SAND2022-10669C

Multimodel Methods for Uncertainty Quantification of Repository Systems

Teresa Portone*, Michael Eldred†, Gianluca Geraci&, Laura Swiler#

*†&#Sandia National Laboratories, P.O. Box 5800, Albuquerque, New Mexico 87185,
*tporton@sandia.gov; †mseldre@sandia.gov; &ggeraci@sandia.gov; #lpswile@sandia.gov

INTRODUCTION

Probabilistic post-closure performance assessment (PA) of systems for deep geologic disposal of nuclear waste requires the repeated evaluation of models which represent complex multiphysics, multiscale phenomena across time and length scales spanning several orders of magnitude. These models include millions of degrees of freedom and require hundreds of core hours to run on a high-performance computer. The sheer cost of running such high-fidelity models limits the number of evaluations available for uncertainty analyses, which impacts statistical accuracy.

Uncertainty quantification (UQ) methods that leverage multiple models, such as multilevel [1] and multifidelity [2] methods and multimodel surrogate methods [3-5] exploit models of varying fidelities and cost to achieve improved computational efficiency and statistical accuracy. This paper presents a survey and demonstration of these methods for a repository model problem in sparsely fractured crystalline rock. The suitability of different methods depending on model ensemble and target uncertainty analysis is discussed, as well as the benefits and challenges of applying these methods to practical PA application problems.

MULTIMODEL METHODS

Sampling-based methods

A common goal in uncertainty analyses is the computation of statistics such as the mean and variance of key quantities of interest, e.g. the mean concentration of a radionuclide, given uncertainties in the modeled system. For a high-fidelity model Q_H, the sample mean is computed as

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

This estimator is unbiased, but its variance is $\mathbb{V}(Q_H)/N$; this means the standard error in the estimator decays slowly, at a rate of $\sqrt{N}$. Multimodel methods aim to reduce the variance of the estimator without incurring additional high-fidelity model evaluations. This variance reduction is achieved by exploiting cheaper, less accurate model evaluations in statistical estimates.

The key concept behind sampling-based multimodel UQ methods is introduced here with the multilevel method in the context of two models, a high-fidelity model Q_H and a low-fidelity model Q_L. The low-fidelity model could be a coarser discretization of Q_H, a more idealized physics model, a surrogate model, etc.

The multilevel (ML) mean estimator $\hat{Q}_H^{ML}$ is defined as:

$$\hat{Q}_H^{ML} = \hat{Q}_L + \widehat{Q_H - Q_L} = \frac{1}{N_L}\sum_{i=1}^{N_L} Q_L^{(i)} + \frac{1}{N_\Delta}\sum_{j=1}^{N_\Delta}\left(Q_H^{(j)} - Q_L^{(j)}\right). \quad (2)$$

This estimator is still unbiased, but its variance is now

$$\mathbb{V}\left[\hat{Q}_H^{ML}\right] = \frac{\mathbb{V}[Q_L]}{N_L} + \frac{\mathbb{V}[Q_H - Q_L]}{N_\Delta}. \quad (3)$$

The first contribution to the variance of the ML estimator can be made small by evaluating the low-fidelity model many times at much lower cost. The second term will be small if $\mathbb{V}[Q_H - Q_L]$ is small; this can occur when Q_H and Q_L are very similar, e.g. as model predictions converge with mesh refinement. If these conditions hold, the variance of the ML estimator can achieve a much smaller (orders-of-magnitude smaller) variance than traditional MC for the same computational cost. Note that while this example focuses on a two-model case, multiple models can be used by combining estimators of the differences between subsequent models in a telescoping sum. See [1] for further details.

Multilevel methods are limited by the fact that they rely on a model hierarchy where the variance in the difference between subsequent models decays as one moves through the hierarchy toward the highest-fidelity model. This often occurs as one refines a mesh, but if the hierarchy is instead defined by, e.g., simplifying assumptions used to define a lower-fidelity model, this may not occur. While the low- and high-fidelity models' outputs may not be very close to each other, they may still be strongly correlated. Multifidelity (MF) methods exploit such correlations to achieve a reduction in estimator variance, thereby relaxing multilevel methods' requirement of decaying variance of discrepancies between models [2].

The multifidelity method yields an analytical expression for the number of evaluations needed for each model in the hierarchy that will yield the minimum variance. However, this requires an assumption that models are ordered in a monotonic hierarchy by fidelity. This may not apply in all cases, for instance if there are two models with simplifying assumptions that are not nested, i.e. one model assumes isothermal reactions and anisotropy, and another does not assume isothermal reactions but assumes isotropy. The approximate control variate (ACV) methods relax the assumption of a strict model hierarchy and allow for a more general ensemble of models of varying fidelity [6]. This increased generality can also result in greater reductions in variance relative to the more restrictive hierarchies assumed

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.

by ML and MF, though each of the methods can outperform the others depending on the nature of the model ensemble.

Multimodel methods determined the optimal number of samples of each model required to achieve maximum variance reduction in the statistical estimator. However, estimates of variances (e.g. of the discrepancies in ML) and/or correlations between models are required for this optimal sample allocation. This is typically done by performing an initial study, called a pilot study, where all models are evaluated for a small number of samples. The sample variances/correlations computed from this process are used in sample allocation in place of their unknown true values. If the computational budget is too low to adequately refine these statistics, it can result in suboptimal sample allocations that lessen the benefit of these methods.

Surrogate-based methods

In contrast to sampling-based methods, surrogate-based multimodel methods focus on constructing a surrogate for the highest-fidelity model of interest at a reduced cost by exploiting evaluations from lower-fidelity models to construct the surrogate. Surrogate-based approaches are attractive in scenarios where the high-fidelity output must be queried many more times than would be computationally feasible, e.g. for calibration or sensitivity analysis. They are currently the only viable option if statistics other than moments of a quantity of interest must be computed, e.g. a failure probability or estimation of a probability distribution (sampling-based methods are currently limited to estimation of the mean, standard deviation, and their combination). In addition, multifidelity surrogates can exploit model relationships that can't currently be exploited by sampling-based methods—for instance, sparsity in the discrepancy between two models.

As with surrogates constructed using a single model's evaluations, multifidelity surrogates have limitations in their applicability. For instance, commonly used surrogates that can resolve complex model outputs, e.g. high-order polynomial chaos expansions (PCEs) or Gaussian processes (GPs), are limited in terms of the number of input parameters they can be constructed over, nominally no more than ~20 inputs. (Sparse grid collocation methods can mitigate this limitation; see e.g. [7,8].) Additionally, the accuracy of commonly used surrogates such as Gaussian processes (GPs) and polynomial chaos expansions (PCEs) is impeded by discontinuities in model outputs. These limitations still exist for the multimodel versions of these surrogates as well. Here we will illustrate how evaluations from multiple models can be combined to create a surrogate for a high-fidelity model at lower cost in the context of multifidelity PCEs [3]; however, other multimodel surrogate methods are in development; see [4,5] for further discussion on some of these methods.

Polynomial chaos expansions (PCE) have gained popularity in uncertainty analyses in recent years and have been documented in detail in [9]. PCE is a stochastic expansion method whereby the model output is expanded in a series of polynomials that are orthogonal with respect to the densities of the random model inputs. One advantage of PCE is that the moments (means, variances, etc.), as well as Sobol' indices for global sensitivity analysis (GSA), can be computed analytically from the expansion coefficients [10].

PCE coefficients are determined either by projecting model responses onto the basis or by solving a regression problem. The number of evaluations required to estimate the coefficients increases with the number of input parameters and the order of the polynomials in the PCE, which in turn increases with complexity of model responses. For model problems with many inputs and complex outputs requiring high-order PCEs, the cost of constructing an adequate surrogate model can quickly become intractable.

To mitigate this, multifidelity PCEs (MF PCEs) exploit a hierarchy of model fidelities and costs to reduce the computational burden of constructing a surrogate for the highest-fidelity model's response [8]. We summarize how this is done here in the context of the previous two-model hierarchy. Let Q_H and Q_L take as inputs the uncertain parameters $\boldsymbol{\xi}$. The goal is to derive a PCE surrogate for Q_H. In a single-fidelity framework, the high-fidelity model is evaluated multiple times to estimate the coefficients of the PCE $\bar{Q}_H$:

$$Q_H(\boldsymbol{\xi}) \approx \bar{Q}_H(\boldsymbol{\xi}) = \beta_0 + \sum_{i=1}^{p} \beta_i \Psi_i(\boldsymbol{\xi}), \qquad (4)$$

where p is the order of the expansion; as p increases, more terms are included in the expansion, requiring more model evaluations to estimate the coefficients β_i.

A MF PCE can also be derived for Q_H which uses evaluations from both models in its construction. The high-fidelity model can be represented as the low-fidelity model corrected with a discrepancy between the two models, which is hopefully easier to approximate than the high-fidelity model directly:

$$Q_H = Q_L + (Q_H - Q_L). \qquad (5)$$

The MF PCE approximates Q_H by constructing a PCE for Q_L ($\bar{Q}_L$) and a PCE for the discrepancy $Q_H - Q_L$ ($\bar{Q}_\Delta$):

$$Q_H(\boldsymbol{\xi}) \approx \bar{Q}_L(\boldsymbol{\xi}) + \bar{Q}_\Delta(\boldsymbol{\xi}). \qquad (6)$$

MF PCE is computationally more efficient if Q_L is similar enough to Q_H that the difference between the two responses is simpler than Q_H alone (e.g. through more rapid decay or sparsity of coefficients). In cases where this is true, fewer terms are required to resolve the $\bar{Q}_\Delta$ than to resolve $\bar{Q}_H$, thereby requiring fewer evaluations of the more costly high-fidelity model.

The MF PCE can be used in any of the ways a PCE constructed in a single-fidelity fashion would be used, e.g., to compute response moments or Sobol' indices. Generalizing to more than two models, a MF PCE can be constructed for a hierarchy of models, indexed from $i = 0$ at the lowest fidelity to $i = H$ at the highest fidelity, as

$$Q_H \approx \bar{Q}_0 + \sum_{i=1}^{H} \bar{Q}_{\Delta_i}, \quad (7)$$

where $\bar{Q}_{\Delta_i}$ is the PCE expansion of $Q_i - Q_{i-1}$.

Computational cost of multimodel methods

The computational costs of multimodel methods are compared to single-fidelity methods applied to the high-fidelity model by computing the number of equivalent high-fidelity model evaluations, denoted N_{equiv}. This is computed by defining the cost of each model relative to the highest fidelity model (C_i/C_H for the i^{th} model, where C_H is the cost of evaluating the highest-fidelity model) by the number of evaluations of each model:

$$N_{equiv} = \sum_{i=1}^{H} \frac{C_i}{C_H} N_i. \quad (8)$$

CASE STUDY

Multimodel methods will be demonstrated and analyzed herein for a simple repository model in fractured granite, based on the well-characterized Forsmark site in Sweden [11]. The problem is constructed such that is reflects many of the properties and challenges of state-of-the-art repository simulations in fractured crystalline rock while being more computationally tractable.

The simplified crystalline problem was defined on a 1000 x 1000 x 480 m^3 domain and was implemented using PFLOTRAN. A discrete fracture network (DFN) was generated to represent spatial uncertainty in the fractured granite. A uniform mesh was applied to the entire domain, and the repository was assumed to be a homogeneous region of disturbed rock zone (DRZ), with a single cell in the center representing buffer material and a single waste package. A leak from the waste package occurs at a certain point in the simulation. A glacial aquifer region was placed at the top of the domain and quantities of interest related to the radionuclide ^{129}I were tracked in the aquifer as a measure of repository performance. For more detail about the problem specification, see [12]. A 2D vertical slice of the domain is shown in Figure 1.

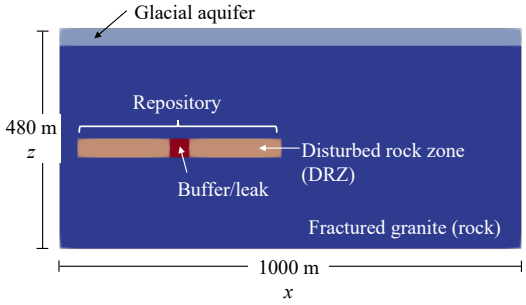

Fig. 1. A vertical slice of the simplified crystalline domain, taken at y = 500 m.

A model hierarchy was designed in terms of the spatial discretization applied to the computational domain. Mesh sizes of 10, 20 and 40 m were used. The mesh size affects the continuum properties which are derived in the process of converting from DFN to equivalent continuous porous medium (ECPM), described in [13]. The continuum values are explicitly scaled by the cell size. Additionally, when two fractures that do not intersect are mapped to the same cell, a false connection is created in the continuum property. The number of false connections increases with cell size, so it is expected that coarser meshes will exhibit increased flow compared to the finer meshes. This effect can be seen in Figure 2, which shows a horizontal slice of the permeability in the *x* direction for the three mesh sizes.

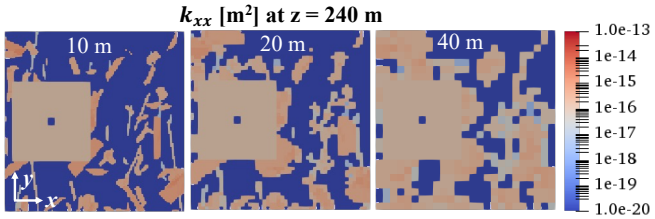

Fig. 2. Horizontal slices of the permeability tensor in the x-direction for meshes with cell sizes d = 10, 20, and 40 m, from left to right.

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

Table 1. Uncertain parameters in case study.

The uncertain parameters are drawn from the crystalline reference case described in [10], except for the time at which the waste package breaches. This is modeled as a uniform random variable ranging from 2500 to 10000 years. Parameter definitions are presented in Table 1 above. Costs of evaluating the three discretizations are reported in Table 2.

Mesh size [m]	Core time [s]	Relative cost
10	9822.8	1.0
20	329.7	3.36e-2
40	22.0	2.24e-3

Table 2. Absolute and relative model costs by discretization.

RESULTS

Sampling-based and surrogate-based results are presented. All analyses were performed using Dakota [14].

Sampling-based methods

Here we consider how different sampling methods perform in terms of their accuracy in estimating the mean of a repository performance quantity of interest (QoI). The QoI under consideration is the peak ^{129}I concentration in the

aquifer, which is the maximum concentration achieved across the timespan of the simulation. This QoI is extracted from the maximum ^{129}I concentration in the aquifer achieved at each time in the simulation, shown in Figure 3 for each of the discretizations in the model hierarchy, evaluated at the means of the uncertain parameter distributions.

It can be seen in Figure 3 that convergence in behavior as the mesh is refined is not observed. This is likely because the finer mesh has fewer (falsely) connected regions of high permeability (and thus rapid flow), leading to decreased dilution of ^{129}I as it is transported throughout the computational domain. This leads to higher concentrations reaching the aquifer as the mesh is refined. An initial 25-sample pilot study indicated that this results in a lack of decay in the variance of discrepancies between models, so we do not expect multilevel to be the most advantageous method in this case, despite the model hierarchy being defined by discretization (a common use-case for the method).

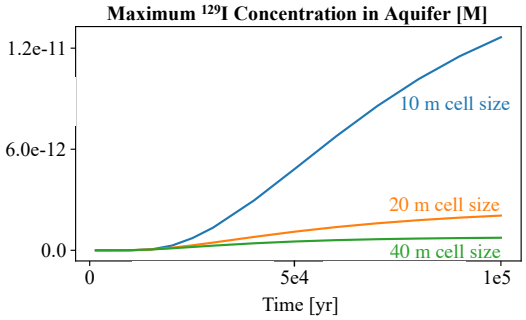

Figure 3. Maximum ^{129}I concentration in the aquifer as a function of time.

Despite the lack of convergence as the mesh is refined, the models are well correlated, which indicates that multifidelity and ACV methods may still be advantageous. This is borne out when we examine the projected estimator variance, computed from the pilot study, that can be achieved for each of these methods relative to standard Monte Carlo (MC) with a computational budget of 100 equivalent high-fidelity evaluations, reported in Table 3.

Method	Projected estimator variance	MC variance / Projected variance
Monte Carlo	1.78e-3	1.0
Multilevel	1.22e-4	15.7
Multifidelity	2.21e-5	80.6
ACV MF	2.85e-5	62.5

Table 3. Projected estimator variances for single-model and multimodel sampling-based methods.

The pilot study suggests that MF will provide the greatest improvement in accuracy with an estimated 80.6x reduction in variance relative to single-model MC. The MF variant of ACV is also expected to provide a significant variance reduction, but not quite as great as MF. Note that, despite its suboptimality, ML is still projected to achieve a 15.7x variance reduction relative to MC, a significant gain in accuracy.

It can be challenging to discern *a priori* which multimodel method will be most advantageous for a given set of models due to the complex interplay of cost with correlation and how model relationships are exploited by each of the estimators. The pilot study can be reused to make projections for each of the methods, so it is recommended to obtain projected variances for each of the methods for selecting one to use in a sampling campaign. See [15] for an algorithmic approach to identifying an optimal estimator.

Surrogate-based methods

Here we consider how the performance of MF PCE depends on the nature of the QoI it is constructed to represent. Here we consider two repository performance QoIs: the peak ^{129}I concentration in the aquifer and the x location at which that peak occurs. Construction of the MF PCE and analysis of other performance QoIs are presented in [10].

As with their single-model counterparts, commonly deployed multimodel surrogates are reliant on continuity of the QoIs they must represent. For this case study it was found that some of the performance QoIs, including the x location of the peak ^{129}I concentration, exhibit discontinuities, where discrete steps are observed for small values of *kGlacial* (see Figure 4). On the other hand, the peak concentration varies smoothly, and the outputs of each of the models exhibit similar trends as a function of the uncertain parameters. Based on these observations it is expected that a surrogate-based approach will not be accurate for the peak x location, but may be advantageous for the peak concentration.

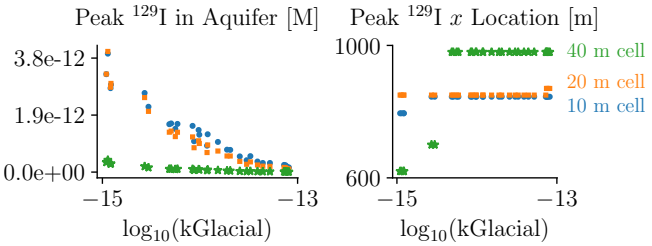

Figure 4. Peak ^{129}I concentration and x location in the aquifer as a function of kGlacial (aquifer permeability).

This expectation is borne out when considering the histograms for each QoI obtained from a direct sampling of the high-fidelity (HF) model, samples from a PCE constructed from 828 HF evaluations, and samples from a MF PCE constructed from evaluations from all model discretizations, shown in Figure 5. Because the surrogate is built on an assumption that the output varies smoothly as a function of the input parameters, we see that for the peak x location in Figure 5 that both the HF and the MF PCE produce a very skewed histogram which is a continuous approximation of the discrete values achieved by the direct sampling of the HF model.

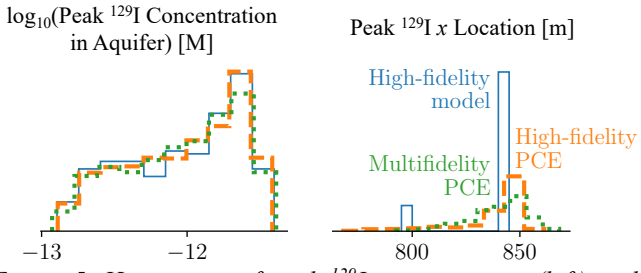

Figure 5. Histograms of peak ^{129}I concentration (left) and peak ^{129}I x location (right) produced from high-fidelity model evaluations, a PCE constructed from high-fidelity model evaluations only, and a MF PCE.

On the other hand, both the HF and MF PCE are able to reproduce the histogram for the peak concentration attained from directly sampling the HF model. This indicates the probability distribution for the peak concentration is well captured by both surrogates. However, the MF PCE achieved this accuracy at a much lower cost of 21.85 equivalent high-fidelity evaluations (18 evaluations of the 10 m discretization, 108 of the 20 m, and 108 of the 40 m). This highlights the potential computational benefit that can be achieved using multimodel surrogate methods, *provided* the conditions necessary for adequate accuracy are achieved and a relationship between models can be exploited to reduce cost.

CONCLUSION

In this work, both sampling-based and surrogate-based multifidelity UQ methods were introduced, and the suitability of the methods based on the target uncertainty analysis and the nature of the models available for use with the methods was discussed. The methods were demonstrated on a relevant PA case study. 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.

ACKNOWLEDGMENTS

This work has been funded by the Geologic Disposal Safety Assessment program under the Spent Fuel and Waste Science and Technology Campaign of the U.S. Department of Energy Office of Nuclear Energy. This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan https://www.energy.gov/downloads/doe-public-access-plan. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

1. M.B. GILES, Multilevel Monte Carlo methods. *Acta Numerica*, 24:259–328, (2015).
2. B. PEHERSTORFER, K. WILLCOX, & M. GUNZBURGER, Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review*, 60(3), 550-591. (2018).
3. L.W.T. NG & M.S. ELDRED, Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation. In 53rd AIAA/ASME/ASCE/AHS/ ASC Structures, Structural Dynamics and Materials Conference (p. 1852) (2012).
4. M.C. KENNEDY & A. O'HAGAN, Predicting the Output from a Complex Computer Code When Fast Approximations Are Available. *Biometrika* **87**, 1–13 (2000).
5. A.A. GORODESTKY, J.D. JAKEMAN & G. GERACI, MFNets: data efficient all-at-once learning of multifidelity surrogates as directed networks of information sources. *Computational Mechanics* **68**, 741–758 (2021).
6. A.A. GORODETSKY, et al., A generalized approximate control variate framework for multifidelity uncertainty quantification. *Journal of Computational Physics*, 408, 109257 (2020).
7. A.L. TECKENTRUP, et al., A Multilevel Stochastic Collocation Method for Partial Differential Equations with Random Input Data. *SIAM/ASA J. Uncertainty Quantification* **3**, 1046–1074 (2015).
8. J.D. JAKEMAN, et al., Adaptive multi-index collocation for uncertainty quantification and sensitivity analysis. *International Journal for Numerical Methods in Engineering* **121**, 1314–1343 (2020).
9. L.P. SWILER et al., Status Report on Uncertainty Quantification and Sensitivity Analysis Tools in the Geologic Disposal Safety Assessment (GDSA) Framework. Sandia Technical Report SAND2019-13835R (2019).
10. B. SUDRET, Global sensitivity analysis using polynomial chaos expansions, *Reliability Engineering and System Safety* 93 (7), 964–979 (2008).
11. S. JOYCE, et al., Multi-scale groundwater flow modeling during temperate climate conditions for the safety assessment of the proposed high-level nuclear waste repository site at Forsmark, Sweden. *Hydrogeology Journal*, 22(6):1233–1249 (2014).
12. L.P. SWILER, et al., Uncertainty and Sensitivity Analysis Methods and Applications in the GDSA Framework (FY2021). Sandia Technical Report SAND2021-9903R.
13. E. STEIN, et al., Modeling Coupled Reactive Flow Processes in Fractured Crystalline Rock. in *Proceedings of the 16th International High-Level Radioactive Waste Management Conference* (2017).

14. B.M. ADAMS, et al., Dakota, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 6.15 User's Manual. Sandia Technical Report SAND2020-12495 (2021).
15. G.F. BOMARITO, et al., On the optimization of approximate control variates with parametrically defined estimators. *Journal of Computational Physics* **451**, 110882 (2022).