nodes connected by edges which can potentially be useful for a comparison such as the one completed in this study since flow and transport is strongly channeled through the fracture network. The graph metrics used in this study are constructed using dfnWork's dfnGraph utility and specific graph metrics were extracted using dfnWorks and NetworkX [2]. The graph metrics are described in more detail in Chapter 3.2 of [1]. They include metrics such as:

- Average degree (average number of intersections a fracture is part of)
- Density
- Length of shortest path
- Number of intersections with repository
- Number of edges (intersections)
- Number of nodes (fractures)
- Shortest travel time between the repository and aquifer

4.2 Results

The information shown in the subsequent tables and plots is representative of all 200 cases, 100 for the correlated constant relationship and 100 for the correlated depth-dependent relationship: outliers were not removed. Interval plots were used to examine the mean values for each type of data and the associated 95% confidence interval on the means computed over the 100 samples for each relationship. Lack of overlap in the confidence intervals indicates the correlated depth-dependent transmissivity relationship influenced the results significantly.

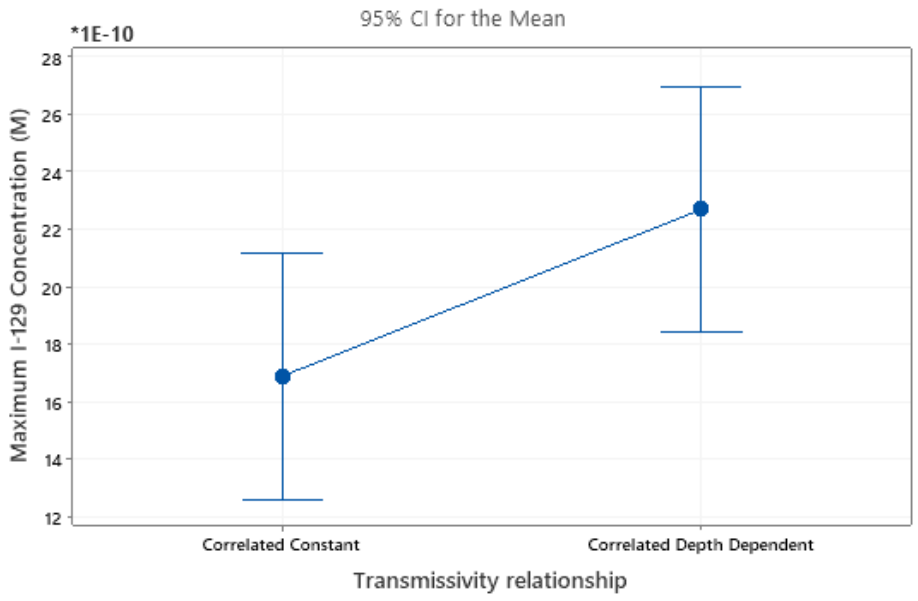
To compare the correlation between QoI and graph metric, scatterplots and a Pearson correlation was used to compare the level of correlation between a specific QoI and graph metric for the both the correlated constant and correlated depth-dependent cases. The Pearson correlation is simply a measure of linear correlation between two sets of data. The higher the *r* value, the stronger the correlation. An *r* value of 1 or -1 means the two data sets are perfectly correlated (whether it is in the positive direction or negative direction) and a *r* value of 0 means no correlation. In practice, a correlation of 0.1 to 0.3 (both positive or negative) is considered to be small, a correlation of 0.30 to 0.50 is considered to be medium, and a correlation of >0.50 is considered to be large.

4.2.1 QoI Comparison

In the FY21 study, the only QoI to show a statistically significant difference between the two transmissivity relationships was the ratio between the mass flow rate from the aquifer to east boundary and the mass flow rate from the rock to east boundary. As can be seen in Figure 4-1, it is still not accurate to say the transmissivity relationship is statistically significant for the maximum ^{129}I concentration in the aquifer due to the large and overlapping intervals.

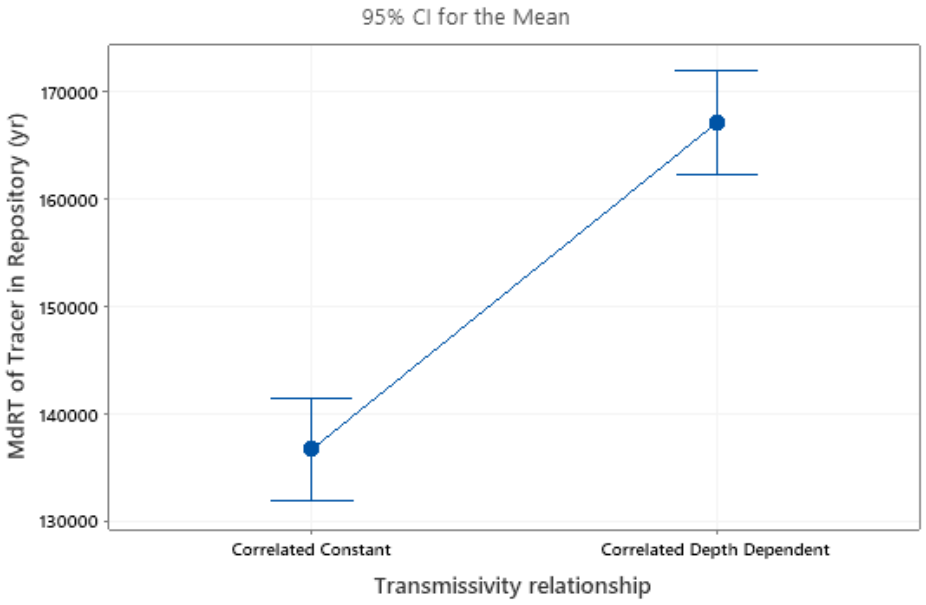
However, there is quite a large difference between the two transmissivity relationships for the QoIs plotted in Figure 4-2 through Figure 4-5. The repository median residence time, fraction of tracer still in the repository, fractional mass flux from the repository and the ratio between the mass flowrate from the rock to aquifer and the mass flow rate from the rock to east boundary all now show a statistically significant difference between the correlated constant and correlated depth-dependent transmissivity relationship. Based on this, it can be assumed the original set of DFNs (which was a total of 20) was too small to draw concrete conclusions about statistical significance. The maximum ^{129}I concentration still shows little significant difference with the larger DFN sample sets, but the timing of this maximum is affected, as the other metrics shown in Figure 4-2 through Figure 4-5 are related to timing.

One additional observation that can be made is with respect to the ratio of the mass flow rates. As can be seen in Figure 4-5, the mass flow rate ratio for the correlated constant transmissivity case is negative while the mass flow rate ratio for the correlated depth dependent transmissivity case is positive. This indicates that the mass flow rate from the rock to aquifer is downward for correlated constant and upward for correlated depth dependent. This is investigated further in Section 4.2.3.



The pooled standard deviation is used to calculate the intervals.

Figure 4-1. Interval plot for the scalar maximum ^{129}I concentration [M] in the aquifer after 1 million years versus transmissivity relationship



The pooled standard deviation is used to calculate the intervals.

Figure 4-2. Interval plot for the time when half the tracer is gone from the repository in years versus transmissivity relationship

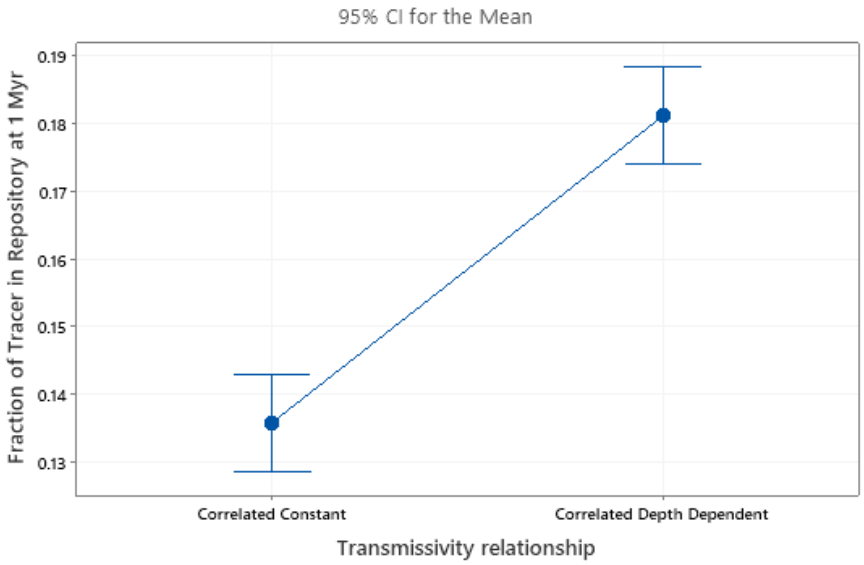


Figure 4-3. Interval plot for the fraction of tracer still in the repository at 1 million years versus transmissivity relationship

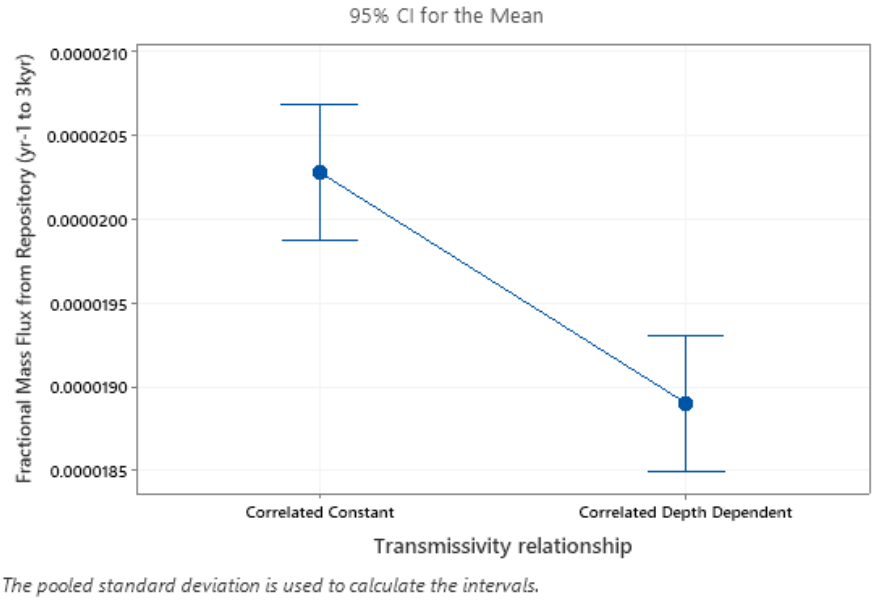


Figure 4-4. Interval plot for the fractional mass flux from the repository at 3 thousand years versus transmissivity relationship

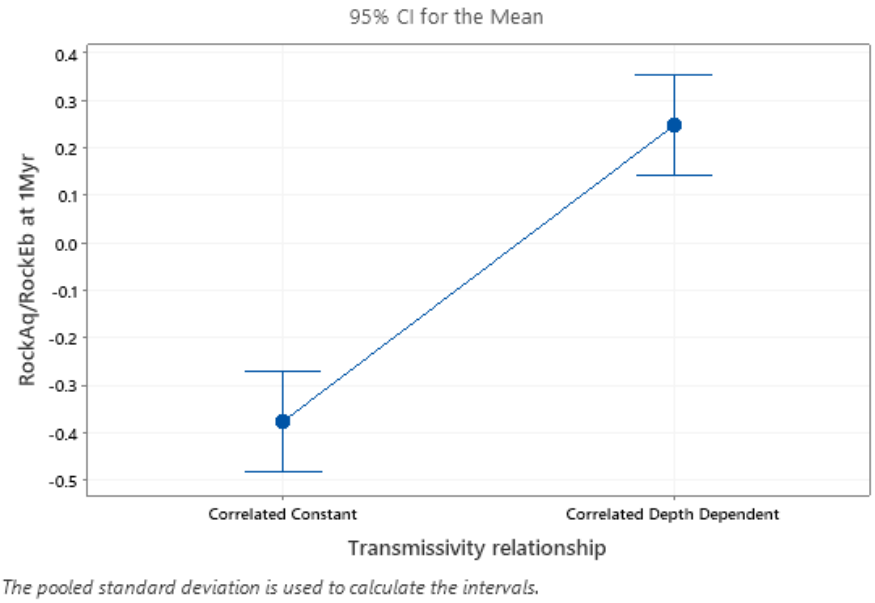


Figure 4-5. Interval plot for the ratio of the mass flow rates for the rock to aquifer and the rock to the east boundary at 1 million years versus transmissivity relationship

4.2.2 Correlation to Graph Metrics

Table 4-1 below shows the Pearson correlation between the maximum ^{129}I in the aquifer and each graph metric for both transmissivity relationships. There are multiple correlations greater than 0.30 which we deem significant, however no correlation above 0.5 was observed. Additionally, it is interesting to note that the highest correlation for the correlated constant relationship is with the number of intersections with the repository, while the highest correlation for the correlated depth-dependent relationship is the number of edges (intersections).

Table 4-1. Maximum ^{129}I in the aquifer correlation with graph metrics

Graph metric	Correlated constant	Correlated depth dependent
Average degree	-0.023	0.104
Density	0.087	0.274
Length of shortest path	-0.016	-0.098
Number of intersections with repository	0.382	0.116
Number of edges	-0.163	-0.325
Number of nodes	-0.125	-0.307
Shortest travel time	0.007	0.003

The number of intersections with repository graph metric seemed to have the best correlation with majority of the other QoIs as well. The strongest correlation was with the repository median residence time, fraction of tracer still in the repository at 1 million years, and fractional mass flux from the repository at 3 thousand years. This correlation is shown in Figure 4-6, Figure 4-7, and

Figure 4-8 respectively. As can be seen in each of the figures, the absolute value of the Pearson correlation value is higher for the correlated constant transmissivity relationship in every case.

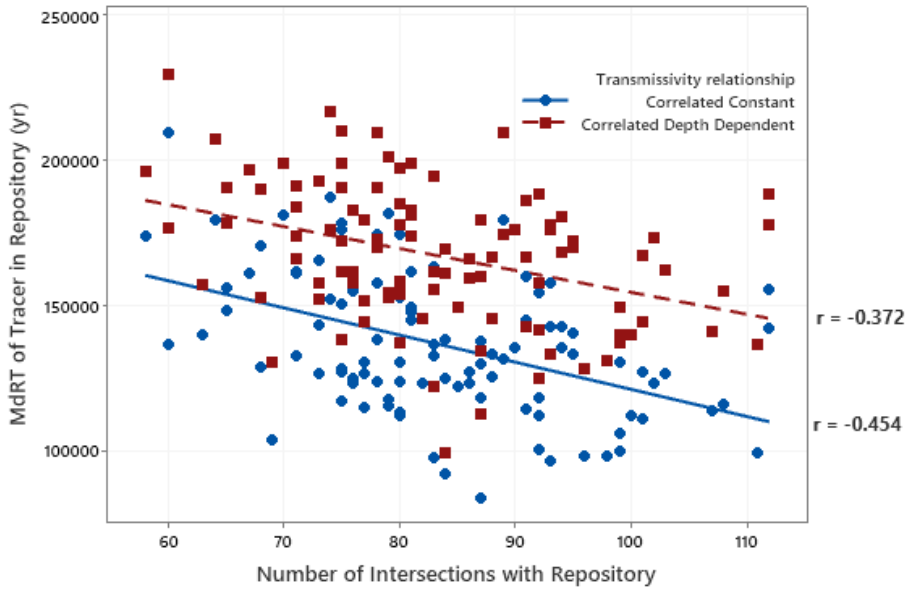


Figure 4-6. The time when half the tracer is gone from the repository in years and number of intersections with repository correlation

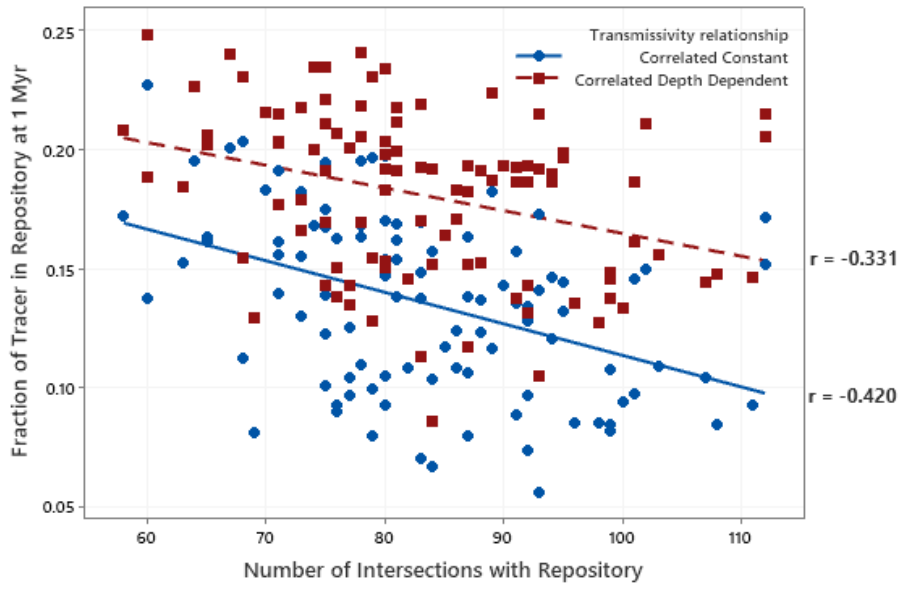


Figure 4-7. Fraction of tracer in repository at 1 million years and number of intersections with repository calculation

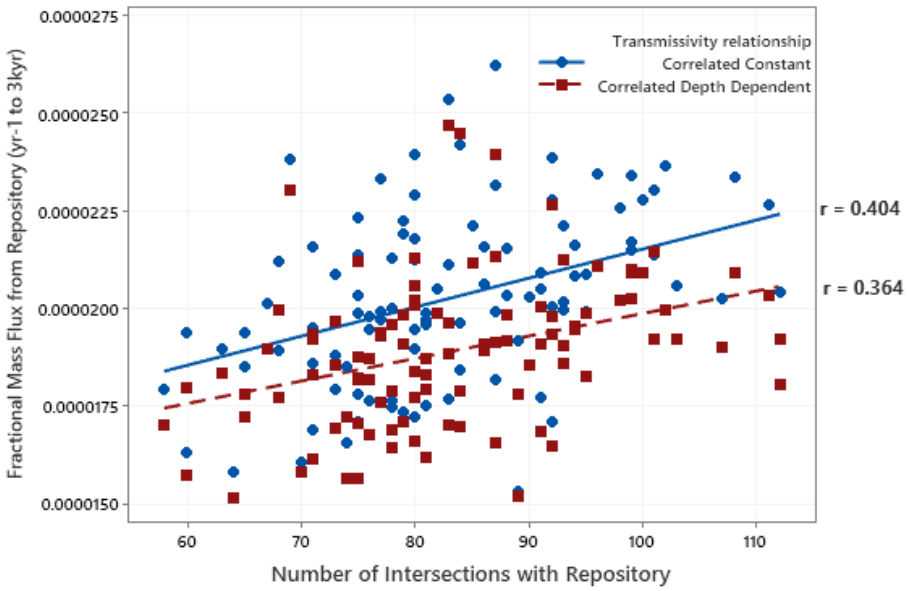


Figure 4-8. Fractional mass flux from repository at 3 thousand years and number of intersections with repository correlation

4.2.3 Time-dependent QoIs

Each of the QoIs assessed was also observed as a function of time using the time history data outputs. A majority of the time dependent results were very similar to the results shown in Section 4.2.1 above. For example, Figure 4-9 below shows a time-dependent main effects plot for the maximum ^{129}I concentration in the aquifer with the correlated depth-dependent transmissivity data shown in red and the correlated constant transmissivity data shown in blue. Figure 4-9 through Figure 4-13 plot the mean value of a QoI for the CC vs CDD transmissivity across time along with the respective confidence intervals in those mean values. For this analysis, the goal is to see if the mean response (across 100 DFNs) with constant correlated transmissivity is different from the mean response (across 100 DFNs) with correlated depth-dependent transmissivity, at all points in time. Significant differences in the mean values for two treatment levels (e.g. two transmissivities) is a form of main effects analysis. As can be seen in Figure 4-9, there is a large overlap of the confidence intervals between the two transmissivity relationships. Although the confidence intervals for the peak maximum ^{129}I values in the aquifer are overlapping (located at approximately 400,000 years for the correlated constant case and 1 million years for the correlated depth-dependent case), it is interesting to note that the correlated depth-dependent case seems to continue to rise in concentration over time and it is worth further investigation.

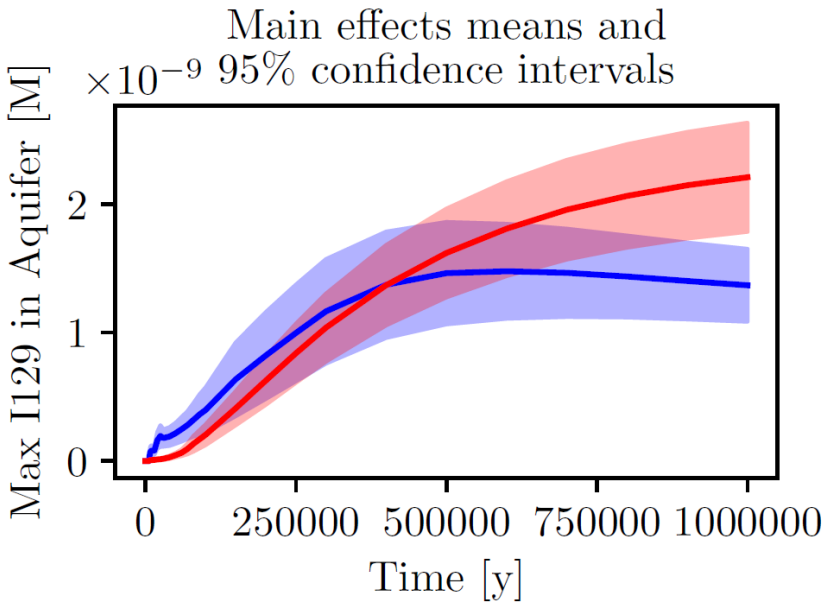


Figure 4-9. Main effects plot for the maximum ^{129}I concentration [M] in the aquifer over time: red is correlated depth-dependent and blue is correlated constant transmissivity

However, some additional interesting observations also came to light when observing the time dependent QoIs. In the 2021 study [1], when observing the difference in permeability, it was found that the correlated constant transmissivity relationship had permeability values of the same order of magnitude in the *x, y,* and *z* directions. In contrast, for the correlated depth-dependent transmissivity relationship, the *x* and *y* directions of the permeability tensor were larger than the *z* direction by two and one orders of magnitude, respectively, for depth zones one and two (the location of the three depth zones for this reference case can be seen in Figure 5-5). Given this information, it was expected that there was an increase in downstream flow and little increase in vertical flow towards the aquifer for the correlated depth-dependent transmissivity relationship. To further investigate this, the time history data for the mass flow rates was compared.

Figure 4-10 and Figure 4-11 below, which display the time dependent data and the corresponding main effects plot for the mass flow rate from the rock to east boundary, show a significant difference between the transmissivity relationships with no overlap of the 95% confidence intervals. The mean mass flow rate for the correlated constant relationship is a little less than 90,000 kg/yr and the mean mass flow rate for the correlated depth-dependent relationships is around 1,000,000 kg/yr. This information follows the assumption made that there would be an increase in downstream flow towards the east boundary for the correlated depth-dependent transmissivity relationship.

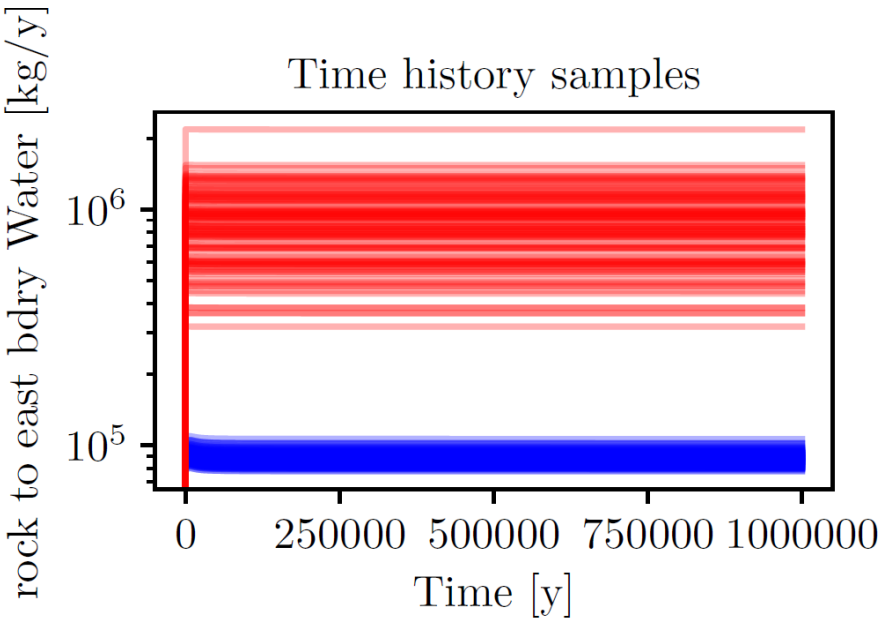


Figure 4-10. Mass flow rate [kg/yr] from the rock to east boundary over time: red is correlated depth-dependent and blue is correlated constant transmissivity

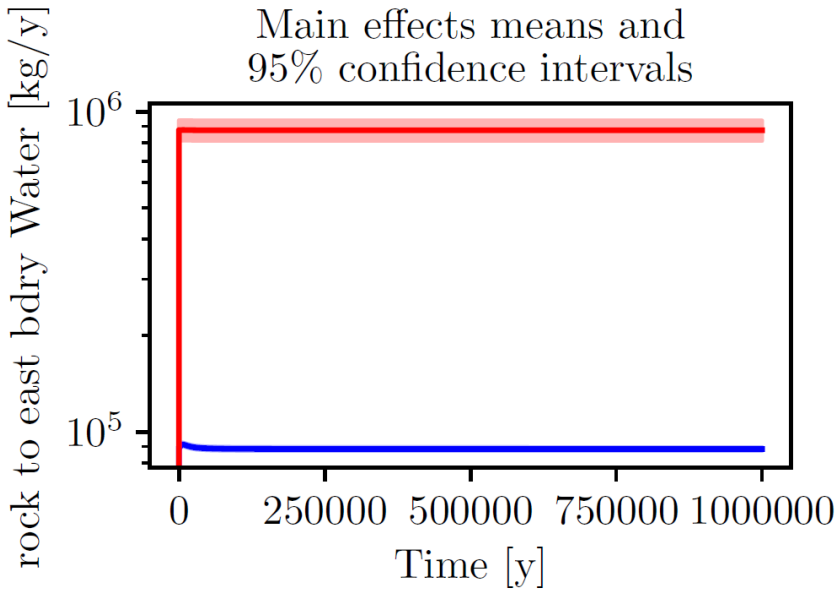


Figure 4-11. Main effects plot for the mass flow rate [kg/yr] from the rock to east boundary over time: red is correlated depth-dependent and blue is correlated constant transmissivity

Figure 4-12 and Figure 4-13 represent the time dependent data and the main effects plot respectively for the mass flow rate from the rock to the aquifer over time. As can be seen in Figure 4-12, for the correlated depth-dependent relationship (shown in red) the mass flow rates range from 1,000,000 kg/yr to -1,000,000 kg/yr (a negative flow rate indicates a downward flow from the aquifer to the rock). This large range is what is influencing the large 95% confidence intervals

in Figure 4-13 and should be investigated further. However, as can be seen in Figure 4-13, the mean mass flow rate for the correlated depth-dependent is slightly higher than the correlated constant relationship. Additionally, the mean mass flow rate for the correlated constant relationship is slightly negative which helps explain the difference shown in Figure 4-5. Nonetheless, due to the large confidence interval for the correlated depth-dependent relationship, it cannot be said that there is a statistically significant difference between the two transmissivity relationships meaning these results also coincide with the expected effect of transmissivity relationship on the streamwise flow rate and vertical flowrates.

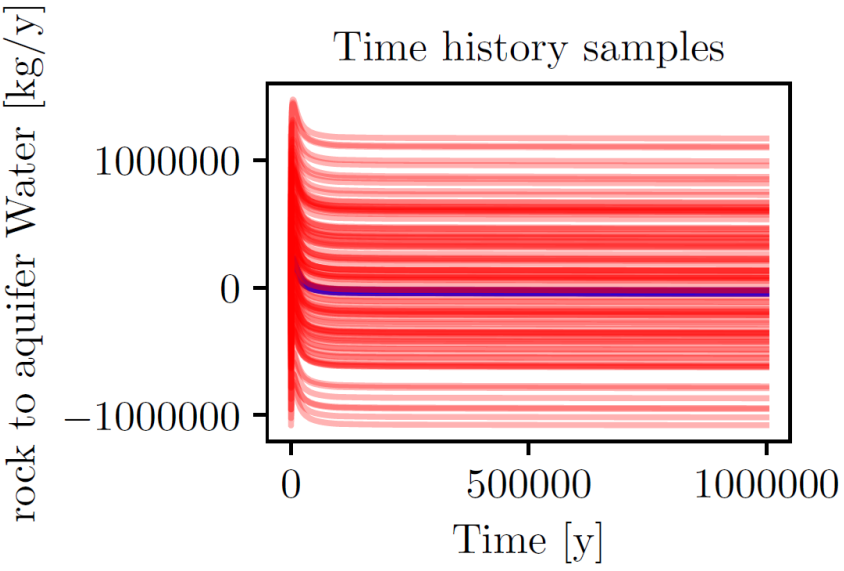


Figure 4-12. Mass flow rate [kg/yr] from the rock to aquifer over time: red is correlated depth-dependent and blue is correlated constant transmissivity

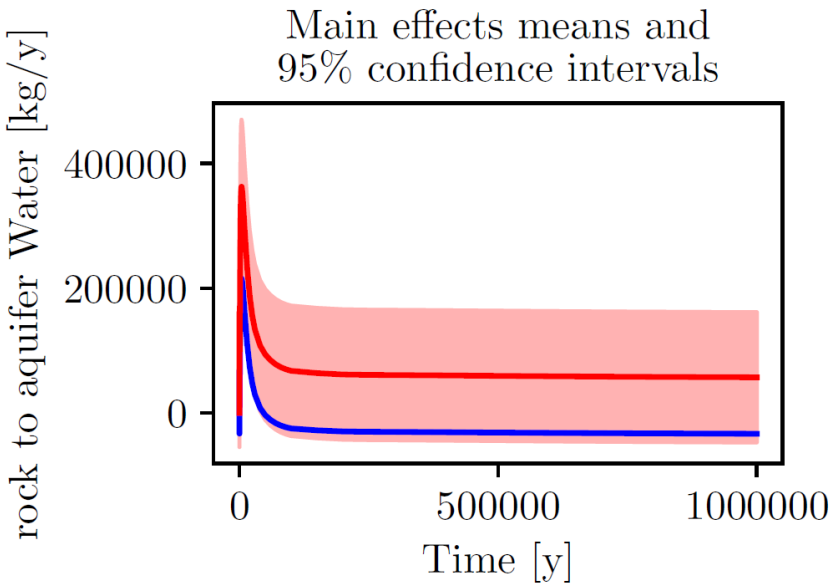


Figure 4-13. Main effects plot for the mass flow rate [kg/yr] from the rock to aquifer over time: red is correlated depth-dependent and blue is correlated constant transmissivity

4.2.4 Additional mass flux analysis

To further supplement the results and information provided previously, additional mass fluxes were extracted for the two transmissivity relationships. Rather than just observing the mass fluxes

from the rock to the aquifer, the rock to east boundary and the aquifer to east boundary, additional mass flux results were developed for the following locations within the domain:

- Horizontal plane between depth zones 1 and 2
- Horizontal plane between depth zones 2 and 3
- Vertical plane for each depth zone at the outflow (east boundary)
- Vertical plane in the middle of the domain (in the x-direction) for each depth zone

Figure 4-14 and Figure 4-15 summarize the information at 1 million years for the correlated constant transmissivity relationship and the correlated depth-dependent transmissivity relationship respectively. Each of the mass flow rate results are reported using the mean value and range for the 100 DFNs. The mass flow rate at the vertical plane in the middle of the domain for each depth zone is shown in orange, the mass flow rate at the vertical plane for each depth zone at the outflow is shown in green, and the mass flow rates at the horizontal planes between the different depth zones and the aquifer is shown in blue. Red arrows are also provided to show the relative magnitude and direction of flow for each of the mass flow rate results.

As can be seen in the two figures below the results agree favorably with the results shown in Section 4.2.3 as well as the permeability differences report in the previous study [1]. For depth zone 3, the mean mass flow rates for the both the middle of the domain and at the outflow are comparable with the correlated constant relationship being slightly higher. As for depth zone 2 and depth zone 1, the mass flow rates for the correlated depth-dependent relationship are significantly higher. The mass flow rate at the horizontal plane between depth zone 1 and the aquifer agree with the results shown in Figure 4-13 and the range in mass flow rates between the rock and aquifer for the correlated depth dependent case is drastically larger than the correlated constant transmissivity case which is consistent with Figure 4-12. Additionally, at the horizontal plane between depth zone 1 and the aquifer, the mean flow rate for the correlated constant transmissivity case is negative while the mean flow rate for the correlated depth dependent transmissivity case is positive but the mean mass flow rates between the depth zones are negative for both transmissivity relationships. These differences in the direction of flow between the transmissivity relationships, especially at certain locations within the domain, should be looked into further.

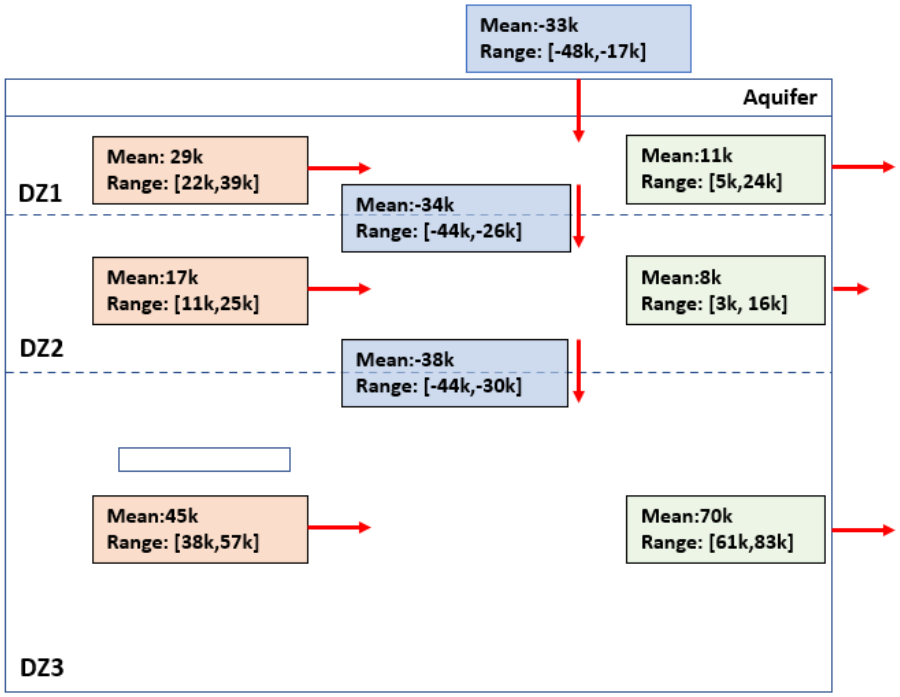


Figure 4-14. Specific mass flow rate information (in kg/yr) for the correlated constant case at 1 million years

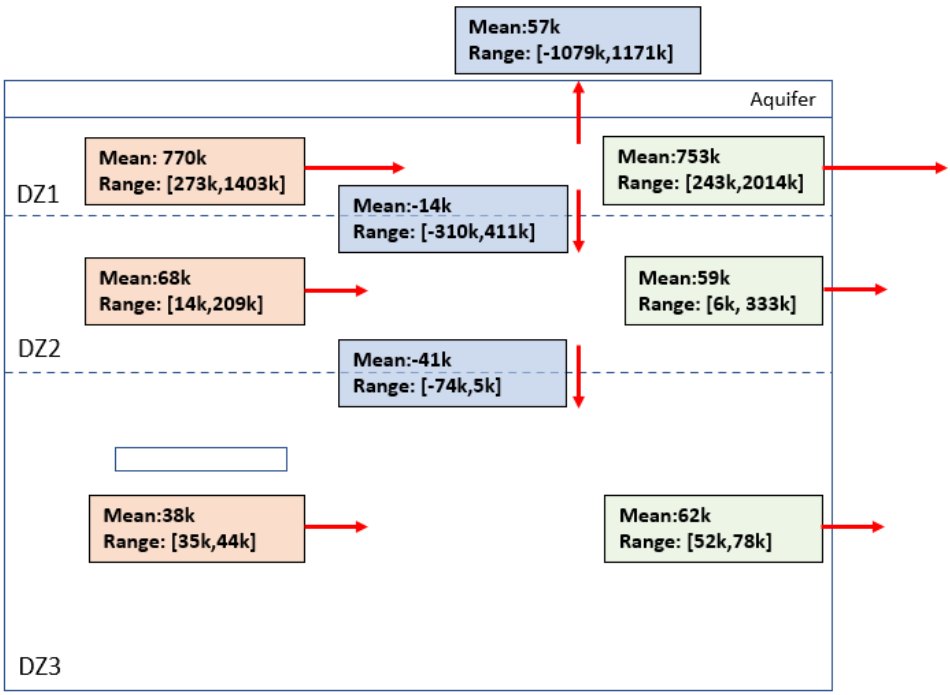


Figure 4-15. Specific flow rate information (in kg/yr) for the correlated depth-dependent case at 1 million years

4.2.5 Comments on DFN analysis

In conclusion, the purpose of this study and the original study was to determine if a correlated depth-dependent transmissivity relationship produces a significant change in the performance quantities for the flow and transport simulations of nuclear repositories in subsurface rock. Unlike the original study, it was found that five out of six quantities of interest assessed showed a statistically significant difference between the two relationships. Although the maximum ^{129}I in the aquifer showed no real change, it is worth noting that there was a considerable difference in timing for the maximum ^{129}I concentration in the aquifer for the two transmissivity relationships. This is worth further investigation and might be correlated to the difference in the direction and range of flow rates seen in Section 4.2.4. In addition, the maximum ^{129}I in the aquifer showed no real correlation with any graph metric for either relationship. Although we did not see strong correlation with respect to the maximum ^{129}I in the aquifer, the number of intersections with repository proved to be the most useful graph metric when considering all of the QoIs. Interestingly, the strongest correlation was seen in the correlated constant relationship as well. Additional graph metrics that are well-correlated with performance quantities of interest will be sought in future work. Lastly, the observations made with respect to the mass flow rates (specifically the rock to east boundary flow rate) may indicate increased flushing for the correlated depth-dependent relationship and this behavior is worth investigating further.

5. 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 (or radionuclides produced within radioactive decay chains) released from a waste package (WP) 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. [29], Mariner et al. [6], and Sevougian et al. [30] except for the use of the correlated depth-dependent transmissivity in the fracture network implementation (discussed in Chapter 4), the design of the uncertainty analyses (Section 5.1), and modifications to the model domain (Section 5.2). Note that this analysis builds extensively on the crystalline reference case uncertainty analyses performed in 2019, 2020, and 2021 which are documented in Chapter 8 of [3], Chapter 4 of [2], and Chapter 4 of [1]. In previous case studies, we have performed sensitivity analysis studies on a generic crystalline reference case, first studying the effects of epistemic uncertain parameters [2] and then extending this to account for the spatial heterogeneity represented by a family of discrete fracture networks (DFNs) [1]. **This work represents a further extension: analyzing the effect of a constitutive chemistry model for nuclear fuel dissolution.** In this chapter, we present a SA case study that seeks to separate the effect of epistemic parameters, spatial heterogeneity as represented by DFNs, and model choice as represented by the form of the FMD alternative model.

5.1 Uncertainty Analysis (UA)

The uncertainty analysis includes multiple types of uncertainty: spatial uncertainty, epistemic parameter uncertainty, and model form uncertainty. Uncertainties described as spatial uncertainty relate to the unknown structure of the repository and surrounding rock, as well as variation in degradation rate between waste packages. The spatial uncertainty is not strictly categorized as epistemic or aleatory, in part because it incorporates some of both types of uncertainties. For a more detailed discussion on spatial uncertainty treatment, see [1]. Uncertainties described as parameter epistemic uncertainty relate to inputs with fixed but unknown scalar values. Uncertainties described as model-form relate to the choice of alternative models used to represent fuel degradation.

High-fidelity fuel degradation modeling is computationally intensive, so alternative models can provide tangible computational benefits, particularly for analyses where many simulations are required. One alternative for such modeling utilizes a fractional dissolution rate (FDR). Another modeling alternative utilizes machine learning to construct a surrogate model of the high-fidelity

fuel matrix degradation (FMD) mechanistic model. Both provide computational efficiency compared to the high-fidelity mechanistic model which is intractable at a detailed repository scale.

This chapter examines how these two modeling approaches to waste form degradation affect performance predictions for a generic crystalline repository reference case. The surrogate model alternative (an Artificial Neural Network or ANN), built from high-fidelity FMD training data, has more detailed predictions of the waste form behavior over time because it incorporates degradation rates changing with time. We examine how these FMD model alternatives influence the behavior of the system.

Section 5.1.1 summarizes the structure and implementation of the UA, Section 5.1.2 describes the fuel matrix degradation alternative options, and Section 5.2 describes the crystalline reference case model set-up. SA results are presented in Section 5.3 with a summary discussion in 5.4.

5.1.1 Uncertainty Implementation

A nested sampling loop was employed in previous work with the crystalline reference case [1, 2, 3]. This is shown in Figure 5-1, where the stochastically generated DFNs in the outer loop represent spatial heterogeneity and the epistemic parameters represent quantities such as corrosion rates, fuel degradation rates, instant release fractions, matrix diffusion, buffer permeability, etc. Nested sampling allows examination of uncertainty/variability within one uncertainty type. For example, we can study the sensitivity of the epistemic parameters across the entire population of samples or for each DFN. This nested sampling structure also can be useful for understanding the relative importance of inherent variability versus reducible uncertainty.

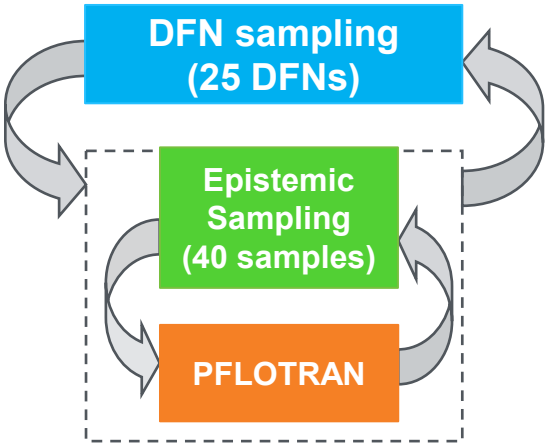


Figure 5-1. Nested Sampling Loop used to sample over Discrete Fracture Networks and then, per each DFN, over epistemic samples

The UA comprises a spatial loop of sample size 25 fracture network realizations, and a parameter loop of sample size 40 for a total of 1000 simulations. Note that both 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 (see Section 5.2.3). The sampling loops are shown in Figure 5-2. Note that in the following figures, green coloring represents epistemic parameters, blue represents spatial heterogeneity, and orange represents the PFLOTRAN model.

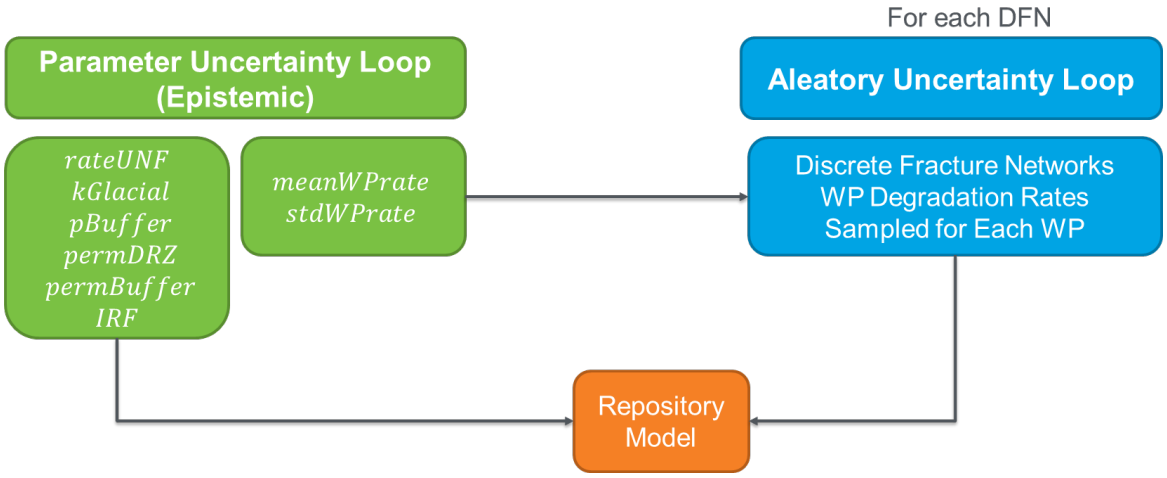


Figure 5-2. Sampling loops used in the Uncertainty Analysis of the Crystalline Reference Case presented in this chapter

To address the different alternatives used for fuel degradation modeling, two separate workflows were run in the Dakota framework. These are shown in Figure 5-3. Each workflow utilized the same samples for the DFNs and for the epistemic samples, so the only difference between the two is the alternative model choice for fuel matrix degradation. One of the epistemic parameters, *rateUNF*, only pertains to the FDR model (left) and so is only included in the workflow for that model.

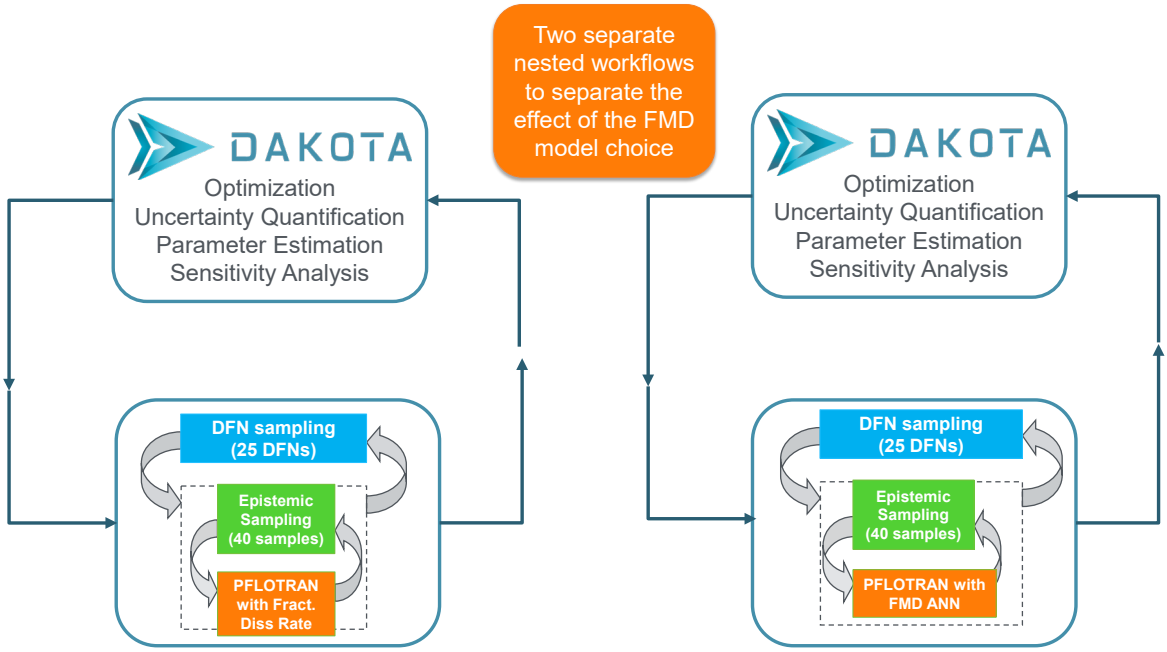


Figure 5-3. Two similar workflows used to run 1000 PFLOTRAN simulations each. The workflow on the left utilizes a FDR model, where the workflow on the right utilizes an Artificial Neural Network model for Fuel Matrix Degradation

Latin hypercube sampling of epistemically uncertain parameters is performed using Dakota [16]. The uncertainty analysis includes *pBuffer, permBuffer, permDRZ, kGlacial*, *rateUNF,* and *IRF* 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 5-1.

We note that the UA consists of 40 different epistemic sample vectors for each of the 25 spatial realizations, resulting in 1000 unique sample vectors for the epistemic variables. This increases coverage of the epistemic sample space compared to repeating the same epistemic samples for each spatial realization and reduces interference from repeated values in the global sensitivity analysis: the unique samples help with surrogate model construction for sensitivity analysis. The cost of this approach (as compared to a case where the same 40 epistemic parameters are used for each DFN) is that it is more difficult to quantify how much of the QoI uncertainty is due to spatial heterogeneity versus epistemic parameter uncertainty. This particular sample design was chosen for the coverage of epistemic parameter space. The epistemically uncertain parameters and their distributions are listed in Table 5-1. We highlight a few things: the waste package (WP) degradation rate varies between waste packages. That is, a mean and standard deviation of the WP corrosion rate model are sampled for each PFLOTRAN run, and then repeated sampling of this distribution is performed for each waste package within the run. Also, in Table 5-1, all eight parameters listed in the Parameter loop are sampled for the simulations using the FDR model. When the ANN surrogate is used, the *rateUNF* parameter is omitted because it is only used in the FDR model; the other seven parameters are still sampled but no uncertainty is included in the ANN surrogate.

Table 5-1. Uncertainty distributions propagated in crystalline reference case UA

Input	Description	Range	Units	Distribution	Sampling Loop
DFN	Particular realization of a discrete fracture network	1-25			Spatial loop
Aleatory_1	Aleatory sampling to generate a fixed ordering of waste package degradation for each DFN	0-1		Uniform	Spatial loop
rateUNF	FDR 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
IRF	Instant release fraction	0.038 – 0.156		Uniform	Parameter loop

Implementation of spatial uncertainty in the PFLOTRAN model via DFN generation is described in Section 4, as well as previous milestone reports [1, 2, 3]. However, this uncertainty must also be incorporated into the SA. The DFNs are not defined parametrically, but the variation in QoIs due to spatial heterogeneity is significant. This spatial heterogeneity is included in SA using graph metrics, which are used in surrogate model construction as if they were model inputs like those in Table 5-1. The graph metrics are the shortest travel time (STT), the number of intersections with the repository (nIntersections), and the average degree (AveDegree), which are described in Section 4.2.2.

5.1.2 Fuel Matrix Degradation Alternatives

As mentioned above, the analyses in this chapter compare two alternative models for fuel degradation modeling. The first alternative is a FDR model. The second modeling alternative utilizes machine learning to construct a surrogate model of the high-fidelity fuel matrix degradation mechanistic model. Both provide computational efficiency compared to the high-fidelity mechanistic model which is intractable at a detailed repository scale.

5.1.2.1 Fractional Dissolution Rate

The FDR model uses a FDR and the radionuclide concentrations in the waste form to determine the mass dissolution rate for each waste form [31, 32]. It is a fairly simple, analytic model which is heavily dependent on the FDR of spent nuclear fuel, denoted as *rateUNF*.

5.1.2.2 Artificial Neural Network (ANN) surrogate

The Fuel Matrix Degradation model [33] is a complex chemistry model for calculating spent fuel degradation rates as a function of radiolysis, alteration layer growth, and diffusion reactants through alteration layer. It incorporates mixed potential and analytical radiolysis models. It must be called at each time step for each waste package, making it a very costly process model. To combat this cost, time-series FMD training data was obtained offline from the full process model and used to train an artificial neural network surrogate model alternative to the FMD model.

The ANN had 400K training points consisting of six inputs (fuel temp, dose rate, chemical species concentrations) and one output (UO_2 surface flux/fuel dissolution rate) [34]. A two-layer ANN with 64 nodes per layer was utilized, resulting in 4673 parameters (weights and bias terms) that were estimated based on the training data. The surrogate ANN alternative to the FMD model was called from PFLOTRAN with a similar application programming interface (API) as the full FMD model. Examples of prediction accuracy with the ANN surrogate are shown in Figure 5-4. Note that the mean absolute error on the testing set was 8.26e-4 mol/m^2/year and there was a relative test error of about 25%. This ANN has some error but matches the overall trend well, as shown in Figure 5-4.

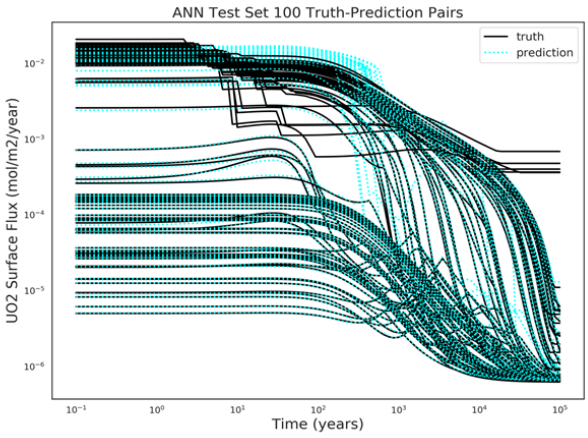


Figure 5-4. Comparison of ANN surrogate to full process model results

5.2 Model Set-up

5.2.1 Model Domain

The current model domain which has been modified from Stein et al. [29] is 3015-m in length, 2025-m in width, and 1260-m in height, partially depicted in Figure 5-5. Overlying the host rock is a 15-m thick overburden of glacial sediments. The repository is located at a depth of 585 m. Forty-two disposal drifts contain 40 12-PWR waste packages each (1680 12-PWR waste packages in total), which is half of the total from the FY20 case [2]. 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 [29] available for download at https://pa.sandia.gov. The WP placement and heat source calculations are discussed extensively in [1]: this discussion is not repeated here but the interested reader is encouraged to view that report.

In the FY21 analysis [1], PFLOTRAN's integral flux card was used in the model to set up three surfaces through which fluxes of all primary dependent variables could be calculated. In the first, three coordinates were used to define a plane between the granite (rock) and glacial (aquifer) materials within the model. The second surface defines a 2D outflow region from the aquifer through the east boundary (downstream of the repository). The third surface defines a 2D outflow region from the rock to the east boundary. Figure 5-5 shows the PFLOTRAN integral flux planes within a cut-away of a single epistemic run mapped to a porous medium grid showing the full repository and the far half of the model domain. Within this figure, the rock to aquifer plane is colored in turquoise, the aquifer to east boundary plane in purple, and the rock to east boundary plane in blue. Also included in the figure are labels for the location of the repository, rock, aquifer, four observation points represented by turquoise spheres, and the location of three depth zones.

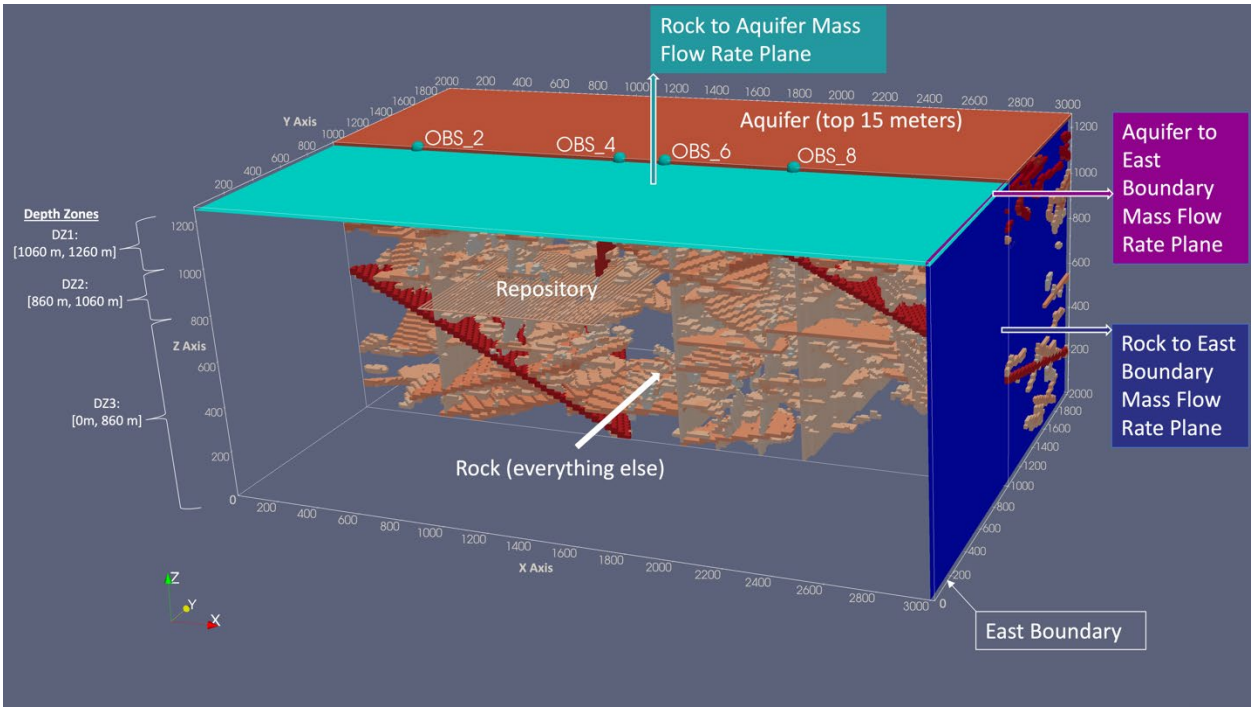


Figure 5-5. FY21 PFLOTRAN integral flux planes and depth zones visualized within the model domain

5.2.2 Discrete Fracture Networks

As described by Mariner et al. [6], the representation of fractured crystalline rock in the GDSA reference case is simplified from the well-characterized, sparsely fractured metagranite at

Forsmark, Sweden [23]. The fracture networks used here are a simplification derived from the Forsmark case: they were customized for this crystalline reference case. 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. Figure 5-5 shows the far half of the model domain for one DFN. This figure cuts the geologic stratigraphy through the midline of the repository, shown in the lower left, but shows the full repository. We note that the repository is near the top of depth zone 3 (or 675m above the bottom of the domain as shown in Figure 5-5.) Fractures of the DFN realization are shown in various shades of orange. Unconnected fractures are not shown. There are four observation points at the top boundary of the host rock located on the vertical plane that intersects the midline of the repository. They are also located where deterministic features intersect the top boundary of the host rock. These user-defined features represent large, mappable fracture zones, such as faults. There are three subvertical fracture zones (in gray) and two fracture zones with a dip of approximately 30 degrees (in red). These features are deterministic and are common to each DFN realization. Overlying the host rock is a 15-m thick overburden of glacial sediments (not shown). For more details of the DFN construction, see [1, 2].

The crystalline host-rock reference case analyzed here [6], based on the Forsmark data set, is modeled using some fixed features and an uncertain fracture network. The case contains five fixed fracture zones and three depth intervals. Twenty-five discrete fracture networks were generated with dfnWorks [28], one for each realization of the spatial uncertainty loop. The depth-dependent transmissivity relationship was used, as described in Chapter 4. The DFNs are 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 [29].

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 [28].

5.2.3 Waste Package Corrosion Model

The waste package corrosion model implemented in PFLOTRAN (Mariner et al. [6], 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. Details of the WP corrosion rate modeling are provided in [1].

5.2.4 Initial Conditions

Initial conditions specified for this case study 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.

5.2.5 Timestep Size

During previous analysis of the crystalline reference case [3], 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 [3], 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.

5.2.6 Quantities of Interest (QoIs)

We examine many Quantities of Interest (QoIs) in the crystalline reference case. To clarify different types of output metrics, see Figure 5-6.

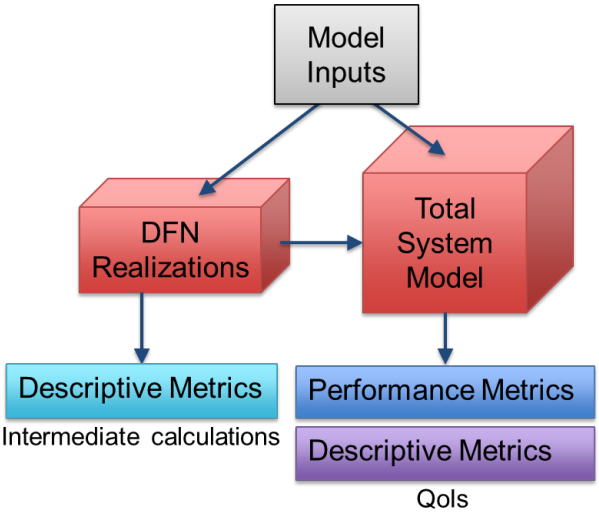


Figure 5-6. Types of model outputs

Performance metrics are QoIs used to assess repository performance. Examples of such metrics include dose, concentrations, and radionuclide release to the biosphere over time. We

define descriptive metrics as either QoIs or quantities from intermediate calculations that describe the system or its behavior, but do not describe repository performance. Examples include bulk quantities related to (or affected by) the stochastically-generated discrete fracture networks such as the number of fractures intersecting the repository or the shortest path to the aquifer. Other descriptive metrics include time-dependent QoIs such as median tracer residence time in the repository, net water fluxes across regions, and water flux ratios (e.g., upward vs. eastward). Finally, there are simple descriptive metrics such as number of waste packages breached.

By defining these different types of outputs, we can describe unique sensitivity analysis structures and the questions each type of analysis seeks to address, shown in Figure 5-7. All three types of analyses have been performed with the crystalline reference case [1, 2] and we continue to define new descriptive metrics to support a better understanding of system behavior. This year's analysis builds upon these analysis types to investigate the effects of the fuel matrix degradation model alternative choice on model behavior and performance.

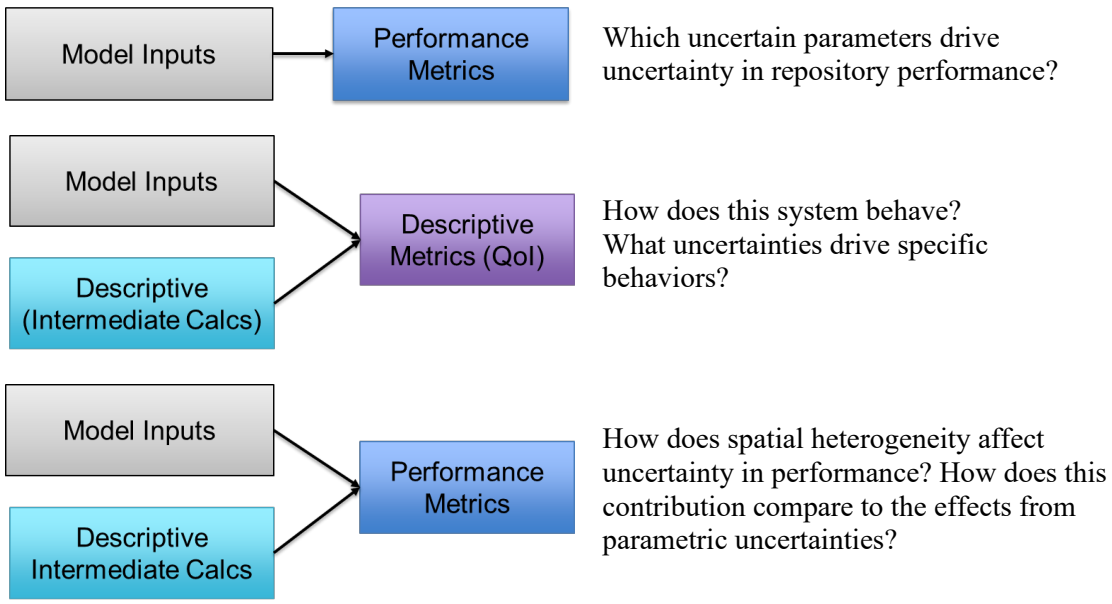


Figure 5-7. Types of sensitivity analysis based on types of model outputs

The QoIs used for the sensitivity analysis of the updated crystalline reference case are listed in Table 5-2. The outputs include maximum ^{129}I concentration in the aquifer over time and location as well as quantities such as mean travel time, median residence time, and flow rate ratios. We note that several quantities of interest were generated for the crystalline reference case in FY20. These are described in [2] and we are still tracking them. We continue to develop additional QoIs that may be better indicators of repository performance, repository leakiness, retention of conservative tracers in the repository, relative flow rates and directions through the rock, and flow connections between the repository and aquifer. Leakiness describes water flux out of the repository.

Table 5-2. FY2022 QoIs for crystalline reference case sensitivity analysis

#	Quantities of Interest (QoIs)	Description
1	Peak ^{129}I concentration in aquifer in mol/L (M)	Scalar with (X,Y,Z) location of peak. This is an indicator of repository performance.
2	MTT for Peak in years	Scalar, MTT as calculated from the tracer concentration ratio, evaluated at the time of Peak ^{129}I concentration
3	Median Residence Time MdRT of Spike in Repository	Median residence of a tracer in the repository. This is the time when half the tracer remains in the repository. It is an indicator of repository retention
4	Fraction of Spike in repository at 3000 yrs	The fraction of a tracer remaining in repository at 3000 years. It is an indicator of repository retention.
5	Fraction of Spike in repository at 1 million years	The fraction of a tracer remaining in repository at 1 million years. It is an indicator of repository retention.
6	Fractional Mass Flux of Tracer from Repository at 3000 yr (1/yr)	The instantaneous fractional loss rate of tracer remaining in repository at 3000 years. It is an indicator of repository retention.
7	Fractional Mass Flux of Tracer from Repository at 1 million years (1/yr)	The instantaneous fractional loss rate of tracer remaining in repository at one million years. It is an indicator of repository retention.
8	Aquifer to East / Rock to East at 3000 years	This is the ratio of two water fluxes: the flux from the aquifer to the east boundary normalized by the flux from the rock to the east boundary. It indicates the multiplication factor on aquifer dominance of East side effluent at 3000yr during thermal pulse.
9	Aquifer to East / Rock to East at 1 million years	This is the ratio of two water fluxes: the flux from the aquifer to the east boundary normalized by the flux from the rock to the east boundary. It indicates the multiplication factor on aquifer dominance of East side effluent at 1 Myr near undisturbed conditions.
10	Rock to Aquifer / Rock to East at 3000 year	This is the ratio of two water fluxes: the rock to the aquifer vs. the rock to the east boundary at 3000 years. It indicates a multiplication factor on upward vs horizontal flow at 3000 years during the thermal pulse.
11	Rock to Aquifer / Rock to East at 1 million years	This is the ratio of two water fluxes: the rock to the aquifer vs. the rock to the east boundary at 1 million years. It indicates a multiplication factor on upward vs horizontal flow at 1 M years in near undisturbed conditions.
12	Rock to aquifer/ Rock to East Boundary	This is the vector of the ratio of two water fluxes: the rock to the aquifer vs. the rock to the east boundary. The vector is used in the time-varying sensitivity analysis.
12	x Location of Peak ^{129}I (m)	The ensemble indicates spatial variation in the x location of peak ^{129}I.
13	y Location of Peak ^{129}I (m)	The ensemble indicates spatial variation in the y location of peak ^{129}I.
14	Rock to Aquifer Flux	Water flux from rock to aquifer (kg/yr). This is a vector of values calculated at different time points.
15	Rock to east flux normalized by the flux at 1 Ma	Water flux from rock to the east boundary, normalized by the flux at 1 Ma. This is a vector of values calculated at different time points (note it has a value of 1 at 1Ma).
16	Aquifer to east flux normalized by the flux at 1 Ma	Water flux from aquifer to the east boundary, normalized by the flux at 1 Ma. This is a vector of values calculated at different time points (note it has a value of 1 at 1Ma).

The main performance metric capability is the peak ^{129}I concentration, as in previous studies. First, the entire aquifer is monitored at each time step for ^{129}I concentration and the maximum ^{129}I concentration is recorded. Then, the maximum value over all time points is determined. This is called the peak ^{129}I concentration in the aquifer over the course of the simulation. The location of the peak ^{129}I concentration (in terms of x and y coordinates in the aquifer layer) is also recorded. The x- and y-location of peak ^{129}I concentration indicate the spatial variability of the system. This metric enables analysis of the model and development of PA tools and techniques. Peak ^{129}I concentration is of interest because it will be a major contributor to dose.

The DFN-related characteristics that are potentially relevant to performance include the mean travel time (MTT) of a conservative tracer from the repository to the aquifer, the median residence time (MdRT) of an initial conservative tracer within the repository, fraction of a conservative tracer remaining in the repository at certain times, fractional mass fluxes of that tracer at certain times, and rock boundary water mass flow rates.

A more detailed description of MTT may be found in Mariner et al. (2020) [32] and Swiler et al. (2020) [2]. 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 [32]. Mean travel 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 decays or ingrows exponentially over time from 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. Note for the QoI studied in this report, we focus on the MTT evaluated when the peak 129I concentration occurs.

The median residence time of a conservative tracer initially present in the repository (MdRT) is the time at which half of the original tracer amount remains in the repository. We note a difference between MdRT and MTT: for MdRT, there is one initial concentration ("spike") of the conservative tracer at the beginning of the simulation and no more of it is injected into the region, then MdRT is obtained by tracking the tracer mass remaining in the repository until half remains. 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. Note additionally that for FY20 the mean residence time (MRT) was tracked, which is also computed by tracking the mass of the "spike," but the estimated MRT is valid only if the mass of the tracer has decreased nearly to zero in the repository. It was found that the fraction of mass of the "spike" remaining in the repository by the end of the simulation was still around 0.25, so the MdRT was deemed a more useful measure of repository retention for this work.

Additional metrics relating to the tracer concentrations include the fraction of the tracer remaining in the repository at 3000 years or at 1 million years. These two times are of interest because 3000 years is a time point during the thermal pulse from the repository and 1 million years is the final performance time. The final tracer-related metric is the fractional mass flux of the tracer from repository. This is also assessed at 3000 and 1 million years. It is an instantaneous fractional loss rate of the tracer remaining in the repository at a particular point in time, in units of (1/yr).

MdRT, the fraction of tracer remaining at certain times, and the fractional mass flux of the tracer from the repository are all indicators of repository retention.

The final group of metrics involves ratios of water fluxes. There are two ratios which we study in this sensitivity analysis. The first is the ratio of the aquifer-to-east-boundary flux to the rock-to-east-boundary flux. This is the ratio of two water fluxes: the flux from the aquifer to the east boundary normalized by the flux from the rock to the east boundary. It indicates the multiplication factor on aquifer dominance of the East side effluent. We evaluate this flux ratio at 3000 years during the thermal pulse which affects the rock to east boundary flux and also at one million years, where the conditions are nearly undisturbed.

The other flux ratio is the ratio of the rock-to-aquifer flux to the rock-to-the-east-boundary flux. This indicates a multiplication factor on upward vs. horizontal flow. It is also evaluated at 3000 years during the thermal pulse as well as at 1M years in nearly undisturbed conditions.

Fluxes and flux ratios were also tracked over time so time-dependent sensitivity analysis could identify the uncertainties driving changes in flow over time. This also allows for a more continuous comparison between the pressure pulse and thermal pulse phases, rather than comparing only two points in time. Time-dependent sensitivity analyses for select fluxes are presented in Section 5.3. We focus on results that are impacted by the alternative FMD model; results for QoI that are not affected by this model choice are documented in [1].

5.3 Results

This uncertainty and sensitivity analysis focused on exploration of the FY22 quantities of interest and their utility for increasing understanding of the repository and its performance. Sensitivity analysis results, time series plots, and scatter plots are presented for the QoIs that proved illustrative. The first set of results (Section 5.3.1) explores the impact that fuel matrix degradation model choice has on sensitivity analysis and the second set of results (Section 5.3.2) explores the effect of spatial heterogeneity on surrogate model construction by presenting separate surrogate models fit to each individual spatial realization.

5.3.1 Fuel Matrix Degradation Model Choice Results

Below we compare two sets of results: one using the FDR model within the Crystalline Reference case, and one using the Artificial Neural Network surrogate model for Fuel Matrix Degradation within the Crystalline Reference Case. For shorthand, results are presented labeled FDR or ANN surrogate, respectively.

As in previous analyses of the crystalline reference case, sensitivity analysis was performed using various surrogate models to estimate main and total effect Sobol' indices using Dakota [3, 16, 35]. Results are presented for second-order polynomial chaos expansion (PCE) surrogate models. This section is drawn from conference paper [36]: it has been edited and includes QoIs not discussed in the conference paper.

The maximum ^{129}I concentration in the aquifer is plotted over time in Figure 5-8 for simulations with the FDR model (black) and with the ANN surrogate (pink). The concentrations are very similar between the simulations with the two alternatives to the FMD model, but some

differences are apparent towards the end of the simulation where there is more variability in concentrations for the FDR model simulations. The mean and standard deviation of the concentrations at 1 million years are 3.66×10^{-9} [M] and 6.74×10^{-9} [M] respectively for the FDR simulations and 3.41×10^{-9}[M] and 5.00×10^{-9} [M] respectively for the ANN simulations.

The difference in uncertainty between the simulations for the two FMD model options can also be seen in the peak ^{129}I concentration empirical cumulative distribution functions (CDFs) plotted in Figure 5-9. The grey curves are the empirical CDFs for each spatial realization; the red/black curves show the empirical CDFs for the full suite of simulations with the ANN surrogate (red) and the FDR model (black). The domains for the FDR empirical CDFs are wider, indicating that there is more variation in peak ^{129}I concentrations within a spatial realization when the FDR model option is used. The increased uncertainty in the FDR simulations is likely due to the incorporation of uncertainty in that model via the *rateUNF* parameter. No analogous uncertainty is currently included in the ANN surrogate model.

Sobol' indices are plotted for the ANN surrogate simulations in Figure 5-10 (left). The top results are for the analysis without the graph metrics, and the bottom results are for the analysis with the graph metrics. These results indicate that *kGlacial* and *IRF* are the dominant parametric uncertainties driving variation in peak ^{129}I concentrations. The Intersections and AveDegree metrics are the dominant spatial uncertainties.

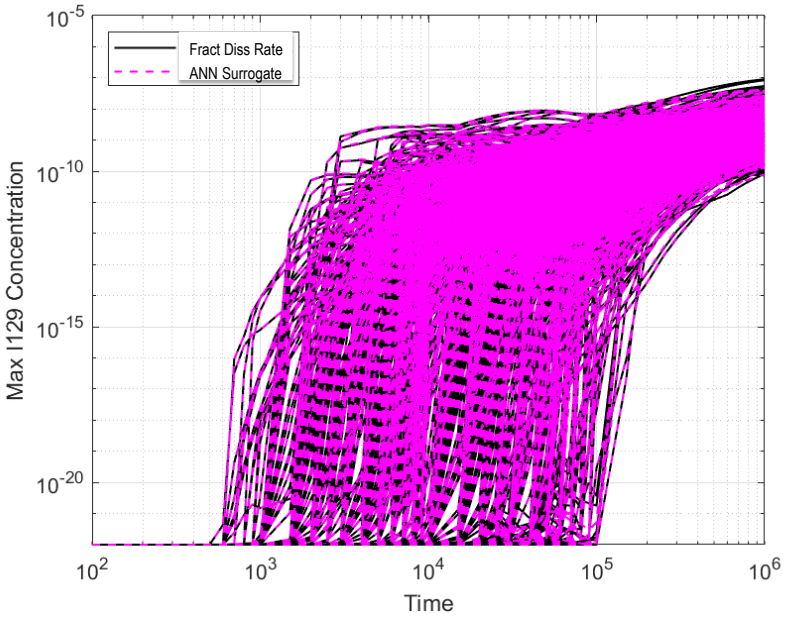


Figure 5-8. Maximum ^{129}I concentrations [M] comparison between the simulations with the ANN and FDR model

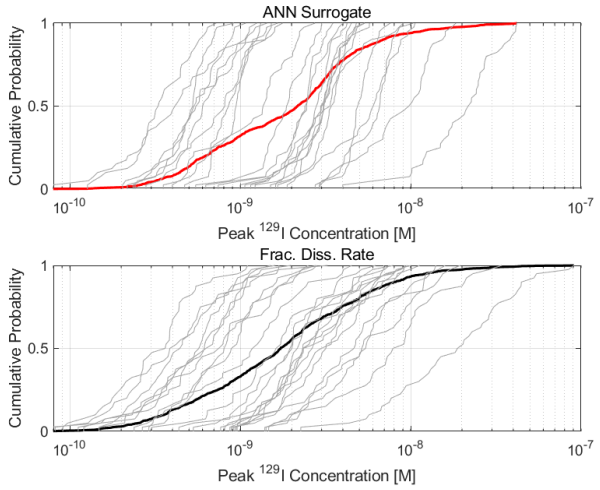


Figure 5-9. Comparison of peak ^{129}I concentration [M] empirical CDF functions for each spatial realization (grey) and the total population of simulations (black/red) for the ANN and FDR simulations

These results can be compared to those for FDR simulations, plotted in Figure 5-10 (right). For these simulations, the dominant uncertainty becomes the FDR parameter uncertainty, *rateUNF,* with *kGlacial* and *IRF* having minor effects. The analysis with the graph metrics identifies the same important spatial uncertainties as for the ANN simulations.

The analysis for peak ^{129}I concentration was repeated for the maximum ^{129}I concentration at each time step. These results are plotted in Figure 5-11 for the ANN simulations and in Figure 5-12 for the FDR simulations. Comparison of the total index results without the graph metrics (top

middle subplot) between the ANN and FDR simulations highlights the importance of the *rateUNF* uncertainty as well as the timing of importance for IRF and *rateUNF*.

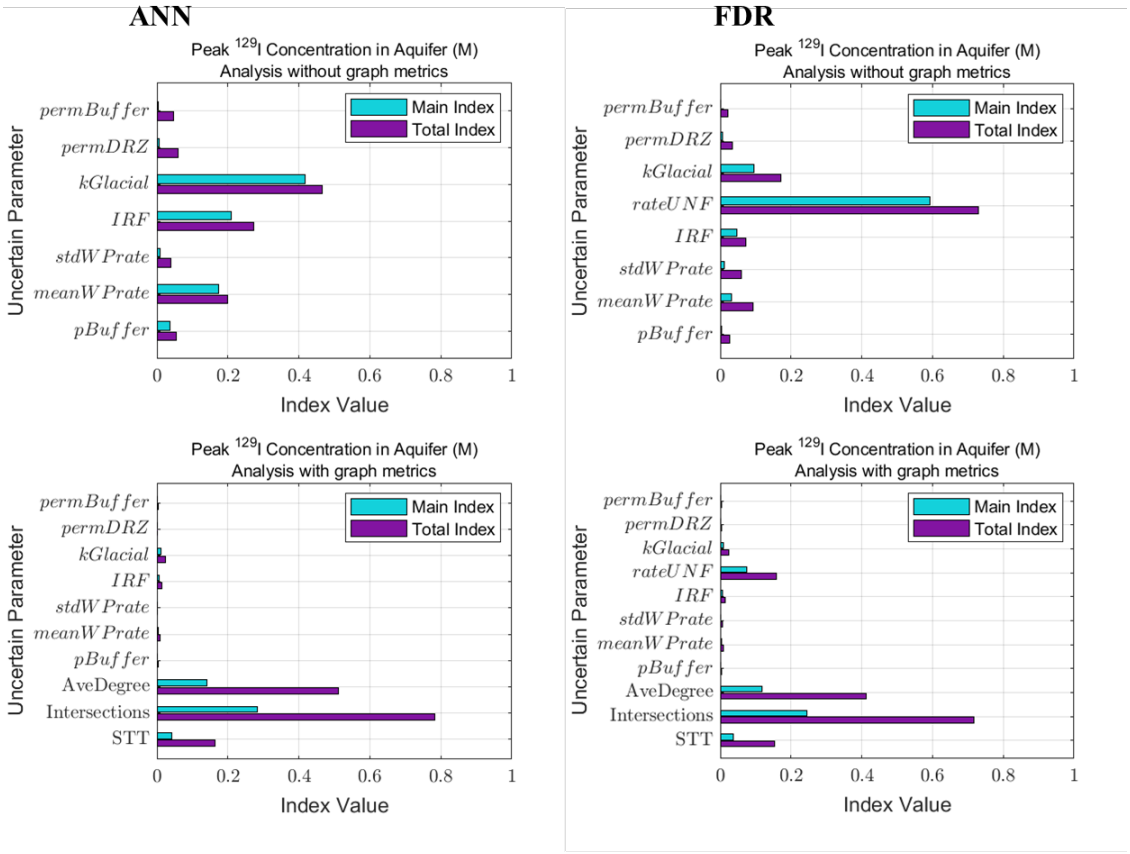


Figure 5-10. Sobol' index estimates for peak ^{129}I concentrations [M] for the simulations with the ANN surrogate model (left) vs. the FDR model (right)

According to these results, the *IRF* parameter significantly contributes to the maximum ^{129}I concentration regardless of which FMD model is used, though the waste package degradation rate is more important for most of the simulation. The *IRF* parameter decreases in importance towards the end of the simulation, where *kGlacial* becomes significant for the ANN simulations and *rateUNF* becomes significant for the FDR simulations. These sensitivity analysis results at the end of the simulation are similar to the peak ^{129}I results in Figure 5-10, which makes sense because the peak concentration is reach at the end of the simulation for most realizations. This analysis provides an explanation for why there is so little difference between the ^{129}I concentrations for the ANN and FDR simulations in Figure 5-8 until the end of simulation; the *IRF* dominates ^{129}I release early on, so the FMD model only begins to have a significant effect in the second half of the simulation time. This behavior can also be observed in the total index plots with graph metrics included (middle plot frame), but the dominance of spatial heterogeneity in these results makes it more difficult to see.

The scatterplots of the peak ^{129}I concentration versus the graph metrics and uncertain inputs are plotted in Figure 5-13 (ANN) and Figure 5-14 (FDR). The trend between *IRF* and peak ^{129}I is

similar between the two sets of simulations but note that the trend between *rateUNF* and peak ^{129}I (Figure 5-14) is more pronounced, which is consistent with the time-dependent sensitivity analysis results at the end of the simulation. Scatterplots of *IRF* and *rateUNF* versus peak ^{129}I at 200,000 years and at 1,000,000 years from the FDR model simulations are plotted in Figure 5-15. These plots show the slightly stronger trend with respect to *IRF* earlier and the stronger trend with respect to *rateUNF* later, as seen in the sensitivity analysis results.

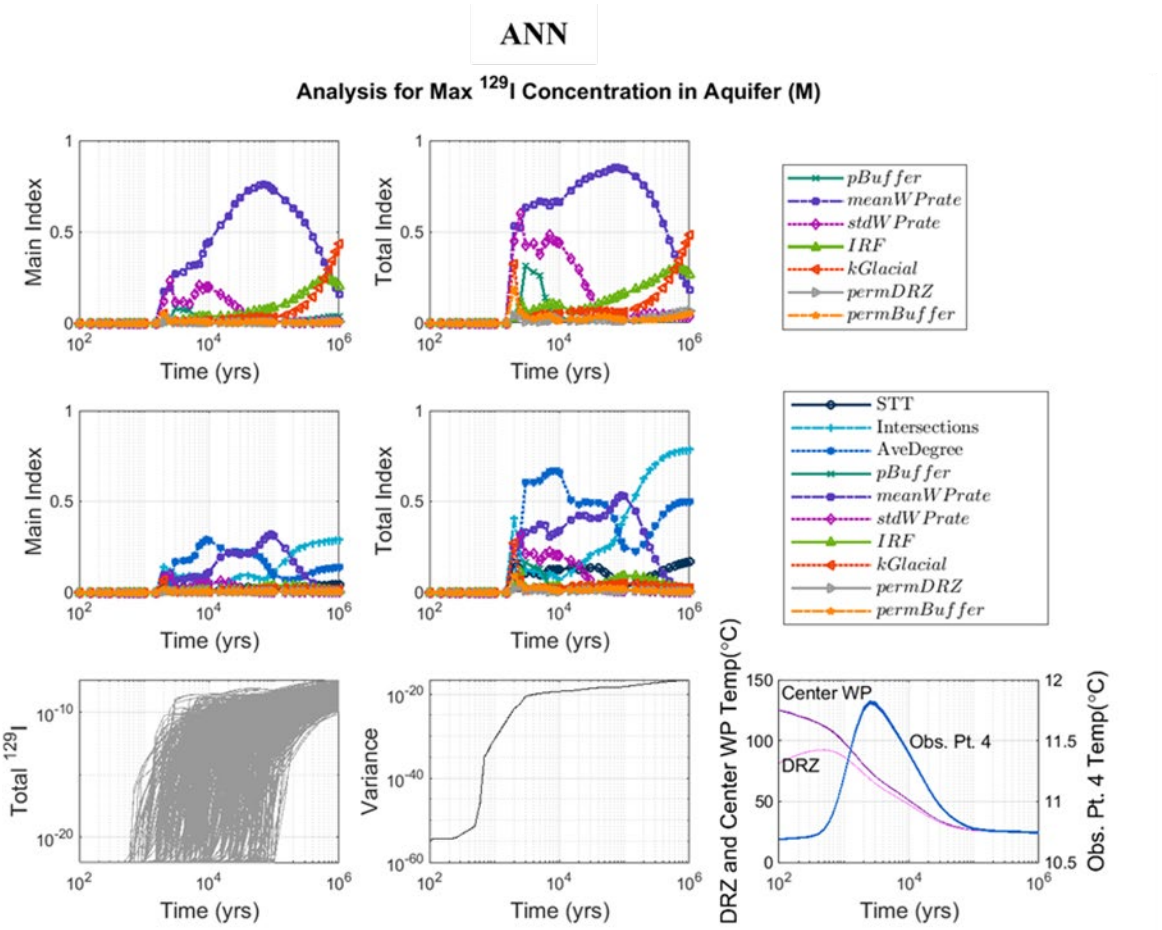


Figure 5-11. Time-dependent sensitivity analysis results for the maximum ^{129}I concentration [M] for ANN simulations

These uncertainty and sensitivity analysis results show that, overall, the choice for FMD model alternative only has a minor effect on ^{129}I concentrations. The timing for release of ^{129}I is similar between the two models, but the mean and variance of the peak ^{129}I concentration are slightly smaller at 1M years for the ANN surrogate alternative. The similar high-level behavior of the models may be due to waste package breach and instantaneous release dominating concentrations for much of the simulation time.

We expect the ANN model to provide a more detailed prediction of the waste form behavior over time because it incorporates degradation rates changing with time and accounts for the chemistry in the local environment. Future work on the FMD model alternatives will include uncertainty in the environmental chemistry for the ANN model. This may affect the relative effects of the different FMD model alternatives, which could be assessed using a similar sensitivity analysis study.

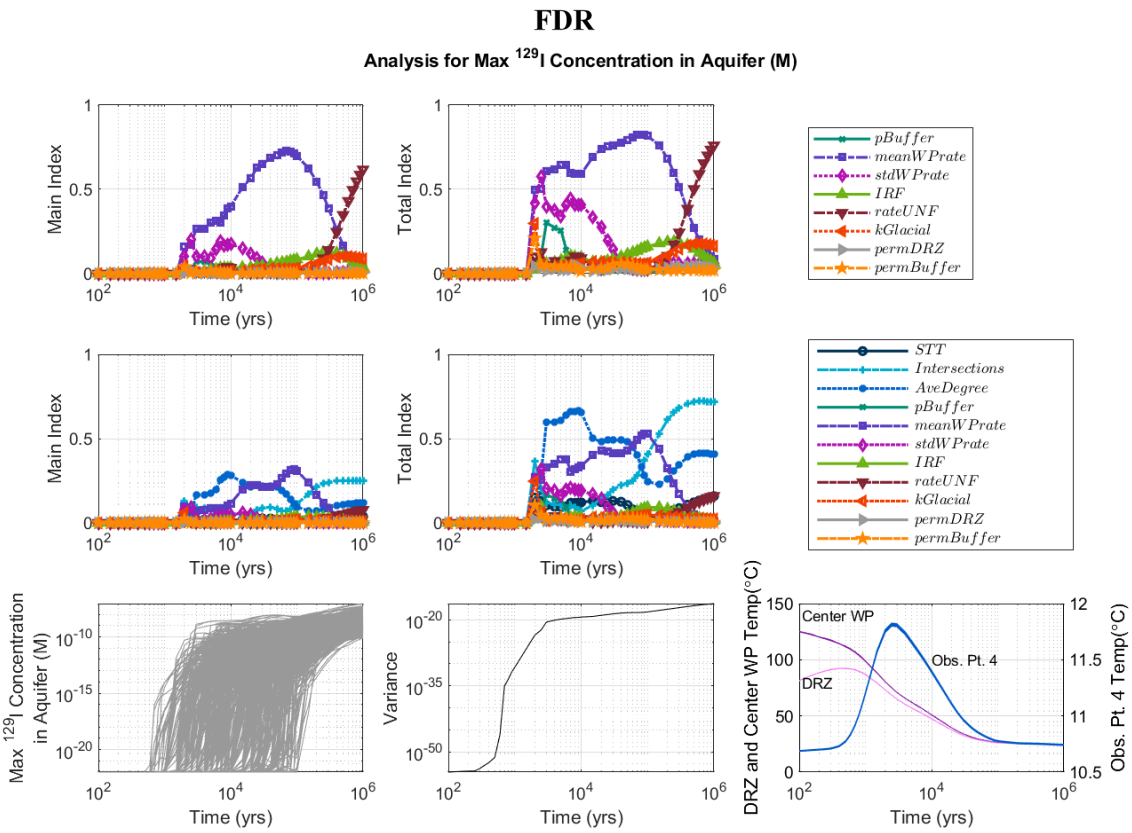


Figure 5-12. Time-dependent sensitivity analysis results for the maximum ^{129}I concentration [M] for ANN simulations

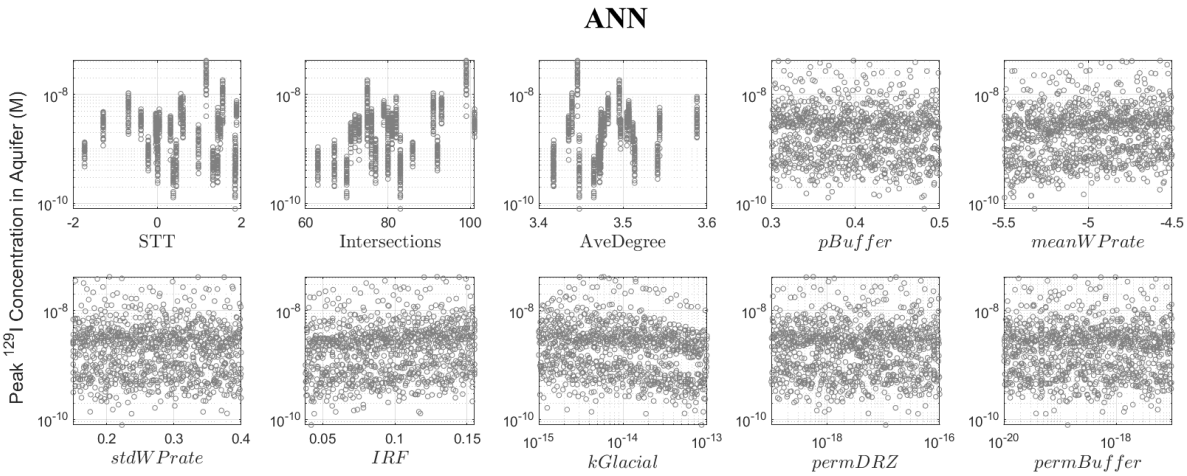


Figure 5-13. Scatterplots of the peak ^{129}I concentration [M] in the aquifer versus the graph metrics and uncertain parameters for the ANN simulations

FDR

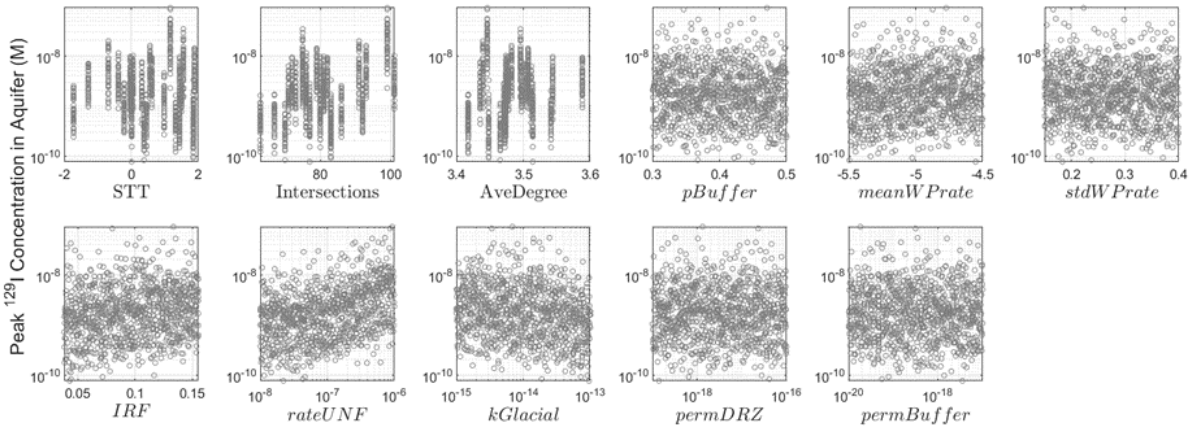


Figure 5-14. Scatterplots of the peak ^{129}I concentration [M] in the aquifer versus the graph metrics and uncertain parameters for the FDR simulations

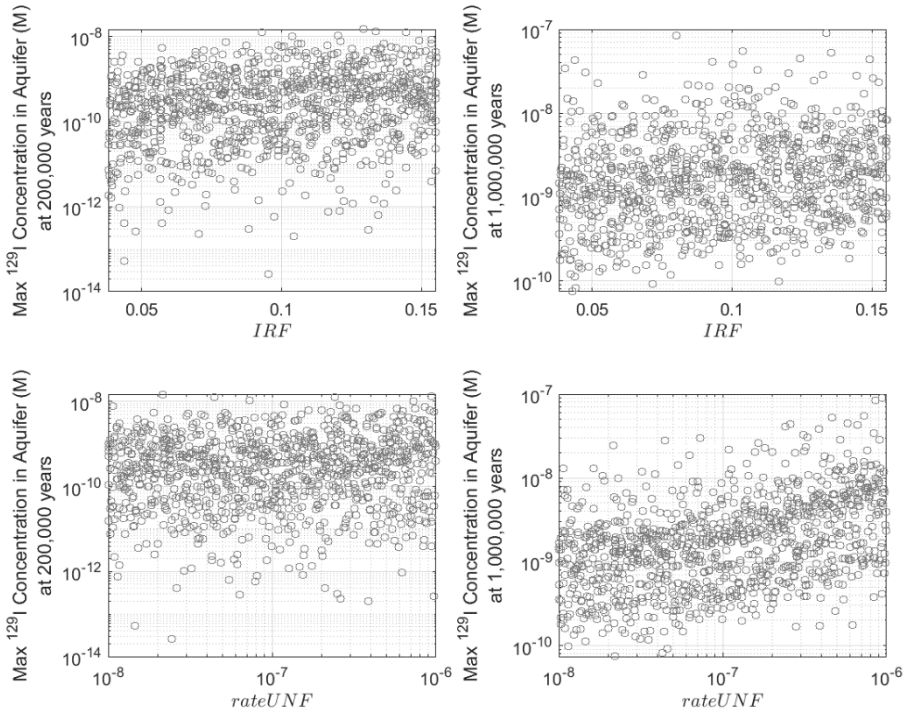


Figure 5-15. Scatterplots of the peak ^{129}I concentration [M] in the aquifer versus *IRF* and *rateUNF* at 200,000 years and at 1,000,000 years from the FDR simulations

There is no difference between sensitivity results from the ANN simulations and the FDR simulation for some of the QoIs. These QoIs are: the fraction of tracer remaining in the repository at 3,000 years, the fraction of tracer remaining in the repository at 1 Ma, the fractional mass flux from the repository at 3,000 years, and the median residence time of tracer in the repository. These results are not presented in this section because there is no change from the FY21 results [2].

For other QoIs, there were differences in the main index results between the ANN simulations and the FDR simulations, but this is a result of overfitting and does not represent a meaningful difference. The y-location of peak ^{129}I is a good example. Sensitivity results for this QoI are plotted in Figure 5-16, with corresponding scatterplots from in Figure 5-17. The sensitivity analysis results for the analyses without graph metrics for both sets of simulations (top) suggest that all of the uncertain variables are important. However, the large total effects indices and small main effects indices for all variables suggest overfitting. When graph metrics are included in either analysis (bottom), they dominate the sensitivity analysis results, which is more consistent with the scatterplots in Figure 5-17 where the horizontal striations (banding) in the scatterplots represent the results per DFN. We saw the same behavior of apparent overfitting for QoIs dominated by spatial heterogeneity for the ratio between the rock to aquifer flux and the rock to east boundary flux at 3,000 years and at 1 Ma (not pictured).

The sensitivity analysis results for the rock to aquifer flux at 3,000 years normalized by the rock to aquifer flux at 1 Ma and for the x-location of peak ^{129}I are similar in that the results are largely unchanged, except from overfitting effects. In these cases, however, there are still important uncertain variables; spatial heterogeneity does not fully dominate the results. The sensitivity analysis results for the rock to aquifer flux at 3,000 years normalized by the rock to aquifer flux at 1Ma are plotted in Figure 5-18. The importance of *IRF* is essentially the same for the ANN and FDR simulations. *rateUNF* appears to be significant in the FDR results, but only has a notable total effects index along with all of the other variables, suggesting this is an artifact of overfitting. The sensitivity analysis results for the x-location of peak ^{129}I are plotted in Figure 5-19. Here, the situation is similar except that the importance of both *IRF* and *rateUNF* appear to be artifacts of overfitting, whereas *kGlacial* has some importance that can be seen in scatterplots (Figure 5-20). *kGlacial* appears to have a slightly larger main effect in the ANN simulations than in the FDR simulations, though this is not clear in the scatterplots. The banding in the scatterplots with *kGlacial* is due to differences between the spatial realizations, but individual bands can be identified for which there is a positive trend with *kGlacial* in both the ANN and FDR realizations.

As was the case for the peak ^{129}I concentration, the sensitivity analysis results for the time of peak ^{129}I and the mean travel time at peak ^{129}I differ significantly between the ANN and FDR realizations due to the importance of *rateUNF* for these QoIs. The sensitivity analysis results for the time of peak are plotted in Figure 5-21, with supporting scatterplots in Figure 5-22. The mean waste package degradation rate is important regardless of the FMD model choice. In the FDR simulation results, *IRF* has almost no importance and this is supported by the scatterplot which shows no discernable trend between *IRF* and the time of peak ^{129}I. These sensitivity results suggest that *rateUNF* is important for the FDR simulations, and this is corroborated by the scatterplot, which shows the variance in time of peak ^{129}I decreasing with increasing *rateUNF*. In the ANN results, however, there is a slight trend with respect to *IRF*, with the time of peak ^{129}I having more variation as *IRF* increases.

The sensitivity analysis results for the mean travel time at peak ^{129}I are plotted in Figure 5-23 with supporting scatterplots in Figure 5-24. These sensitivity analysis results are nearly identical to those for the time of peak ^{129}I, with the same relationships, because these QoIs are highly correlated to each other. The correlation coefficient between these QoIs is 0.97 in the FDR simulations and 0.95 in the ANN simulations.

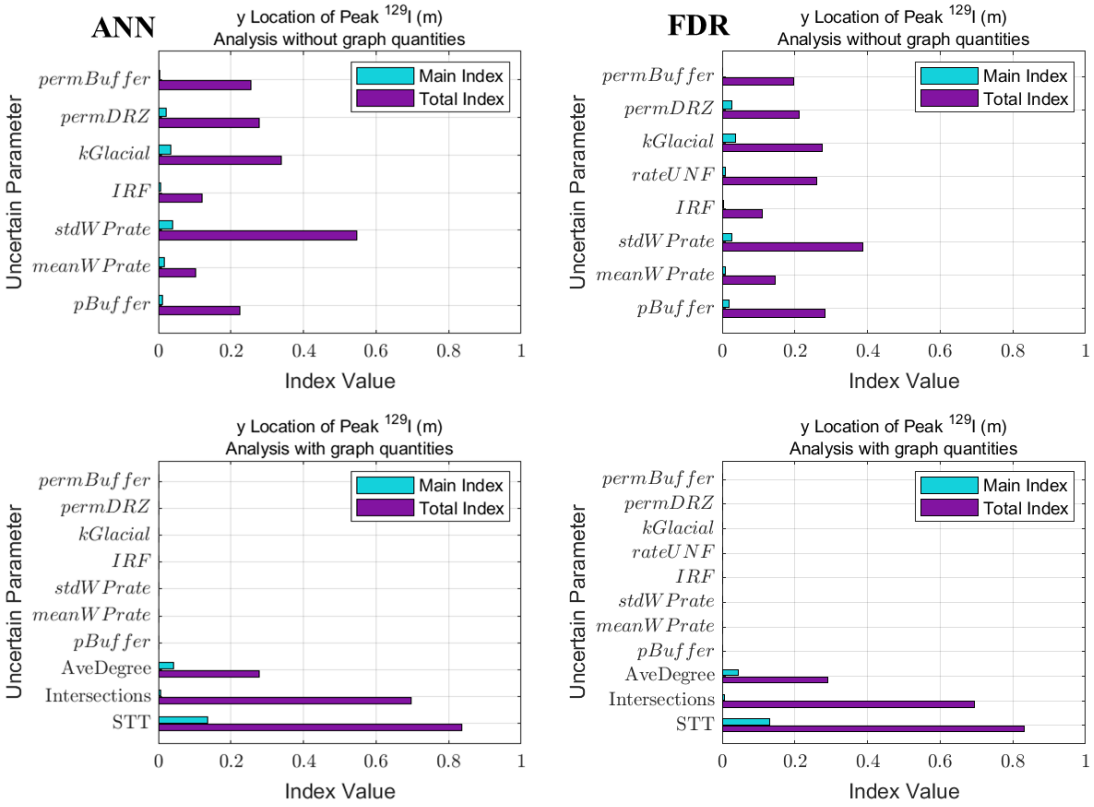


Figure 5-16. Sensitivity analysis results for the y-location [m] of peak ^{129}I from the ANN simulations (left) and the FDR simulations (right)

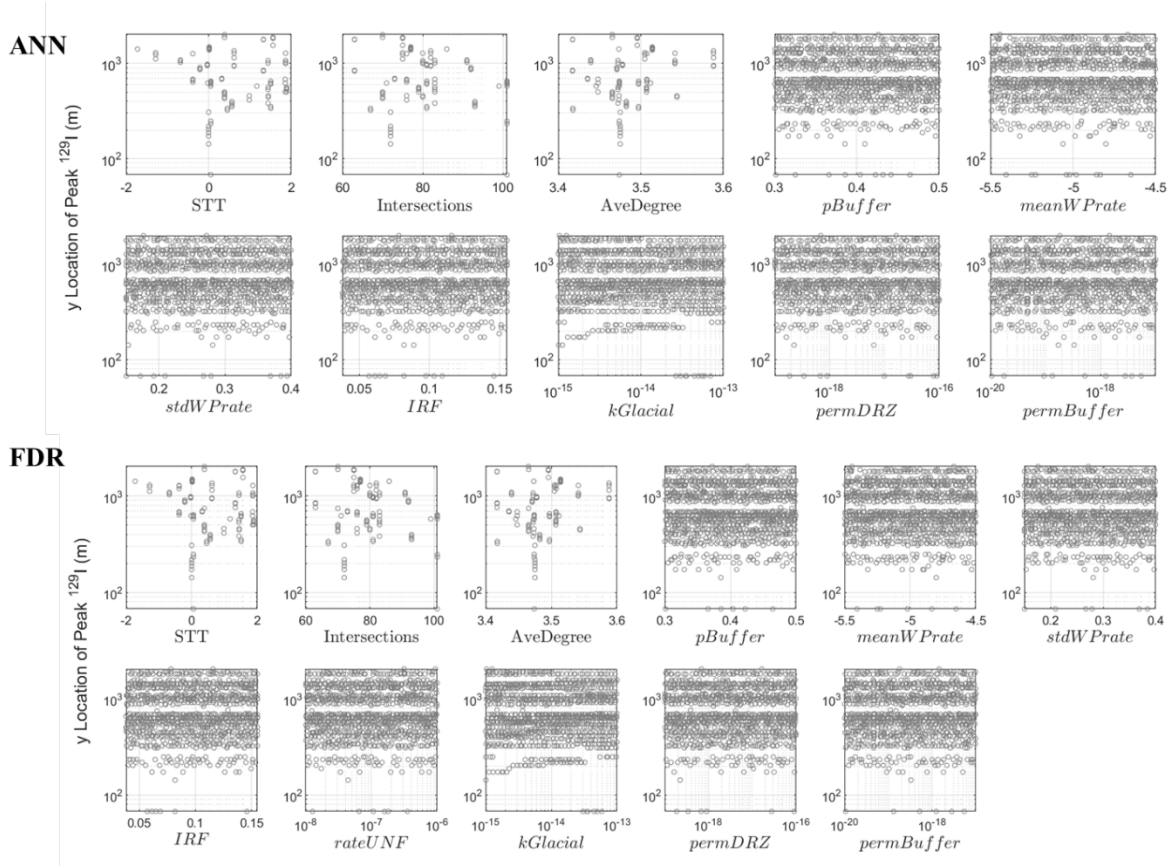


Figure 5-17. Scatter plots of the y-location [m] of peak ^{129}I versus the graph metrics and the uncertain parameters for the ANN simulations (top) and the FDR simulations (bottom)

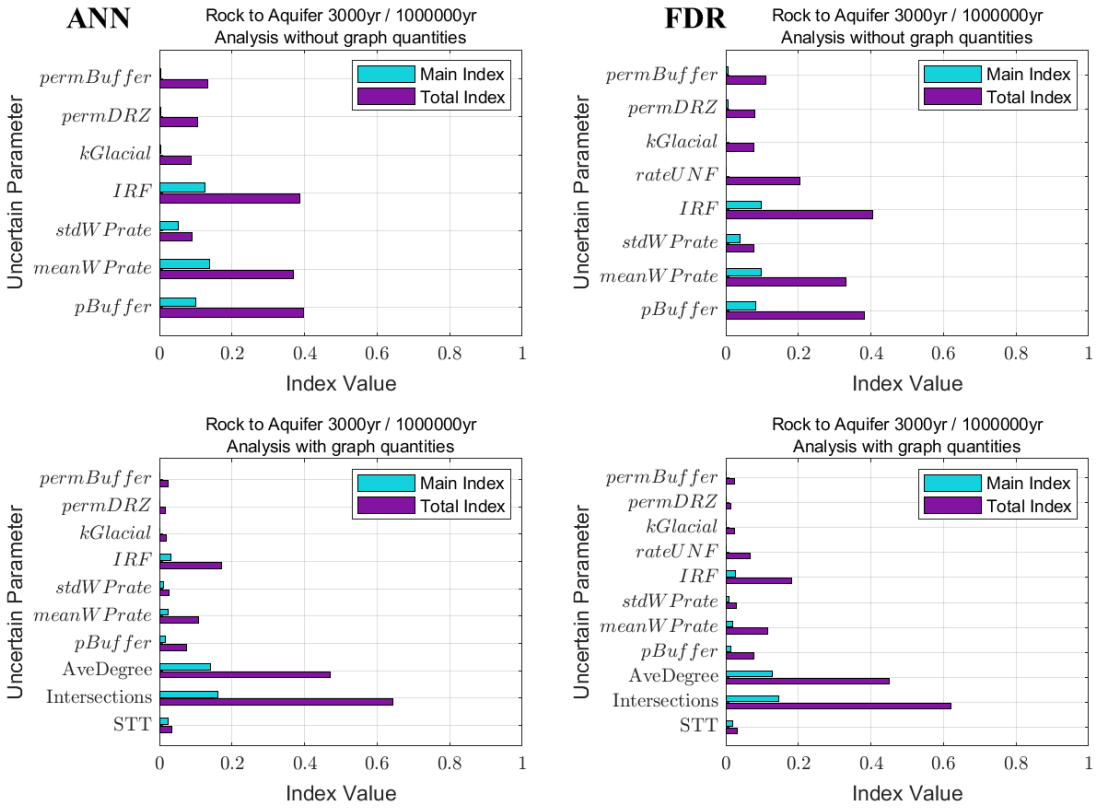


Figure 5-18. Sensitivity analysis results for the ratio between the rock to aquifer flux at 3,000 years and at 1,000,000 years from the ANN simulations (left) and the FDR simulations (right)

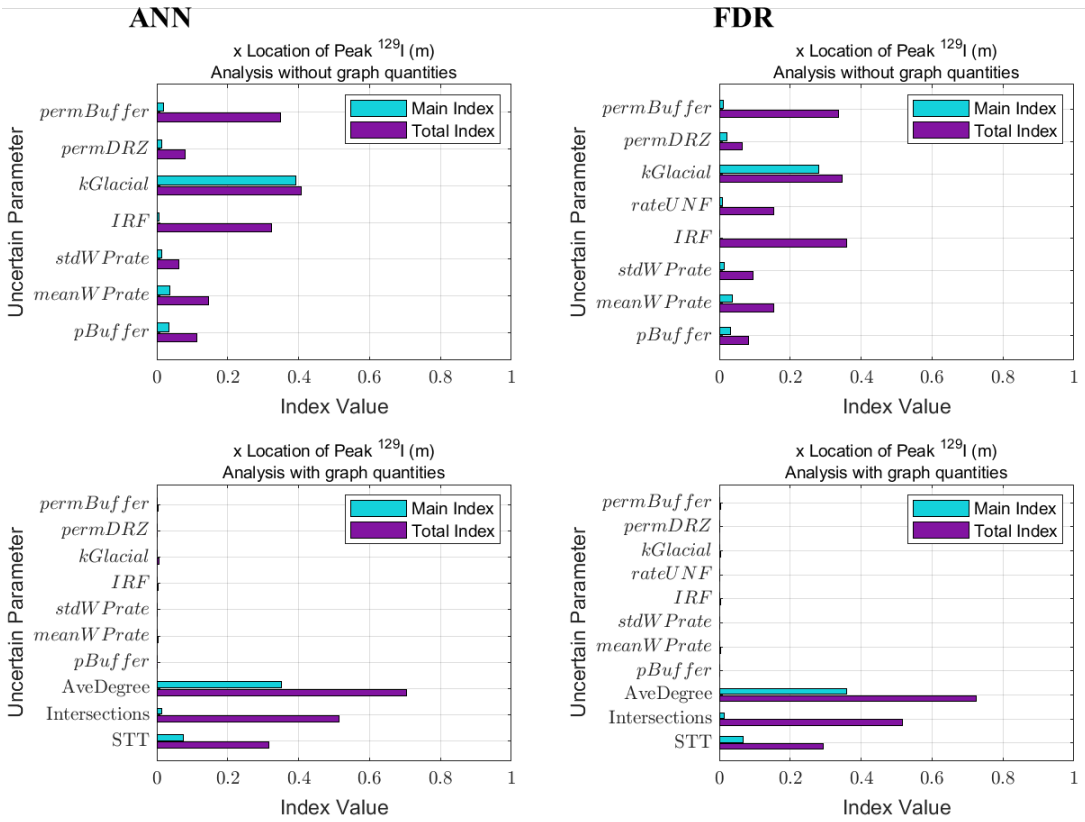


Figure 5-19. Sensitivity analysis results for the x-location [m] of peak ^{129}I from the ANN simulations (left) and the FDR simulations (right)

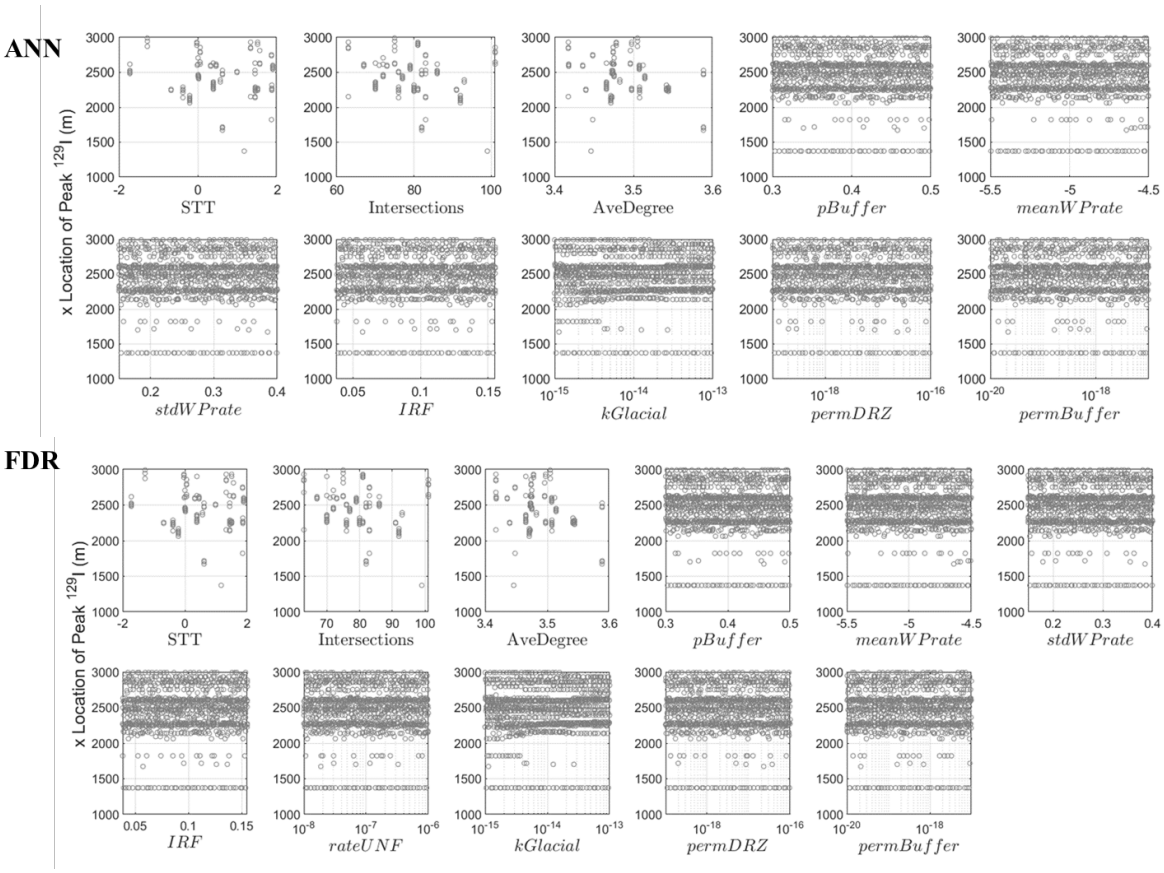


Figure 5-20. Scatter plots of the x-location [m] of peak ^{129}I versus the graph metrics and the uncertain parameters for the ANN simulations (top) and the FDR simulations (bottom)

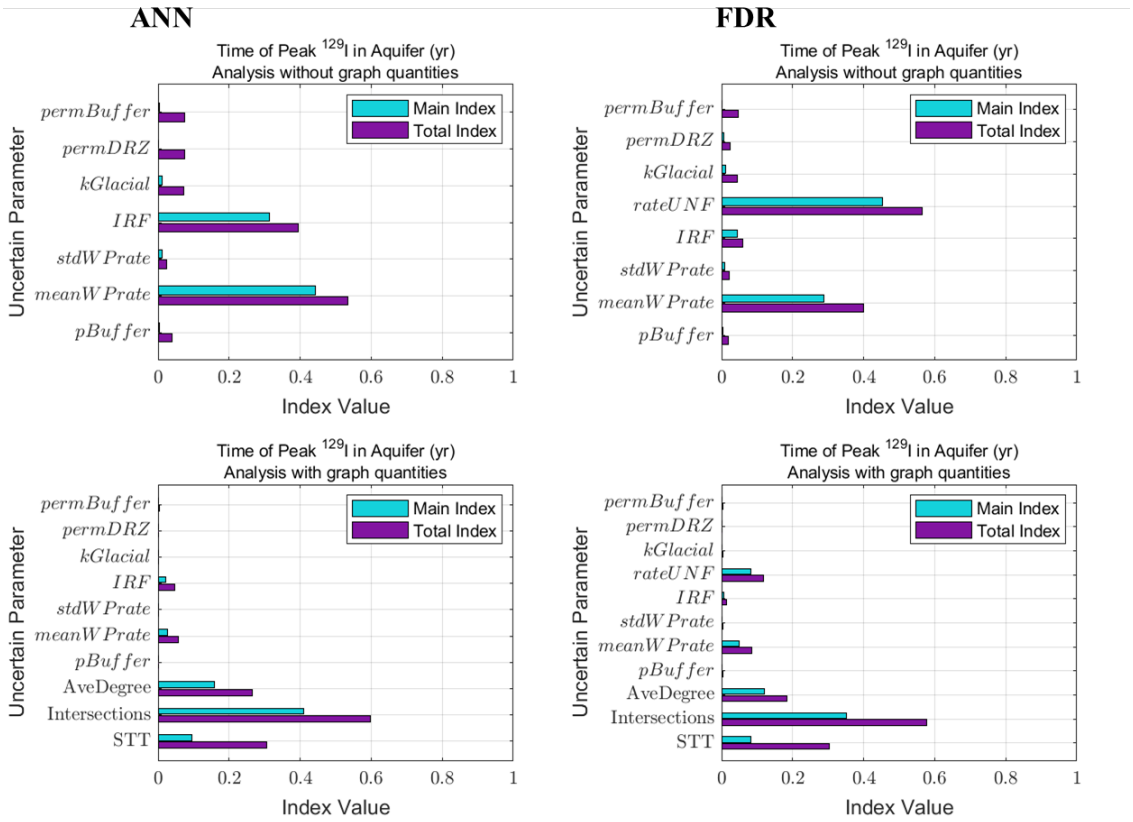


Figure 5-21. Sensitivity analysis results for the time [yr] of peak ^{129}I from the ANN simulations (left) and the FDR simulations (right)

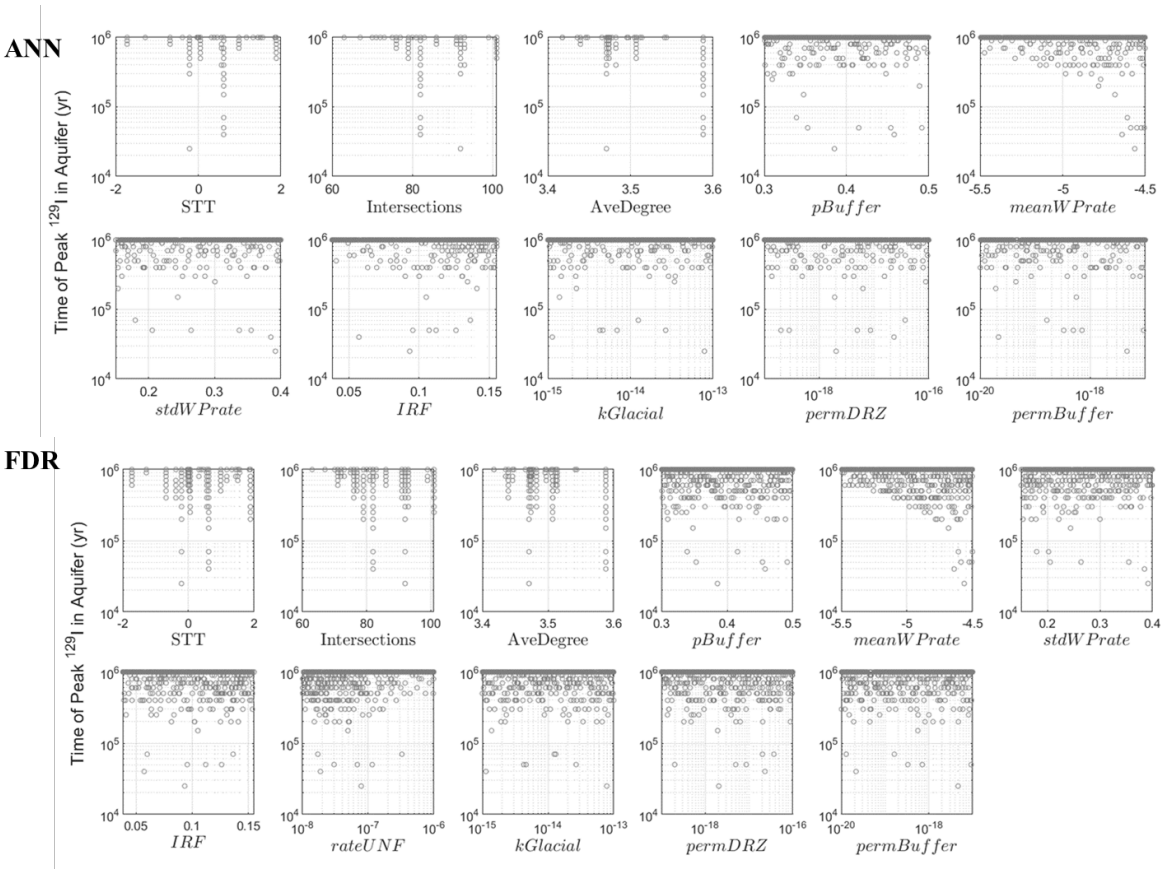


Figure 5-22. Scatter plots of the time [yr] of peak ^{129}I versus the graph metrics and the uncertain parameters for the ANN simulations (top) and the FDR simulations (bottom)

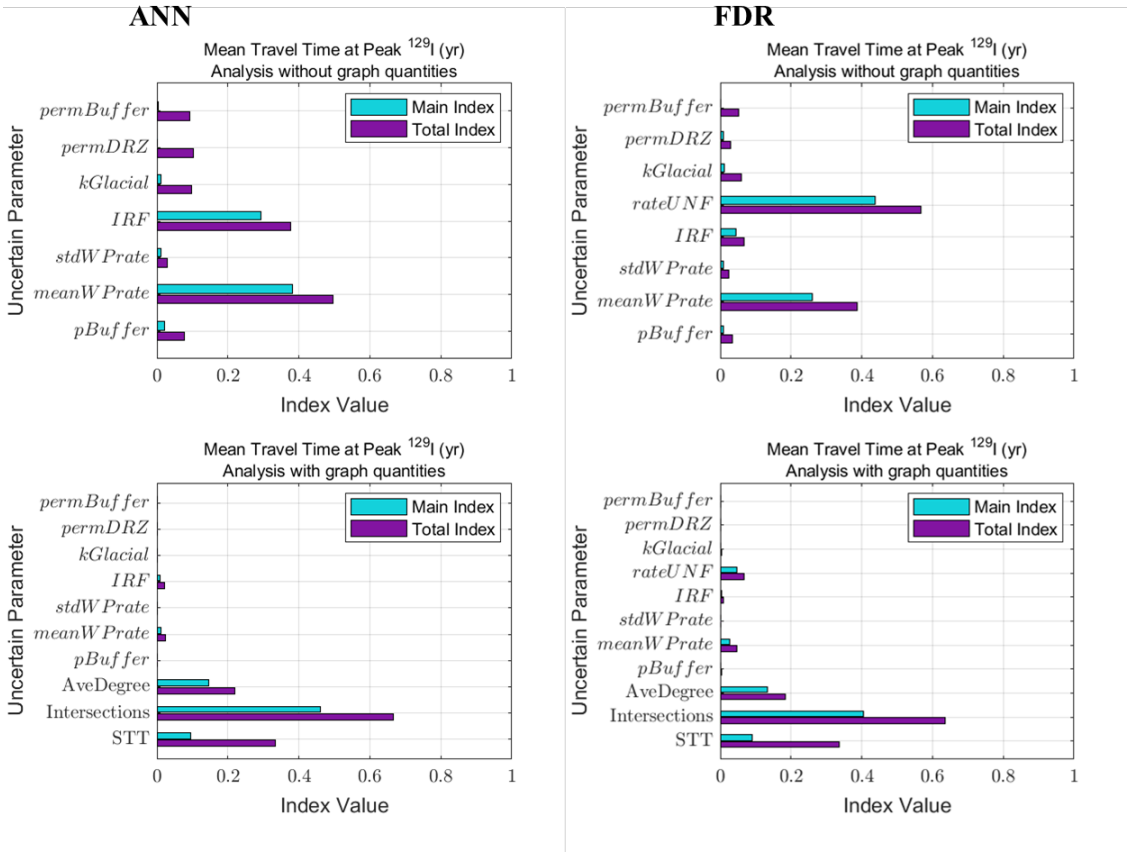


Figure 5-23. Sensitivity analysis results for the mean travel time [yr] at peak ^{129}I from the ANN simulations (left) and the FDR simulations (right)

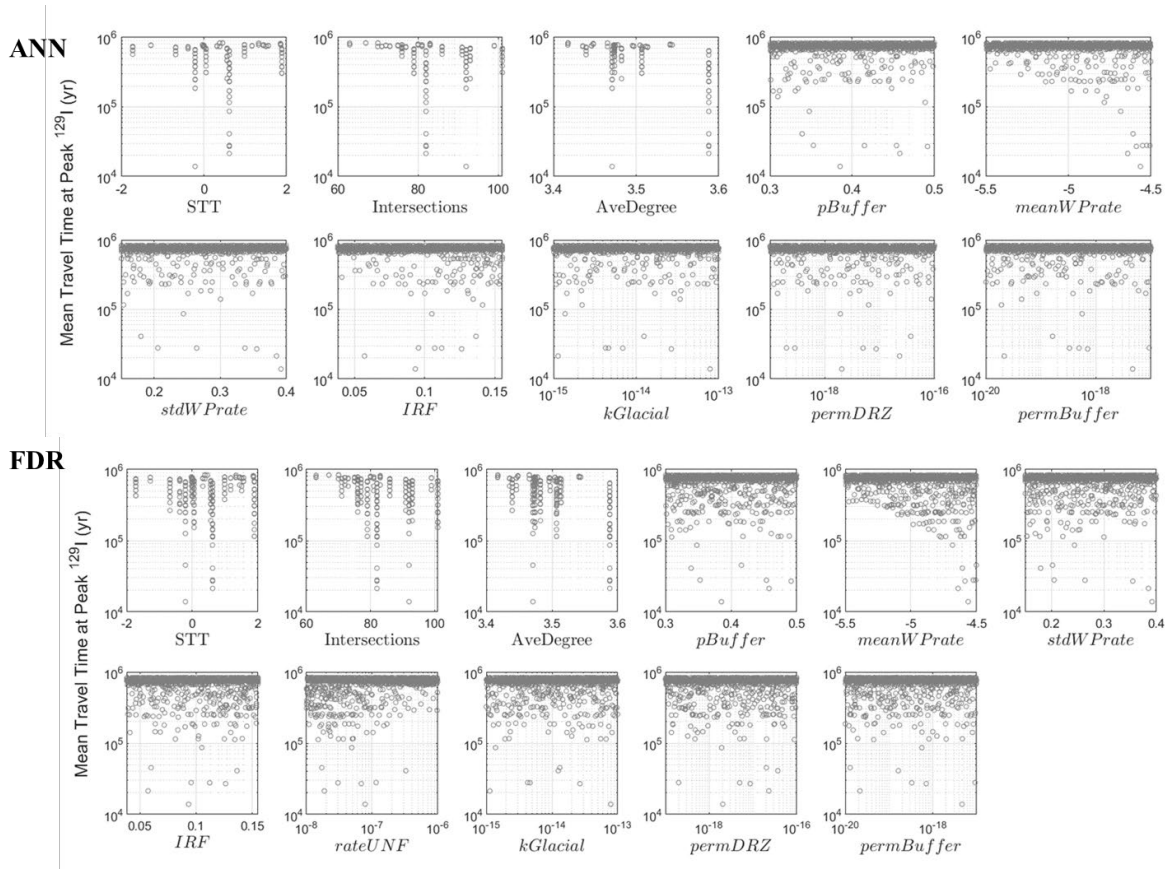


Figure 5-24. Scatter plots of the mean travel time [yr] at peak ^{129}I versus the graph metrics and the uncertain parameters for the ANN simulations (top) and the FDR simulations (bottom)

5.3.2 Spatial Realization Results

As detailed in Section 3, we use surrogate models to overcome computational cost when estimating Sobol' indices for sensitivity analysis. The quality of such estimates, however, depends on the quality of the surrogate model, and spatial heterogeneity presents a significant challenge for surrogate models that rely on parametric characterizations of uncertainty. Though graph metrics have enabled us to incorporate some spatial heterogeneity in sensitivity analyses, the graph metrics are incomplete tools that, necessarily, cannot fully summarize all important spatial factors.

To better understand the extent to which spatial heterogeneity affects surrogate model construction, we fit individual surrogate models to each spatial realization. That is, we trained a separate second order polynomial chaos expansion using the 40 samples for each spatial realization. The R^2 values were calculated for each of the models, and this is the metric we used to assess surrogate quality. This analysis was performed with the ANN simulations and focuses on peak ^{129}I since it is one of the most challenging QoIs to predict accurately with surrogate models (see Section 3).

The sensitivity analysis results per spatial realization for peak ^{129}I concentration are plotted in Figure 5-25. The R^2 values for each of the surrogate models are shown on the right-hand side of the figure. This shows that the surrogate model can account for the majority of the variance in peak

^{129}I concentration when spatial heterogeneity is excluded from the data and the surrogate construction; the average R^2 is about 0.95. Though the surrogate model built over all 1,000 simulations (Figure 5-10) does not account for as much of the variance as each individual surrogate in Figure 5-25, it does identify *kGlacial*, *IRF*, and *meanWPrate* as the dominant parameter uncertainties, so the conclusions regarding ranking of parametric uncertainties are the same.

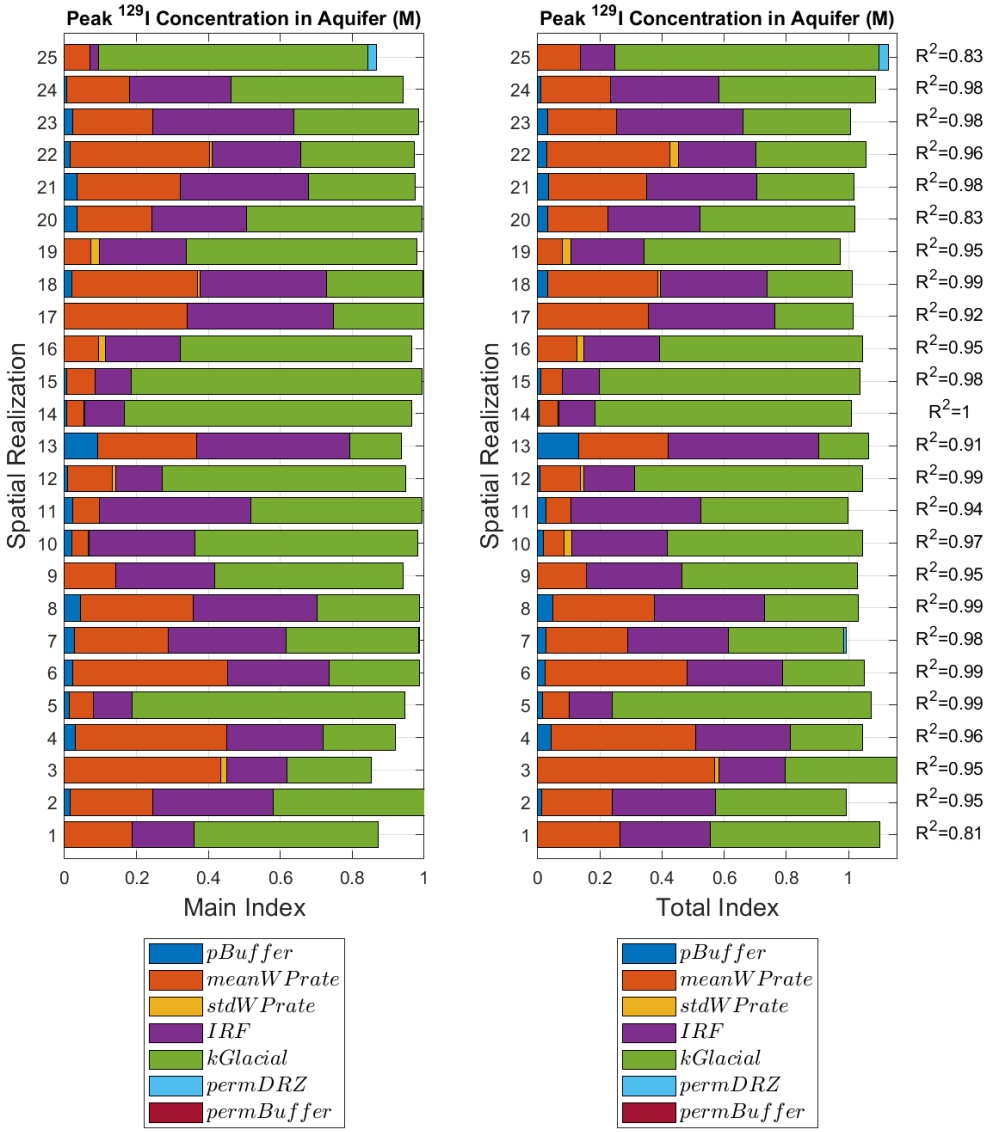


Figure 5-25. Sensitivity analysis results per-spatial realization for the peak ^{129}I concentration [M] in the aquifer, for the ANN model case

The high quality of these surrogate model fits supports our conclusions from Chapter 3 that spatial heterogeneity, rather than scale differences or non-monotonic relationships, presents the most significant challenge for our use of surrogate modeling SA. Differences between the SA

results for each spatial realization demonstrate that spatial heterogeneity has a significant impact on this QoI, and that the spatial heterogeneity interacts with the parameter uncertainties.

The analysis for mean travel time also demonstrates the importance of spatial heterogeneity. The sensitivity analysis results for mean travel time for each spatial realization are plotted in Figure 5-26. The most important variable differs between spatial realizations, with *pBuffer*, *meanWPrate*, *IRF*, and *kGlacial* all dominating for some of the realizations. The R^2 values for each model are, on average, fairly high. There are seven realizations for which the model R^2 is less than 0.8, and ten realizations for which the model R^2 is less than 0.9. Scatterplots of the mean travel time versus the input variable for each spatial realization are plotted in Figure 5-27 (*pBuffer*), Figure 5-28 (*IRF*), Figure 5-29 (*meanWPrate*) and Figure 5-30 (*kGlacial*). Highlighting in the plots identifies the spatial realizations for which that variable was the most important in the sensitivity analysis. Red markings in Figure 5-27 and Figure 5-28 denote important points for comparison as discussed below.

In some cases, scatterplots look very similar for two different spatial realizations, but the surrogate model did not rank the same variable as the most influential. This is the case for realizations 14 and 23 in Figure 5-28 (*IRF*), for example. The plots for these two realizations have the same y-axis and the plots look nearly identical, however *IRF* is the top ranked parameter for realization 14 but only the 3rd most important variable for realization 23. The scatterplots for realizations 14 and 23 versus *pBuffer* may give some insight into why the sensitivity results for these realizations differ. In Figure 5-27, there is more spread in the points that deviate from the dominant trend for realization 14 than for realization 23 (see the red regions in the figures). This slight difference in points that deviate from the main trend is enough to substantially change the sensitivity analysis results, even though the slope of the main trend between *pBuffer* and MTT is stronger in both realizations than the slope of the main trend between *IRF* and MTT (see the red arrows in the figures). For this particular spatial realization, the sensitivity analysis results are more influenced by a few points than by the trend that the majority of the points follow.

This comparison does not mean that the sensitivity analysis results are incorrect; it is important to remember that the sensitivity analysis method is a variance decomposition, not just a regression model. Because the points with lower MTT are concentrated towards lower values of *pBuffer* for realization 23 (Figure 5-27), there is higher variance in the QoI at low *pBuffer* values than at high *pBuffer* values. For realization 14, the lower MTT points are spread over more of the range of *pBuffer* so it is not clear that we would expect more variance in the MTT at certain values of *pBuffer*. The sensitivity results for these realizations make sense, but the sensitivity to such a small number of points raises an important question of whether the behavior of points that deviate from the main trend is influenced by differences in the DFNs or whether it is a result of our limited epistemic sample size. This is another important point for consideration as we seek to improve our methods for incorporating spatial heterogeneity into sensitivity analysis.

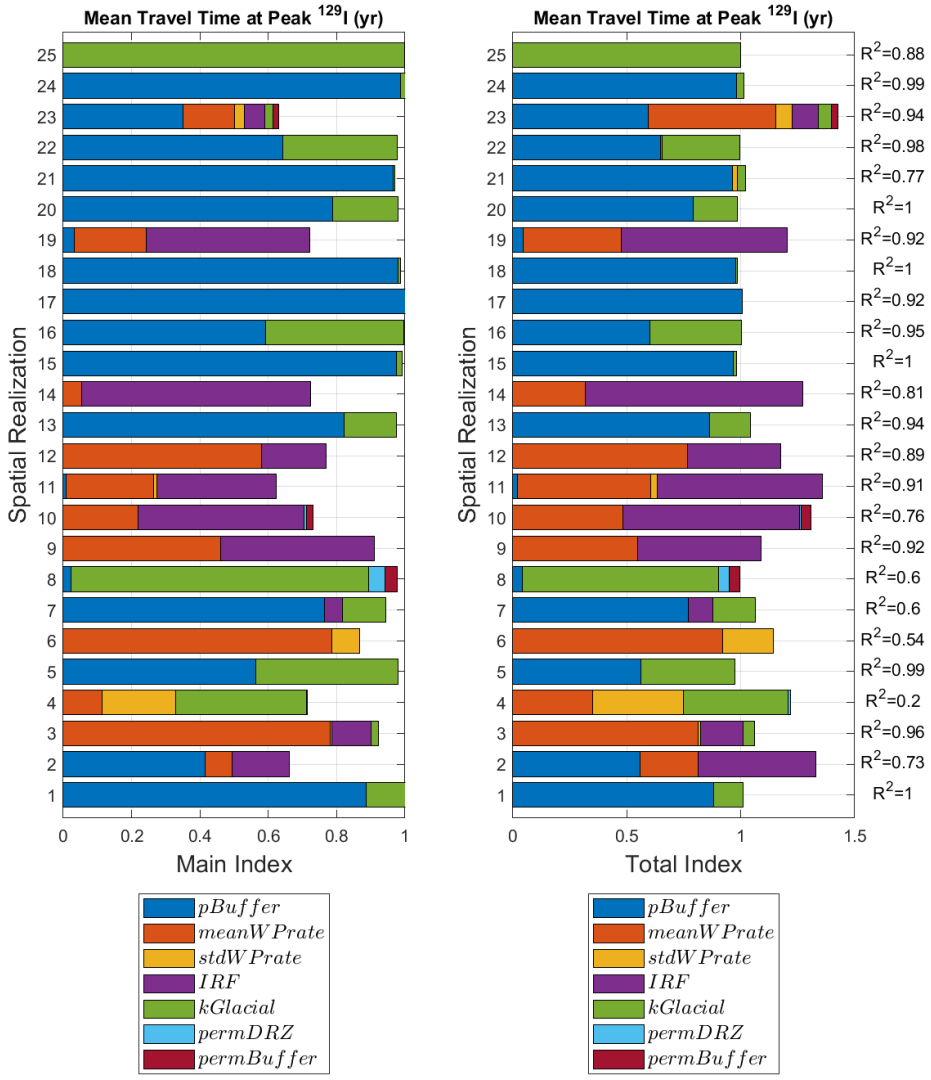


Figure 5-26. Sensitivity analysis results per-spatial realization for the mean travel time [yr] at peak ^{129}I concentration in the aquifer, for the ANN model case

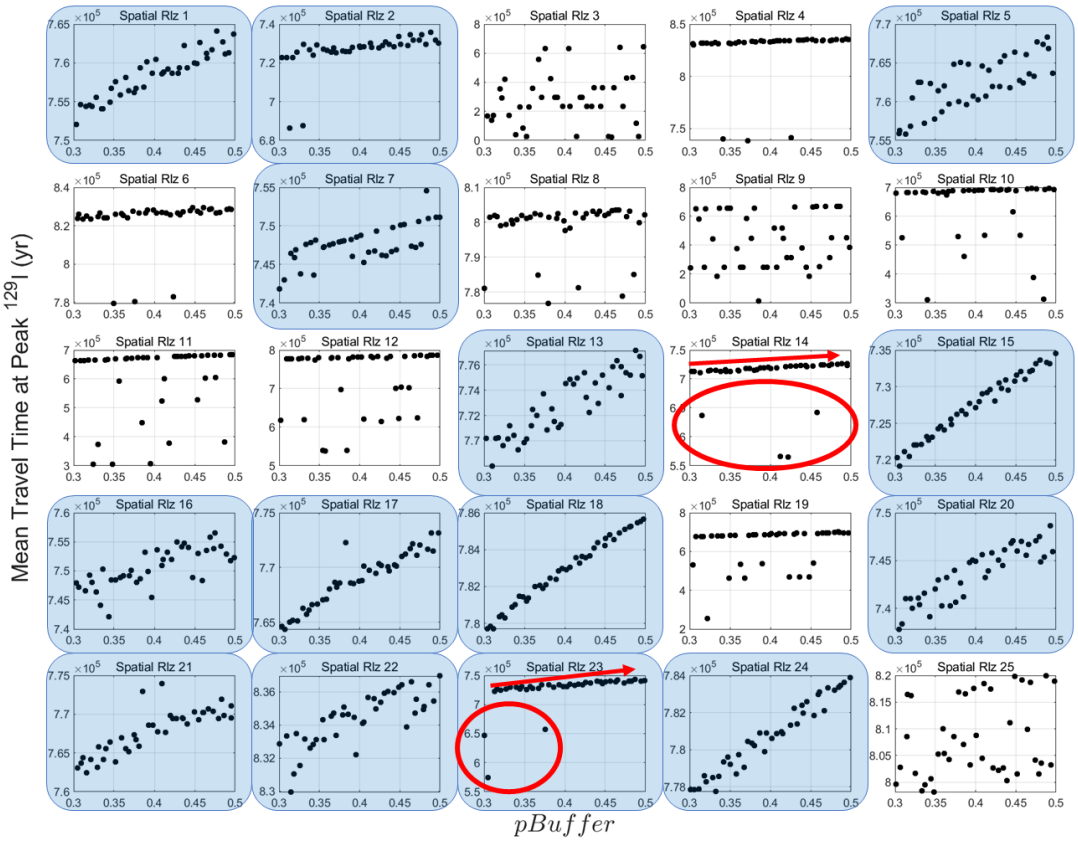


Figure 5-27. Scatterplots of the mean travel time [yr] for each spatial realization versus pBuffer; blue highlighting indicates that the surrogate model ranked pBuffer as the dominant parametric uncertainty for that spatial realization and the red marks identify important comparisons

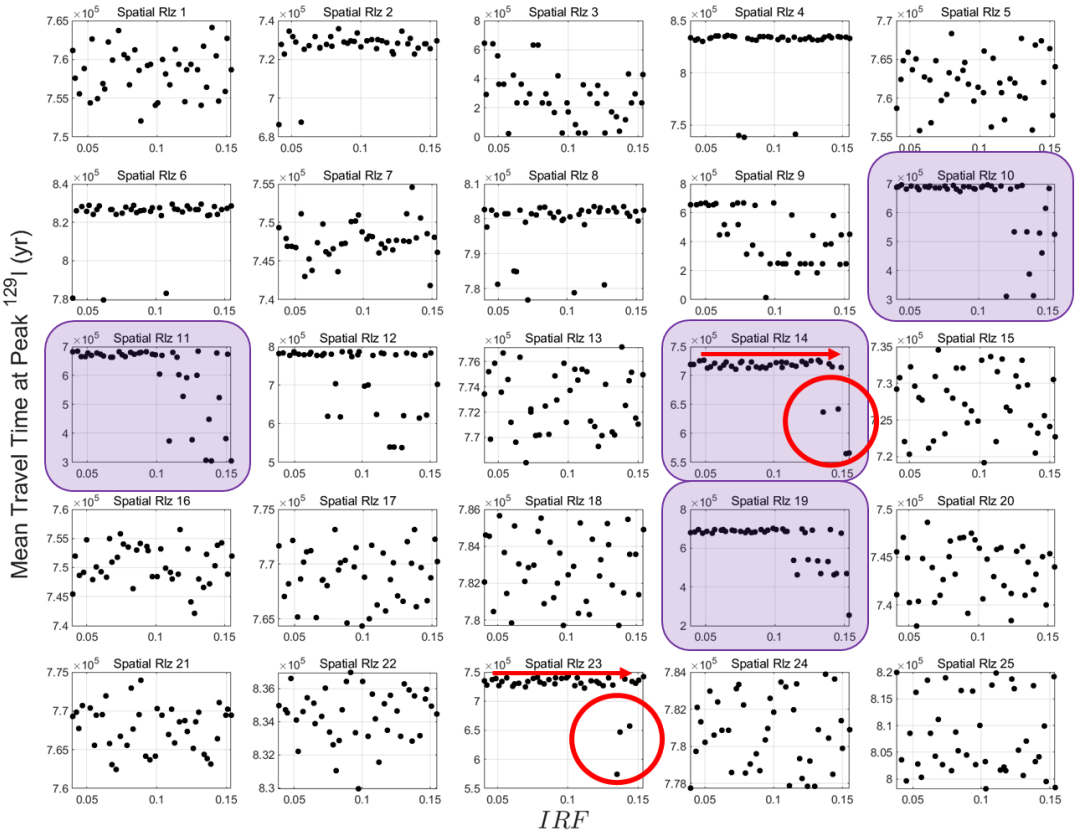


Figure 5-28. Scatterplots of the mean travel time [yr] for each spatial realization versus *IRF*; highlighting indicates that the surrogate model ranked *IRF* as the dominant parametric uncertainty for that spatial realization and the red marks identify important comparisons

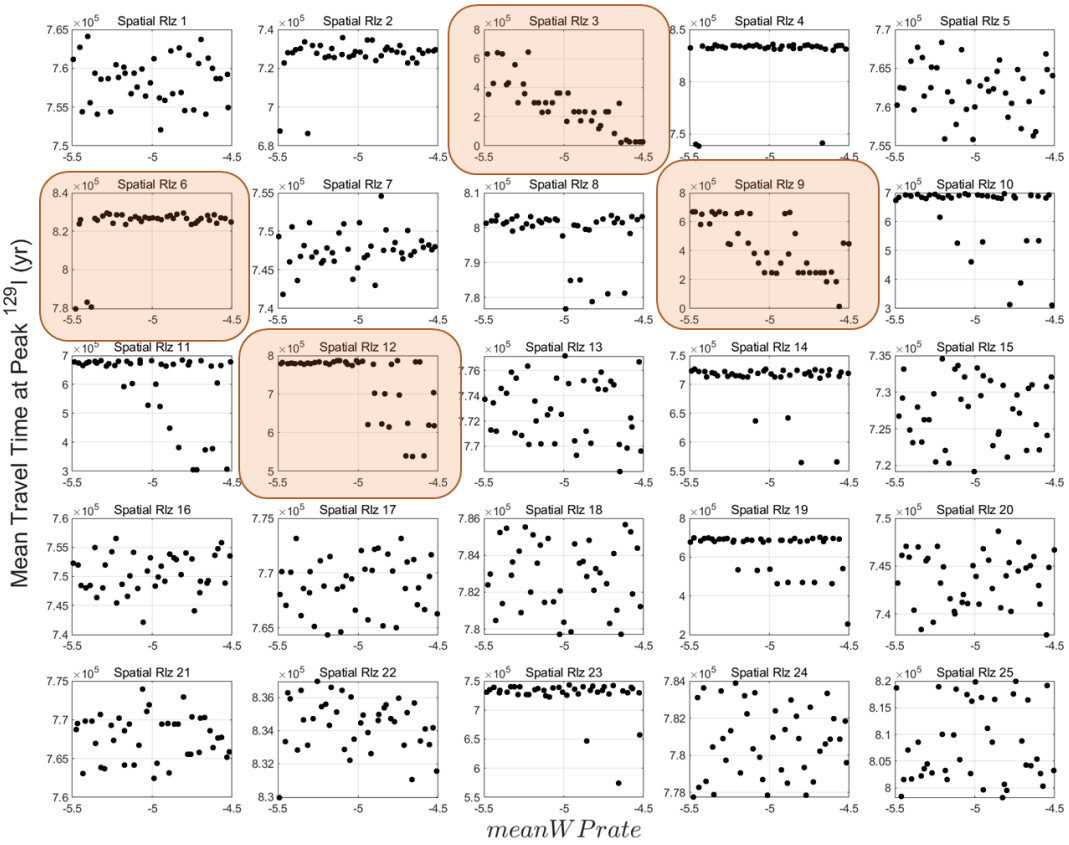


Figure 5-29. Scatterplots of the mean travel time [yr] for each spatial realization versus *meanWPrate*; highlighting indicates that the surrogate model ranked *meanWprate* as the dominant parametric uncertainty for that spatial realization

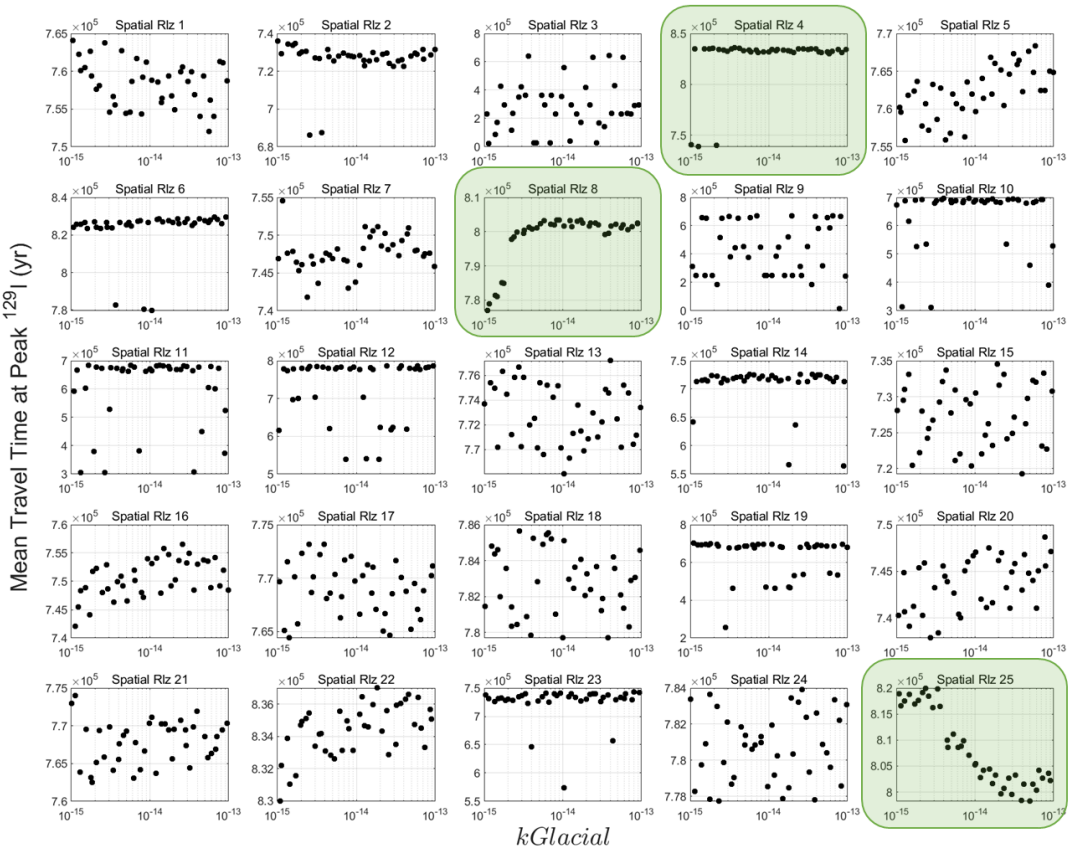


Figure 5-30. Scatterplots of the mean travel time [yr] for each spatial realization versus *kGlacial*; highlighting indicates that the surrogate model ranked *kGlacial* as the dominant parametric uncertainty for that spatial realization

5.4 Summary of Results

Comparison between the sensitivity analyses for the crystalline reference case simulations with either the FDR alternative model or the ANN alternative model for FMD showed that the choice of FMD model alternative has a minimal effect on QoIs. Some differences were detected in the maximum ^{129}I concentrations in the aquifer over time, with the FDR model alternative leading to more variation between simulations near 1 million years. The increase in variation is likely due to the incorporation of uncertainty in the FDR model via *rateUNF*; no comparable parametric uncertainty is currently included in the ANN model. When not attributable to *rateUNF*, differences in sensitivity analysis results between simulations using either FMD alternative model were due to overfitting behavior in the surrogate models. This overfitting behavior was shown to occur with all of the surrogate model options investigated in Chapter 3 and is most likely to occur at smaller sample sizes, such as the 1,000 model evaluations we can feasibly perform with this full-scale crystalline reference case model.

The effect of FMD model choice was also small because of the significance of the instantaneous release fraction (*IRF*) up until around 200,000. Time-dependent sensitivity analysis showed that, among the parametric uncertainties, the waste package degradation dominates early, then *IRF* becomes significant, and *rateUNF* only gains importance after most waste packages have breached. Both analyses included the same uncertainties for *IRF*, *meanWPrate*, and *stdWPrate* (as well as spatial heterogeneity), so the dominance of these parameters for much of the simulation resulted in similar uncertainties in the QoIs towards the end of the simulation. *kGlacial* also gained some importance towards the end of the simulation, but this was more pronounced in the simulations with the ANN surrogate; *rateUNF* is significantly more important than *kGlacial* in the simulations with the FDR model alternative.

Future analysis may incorporate additional uncertainty into the ANN surrogate model so that it is more directly comparable to the FDR model. Additionally, the *IRF* uncertainty is fairly high, so reductions in that uncertainty could also influence how early the FMD model alternative becomes important. However, uncertainty in this variable is not purely epistemic. *IRF* should vary, for example, due to variability in fuel burnup, so there is a limit to how much this uncertainty can be reduced in a PA.

SA was performed on the simulations that used the ANN surrogate for each spatial realization separately. This analysis was not aimed at comparing the effects of FMD model choice; it was motivated by the results of Section 3, which identified spatial heterogeneity as the dominant factor preventing high accuracy in SA. Comparison of SA results for the peak ^{129}I concentrations for each spatial realization showed that, in general, the surrogate model can characterize more of the variance in QoIs when spatial heterogeneity is absent. Differences in the SA results from these high-quality surrogates between spatial realizations showed that spatial heterogeneity has a significant impact on the behavior of ^{129}I in simulations and interacts significantly with the parameter uncertainties. This conclusion motivates further advancement to our treatment of spatial heterogeneity to improve SA accuracy in performance assessment with the GDSA Framework.

6. DECOVALEX CASE

In this chapter, we present some early results from a sensitivity analysis of the DECOVALEX Task F crystalline case. The Task F case has been documented in [37]: we present a brief summary below. The results presented here differ from other analyses in that this is a sensitivity analysis: we have varied five parameters of the DECOVALEX case and run 300 samples. We show these preliminary results to communicate the generality of the *GDSA Workflow*, which was used to generate the 300 runs (10 DFNs x 30 epistemic samples). We caveat these results because they are very preliminary: we wish to investigate the model and results in more detail, so we caution against over-analyzing these results. However, we present these initial results as a demonstration of what might be possible with additional investigations of the DECOVALEX case and to generate ideas for subsequent analyses.

6.1 Description of case study

In the DECOVALEX Task F crystalline case, the system of comparison is a generic repository for commercial spent nuclear fuel (SNF) in a fractured crystalline host rock. The reference case repository is located beneath a gently sloping hill in a domain 5 km in length, 2 km in width, and ~1 km in depth (Figure 6-1). Fracture intensity and fracture transmissivity decrease with depth. The repository is located in the west (left) side of the domain at a depth of approximately 450 m, comprises 50 deposition drifts branching off two parallel access tunnels. The deposition drifts are spaced 40 m center-to-center; 50 deposition holes within each tunnel are spaced 6 m center-to-center. The area of lowest elevation is located on the east (right) side of the domain, representing the location where water would collect at the surface forming a feature such as a lake or wetland.

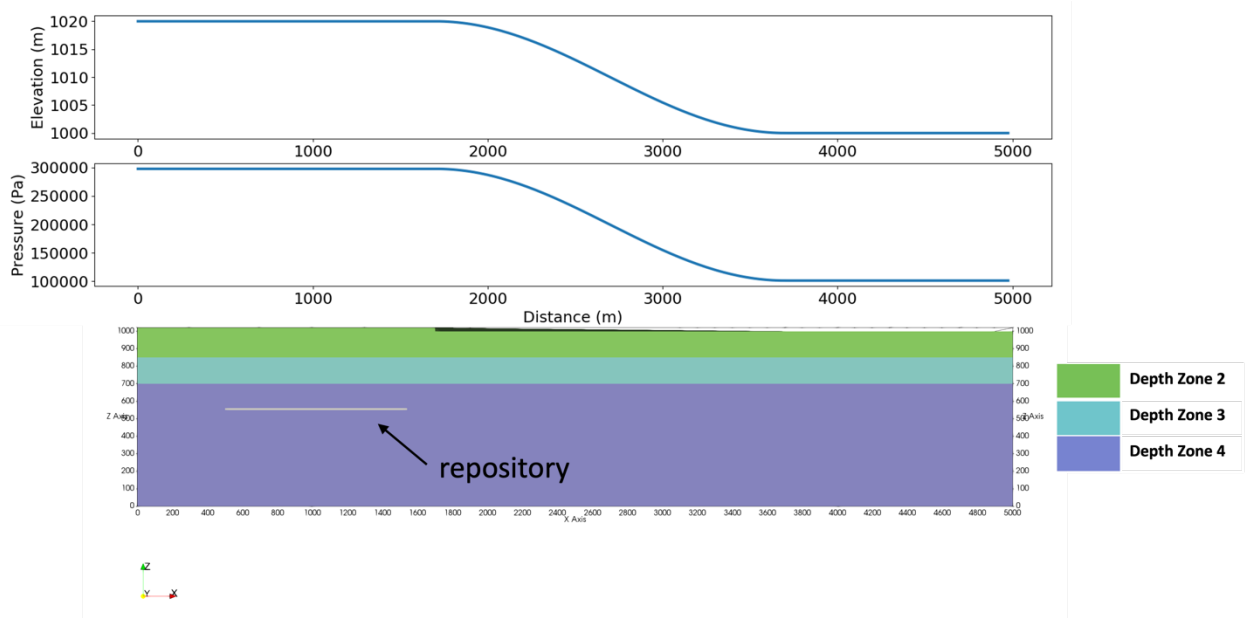


Figure 6-1. Elevation profile and corresponding pressure boundary condition (top) and depth zones in the domain (bottom)

The crystalline host rock consists of a set of large-scale, highly fractured brittle deformation zones (deterministic fractures) and intervening masses of competent rock containing sparse networks of connected fractures (stochastic fractures). Stochastic and deterministic fractures were generated using Los Alamos National Laboratory's (LANL) software dfnWorks. Ten sets of the stochastic fracture network were created. Fractures were then upscaled to an Equivalent Continuous Porous Medium via the Python script mapdfn.py. The far field fractures were upscaled to a 25 m length and the fractures in the repository were then upscaled to a 25/3 m length.

The initial reference case assumes steady state flow and transient transport of two conservative tracers (Tracer 1 and 2) upon simultaneous breach of all the canisters in the repository. The inventory per waste package of Tracer 1 is 0.545 g (0.00423 moles) and 100% of the mass is instantly released at the start of the transport simulation. The inventory per waste package of Tracer 2 is 4.90 g (0.038 moles) and is released at a fractional rate of 10^{-7}/y throughout the transport simulation. Transport of the two tracers is simulated for 100,000 years. The tracer advects out of the top boundary while no-flow boundary conditions are applied to all other faces. Three different surfaces of interest were defined at the top of the domain (z=1000 m): the surface of the high point (0 m < x < 1700 m), the surface of the hillslope (1700 m < x < 3700 m), and the surface of the low point (3700 m < x < 5000 m) (Figure 6-2). Preliminary results for the cumulative mass flow out of the low point can be seen in Figure 6-3 for the 10 realizations.

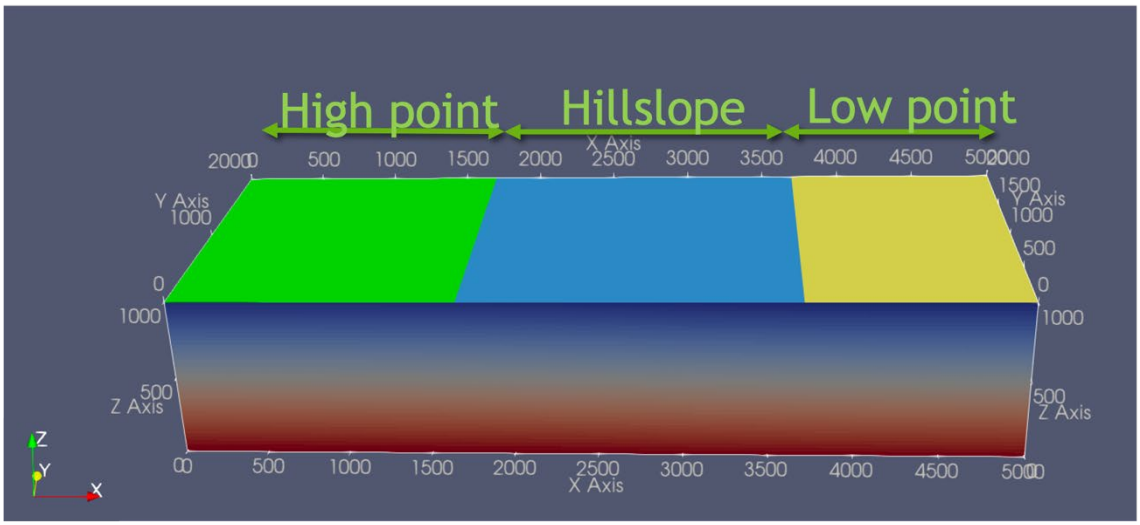


Figure 6-2. Surfaces of interest for initial reference case

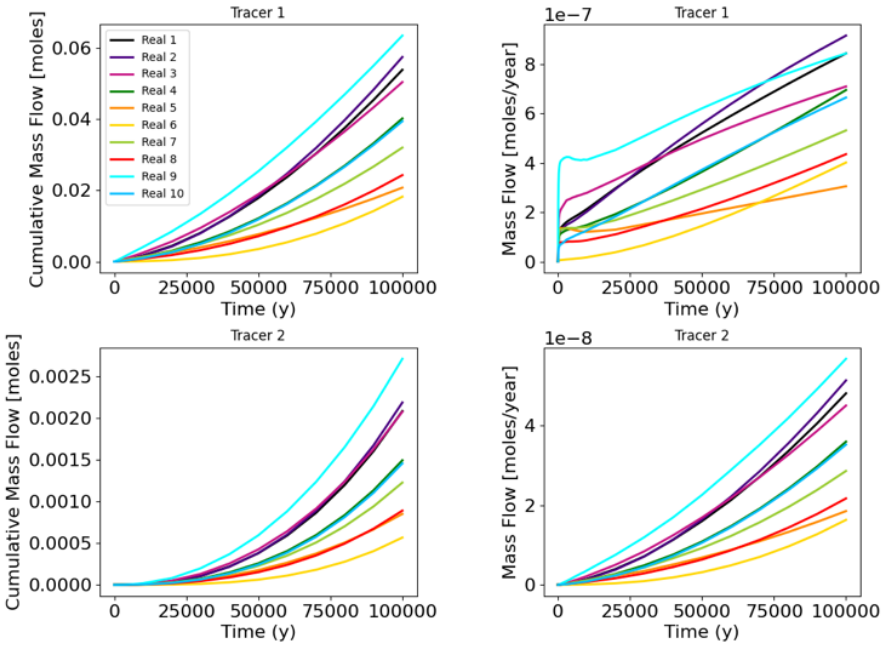


Figure 6-3. Cumulative mass flow (left) and mass flow (right) across the low point for the 10 realizations

6.2 Results

Five parameters were varied for this study. The parameter descriptions, ranges, units, and distributions are shown in Table 6-1.

Table 6-1. Parameter Descriptions for the DECOVALEX Sensitivity Analysis Case Study

Input	Description	Range	Units	Distribution
pBuffer	Buffer porosity	0.41 – 0.466	-	Uniform
pBackfill	Backfill porosity	0.3 – 0.5	-	Uniform
permBuffer	Buffer permeability	10^{-22} – 10^{-20}	m^2	Log uniform
permBackfill	Backfill permeability	10^{-22} – 10^{-19}	m^2	Log uniform
rateUNF	Fractional dissolution rate of spent nuclear fuel	10^{-8} – 10^{-6}	yr^{-1}	Log uniform
DFN	Particular realization of a Discrete Fracture Network	1-10		

Seven QoIs were analyzed for this study. Each of these was output at both 50,000 and 100,000 years, for a total of fourteen QoIs. The seven QoIs are shown in Table 6-2. Note that the first three QoIs (the water fluxes at different surface locations) had the same values at 50K and 100K years for each DFN. For example, the low_point_water flux had the same values per DFN at 50K years and 100K years. This is because the water flux remained the constant over the simulation.

Table 6-2. Quantities of Interest for DECOVALEX Sensitivity Analysis Case Study. Note that each QoI was calculated at 50K years and 100K years

Quantity of Interest	Description
Low_point Water	Water flux reported over the low point surface of the domain (kg/yr)
HillSlope Water	Water flux reported over the hill slope surface of the domain (kg/yr)
High_point water	Water flux reported over the high point surface of the domain (kg/yr)
Low_point Tracer 1	Mass flow of tracer 1 over the low point surface (mol/yr)
Low_point Tracer 2	Mass flow of tracer 2 over the low point surface (mol/yr)
HillSlope Tracer 1	Mass flow of tracer 1 over the hill slope surface (mol/yr)
HillSlope Tracer 2	Mass flow of tracer 2 over the hill slope surface (mol/yr)

We now present some initial results. As mentioned above, these are very preliminary, but show some interesting behavior. Correlations of the output QoIs (shown in the rows) and the inputs (shown in the columns) are presented in Table 6-3, which shows that four of the inputs, the porosity and permeability of both the buffer and backfill, have near zero correlation with any of the outputs. This means that the variation seen in these outputs is solely due to variation in the DFNs. The only parameter that has a significant correlation is *rateUNF*, and that is only with respect to the Tracer 2 concentrations. This is expected behavior, as Tracer 1 is released instantly upon waste package failure; the fractional dissolution rate of spent nuclear fuel is irrelevant for Tracer 1.

Table 6-3. Correlations of Input Parameters (columns) and QoIs (rows) for the DECOVALEX case study. Significant correlations highlighted in yellow

Variable Name	*pBuffer*	*pBackfill*	*permBuffer*	*permBackfill*	*rateUNF*
low_pointWater_kg_y_50kyr	0.004	0.001	0.003	0.000	-0.010
low_pointWater_kg_y_100kyr	0.004	0.001	0.003	0.000	-0.010
low_pointTracer1_mol_y_50kyr	0.003	0.003	0.000	0.006	0.002
low_pointTracer1_mol_y_100kyr	0.004	0.003	-0.002	0.001	-0.002
low_pointTracer2_mol_y_50kyr	0.011	0.005	0.008	-0.004	0.849
low_pointTracer2_mol_y_100kyr	0.009	0.009	0.006	-0.009	0.895
hillslopeWater_kg_y_50kyr	-0.002	0.003	-0.006	-0.004	0.001
hillslopeWater_kg_y_100kyr	-0.002	0.003	-0.006	-0.004	0.001
hillslopeTracer1_mol_y_50kyr	0.001	0.001	0.003	0.013	0.009
hillslopeTracer1_mol_y_100kyr	0.001	0.001	0.002	0.012	0.009
hillslopeTracer2_mol_y_50kyr	0.015	-0.009	0.012	0.022	0.668
hillslopeTracer2_mol_y_100kyr	0.014	-0.007	0.011	0.019	0.700
high_pointWater_kg_y_50kyr	-0.001	-0.004	0.003	0.004	0.008
high_pointWater_kg_y_100kyr	-0.001	-0.004	0.003	0.004	0.008

Scatterplots are now presented for various QoIs vs. inputs. The purpose of these scatterplots is to verify the correlation data shown above and to better understand the behavior of the model as the parameters vary.

Scatterplots of the input parameters vs. the low point water fluxes are shown in Figure 6-4. Note that the horizontal lines colored for each DFN indicate that per DFN, these low point water fluxes are constant as the input parameters vary across their ranges: this indicates that the parameters have negligible effect on these outputs.

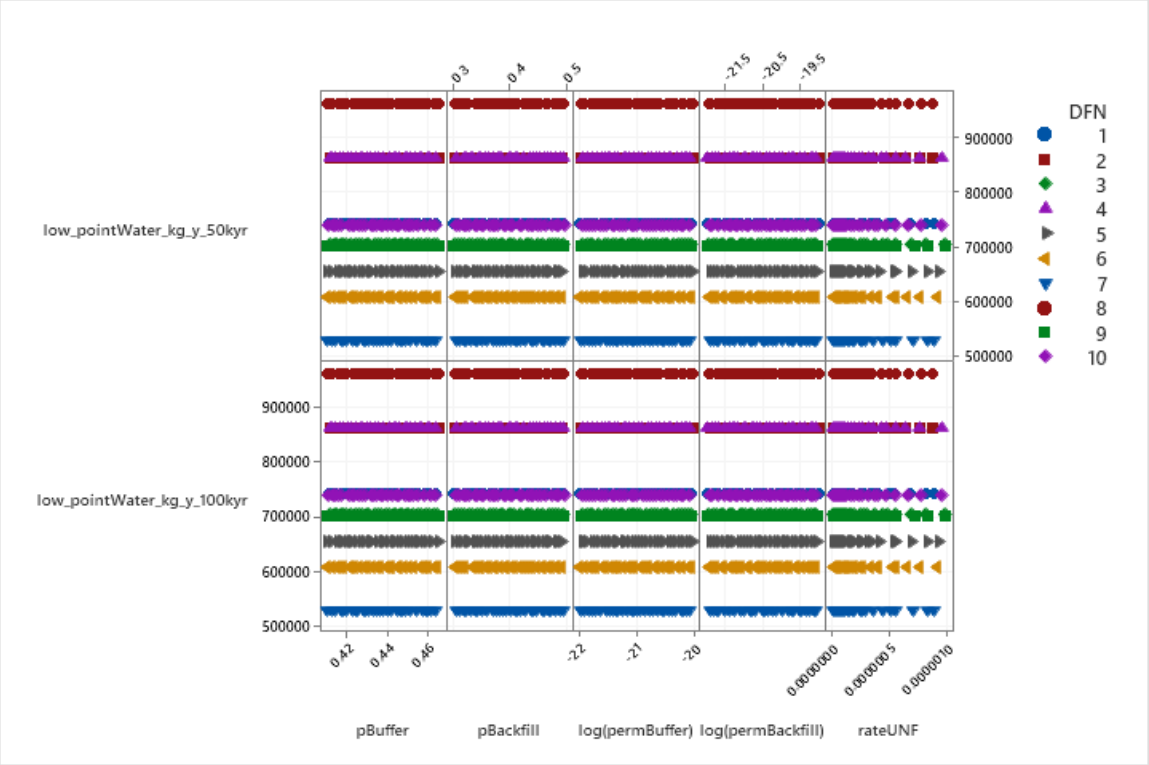


Figure 6-4. Scatterplots of inputs vs. low point water fluxes, colored by DFN

Similar scatterplots of the input parameters vs. the hillslope water and high point water fluxes are shown in Figure 6-5 and **Figure 6-6**. Note that the horizontal lines colored for each DFN indicate that per DFN, these low point water fluxes are constant as the input parameters vary across their ranges: this indicates that the parameters have negligible effect on these outputs.

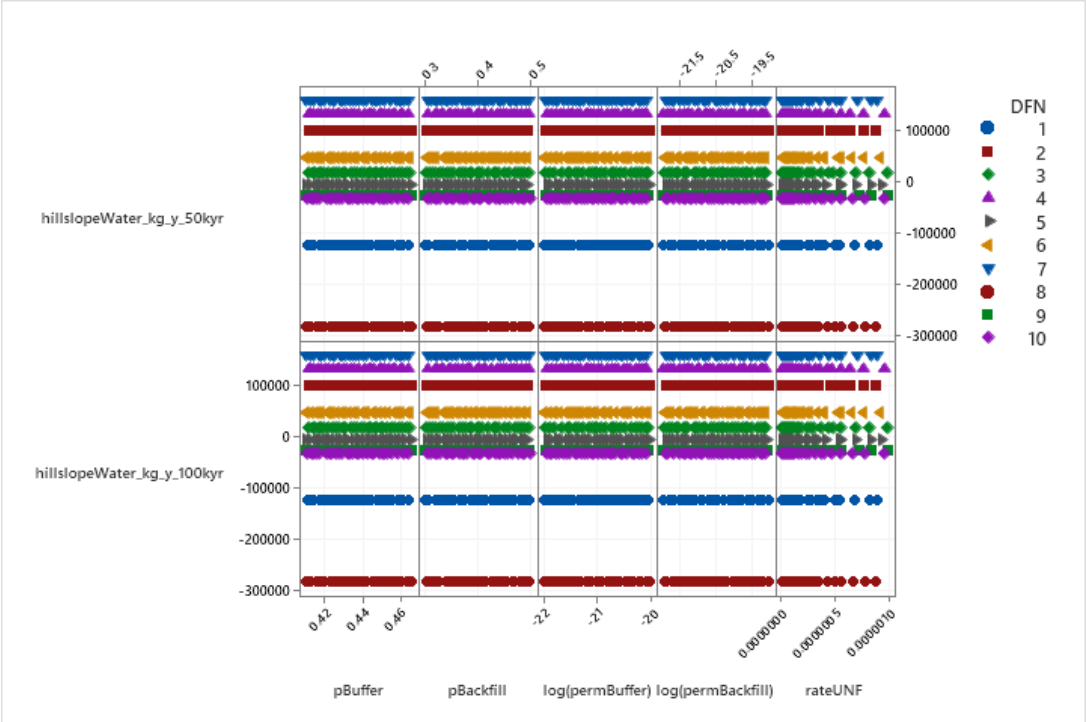


Figure 6-5. Scatterplots of inputs vs. hillslope water fluxes, colored by DFN

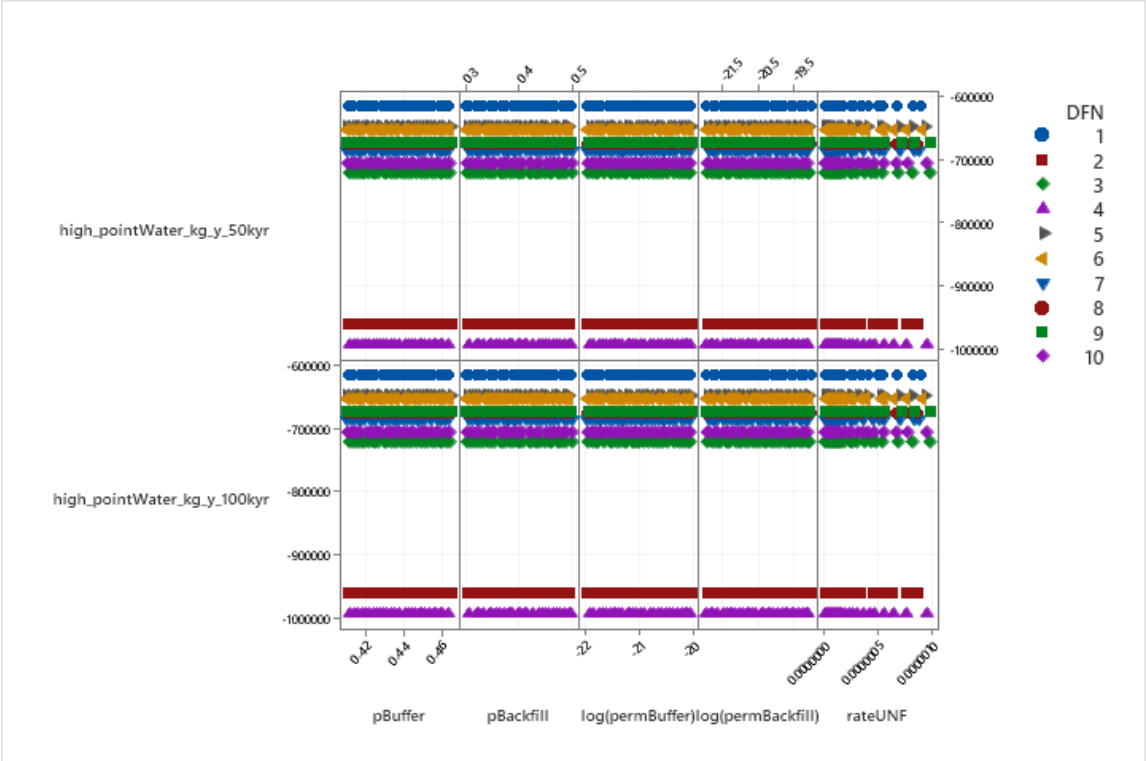


Figure 6-6. Scatterplots of inputs vs. high point water fluxes, colored by DFN

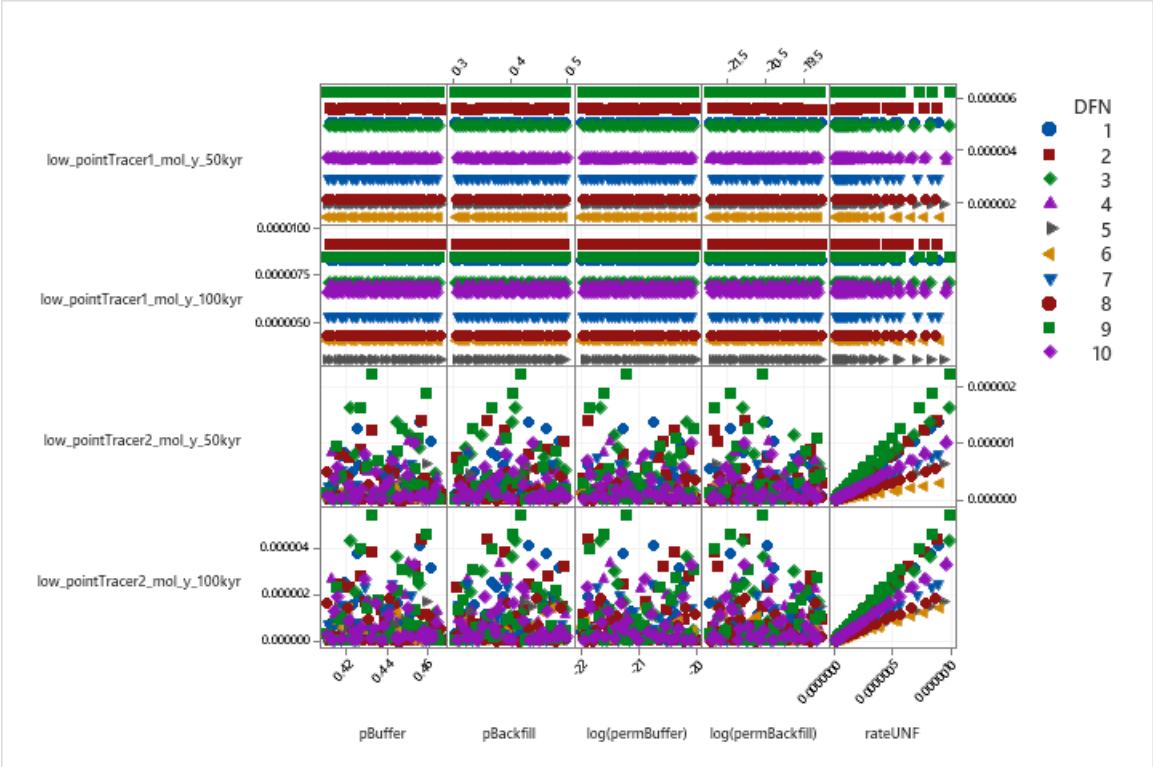


Figure 6-7. Scatterplots of inputs vs. low point tracer concentrations, colored by DFN

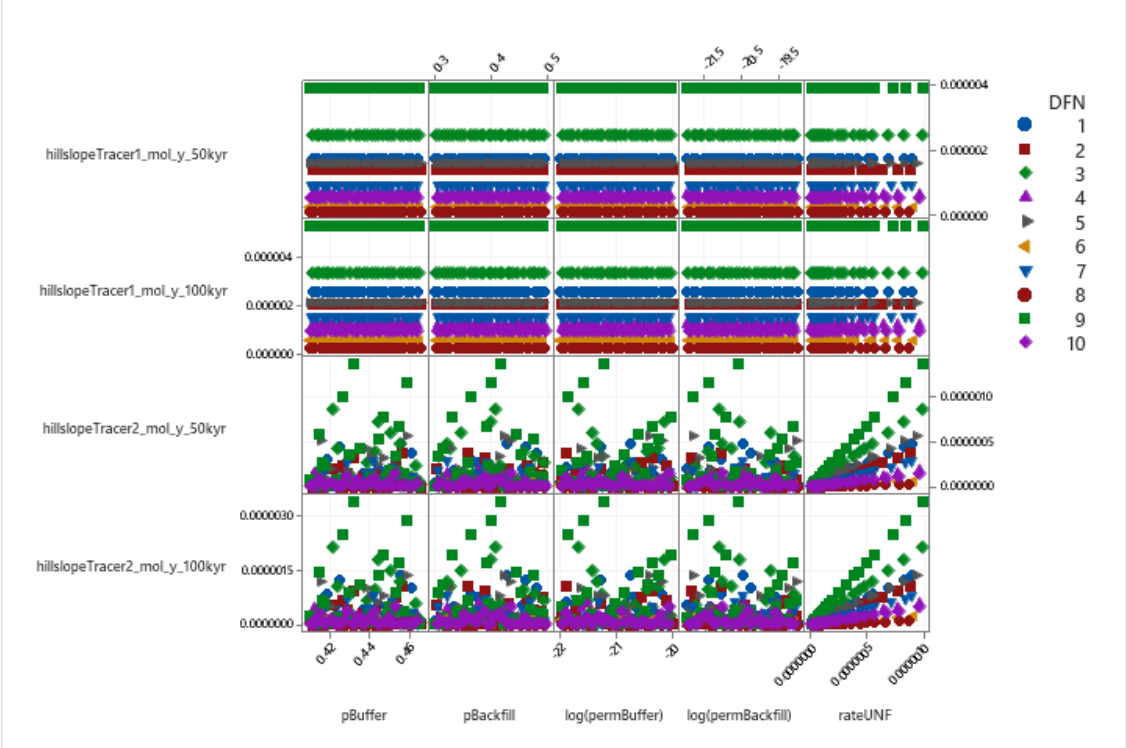


Figure 6-8. Scatterplots of inputs vs. hill slope tracer concentrations, colored by DFN

Scatterplots of the input parameters vs. the tracer concentrations at both the low point and the hillslope are shown in Figure 6-7 and Figure 6-8, respectively. Note that the horizontal lines colored for each DFN indicate that per DFN, the input parameters do not have an effect on Tracer 1 concentrations either at the low point or at the hillslope. However, some of the input parameters DO have a significant effect on Tracer 2 concentrations at both these locations. In particular, note that the Tracer 2 concentration (per DFN) is nearly linear with respect to the *rateUNF* parameter for both locations. This is corroborated by the strong correlations listed in Table 6-3.

7. CONCLUSIONS

This report covers the main topics addressed in the GDSA UQ/SA work performed in 2022.

The report presents the results of a new uncertainty quantification method involving models at multiple levels of fidelity applied to a simplified version of the crystalline reference case. This simplified reference case retains many of the key qualities of the full reference case. The multifidelity analysis showed that the variance in a key Quantity of Interest, peak ^{129}I concentration, could be reduced by almost an order of magnitude using a set of runs at different fidelities which had the same cost as 500 high fidelity runs. This year, we explored multifidelity surrogate-based and sampling-based methods in the context of this simplified reference case. These methods show promise to improve the efficiency of uncertainty analyses for PA, but additional tuning of model ensembles and surrogate construction algorithms will be needed to realize their full potential.

We also examined the possible errors and biases introduced by the use of surrogates when calculating the Sobol' variance-based sensitivity indices. Surrogates are typically used to calculate these sensitivity indices given the large number of samples required to get good estimates of conditional expectations and variances. Chapter 3 presented a comprehensive case study, again using the simplified crystalline reference case. A Sobol' index calculation was performed for 7 variables involving 9000 sample runs, where each run involved a DFN generated by dfnWorks and PFLOTRAN. This entire workflow was repeated five times to generate five entire replicate sets. This allowed detailed comparison of surrogate models within and across replicates. It also allowed comparison of the sensitivity indices using only the samples and using surrogates based on the samples. We take the "samples only" case as the better estimate because the direct effect of DFNs can be obtained. Surrogates required a proxy metric for the DFN spatial heterogeneity to be calculated and used as an independent predictor in the surrogate.

We found that surrogate choice has some effect on the sensitivity analysis results, but it is minimal for most QoIs. All surrogate models included in these studies demonstrated overfitting behavior. A significant finding was that the calculation of the Sobol' indices based on samples only vs. using the surrogates resulted in significant difference in the actual sensitivity index values and in their ranking. Typically, the sample-only calculation of the sensitivity indices ranked the effect of spatial heterogeneity much higher than the surrogate-based calculations. This is likely due to proxies or metrics for the DFNs that do not completely capture their spatial heterogeneity. We considered the following proxies: y-location of peak ^{129}I, number of intersections between the repository and aquifer, shortest travel time, and number of intersections in the entire discrete fracture network. The graph metrics proved to be our best tool to date for improving surrogate model performance but are demonstrably incomplete tools. These studies showed that improvement in our treatment of spatial heterogeneity is the most promising avenue for accomplishing better surrogate models.

We continue to study various aspects of the Discrete Fracture Networks. This year, we performed a detailed study involving 100 DFNs each modeling two transmissivity relationships: correlated constant vs. correlated depth-dependent transmissivity. The purpose was to determine if a correlated depth-dependent transmissivity relationship produces a significant change in the performance quantities for the flow and transport simulations of nuclear repositories in subsurface

rock as compared to a correlated constant transmissivity. The findings showed that the main QoI, the maximum ^{129}I concentration in the aquifer at 1 Ma, was not significantly different but the timing of the peak ^{129}I concentration did differ. Other QoIs such as median residence time and mass fluxes did show statistically significant differences between the two transmissivities. We also observed that the maximum ^{129}I in the aquifer showed no real correlation with any graph metric for either relationship. The study highlighted differences with respect to the mass flow rates (specifically the rock to east boundary flow rate) which may indicate increased flushing for the correlated depth dependent relationship: this behavior is worth investigating further.

This year, we extended the sensitivity analysis of the crystalline reference case performed in FY20 and FY21. The FY22 analysis uses the FY21 crystalline case and adds another dimension to the sensitivity analysis: that of model form. The FY22 analysis addresses uncertainty from spatial heterogeneity represented by DFNs, epistemic parameter uncertainty, and model form uncertainty in the treatment of fuel matrix degradation (FMD). Two alternatives for FMD are investigated: a fractional dissolution rate model and an Artificial Neural Network surrogate model of the FMD model.

Comparison between the sensitivity analyses for the crystalline reference case simulations with either the FDR alternative model or the ANN alternative model for FMD showed that the choice of FMD model alternative has a minimal effect on most QoIs. Some differences were detected in the maximum ^{129}I concentrations in the aquifer over time, with the FDR model alternative leading to more variation between simulations near 1 million years. The increase in variation is likely due to the incorporation of uncertainty in the FDR model via *rateUNF*; no comparable parametric uncertainty is currently included in the ANN model. The effect of FMD model choice was also small because of the significance of the instantaneous release fraction (*IRF*) up until around 200,000 years. Time-dependent sensitivity analysis showed that, among the parametric uncertainties, the waste package degradation dominates early, then *IRF* and *kGlacial* become significant, and *rateUNF* only gains importance after most waste packages have breached. Future analysis may incorporate additional uncertainty into the ANN surrogate model so that it is more directly comparable to the FDR model. Additionally, the IRF uncertainty is fairly high, so reductions in that uncertainty could also influence how early the FMD model alternative becomes important.

The sensitivity analyses of the FY22 crystalline reference case again highlighted the important effect that spatial heterogeneity has on the QoIs, including on the behavior of ^{129}I. This spatial heterogeneity also interacts significantly with the parameter uncertainties. Again, we saw that spatial heterogeneity presents the most significant challenge for our use of surrogate modeling SA. This conclusion motivates further advancement to our treatment of spatial heterogeneity to improve SA accuracy in performance assessment with the GDSA Framework.

This year, we utilized the *GDSA Workflo*w to perform a sensitivity analysis on the DECOVALEX crystalline case. Preliminary results are discussed in Chapter 6. We note that these results need further investigation, but it was interesting to see that the permeabilities and porosities of both the buffer and the backfill had very little influence on the QoIs which were water fluxes at different surface locations and tracer concentrations. Only the *rateUNF* dissolution rate parameter had a significant effect on the Tracer2 concentrations: the uncertainty in most of the QoIs was

entirely due to spatial heterogeneity as seen by the effect of the DFNs. We plan to extend this analysis as the DECOVALEX case develops further.

Another activity continued this year was the *GDSA Workflow*. This workflow couples Dakota, PFLOTRAN, and NGW (the Next Generation Workflow) to provide analysts with a capability to construct, execute, and communicate end-to-end computational simulation analysis workflows. The *GDSA Workflow* also allows greater reproducibility and traceability of the actual files and scripts used for a particular study. The *GDSA Workflow* was extended and generalized this year to support the DECOVALEX case study.

Finally, we continue to engage with the international community on sensitivity analysis topics. This year, we started working on a new set of case studies which are more complex reference cases. Each participating organization in our working group separately performs sensitivity analyses on these case studies. Then, we meet to compare results and evaluate the performance of different sensitivity analysis methods, with the goal of developing consensus recommendations on some best practices for sensitivity analysis in the context of repository performance assessment.

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. 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. We seek to improve 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 better understanding and quantification of the effect of spatial heterogeneity, the use of optimization in determining uncertainty bounds, further work with multifidelity UQ approaches, more detailed examination of surrogates and machine learning methods, model form uncertainty, and density-based sensitivity analysis. Future work will involve further development of these potential topics as they align with the *GDSA Framework* and the repository performance assessment needed.

8. REFERENCES

[1] L. Swiler, E. Basurto, D. Brooks, A. Eckert, R. Leone, P. Mariner, T. Portone, M. Smith and E. Stein, "Uncertainty and Sensitivity Analysis Methods and Applications in the GDSA Framework (FY2021)," SAND2021-9903R. Sandia National Laboratories, Albuquerque, New Mexico, 2021.

[2] L. Swiler, E. Basurto, D. Brooks, A. Eckert, P. Mariner, T. Portone and E. Stein, "Advances in Uncertainty Quantification and Sensitivity Analysis Methods and Applications in GDSA Framework," SAND2020-10802R. Sandia National Laboratories, Albuquerque, New Mexico, 2020.

[3] L. Swiler, J. Helton, E. Basurto, D. Brooks, P. Mariner, L. Moore, S. Mohanty, S. Sevougian and E. Stein, "Status Report on Uncertainty Quantification and Sensitivity Analysis Tools in the Geologic Disposal Safety Assessment (GDSA) Framework," SAND2019-13835R. Sandia National Laboratories, Albuquerque, New Mexico, 2019.

[4] P. Mariner, E. Stein, S. Sevougian, L. Cunningham, J. Frederick, G. Hammond, T. Lowry, S. Jordan and E. Basurto, "Advances in Geologic Disposal Safety Assessment and an Unsaturated Alluvium Reference Case," SFWST-2018-000509; SAND2018-11858R. Sandia National Laboratories, Albuquerque, New Mexico, 2018.

[5] P. Mariner, E. Stein, J. Frederick, S. Sevougian and G. Hammond, "Advances in Geologic Disposal System Modeling and Shale Reference Cases," SFWST-2017-000044; SAND2017-10304R. Sandia National Laboratories, Albuquerque, New Mexico, 2017.

[6] P. Mariner, E. Stein, J. Frederick, S. Sevougian, G. Hammond and D. Fascitelli, "Advances in Geologic Disposal System Modeling and Application to Crystalline Rock," FCRD-UFD-2016-000440; SAND2016-9610R. Sandia National Laboratories, Albuquerque, New Mexico, 2016.

[7] P. Mariner, W. Gardner, G. Hammond, S. Sevougian and E. Stein, "Application of Generic Disposal System Models," FCRD-UFD-2015-000126; SAND2015-10037R. Sandia National Laboratories, Albuquerque, New Mexico, 2015.

[8] U.S. DOE, "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," U.S. Department of Energy, Washington, D.C. , 2008.

[9] U.S. DOE, "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, 2014.

[10] J. Helton and M. Marietta (Editors), "Special Issue: The 1996 Performance Assessment for the Waste Isolation Pilot Plant," *Reliability Engineering and System Safety,* no. 69, pp. 1-451, 2000.

[11] J. Helton, C. Hansen and P. Swift (Editors), "Special Issue: Performance Assessment for the Proposed High-Level Radioactive Waste Repository at Yucca Mountain, Nevada," *Reliability Engineering and System Safety,* vol. 122, pp. 1-456, 2014.

[12] P. Lichtner and G. Hammond, "Quick Reference Guide: PFLOTRAN 2.0 (LA-CC-09-047) Multiphase-Multicomponent-Multiscale Massively Parallel Reactive Transport Code," LA-UR-06-7048. Los Alamos National Laboratory, Los Alamos, New Mexico, 2012.

[13] G. Hammond, P. Lichtner, R. Mills and C. Lu, "Toward petascale computing in geosciences: application to the Hanford 300 Area," *Journal of Physics: Conference Series, SciDAC 2008: Scientific Discovery through Advanced Computing,* vol. 125, no. 012051, 2008.

[14] G. Hammond, P. Lichtner and R. Mills, "Evaluating the performance of parallel subsurface simulators: Ann illustrative example with PFLOTRAN," *Water Resources Research,* vol. 501, no. 1, pp. 208-228, 2014, http://doi.org/10.1002/2012wr013483.

[15] S. Sevougian, G. Hammond, P. Mariner, E. F. J. Stein and R. MacKinnon, "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.

[16] B. Adams, W. Bohnhoff, K. Dalbey, M. Ebeida, J. Eddy, M. Eldred, R. Hooper, P. Hough, K. Hu, J. Jakeman, M. Khalil, K. Maupin, J. Monschke, E. Ridgway, A. Rushdi, D. Seidl, J. Stephens, L. Swiler and J. Winokur, "Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.16 User's Manual," SAND2022-6171, Sandia National Laboratories, Albuquerque, New Mexico, 2022.

[17] A. Eckert, L. Swiler, T. Portone, E. Basurto, R. Leone and D. Brooks, "A Graphical Workflow Supporting Deep Geological Repository Performance Assessment," In Preparation.

[18] B. Peherstorfer, K. Willcox and M. Gunzburger, "Optimal Model Management for Multifidelity Monte Carlo Estimation," *SIAM Journal of Scientific Computing,* vol. 38, no. 5, pp. A3163-A3194, 2016, http://doi.org/10.1137/15M1046472.

[19] M. Giles, "Multilevel Monte Carlo Methods," *Acta Numerica,* vol. 24, pp. 259-328, 2015, http://doi.org/10.1017/S096249291500001X.

[20] A. A. Gorodetsky, G. Geraci, M. Eldred and J. Jakeman, "A generalized approximate control variate framework for multifidelity uncertainty quantification," *Journal of Computational Physics,* vol. 408, p. 109257, 2020, http://doi.org/10.1016/j.jcp.2020.109257.

[21] E. Qian, B. Peherstorfer, D. O'Malley, V. Vesselinov and K. Willcox, "Multifidelity Monte Carlo Estimation of Variance and Sensitivity Indices," *SIAM/ASA Journal of Uncertainty Quantification,* vol. 6, no. 2, pp. 683-706, 2018, http://doi.org/10.1137/17M1151006.

[22] B. Peherstorfer, T. Cui, Y. Marzouk and K. Willcox, "Multifidelity importance sampling," *Computer Methods in Applied Mechanics Engineering,* vol. 300, pp. 490-509, 2016, http://doi.org/10.1016/j.cma.2015.12.002.

[23] S. Joyce, L. Harley, D. Applegate, J. Hoek and P. Jackson, "Multiscale 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,* vol. 22, no. 6, pp. 1233-1249, 2014.

[24] T. Portone, G. Geraci, M. Eldred and L. Swiler, "Multifidelity uncertainty quantification for deep geologic nuclear waste repositories," *In preparation.*

[25] I. Sobol', "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Mathematics and Computers in Simulation,* vol. 55, no. 1-3, pp. 271-280, 2001, http://doi.org/10.1016/S0378-4754(00)00270-6.

[26] A. Saltelli, "Making best use of model evaluations to compute sensitivity indices," *Computer Physics Communications,* vol. 145, no. 2, pp. 280-297, 2002, http://doi.org/10.1016/S0010-4655(02)00280-1.

[27] M. Smith, T. Portone and L. Swiler, "Effects of Discrete Fracture Network Modeling Choices on Repository Performance Characteristics," in *International High-Level Radioactive Waste Management Conference*, American Nuclear Society, Phoenix, November 2022, SAND2022-10610C.

[28] J. Hyman, S. Karra, J. Makedonska, C. Gable, S. Painter and H. Viswanathan, "dfnWorks: A discrete fracture network framework for modeling subsurface flow and transport," *Computers & Geoscience,* vol. 84, pp. 10-19, 2015.

[29] E. Stein, J. Frederick, G. Hammond, K. Kuhlman, P. Mariner and S. Sevougian, "Modeling Coupled Ractive Flow Processes in Fractured Crystalline Rock," in *Proceedings of the 16th International High-Level Radioactive Waste Management (IHLRWM 2017) Conference*, Charlotte, North Carolina, 2017, https://www.osti.gov/servlets/purl/1417242.

[30] S. Sevougian, E. Stein, G. Hammond, P. Mariner, J. Frederick and E. Basurto, "Simulating the Effect of Fracture Connectivitiy on Repository Performance with GDSA Framework - 18589," in *Proceedings of WM2018 Conference*, Phoenix, Arizona, 2018.

[31] J. M. Frederick, G. E. Hammond, P. E. Mariner, E. R. Stein and D. Sevougian, "Development of a Waste Form Process Model in PFLOTRAN," in *Proceedings of the 16th International High-Level_Radioactive Waste Management Conference (IHLRWM 2017)*, Charlotte, North Carolina, 2017.

[32] P. Mariner, D. Seidl, B. Debusschere, J. Vo, J. Frederick and L. Swiler, *High Fidelity Surrogate Modeling of Fuel Dissolution for Probabilistic Assessment of Repository Performance,* Knoxville, TN: International High-Level Radioactive Waste Conference, 2019.

[33] J. Jerden, V. Gattu and E. W., "Progress Report on Development of the Spent Fuel Degradation and Waste Package Degradation Models and Model Integration.," Argonne National Laboratory, Lemont, Illinois.

[34] P. Mariner, T. Berg, K. Chang, B. Debusschere, R. Leone and D. Seidl, "Surrogate Model Development of Spent Fuel Degradation for Repository Performance Assessment," SAND2020-10797R. Sandia National Laboratories, Albuquerque, New Mexico, 2020.

[35] P. Mariner, E. Basurto, D. Brooks, E. Stein and L. Swiler, "A Study of Epistemic and Spatial Aleatory Uncertainties on Crystalline Repository Performance," SAND2019-13399PE. Sandia National Laboratories, Albuquerque, New Mexico, 2019.

[36] D. Brooks, L. Swiler, P. Mariner, T. Portone, E. Basurto and R. Leone, "Sensitivity and uncertainty analysis of FMD model choice for a generic crystalline repository," in *2022 International High Level Radioactive Waste Management Conference (IHLRWM)*, Pheonix, Arizona, To be presented in November 2022.

[37] E. Stein, R. Jayne, T. LaForce, R. Leone and S. Nguyen, "DECOVALEX-2023 Task F Specification, Revision 7," SAND2021-13423O. Sandia National Laboratories, Albuquerque, New Mexico, 2021.