



Bayesian Analysis of TRISO Fuel: Quantifying Model Inadequacy, Incorporating Lower-Length-Scale Effects, and Developing Parallel Active Learning Capabilities

Technical Report

Somayajulu L. N. Dhulipala¹, Pierre-Clément A. Simon¹, and Stephen R. Novascone¹

¹Idaho National Laboratory



*INL is a U.S. Department of Energy National Laboratory
operated by Battelle Energy Alliance, LLC*

DISCLAIMER

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

Bayesian Analysis of TRISO Fuel: Quantifying Model Inadequacy, Incorporating Lower-Length-Scale Effects, and Developing Parallel Active Learning Capabilities

Technical Report

Somayajulu L. N. Dhulipala¹, Pierre-Clément A. Simon¹, and Stephen R. Novascone¹

¹Idaho National Laboratory

July 2024

**Idaho National Laboratory
Computational Mechanics and Materials Department
Idaho Falls, Idaho 83415**

<http://www.inl.gov>

**Prepared for the
U.S. Department of Energy
Office of Nuclear Energy
Under U.S. Department of Energy-Idaho Operations Office
Contract DE-AC07-05ID14517**

Page intentionally left blank

ABSTRACT

The U.S. Department of Energy (DOE)'s Nuclear Energy Advanced Modeling and Simulation (NEAMS) program aims to develop predictive capabilities by applying computational methods to the analysis and design of advanced reactor and fuel-cycle systems. This program has been providing engineering-scale support for the continued development of BISON, a high-fidelity, high-resolution fuel performance tool. Fuel behavior in nuclear reactors is governed by a complex network of mechanisms that interact with various other physics aspects in the reactor system. Any model developed to represent fuel behavior will likely be idealized, resulting in uncertainties when comparing their predictions against the observed data. In Fiscal Year (FY)-23, we initiated the Uncertainty Quantification (UQ) work by using Bayesian methods to establish a level of model trustworthiness and further improve it, with a particular emphasis on TRI-Structural isotropic (TRISO) nuclear fuel. This year, we further expanded on that UQ work by investigating an approach to quantifying model inadequacy and accounting for lower-length scale (LLS) effects in TRISO silver (Ag) release modeling. Furthermore, we are implementing parallel active learning capabilities to reduce the computational cost (i.e., required computational resources and elapsed time) of performing UQ.

Specifically, we utilized The Kennedy O'Hagan framework for Bayesian uncertainty quantification (KOH) to account for model inadequacy in TRISO Ag release predictions made by BISON. The KOH framework represents an improvement over the standard Bayesian framework used in FY-23. Explicitly accounting for model inadequacy in the Bayesian framework helps establish the level of experimental noise uncertainty in the Advanced Gas Reactor (AGR) data. We compared the inverse UQ results obtained from both the standard Bayesian and KOH frameworks in light of the AGR-2/3/4 data, and also compared the predictive UQ results obtained from these two frameworks in light of the AGR-1 data.

Next, we investigated the impact of considering LLS effects in the Ag release simulations. We developed an expanded database of LLS simulated effective diffusivities for Ag, covering a wide range of microstructures and temperatures. Using this database, we developed a framework for incorporating LLS effects into the engineering-scale Ag release UQ. We developed both parametric and non-parametric approaches for bridging the length scales. We then investigated the inverse UQ results in light of the AGR-2/3/4 data and the predictive UQ results in light of the AGR-1 data, and compared the LLS-informed approach and the Arrhenius equation, which does not include microstructure information.

Finally, we discussed implementing parallel active learning capabilities in the Multiphysics Object-Oriented Simulation Environment (MOOSE)/BISON to reduce the computational cost (i.e., computational resources and elapsed time) of Bayesian UQ. For verification purposes, we first tested these new capabilities on a species interaction problem. We then demonstrated them on the TRISO Ag release application, showing that parallel active learning capabilities can enhance the accuracy of UQ while also substantially reducing the computational cost in comparison to the reference methods developed in FY-23.

ACKNOWLEDGMENT




This report was authored by a contractor of the U.S. Government under contract no. DE-AC07-05ID14517. Accordingly, the U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.

This research made use of the resources of the High Performance Computing Center at Idaho National Laboratory (INL), which is supported by the Office of Nuclear Energy of the U.S. Department of Energy and the Nuclear Science User Facilities under contract no. DE-AC07-05ID14517.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could appear to have influenced the work reported in this technical report.

ORCID

Somayajulu L. N. Dhulipala		0000-0002-0801-4250
Pierre-Clément A. Simon		0000-0001-7083-2628
Stephen R. Novascone		0000-0002-3938-4326

CONTENTS

ABSTRACT	iv
LIST OF FIGURES	viii
LIST OF TABLES	x
LIST OF CODE LISTINGS	xi
LIST OF ALGORITHMS	xii
ACRONYMS	xiii
LIST OF SYMBOLS	1
I INTRODUCTION	2
II Decomposing Total Predictive Uncertainty and Quantifying the Uncertainty due to Model Inadequacy	4
II-A Mathematical Formulation of the Kennedy-O’Hagan Framework for Decomposing Predictive Uncertainties	4
II-B Implementation Details for the Kennedy-O’Hagan Framework	7
II-C Inverse Uncertainty Quantification Considering the AGR-2/3/4 Experimental Datasets	9
II-D Forward Prediction and Uncertainty Quantification in Light of the AGR-1 Experimental Dataset	14
II-E Summary and Conclusions	19
III Incorporating Lower-Length-Scale Effects into the Uncertainty Quantification Framework	21
III-A Strategies for Incorporating Lower-Length-Scale Effects	21
III-B Updated Lower-Length-Scale Simulation Database	21
III-C Models for Bridging the Lower-Length-Scale and the Engineering-Scale	24
III-D Inverse Uncertainty Quantification on the AGR-2/3/4 Experimental Datasets	27
III-E Forward Prediction and Uncertainty Quantification on the AGR-1 Experimental Dataset	30
III-F Summary and Conclusions	33
IV Parallel Active Learning for Accelerating Bayesian Inverse Uncertainty Quantification in MOOSE	38
IV-A Parallel Active Learning Approach	39
IV-B MOOSE Implementation Details	42
IV-C Application to a Test Case: Species Interaction Problem	44
IV-D Application to TRISO Fuel, Using the AGR-2 Experimental Dataset	46
IV-E Summary and Conclusions	49

V	FUTURE WORK	52
	REFERENCES	53

LIST OF FIGURES

1	Implementation details of the KOH while quantifying model inadequacy	7
2	Temperature histories from the AGR-2 and AGR-3/4 series	10
3	Comparison of the Gaussian process (GP) predicted and true model discrepancy/inadequacy considering the training set	11
4	Comparison of the GP predicted and true model discrepancy/inadequacy considering the testing set	12
5	Correlation matrix plot of the TRISO silver release model parameters comparing the base (i.e., standard Bayesian) and the KOH frameworks	13
6	Sigma values of the TRISO silver release model comparing the base (i.e., standard Bayesian) and the KOH frameworks	14
7	Forward prediction and UQ and decomposition on the AGR-1 dataset with the KOH framework with results shown for two compacts	16
8	Comparison of the predictive uncertainty estimates obtained from the KOH framework and the standard Bayesian framework on the AGR-1 dataset considering two compacts	17
9	Median predictions and 5-95 percentile confidence bands for the AGR-1 dataset obtained using the KOH and standard Bayesian frameworks	18
10	Strategies for incorporating LLS effects in the UQ framework	22
11	(a) Simulated grain dimensions (black dots) compared against the AGR-1 as-fabricated SiC grain dimensions, with twins included for all batches provided by Gerczak et al. (2016b). The simulated microstructures cover the variations in grain dimensions found in the AGR-1 fuel batches. (b) and (c) reflect two examples of polycrystalline SiC microstructures created for this study, corresponding to the (b) bottom-left and (c) top-right points shown in (a). The domain size is $9 \times 9 \mu\text{m}$. Grains are shown in red, the centers of the 60-nm-thick GBs are in blue, and the continuous transition between the two are shown in shades of white. This figure was reproduced from Jiang et al. (2021); Simon et al. (2022).	23
12	Augmented LLS simulation database for the effective Ag diffusivity	24
13	Derivation of the thermal contribution to the effective Ag diffusion coefficient as a function of the temperature T , the grain minor axis length m_i , and the grain major axis length m_a	26
14	Comparison between the LLS simulated effective diffusivity and the prediction made by the different LLS-informed models	27
15	Posterior distributions of the model parameters θ obtained using the standard Bayesian framework considering LLS effects in the Silicon Carbide (SiC) diffusivity	29
16	Posterior distribution of the sigma term σ obtained using the standard Bayesian framework considering LLS effects in the SiC diffusivity	30
17	Comparison of the GP predicted and true model discrepancy/inadequacy considering the training set and the LLS-informed SiC diffusivity model	31
18	Comparison of the GP predicted and true model discrepancy/inadequacy considering the testing set and the LLS-informed SiC diffusivity model	32
19	Correlation matrix plot of the TRISO silver release model parameters comparing the base (i.e., standard Bayesian) and the KOH frameworks when considering the LLS-informed SiC diffusivity model	33

20	Sigma values of the TRISO silver release model considering LLS effects comparing the base (i.e., standard Bayesian) and the KOH frameworks	34
21	Comparison of the predictive uncertainty estimates obtained from the KOH framework and the standard Bayesian framework on the AGR-1 dataset considering two compacts	35
22	Median predictions and 5-95 percentile confidence bands for the AGR-1 dataset obtained using the KOH and standard Bayesian frameworks considering microstructure information	36
23	Computational resources required for Bayesian inverse UQ as a function of the number of parallel Markov chains and the number of experimental configurations	38
24	Schematic describing the parallel active learning approach	39
25	Comparison of the Expected Improvement for Global Fit (EIGF) and the posterior targeted acquisition functions (Equations (23) and (24), respectively)	41
26	MOOSE objects and their dependencies to perform the training component of parallel active learning	43
27	MOOSE objects and their dependencies to perform the evaluation component of parallel active learning	44
28	Metrics to monitor the active learning process at each iteration for the species interaction test case	45
29	Comparison of the posterior distributions of the model parameters θ between parallel active learning and parallel Markov Chain Monte Carlo (MCMC) approaches for the species interaction problem	46
30	Comparison of the posterior distributions of the sigma term σ between parallel active learning and parallel MCMC approaches for the species interaction problem	47
31	Comparison of the computational cost of performing Bayesian UQ between the parallel active learning and parallel MCMC approaches for the species interaction test case	47
32	Metrics to monitor the active learning process at each iteration for the TRISO Ag release application	48
33	Comparison of the posterior distributions of the model parameters θ between parallel active learning and parallel MCMC approaches for the TRISO Ag release application	49
34	Comparison of the posterior distributions of the sigma term σ between parallel active learning and parallel MCMC approaches for the TRISO Ag release application	50
35	Comparison of the computational cost of performing Bayesian UQ between the parallel active learning and parallel MCMC approaches for the species interaction test case	50

LIST OF TABLES

I	Summary statistics of the temperature histories (i.e., experimental configurations) used for fitting the GP inadequacy term.	10
II	The most likely σ_ε values from five iterations of the KOH framework. The mean σ_ε value across all iterations of the KOH framework is also shown, along with the σ value from the base UQ framework.	14
III	Mean Squared Error (MSE)s between the median predictions and experimental values from the AGR-1 dataset, demonstrating the accuracy of the UQ approaches. Median predictions were obtained via the KOH framework (five iterations) and standard Bayesian framework. The average MSE across the five iterations is 32.4% less than that for the standard Bayesian framework, indicating greater accuracy.	19
IV	Calibration error of the uncertainty estimates (CAL) values between the median predictions and the experimental values from the AGR-1 dataset, demonstrating the quality of the uncertainty estimates of the UQ approaches by comparing against the experimental values. CAL values were computed for the KOH framework (five iterations) and standard Bayesian framework. The average CAL value across the five iterations was 42.5% less than that for the standard Bayesian framework, indicating better uncertainty estimates.	19
V	Prior distributions over the parameters, along with the σ (or σ_ε) term for the TRISO model, when considering the LLS-informed GP model for Ag diffusivity in SiC.	28
VI	The most likely σ_ε values from the five iterations of the KOH framework, comparing the Arrhenius equation and the LLS-informed GP approach. The mean σ_ε value across all iterations of the KOH framework, along with the σ value from the base UQ framework, are also shown.	34
VII	The MSE between the median predictions and experimental values from the AGR-1 dataset, demonstrating the accuracy of the UQ approaches. The median predictions were obtained from both the KOH framework (five iterations) and standard Bayesian framework in light of the Arrhenius and LLS-informed GP models.	35
VIII	The CAL values between the median predictions and experimental values from the AGR-1 dataset, demonstrating the quality of the uncertainty estimates of the UQ approaches in terms of being able to encompass the experimental values. The CAL values were computed for the KOH framework (five iterations) and the standard Bayesian framework, in light of both the Arrhenius equation and LLS-informed GP models.	37
IX	Prior distributions for the species interaction test case for parallel active learning capabilities.	44
X	Prior distributions for the TRISO Ag release application with parallel active learning capabilities.	48

LISTINGS

LIST OF ALGORITHMS

ACRONYMS

Ag	silver
AGR	Advanced Gas Reactor
CAL	Calibration error of the uncertainty estimates
DOE	U.S. Department of Energy
DTF	designed to failed
DES	Differential Evolution Sampler
EIGF	Expected Improvement for Global Fit
FIMA	fissions per initial metal atom
FY	Fiscal Year
GP	Gaussian process
INL	Idaho National Laboratory
IPyC	inner Pyrolytic Carbon (PyC)
IMH	Independent Metropolis-Hastings
KOH	The Kennedy O'Hagan framework for Bayesian uncertainty quantification
LLS	lower-length scale
MCMC	Markov Chain Monte Carlo
MOOSE	Multiphysics Object-Oriented Simulation Environment
MSE	Mean Squared Error
NEAMS	Nuclear Energy Advanced Modeling and Simulation
PIE	Post-Irradiation Examination
PyC	Pyrolytic Carbon
PI	Probability of Improvement
SiC	Silicon Carbide
Ag	silver
STM	Stochastic Tools Module
TAVA	time-average volume-average
TRISO	TRI-Structural isotropic
UCO	Uranium oxycarbide
UQ	Uncertainty Quantification
UCB	Upper Confidence Bound

LIST OF SYMBOLS

α_{xy}	Independent Metropolis-Hastings (IMH) acceptance probability
\mathcal{G}	The proposal distribution in IMH
\mathcal{L}	Likelihood function for assessing the model output against experiments
\mathcal{M}	The computational model
$\mathcal{N}(a, b)$	Normal distribution with mean a and standard deviation b
\mathcal{P}	Prior distribution for the model parameters θ
\mathcal{S}	State space of a Markov chain
$\mathcal{TN}(a, b, l, u)$	Truncated Normal distribution with mean a , standard deviation b , lower bound l , and upper bound u
$\mathcal{U}(l, u)$	Uniform distribution with lower bound l and upper bound u
$\pi = \{\pi_x\}_{x \in \mathcal{S}}$	Stationary distribution of a Markov chain with state space \mathcal{S}
\mathcal{D}	Experimental data
Θ	Vector of experimental configurations
θ	Vector of model parameters
$T = \{t_{xy}\}_{x,y \in \mathcal{S}}$	One step transition probability matrix
σ	Term which captures both the model inadequacy and experimental noise uncertainties
τ_p	Integrated autocorrelation time with respect to a function $p(\cdot)$
$E(\cdot)$	Expectation function
E_S	Effective sample size
$f(\cdot)$	Probability density function
M	Number of model parameters
N	Number of experimental configurations or data points
P	Number of parallel proposals
S	Number of Markov chain Monte Carlo samples
$t_{xy}^{(n)}$	n step transition probability
t_{xy}	One step transition probability
$Var(\cdot)$	Variance function

I. INTRODUCTION

Beyond the possibility of some known inadequacy, however, models will virtually always be wrong in unknown ways and to unknown degrees.

–MacCallum and O’Hagan (2015)

As one of the main programs supporting modeling and simulation activities under the DOE Office of Nuclear Energy, NEAMS is committed to conducting model validation and calibration based on experimental data. The real-world phenomena that take place in nuclear reactors are complex and entail many different physics aspects all occurring in tandem and interacting with each other. While these physics aspects may each be well understood in isolation, their interactions create challenges in terms of developing models to replicate the behavior of nuclear reactor components. Thus, any model should be viewed as a simplified representation of the actual phenomena observed in a nuclear reactor. This simplification creates fundamental model deficiencies in light of actual observations, and these manifest as model inadequacy. Moreover, the development of models introduces model parameters perhaps not directly observable. These parameters must be calibrated based on experimental observations. Finally, the experimental observations themselves may not be completely trustworthy. The process of collecting experimental data using instruments can introduce random noise. Even the data collection process itself can alter the natural course of a physical process in a reactor, introducing random noise. As a result of these model parameters, model inadequacy, and noise in the experimental data, any comparison between model predictions and experiment results will be uncertain.

In FY-23, UQ was performed based on the standard Bayesian framework and applied to the case of Ag release in TRISO fuel. In the standard Bayesian framework, uncertainties stemming from experimental noise and model inadequacy are characterized in lumped fashion, meaning they cannot be measured individually. Moreover, failing to account for model inadequacy in the UQ framework can explicitly lead to inaccuracies in the quantified uncertainties in the model parameters, as well as when performing forward UQ (MacCallum and O’Hagan, 2015; Brynjarsdottir and O’Hagan, 2014; Wu et al., 2024; Sung and Tuo, 2024; Maupin and Swiler, 2020; White and Mahadevan, 2023; Leoni et al., 2024; Bayarri et al., 2009). Therefore, in this FY, we explicitly accounted for model inadequacy by using the KOH framework. This framework, proposed by Kennedy and O’Hagan (2001), characterizes model inadequacy as a GP. The GP model identifies hidden experimental-data trends not captured by the computational model. In doing so, it aims to isolate the amount of experimental noise uncertainty during the inverse UQ process of calibrating the model parameters. In the forward UQ process, it also adds a correction term to the model predictions, potentially enhancing their accuracy and quality.

In FY-23, we used the Arrhenius equation to model Ag release in TRISO fuel. However, this model does not consider microstructure effects—particularly in the SiC layer, which is mostly responsible for acting as a barrier to Ag release. Thus, in the current FY, we consider LLS effects in modeling Ag diffusivity through the SiC layer. For this purpose, we simulate effective Ag diffusivities in consideration of a wide range of microstructures and temperatures. We then develop a framework by using both parametric and non-parametric approaches to integrate the LLS simulation data into the engineering-scale UQ of TRISO Ag release. We use both the standard Bayesian and KOH frameworks to perform the UQ, considering both the Arrhenius equation and the LLS-informed models. In this regard, AGR-2/3/4 data are used to perform the inverse UQ, and AGR-1 data to test the predictive uncertainties.

With computational models, Bayesian inverse UQ is computationally expensive in terms of computational resources and elapsed time. In FY-23, we developed massively parallelizable algorithms in MOOSE/BISON to address the significant amount of elapsed time necessary to perform Bayesian inverse UQ with computational models (Dhulipala et al., 2023). Still, the computational resources required to solve inverse UQ problems remain highly significant. For example, as the number of experimental configurations and the complexity of the UQ problem rise, the number of processors required to perform Bayesian UQ increases exponentially. Thus, in this FY, we discussed the development of new parallel active learning capabilities in MOOSE/BISON to reduce the overall computational cost of solving the Bayesian UQ problem, as measured in terms of processor hours. We then demonstrated these new capabilities and the associated reduction in computational resources by applying them to both a species interaction test problem and the TRISO Ag release application.

II. DECOMPOSING TOTAL PREDICTIVE UNCERTAINTY AND QUANTIFYING THE UNCERTAINTY DUE TO MODEL INADEQUACY

This chapter discusses the KOH framework for decomposing uncertainties and quantifying model inadequacy, with application to TRISO fractional Ag release. The KOH framework is an improvement over the standard Bayesian framework used in FY-23 (Dhulipala et al., 2023). Explicitly accounting for model inadequacy in the Bayesian framework helps establish the level of experimental noise uncertainty in the AGR data, and can improve BISON predictions. Accounting for model inadequacy via the KOH framework, its use for inverse UQ based on the AGR-2/3/4 experimental data, and its application to forward predictive UQ based on the AGR-1 data are discussed. A comparison between the KOH and standard Bayesian frameworks is also given.

A. Mathematical Formulation of the Kennedy-O'Hagan Framework for Decomposing Predictive Uncertainties

1) Generative model

A generative model defines the relationship between the experimental data and the model predictions, while accounting for any uncertainties. In FY-23, we used a generative model that took the following form:

$$\begin{aligned} \mathcal{D}(\Theta_i) &= \mathcal{M}(\boldsymbol{\theta}, \Theta_i) + \varepsilon_t \\ \text{where, } \varepsilon_t &\sim \mathcal{L}(\cdot) \end{aligned} \quad (1)$$

where Θ_i is the i^{th} experimental configuration; $\mathcal{D}(\cdot)$ and $\mathcal{M}(\cdot)$ are, respectively, the experimental observation and model prediction corresponding to Θ_i ; $\boldsymbol{\theta}$ is the model parameter vector; and ε_t is the correction factor assumed to originate from the probability distribution $\mathcal{L}(\cdot)$. ε_t is generally assumed to originate from a distribution of the form:

$$\varepsilon_t \sim \mathcal{L}(\sigma) \quad (2)$$

where σ defines the combined uncertainty stemming from both model inadequacy and experimental noise. In other words, in FY-23, we modeled the uncertainties due to model inadequacy and experimental noise in lumped fashion, via the σ term in Equation (2).

However, in this FY, by adopting the KOH framework, we aim to decompose the uncertainties stemming from model inadequacy and experimental noise. Mathematically, the generative model in this case is given by (Kennedy and O'Hagan, 2001; Arendt et al., 2012):

$$\begin{aligned} \mathcal{D}(\Theta_i) &= \mathcal{M}(\boldsymbol{\theta}, \Theta_i) + \delta(\Theta_i) + \varepsilon \\ \text{where, } \varepsilon &\sim \mathcal{L}(\sigma_\varepsilon) \end{aligned} \quad (3)$$

where $\delta(\cdot)$ is the model discrepancy or inadequacy term dependent on the experimental configuration Θ_i , and ε is the correction term representing the experimental noise. ε is assumed to originate from a distribution $\mathcal{L}(\cdot)$, with σ_ε representing the experimental noise uncertainty. The inadequacy term is modeled, for all experimental configurations Θ_i and Θ'_i , as a GP defined by:

$$\delta(\Theta_i) \sim \mathcal{N}(m(\Theta_i), k(\Theta_i, \Theta'_i) | \boldsymbol{\gamma}_\delta) \quad (4)$$

where γ_δ represents the GP hyperparameters to be solved for by utilizing an optimization process. Further details on the GP model for simulating the inadequacy term are given in Section II-A2.

2) Gaussian processes for modeling the model inadequacy term

Under the GP framework, the model inadequacy term $\delta(\cdot)$ is treated as a probability distribution over the function space of the experimental configurations in a finite domain (Rasmussen, 2004). A function $\delta(\Theta)$ is said to be a GP if it follows a joint normal distribution with mean and covariance functions $m(\Theta)$ and $k(\Theta, \Theta')$ (Rasmussen, 2004):

$$\delta(\Theta) \sim \mathcal{N}(m(\Theta), k(\Theta, \Theta') | \gamma_\delta) \quad (5)$$

Given the general flexibility of a GP to model the relation between input and output data, GPs are often used as surrogate models to predict new output values at previously unsampled input values. That is, given some training data $\{\Theta, \delta\}$, a GP can predict the output δ_* at a new input value Θ_* by exploiting the joint Gaussian distribution between the training data and the new sample points:

$$\begin{Bmatrix} \delta \\ \delta_* \end{Bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} k(\Theta, \Theta) & k(\Theta, \Theta_*) \\ k(\Theta_*, \Theta) & k(\Theta_*, \Theta_*) \end{bmatrix} \middle| \gamma_\delta\right) \quad (6)$$

The posterior predictive distribution of δ_* , given the training/new inputs and training outputs, is:

$$p(\delta_* | \Theta, \Theta_*, \delta, \gamma_\delta) \sim \mathcal{N}\left(k(\Theta_*, \Theta) k(\Theta, \Theta)^{-1} \delta, \right. \\ \left. k(\Theta_*, \Theta_*) - k(\Theta_*, \Theta) k(\Theta, \Theta)^{-1} k(\Theta, \Theta_*) | \gamma_\delta\right) \quad (7)$$

To determine the precise mean and variance of δ_* , it is necessary to infer/learn a set of hyperparameters γ_δ for the covariance function $k(\Theta, \Theta')$. This parameter learning is often accomplished by minimizing the negative marginal log-likelihood with respect to the hyperparameters:

$$-\ln p(\delta | \Theta, \gamma_\delta) \propto \frac{1}{2} \ln |k(\Theta, \Theta)| + \frac{1}{2} \delta^T k(\Theta, \Theta)^{-1} \delta \quad (8)$$

The covariance function $k(\cdot, \cdot)$ is defined through a kernel function $\kappa(\cdot, \cdot)$ (Rasmussen, 2004). Here, we specifically adopted the squared exponential kernel:

$$\kappa(\Theta_i, \Theta'_i) = \tau^2 \exp\left(-\frac{1}{2} \sum_{d=1}^D \frac{(\Theta_{i,d} - \Theta'_{i,d})^2}{l_d^2}\right) \quad (9)$$

where $\gamma_\delta = \{\tau^2, l_d\} \forall d \in \{1, \dots, D\}$, with D representing the dimensionality of each experimental configuration. These hyperparameters are solved for by optimizing Equation (8).

3) Bayesian inverse analysis

With the generative model now defined by Equation (3) and the model inadequacy/discrepancy term defined by a GP described in Equations (5)–(7), the goal of the Bayesian inverse analysis is to quantify the unknown θ , σ_ε , and γ_δ given the experimental observations. Formally, we aim to infer the following joint probability distribution:

$$f(\theta, \sigma_\varepsilon, \gamma_\delta | \Theta, \mathcal{M}, \mathcal{D}) \propto \mathcal{L}(\theta, \sigma_\varepsilon, \gamma_\delta | \Theta, \mathcal{M}, \mathcal{D}) f(\theta, \sigma_\varepsilon) f(\gamma_\delta) \quad (10)$$

where $\mathcal{L}(\boldsymbol{\theta}, \sigma_\varepsilon, \boldsymbol{\gamma}_\delta | \boldsymbol{\Theta}, \mathcal{M}, \mathcal{D})$ is the likelihood function describing the relation between the experimental observations and the model predictions—in consideration of the model inadequacy term—and $f(\boldsymbol{\theta}, \sigma_\varepsilon)$ and $f(\boldsymbol{\gamma}_\delta)$ are prior distributions over $\{\boldsymbol{\theta}, \sigma_\varepsilon\}$ and $\boldsymbol{\gamma}_\delta$, respectively. The likelihood function is further defined as:

$$\mathcal{L}(\boldsymbol{\theta}, \sigma_\varepsilon, \boldsymbol{\gamma}_\delta | \boldsymbol{\Theta}, \mathcal{M}, \mathcal{D}) \propto \prod_{i=1}^N \mathcal{L}(\boldsymbol{\theta}, \sigma_\varepsilon, \boldsymbol{\gamma}_\delta | \Theta_i, \mathcal{M}, \mathcal{D}_i) \quad (11)$$

where N is the number of experimental configurations. Inferring the joint distribution of $\{\boldsymbol{\theta}, \sigma_\varepsilon\}$ and $\boldsymbol{\gamma}_\delta$, as shown in Equation (10), implies that the model parameters, experimental noise, and GP hyperparameters should be jointly inferred given the experimental data. If the prior distribution over these parameters is non-informative or poorly constrained, it will be highly complex to jointly infer these parameters, both from a numerical convergence perspective and in terms of the practical utility of these joint distribution (Kennedy and O'Hagan, 2001; Jiang et al., 2020). Thus, several studies adopting the KOH framework (e.g., Jiang et al. (2020); Liu et al. (2019); Wu et al. (2018)) first infer the GP inadequacy term hyperparameters $\boldsymbol{\gamma}_\delta$, then infer the joint distribution of only $\{\boldsymbol{\theta}, \sigma_\varepsilon\}$. With this practical consideration, the inference problem boils down to characterizing the following distribution:

$$f(\boldsymbol{\theta}, \sigma_\varepsilon | \boldsymbol{\Theta}, \mathcal{M}, \mathcal{D}, \boldsymbol{\gamma}_\delta) \propto \mathcal{L}(\boldsymbol{\theta}, \sigma_\varepsilon | \boldsymbol{\Theta}, \mathcal{M}, \mathcal{D}, \boldsymbol{\gamma}_\delta) f(\boldsymbol{\theta}, \sigma_\varepsilon) \quad (12)$$

where, $\mathcal{L}(\boldsymbol{\theta}, \sigma_\varepsilon | \boldsymbol{\Theta}, \mathcal{M}, \mathcal{D}, \boldsymbol{\gamma}_\delta) \propto \prod_{i=1}^N \mathcal{L}(\boldsymbol{\theta}, \sigma_\varepsilon | \Theta_i, \mathcal{M}, \mathcal{D}_i, \boldsymbol{\gamma}_\delta)$

Details on inferring the GP inadequacy term hyperparameters a priori and performing the inverse UQ are given in Section II-B.

Specifically, for the TRISO fuel fractional Ag release case, the likelihood function is defined as:

$$\mathcal{L}(\boldsymbol{\theta}, \sigma_\varepsilon | \boldsymbol{\Theta}, \mathcal{M}, \mathcal{D}, \boldsymbol{\gamma}_\delta) = \prod_{i=1}^N \mathcal{TN}(\mathcal{D}(\Theta_i) - \mathcal{M}(\boldsymbol{\theta}, \Theta_i) - \delta(\Theta_i), \sigma_\varepsilon, 0.0, 1.0) \quad (13)$$

where \mathcal{TN} is a truncated normal distribution between 0.0 and 1.0, representing the upper and lower bounds, respectively, of the fractional Ag release.

4) Forward prediction and uncertainty quantification

The previous subsection discussed calibrating the model inadequacy term $\delta(\Theta)$, model parameters, and experimental noise $\{\boldsymbol{\theta}, \sigma_\varepsilon\}$ given the experimental data. For any new experimental configuration $\hat{\Theta}$, this section discusses calculating the posterior predictive distribution, given mathematically by:

$$f(\mathcal{M}(\hat{\Theta}, \boldsymbol{\theta}) | \boldsymbol{\Theta}, \mathcal{D}, \boldsymbol{\gamma}_\delta) = \int_{\delta(\hat{\Theta})} \int_{\sigma_\varepsilon} \int_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \sigma_\varepsilon | \hat{\Theta}, \mathcal{M}, \boldsymbol{\gamma}_\delta) f(\boldsymbol{\theta}, \sigma | \boldsymbol{\Theta}, \mathcal{M}, \mathcal{D}) f(\delta(\hat{\Theta}) | \boldsymbol{\gamma}_\delta) d\delta(\hat{\Theta}) d\boldsymbol{\theta} d\sigma \quad (14)$$

where $\mathcal{L}(\boldsymbol{\theta}, \sigma_\varepsilon | \hat{\Theta}, \mathcal{M}, \boldsymbol{\gamma}_\delta)$ has the same form as the likelihood function as in Equations (12) and (13), which describes the probability distribution over the model prediction \mathcal{M} for a new experimental configuration $\hat{\Theta}$, and is given by:

$$\mathcal{L}(\boldsymbol{\theta}, \sigma_\varepsilon | \hat{\Theta}, \mathcal{M}, \boldsymbol{\gamma}_\delta) = \mathcal{TN}(\mathcal{M}(\boldsymbol{\theta}, \hat{\Theta}) + \delta(\hat{\Theta}), \sigma_\varepsilon, 0.0, 1.0) \quad (15)$$

From the probability distribution of the model prediction described in Equation (14), statistics such as the median prediction and confidence bands can be inferred. Although the above equation is a multi-dimensional integral involving a computational model, it can be solved by a simple Monte Carlo simulation because we have samples from the posterior distribution $f(\boldsymbol{\theta}, \sigma_\varepsilon | \hat{\Theta}, \mathcal{M}, \boldsymbol{\mathcal{D}}, \boldsymbol{\gamma}_\delta)$ as well as a calibrated GP model for the inadequacy term as a function of the experimental configurations. Using these samples, we can perform a simple re-sampling, either directly or via kernel techniques, and compute the corresponding model outputs and inadequacy GP term. Each of these model outputs—plus the inadequacy term—can be treated as the mean of the probability distribution \mathcal{L} given in Equation (15) for the model parameter realization $\boldsymbol{\theta}$, and the scale of this distribution is the realization σ_ε . Therefore, for each Monte Carlo output, a random draw is made from the distribution \mathcal{L} while also randomizing the model inadequacy term drawn from the GP distribution $f(\delta(\hat{\Theta}) | \boldsymbol{\gamma}_\delta)$. The resulting samples stem from the posterior predictive distribution in Equation (14).

B. Implementation Details for the Kennedy-O'Hagan Framework

The mathematical details of the KOH framework that were discussed in Section II-A will be now explained from an implementation perspective. Figure 1 presents a schematic of the KOH framework implementation.

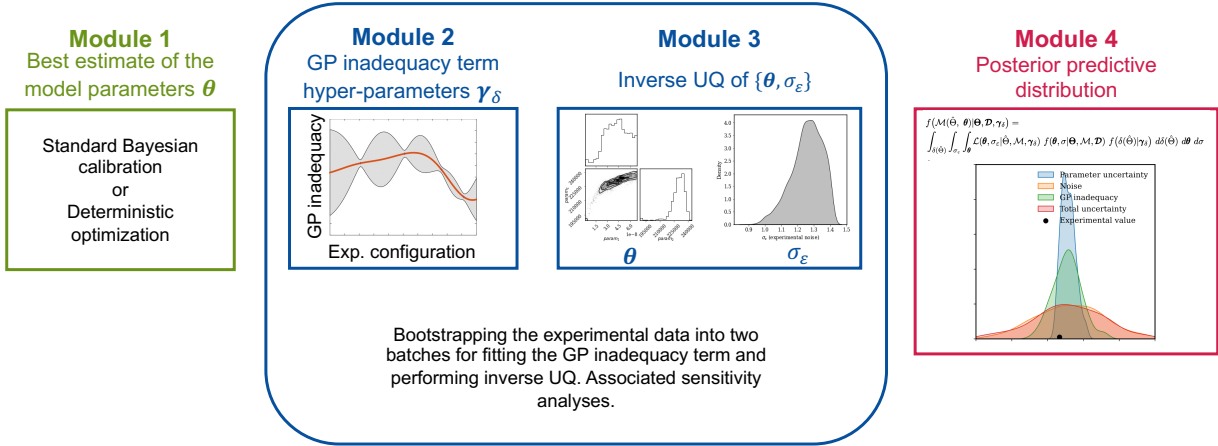


Fig. 1. Implementation of the KOH framework for Bayesian UQ while quantifying model inadequacy.

From an implementation perspective, the KOH framework can be divided into the following four modules:

- **Module 1:** A key component of the KOH framework is fitting the GP to capture the model inadequacy term in Equation (3). For this purpose, we must compute the differences between the experimental data and the corresponding model predictions. However, to compute the model predictions, we

require the model parameters to be used, and we do not know these a priori. Several studies that have implemented the KOH framework (e.g., Jiang et al. (2020); Liu et al. (2019); Wu et al. (2018)) proposed strategies such as using the best estimate of the model parameters from the prior distribution or finding the best parameter estimate by performing a deterministic optimization. The strategy of using the best estimate of the model parameters from the prior distribution can lead to significant inaccuracies in the calibrated GP inadequacy term when the priors are poorly designed or non-informative (Wu et al., 2024; Sung and Tuo, 2024). And while performing a deterministic optimization to find the best parameter estimate seems reasonable, it usually requires gradient information pertaining to the model output, which is difficult to compute for complex multiphysics models such as TRISO fuel models. Hence, we used the standard Bayesian approach, with the generative model being defined per Equation (1), and considered all the available experimental data in order to find the best estimate of the model parameters θ .

- **Module 2:** The available experimental data are divided randomly into two equal batches. Using one batch in tandem with the best estimate of the model parameters that was determined in Module 1 allows for obtaining the difference between the experimental data and the model predictions. With these differences and the corresponding experimental configurations as the training outputs and inputs, respectively, the GP inadequacy term is fit by optimizing the hyperparameters γ_δ . The second batch of experimental data can then be used as a testing set to evaluate the fitness of the GP inadequacy term predictions.
- **Module 3:** The second batch of experimental data is used to inversely quantify the model parameters and the experimental noise term $\{\theta, \sigma_\varepsilon\}$. Their posterior distribution is described in Equation (12). To perform this inverse UQ, massively parallel MCMC methods implemented in the MOOSE Stochastic Tools Module (STM) in FY-23 were utilized. Thanks to these MCMC methods, the elapsed time for the inverse UQ has been significantly reduced (see Chapters II and III of the FY-23 milestone report (Dhulipala et al., 2023) for further computational details).
- **Module 4:** Once the GP inadequacy term is fit and the model parameters and experimental noise uncertainty inversely quantified, the posterior predictive distribution is computed for a new experimental configuration for which the experimental data are unavailable. In this case, Equation (14) was solved using the Monte Carlo method. From the quantified posterior of the model parameters and experimental noise (Equation (12)), random samples were selected. The GP inadequacy values were also randomly sampled from the underlying distribution, given the new experimental configuration. The corresponding model predictions for the new experimental configuration were computed with the GP inadequacy samples added in a manner consistent with the likelihood function in Equation (15). The resulting samples stemmed from the posterior predictive distribution in Equation (14).
- **Bootstrapping and sensitivity analyses:** In Modules 2 and 3, the available experimental data are randomly divided into two batches, one for fitting the GP model inadequacy term and the other for inversely quantifying the posterior distribution of $\{\theta, \sigma_\varepsilon\}$. Thus, it is practically advisable to create several bootstrapped batches of the experimental data and to analyze the corresponding sensitivity in fitting the GP inadequacy term and quantify the posterior of $\{\theta, \sigma_\varepsilon\}$. The sensitivity of the posterior predictive distribution for a new experimental configuration can also be analyzed. Reporting the average values obtained across the bootstrapping sensitivity analysis results will contribute to the robustness of the KOH framework and likely mitigate any over- or under-fitting.

C. Inverse Uncertainty Quantification Considering the AGR-2/3/4 Experimental Datasets

The present report especially focuses on fractional Ag release for TRISO fuel. The multiphysics modeling approach for simulating fractional Ag release using BISON is given in Hales et al. (2021); Toptan et al. (2022); Dhulipala et al. (2023). The details and results of the inverse UQ process following the KOH framework are discussed below.

1) Experimental datasets

AGR-2 is the second of a series of irradiation tests sponsored by the DOE AGR program. Details on AGR-2 can be found in Collin (2018). The experiment involved six capsules, each with three stacks of four compacts—for a total of 72 compacts. Surface temperature, power level, fast neutron flux, and experiment duration were obtained from the AGR program and used for the BISON analysis. Solely focusing on Ag release and the Uranium oxycarbide (UCO) fuel kernel gives us 36 experimental data points from the AGR-2 series.

AGR-3/4 is a combination of the third and fourth of a series of irradiation tests sponsored by the DOE AGR program. Details on AGR-3/4 are provided in Skerjanc and Jiang (2022). The experiment involved a multi-monitored capsule test train with compacts containing “driver” TRISO fuel particles and 20 designed to failed (DTF) fuel particles, surrounded by rings of carbonous material. The AGR-3/4 fuel compacts were irradiated up to a burnup of 5.35–15.24% fissions per initial metal atom (FIMA), with a fast neutron fluence of $1.50\text{--}5.31 \times 10^{25} \text{ n/m}^2$ ($E > 0.18\text{MeV}$) and a time-average volume-average (TAVA) temperature of 854–1345°C. The experiment involved eight capsules, each with three stacks of four compacts—for a total of 32 compacts (data points)—along with the UCO fuel kernel.

In total, the AGR-2 and AGR-3/4 series led to 68 experimental observations of fractional Ag release when considering the UCO fuel kernel. In utilizing BISON, we made sure to use the correct experimental configurations for the AGR-2 and AGR-3/4 series, including aspects such as temperature, power level, and fast neutron flux. Furthermore, for the AGR-3/4 series, we re-weighted the BISON output to factor in the DTF particles, and excluded the ring simulations. Moreover, before performing the Bayesian UQ with the KOH framework, we verified that the outputs of the BISON input file used for the analysis in this report matched the BISON assessment results for both the AGR-2 and AGR-3/4 series.

2) Defining the experimental configuration for fitting the GP inadequacy term

To fit the GP inadequacy term, we must define the experimental configuration. Here, temperature history served as the experimental configuration. Figure 2 shows the temperature histories from the AGR-2 and AGR-3/4 series. One difficulty in using these in their entirety is that they are not consistently discretized with time across the AGR-2 and AGR-3/4 series. Temperature history interpolation and/or extrapolation led to significant errors in fitting the GP inadequacy term. On the other hand, selecting only the mean temperature for each experiment is insufficient, as the entire temperature history may be important for predicting fractional Ag release in BISON. Thus, we used temperature history summary statistics for fitting the GP inadequacy term. For each experiment’s temperature history, we computed summary statistics such as those presented in Table I. Also included among these summary statistics is the number of sudden temperature changes reflected in the history. These summary statistics generally led to good fitting of the GP inadequacy term, as assessed on the test sets in Section II-C3.

3) GP inadequacy term results

As discussed in Section II-B, the 68 available experimental data points were randomly divided into two equal batches. One batch was used for fitting the GP inadequacy term. To accomplish this, a standard Bayesian analysis was performed that considered all the available experimental data so as to obtain the

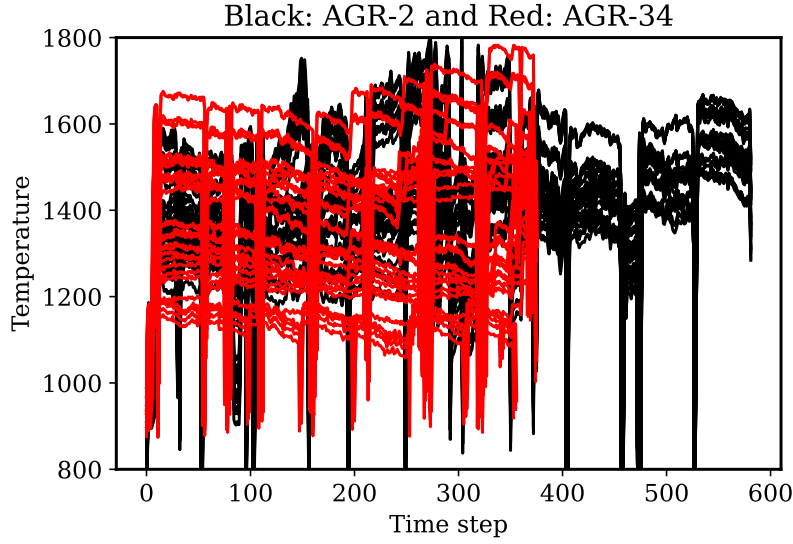


Fig. 2. Temperature histories from the AGR-2 and AGR-3/4 series.

TABLE I. Summary statistics of the temperature histories (i.e., experimental configurations) used for fitting the GP inadequacy term.

Summary statistic index	Value
1	Minimum temperature
2	Maximum temperature
3	Starting temperature
4	Ending temperature
5	Mean temperature
6	Standard deviation of temperature
7	Number of sudden temperature changes

best estimate of the model parameters. This best estimate was then used to compute the residuals for fitting the GP inadequacy term. Five iterations were performed by randomizing the division of the experimental data into two batches and fitting the GP inadequacy term on one of those batches in each iteration. This bootstrapping was performed to improve the robustness of the KOH framework and mitigate any over- or under-fitting, as has been explained.

Figure 3 shows the performance of the GP inadequacy/discrepancy term predictions, based on comparison against the true inadequacy/discrepancy for the training set, meaning the batch from the experimental dataset used for fitting the GP term. Overall, for the training set in each iteration, the GP predictions compare very satisfactorily with the true values, as indicated by the MSE values.

Figure 4 presents the performance of the GP inadequacy/discrepancy term predictions, based on comparison against the true inadequacy/discrepancy for the testing set, meaning the batch from the experimental dataset not used for fitting the GP term. Overall, for the testing set in each iteration, the GP predictions also compare satisfactorily with the true values, as indicated by the MSE values. In the

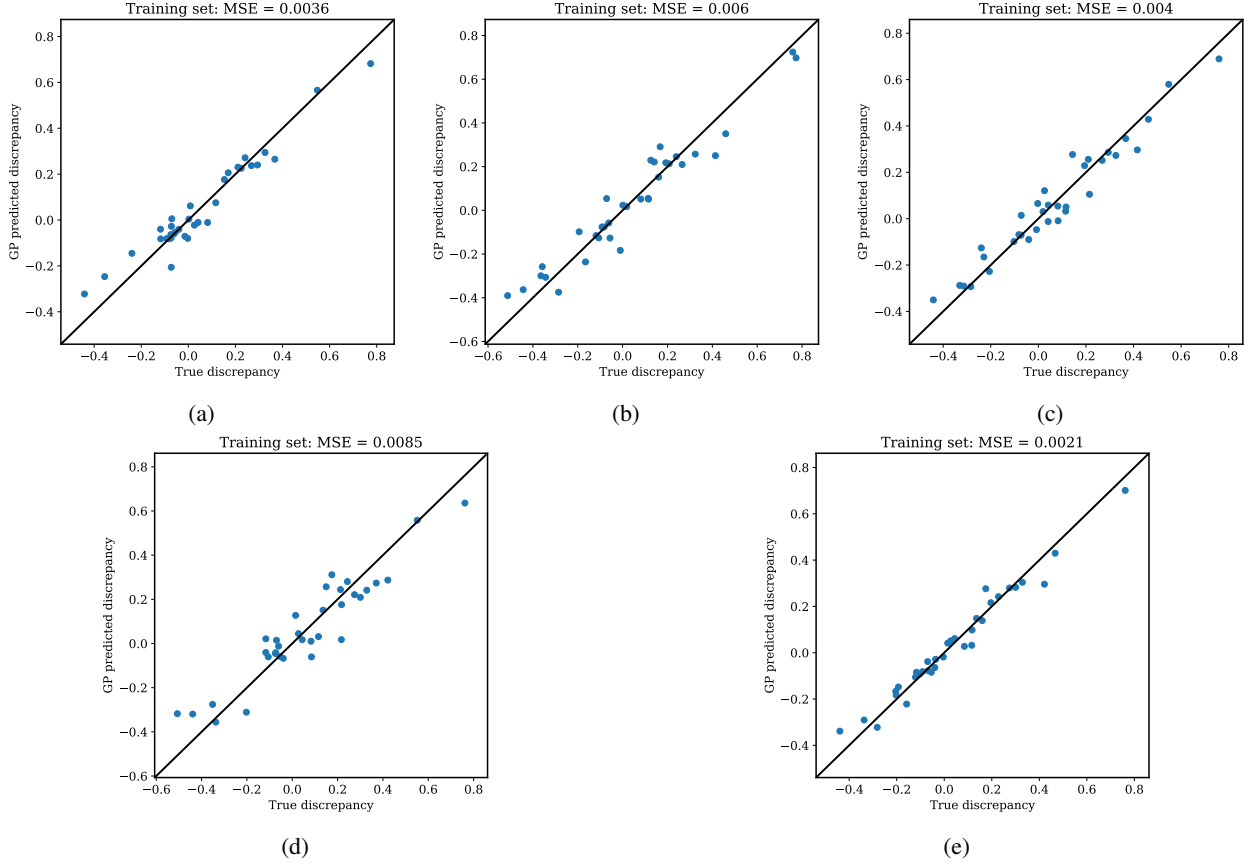


Fig. 3. Comparison of the GP-predicted and true model discrepancies when considering the training sets for (a) iteration 1, (b) iteration 2, (c) iteration 3, (d) iteration 4, and (e) iteration 5. The five iterations represent random selection of the 34 data points for training the GP discrepancy/inadequacy term on the AGR-2/3/4 experimental datasets, which encompass 68 data points in total. The MSE values are also shown in each plot.

case of iteration 2 (Figure 4b), the MSE value is slightly higher, due to the random selection of the training and testing sets. This is as expected, since during such randomization or bootstrapping we can sometimes end up with an inferior training set for the GP inadequacy term. Thus, it is recommended to pursue multiple iterations of the KOH framework with random selection of the training set for the GP term, then average the results across these iterations.

4) Parameter and experimental noise term posterior results

For each iteration of the KOH framework, upon fitting the GP inadequacy term using one batch of the random experimental data, the other batch is then used to infer the posterior distribution of $\{\theta, \sigma_\varepsilon\}$. Figure 5 presents the posterior distribution of the model parameters θ for all five KOH iterations. Also shown in this figure is the θ posterior obtained by applying the standard Bayesian framework, using the entire experimental dataset denoted as the “Base” dataset in the figure. Note that the θ posteriors across the five iterations of the KOH framework are consistent with the posterior obtained via the standard

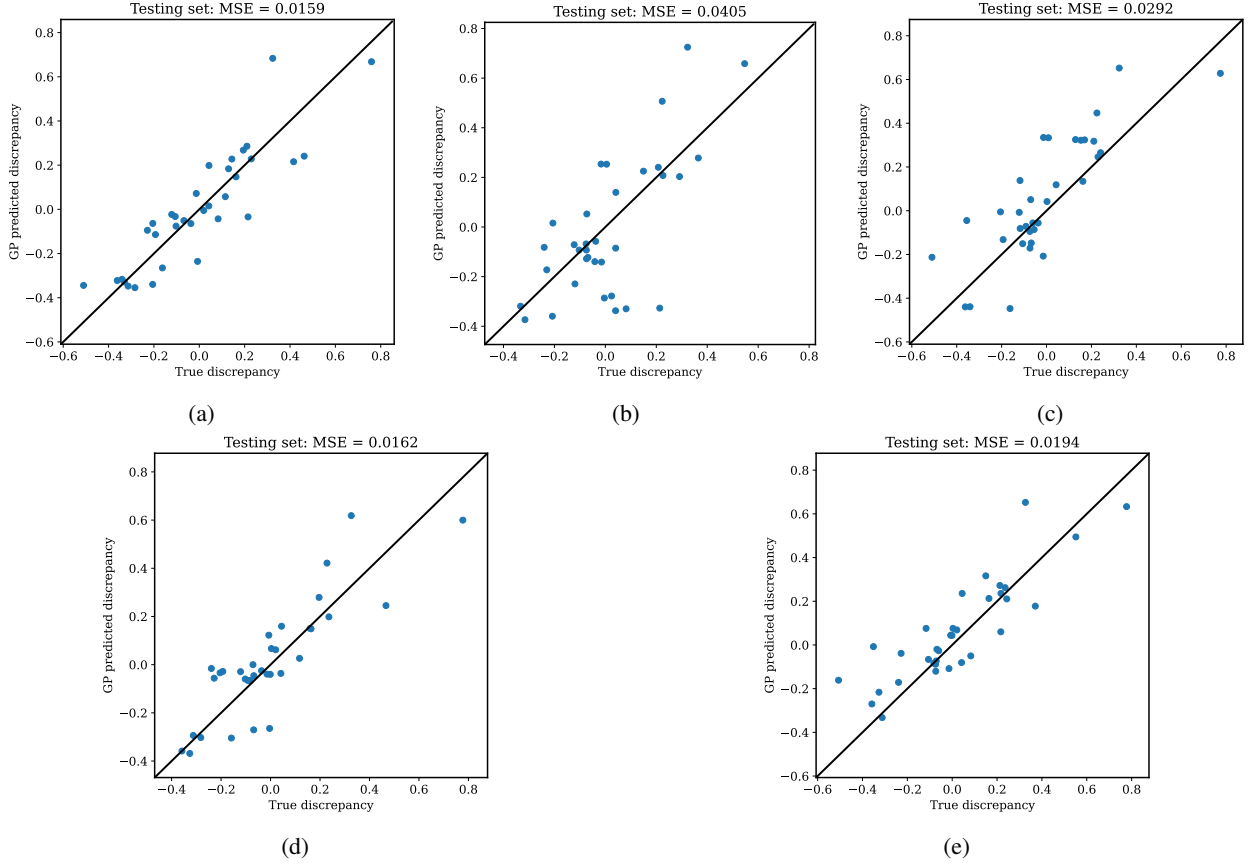


Fig. 4. Comparison of the GP-predicted and true model discrepancies when considering the testing set for (a) iteration 1, (b) iteration 2, (c) iteration 3, (d) iteration 4, and (e) iteration 5. The five iterations represent random selection of 34 data points not used for training but rather for testing the GP discrepancy/inadequacy term from the AGR-2/3/4 experimental datasets, which encompass 68 data points in total. The MSE values are also shown in each plot.

Bayesian framework. This consistency justifies using the standard-Bayesian-framework best estimate of the model parameters to fit the GP inadequacy term (see Section II-C3).

Figure 6 presents the posterior distribution of σ_ϵ for each of the five KOH iterations. Also shown in this figure is the σ posterior obtained based on the standard Bayesian framework, using the entire experimental dataset denoted as the “Base” dataset in the figure. Note that σ from the standard Bayesian framework represents the uncertainties stemming from model inadequacy and experimental noise. In contrast, the sigma value σ_ϵ from the KOH framework represents the experimental noise uncertainty, thus it is smaller than σ in each case. Table II presents the most likely sigma values from the five iterations of the KOH and the standard Bayesian approach. The mean sigma value over all five iterations of the KOH framework is $\sigma_\epsilon = 0.149$. The mostly likely sigma value from the standard Bayesian framework is $\sigma = 0.254$. Since σ_ϵ is supposed to represent the experimental noise uncertainty and σ represents both the model inadequacy and experimental noise uncertainties, it is interesting to decompose

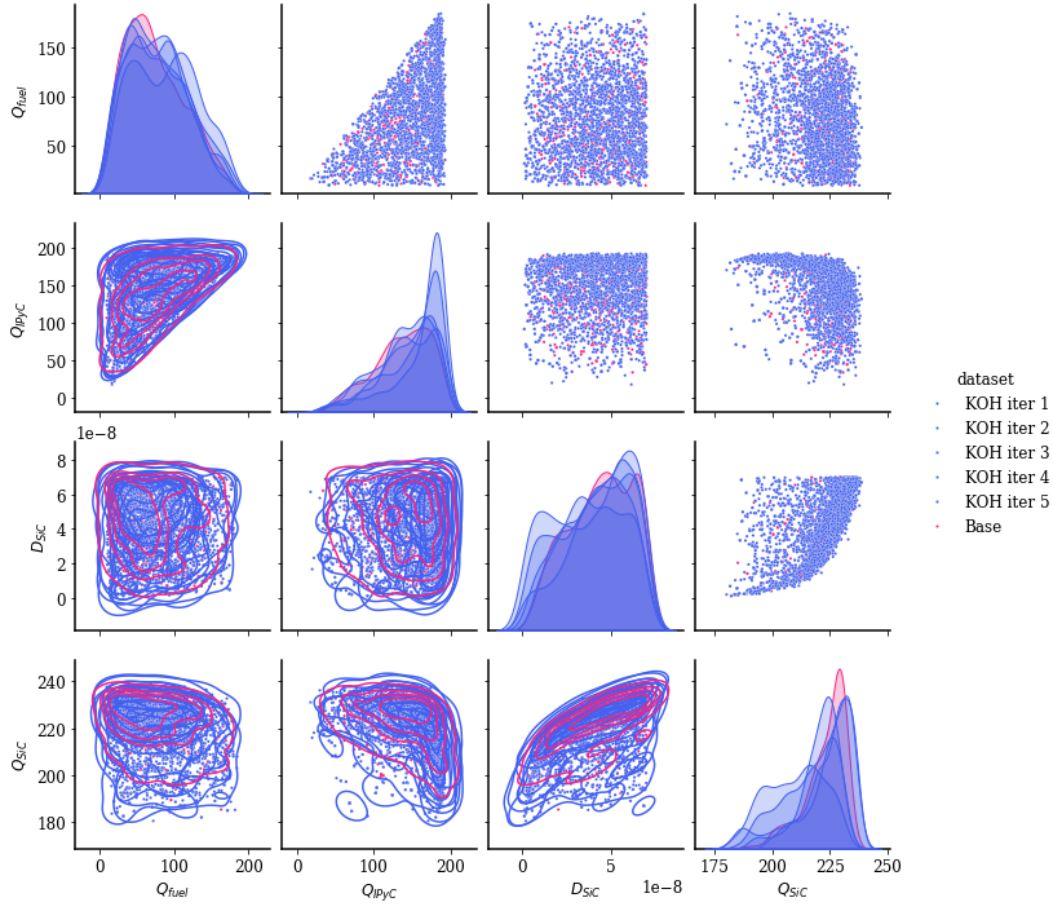


Fig. 5. Correlation matrix plot of the TRISO silver release model parameters when comparing the base (i.e., standard Bayesian) and the KOH Bayesian UQ frameworks.

the percentage that each of the two component uncertainties contributed to the total predictive uncertainty. In the TRISO fractional Ag release case, the average percentage of total uncertainty due to experimental noise was $\frac{\sigma_n^2}{\sigma^2} \times 100 = \frac{0.149^2}{0.254^2} \times 100 = 35\%$. And the average percentage of total uncertainty due to model inadequacy was $100 - 35 - 5 = 60\%$, with the 5% being assumed to represent the parametric uncertainty, which in prior studies (e.g., (Dhulipala et al., 2023)) is usually quite small. While this type of uncertainty decomposition analysis is quite interesting, care must be taken to ensure that the model inadequacy term is properly accounted for, and sensitivity analyses should be conducted to mitigate any possible over- or under-representation of this term. If adequate care is not taken, there is potential for confounding between model inadequacy and experimental noise uncertainties and a misrepresentation in the uncertainty decomposition analysis. Mitigation of such confounding of uncertainties in the KOH framework is an active area of research (e.g., MacCallum and O'Hagan (2015); Brynjarsdottir and O'Hagan (2014); Wu et al. (2024); Sung and Tuo (2024); Maupin and Swiler (2020); White and Mahadevan (2023); Leoni et al. (2024); Bayarri et al. (2009)).

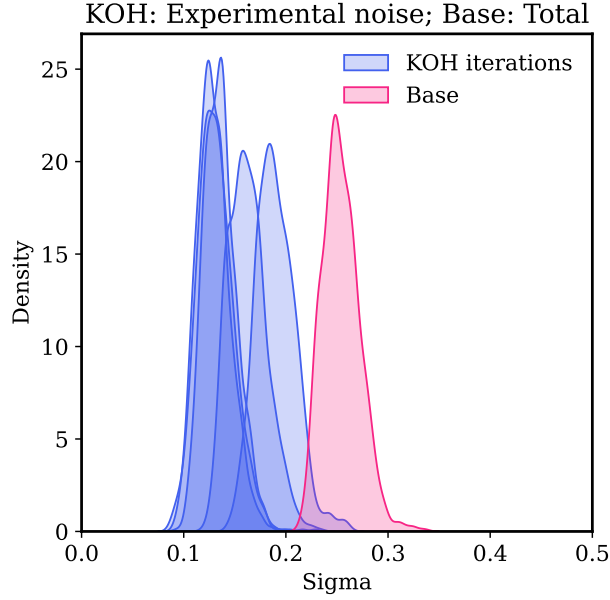


Fig. 6. Sigma values of the TRISO Ag release model in comparison to the base (i.e., standard Bayesian) and KOH Bayesian UQ frameworks. Note that the sigma value (σ) from the base framework represents both the model inadequacy and experimental noise uncertainties. In contrast, since the sigma value (σ_ε) from the KOH framework represents the experimental noise uncertainty only, it is smaller than σ in each case.

TABLE II. The most likely σ_ε values from five iterations of the KOH framework. The mean σ_ε value across all iterations of the KOH framework is also shown, along with the σ value from the base UQ framework.

	Value
KOH iteration 1	$\sigma_\varepsilon = 0.129$
KOH iteration 2	$\sigma_\varepsilon = 0.189$
KOH iteration 3	$\sigma_\varepsilon = 0.162$
KOH iteration 4	$\sigma_\varepsilon = 0.131$
KOH iteration 5	$\sigma_\varepsilon = 0.135$
KOH mean	$\sigma_\varepsilon = 0.149$
Standard Bayesian (Base)	$\sigma = 0.254$

σ_ε represents the experimental noise uncertainty.

σ represents both the model inadequacy and experimental noise uncertainties.

D. Forward Prediction and Uncertainty Quantification in Light of the AGR-1 Experimental Dataset

The forward UQ was applied to the AGR-1 dataset that was not part of the experimental data, in order to perform inverse analysis based on either the KOH framework or standard Bayesian framework. AGR-1 is the first in a series of irradiation tests sponsored by DOE's AGR program. Details on AGR-1 are given in Collin (2015). Surface temperature, power level, fast neutron flux, and experiment duration data were extracted from a data file prepared by the AGR program, and are available in BISON. For the UCO fuel kernel, 16 experimental values are available for fractional Ag release, disregarding two compacts for

which the corrected Post-Irradiation Examination (PIE) values are unavailable. These 16 experimental values were used to evaluate both the quality and predictive accuracy of uncertainties obtained via the KOH and standard Bayesian frameworks.

Posterior predictive distributions were obtained by solving Equation (14) via a Monte Carlo approach, based on the likelihood value defined in Equation (15). Predictive distributions were also computed for specific components of the total uncertainty. The (1) *parametric uncertainty* was computed by evaluating the BISON model over the parameter sets obtained from the posterior distribution of θ , then adding the mean GP inadequacy term; the (2) *GP inadequacy term uncertainty* was computed by randomly sampling the GP inadequacy term from the underlying distribution $f(\delta(\hat{\Theta})|\gamma_\delta)$, then adding to the BISON model prediction while ignoring the uncertainty due to σ_ϵ ; and the (3) *experimental noise uncertainty* was computed by randomly sampling from the likelihood value determined via Equation (15) using σ_ϵ but ignoring the uncertainty in the GP inadequacy term. Figure 7 presents the posterior predictive distributions representing the total uncertainty and its various components for 2 of the 16 AGR-1 compacts: namely, 5-3-3 (Figures 7a and 7c) and 6-4-1 (Figures 7b and 7d). Two iterations of the KOH framework are represented. Figures 7a and 7b reflect iteration 1. Figures 7c and 7d reflect iteration 3. Generally, as concluded in Dhulipala et al. (2023) and Dhulipala et al. (2024), the parametric uncertainty is a very small component of the total uncertainty. Also, in each case, the uncertainty in the GP inadequacy term is less than that due to experimental noise, as indicated by the flatness of the corresponding probability distributions. The probability distribution due to the experimental noise is almost as flat as the total uncertainty distribution, thus identifying the primary diver behind the total uncertainty.

Figure 8 presents the total uncertainty from the KOH and standard Bayesian frameworks for 2 of the 16 AGR-1 compacts: namely, 5-3-3 (Figures 8a and 8c) and 6-4-1 (Figures 8b and 8d). Two iterations of the KOH framework are represented. Figures 8a and 8b reflect iteration 1. Figures 8c and 8d reflect iteration 3. Generally, the most likely value (i.e., the fractional Ag release with the highest probability density) is closer to the experimental value predicted by the KOH framework than the one predicted by the standard Bayesian framework. Moreover, for Compact 5-3-3 (Figures 8a and 8c), where the predicted most likely values are far from the experimental value, the distributions from the KOH framework more adequately capture the experimental value than do those from the standard Bayesian framework. Next, a more quantitative comparison is made in terms of both the median accuracy and quality of the uncertainty estimates.

Figure 9 presents the median posterior predictions and 5-95 percentile confidence bands—along with the experimental values—for the KOH and standard Bayesian frameworks. More specifically, Figures 9a and 9b reflect KOH framework iterations 1 and 3, respectively, and Figure 9c reflects the standard Bayesian framework. It is generally argued that the KOH framework (all iterations) not only gives median predictions that more accurately reflect the experimental values, but also confidence bands that better represent those experimental values. To reinforce this argument, the MSE and CAL values were compared. The CAL value represents how well the confidence bands capture the experimental data for C possible confidence levels. If the C possible confidence levels are defined as $0 \leq p_1 \leq p_2 \leq \dots \leq p_C \leq 1$, the empirical frequency for each p_j is computed as (Kuleshov et al., 2018; Tran et al., 2020):

$$\hat{p}_j = \frac{|\{\mathcal{D}(\hat{\Theta}_t) | F(\mathcal{M}(\hat{\Theta}_t, \theta) | \Theta, \mathcal{D}, \gamma_\delta) \leq p_j \text{ s.t. } t = 1, \dots, T\}|}{T} \quad (16)$$

where t is the index of the experimental testing set, T is the total number of experimental points for

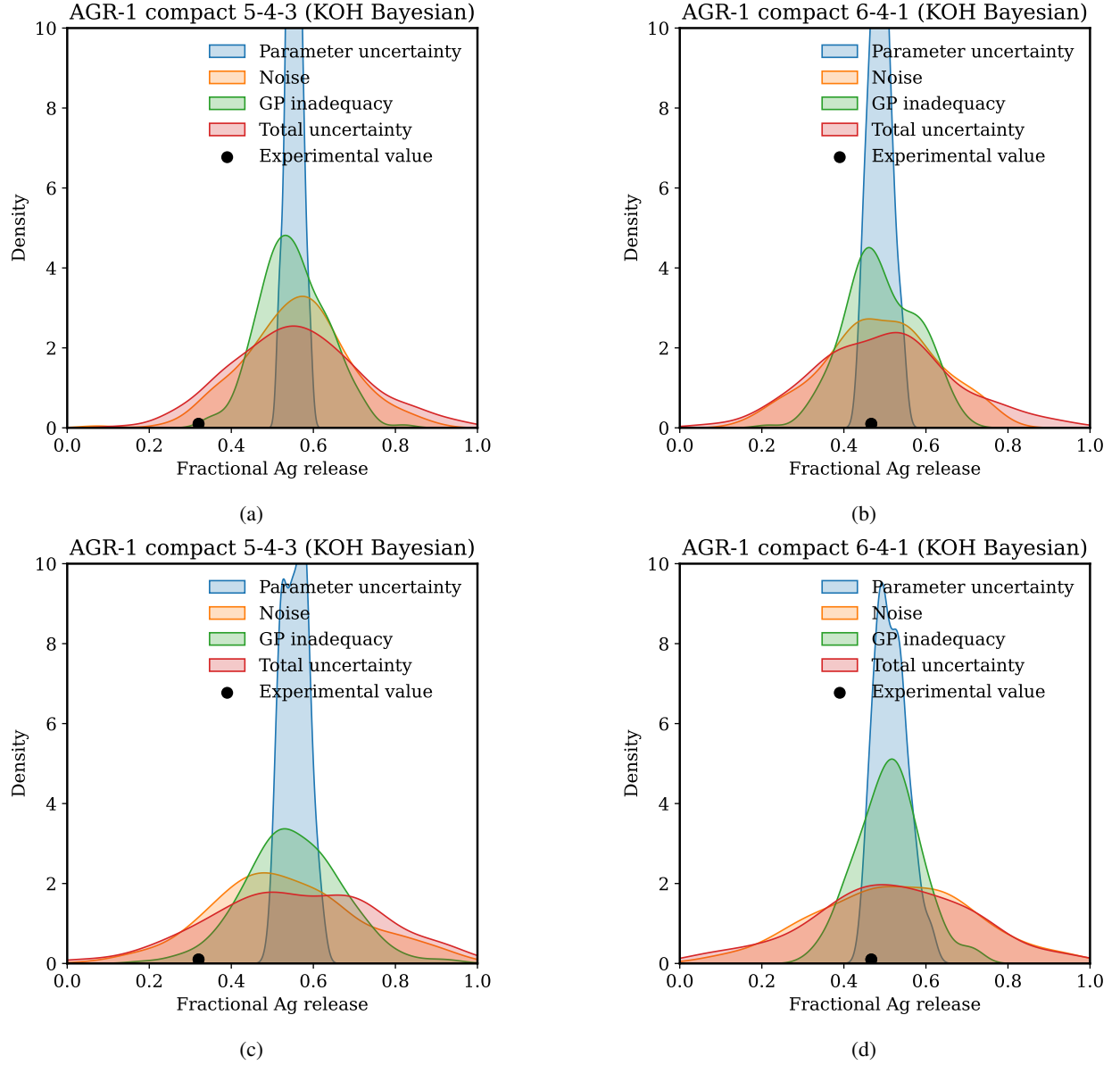


Fig. 7. Forward prediction, UQ, and decomposition on the AGR-1 dataset when using the KOH framework, with results shown for two compacts: (a) Compact 5-3-3, iteration 1; (b) Compact 6-4-1, iteration 1; (c) Compact 5-3-3, iteration 3; and (d) Compact 6-4-1, iteration 3. Each sub-plot shows the predictive uncertainty contributions stemming from different sources (model parameters, GP inadequacy, and experimental noise), the total predictive uncertainty, and the recorded experimental value.

testing, $\mathcal{D}(\hat{\Theta}_t)$ is the experimental value for the t^{th} configuration in the testing set, and $F(\cdot)$ is the cumulative distribution function of the probability distribution in Equation (14). From Equation (16),

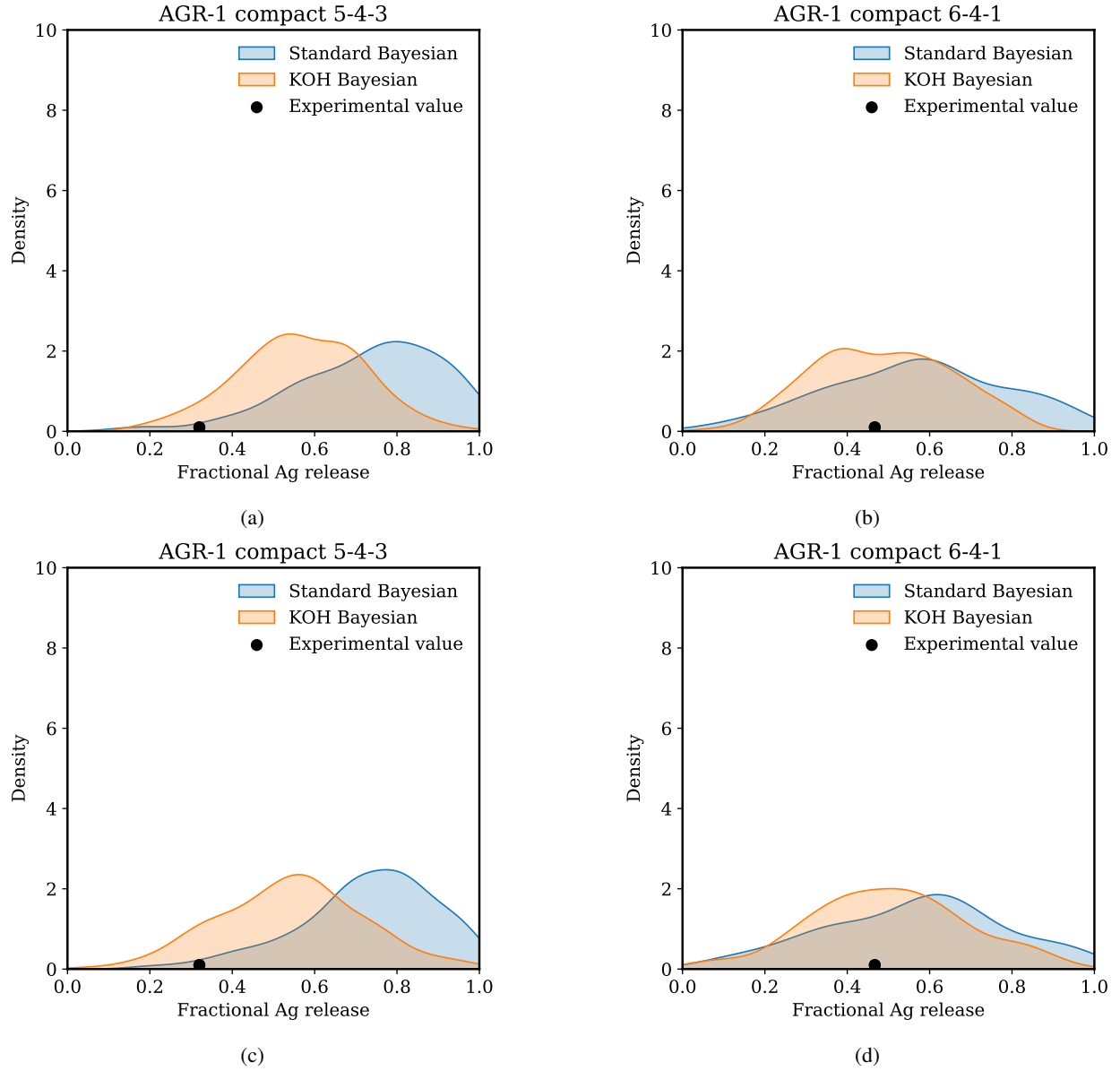


Fig. 8. Comparison of the predictive uncertainty estimates obtained from the KOH framework and standard Bayesian framework on the AGR-1 dataset when considering two compacts: (a) Compact 5-3-3 KOH, iteration 1; (b) Compact 6-4-1 KOH, iteration 1; (c) Compact 5-3-3 KOH, iteration 3; and (d) Compact 6-4-1 KOH, iteration 3. In general, the KOH framework leads to better median predictions and uncertainty estimates when comparing against the experimental values.

we can compute the CAL value so as to assess the quality of the predictive uncertainty estimates, per: (Kuleshov et al., 2018; Tran et al., 2020):

$$CAL = \frac{\sum_{j=1}^C (p_j - \hat{p}_j)^2}{C} \quad (17)$$

A lower CAL value indicates higher quality of the uncertainty estimates in terms of encompassing the experimental values.

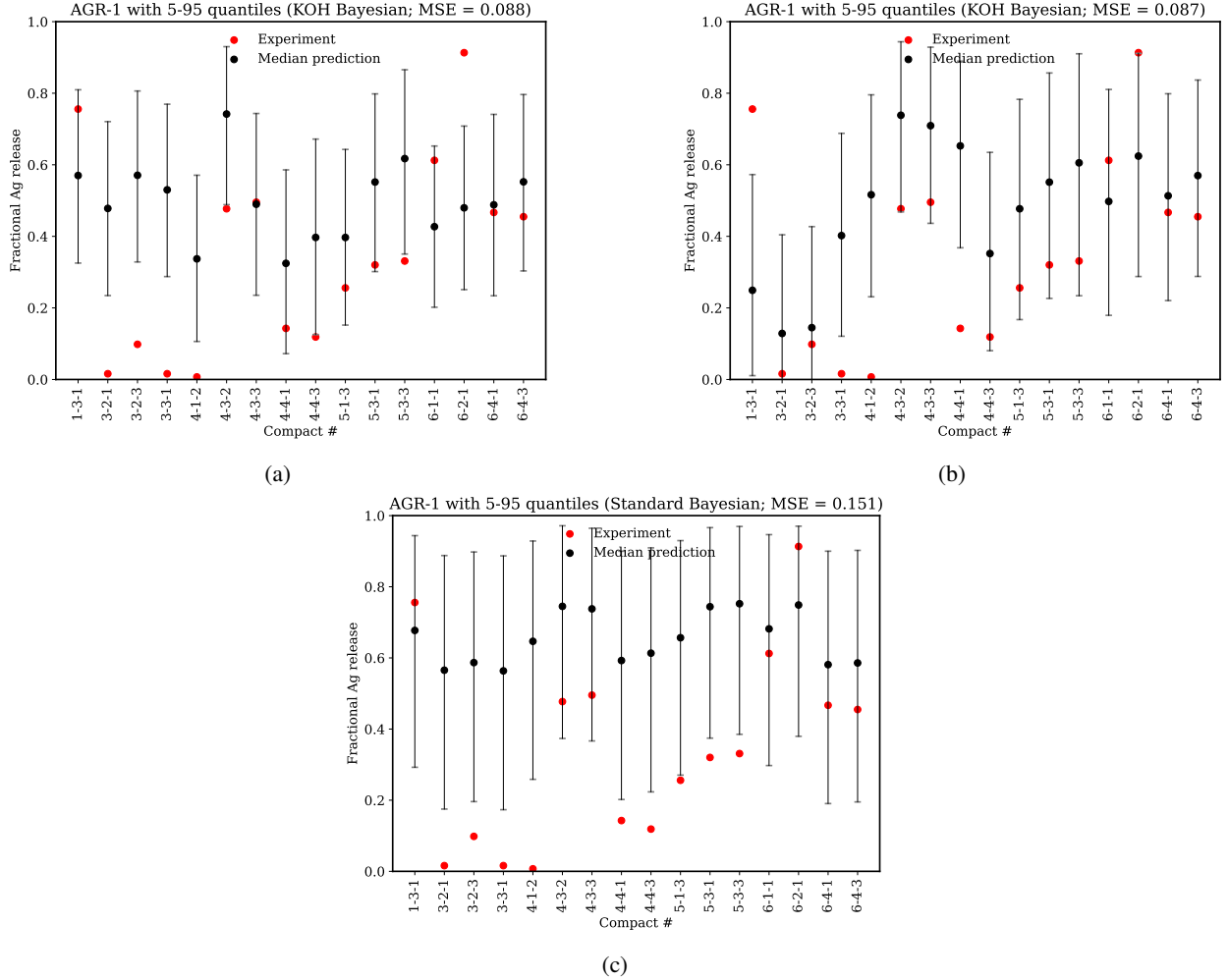


Fig. 9. Median predictions and 5-95 percentile confidence bands for the AGR-1 dataset, obtained using: (a) KOH framework, iteration 1; (b) KOH framework, iteration 3; and (c) the standard Bayesian framework. In general, the KOH framework leads to better median predictions and confidence bands when comparing against the experimental values.

Table III presents the MSEs between the median posterior predictions and the experimental values for the KOH and standard Bayesian frameworks. The average MSE value across the five KOH iterations is 32.4% less than that for the standard Bayesian framework. Table IV presents the CAL values quantifying the uncertainty quality. The average CAL value across the five KOH iterations is 42.5% less than that for

the standard Bayesian framework. These comparisons indicate that the KOH framework generally leads to better median predictions and uncertainty estimates than does the standard Bayesian framework.

TABLE III. MSEs between the median predictions and experimental values from the AGR-1 dataset, demonstrating the accuracy of the UQ approaches. Median predictions were obtained via the KOH framework (five iterations) and standard Bayesian framework. The average MSE across the five iterations is 32.4% less than that for the standard Bayesian framework, indicating greater accuracy.

	MSE
KOH iteration 1	0.088
KOH iteration 2	0.108
KOH iteration 3	0.087
KOH iteration 4	0.119
KOH iteration 5	0.112
KOH mean	0.102
Standard Bayesian (Base)	0.151

TABLE IV. CAL values between the median predictions and the experimental values from the AGR-1 dataset, demonstrating the quality of the uncertainty estimates of the UQ approaches by comparing against the experimental values. CAL values were computed for the KOH framework (five iterations) and standard Bayesian framework. The average CAL value across the five iterations was 42.5% less than that for the standard Bayesian framework, indicating better uncertainty estimates.

	CAL
KOH iteration 1	0.0394
KOH iteration 2	0.0396
KOH iteration 3	0.0568
KOH iteration 4	0.0679
KOH iteration 5	0.0511
KOH mean	0.0509
Standard Bayesian (Base)	0.0886

E. Summary and Conclusions

This chapter discussed the theory of accounting for model inadequacy through application of the KOH framework, which was used for inverse UQ based on the AGR-2/3/4 experimental data, and for forward predictive UQ based on the AGR-1 data. The results obtained from comparing the KOH and standard Bayesian frameworks were also presented. The following main conclusions were drawn:

- The model parameter θ posteriors are consistent between the KOH framework and the standard Bayesian framework.
- The experimental noise uncertainty from the KOH framework averages about $\sigma_\epsilon = 0.149$. The sigma value from the standard Bayesian framework is about $\sigma = 0.254$. The average percentage of total uncertainty due to experimental noise is $\frac{\sigma_\epsilon^2}{\sigma^2} \times 100 = \frac{0.149^2}{0.254^2} \times 100 = 35\%$. The average percentage of total uncertainty due to model inadequacy is $100 - 35 - 5 = 60\%$, with the 5% being assumed to represent parametric uncertainty, which prior studies have revealed to usually be quite small.

However, note that the KOH framework is affected by potential confounding of uncertainties. Thus, to limit the potential for such confounding, we ran multiple iterations of the KOH framework and then used the average values.

- When predicting on the AGR-1 dataset, the KOH framework led to both higher predictive accuracy and uncertainty quality than did the standard Bayesian framework. The KOH framework reflected a 32.4% and 42.5% improvement in predictive accuracy and uncertainty quality, respectively, as assessed in terms of the MSE and CAL metrics, respectively.

III. INCORPORATING LOWER-LENGTH-SCALE EFFECTS INTO THE UNCERTAINTY QUANTIFICATION FRAMEWORK

This chapter discusses the incorporation of LLS effects into the Bayesian UQ framework. Specifically, we discuss the generation of LLS simulation data pertaining to effective diffusivity as a function of temperature and the microstructure information and explore the development of models to bridge the scales. Next, inverse UQ on the AGR-2/3/4 data and forward UQ on the AGR-1 data are compared in light of the Arrhenius equation and two LLS-informed models for Ag diffusivity in the SiC layer. Inverse UQ is performed using both the standard Bayesian and KOH approaches. Forward predictive UQ is compared in regard to each of the different approaches (i.e., the Arrhenius equation and LLS-informed models for the diffusivity and standard Bayesian, and KOH for the UQ) in terms of both the accuracy and quality of the UQ, using the MSE and CAL metrics, respectively.

A. *Strategies for Incorporating Lower-Length-Scale Effects*

Figure 10 presents two strategies for incorporating LLS effects into the UQ framework. We simulated effective diffusivity at the LLS by considering a wide range of microstructures, which are characterized by their average minor and major axis lengths. Details on the LLS simulations are provided in Section III-B. Using the LLS data, we fit a parametric equation of the effective diffusivity as a function of temperature and the microstructure information (particularly, the minor axis length). We also fit a non-parametric GP model of the effective diffusivity as a function of temperature and the microstructure information (i.e., both the minor and major axis lengths). Using these LLS-informed models for Ag diffusivity in the SiC layer, we performed a Bayesian inverse UQ analysis on the AGR-2/3/4 datasets. We used both the standard Bayesian and KOH frameworks to perform the inverse UQ. We then performed a forward UQ by computing the posterior predictive distributions for the AGR-1 dataset, which was not used in the inverse UQ process. Further details on these strategies, along with the results thereof, are presented next.

B. *Updated Lower-Length-Scale Simulation Database*

In previous years, the NEAMS program has supported development of multi-scale models (called LLS models) of Ag diffusion in SiC in order to increase BISON's fidelity at predicting fission product release from TRISO particles. This work was first described in Simon et al. (2022), in which a temperature- and microstructure-dependent effective Ag diffusivity model was developed that improved BISON's predictions of Ag release from TRISO particles under AGR-1 conditions. Using atomistic simulations of bulk and grain boundary (GB) diffusivity and a mesoscale evaluation of the effect of grain size, performed using the fuel performance code MARMOT, the new model captured and quantified microstructural effects that empirical fits could not (Simon et al., 2022).

However, in this first version of the model, only a limited number of SiC microstructures (20 in total) were used to capture microstructure effects. The effect of microstructure was captured, but with significant uncertainty. To decrease this uncertainty and increase the model's range of definition, additional microstructures were generated and the effective diffusivity model was fitted to a wider, more refined range of grain sizes. This recent update is described in Simon et al. (2024), and is detailed below for the sake of completeness.

The same approach as implemented in Simon et al. (2021) was used to generate polycrystalline structures. However, as stated above, more microstructures were used than in Jiang et al. (2021); Simon

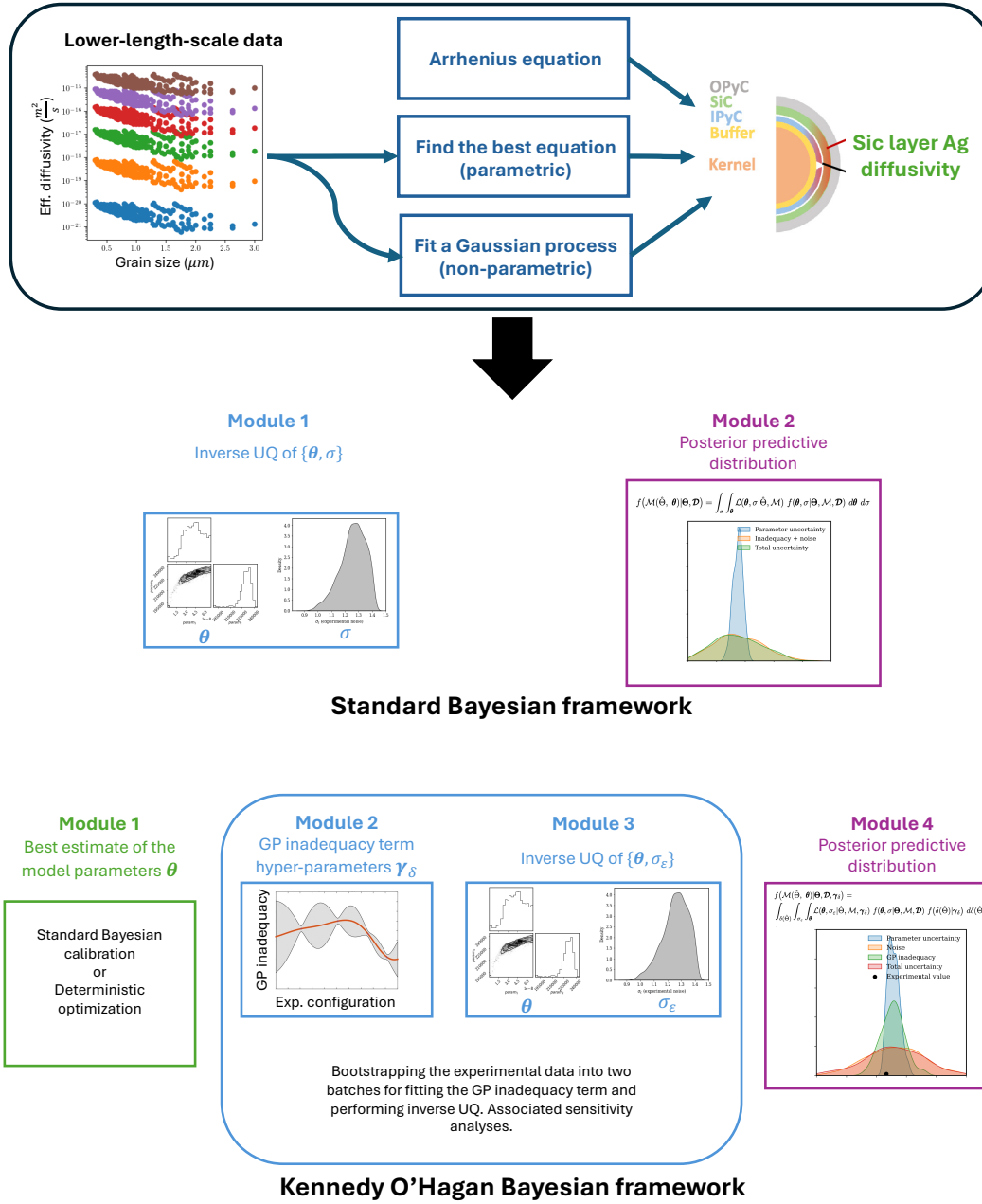


Fig. 10. Strategies for incorporating LLS effects into the UQ framework.

et al. (2022) so as to better capture the microstructure effects. We used 169 2-D domains of dimensions $9 \mu\text{m} \times 9 \mu\text{m}$ generated with the desired grain sizes, grain aspect ratios, and GB widths (as opposed to the 20 domains in Jiang et al. (2021); Simon et al. (2022)). Their grain major/minor axis lengths vary from

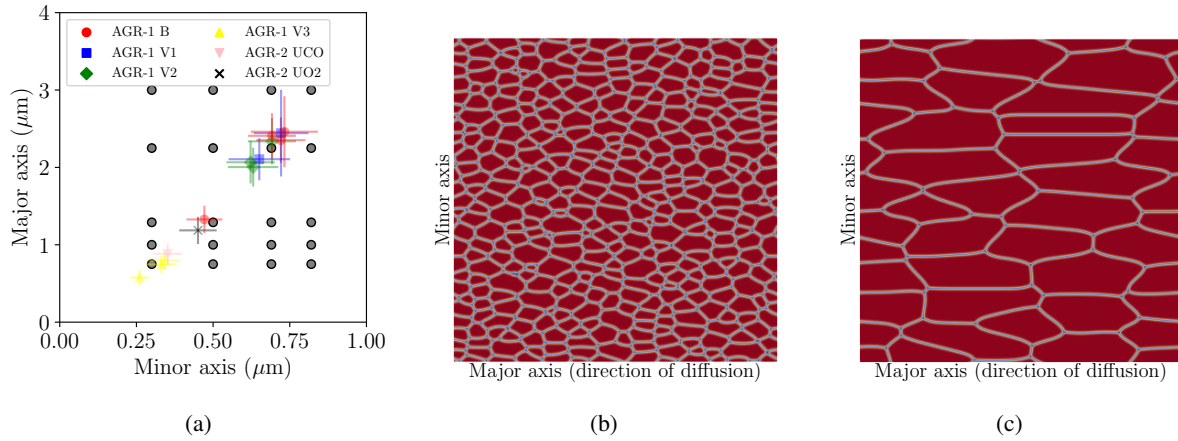


Fig. 11. (a) Simulated grain dimensions (black dots) compared against the AGR-1 as-fabricated SiC grain dimensions, with twins included for all batches provided by Gerczak et al. (2016b). The simulated microstructures cover the variations in grain dimensions found in the AGR-1 fuel batches. (b) and (c) reflect two examples of polycrystalline SiC microstructures created for this study, corresponding to the (b) bottom-left and (c) top-right points shown in (a). The domain size is $9 \times 9 \mu\text{m}$. Grains are shown in red, the centers of the 60-nm-thick GBs are in blue, and the continuous transition between the two are shown in shades of white. This figure was reproduced from Jiang et al. (2021); Simon et al. (2022).

0.3 to $3 \mu\text{m}$ to cover the grain sizes measured from the AGR-1 and AGR-2 SiC layers Gerczak et al. (2016b), as shown in Fig. 11. Though the GB size of the polycrystalline SiC is approximately equal to 1 nm, with variations between high- and low-angle GBs, the GB width was scaled up to 60 nm to reduce the computational costs. To compensate for the artificially expanded GBs, the GB and bulk diffusivities were carefully scaled down and up, respectively, using the approach developed in Jiang et al. (2021); Simon et al. (2022). This approach ensures lower computational costs while also minimizing scaling errors. As mentioned above, the simulated microstructures cover the variations in grain dimensions found in the AGR-1 and AGR-2 fuel batches, thus affording results relevant to realistic SiC grain sizes for TRISO fuel particles.

Moreover, we extended the temperature range of the LLS model to better capture the temperature regimes of AGR-1 and AGR-2. Simulations were performed for all 169 microstructures, with temperatures ranging from 1000 to 2000 K, at 200 K increments, significantly increasing the temperature range from Jiang et al. (2021); Simon et al. (2022) (i.e., 1450–2073 K) toward temperatures relevant to AGR-1 and AGR-2 (Hunn et al., 2015; Stempien et al., 2021; Collin, 2015, 2016; Collin et al., 2015; Gerczak and Hunn, 2022). Additional details on the LLS simulations and the derivation and validation of the LLS model are given in Simon et al. (2022, 2024).

This new augmented and refined LLS simulation database is shown in Fig. 12, with the Ag effective diffusivity being plotted as a function of temperature, grain minor/major axis length, and grain size (i.e., average grain diameter). The effect of temperature and grain minor axis length is clearly visible in Fig. 12a. The lack of impact of the major axis length is apparent in Fig. 12b.

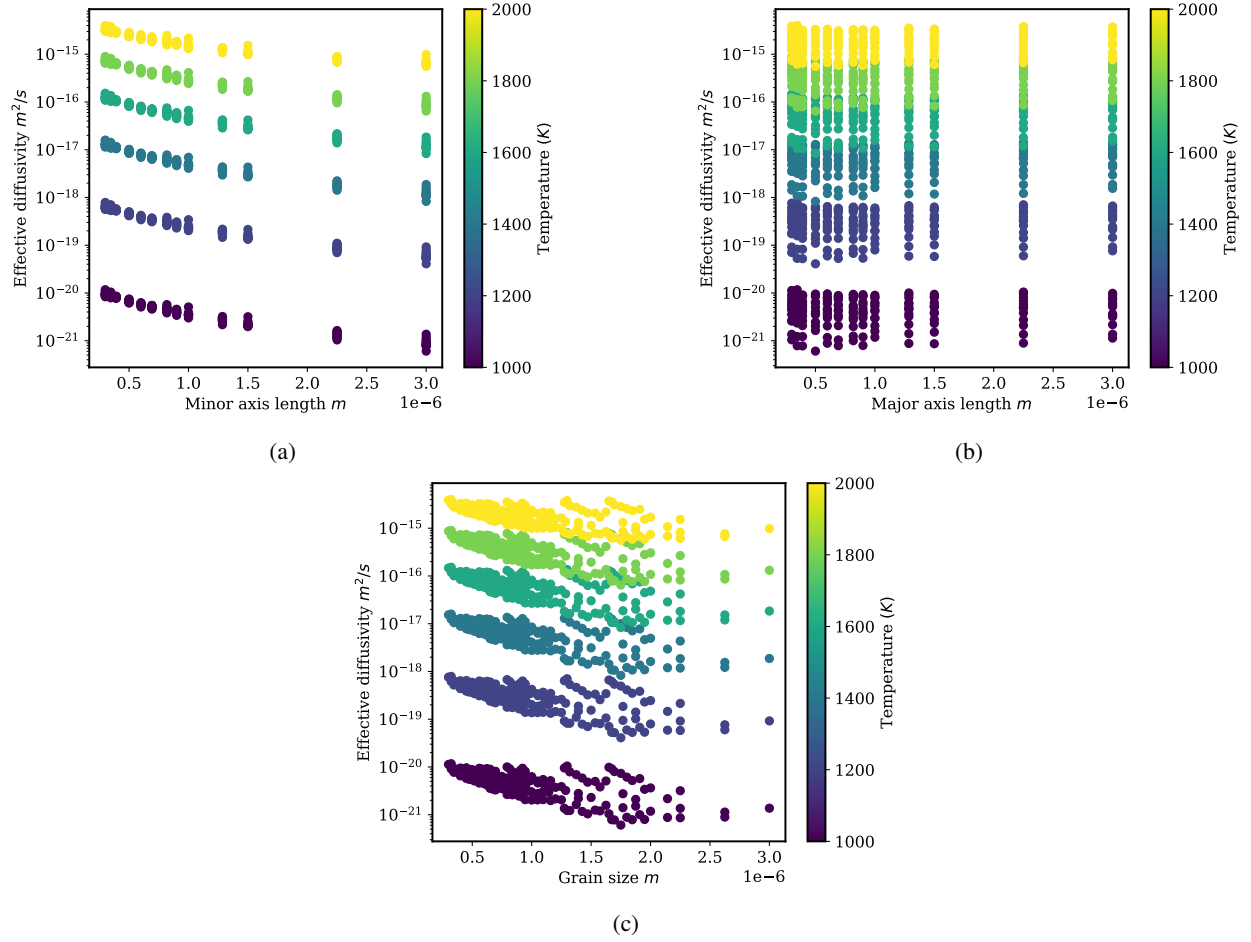


Fig. 12. Augmented LLS simulation database for the effective Ag diffusivity. The diffusivity is presented as a function of temperature, grain (a) minor and (b) major axis length, and (c) size.

C. Models for Bridging the Lower-Length-Scale and the Engineering-Scale

Bridging the gap between the LLS and the engineering scale required development of a function to capture the LLS data presented in Section III-B. Such a function could then be implemented in BISON to accurately capture multiscale, high-fidelity insights while limiting computational costs. To that end,

Simon et al. (2024) captured the LLS data as:

$$\begin{aligned}
D_{th}(T, m_i) &= D_{m,0}(m_i/m_0)^{D_{m,1}} \exp(-(Q_{m,0} + m_i Q_{m,1})/RT) \\
\text{where: } D_{m,0} &= \exp(-30.49) \text{ m}^2/\text{s} \\
D_{m,1} &= -0.65 \text{ (-)} \\
Q_{m,0} &= 210620.825 \text{ J/mol} \\
Q_{m,1} &= 3328326006.32 \text{ J/mol/m} \\
m_0 &= 1.0 \text{ m} \\
m_i &= \text{the grain minor axis length.}
\end{aligned} \tag{18}$$

This function appropriately captures the LLS data, as shown in Fig. 13.

Another approach we took to bridge the LLS and engineering scale was to fit a GP model of the thermal diffusivity as a function of temperature and both the minor and major axis lengths. Finding the best parametric functional form given the training dataset can be challenging. Thus, a GP is a data-driven approach to link the thermal diffusivity to the temperature and the microstructure information. It can serve as an alternative to the traditional parametric functional form approach followed in Simon et al. (2022, 2024). The theory of GP was presented in Chapter II, Section II-A2. Essentially, thermal diffusivity is expressed as:

$$\begin{aligned}
D_{th}(T, m_i, m_a) &\sim \mathcal{N}(m(\{T, m_i, m_a\}), k(\{T, m_i, m_a\}, \{T, m_i, m_a\}') | \gamma_{D_{th}}) \\
\text{where, } \gamma_{D_{th}} &= \{1.416, 1.768, 8.965, 28.396\}
\end{aligned} \tag{19}$$

where \mathcal{N} is a normal distribution with mean $m(\cdot)$ and covariance $k(\cdot, \cdot)$, and $\gamma_{D_{th}}$ are the GP hyperparameters to be optimized from the LLS simulation dataset. The optimized values of the hyperparameters are also presented in Equation (19); the first value represents the amplitude scale, and the next three represent the length scales for temperature, minor axis length, and major axis length. We again used the squared exponential kernel for the GP. Its definition was presented in Equation (9).

For comparison purposes, we also fit the standard Arrhenius equation to the LLS simulation data as a function of the temperature only. The functional form and the optimized parameters are:

$$\begin{aligned}
D_{th}(T) &= D_0 \exp\left(-\frac{Q_0}{RT}\right) \\
\text{where, } D_0 &= 1.013e-10 \text{ m}^2/\text{s} \\
Q_0 &= 193153.66 \text{ J/mol}
\end{aligned} \tag{20}$$

We fit this standard Arrhenius equation to compare the impacts of including microstructure information when predicting effective diffusivity at the LLS.

Figure 14 compares the LLS simulated effective diffusivity against the prediction made by the model previously discussed. Specifically, Figures 14a, 14b, and 14c correspond to the Arrhenius equation (20), LLS-informed equation (18), and LLS-informed GP (Equation (19)), respectively. Also presented in these figures are the MSE values computed in the logarithmic space. Both the LLS-informed models have a substantially low MSE value in the logarithmic space as compared to the Arrhenius equation. This indicates that the two LLS-informed equations better capture the effective diffusivities, due to incorporating the microstructure information. The impact of including microstructure information in the

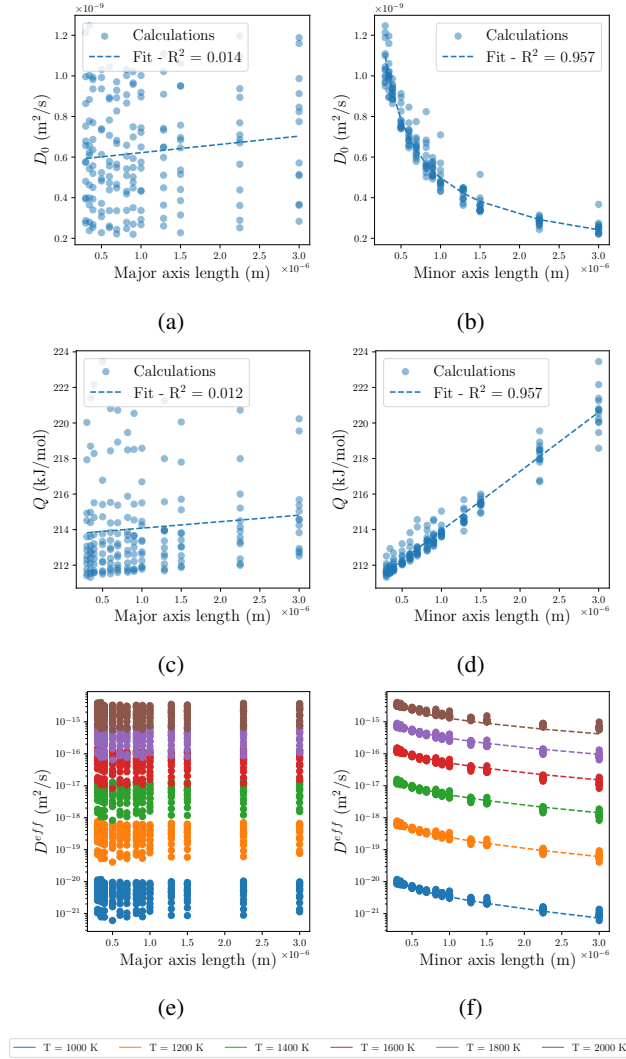


Fig. 13. Derivation of the thermal contribution to the effective Ag diffusion coefficient as a function of temperature T , grain minor axis length m_i , and grain major axis length m_a . (a,b) show how D_0 depends on (a) m_a and (b) m_i , and (c,d) show how Q depends on (c) m_a and (d) m_i . While both D_0 and Q are independent of m_a , D_0 decreases with increasing m_i , and Q increases linearly with m_i . (e,f) show how the thermal contribution to the effective diffusion coefficient $D^{eff,th}$ depends on the (e) m_a and (f) m_i for different temperatures. In (a, c, e), the points provided for one m_a value (and one temperature for [e]) correspond to the different minor axis lengths. Similarly, in (b,d,f), the points provided for a given value of m_i (and one temperature for [f]) correspond to the different major axis lengths. This figure was reproduced from Jiang et al. (2021); Simon et al. (2022).

diffusivity model when it comes to engineering-scale Ag release predictions is investigated next. Going forward, we will give particular consideration to the LLS GP model, as it and the LLS functional form model both have very similar MSE values, as shown in Figure 14.

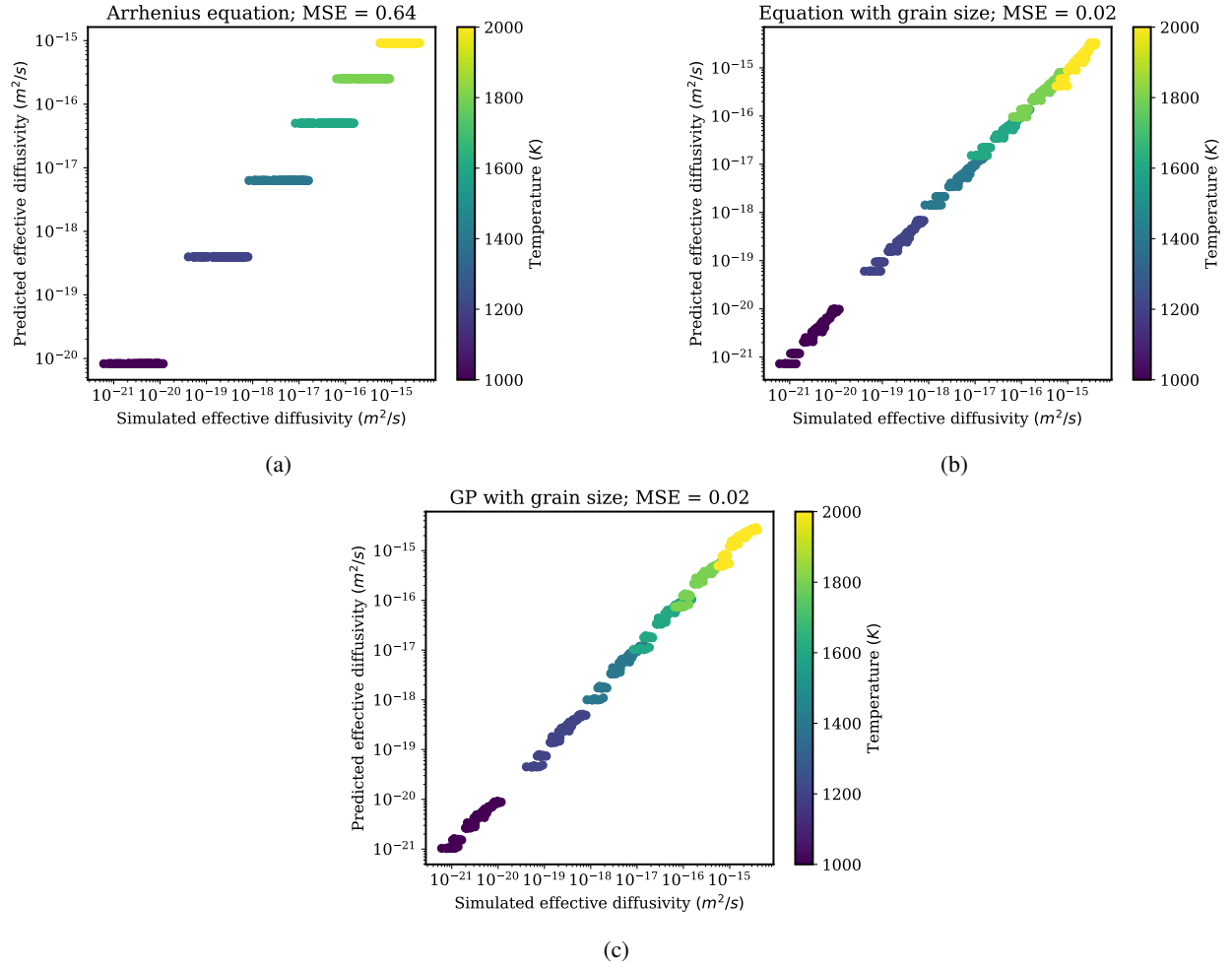


Fig. 14. Comparison between the LLS simulated effective diffusivity and the predictions made by the different LLS-informed models: (a) standard Arrhenius equation (no microstructure information considered), (b) parametric equation as a function of temperature and minor axis length, and (c) GP as a function of temperature, minor axis length, and major axis length. The MSE value, computed in the logarithmic space, for each model is also shown.

D. Inverse Uncertainty Quantification on the AGR-2/3/4 Experimental Datasets

Using the LLS-informed GP model for Ag diffusivity in the SiC layer, we first performed a standard Bayesian analysis. That is, we lumped together the uncertainty components for model inadequacy and experimental noise. We included a multiplicative factor, which is to be calibrated, to the LLS-informed GP model to help add flexibility when making engineering-scale predictions. The prior distributions used for performing the standard Bayesian analysis are presented in Table V. Note that the minor and major axis lengths are also included as model parameters for the Bayesian analysis. The priors for these axis lengths were obtained from experimental measurements of TRISO SiC microstructure Gerczak et al. (2016a). Unlike the type of Bayesian analysis that utilizes the standard Arrhenius equation without microstructure

information (see the FY-23 report Dhulipala et al. (2023)), we did not include inequality constraints on the activation energies for the fuel kernel, inner PyC (IPyC), and SiC. That is, $Q_{kernel} < Q_{pyc} < Q_{sic}$ is not imposed when using the LLS-informed GP model, due to the lack of an explicit activation energy term for the SiC layer. The associated impact on the quantified uncertainties when making predictions will be negligible, as is discussed next.

TABLE V. Prior distributions over the parameters, along with the σ (or σ_ϵ) term for the TRISO model, when considering the LLS-informed GP model for Ag diffusivity in SiC.

Model parameter	Distribution
Q_{kernel}	$\mathcal{U}(10.0, 300.0)$
Q_{pyc}	$\mathcal{U}(10.0, 300.0)$
Factor	$\mathcal{U}(0.0, 1000.0)$
Minor axis length	$\mathcal{TN}(0.35e-6, 0.05e-6, 0.05e-6, 0.65e-6)$
Major axis length	$\mathcal{TN}(0.85e-6, 0.1e-6, 0.65e-6, 1.2e-6)$
σ^2 (or σ_ϵ^2)	$\mathcal{TN}(0.01, 0.02, 0.0, 1.0)$

Figure 15 presents the prior and posterior distributions of the parameters θ obtained via the standard Bayesian framework. The parameters Q_{kernel} , Q_{sic} , and $Factor$ seem to be concentrated in one or two regions of the parameter space compared to their corresponding priors. The posteriors for the minor and major axis lengths closely match their priors. Figure 16 presents the σ value posterior. Also presented for comparison in this figure is the σ value posterior obtained when using the standard Arrhenius equation. The two σ posteriors closely match each other, indicating that the LLS-informed GP model for SiC Ag diffusivity has a similar level of predictive uncertainty in comparison to the Arrhenius equation. Although the LLS-informed GP model outperforms the Arrhenius equation in predicting the effective diffusivity at the LLS, at the engineering-scale, temperature seems much more dominant than SiC microstructure in terms of controlling the Ag diffusivity. Next, uncertainty decomposition via the KOH framework is performed and the results compared to those of the standard Arrhenius equation from Chapter II.

As discussed in Chapter II, Sections II-B and II-C, the same procedure was adopted to model the inadequacy term in the KOH framework. The 68 experimental data points were once again split into two batches, one for fitting the inadequacy term and the other for inversely estimating $\{\theta, \sigma_\epsilon\}$ for five iterations of the KOH framework (see Chapter II for more details of the KOH framework). Figure 17 compares the performance of the GP inadequacy/discrepancy term predictions against the true inadequacy/discrepancy for the training set. Overall, for the training set in each iteration, the GP inadequacy predictions compare very satisfactorily with the true values, as indicated by the MSEs.

Figure 18 presents the performance of the GP inadequacy/discrepancy term predictions by comparing them against the true inadequacy/discrepancy for the testing set, meaning the batch from the experimental dataset not used for fitting the GP term. Overall, for the testing set in each iteration, the GP predictions also compare satisfactorily with the true values, as indicated by the MSEs. As with the Arrhenius equation in Chapter II, for iteration 2 (Figure 18b), the MSE is slightly higher, due to the random selection of the training and testing sets. As was recommended, multiple iterations of the KOH framework should be pursued, with random selection of the training set for the GP term, then averaging the results across these iterations.

For each iteration of the KOH framework, upon fitting the GP inadequacy term using one batch of the random experimental data, the other batch was used to infer the posterior distribution of $\{\theta, \sigma_\epsilon\}$. Figure

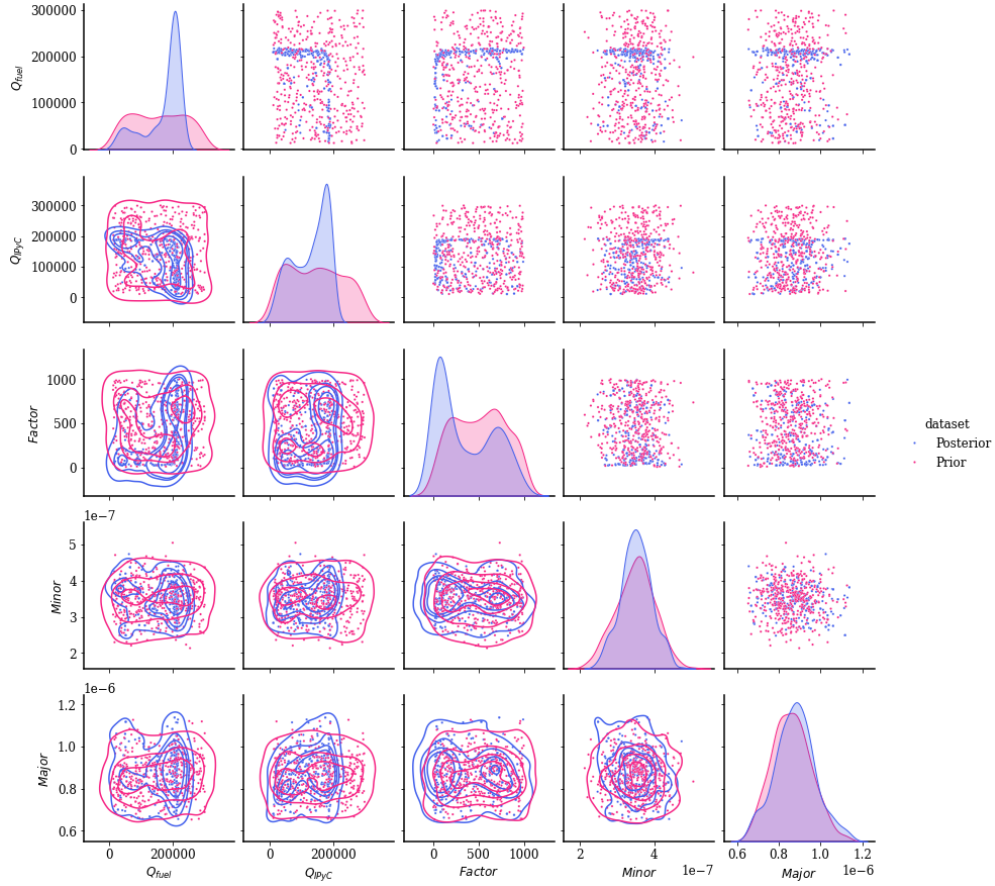


Fig. 15. Posterior distributions of the model parameters θ obtained using the standard Bayesian framework when considering LLS effects on the SiC diffusivity through a GP model. Also shown for comparison are the prior distributions.

19 presents the posterior distribution of the model parameters θ for all five KOH iterations, considering the LLS-informed SiC diffusivity model. Also shown in this figure is the θ posterior obtained based on the standard Bayesian framework when using the entire experimental dataset denoted as the “Base” dataset in the figure, and again considering the LLS-informed SiC diffusivity model. Note that the θ posteriors across the five iterations of the KOH framework are consistent with the posterior obtained via the standard Bayesian framework.

Figure 20 presents the posterior distribution of σ_ε for all five KOH iterations when considering the LLS-informed SiC diffusivity model. Also shown in this figure is the σ posterior obtained based on the standard Bayesian framework when using the entire experimental dataset denoted as the “Base” dataset in the figure. Table II presents the most likely sigma values from the five iterations of the KOH and from the standard Bayesian approach. The sigma values from the five iterations of the KOH framework are well constrained. The mean sigma value from the five iterations of the KOH framework is $\sigma_\varepsilon = 0.155$. The mostly likely sigma value from the standard Bayesian framework is $\sigma = 0.253$. When considering

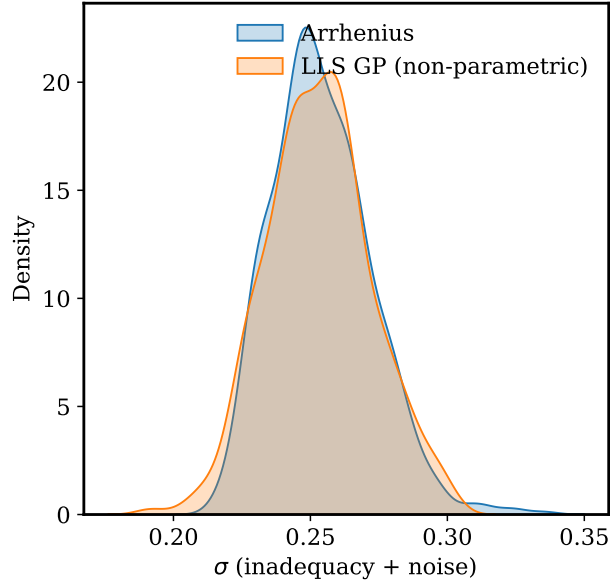


Fig. 16. Posterior distribution of the sigma term σ obtained by using the standard Bayesian framework and considering LLS effects in the SiC diffusivity through a GP model. Also shown for comparison is the posterior obtained when considering the Arrhenius equation for the SiC layer.

the LLS-informed model, these values are very consistent with those obtained via the Arrhenius equation. Thus, two important conclusions can be drawn:

- There is a high degree of similarity between the most likely σ_ε for both the Arrhenius equation and the LLS-informed model. Theoretically, the inferred experimental noise uncertainty should not be influenced by the model used, given the same experimental data. Thus, this is numerically demonstrated in Table II—specifically when comparing the KOH mean values from the Arrhenius equation against those from the LLS-informed model.
- The total sigma σ values from the standard Bayesian framework are also quite consistent when comparing the Arrhenius equation and the LLS-informed model. Given the same/similar amount of experimental noise uncertainty produced by these two modeling approaches, the similarity between the total sigma σ values indicates a similar degree of modeling inadequacy.

As emphasized in Chapter II, while such an uncertainty decomposition and comparison across modeling approaches is quite interesting, care must be taken to ensure that the model inadequacy term is properly accounted for and that sensitivity analyses are conducted to mitigate any possible over- or under-representation of this term. There is potential for confounding between the model inadequacy and experimental noise uncertainties and a misrepresentation in the uncertainty decomposition analysis. Mitigation of such confounding of uncertainties in the KOH framework is an active area of research.

E. Forward Prediction and Uncertainty Quantification on the AGR-1 Experimental Dataset

Forward UQ was applied to the AGR-1 dataset, which was not part of the experimental data, so as to perform inverse analysis using either the KOH framework or the standard Bayesian framework. For the

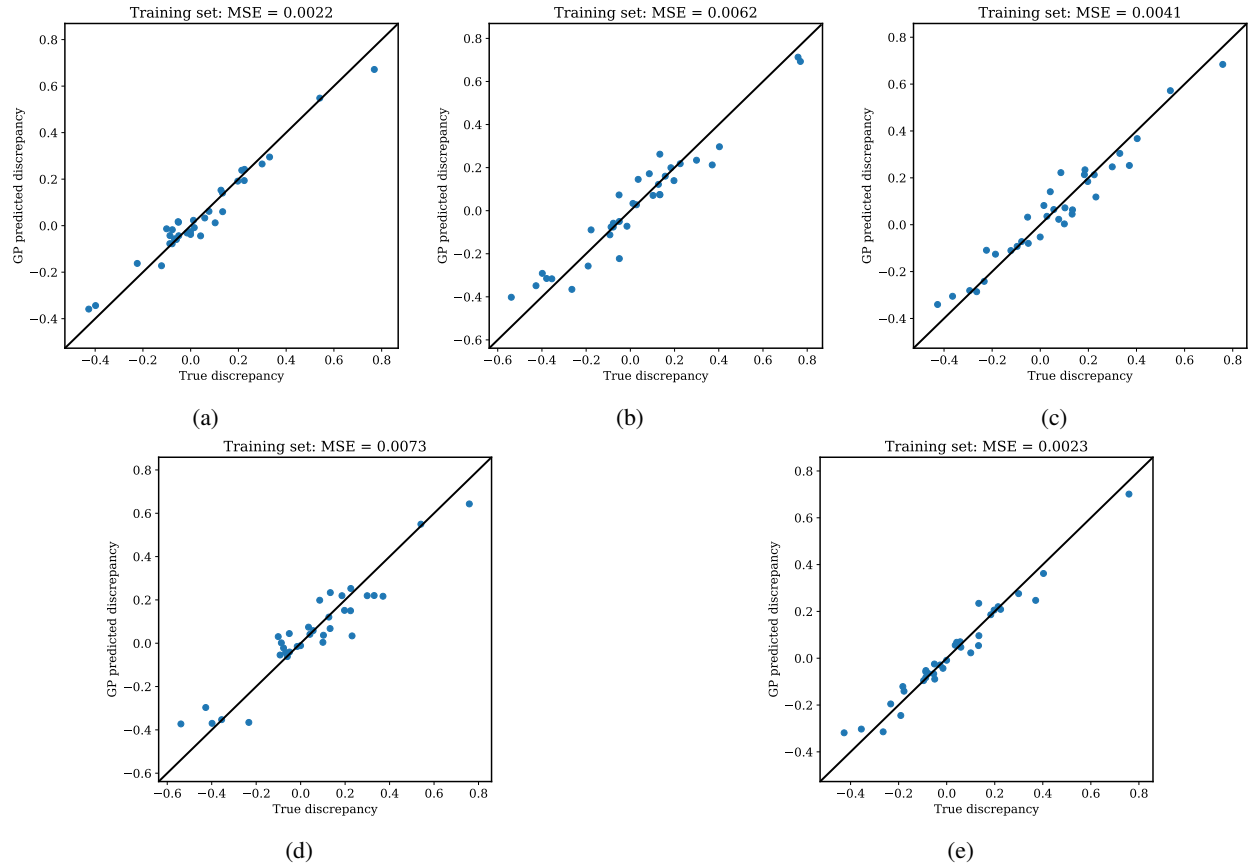


Fig. 17. Comparison of the GP-predicted and true model discrepancy/inadequacy when considering the training set and the LLS-informed SiC diffusivity model: (a) iteration 1, (b) iteration 2, (c) iteration 3, (d) iteration 4, and (e) iteration 5. The five iterations represent random selection of 34 data points for training the GP discrepancy/inadequacy term on the AGR-2/3/4 experimental datasets, which consisted of 68 data points in total. The MSE values are also shown in each plot.

UCO fuel kernel, 16 experimental values of the fractional Ag release were available (disregarding two compacts for which the corrected PIE values were unavailable). These 16 experimental values were used to evaluate both the quality and predictive accuracy of uncertainties obtained using the KOH and standard Bayesian frameworks. Microstructure information for performing the predictive UQ was obtained from Gerczak et al. (2016a). More specifically, microstructure information on the AGR-1 V3 and AGR-1 B-45T variants were independently used, having major and minor axis lengths of $\{0.71, 0.31\} \mu m$ and $\{1.33, 0.47\} \mu m$, respectively.

Figures 21a and 21b present the predictive distributions for Compacts 5-3-3 and 6-4-1 when using the KOH and standard Bayesian frameworks, respectively. Results corresponding to both the AGR-1 V3 and AGR-1 B-45T microstructure variants are presented. For the KOH framework, only the iteration 3 results are shown. The predictive distributions for the KOH and standard Bayesian frameworks are consistent across both microstructure variants. In comparing the KOH and standard Bayesian frameworks, the former

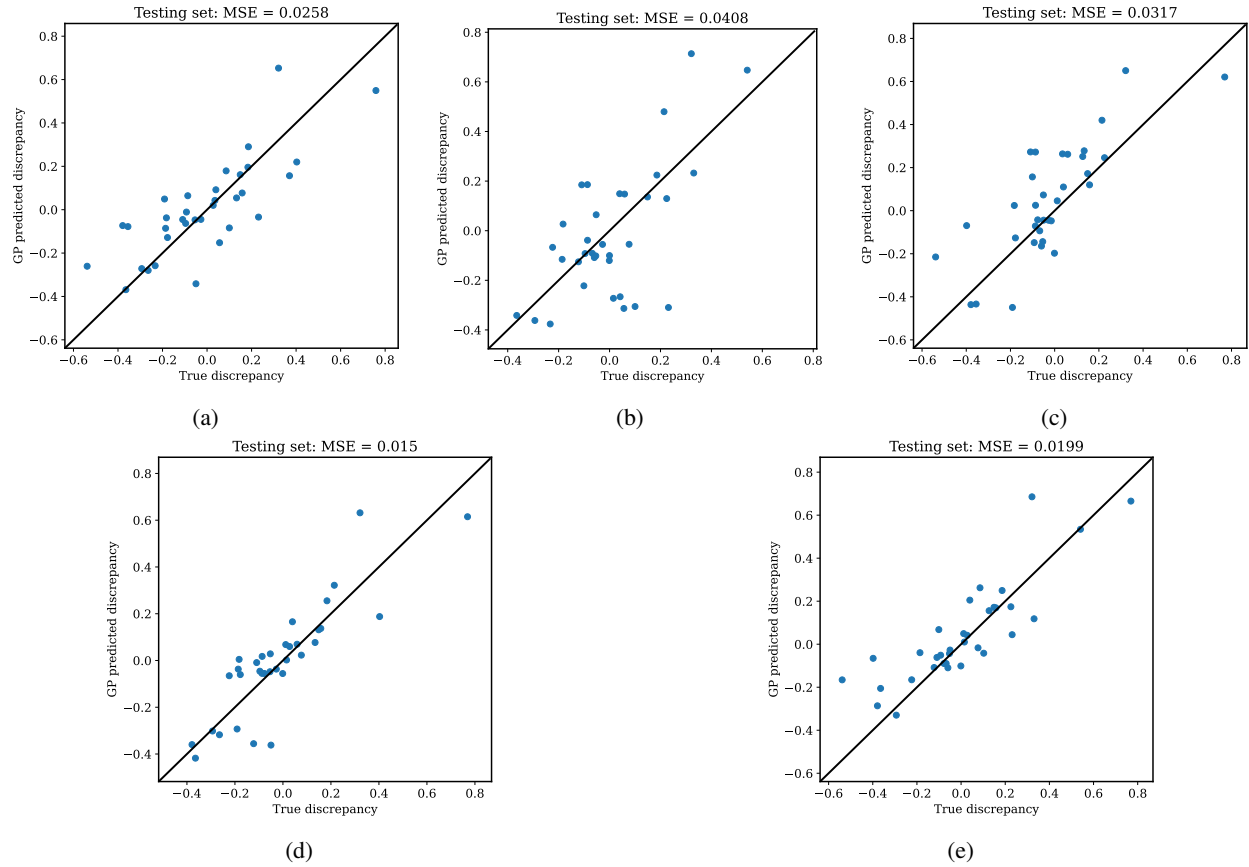


Fig. 18. Comparison of the GP-predicted and true model discrepancy/inadequacy when considering the testing set and the LLS-informed SiC diffusivity model: (a) iteration 1, (b) iteration 2, (c) iteration 3, (d) iteration 4, and (e) iteration 5. The five iterations represent random selection of 34 data points for training the GP discrepancy/inadequacy term on the AGR-2/3/4 experimental datasets, which consisted of 68 data points in total. The MSE values are also shown in each plot.

leads to better predictive distributions that encompass the experimental values. This is especially true for Compact 5-3-3 in Figure 21a. Figure 22 presents the median predictions and 5-95 percentile confidence bands for four cases: (a) KOH, iteration 3, V3 microstructure; (b) KOH, iteration 3, B-45T microstructure; (c) standard Bayesian V3 microstructure; and (d) standard Bayesian B-45T microstructure. It is generally argued that the KOH results for both these microstructures increase the accuracy and quality of uncertainty estimates, based on comparison to the experimental values. This statement is analyzed more rigorously next.

Tables VII and VIII, respectively, present the MSE and CAL values for the Arrhenius and LLS-informed GP (both V3 and B-45T microstructures) models for the five KOH iterations. Results corresponding to the standard Bayesian framework are also shown. Both microstructures lead to similar values of MSE and CAL, as averaged across the five KOH iterations. For both microstructures, the mean KOH MSE and CAL values are less than those corresponding to the standard Bayesian framework, indicating better

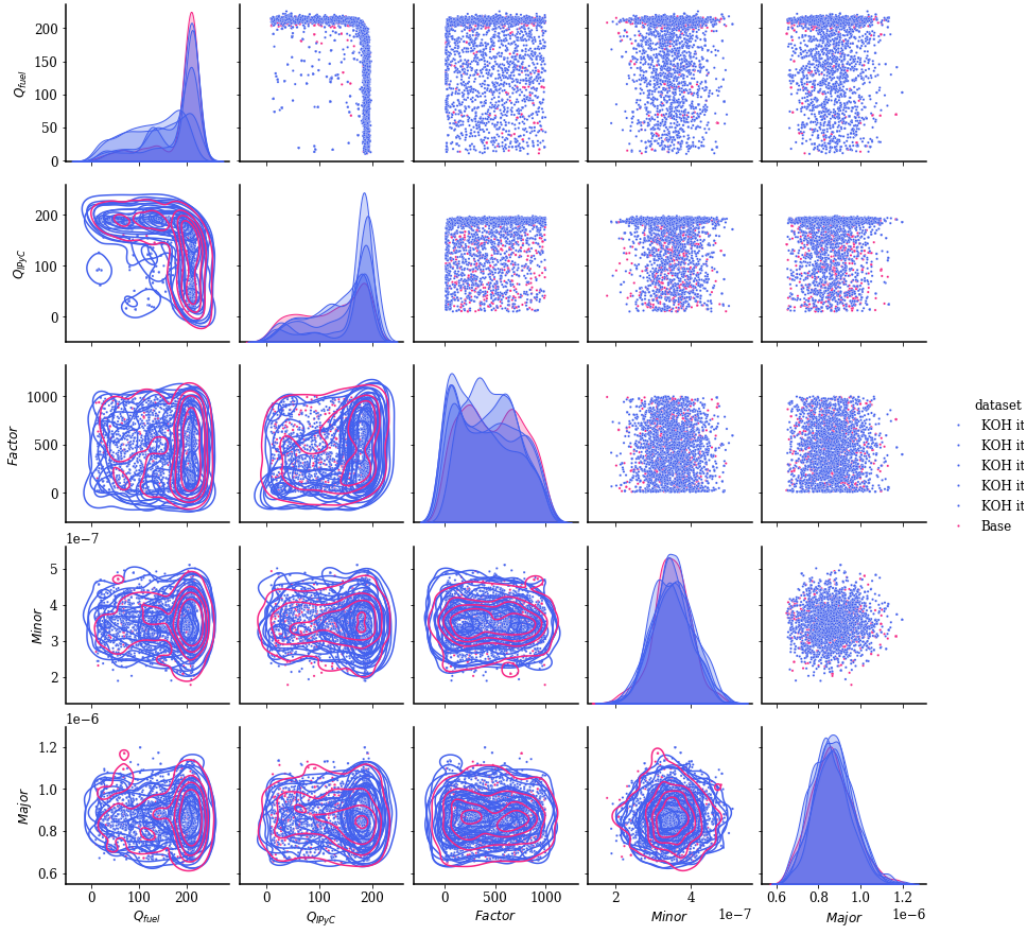


Fig. 19. Correlation matrix plot of the TRISO silver release model parameters for comparing the base (i.e., standard Bayesian) and the KOH Bayesian UQ frameworks when considering the LLS-informed SiC diffusivity model.

accuracy and uncertainty estimates provided by the KOH framework. This conclusion is consistent with what is presented in Chapter II when considering the Arrhenius equation. In comparing the Arrhenius equation and the LLS-informed GP model, both seem to reflect similar performance in terms of the MSE and CAL values for both the KOH and standard Bayesian frameworks. However, being the simpler of the two models, the Arrhenius equation has slightly better MSE and CAL values.

F. Summary and Conclusions

In this chapter, comparison was made between the inverse UQ with AGR-2/3/4 data and the forward UQ results with AGR-1 data, in light of the Arrhenius equation and the two LLS-informed models for Ag diffusivity in the SiC layer. Inverse UQ was performed using both the KOH and standard Bayesian approaches. Forward predictive UQ was compared in light of the different approaches (i.e., Arrhenius equation and LLS-informed models for the diffusivity, and standard Bayesian and KOH for the UQ),

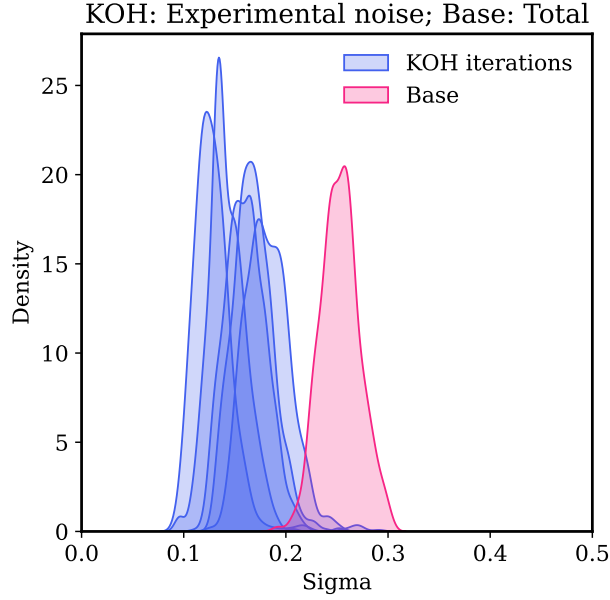


Fig. 20. Sigma values of the TRISO Ag release model that considers LLS effects in comparing the base (i.e., standard Bayesian) and KOH Bayesian UQ frameworks. Note that the sigma value (σ) from the base framework represents both the model inadequacy and experimental noise uncertainties. In contrast, the sigma value (σ_ϵ) from the KOH framework represents the experimental noise uncertainty only and is thus smaller than σ in each case.

TABLE VI. The most likely σ_ϵ values from the five iterations of the KOH framework, comparing the Arrhenius equation and the LLS-informed GP approach. The mean σ_ϵ value across all iterations of the KOH framework, along with the σ value from the base UQ framework, are also shown.

	Arrhenius	LLS-informed GP
KOH iteration 1	$\sigma_\epsilon = 0.129$	$\sigma_\epsilon = 0.167$
KOH iteration 2	$\sigma_\epsilon = 0.189$	$\sigma_\epsilon = 0.181$
KOH iteration 3	$\sigma_\epsilon = 0.162$	$\sigma_\epsilon = 0.160$
KOH iteration 4	$\sigma_\epsilon = 0.131$	$\sigma_\epsilon = 0.129$
KOH iteration 5	$\sigma_\epsilon = 0.135$	$\sigma_\epsilon = 0.142$
KOH mean	$\sigma_\epsilon = 0.149$	$\sigma_\epsilon = 0.155$
Standard Bayesian (Base)	$\sigma = 0.254$	$\sigma = 0.253$

σ_ϵ represents the experimental noise uncertainty.

σ represents both the model inadequacy and experimental noise uncertainties.

based on both the accuracy and quality of UQ, with MSE and CAL, respectively, serving as the metrics. The following observations were made:

- For fitting the LLS simulation data, both the parametric approach of finding the best functional form and a non-parametric approach using GP performed equally well. The Arrhenius equation, lacking microstructure information, had larger predictive errors when considering the LLS data.
- For the inverse UQ at the engineering scale, when considering fractional Ag release data from

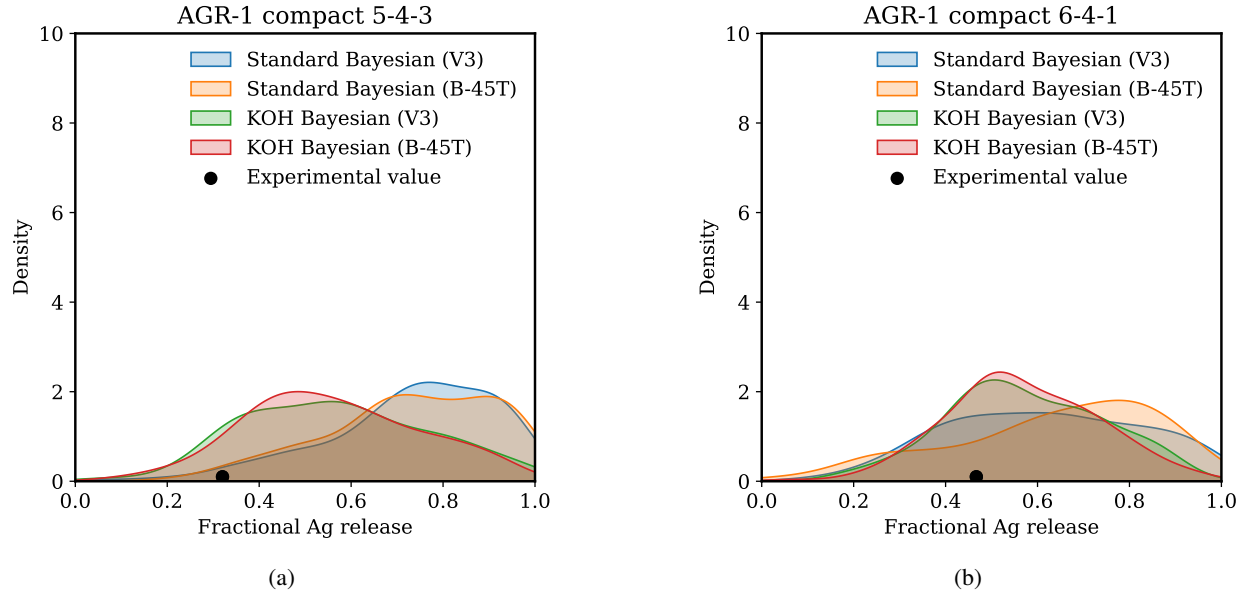


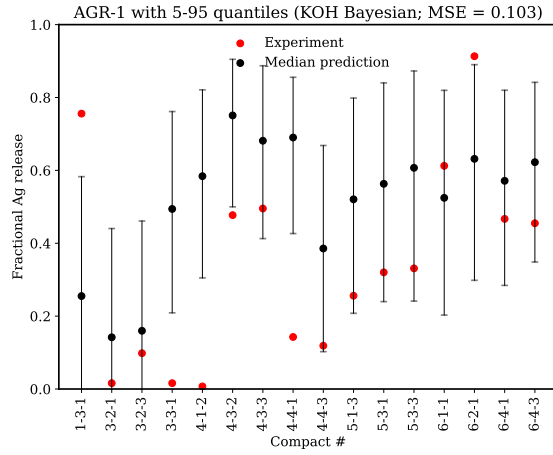
Fig. 21. Comparison of the predictive uncertainty estimates obtained from the KOH and standard Bayesian frameworks on the AGR-1 dataset, considering two compacts: (a) Compact 5-3-3, KOH iteration 3; and (b) Compact 6-4-1, KOH iteration 3. In general, the KOH framework enables better median predictions and uncertainty estimates, based on comparison with the experimental values.

TABLE VII. The MSE between the median predictions and experimental values from the AGR-1 dataset, demonstrating the accuracy of the UQ approaches. The median predictions were obtained from both the KOH framework (five iterations) and standard Bayesian framework in light of the Arrhenius and LLS-informed GP models.

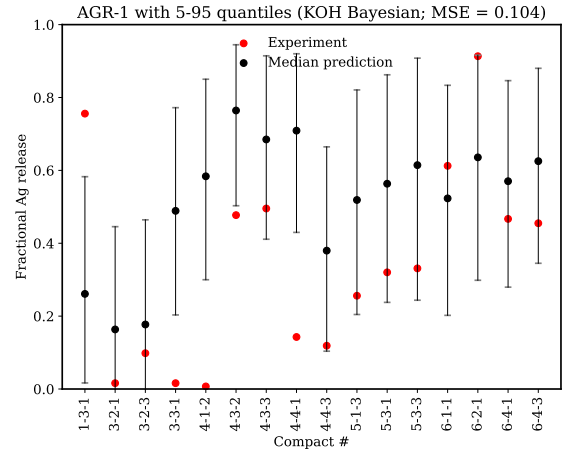
	MSE Arrhenius	MSE LLS-informed GP (V3 microstructure)	MSE LLS-informed GP (B-45T microstructure)
KOH iteration 1	0.088	0.131	0.128
KOH iteration 2	0.108	0.125	0.125
KOH iteration 3	0.087	0.103	0.104
KOH iteration 4	0.119	0.148	0.141
KOH iteration 5	0.112	0.131	0.127
KOH mean	0.102	0.127	0.125
Standard Bayesian (Base)	0.151	0.176	0.173

AGR-2/3/4 and using the standard Bayesian framework, both the Arrhenius equation and the LLS-informed model had a similar value of σ , which represents experimental noise plus model inadequacy uncertainty.

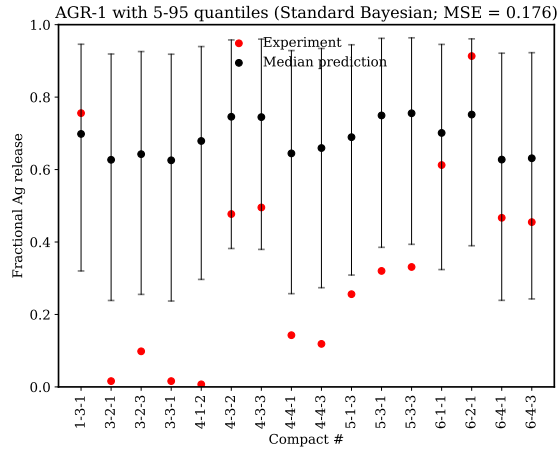
- When using the KOH Bayesian framework, both the Arrhenius equation and the LLS-informed model produced a similar level of experimental noise uncertainty. Theoretically, the inferred experimental noise uncertainty should not be influenced by the model used, given the same experimental data. Here, this theory was numerically demonstrated.



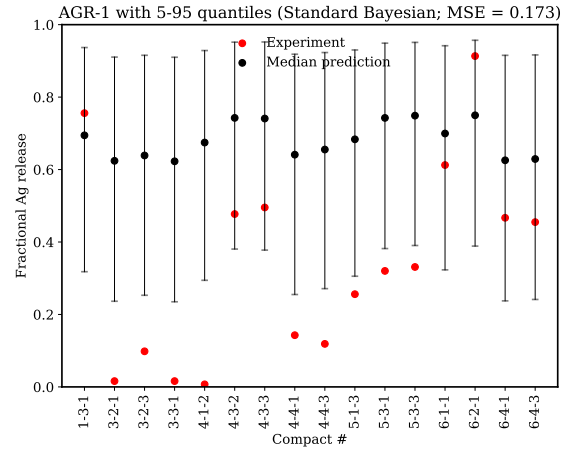
(a)



(b)



(c)



(d)

Fig. 22. Median predictions and 5-95 percentile confidence bands for the AGR-1 dataset, as obtained using the KOH and standard Bayesian frameworks and considering microstructure information: (a) KOH iteration 3, with V3 microstructure; (b) KOH iteration 3, with B-45T microstructure; (c) standard Bayesian with V3 microstructure; and (d) standard Bayesian with B-45T microstructure.

- The total sigma σ values from the standard Bayesian framework are also quite consistent when comparing the Arrhenius equation and the LLS-informed model. Given the same/similar amount of experimental noise uncertainty between these two modeling approaches, similarity between the total sigma σ values indicates a similar amount of modeling inadequacy.
- For the forward UQ at the engineering scale when considering fractional Ag release data from AGR-1, the KOH framework resulted in better predictive accuracy and UQ quality than did the standard Bayesian framework, as indicated by the MSE and CAL values, respectively. Specifically, when using the LLS-informed model, the improvements offered by the KOH framework exceeded those when using the Arrhenius equation.

TABLE VIII. The CAL values between the median predictions and experimental values from the AGR-1 dataset, demonstrating the quality of the uncertainty estimates of the UQ approaches in terms of being able to encompass the experimental values. The CAL values were computed for the KOH framework (five iterations) and the standard Bayesian framework, in light of both the Arrhenius equation and LLS-informed GP models.

	CAL Arrhenius	CAL LLS-informed GP (V3 microstructure)	CAL LLS-informed GP (B-45T microstructure)
KOH iteration 1	0.0394	0.039	0.037
KOH iteration 2	0.0396	0.053	0.05
KOH iteration 3	0.0568	0.077	0.0747
KOH iteration 4	0.0679	0.067	0.066
KOH iteration 5	0.0511	0.0639	0.0616
KOH mean	0.0509	0.05990	0.0578
Standard Bayesian (Base)	0.0886	0.103	0.102

- In comparing the Arrhenius equation and the LLS-informed model in terms of predictive forward UQ on the AGR-1 data, both resulted in somewhat similar accuracy and uncertainty quality. The Arrhenius equation, being the simpler model of the two models, fared slightly better.

IV. PARALLEL ACTIVE LEARNING FOR ACCELERATING BAYESIAN INVERSE UNCERTAINTY QUANTIFICATION IN MOOSE

This chapter discusses the parallel active learning capabilities currently being implemented in MOOSE to accelerate Bayesian inverse UQ. Bayesian inverse UQ with computational models is computationally expensive in terms of both the elapsed time and computational resources. In FY-23, we developed massively parallelizable algorithms in MOOSE to address the large amount of elapsed time necessary to perform Bayesian inverse UQ with computational models (Dhulipala et al., 2023). Still, the computational resources required to solve inverse UQ problems remains very significant. For example, Figure 23 presents the number of processors required for Bayesian inverse UQ as a function of the number of both parallel Markov chains and experimental configurations. In this figure, the factor n on the y-axis represents the number of processors required for a single evaluation of the MOOSE model. It is seen that the number of processors required increases by over several orders of magnitude with the number of experimental configurations and parallel Markov chains. (Essentially, this is a function of the problem complexity). In multiplying the number of processors to the elapsed time in order to solve the problem, the resulting computational cost can be very significant. Specifically, for the TRISO case ($n = 1$) considering the AGR-2/3/4 experimental dataset, around 3,400 processors are required, for a period of 72 hours. This makes the computational cost 244,800 processor hours. This chapter thus discusses the development of new parallel active learning capabilities to reduce the overall computational cost of solving the Bayesian inverse UQ problem, measured in terms of processor hours.

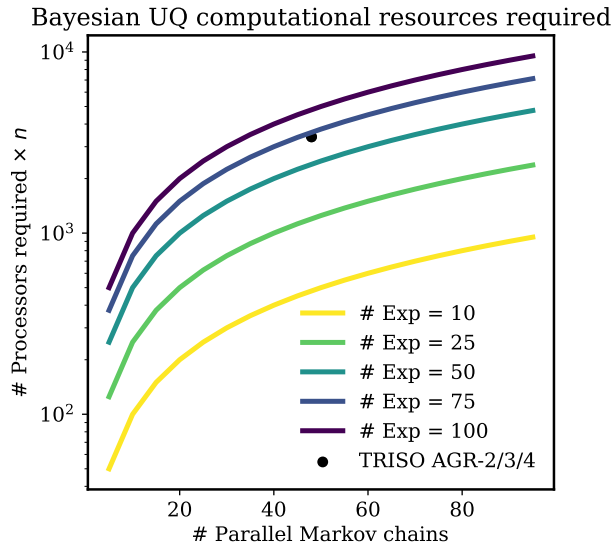


Fig. 23. Computational resources required for Bayesian inverse UQ as a function of the number of both parallel Markov chains and experimental configurations. The factor n on the y-axis is the number of processors required for a single evaluation of the MOOSE application. For TRISO Ag release simulations in BISON, $n = 1$.

A. Parallel Active Learning Approach

Figure 24 presents a flowchart describing the parallel active learning approach. In traditional active learning, the input space of the model is described in terms of probability distribution. This input space is supplied to a machine learning (ML) model (e.g., a GP model) that quantifies the uncertainties in its predictions. Then the ML model, using the uncertainty information, computes an acquisition function to determine the best input for evaluating the computational MOOSE model. The model is then evaluated and the output sent back to the ML model for re-training. With each iteration, the ML model improves its predictive ability by selecting the best input-output pair, aided by the acquisition function. However, this traditional approach to active learning is very sequential, and its computational efficiency can be hindered by evaluating the many serial evaluations of the expensive computational model. Therefore, we here adopt a parallel version of the active learning approach. Instead of selecting one best input-output pair in each iteration, we select P best input-output pairs. This is achieved by modifying the acquisition function via a local penalization so as to prevent clustering of the P best input-output pairs. The mathematical details of this approach are discussed next.

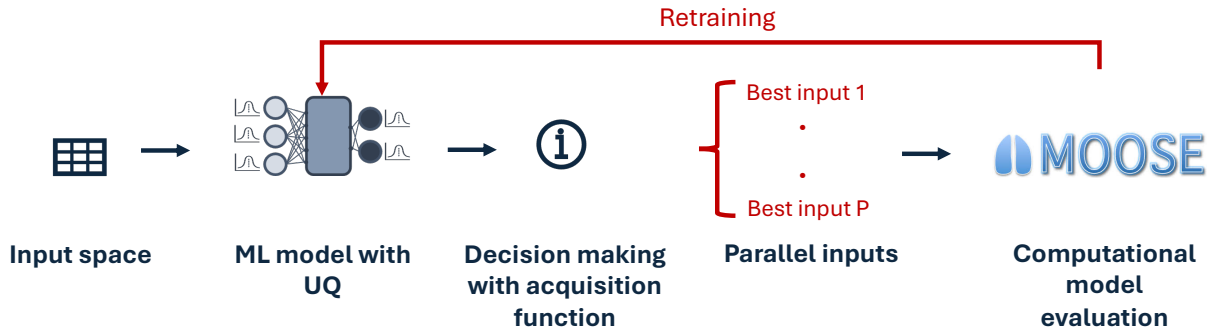


Fig. 24. Schematic of the parallel active learning approach.

1) Machine learning model with uncertainty quantification: Gaussian process

ML models that can quantify their prediction uncertainties are known as probabilistic ML models. A GP is a probabilistic ML model that works very well in active learning settings, thanks to its robust UQ abilities (El Gammal et al., 2023; Wang and Li, 2018; Dhulipala et al., 2022; Chakroborty et al., 2023; Shields et al., 2023; Peherstorfer et al., 2018; Ginsbourger et al., 2010; Chen et al., 2022). We used a GP model to approximate the log-likelihood as a function of model parameters and sigma term $\{\theta, \sigma\}$, considering the experimental data \mathcal{D} , corresponding configurations Θ , and model predictions $\mathcal{M}(\Theta)$:

$$\ln \mathcal{L}(\theta, \sigma | \Theta, \mathcal{M}, \mathcal{D}) = \sum_{i=1}^N \ln \mathcal{L}(\theta, \sigma | \Theta_i, \mathcal{M}, \mathcal{D}_i) \quad (21)$$

As such, the GP training and testing is only a function of $\{\theta, \sigma\}$ and not the experimental configurations due to the aggregation performed in the above equation. The theory behind GP was described in Chapter II, Section II-A2. To put it succinctly, given the model parameters $\{\theta, \sigma\}$, a GP approximation of the log-likelihood is described as:

$$\ln \mathcal{L}(\{\boldsymbol{\theta}, \sigma\}) \sim \mathcal{N}(m(\{\boldsymbol{\theta}, \sigma\}), k(\{\boldsymbol{\theta}, \sigma\}, \{\boldsymbol{\theta}, \sigma\}')) \quad (22)$$

where $m(\cdot)$ is the mean function and $k(\cdot, \cdot)$ is the covariance function. The GP model outputs both the mean prediction of $\ln \mathcal{L}$ and the standard deviation denoted as μ_{GP} and σ_{GP} .

2) Acquisition function

The acquisition function is a critical component of the active learning framework, as it determines the inputs under which to evaluate the expensive computational model so as to re-train the GP. As such, the acquisition function is responsible for improving the GP model with each iteration of the active learning. Several acquisition functions are available in the literature. Among the classic examples are EIGF, Upper Confidence Bound (UCB), and Probability of Improvement (PI) Zhan et al. (2017); Chen et al. (2022); Contal et al. (2013); Wang et al. (2016). For example, the EIGF acquisition function is defined as (Lam and Notz, 2008):

$$EIGF(\{\boldsymbol{\theta}, \sigma\}) = (\mu_{GP} - \ln \mathcal{L}(\{\boldsymbol{\theta}, \sigma\}_*))^2 + \sigma_{GP}^2 \quad (23)$$

where $\{\boldsymbol{\theta}, \sigma\}_*$ is the point the training set that is closest to $\{\boldsymbol{\theta}, \sigma\}$ in terms of the Euclidean distance. While EIGF is the good choice for an acquisition function, it prioritizes improving the GP model throughout the parameter space. For Bayesian UQ, however, we are interested in sampling from regions of high log-likelihood. El Gammal et al. (2023) proposed an acquisition function that aims to improve the GP model in regions of high log-likelihood or log-posterior. This function is mathematically described as:

$$a_{\mathcal{L}}(\{\boldsymbol{\theta}, \sigma\}) = \exp(2\xi \mu_{GP}) (\exp(\sigma_{GP}) - 1) \quad (24)$$

where ξ is a factor to boost the exploratory behavior, and is set to $M^{-0.85}$, with M being the number of model parameters. Figure 25 compares the EIGF and the posterior targeted acquisition functions (Equations (23) and (24), respectively) by plotting the number of re-training points vs. the corresponding log-likelihood values. It is observed that the EIGF results in re-training points with both small and large log-likelihood values. In contrast, after some initial iterations, the posterior targeted acquisition function results in small log-likelihood values that correspond to the important regions for Bayesian inverse UQ. Hence, we adopt the posterior targeted acquisition function described in Equation (24).

3) Parallelization with local penalization

Instead of selecting one best point for re-training the GP model by optimizing the acquisition function, we selected P best points by modifying the acquisition function. There are several approaches to accomplishing parallelization. For example, Ginsbourger et al. (2010) proposed a Kriging believer approach to parallelize active learning. More recently, Zhan et al. (2017) proposed the pseudo acquisition function criterion, which essentially prevents clustering of the selected P re-training points while optimizing the acquisition function. We adopted this approach for its simplicity. To illustrate it, the correlation function between two inputs is first defined as:

$$Corr(\{\boldsymbol{\theta}, \sigma\}, \{\boldsymbol{\theta}, \sigma\}') = 1 - \exp\left(\frac{-||(\{\boldsymbol{\theta}, \sigma\} - \{\boldsymbol{\theta}, \sigma\}')/\boldsymbol{\gamma}_l||}{2}\right) \quad (25)$$

where $\boldsymbol{\gamma}_l$ are the length scales for $\{\boldsymbol{\theta}, \sigma\}$, and these are a part of the GP hyperparameters obtained through optimization (i.e., $\boldsymbol{\gamma}_l \subset \boldsymbol{\gamma}$). Now, the first optimal re-training point is defined as:

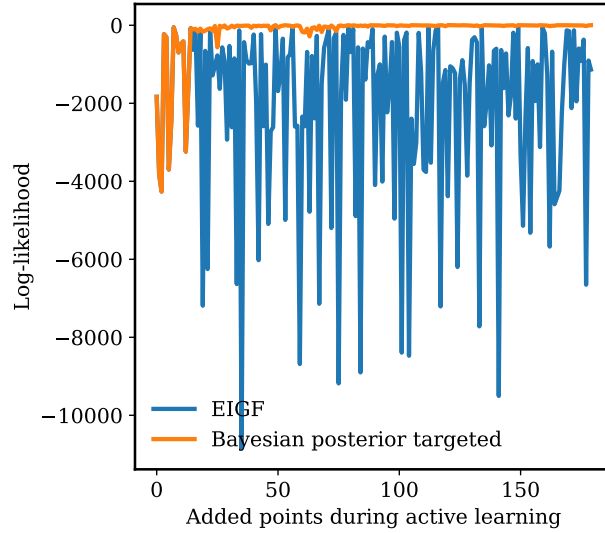


Fig. 25. Comparison of the EIGF and the posterior targeted acquisition functions (Equations (23) and (24), respectively). EIGF results in re-training points that have both small and large values of log-likelihood, whereas the posterior targeted results in points that mostly have small values of log-likelihood and are important for Bayesian UQ.

$$\{\boldsymbol{\theta}, \sigma\}^1 = \operatorname{argmax} a_{\mathcal{L}}(\{\boldsymbol{\theta}, \sigma\}) \quad (26)$$

The second optimal re-training point is defined as:

$$\{\boldsymbol{\theta}, \sigma\}^2 = \operatorname{argmax} a_{\mathcal{L}}(\{\boldsymbol{\theta}, \sigma\}) \operatorname{Corr}(\{\boldsymbol{\theta}, \sigma\}, \{\boldsymbol{\theta}, \sigma\}^1) \quad (27)$$

The p^{th} optimal re-training point is defined as:

$$\{\boldsymbol{\theta}, \sigma\}^p = \operatorname{argmax} a_{\mathcal{L}}(\{\boldsymbol{\theta}, \sigma\}) \prod_{i=1}^{p-1} \operatorname{Corr}(\{\boldsymbol{\theta}, \sigma\}, \{\boldsymbol{\theta}, \sigma\}^i) \quad (28)$$

In this manner, we can select P re-training points sequentially within each iteration of active learning by performing local penalization to mitigate any clustering of those points.

4) Re-parameterization to account for the sigma term

The log-likelihood, which we intend to emulate as a GP, is a function of both the model parameters $\boldsymbol{\theta}$ and the sigma term σ . However, a varying σ term can make the active learning process quite complex, as it influences the spread of the log-likelihood. Tests conducted on some example problems showed that considering σ as an input parameter to the GP for active learning can significantly increase the active learning iterations but also lead to errors when performing subsequent Bayesian UQ. Thus, we designed a strategy that uses re-parameterization to avoid including σ as an input parameter during active learning. This strategy operates under the assumption that the likelihood is Gaussian distributed. During the active learning process, a value of $\sigma = 1$ is used. Essentially, σ is then excluded from the input parameters for

the GP. During the evaluation stage, where the trained GP model is used for performing Bayesian UQ, the following re-parameterization steps are followed in the same sequence in order to account for the correct σ value:

$$\begin{aligned}
\ln \hat{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{\Theta}, \mathcal{M}, \mathcal{D}) &\leftarrow \ln \hat{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{\Theta}, \mathcal{M}, \mathcal{D}) - N \ln\left(\frac{1}{\sqrt{2} \pi}\right) \\
\ln \hat{\mathcal{L}}(\boldsymbol{\theta}, \sigma|\boldsymbol{\Theta}, \mathcal{M}, \mathcal{D}) &\leftarrow \ln \hat{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{\Theta}, \mathcal{M}, \mathcal{D}) \frac{1}{\sigma^2} \\
\ln \hat{\mathcal{L}}(\boldsymbol{\theta}, \sigma|\boldsymbol{\Theta}, \mathcal{M}, \mathcal{D}) &\leftarrow \ln \hat{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{\Theta}, \mathcal{M}, \mathcal{D}) + N \ln\left(\frac{1}{\sigma\sqrt{2} \pi}\right)
\end{aligned} \tag{29}$$

where $\hat{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{\Theta}, \mathcal{M}, \mathcal{D})$ is the log-likelihood, with $\sigma = 1$ predicted by the GP (i.e., μ_{GP}). Through the steps listed in Equation (29), $\hat{\mathcal{L}}(\boldsymbol{\theta}|\boldsymbol{\Theta}, \mathcal{M}, \mathcal{D})$ is transformed into $\ln \hat{\mathcal{L}}(\boldsymbol{\theta}, \sigma|\boldsymbol{\Theta}, \mathcal{M}, \mathcal{D})$, which is now a function of both $\{\boldsymbol{\theta}, \sigma\}$.

B. MOOSE Implementation Details

We will now discuss the parallel active learning capabilities being implemented in MOOSE. These capabilities build upon the parallel MCMC capabilities that were implemented in FY-23 and discussed in the associated milestone report Dhulipala et al. (2023). Readers are recommended to review these previously implemented capabilities first. Parallel active learning in MOOSE consists of two components: training and evaluation. Figures 26 and 27 present the MOOSE objects and their dependencies in regard to performing the training and evaluation components of active learning, respectively. The associated details are discussed next.

The training component of parallel active learning in MOOSE contains the following important objects:

- **BayesianActiveLearningSampler:** This sampler object is derived from a `GenericActiveLearningSampler` object. It is tailored for Bayesian inverse UQ applications in that it considers the experimental configurations. At each iteration of active learning, this object proposes a set of P parallel input samples and combines them with the experimental configurations, then finally sends these samples to the `MultiApp` object for model evaluations. This object also creates a large population of input samples at each iteration, which are retrieved by the `BayesianActiveLearner` reporter object in order to facilitate optimization of the acquisition function.
- **MultiApp:** This object facilitates evaluation of the computational model under parallel inputs sent by the `Sampler` object.
- **BayesianActiveLearner:** This is a reporter object that derives off of a `GenericActiveLearner` object. Its function is to assemble the inputs and outputs from the `BayesianActiveLearningSampler` and `MultiApp`, respectively, and compute the log-likelihood function that serves as the training data for the GP. Also, this object performs the important function of optimizing the acquisition function and selecting the next best set of inputs to the `BayesianActiveLearningSampler` object. Currently, the acquisition function is optimized by selecting the best P inputs out of the large population of samples created earlier in the iteration by the `BayesianActiveLearningSampler`.
- **Support objects:** The `BayesianActiveLearner` object relies on several support objects: `ActiveLearningGaussianProcess` to facilitate re-training of the GP model object;

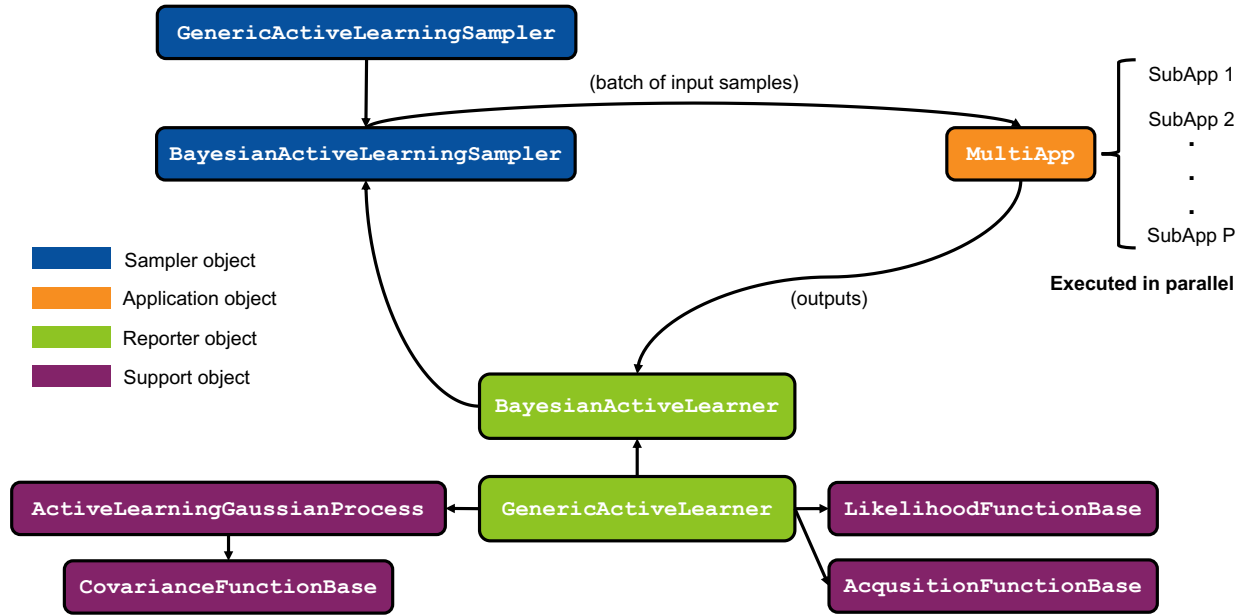


Fig. 26. MOOSE objects and their dependencies in regard to performing the training component of parallel active learning.

CovarianceFunctionBase to construct covariances for the GP model object, based on the kernel specified by the user; LikelihoodFunctionBase to evaluate the likelihood function given inputs and model outputs based on the distribution specified by the user; and AcquisitionFunctionBase to compute the acquisition function specified by the user and to perform local penalization when selecting the best P input samples.

The evaluation component of parallel active learning in MOOSE contains the following important objects:

- **AffineInvariantDifferentialEvolution:** This is the same parallel MCMC sampler implemented in FY-23. It can be directly used for active learning since the primary job of this object is to propose new samples according to the differential evolution sampler algorithm.
- **GPDifferentialEvolutionDecision:** This object derives off of ParallelMCMCDecision, and its job is to make accept/reject decisions for the samples proposed by AffineInvariantDifferentialEvolution. While this object's function is similar to that of the AffineInvariantDifferentialEvolutionDecision object, a key difference is it uses the GP predictions of log-likelihood and does not depend on the outputs from a MultiApp. Hence, this object completely avoids the need to call the computational model.
- **GaussianProcess:** This existing object in MOOSE facilitates the load of a previously trained GP model, as well as its evaluation by the GPDifferentialEvolutionDecision reporter object.

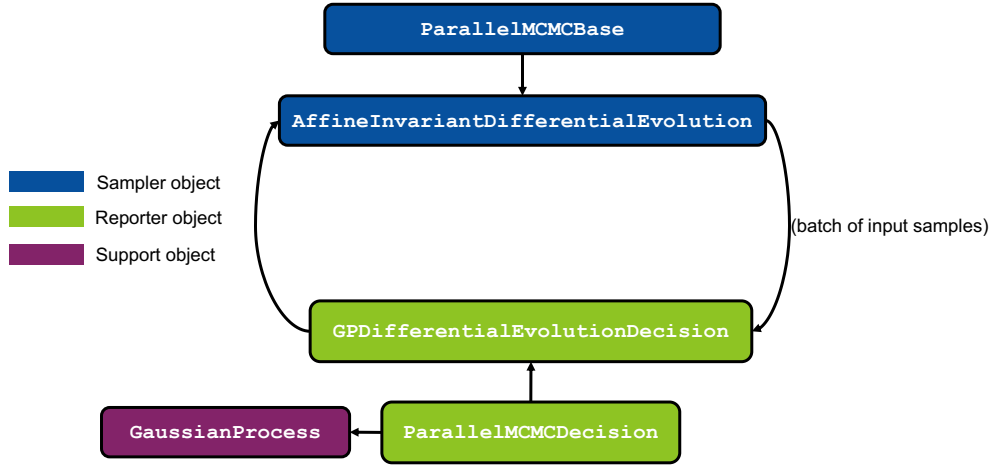


Fig. 27. MOOSE objects and their dependencies to execute the evaluation component of parallel active learning.

C. Application to a Test Case: Species Interaction Problem

We first tested the parallel active learning capabilities on a species interaction problem from FY-23. Species interaction is described by the following Lotka-Volterra equations:

$$\begin{aligned}
 \frac{du}{dt} &= -\alpha u + \beta u v \\
 \frac{dv}{dt} &= \gamma v - \delta u v \\
 \text{with, } u(0) &= 10.0 \text{ and } v(0) = 10.0
 \end{aligned} \tag{30}$$

where u and v are the variables of interest, and α , β , γ and δ are the model parameters. To simulate the “experimental” data, we considered 10 different time steps and computed the u and v values with known values of the four model parameters. To the computed u and v values, a known Gaussian random noise of 0.25 was added in the logarithmic space. Note that u and v are strictly positive; hence, we will operate in the logarithmic space to ensure their positivity. The milestone report from FY-23 lists the experimental data and the experimental configurations (Dhulipala et al., 2023). The prior distributions over the model parameters α , β , γ , and δ and the σ term are presented in Table IX.

TABLE IX. Prior distributions for the species interaction test case for parallel active learning capabilities.

Model parameter	Distribution
$\theta_1 = \alpha$	$\mathcal{TN}(0.8, 0.05, 0.0, 2.0)$
$\theta_2 = \beta$	$\mathcal{TN}(0.5, 0.05, 0.0, 2.0)$
$\theta_3 = \gamma$	$\mathcal{TN}(0.5, 0.05, 0.0, 2.0)$
$\theta_4 = \delta$	$\mathcal{TN}(0.25, 0.05, 0.0, 2.0)$
σ^2 (inadequacy + noise)	$\mathcal{U}(0.0, 0.5)$

We performed parallel active learning for 100 iterations, with 8 new training points identified at each.

As already described, we set $\sigma = 1$ during the active learning process and trained the GP model only as a function of θ . During the active learning, we monitored a convergence metric; namely, the root mean squared error between the GP predictions and the true values of the log-likelihood before re-training the GP in the current iteration. We also monitored the maximum value of the acquisition function throughout the active learning. Figures 28a and 28b present the convergence metric and maximum value of the acquisition function at each iteration of active learning. Generally, both these metrics showed a downward trend as the number of active learning iterations increased. While the convergence metric occasionally demonstrated sudden jumps in its values, after iteration 60, these jumps were essentially absent.

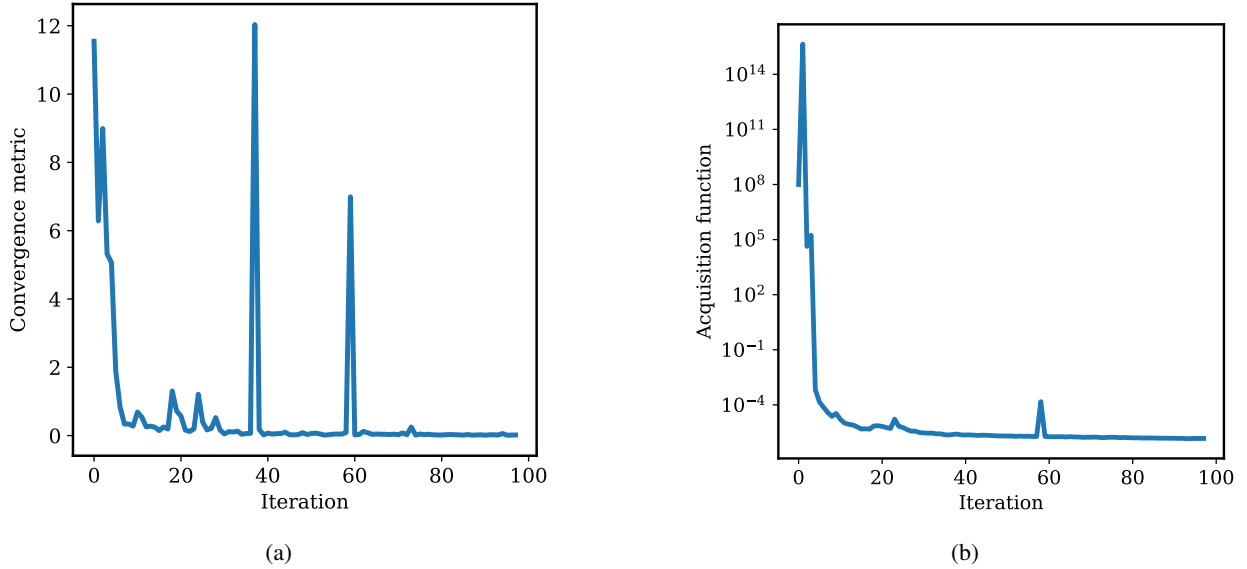


Fig. 28. Metrics for monitoring the active learning process at each iteration for the species interaction test case: (a) a convergence metric defined as the root mean squared error between the GP predictions and the true values of the log-likelihood before re-training the GP in the current iteration, and (b) the maximum value of the acquisition function.

Once the GP model was actively trained for 100 iterations, we used it in a parallel MCMC scheme—namely the Differential Evolution Sampler (DES) algorithm—to approximate the log-likelihood. Per the discussion in Section IV-A4, the approximated log-likelihood was corrected to use the appropriate σ value and predict $\ln \hat{\mathcal{L}}(\theta, \sigma | \Theta, \mathcal{M}, \mathcal{D})$ for a given $\{\theta, \sigma\}$. We performed DES with the GP-approximated log-likelihood for 500 iterations, with 48 parallel chains. We also performed the traditional DES, again using 48 parallel chains, to obtain the reference solutions of the posterior space. Figures 29 and 30 compare the parallel active learning and parallel MCMC posterior distributions of θ and σ , respectively. In general, there is a very good comparison between the parallel active learning and parallel MCMC results.

Figure 31 compares the computational cost of performing Bayesian UQ based on the parallel active learning and parallel MCMC approaches. Computational cost is defined as the number of processors used times the elapsed time necessary to solve the Bayesian UQ problem. Since the species interaction problem is a simple test case with negligible model evaluation time, we measured computational cost for this case

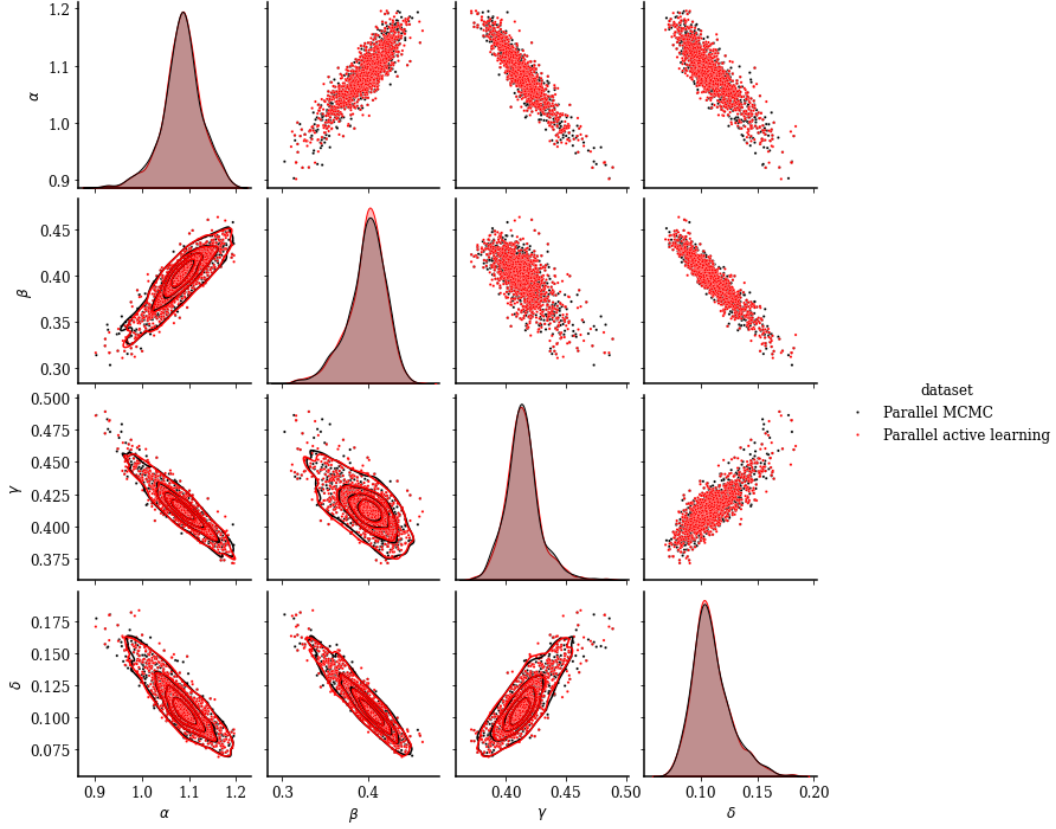


Fig. 29. Comparison of the posterior distributions of the model parameters θ when considering parallel active learning and parallel MCMC approaches for the species interaction problem.

as being the number of processors used times the serial model calls. This metric ignores the GP re-training time, which can be substantial. Nevertheless, parallel active learning is seen to substantially reduce the computational cost in comparison to active learning. However, the computational cost comparison is more genuine for the TRISO application next discussed next.

D. Application to TRISO Fuel, Using the AGR-2 Experimental Dataset

We then applied the parallel active learning capabilities to the TRISO Ag release problem, considering the AGR-2 experimental data, which consisted of 36 data points for the UCO fuel kernel. The parameters of interest were the activation energies for three layers and the diffusion coefficient for the SiC layer; that is, $\theta = \{Q_{kernel}, Q_{ipy}, D_{sic}, Q_{sic}\}$. The sigma term posterior was also of interest. The prior distributions are presented in Table X. The likelihood function for this case has historically been a truncated normal to obey the constraints on fractional Ag release, which is bounded between 0 and 1. However, using a truncated normal violates the Gaussianity assumption for the re-parameterization approach (discussed in Section IV-A4) to approximate the posterior of the σ term. Thus, we transformed the experimental

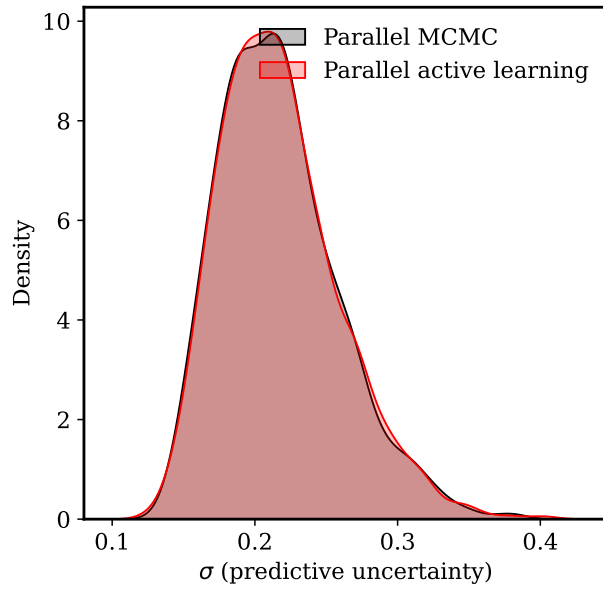


Fig. 30. Comparison of the posterior distributions of the sigma term σ when considering parallel active learning and parallel MCMC approaches for the species interaction problem.

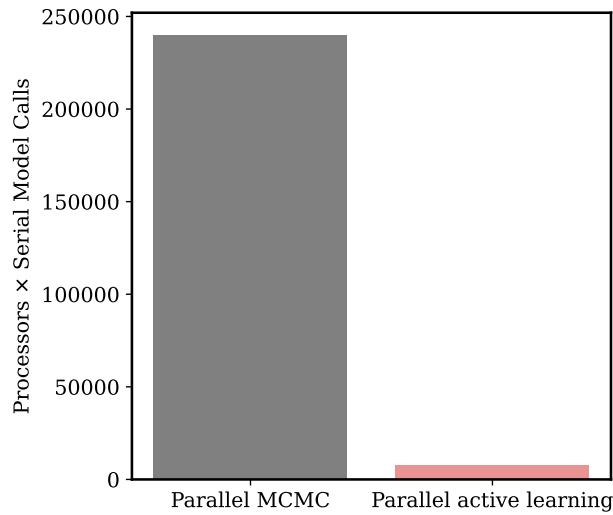


Fig. 31. Comparison of the computational cost of performing Bayesian UQ when considering the parallel active learning and parallel MCMC approaches for the species interaction test case.

fractional Ag release values to an inverse logistic space that had support $(-\infty, \infty)$, and used a normal likelihood function.

We performed parallel active learning for 80 iterations, with 10 new training points identified at each.

TABLE X. Prior distributions for the TRISO Ag release application with parallel active learning capabilities.

Model parameter	Distribution
$\theta_1 = Q_{kernel}$	$\mathcal{U}(10e3, 90e3)$
$\theta_2 = Q_{ipyg}$	$\mathcal{U}(90e3, 180e3)$
$\theta_3 = D_{sic}$	$\mathcal{U}(0.1e-8, 7e-8)$
$\theta_4 = Q_{sic}$	$\mathcal{U}(180e3, 300e3)$
σ^2 (inadequacy + noise)	$\mathcal{U}(0.0, 2.0)$

As already described, we set $\sigma = 1$ during the active learning process and trained the GP model only as a function of θ . During the active learning, we monitored the convergence metric and the maximum value of the acquisition function. Figures 32a and 32b present the convergence metric and maximum value of the acquisition function at each iteration of active learning. In general, both these metrics revealed a downward trend as the number of active learning iterations increased and eventually became almost constant.

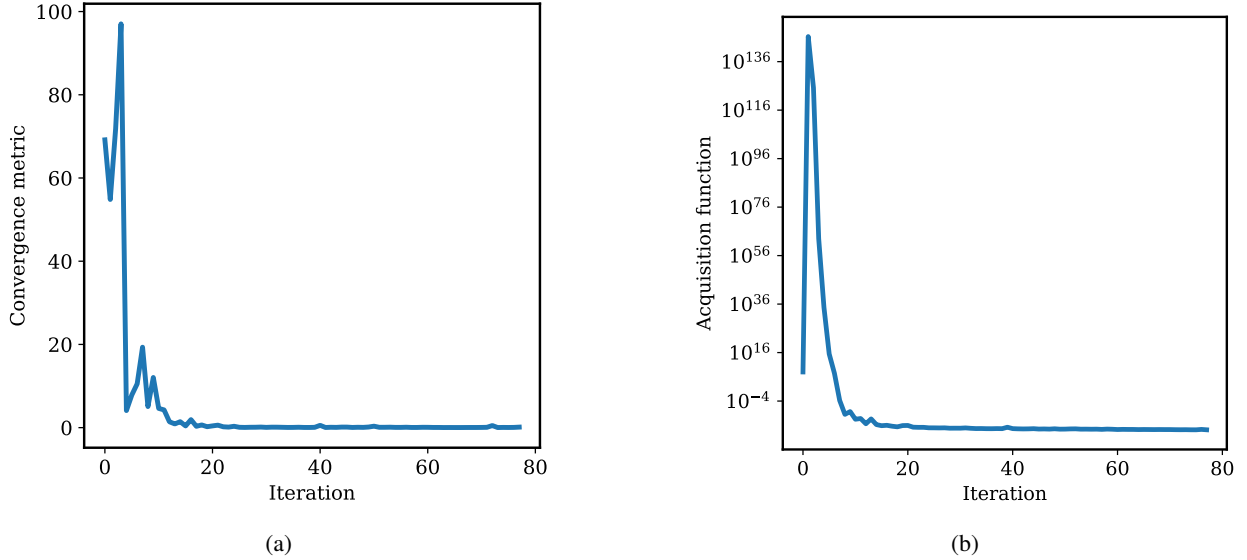


Fig. 32. Metrics for monitoring the active learning process at each iteration for the TRISO Ag release application: (a) a convergence metric defined as the root mean squared error between the GP predictions and the true values of the log-likelihood before re-training the GP in the current iteration, and (b) the maximum value of the acquisition function.

Once the GP model was actively trained for 80 iterations, we used it in a parallel MCMC scheme—namely the DES algorithm—to approximate the log-likelihood. Based on the discussion in Section IV-A4, the approximated log-likelihood was corrected to use the appropriate σ value and predict $\ln \hat{\mathcal{L}}(\theta, \sigma | \Theta, \mathcal{M}, \mathcal{D})$ for a given $\{\theta, \sigma\}$. We performed DES with the GP-approximated log-likelihood for 500 iterations, with 50 parallel chains. We also performed the traditional DES, again using 50 parallel chains, to obtain the reference solutions of the posterior space. Figures 33 and 34 compare the parallel active learning and parallel MCMC posterior distributions of θ and σ , respectively. In general, there is a very good

comparison between the parallel active learning and parallel MCMC results. In examining the correlation between D_{sic} and Q_{sic} in Figure 33 more closely, there seems to be a strong exponential relationship, as demonstrated by the parallel MCMC approach. The parallel active learning approach is able to capture this exponential relationship for the most part. Near the sharp tail region where the probability density is low, however, parallel active learning is unable to explore as well as parallel MCMC. Increasing the number of active learning iterations could potentially fix this issue, if it is deemed important.

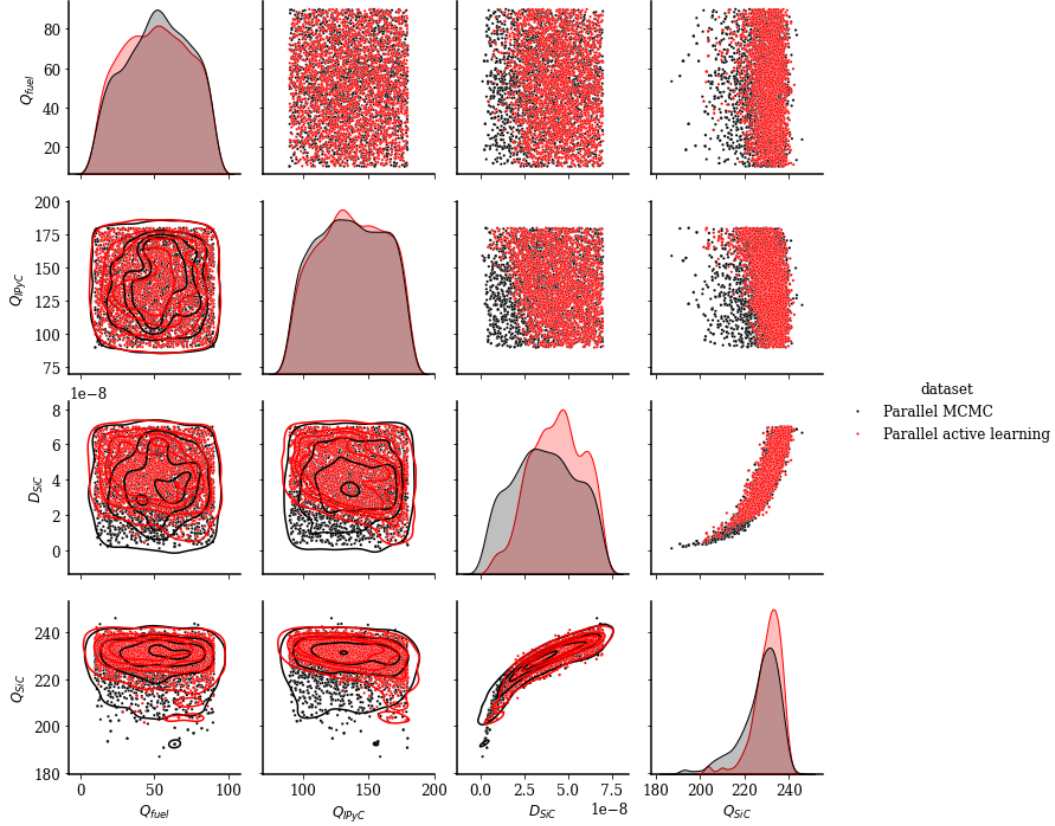


Fig. 33. Comparison of the posterior distributions of the model parameters θ when considering parallel active learning and parallel MCMC approaches for the TRISO Ag release application.

Figure 35 compares the computational cost of performing Bayesian UQ in regard to the parallel active learning and parallel MCMC approaches. Computational cost is defined as the number of processors used times the elapsed time necessary to solve the Bayesian UQ problem. Parallel active learning is seen to substantially reduce the computational cost in comparison to active learning, by about a factor of 20.

E. Summary and Conclusions

This chapter discussed the implementation of parallel active learning capabilities in MOOSE so as to reduce the computational cost of performing Bayesian inverse UQ, as measured in terms of processor hours. We discussed the theory behind parallel active learning, especially in terms of GP,

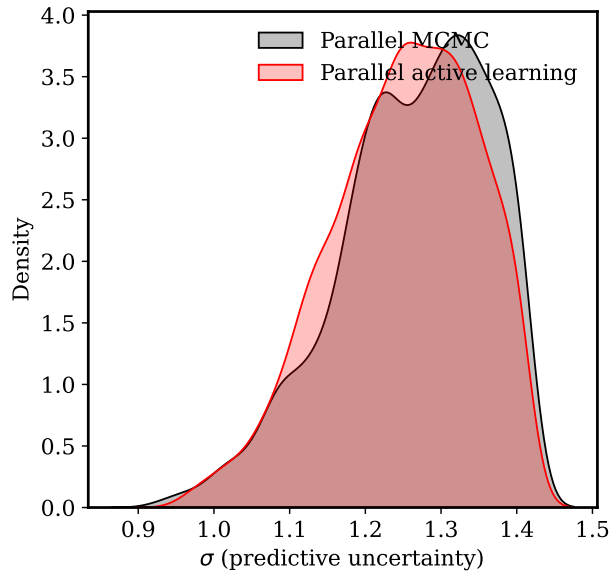


Fig. 34. Comparison of the posterior distributions of the sigma term σ when considering parallel active learning and parallel MCMC approaches for the TRISO Ag release application.

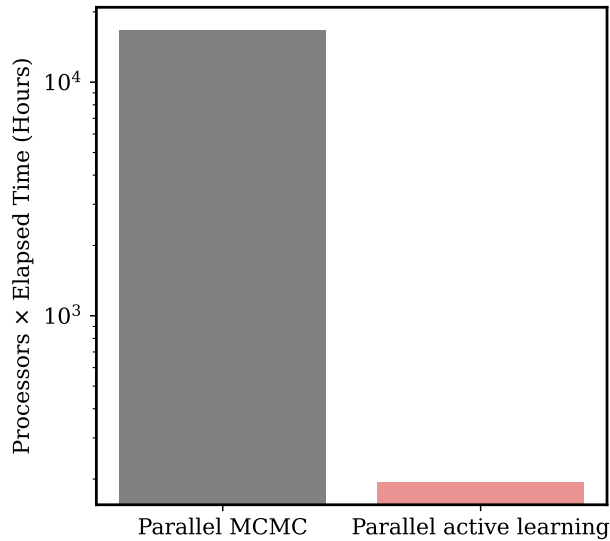


Fig. 35. Comparison of the computational cost of performing Bayesian UQ in regard to the parallel active learning and parallel MCMC approaches for the species interaction test case.

acquisition functions, and local penalization for facilitating it. We also discussed details on the MOOSE implementation and the different objects (and interactions between them) utilized to perform parallel active learning. We then applied parallel active learning to two cases, leading to the following conclusions:

- For the species interaction test case, parallel active learning gave highly satisfactory posteriors for the model parameters and the sigma term—in comparison to parallel MCMC (i.e., the reference solution)—despite the model parameters having strong correlations and anti-correlations between them. In terms of computational cost, parallel active learning was more efficient by a factor of at least 10.
- For the TRISO Ag release application when considering the AGR-2 data, parallel active learning again gave satisfactory posteriors for the model parameters and the sigma term, as compared to parallel MCMC. In terms of computational cost, parallel active learning was more efficient by a factor of at least 20.

V. FUTURE WORK

Future work for the UQ of computational models may focus on one or more of the following topics:

- ***High-dimensional problems:*** Some computational models can contain tens of model parameters to calibrate. Moreover, in some cases, it may be of interest to infer from experimental data model properties that are field variables. This is essentially a high-dimensional inference problem. For these cases, a combination of sensitivity analyses, dimensionality reduction, and superior Bayesian inference algorithms can be investigated. To do so, the underlying assumptions made by a particular approach must be understood and the results carefully evaluated. For example, sensitivity analyses based on variance-based approaches only capture linear correlations between parameters. However, lack of a linear correlation does not imply lack of a nonlinear one. Concurrently, dimensionality reduction starting from linear methods and then moving toward nonlinear methods such as kernel principal component analysis can be investigated as well.
- ***Identifying missing model physics:*** One interesting pursuit is to identify what physics missing from the model could be included to drive down the predictive uncertainties. In this FY, the KOH framework was explored in terms of capturing model inadequacy by using a GP. Future work can focus on including physics in the inadequacy modeling approach by considering methods such as formulating inadequacy from a dictionary of physics terms in combination with a data-driven term.
- ***Considering time-dependent quantities of interest:*** Modeling the uncertainties in time-dependent quantities in nuclear fuels and materials represents a very interesting expansion of the current work. Time-dependent quantities present interesting features to be captured in the UQ framework, such as non-stationarity and heteroscedasticity. Developing computational tools to capture these features and demonstrating on nuclear fuels and materials is an avenue of future work.
- ***Expanding to other fuels and materials applications:*** Expanding the UQ capabilities to other fuel and material applications, while including phenomena such as LLS effects, time-dependent quantities, and high-dimensional phenomena (if required), is another very interesting avenue for future work.

REFERENCES

- P. D. Arendt, D. W. Apley, and W. Che. Quantification of Model Uncertainty: Calibration, Model Discrepancy, and Identifiability. *ASME Journal of Mechanical Design*, 134(10):100908, 2012. doi:10.1115/1.4007390.
- M. J. Bayarri, J. O. Berger, and F. Liu. Modularization in Bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4(1):119–150, 2009. doi:10.1214/09-BA404.
- J. Brynjarsdottir and A. O’Hagan. Learning about physical parameters: The importance of model discrepancy. *Inverse problems*, 30(11):114007, 2014. doi:10.1088/0266-5611/30/11/114007.
- P. Chakroborty, S. L. N. Dhulipala, Y. Che, W. Jiang, B. W. Spencer, J. D. Hales, and M. D. Shields. General multifidelity surrogate models: Framework and active-learning strategies for efficient rare event simulation. *Journal of Engineering Mechanics*, 149(12):04023096, 2023. doi:10.1061/JENMDT.EMENG-7111.
- C. Chen, J. Liu, and P. Xu. Comparison of parallel infill sampling criteria based on kriging surrogate model. *Scientific Reports*, 12(1):678, 2022. doi:10.1137/16M1082469.
- B. P. Collin. AGR-2 irradiation test final as-run report. Technical Report INL/EXT-14-32277 (Rev.4), Idaho National Laboratory, Idaho Falls, ID United States, 2018.
- Blaise P. Collin. AGR-1 irradiation test final as-run report. Technical report, Idaho National Laboratory (INL) ID. INL/EXT-10-18097, 1 2015. URL https://art.inl.gov/trisofuels/TRISO%20Fuels%20Documents/AGR-1/INL_EXT-10-18097_AGR-1_Irradiation_Test_Final_As-Run_Report.pdf.
- Blaise P Collin. Diffusivities of Ag, Cs, Sr, and Kr in TRISO Fuel Particles and Graphite. Technical report, Idaho National Laboratory (INL) ID. INL/EXT-16-39548, 9 2016. URL <http://www.inl.gov>.
- Blaise P. Collin, David A. Petti, Paul A. Demkowicz, and John T. Maki. Comparison of silver, cesium, and strontium release predictions using PARFUME with results from the AGR-1 irradiation experiment. *Journal of Nuclear Materials*, 466:426–442, 11 2015. ISSN 0022-3115. doi:10.1016/J.JNUCMAT.2015.08.033.
- E. Contal, D. Buffoni, A. Robicquet, and N. Vayatis. *Parallel Gaussian process optimization with upper confidence bound and pure exploration*. Springer Berlin Heidelberg, 2013.
- S. L. N. Dhulipala, M. D. Shields, B. W. Spencer, C. Bolisetti, A. E. Slaughter, V. M. Laboure, and P. Chakroborty. Active learning with multifidelity modeling for efficient rare event simulation. *Journal of Computational Physics*, 468:111506, 2022. doi:10.1016/j.jcp.2022.111506.
- S. L. N. Dhulipala, D. Schwen, Y. Che, R. T. Sweet, A. Toptan, Z. M. Prince, P. German, and S. R. Novascone. Massively parallel bayesian model calibration and uncertainty quantification with applications to nuclear fuels and materials. Technical Report INL/RPT-23-73383-Rev000, Idaho National Laboratory, Idaho Falls, ID United States, 2023.
- S. L. N. Dhulipala, A. Toptan, Y. Che, D. Schwen, R. T. Sweet, J. D. Hales, and S. R. Novascone. Bayesian uncertainty quantification of tristructural isotropic particle fuel silver release: Decomposing model inadequacy plus experimental noise and parametric uncertainties. *Journal of Nuclear Materials*, 588:154790, 2024. doi:10.1016/j.jnucmat.2023.154790.
- J. El Gammal, N. Schoneberg, J. Torrado, and C. Fidler. Fast and robust bayesian inference using gaussian processes with gprry. *Journal of Cosmology and Astroparticle Physics*, 2023(10):021, 2023. doi:10.1088/1475-7516/2023/10/021.
- T. J. Gerczak, J. D. Hunn, R. A. Lowden, and T. R. Allen. Sic layer microstructure in agr-1 and agr-2

- triso fuel particles and the influence of its variation on the effective diffusion of key fission products. *Journal of Nuclear Materials*, 480:257–270, 2016a. doi:10.1016/j.jnucmat.2016.08.011.
- Tyler Gerczak and John D. Hunn. AGR program methodology for examining TRISO fuel performance and particle failure. In *15th International Conference on CANDU Fuel*. Oak Ridge National Laboratory (ORNL), 15th International Conference on CANDU Fuel, Ajax, Canada, 8 2022.
- Tyler J. Gerczak, John D. Hunn, Richard A. Lowden, and Todd R. Allen. SiC layer microstructure in AGR-1 and AGR-2 TRISO fuel particles and the influence of its variation on the effective diffusion of key fission products. *Journal of Nuclear Materials*, 480:257–270, 11 2016b. ISSN 0022-3115. doi:10.1016/J.JNUCMAT.2016.08.011.
- D. Ginsbourger, R. Le Riche, and L. Carraro. *Kriging is well-suited to parallelize optimization*. Springer Berlin Heidelberg, 2010.
- J. D. Hales, W. Jiang, A. Toptan, and K. A. Gamble. Modeling fission product diffusion in triso fuel particles with bison. *Journal of Nuclear Materials*, 548:152840, 2021. doi:10.1016/j.jnucmat.2021.152840.
- J.D. Hunn, R.N. Morris, I. van Rooyen, T. Gerczak, J.M. Harp, S.A. Ploger, and P.A. Demkowicz. AGR-1 post irradiation examination final report. Technical report, Idaho National Laboratory (INL) ID. INL/EXT-15-36407, 2015.
- C. Jiang, Z. Hu, Y. Liu, Z. P. Mourelatos, D. Gorsich, and P. Jayakumar. A sequential calibration and validation framework for model uncertainty quantification and reduction. *Computer Methods in Applied Mechanics and Engineering*, 368:113172, 2020. doi:10.1016/j.cma.2020.113172.
- C. Jiang, J.-H. Ke, P.-C. A. Simon, W. Jiang, and L. K. Aagesen. Atomistic and mesoscale simulations to determine effective diffusion coefficient of fission products in SiC. Technical report, Idaho National Laboratory (INL) ID. INL/EXT-21-64633-Rev000, 10 2021. URL <https://www.osti.gov/servlets/purl/1825508>.
- M. C. Kennedy and A. O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001. doi:10.1111/1467-9868.00294.
- V. Kuleshov, N. Fenner, and S. Ermon. Accurate uncertainties for deep learning using calibrated regression. *International conference on machine learning*, pages 2796–2804, 2018.
- C. Q. Lam and W. I. Notz. Sequential adaptive designs in computer experiments for response surface model fit. *Statistics and Applications*, 6(1):207–233, 2008.
- N. Leoni, O. Le Maitre, M. G. Rodio, and P. M. Congedo. Bayesian calibration with adaptive model discrepancy. *International Journal for Uncertainty Quantification*, 14(1):19–41, 2024. doi:10.1615/Int.J.UncertaintyQuantification.2023046331.
- Y. Liu, N. T. Dinh, R. C. Smith, and X. Sun. Uncertainty quantification of two-phase flow and boiling heat transfer simulations through a data-driven modular Bayesian approach. *International Journal of Heat and Mass Transfer*, 138:1096–1116, 2019. doi:10.1016/j.ijheatmasstransfer.2019.04.075.
- R. C. MacCallum and A. O’Hagan. Advances in modeling model discrepancy: Comment on Wu and Browne. *Psychometrika*, 80:601–607, 2015. doi:10.1007/s11336-015-9452-2.
- K. A. Maupin and L. P. Swiler. Model discrepancy calibration across experimental settings. *Reliability Engineering & System Safety*, 200:106818, 2020. doi:10.1016/j.res.2020.106818.
- B. Peherstorfer, K. Willcox, and M. Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review*, 60(3):550–591, 2018. doi:10.1137/16M1082469.
- C. E. Rasmussen. *Gaussian Processes in Machine Learning*. Springer Berlin Heidelberg, 2004.
- M. D. Shields, K. Gurley, R. Catarelli, M. Chauhan, M. Ojeda-Tuz, and F. J. Masters. Active learning

- applied to automated physical systems increases the rate of discovery. *Scientific Reports*, 13(1):8402, 2023. doi:10.1038/s41598-023-35257-7.
- P. C. Simon, L. K. Agesen, C. Jiang, W. Jiang, and J. H. Ke. Mechanistic calculation of the effective silver diffusion coefficient in polycrystalline silicon carbide: Application to silver release in AGR-1 TRISO particles. *Journal of Nuclear Materials*, 563:153669, 2022. doi:10.1016/j.jnucmat.2022.153669.
- P. C. Simon, J. H. Ke, C. Jiang, L. K. Agesen, , and W. Jiang. Multiscale, mechanistic modeling of irradiation-enhanced silver diffusion in TRISO particles. *Journal of Nuclear Materials*, Submitted, XX: XX, 2024.
- Pierre-Clément Simon, Larry Agesen, Chao Jiang, Wen Jiang, and Jia-Hong Ke. Jupyter script for automatic polycrystalline generation with desired grain size and elongation. *Mendeley Data*, V1, doi: 10.17632/vh78xyy85f.1, 2021. doi:10.17632/VH78XY85F.1.
- W. F. Skerjanc and W. Jiang. Comparison of fission product release predictions using perfume and bison with results from the agr-3/4 irradiation experiment. Technical Report INL/RPT-22-69003, Idaho National Laboratory, Idaho Falls, ID United States, 2022.
- John D. Stempien, John D. hunn, Robert N. Morris, Tyler J. Gerczak, and Paul A Demkowicz. AGR-2 TRISO fuel post-irradiation examination final report. Technical report, Idaho National Laboratory, INL/EXT-21-64279, 9 2021. URL <https://www.osti.gov/biblio/1822447>.
- C. L. Sung and R. Tuo. A review on computer model calibration. *Wiley Interdisciplinary Reviews: Computational Statistics*, 16(1):e1645, 2024. doi:10.1002/wics.1645.
- A. Toptan, W. Jiang, G. Singh, S. L. N. Dhulipala, Y. Che, J. D. Hales, and S. R. Novascone. Assessment and improvement of fission product transport predictions of particle fuel in bison. Technical Report INL/RPT-22-69706-Rev000, INL, Idaho Falls, ID United States, 2022.
- K. Tran, W. Neiswanger, J. Yoon, Q. Zhang, E. Xing, and Z. W. Ulissi. Methods for comparing uncertainty quantifications for material property predictions. *Machine Learning: Science and Technology*, 1(2): 025006, 2020. doi:10.1088/2632-2153/ab7e1a.
- H. Wang and J. Li. Adaptive gaussian process approximation for bayesian inference with expensive likelihood functions. *Neural computation*, 30(11):3072–3094, 2018. doi:10.1162/neco_a_01127.
- J. Wang, S. C. Clark, E. Liu, and P. I. Frazier. Parallel bayesian global optimization of expensive functions. *Operations Research*, 68(6):1850–1865, 2016. doi:10.1287/opre.2019.1966.
- A. White and S. Mahadevan. Discrepancy modeling for model calibration with multivariate output. *International Journal for Uncertainty Quantification*, 13(6):1–23, 2023. doi:10.1615/Int.J.UncertaintyQuantification.2023044543.
- J. L. Wu, M. E. Levine, T. Schneider, and A. Stuart. Learning about structural errors in models of complex dynamical systems. *Journal of Computational Physics*, 513:113157, 2024. doi:10.1016/j.jcp.2024.113157.
- X. Wu, T. Kozlowski, H. Meidani, and K. Shirvan. Inverse uncertainty quantification using the modular Bayesian approach based on Gaussian process, Part 1: Theory. *Nuclear Engineering and Design*, 335: 339–355, 2018. doi:10.1016/j.nucengdes.2018.06.004.
- D. Zhan, J. Qian, and Y. Cheng. Pseudo expected improvement criterion for parallel ego algorithm. *Journal of Global Optimization*, 68:641–662, 2017. doi:10.1007/s10898-016-0484-7.