LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-778000-DRAFT

Estimating Posterior Quantity of Interest Expectations in a Multilevel Scalable Framework

H. R. Fairbanks, S. Osborn, P. Vassilevski

June 14, 2019

Numerical Linear Algebra with Applications

Disclaimer

ESTIMATING POSTERIOR QUANTITY OF INTEREST EXPECTATIONS IN A MULTILEVEL SCALABLE FRAMEWORK

HILLARY R. FAIRBANKS[1] AND SARAH OSBORN[1] AND PANAYOT VASSILEVSKI[1,2]

[1] *Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, CA, USA*
[2] *Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, OR, USA*

Abstract. Scalable approaches for uncertainty quantification are necessary for characterizing prediction confidence in large-scale subsurface flow simulations with uncertain permeability. To this end we explore a multilevel Monte Carlo approach for estimating posterior moments of a particular quantity of interest, where we employ an element-agglomerated algebraic multigrid (AMG) technique to generate the hierarchy of coarse spaces with guaranteed approximation properties for both the generation of spatially correlated random fields and the forward simulation of Darcy's law to model subsurface flow. In both these components (sampling and forward solves), we exploit solvers that rely on state-of-the-art scalable AMG. To showcase the applicability of this approach, numerical tests are performed on two 3D examples – a unit cube and an egg-shaped domain with an irregular boundary – where the scalability of each simulation as well as the scalability of the overall algorithm are demonstrated.

Keywords. scalable uncertainty quantification, Bayesian inference, multilevel methods, multilevel Monte Carlo, algebraic multigrid

1. Introduction

A major challenge in large-scale (and extreme-scale) predictive modeling is the scalability of numerical algorithms. While the last several decades have leant themselves to the development of scalable solvers, such as multigrid, for solving systems of partial differential equations (PDEs), there has been a more recent push for forming highly parallelizable and scalable methods for performing uncertainty quantification (UQ). In particular, with the growing ability to incorporate high-dimensional uncertainties of model parameters, e.g., spatially varying coefficients, in large-scale simulations, there is a need to formulate algorithms that are scalable with increasing dimension of the uncertain parameter. While the probability distribution of the uncertain parameter (referred to as the prior) may be assumed based on expert knowledge, further improvement in the uncertainty characterization may be made by performing Bayesian inference, where the probability distribution of the uncertainty is conditioned on available observational data (referred to as the posterior). A drawback of this approach is that many methods are intractable for large-scale simulations.

Date: September 21, 2020.

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 (LLNL-JRNL-778000).

As an example, and the focus of this work, Bayesian inference is applied to Darcy's law, which relates the underlying permeability field, pressure, and velocity through a porous medium, where the permeability field is treated as uncertain. In many cases we may rely on expert opinion to describe the underlying subsurface field structure and thus permeability field; however, given the heterogeneity of the soil and limited availability of data, treating the permeability as a spatially correlated random field to account for this lack of knowledge will provide more informative predictions. Furthermore, observational data, in the form of local pressure measurements, may be used to further improve how we characterize the uncertainty in the permeability field via Bayesian inference. For moderate sized discretization problems, performing Bayesian inference on Darcy's law is already computationally challenging, and many approaches have been developed to accelerate performing inverse UQ. Given the expansive physical domain of interest for such problems that we consider, additionally we require large-scale simulations to account for large, irregular 3D spatial domains with a finely resolved unstructured mesh, resulting in extremely computationally demanding UQ approaches, where the complete algorithm must be scalable.

To this end, the goal of this work is to estimate the statistical moments of a particular quantity of interest (QoI) Q that is estimated from output solutions of the numerical model. In general, Q is a functional of the uncertain random variable u with prior density $\pi_{\text{prior}}(u)$, and we seek to estimate the statistical moments of Q with respect to the posterior density $\pi(u|y)$, that is, the density of u conditioned on available observational data denoted as y. While there is no closed form for the posterior density for our problems of interest, Bayes' law provides the relationship between the prior and posterior that aids in sampling approaches. However, as we consider the unknown u to be a spatially varying field (related to the permeability field), our sampling strategies are limited to Monte Carlo-based approaches.

Often, when seeking posterior moments of a QoI, we may consider approaches that target samples of Q from the posterior distribution, such as Markov chain Monte Carlo (MCMC) [37, 21, 18]. For large-scale 3D problems, standard MCMC is not feasible as it typically requires searching the parameter space and forming a possibly infeasible number of time-intensive forward simulations. An active area of research focuses on developing acceleration approaches for MCMC. Such methods include utilizing gradient and Hessian information to modify the MCMC proposal (requiring additional solvers beyond the forward PDE), as in [36, 42]. Other methods include multilevel approaches that utilize coarse grid solvers to accelerate mixing, as in [24, 14] or – in addition to accelerating MCMC mixing – utilize a telescoping sum to perform variance reduction (similar to multilevel Monte Carlo [22, 17, 7, 46]) resulting in fewer fine grid samples [26, 11, 25].

An alternative approach to target the posterior is to utilize importance sampling as is done in sequential Monte Carlo [12, 38]. After drawing samples according to an initial distribution, simulations are resampled according to a sequence of distributions. The resampling requires calculating weights associated with each sample for each distribution in the sequence. Multilevel sequential Monte Carlo strategies are developed in [30, 5, 33], where the authors build the sequence of distributions from coarse approximations of the posterior, and in [5] they also employ a variance reduction strategy similar to multilevel Monte Carlo.

Of interest in this work is the ratio estimator [44, 9, 8, 43] as it enables us to form the posterior estimates of the QoI while sampling from only the prior distribution, avoiding the expensive acceptance-rejection process of MCMC. Furthermore, forming estimates from

the prior distribution may be done with standard forward UQ methodologies and leveraging existing codebases. The works of [8, 43] compare Monte Carlo (MC), quasi-Monte Carlo (QMC), and multilevel Monte Carlo (MLMC) methods to estimate the components of this ratio estimator, indicating that QMC and MLMC see improved performance over MC. However, for 3D problems of interest, the implementation of their work does not scale in the overall simulation process, as they use a Karhunen–Loève expansion (KLE) to form realizations of the permeability field. The primary issue when forming the KLE is that the construction requires obtaining a (possibly low-rank) eigenvalue decomposition of the covariance matrix. For large-scale problems of interest, this is not feasible, as naïve implementations scale cubically with the dimension of the unknown u. Improvements may be achieved with hierarchical matrix approaches resulting in log-linear scaling; however, storage may become an issue. Another approach is circulant embedding [20]; and while possible to perform in parallel, there is not software available that will perform circulant embedding on unstructured meshes. For a parallelizable and scalable approach, we employ algebraic multigrid (AMG) techniques – in particular element-agglomerated AMG (AMGe) – to both generate the unknown random field realizations (as in developed in [39, 40]) and to perform forward solves of Darcy's model problem. Utilizing AMGe will allow for unstructured meshes and simulations that scale linearly with the problem size.

In this paper, motivated by the theoretical derivations in [43], we extend their work to solving the mixed Darcy equations using the multilevel, hierarchical sampler method from [39, 40], resulting in a Bayesian inference method for estimating moments of a QoI with respect to the posterior distribution with efficient overall scaling. The remainder of the paper is organized into five main sections. In Section 2 we discuss Bayes' law and provide the framework for the ratio estimator as in [43]. In Section 3 the weak formulation of the mixed Darcy equations is presented, followed by a discussion of the discrete form and numerical solver. Section 4 provides background on the stochastic PDE implementation of [39, 40], as well as the numerical solver applied in this work. In Section 5 we discuss notation and main steps of MC and MLMC for performing UQ, and how these methods are applied to the ratio estimator as in [43]. Finally, Section 6 presents the numerical results, where we compare ratio estimates of MC and MLMC on the mixed Darcy equations. To this end, we consider two 3D examples, to showcase the computational speed up and scalability of the method.

2. Mathematical Framework of Bayesian Inverse Problems

Following the Bayesian approach to inverse problems, we seek to estimate moments of a QoI with respect to the *posterior* probability density of the unknown parameter conditioned on observational data. Let $u \in X$ denote the unknown parameter, with X a function space defined over the spatial domain D, an open and bounded subset of $\mathbb{R}^d$, $d = 1, 2, 3$. In particular, given the probability space (Ω, A, μ_0), with sample space Ω, σ-Algebra A, and prior measure μ_0, we consider the unknown u to be a random field on $D \times \Omega$ with $X = L^2(D)$, the Hilbert space of square-integrable functions, and $\mu_0(X) = 1$. Let $y \in Y$ denote the observational data, where $Y = \mathbb{R}^m$ for some $m \in \mathbb{N}$.

Let V be another function space, and define the *forward response map* as $\mathcal{B} : X \to V$, which maps the uncertain input $u \in X$ to a response. Furthermore, the bounded linear *observational functional* is given by $\mathcal{H} : V \to Y$, which maps the response to the space of observational data. Then the *uncertainty-to-observation* map is the composition of the

forward model with its projection onto the observed data and is denoted as $\mathcal{G} := \mathcal{H} \circ \mathcal{B} : X \to Y$.

In practice, it is not assumed that the unknown parameter is mapped directly to the observations; rather, the observational data have been corrupted by additive Gaussian observational noise, such that

$$y = \mathcal{G}(u) + \eta, \quad \eta \sim \mathcal{N}(0, \Gamma_\eta), \tag{1}$$

where η is zero-mean with covariance $\Gamma_\eta = \sigma_\eta^2 I_m$, for some $\sigma_\eta^2 > 0$. More details on the operator $\mathcal{G}$ of this work will be discussed in Section 3. Given (1), the *likelihood* is defined as

$$\pi_{like}^y := \exp\left[-\Phi(u; y)\right] = \exp\left[-\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2\right], \tag{2}$$

where $\Phi(u; y) := -\frac{1}{2\sigma_\eta^2} \|y - \mathcal{G}(u)\|^2$ denotes the *log-likelihood*.

The infinite-dimensional version Bayes' formula establishes the relation between the posterior measure μ^y and the prior measure μ_0 of the parameter [45], and is given by

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z}\pi_{like}^y, \tag{3}$$

which means that for any measurable function ψ, we have

$$\int_X \psi(u)\, d\mu^y(u) = \frac{1}{Z} \int_X \psi(u)\pi_{like}^y\, d\mu_0(u). \tag{4}$$

The *normalization* constant Z above is given by

$$Z := \int_X \exp\left[-\Phi(\cdot\,; y)\right] d\mu_0(u). \tag{5}$$

2.1. **Computation of Posterior Expectations.** We seek the expected value of a QoI, $Q(u) = \mathcal{Q}[\mathcal{B}(u)]$, $\mathcal{Q} : V \to \mathbb{R}$, under the posterior distribution μ^y. Typically, we do not have a closed form expression for the posterior distribution μ^y, since the normalizing constant Z is not known explicitly. However, using Bayes' formula (or (4) for $\psi = Q$), it can be shown that the posterior QoI expectation may be expressed as the ratio of two prior expectations:

$$\mathbb{E}_{\mu^y}[Q(u)] = \frac{\mathbb{E}_{\mu_0}\left[Q \cdot \pi_{like}^y\right]}{\mathbb{E}_{\mu_0}\left[\pi_{like}^y\right]} := \frac{R}{Z}, \tag{6}$$

As we are able to sample from the prior, the form of (6) provides a method to compute the posterior expectation; see equation (2.5) of [43], which utilizes a derivation of equation (6.24) in [45].

For problems where the discrete approximation of the unknown u is high-dimensional, as in this work, approximating R and Z must be done with care. In particular, we consider random sampling strategies, i.e., Monte Carlo and its variants as in [43], to approximate the prior expectations in (6). The details of MC and MLMC sampling methods will be presented in Section 5. Now that we have defined posterior QoI estimates with respect to the prior distribution and have motivated sample-based strategies, we move to discuss the model problem with uncertain input coefficient.

3. Forward Model Formulation with Uncertain Permeability

The application focus of this work is Darcy's law to model subsurface flow where the permeability $k(\mathbf{x}, \omega)$ is subject to uncertainty and is modeled as a lognormal random field. For a fixed $\omega \in \Omega$, velocity $\boldsymbol{q}$ and pressure p may be determined by solving

$$\begin{aligned} \tfrac{1}{k(\mathbf{x},\omega)}\boldsymbol{q}(\mathbf{x},\omega) + \nabla p &= \mathbf{f} && \text{in } D \\ \nabla \cdot \boldsymbol{q} &= 0 && \text{in } D, \end{aligned} \tag{7}$$

with Dirichlet boundary condition $p = p_D$ on Γ_D, enforced by the right-hand side $\mathbf{f}$, and Neumann boundary condition $\mathbf{q} \cdot \mathbf{n} = 0$ on Γ_N, where Γ_D and Γ_N are non-overlapping partitions of ∂D. The scalar QoI Q will depend on $\boldsymbol{q}$ and the likelihood will depend on p; however, we postpone defining Q and the likelihood until Section 6.

3.1. Discrete Spaces and Associated Darcy Solver.

Define the spaces $W = L^2(D)$ with inner product $(u, v) = \int_D uv d\mathbf{x}$ for all $u, v \in W$ and $\boldsymbol{R} = H(\text{div}; D) := \{\mathbf{q} \in [L^2(D)]^d | \text{ div } \mathbf{q} \in L^2(D),\ \mathbf{q} \cdot \mathbf{n} = 0 \text{ on } \partial D\}$ with inner product $(\mathbf{q}, \mathbf{s}) = \int_D \mathbf{q} \cdot \mathbf{s} d\mathbf{x}$ for all $\mathbf{q}, \mathbf{s} \in \boldsymbol{R}$.

For the discretization scheme, define the discrete space $\boldsymbol{R}_h \subset \boldsymbol{R}$ to be the lowest order Raviart-Thomas element, and $W_h \subset W$ to be the space of piecewise constants, both on an unstructured triangulation $\mathcal{T}_h$ of D with mesh size h. For a single discrete realization of k, denoted k_h, solutions of the forward problem $(\mathbf{q}_h, p_h) \in \boldsymbol{R}_h \times W_h$ are calculated via the discretized weak formulation

$$\begin{aligned} (k_h^{-1}\mathbf{q}_h, \mathbf{s}_h) - (\text{div}\, \mathbf{s}_h, p_h) &= (\mathbf{f}_h, \mathbf{s}_h) && \forall \mathbf{s}_h \in \mathbf{R}_h, \\ (\text{div}\, \mathbf{q}_h, v_h) &= 0 && \forall v_h \in W_h, \end{aligned} \tag{8}$$

subject to the corresponding boundary conditions of (7). Here we consider the mixed formulation due to the availability of scalable and robust solvers. More importantly, it allows for direct approximation of the velocity needed for the QoI, and piecewise constant pressure needed in the likelihood calculation.

The reader is directed to [19] for discretization error analysis of a general mixed formulation of Darcy's equations using the Raviart-Thomas finite elements, when k_h is a lognormal random field. Here we do not discuss convergence theory, but rather focus on the numerical results of Section 6 to illustrate convergence.

The associated linear system for (8) is constructed as follows. Define $M(k)_h$ to be the mass matrix associated with the inner product $(k_h^{-1}\mathbf{q}_h, \mathbf{s}_h)$ for a fixed k_h and B_h to be the matrix associated with the bilinear form $(\text{div}\, \mathbf{q}_h, v_h)$. Then solving (8) amounts to solving the linear system

$$\begin{bmatrix} M(k)_h & B_h^T \\ B_h & 0 \end{bmatrix} \begin{bmatrix} \mathbf{q}_h \\ p_h \end{bmatrix} = \begin{bmatrix} \mathbf{f}_h \\ 0 \end{bmatrix}. \tag{9}$$

In this work, (9) is solved by a block Jacobi-type preconditioned MINRES, where the system is preconditioned with

$$B_1 = \begin{bmatrix} \tilde{M}_h & 0 \\ 0 & \tilde{S}_h \end{bmatrix}, \tag{10}$$

where $\tilde{S}_h = B_h D_M^{-1} B_h^T$ and D_M is the diagonal of $M(k)_h$. In our numerical experiments, the action of applying the inverse of $\tilde{M}_h$ is 3 Gauss-Seidel iterations, and applying the inverse of $\tilde{S}_h$ is approximated with a single algebraic multigrid V-cycle; specifically, we use the scalable

BoomerAMG [23] from hypre [2] constructed on the basis of $\tilde{S}_h$. We note that $\tilde{S}_h$ (and the AMG preconditioner) must be recomputed for each new realization of k_h.

In practice, each forward simulation requires a new realization of the permeability $k(\mathbf{x},\omega) = k^* + \exp(u(\mathbf{x},\omega))$, which is formed from independent realizations of the Gaussian random field $u(\mathbf{x},\omega)$. Common approaches to forming the field $u(\mathbf{x},\omega)$ include using a truncated KLE [16], circulant embedding [20], as well as SPDE approach [35]. As discussed in Section 1, a naïve KLE or circulant embedding approach is not directly applicable for large-scale problems where the problems are posed on unstructured meshes. Note, however, there are KLE formulations that may work in this environment, e.g., [6], and comparisons with SPDE are a consideration for future work. In this work we have chosen to form the Gaussian random fields via SPDE due to its efficient scalability, as well as the availability of tools to form these realizations on unstructured meshes (see, e.g., [3, 4]).

4. Prior Random Field Sampling via Stochastic Partial Differential Equation

Following work in [35], a stochastic PDE method may be employed to form Gaussian random field realizations based on the exponential covariance function:

$$\operatorname{cov}(\mathbf{x},\mathbf{y}) = \sigma^2 \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|}{\lambda}\right), \tag{11}$$

with correlation length λ. The associated stochastic PDE from which realizations of the random field u are formed is given by

$$(\kappa^2 - \Delta)u(\mathbf{x},\omega) = g\mathcal{W}(\mathbf{x},\omega), \quad \mathbf{x}\in D, \kappa > 0, \tag{12}$$

where $\kappa = 1/\lambda$, $\mathcal{W}(\mathbf{x},\omega)$ is a realization of the standard Gaussian *white noise* with unit variance, and g is a scaling factor used to impose unit variance on u. Solutions to (12) form realizations of a continuous Gaussian random field. In practice, finite dimensional, discrete approximations of u are formed via numerical methods.

4.1. Mixed Finite Element Approach to Form Realizations.

Recall the discrete space $\boldsymbol{R}_h \subset \boldsymbol{R}$ is the lowest order Raviart-Thomas element, and $W_h \subset W$ is the space of piecewise constants, both on an unstructured triangulation $\mathcal{T}_h$ of D with mesh size h. We seek solutions $(\boldsymbol{\rho}_h, u_h) \in \mathbf{R}_h \times W_h$ from the system

$$\begin{aligned} (\boldsymbol{\rho}_h, \mathbf{s}_h) + (\operatorname{div}\mathbf{s}_h, u_h) &= 0 && \forall \mathbf{s}_h \in \mathbf{R}_h \\ (\operatorname{div}\boldsymbol{\rho}_h, v_h) - \kappa^2 (u_h, v_h) &= -g(\mathcal{W}, v_h) && \forall v_h \in W_h. \end{aligned} \tag{13}$$

with essential boundary conditions $\boldsymbol{\rho}_h \cdot \mathbf{n} = 0$. Similar to the discrete form of Darcy's equations in (8), we employ a mixed system to solve (12), as it allows for element-by-element discontinuous (e.g., piecewise constant) approximate solutions for u_h which is of our main interest.

Similar to Section 3.1 let M_h be the mass matrix associated with inner product $(\boldsymbol{\rho}_h, \mathbf{s}_h)$, Λ_h with the inner product (u_h, v_h) which is diagonal (due to the piecewise discontinuity of the spaces), and B_h with the bilinear form $(\operatorname{div}\mathbf{s}_h, u_h)$. A realization of the righthand side of (13) may be calculated, as in [35], using the fact that $\mathbb{E}[(\mathcal{W}, v_i)] = 0$ and $\operatorname{cov}((\mathcal{W}, v_i), (\mathcal{W}, v_j)) = (v_i, v_j)$ for $v_i, v_j \in W$. For piecewise constant $v_h \in W_h$ it follows that a realization of $(\mathcal{W}, v_h)$ is given as $\Lambda_h^{1/2}\xi_h(\omega)$, with $\xi_h(\omega) \sim \mathcal{N}(0, I)$.

For a fixed $\omega \in \Omega$, solving (13) amounts to solving the linear system given as

$$\mathcal{A}_h U_h = \begin{bmatrix} M_h & B_h^T \\ B_h & -\kappa^2 \Lambda_h \end{bmatrix} \begin{bmatrix} \boldsymbol{\rho}_h \\ u_h \end{bmatrix} = \begin{bmatrix} 0 \\ -g\, \Lambda_h^{1/2} \xi_h(\omega) \end{bmatrix} = F_h(\omega). \tag{14}$$

Our framework allows for different preconditioners to solve this system [4]; and within this framework, we utilize the hybridization AMG method of [34, 10], which allows for a scalable parallelizable implementation. In particular, an equivalent problem is constructed with a modified (broken) lowest order Raviart-Thomas space, such that the basis functions of the modified space have support within only a single element; this results in a highly parallelizable system, where M_h and B_h become decoupled blocks in the modified space, leading to block diagonal matrices. In practice, the setup of the hybridization AMG solver is computed only once, and can be reapplied to different realizations of the right-hand side $g\, \Lambda_h^{1/2} \xi_h(\omega)$. In our experiments the resulting hybridization linear system is solved with the conjugate gradient method preconditioned with hypre's BoomerAMG [23].

Remark 1 (Artificial Boundary Effects)**.** *As discussed in* [35] *numerical solutions to (12) suffer from the variance being artificially inflated along the boundary. A standard remedy is finding solutions on an extended domain; however, this is not straightforward when the mesh is unstructured, as we will consider in this work. To mitigate this issue, solutions are formed by applying the mesh embedding technique from* [40]*. In* [40]*, the domain D is embedded in a larger mesh $\bar{D}$ in a manner that does not require vertices of the unstructured meshes of either domain to match. More precisely, solutions to (13) are solved on the domain $\bar{D}$, then projected onto the mesh of interest D; as solutions u_h are in W_h, the space consisting of piecewise constants, the transfer is scalable and the meshes can be arbitrarily distributed among processors.*

Now that we have discussed the implementation for a single level of mesh refinement, we discuss the methodology in a multilevel setting.

4.2. **Multilevel Realizations.** We now consider a sequence of finite element spaces $\boldsymbol{R}_\ell$ and W_ℓ on an unstructured triangulation $\mathcal{T}_\ell$ of size h_ℓ, for $\ell = 0, 1, \dots L$. We assume that the hierarchy of unstructured meshes has been generated by recursively agglomerating finer level elements. Here different values of ℓ refer to different levels of mesh widths (characteristic diameter of agglomerated elements) with $h_0 < h_1 < \cdots < h_L$, e.g., h_0 corresponds to the level with finest mesh, and h_L to that with the coarsest mesh. The sequence of finite element spaces, associated with the (agglomerated) mesh, is constructed using methodology from element-agglomerated AMG (AMGe) methods, so that we are able to construct operator-dependent coarse spaces for $H(\mathrm{div})$ problems with guaranteed approximation properties on general, unstructured grids; see [31, 32, 41, 29] for technical details.

Consider the linear system in (14) where a subscript ℓ indicates mesh size h_ℓ. For a single level realization of the Gaussian random field, it is necessary to only solve $\mathcal{A}_\ell U_\ell = F_\ell(\omega^{(i)})$ for a fixed $\omega^{(i)}$. To obtain a multilevel realization, as is needed for MLMC, furthermore requires the corresponding coarse solution to $\mathcal{A}_{\ell+1} U_{\ell+1} = F_{\ell+1}(\omega^{(i)})$. This is done following the work of [39]. Define the interpolation operators $P_u : W_{\ell+1} \to W_\ell$ and $P_{\boldsymbol{\rho}} : \boldsymbol{R}_{\ell+1} \to \boldsymbol{R}_\ell$ constructed using techniques from AMGe [31, 32, 41]. First, a level ℓ realization of $\xi_\ell(\omega^{(i)}) \sim \mathcal{N}(0, I)$ is generated, and second the coarse system is solved

$$\mathcal{A}_{\ell+1} U_{\ell+1} = \begin{bmatrix} 0 \\ -g P_u^T \Lambda_\ell^{1/2} \xi_\ell(\omega^{(i)}) \end{bmatrix}, \tag{15}$$

using the hybridization and AMG technique as discussed in Section 4.1. Third, to the solve for corresponding fine level U_ℓ, the interpolated coarse solution,

$$\begin{bmatrix} P_{\boldsymbol{\rho}} & 0 \\ 0 & P_u \end{bmatrix} U_{\ell+1},$$

is applied as an initial guess fine level ℓ solver. Work showing that both level ℓ and level $\ell+1$ solutions produce Gaussian random fields is found in Proposition 3.1 of [39].

5. Monte Carlo and Multilevel Monte Carlo for Posterior QoI Mean Estimation

In this section we briefly review key concepts of Monte Carlo and multilevel Monte Carlo methods [22, 17, 7, 46] for estimating the posterior mean $\mathbb{E}_{\mu^y}[Q(u)] = R/Z$ from (6).

Since the forward problem is a PDE, the quantities $Q(u)$ and $\Phi(u;y)$ cannot be computed exactly, thus we consider the numerical approximations $Q_h(u_h) = \mathcal{Q}_h[\mathcal{B}_h(u_h)]$ and

$$\Phi_h(u_h;y) = \frac{1}{2\sigma_\eta^2}\left\|y - \mathcal{H}_h(\mathcal{B}_h(u_h))\right\|^2 \tag{16}$$

where the forward problem is discretized on the triangulation $\mathcal{T}_h$. Subsequently, we consider the prior expectations of the numerical approximations:

$$R_h = \mathbb{E}_{\mu_0}\big[Q_h \cdot \exp[-\Phi_h]\big] \text{ and } Z_h = \mathbb{E}_{\mu_0}\big[\exp[-\Phi_h]\big]. \tag{17}$$

We can now define the MC and MLMC estimators for R_h and Z_h. For the purpose of clarity, we review MC and MLMC estimators for the expected value for some functional $f_h(u_h) : X_h \to \mathbb{R}$, where the formulation and notation will directly apply to both R_h and Z_h.

5.1. Standard Monte Carlo methods.

In standard Monte Carlo, the expectation $\mathbb{E}_{\mu_0}[f_h]$ is estimated via a sample mean. Given N independently identically distributed samples, ${u_h}^{(i)}$ for $i = 1, \ldots, N$, according to the prior distribution, the MC estimate of $\mathbb{E}_{\mu_0}[f_h]$ is defined as

$$\widehat{f}_{h,N}^{MC,\mu_0} = \frac{1}{N}\sum_{i=1}^{N} f_h({u_h}^{(i)}). \tag{18}$$

It can be shown that the mean square error (MSE) of the estimator is

$$\mathbb{E}\left[\left(\widehat{f}_{h,N}^{MC,\mu_0} - \mathbb{E}_{\mu_0}[f]\right)^2\right] = \frac{\mathbb{V}_{\mu_0}[f_h]}{N} + \left(\mathbb{E}_{\mu_0}[f - f_h]\right)^2, \tag{19}$$

where $\mathbb{V}_{\mu_0}[\cdot]$ is the variance. In (19), the first term is the sampling error and the second term is related to the discretization error. For a desired MSE tolerance of ε^2, one often requires $\mathbb{V}_{\mu_0}[f_h]/N \leq \varepsilon^2/2$, and thus the total number of samples N scales as $2\mathbb{V}_{\mu_0}[f_h]/\varepsilon^2$. For large $\mathbb{V}_{\mu_0}[f_h]$, a larger number of samples and thus more forward PDE solves are required for a desired MSE tolerance, which can be cost prohibitive. Alternatively, more cost efficient approaches may be considered, e.g., variance reduction methods, where by reducing the variance of f_h, fewer forward simulations are required, thus reducing the overall cost. To this end, multilevel Monte Carlo (MLMC) is considered.

5.2. **Multilevel Monte Carlo methods.** In MLMC, the expectation $\mathbb{E}_{\mu_0}[f_h]$ is estimated using a hierarchy of spatial discretizations as introduced in Section 4.2, called levels, wherein most of the computational burden is placed on forward PDE solves using coarser levels. Assume we have a sequence of approximations $f_L, \ldots, f_1$ of the QoI f_h, where $f_\ell := f_{h_\ell}$ for $\ell := L, \ldots, 1$ denotes the QoI obtained from the level ℓ solver and $f_h := f_0$, then we have

$$\mathbb{E}_{\mu_0}[f_h] = \mathbb{E}_{\mu_0}[Y_L] + \sum_{\ell=0}^{L-1} \mathbb{E}_{\mu_0}[Y_\ell], \tag{20}$$

where $Y_\ell = f_\ell - f_{\ell+1}$ for $\ell = 0, \ldots, L-1$, and $Y_L = f_L$. Instead of applying MC to directly estimate $\mathbb{E}_{\mu_0}[f_h]$, a MC estimator is employed to estimate to each expectation in (20) to obtain the estimator for Y_ℓ:

$$\widehat{Y}_{\ell,N_\ell} = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \left(f_\ell({u_\ell}^{(i)}) - f_{\ell+1}({u_{\ell+1}}^{(i)}) \right), \tag{21}$$

where N_ℓ is the number of samples for level ℓ. Then it follows that the MLMC estimator of $\mathbb{E}_{\mu_0}[f_h]$ is

$$\widehat{f}_h^{ML,\mu_0} = \sum_{\ell=0}^{L} \widehat{Y}_{\ell,N_\ell}. \tag{22}$$

It can be shown that the MSE of the MLMC estimator is given by,

$$\mathbb{E}\left[\left(\widehat{f}_h^{ML,\mu_0} - \mathbb{E}_{\mu_0}[f] \right)^2 \right] = \sum_{\ell=0}^{L} \frac{\mathbb{V}_{\mu_0}[Y_\ell]}{N_\ell} + \left(\mathbb{E}_{\mu_0}[f - f_h] \right)^2. \tag{23}$$

The cost reduction lies in the sampling error of MLMC; as compared to the sampling error of MC (in (19)), the MLMC variance associated with the finest level QoI, level $\ell = 0$, is $\mathbb{V}_{\mu_0}[Y_0]$, i.e., the variance of the difference between f_0 and f_1. As $h \to 0$, $Y_\ell \to 0$, and the assumption (which can be theoretically or numerically verified) is that $\mathbb{V}_{\mu_0}[Y_\ell]$ approaches zero, resulting in fewer required samples N_ℓ on finer levels. Based on work in [17] it can be shown that the optimal number of samples to obtain a desired sampling error of $\varepsilon^2/2$ is given by

$$N_\ell \gtrsim \frac{2}{\varepsilon^2} \left[\sum_{k=0}^{L} \sqrt{\mathbb{V}_{\mu_0}[Y_k]\mathcal{C}_k} \right] \sqrt{\frac{\mathbb{V}_{\mu_0}[Y_\ell]}{\mathcal{C}_\ell}}, \tag{24}$$

where $\mathcal{C}_\ell$ is the cost per sample of Y_ℓ. In practice, the variance for each level correction Y_ℓ must be estimated in order to calculate (24).

Estimating the discretization error in (23) requires more care. In particular, it is important to verify that the finest level mesh size $h(= h_0)$ is small enough to obtain an accurate QoI. To this end, we employ the MLMC theory for PDEs to verify a bound on the both terms of the MSE; see, e.g., [7, 46]. Let M_ℓ be the spatial degrees of freedom (DOFs) for the mesh on level ℓ (i.e., the number of unknowns in the linear solver). Briefly stated, Theorem 1 of [7] states, that, for constants $\alpha, \beta, \gamma > 0$ such that $\alpha \geq \frac{1}{2}\min(\beta, \gamma)$ and

(1) $|\mathbb{E}_{\mu_0}[f_\ell - f]| \lesssim M_\ell^{-\alpha}$,
(2) $\mathbb{V}_{\mu_0}[Y_\ell] \lesssim M_\ell^{-\beta}$,
(3) $\mathcal{C}_\ell \lesssim M_\ell^{\gamma}$,

there exists a level of mesh refinement h, and number of levels $L+1$ with N_ℓ samples per level, such that, for a given $\varepsilon < e^{-1}$ it follows that,

$$\mathbb{E}\left[\left(\widehat{f}_h^{ML,\mu_0} - \mathbb{E}_{\mu_0}[f]\right)^2\right] < \varepsilon^2,$$

and

$$C(\widehat{f}_h^{ML,\mu_0}) \lesssim \begin{cases} \varepsilon^{-2}, \text{ for } \beta > \gamma \\ \varepsilon^{-2}(\log \varepsilon)^2, \text{ for } \beta = \gamma \\ \varepsilon^{-2-(\gamma-\beta)/\alpha}, \text{ for } \beta < \gamma, \end{cases}$$

where $C(\widehat{f}_h^{ML,\mu_0})$ is the total cost of performing MLMC to calculate $\widehat{f}_h^{ML,\mu_0}$. More concisely, given a value of ε, and values α, β, and γ, a specific sequence of levels exists that guarantees (up to a constant) bounds on both the MSE and total cost of performing MLMC. For comparison, the cost to perform standard MC will grow as $\varepsilon^{-2-\gamma/\alpha}$. In practice, the key is to estimate values α, β, and γ, to determine whether or not a value of ε is suitable for a particular sequence of levels.

5.3. **MC and MLMC Ratio Estimators for Posterior QoI Mean Estimation.** Using MC and MLMC estimators of R_h and Z_h (defined in (17)) as in (18) and (22), respectively, we can define the following estimators for the ratio estimator in (6). Define the MC posterior estimator to be

$$\widehat{Q}_{h,N}^{MC,\mu^y} := \frac{\widehat{R}_{h,N}^{MC,\mu_0}}{\widehat{Z}_{h,N}^{MC,\mu_0}}, \tag{25}$$

and the MLMC posterior estimator to be

$$\widehat{Q}_h^{ML,\mu^y} := \frac{\widehat{R}_h^{ML,\mu_0}}{\widehat{Z}_h^{ML,\mu_0}}. \tag{26}$$

Using $\widehat{R}_h$ and $\widehat{Z}_h$ to denote estimators – either MC estimator as defined in (18) or MLMC estimator in (22) – of R and Z, respectively, a bound on the MSE of the posterior estimator is provided in [43], where the authors show

$$\mathbb{E}\left[\left(\frac{\widehat{R}_h}{\widehat{Z}_h} - \frac{R}{Z}\right)^2\right] \leq \frac{2}{Z^2}\max\{1, \|\widehat{R}_h/\widehat{Z}_h\|_{L^\infty}^2\}\left(\mathbb{E}\left[\left(\widehat{R}_h - R\right)^2\right] + \mathbb{E}\left[\left(\widehat{Z}_h - Z\right)^2\right]\right). \tag{27}$$

For bounded $\widehat{R}_h/\widehat{Z}_h$ and Z away from zero, this indicates that the MSE of the posterior mean scales as the sum of the MSE of $\widehat{R}_h$ and the MSE of $\widehat{Z}_h$; in particular, MC and MLMC theory can be directly applied to the individual errors $\mathbb{E}[(R - \widehat{R}_h)^2]$ and $\mathbb{E}[(Z - \widehat{Z}_h)^2]$ to estimate the MSE of the posterior estimator.

Remark 2 (Scaling of Variance)**.** *In practice, the sample variances of R and Z are quite small, and thus we employ ideas similar to [47] with regard to the sampling error. That is, we instead use the squared coefficient of variation estimate to indicate the number of needed samples, where the squared coefficient of variation of R is defined as*

$$\delta^2 := \frac{\mathbb{V}_{\mu_0}[R]}{\mathbb{E}_{\mu_0}[R]^2},$$

and similarly for Z. For the numerical results, estimates of the variance are in fact estimates of the squared coefficient of variation, which are then used to estimate the number of required samples per level N_ℓ.

6. Numerical Results for Posterior QoI Mean Estimation

In this section we discuss the numerical performance of the ratio estimator when using MC and MLMC on the Darcy model problem with the SPDE sampler to generate Gaussian random field realizations. To this end we consider two test cases; first, we compare MC and MLMC versions of the ratio estimator on the unit cube to confirm MLMC outperforms MC, and second, we apply the ratio estimator with MLMC to the Egg model, as in [27], to test the estimator in a large-scale setting.

In both tests, we seek the posterior statistical moments, e.g., $\mathbb{E}_{\mu^y}[Q(u)]$ (as in (6)), of the effective permeability along the outflow boundary Γ_{out}, where the QoI is defined as

$$Q(u) = \frac{1}{|\Gamma_{\text{out}}|} \int_{\Gamma_{\text{out}}} \mathbf{q} \cdot \mathbf{n} \, dS, \tag{28}$$

with $\mathbf{q}$ a solution of (7). To form the ratio estimate, we utilize independent simulations of R_ℓ and Z_ℓ; both which require us to simulate the forward model to obtain velocity and pressure solutions $(\mathbf{q}_\ell, p_\ell)$. A sample of Z_ℓ requires the calculation of the pressure p_ℓ to estimate Φ_ℓ (see (16)), where $\|y - \mathcal{H}_\ell(\mathcal{B}_\ell(u_\ell))\|$ is the error between simulated and observed local pressure measurements. A sample of R_ℓ requires, in addition, the calculation of Q_ℓ (a discrete estimation of (28)). Algorithm 1 outlines the different steps taken to form scalable samples of R_ℓ or Z_ℓ.

Algorithm 1: Algorithm for simulating R_ℓ or Z_ℓ

i. Sample $\xi_\ell(\omega) \sim \mathcal{N}(0, I)$
ii. Form Gaussian random field realization u_ℓ via (14)
iii. Solve Darcy's equations to obtain $(\mathbf{q}_\ell, p_\ell)$ via (9), with $k_\ell = k^* + \exp(u_\ell)$
iv. Calculate $\Phi_\ell(u_\ell; y)$ via (16)
v. If estimating R_ℓ, calculate Q_ℓ via (28); otherwise skip
vi. Form a realization of R_ℓ or Z_ℓ via (17)

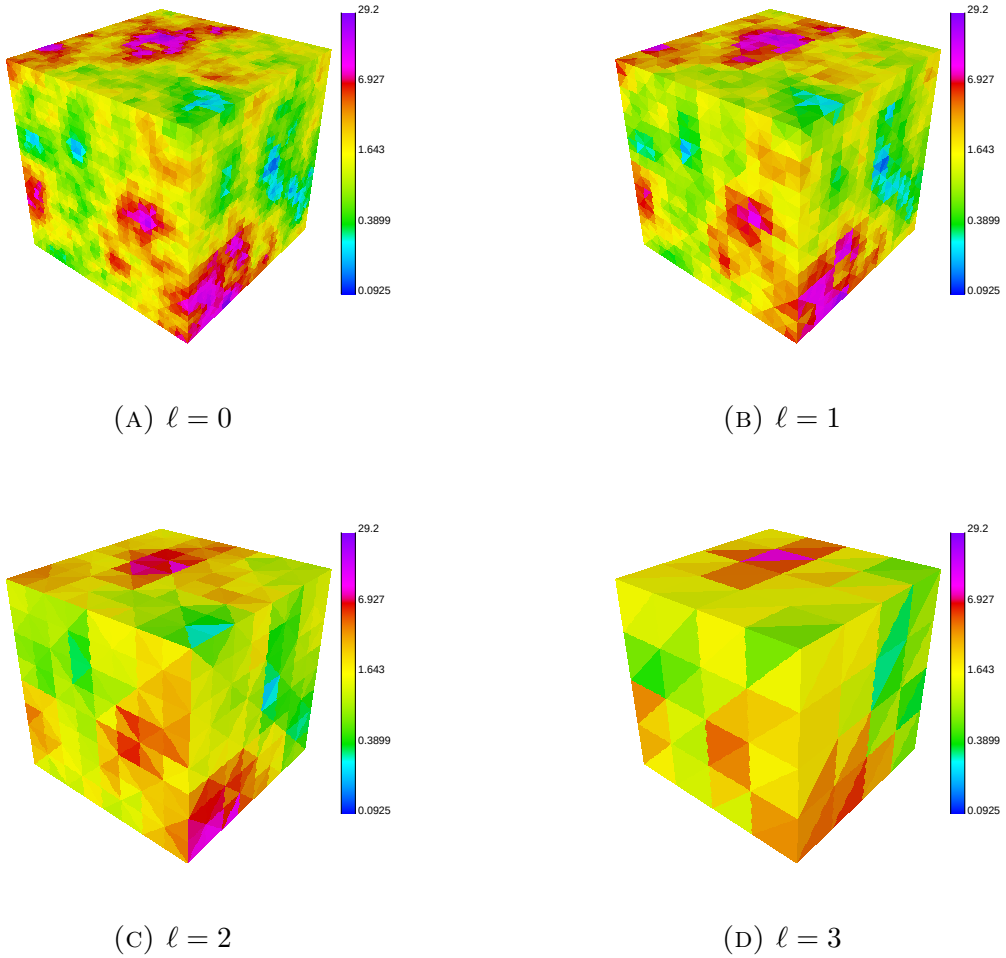
Numerical simulations – of both the SPDE and forward Darcy solvers – were performed using tools developed in ParELAG [3], a parallel C++ library for performing numerical upscaling of finite element discretizations using specialized element-based agglomeration techniques. ParELAG makes use of several scalable multilevel solvers and preconditioners from the hypre library. In particular for our results, BoomerAMG is used in both the solvers for the SPDE and forward Darcy problems. ParELAGMC [4], a parallel element agglomeration MLMC library, was applied to generate all MLMC results, with images generated with GLVis [1]. All timing results were executed on the Quartz cluster at Lawrence Livermore National Laboratory, consisting of 2,688 nodes where each node has two 18-core Intel Xeon E5-2695 processors. For the Egg model results, we use 36 MPI processes per node.

6.1. 3D Mixed Darcy Equations: Unit cube.

The first test case applies the ratio estimator to the Darcy model problem on a unit cube domain, where we consider $h_0 = 1/32$ to be the finest mesh size. For MLMC we consider three coarser levels with mesh sizes

$h_1 = 1/16$, $h_2 = 1/8$, and $h_3 = 1/4$ – all with tetrahedral elements. Furthermore, to avoid inverse crimes (see [28]), reference observational data are generated via (1), by performing the forward solve on a highly refined mesh of size $h_{\text{obs}} = 1/64$, and adding observational noise with variance $\sigma_\eta^2 = 0.1$. We note, initial results consider $m = 9$ uniformly spaced data points measuring local average pressure.

Algorithm 1 outlines the steps taken to generate samples of R_ℓ and Z_ℓ on each level. In step (ii), a realization of u_ℓ is formed with correlation length $\lambda = 0.3$, where the corresponding coarser realization is obtained via (14) when utilizing the same $\xi_\ell(\omega)$ from step (i). Next, the lognormal random field k on each level is calculated as $k_\ell = k^* + \exp(u_\ell)$ with $k^* = 0.001$ to enforce uniform ellipticity. Figure 1 displays a single realization of k_ℓ on the four different levels. For step (iii) of Algorithm 1, the Darcy solver is completed with boundary conditions



(A) $\ell = 0$ (B) $\ell = 1$

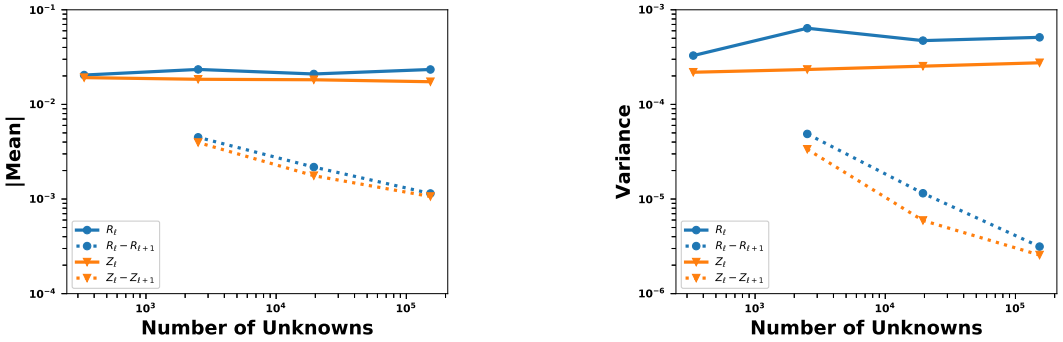
(C) $\ell = 2$ (D) $\ell = 3$

FIGURE 1. Corresponding realizations of the lognormal random field k on the four levels of refinement: (A) finest level $\ell = 0$, (B) level $\ell = 1$, (C) level $\ell = 2$, and (D) coarsest level $\ell = 3$.

of

$$\begin{cases} -p = 1 & \text{in } \Gamma_{in} := (0,1) \times (0,1) \times \{0\}, \\ -p = 0 & \text{in } \Gamma_{out} := (0,1) \times (0,1) \times \{1\}, \\ \mathbf{q} \cdot \mathbf{n} = 0 & \text{in } \partial D \backslash \{\Gamma_{in} \cup \Gamma_{out}\}. \end{cases}$$

To show that MLMC may be applied to estimate the moments of R and Z, we first consider the decay of mean and variance estimates of $R_\ell - R_{\ell+1}$ and $Z_\ell - Z_{\ell+1}$ compared to single level estimates for R_ℓ and Z_ℓ. Figure 2 (A) provides the mean estimates as a function of the number of unknowns in the linear solver. The decay of the mean helps indicate convergence in the discretization error (see Section 5.2 for theoretical error discussion). To estimate the discretization error of R and Z MLMC estimators we must estimate the value of α. In particular, we estimate $\alpha \approx 0.3$ from the slopes of $R_\ell - R_{\ell+1}$ and $Z_\ell - Z_{\ell+1}$. Figure 2 (B) provides the variance estimates as a function of the number of unknowns. The decay in variance for $R_\ell - R_{\ell+1}$ and $Z_\ell - Z_{\ell+1}$ indicates that fewer samples will be needed on the finer levels, as per the definition of N_ℓ (see (24)). The variance for both these correction terms decays with a slope of $\beta \approx 0.6$. These values of α and β correspond to the MLMC error and cost bounds discussed in Section 5, and will be used later in this section as a comparison with the computational cost result.



(A) Sample mean (B) Sample variance

FIGURE 2. (A) Absolute value of sample mean and (B) sample variance values for R_ℓ and Z_ℓ estimates in a single level, and $R_\ell - R_{\ell+1}$ and $Z_\ell - Z_{\ell+1}$ in a multilevel setting. The x-axis provides the number of unknowns in the linear solver. All data points determined from $N = 128$ samples on each level.

While the previous figure indicates that MLMC may be applied to the estimates of R and Z, the same does not necessarily hold for the QoI Q. Since Q is the ratio of the R and Z estimates, it must be confirmed that the MSE of Q scales according to the MSE of both R and Z estimates. To do this we approximate the discretization error and sampling error of R, Z, and Q. Figure 3 displays the decay of $R_\ell - R_{\ell+1}$, $Z_\ell - Z_{\ell+1}$, and $\hat{Q}_\ell - \hat{Q}_{\ell+1}$ estimates using the average of 500 samples. The decay lines indicate that as the discretization error of Q decays with the error decay of R and Z.

For Figures 4 and 5, we perform only a two-level MLMC, with the coarsest two levels; that is, $h_2 = 1/8$ is the fine level mesh size, and $h_3 = 1/4$ is the coarse level mesh size. Figure 4 provides the estimated sampling error decay, as a function of sample size (see (19) for MC and (23) for MLMC sampling error definitions). In particular, Figure 4 (A) displays this sampling error when using N MC samples. For MLMC, the number of samples is

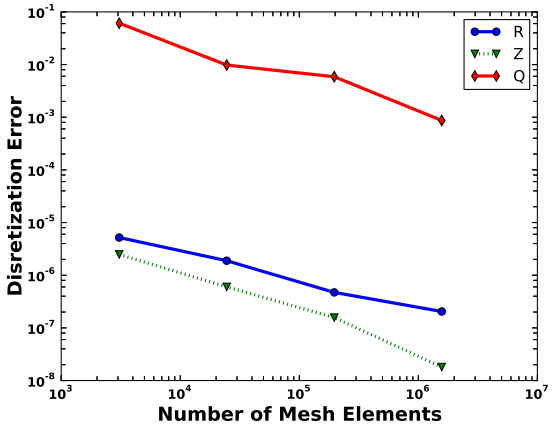
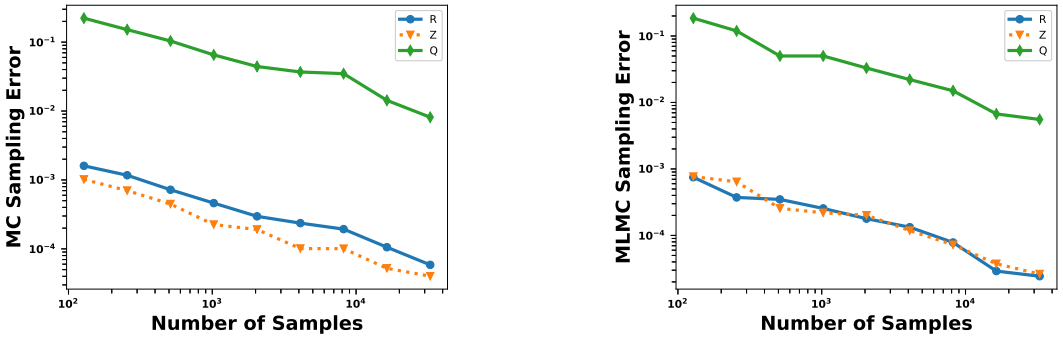


FIGURE 3. Decay of the discretization error for Q_ℓ, R_ℓ, and Z_ℓ for $N = 500$ samples on each level. The discretization error convergence rate of the ratio estimator Q_ℓ is approximately equal to the convergence rates of individual prior estimators R_ℓ and Z_ℓ.

selected to match the equivalent cost of N MC samples. For this comparison, we select the number of samples on each level, $\tilde{N}_\ell$, such that the MSE is optimized constraint to the cost $\mathcal{C}_{MLMC} = N\mathcal{C}_0$, resulting in

$$\tilde{N}_\ell = \frac{N\mathcal{C}_0}{\sum_{k=0}^{L}\sqrt{V_k C_k}}\sqrt{\frac{V_\ell}{C_\ell}}, \tag{29}$$

on each level. Figure 4 (B) provides the sampling error of a 2-level MLMC as a function of N, so that it is comparable to MC.

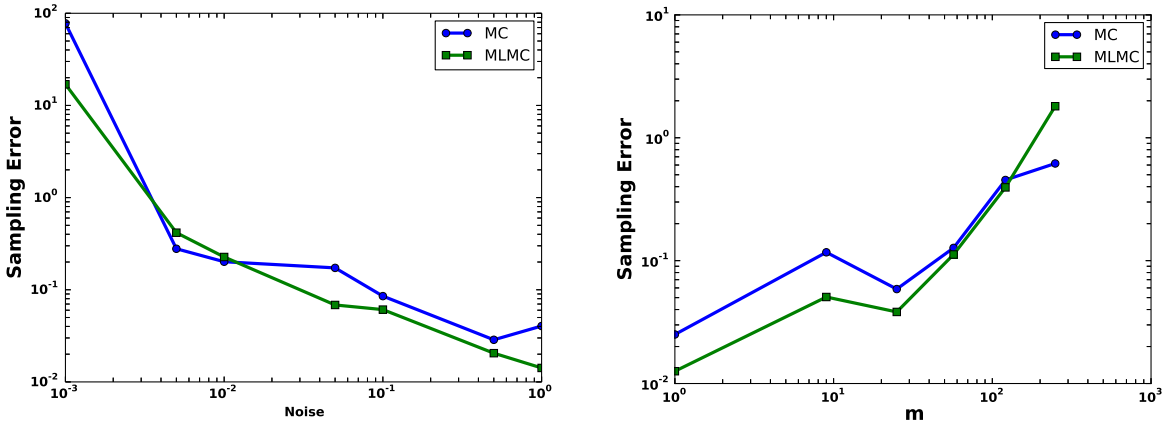


(A) MC Sampling Error (B) MLMC Sampling Error

FIGURE 4. (A) Decay of the MC sampling error (sample average of error for 10 runs). (B) Decay of the MLMC sampling error (sample average of error for 10 runs).

The next set of results compares the effect of noise and number of observations on the sampling error for ratio estimates of MC and MLMC, again treating h_2 as the fine level mesh size. Figure 5 (A) displays the estimated sampling error as a function of the observation noise variance σ_η^2 (see (1)) for $m = 9$ observations. These estimates are calculated by performing 10 runs of MC and MLMC and averaging the associated MSE estimates. The

sampling error improves with increased noise parameter. An important point to highlight is that for small values of the noise parameter, this method does not perform as well, as the probability density is compressed over a region that is difficult to sample from accurately. On the right, Figure 5 (B) provides the estimated sampling error as a function of the number of observations calculated in (1). For values $m < 100$, MC and MLMC sampling errors are approximately less than 0.01, while a larger number of observations results in sampling errors greater than 0.4.



(A) Vary noise variance σ_η^2

(B) Vary number of observation points

FIGURE 5. (A) Comparison of MC and MLMC sampling errors, with fixed $N = 1048$ and $m = 9$, as a function of noise parameter (sample average of error for 10 runs). (B) Comparison of MC and MLMC sampling errors, with fixed $N = 1048$ and $\sigma_\eta^2 = 0.1$, as a function of noise parameter (sample average of error for 10 runs)

The final result in Figure 6 compares the total wall time of MC and MLMC for various values of ε. Note that the calculated cost is determined from the R and Z MLMC and MC estimators to achieve specified MSE errors. As previously demonstrated, the sampling error and discretization error of Q scales similarly to that of R and Z estimators. As such, the cost scaling of the ratio estimator will perform similarly to the results in Figure 6.

In Figure 6 each MC and MLMC estimation performed, i.e., each data point, was completed on 4 MPI processors on Quartz. For $\varepsilon = 0.25$, MLMC requires one level (to meet the discretization error estimate), thus MC and MLMC have the same costs. For each smaller value of ε, an additional level in incorporated into MLMC so that the smallest value ε is associated with the 4-level MLMC. Furthermore, as the value of ε decreases, the cost of MLMC is clearly better than the cost of performing MC. For comparison, we also plot the theoretical cost increase of MLMC and MC as a function of ε (as discussed in Section 5). These values come from the $\beta < \gamma$ scenario, where the MLMC cost increases as $\varepsilon^{-2-(\gamma-\beta)/\alpha}$. Setting $\alpha = 0.3$, $\beta = 0.6$, and $\gamma = 1$ – as estimated in this numerical results section – we estimate that the cost of MLMC grows as $\varepsilon^{-10/3}$. For single-level MC, the cost grows as $\varepsilon^{-2-\gamma/\alpha} = \varepsilon^{-16/3}$.

These results demonstrate the cost improvement of MLMC compared to MC. In particular, MLMC has improved wall time scaling for decreased values of ε. Next, we investigate the scalability of this MLMC approach for increased problem size.

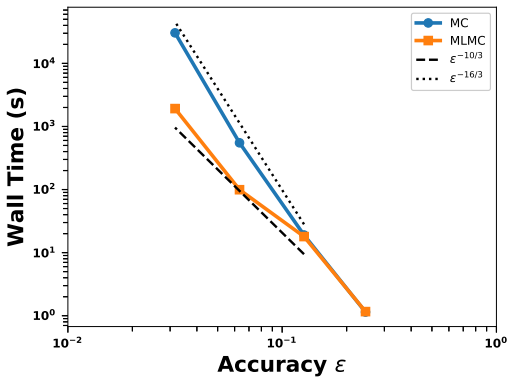


FIGURE 6. Computational cost of performing MC vs MLMC for various values of the root MSE accuracy ε. Both MC and MLMC runs completed on 4 processors.

6.2. **Cost scaling with the Egg Model.** In order to show that this approach is amenable to large-scale settings, we consider the same forward problem, but within a larger, irregular domain, as described by the 'Egg model' [27]. We consider the performance of computing the MLMC ratio estimator using a weak scaling perspective, i.e., increasing the fine level problem size proportionally to the processor count, for 3 different MLMC computations.

The Egg model is a 3D synthetic reservoir model containing an ensemble of synthetic realizations that contain channels of high permeability to represent subsurface river patterns. With regards to the discretization, the irregular Egg domain is contained by a 480 m $\times$ 480 m $\times$ 28 m bounding box. For the SPDE solver (step (ii) of Algorithm 1), the Egg model domain is embedded within a 512 m $\times$ 512 m $\times$ 44 m domain (see Remark 1). Figure 7 displays both the original Egg model mesh and enlarged mesh (in which it is embedded) for the coarsest level, both with hexahedral elements of size 8 m $\times$ 8 m $\times$ 4 m. For finer levels, the mesh of Figure 7 is uniformly refined by a factor of two in each direction.

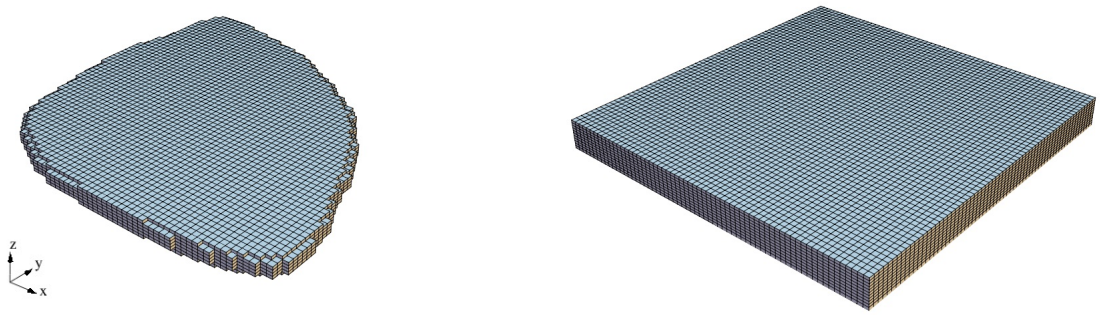
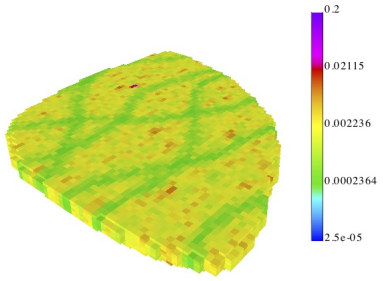
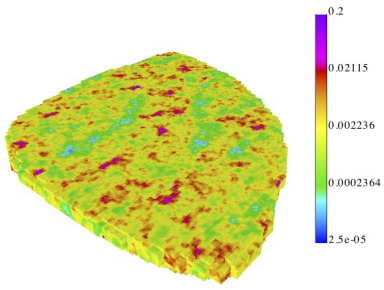


FIGURE 7. (A) Original Egg model mesh containing 18,553 elements. (B) Enlarged mesh, in which the Egg model mesh is embedded, extends two elements in each direction beyond the Egg model mesh bounding box. Both meshes displayed correspond to the coarsest level.

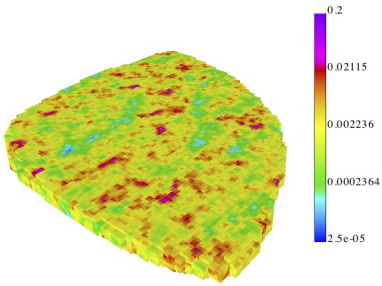
To generate realizations of R_ℓ and Z_ℓ, we following Algorithm 1. In step (ii), realizations of u_ℓ are formed with correlation length $\lambda = 10$ m. Dissimilar to the unit cube example, after calculating u_ℓ in step (ii), the permeability field is formed using data from [27]. That is, we employ the base permeability field k_{data}, with its inverse displayed in Figure 8 (A), and define the permeability as $k_\ell = k^* + k_{data}\exp(u_\ell)$ with $k^* = 0$, so that the synthetic permeability data is the mean of the lognormal field. It is important to note, that with this construction of k, uniform ellipticity is no longer enforced. Figure 8 (B)-(D) displays realizations of $\exp(u_\ell)$ on three different levels. These images represent solutions calculated on the coarsest mesh, and well as corresponding solutions on meshes with one and two additional levels of refinement. For step (iii) of Algorithm 1, the Darcy solver is completed



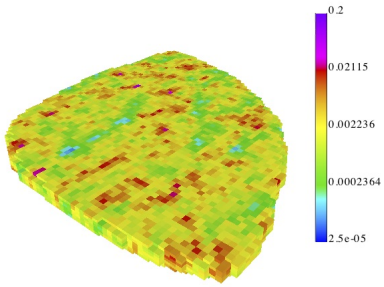
(A) Inverse permeability data k_{data}^{-1}



(B) $\ell = 0$



(C) $\ell = 1$



(D) $\ell = 2$

FIGURE 8. (A) Inverse of permeability field 18 from [27]. (B)-(D) A single realization of the coefficient k_ℓ^{-1} on levels $\ell = 0, 1$, and 2 on the smallest problem with 36 processors.

with boundary conditions of

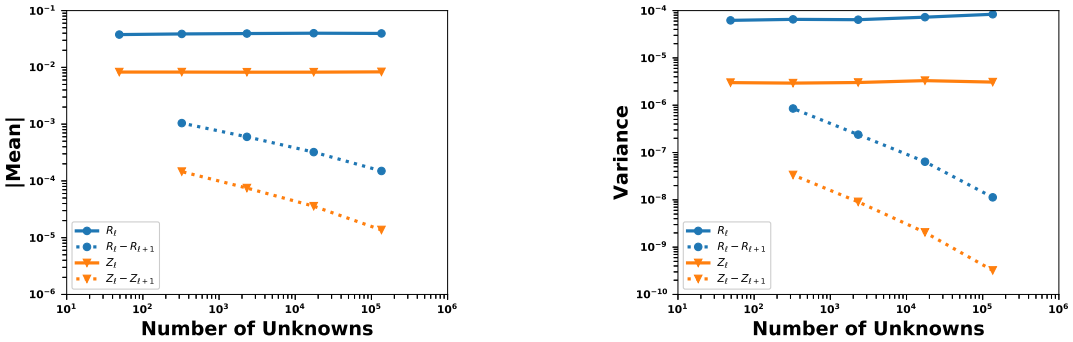
$$\begin{cases} -p = 1 & \text{on } \Gamma_{in} := \{0\} \times (0, 480) \times (0, 28), \\ -p = 0 & \text{on } \Gamma_{out} := (0, 480) \times \{0\} \times (0, 28), \\ \mathbf{q} \cdot \mathbf{n} = 0 & \text{on } \partial D \backslash \{\Gamma_{in} \cup \Gamma_{out}\}. \end{cases}$$

The reference observational data – localized average pressure – are extracted about $m = 12$ locations in the domain at $z = 12$ and (x, y) coordinates provided in Table 1. We consider observational noise with $\sigma_\eta^2 = 1.0$.

TABLE 1. Observational Data Coordinates

m	Coordinates
12	$(40, 399)$, $(240, 432)$, $(16, 280)$, $(216, 232)$, $(400, 280)$, $(64, 72)$, $(256, 16)$, $(399, 42)$, $(128, 344)$, $(280, 320)$, $(184, 128)$, $(344, 144)$

The first set of tests performed mimics those of standard MLMC; that is, we investigate the decay of the mean and variance of $R_\ell - R_{\ell+1}$ and $Z_\ell - Z_{\ell+1}$. This is done using a discretization with 7.6e+07 elements (corresponding to the largest problem in the weak scaling results below), with five levels using $2,304$ MPI processors. Figure 9 provides these results. The slopes indicate we have similar results to the cube model, in particular, we have values $\alpha \approx 0.3$ and $\beta \approx 0.6$. These results aid in the selection of ε for the remaining results.



(A) Sample mean (B) Sample variance

FIGURE 9. (A) Absolute value of sample mean and (B) sample variance values for R_ℓ and Z_ℓ estimates in a single level, and $R_\ell - R_{\ell+1}$ and $Z_\ell - Z_{\ell+1}$ in a multilevel setting. The x-axis provides the number of unknowns per processor.

For the scaling analysis, we consider the performance of adaptive MLMC for computing the ratio estimator with two smaller problem sizes (in addition to the five level formulation above), with the details provided in Table 2. For the each problem size, starting with the smallest, the number of MPI processors (NP) is increased by a factor of 8 to maintain the same number of spatial degrees of freedom per processor, approximately 1.39e+05 DOFs, on the fine level. Then the target MSE value ε^2 is reduced by a factor of 4, for each increase in problem size, to account for the more accurate spatial discretization. The smallest problem, with NP $= 36$, uses a three-level MLMC estimator for R and Z. As the problem size is increased, the coarse level is held fixed, so the number of levels increases as well. The largest problem size utilizes all five levels in the MLMC estimators. We note, the number of samples for each level (see (24)) is the combined total for R and Z multilevel estimators.

TABLE 2. MLMC Problem Sizes

NP	Global DOFs ($\ell = 0$)	Target MSE	Wall Time (s)	N_0	N_1	N_2	N_3	N_4
36	4.8063e+06	4.00e-05	869	35	284	7783		
288	3.8223e+07	1.00e-05	2,817	61	470	1746	28199	
2304	3.0488e+08	2.50e-06	16,824	111	813	2764	6892	102931

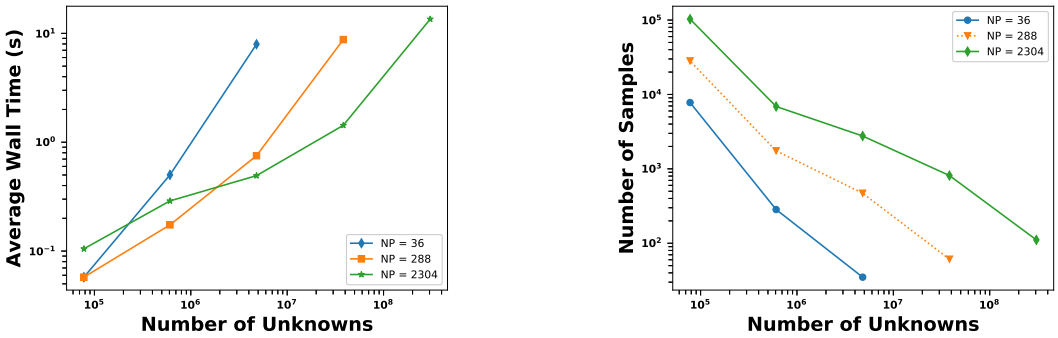
The first scaling results we consider are provided in Table 3 and Figure 10 (A). Table 3 provides the average computational time to generate a sample (averaged over the combined cost for R and Z, as the costs are comparable) on each level for the three different number of processors used in the experiments. We note this is both the cost to generate the random field input and to solve Darcy's equations. For $\ell = 0, 1, 2, 3$, each column provides individual weak scaling results, since each processor has approximately the same number of DOFs. As expected, level 0 performs the best, with efficiency decreasing to 90% on the middle problem (np=288), and 58% (np=2304) on the largest problem, relative to the smallest problem. As the levels are coarsened, the efficiency decreases, also to be expected; for example, for level 2 the efficiency decreases from 33% to 11%. For the coarsest problem on each row, the timings provide strong scaling results. Note that problems at a given level that have about the same number of global DOFs are distributed differently (as it is, for example, at the coarsest level in our tests). The different distribution implies different communication patterns which affects the performance at the coarse levels. In particular, the coarsest problem in the third row is about two times slower than the same problem in the second and first rows. This decreased scaling is expected for the coarser levels as the number of processors increases - a fact well known in strong scaling studies. As we are using BoomerAMG from the hypre library within the solve of both linear systems (sampler and forward problem), we are limited by the performance of the solver and AMG's performance degradation on coarse grids. For more detailed study on the increased communication complexity at coarse levels of the AMG solver we use, BoomerAMG, we refer to [15]. However, the decreased parallel efficiency at coarse levels does not substantially affect the overall parallel performance of the simulation. The latter can be improved by either redistributing the data on coarse levels by using less processors and hence improve the communication pattern, and/or incorporate parallelism in the sampling at these levels (cf. [13]).

Figure 10 (A) displays the average computational time per sample (averaged over the combined cost for R and Z) compared to the number of global unknowns in the forward problem. For optimal scaling, this cost will grow linearly, and be maintained as the number of processors are increased. On the finest levels, for each NP = 36, 288, and 2304, these results display the desired scaling property; however, on the coarser levels scaling deteriorates due to the reason explained above.

With regards to the adaptive MLMC results, Figure 10 (B) provides the number of samples N_ℓ (for combined R and Z) for each level. For a fixed number of processors, the number of samples increases for larger ℓ, that is, a larger number of samples are needed on the coarser levels. As the number of processors increases along with increased problem sizes, the number of samples needed across all levels increases. This is because these larger problems have a smaller target ε^2, and thus will need more samples to achieve the associated target sampling error.

TABLE 3. Average wall time (s) per simulation, separated into the three problem sizes (NP), on each level. Levels $\ell = 0, 1, 2, 3$ provide weak scaling results, where each column maintains the same number of DOFs per processor.

NP	$\ell = 0$	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$
36	7.95	0.501	0.0576		
288	8.77	0.752	0.174	0.0576	
2304	13.5	1.43	0.493	0.289	0.105

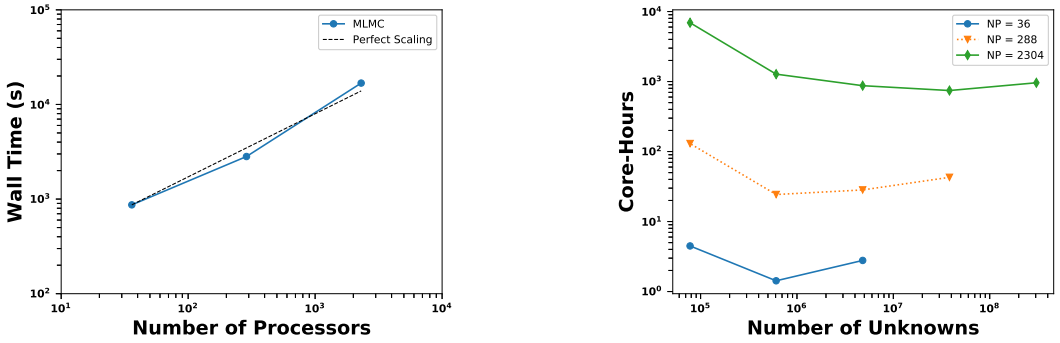


(A) Average time per simulation (B) Total number of samples

FIGURE 10. (A) Average wall time per forward simulation on each level, for each of the three problem sizes. (B) Total number of simulations N_ℓ (for combined R and Z) on each level.

The final results we consider correspond to adaptive MLMC scaling for all three problems. These results take into account total wall time to run all simulations required for the MLMC estimators. Figure 11 (A) displays the total wall time (in seconds) to perform MLMC on each problem size, in comparison to the *ε-adjusted* perfect scaling. Here the *ε-adjusted* cost growth takes into account that as we increase the problem size, we are also decreasing ε; as such, the number of samples on each level (see (24)) and thus the total cost increases accordingly (in contrast to traditional weak scaling results). While this indicates reasonable scaling, Figure 11 (B) breaks down this cost into the total number of core-hours spent on each level, for each problem size. This result indicates that the work is balanced well across most levels. The finer levels (those with larger number of unknowns in these results) have consistently good scaling performance. On the coarsest level, however, this is not that case. We observe degraded scaling, where more time is spent generating coarse level samples relative to the finer levels.

In our implementation the same number of processors are used for generating samples on each level. This results in a decreased scaling efficiency on the coarser levels, due to an increase in communication costs (relative to problem size) as well as the fact that more time per solve is typically spent on the coarse level in algebraic multigrid (see e.g., [15]). An option to improve on the coarse level scaling is to use level-dependent values of NP such that we can achieve scaling similar to that of the finest level across all the levels, as is done in [13].



(A) (B)

FIGURE 11. (A) Total wall time for the full MLMC algorithm across the different problem sizes. Here perfect scaling takes the change in ε – and thus the increase in number of samples on each level – into account. (B) Total number of core-hours spent on each level, for each problem size.

7. CONCLUSIONS

In this work we present a scalable approach to perform nonlinear Bayesian inference for large-scale 3D problems. The key components combine work from [43] and [39, 40] to form a multilevel sampling approach where both the formation of the spatially varying random field and forward PDE solver employ AMGe techniques to generate the hierarchy of coarse discretization spaces with guaranteed approximation properties and rely on the scalability of the utilized AMG solvers (such as the BoomerAMG in our case), and thereby scale as the solvers do with increase in problem size. We apply this approach to two 3D examples, a unit cube problem and the Egg model [27], to show the success of this approach. While each simulation is scalable, the key result is that within the implementation of the MLMC ratio estimate, the finest levels scale with increase in problem size, enabling us to perform Bayesian inference on large-scale simulations with complex spatial domains. While the forward PDEs solves were performed in parallel, further scheduling improvements can be made to increase the coarse level efficiencies, as well as implementing the MC simulations in parallel (see, e.g., [13]).

From these results, we see that the errors incurred by performing this ratio estimate are better in the parameter regime of large observational noise and small number of observations (as discussed in [43]). For applications where there is smaller observational noise, an MCMC approach will have improved accuracy. As the computational effort of MCMC is extremely demanding for large-scale problems, ongoing research is focused on extending the multilevel MCMC work in [11] to incorporate the already scalable features that are present in the current work, that is, scalable sampling and scalable forward PDE solves. Of interest will be to compare these two methods, given scalable generation of the spatially varying random field.

References

[1] GLVis: OpenGL finite element visualization tool. glvis.org.

[2] HYPRE: High performance preconditioners. http://www.llnl.gov/CASC/hypre/.

[3] ParELAG: Element-agglomeration algebraic multigrid and upscaling library, version 2.0. http://github.com/LLNL/parelag, 2015.

[4] ParELAGMC: Parallel element agglomeration multilevel Monte Carlo library. http://github.com/LLNL/parelagmc, 2018.

[5] A. Beskos, A. Jasra, K. Law, R. Tempone, and Y. Zhou. Multilevel sequential Monte Carlo samplers. *Stochastic Processes and their Applications*, 127(5):1417–1440, 2017.

[6] D. Cai, E. Chow, L. Erlandson, Y. Saad, and Y. Xi. SMASH: Structured matrix approximation by separation and hierarchy. *Numerical Linear Algebra with Applications*, 25(6):e2204, 2018.

[7] K.A. Cliffe, M.B. Giles, R. Scheichl, and A.L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, 14(1):3–15, 2011.

[8] J. Dick, R.N. Gantner, Q.T. Le Gia, and Christoph C. Schwab. Multilevel higher-order quasi-Monte Carlo Bayesian estimation. *Mathematical Models and Methods in Applied Sciences*, 27(05):953–995, 2017.

[9] J. Dick, R.N. Gantner, Q.T. Le Gia, and C. Schwab. Higher order quasi-Monte Carlo integration for Bayesian estimation. *arXiv preprint arXiv:1602.07363*, 2016.

[10] V. Dobrev, T. Kolev, C.S. Lee, V. Tomov, and P.S. Vassilevski. Algebraic hybridization and static condensation with application to scalable H(div) preconditioning. *SIAM Journal on Scientific Computing*, 41(3):B425–B447, 2019.

[11] T.J. Dodwell, C. Ketelsen, R. Scheichl, and A.L. Teckentrup. A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):1075–1108, 2015.

[12] A. Doucet, N. De Freitas, and N. Gordon. An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo methods in practice*, pages 3–14. Springer, 2001.

[13] D. Drzisga, B. Gmeiner, U. R"ude, R. Scheichl, and B. Wohlmuth. Scheduling massively parallel multigrid for multilevel Monte Carlo methods. *SIAM Journal on Scientific Computing*, 39(5):S873–S897, 2017.

[14] Y. Efendiev, T. Hou, and W. Luo. Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. *SIAM Journal on Scientific Computing*, 28(2):776–803, 2006.

[15] H. Gahvari, A.H. Baker, M. Schulz, U.M. Yang, K. E. Jordan, and W. Gropp. Modeling the performance of an algebraic multigrid cycle on hpc platforms. In *Proceedings of the international conference on Supercomputing*, pages 172–181. 2011.

[16] R.G. Ghanem and P.D. Spanos. *Stochastic finite elements: a spectral approach*. Courier Corporation, 2003.

[17] M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[18] W.R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC, 1995.

[19] I. G. Graham, R. Scheichl, and E. Ullmann. Mixed finite element analysis of lognormal diffusion and multilevel Monte Carlo methods. *Stochastics and Partial Differential Equations Analysis and Computations*, 4(1):41–75, Mar 2016.

[20] I.G. Graham, F.Y. Kuo, D. Nuyens, R. Scheichl, and I.H. Sloan. Analysis of circulant embedding methods for sampling stationary random fields. *SIAM Journal on Numerical Analysis*, 56(3):1871–1895, 2018.

[21] W Keith Hastings. Monte Carlo sampling methods using markov chains and their applications. 1970.

[22] S. Heinrich. Multilevel Monte Carlo methods. In *Large-scale scientific computing*, pages 58–67. Springer, 2001.

[23] V.E. Henson and U.M. Yang. BoomerAMG: a parallel algebraic multigrid solver and preconditioner. *Applied Numerical Mathematics*, 41:155–177, 2000.

[24] D. Higdon, H. Lee, and Z. Bi. A Bayesian approach to characterizing uncertainty in inverse problems using coarse and fine-scale information. *IEEE Transactions on Signal Processing*, 50(2):389–399, 2002.

[25] V.H. Hoang, J.H. Quek, and C. Schwab. Analysis of multilevel MCMC-fem for Bayesian inversion of log-normal diffusions. *Inverse Problems*, 2019.
[26] V.H. Hoang, C. Schwab, and A.M. Stuart. Complexity analysis of accelerated MCMC methods for Bayesian inversion. *Inverse Problems*, 29(8):085010, 2013.
[27] J.D. Jansen, R.M. Fonseca, S. Kahrobaei, M.M. Siraj, G.M. Van Essen, and P.M.J. Van den Hof. The egg model–a geological ensemble for reservoir simulation. *Geoscience Data Journal*, 1(2):192–195, 2014.
[28] J. Kaipio and E. Somersalo. Statistical inverse problems: discretization, model reduction and inverse crimes. *Journal of computational and applied mathematics*, 198(2):493–504, 2007.
[29] D.Z. Kalchev, C.S. Lee, U. Villa, Y. Efendiev, and P.S. Vassilevski. Upscaling of mixed finite element discretization problems by the spectral AMGe method. *SIAM J. Sci. Comput.*, 38(5):A2912–A2933, 2016.
[30] P.S. Koutsourelakis. A multi-resolution, non-parametric, Bayesian framework for identification of spatially-varying model parameters. *Journal of computational physics*, 228(17):6184–6211, 2009.
[31] I. Lashuk and P.S. Vassilevski. The construction of the coarse de Rham complexes with improved approximation properties. *Comput. Meth. in Appl. Math.*, 14(2):257–303, 2014.
[32] I.V. Lashuk and P.S. Vassilevski. Element agglomeration coarse Raviart–Thomas spaces with improved approximation properties. *Numer. Linear Alg. Appl.*, 19(2):414–426, 2012.
[33] J. Latz, I. Papaioannou, and E. Ullmann. Multilevel sequential^2 Monte Carlo for Bayesian inverse problems. *Journal of Computational Physics*, 368:154–178, 2018.
[34] C.S Lee and P.S. Vassilevski. Parallel solver for $\mathbf{H}$(div) problems using hybridization and AMG. In *Domain Decomposition Methods in Science and Engineering XXIII*, pages 69–80. Springer, 2017.
[35] F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.
[36] J. Martin, L.C. Wilcox, C. Burstedde, and O. Ghattas. A stochastic newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460–A1487, 2012.
[37] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.
[38] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.
[39] S. Osborn, P.S. Vassilevski, and U. Villa. A multilevel, hierarchical sampling technique for spatially correlated random fields. *SIAM J. Sci. Comput.*, 39(5):S543–S562, 2017.
[40] S. Osborn, P. Zulian, T. Benson, U. Villa, R. Krause, and P.S. Vassilevski. Scalable hierarchical PDE sampler for generating spatially correlated random fields using nonmatching meshes. *Numerical Linear Algebra with Applications*, 25(3):e2146, 2018.
[41] J.E. Pasciak and P.S. Vassilevski. Exact de Rham sequences of spaces defined on macro-elements in two and three spatial dimensions. *SIAM J. Sci. Comput.*, 30(5):2427–2446, 2008.
[42] N. Petra, J. Martin, G. Stadler, and O. Ghattas. A computational framework for infinite-dimensional Bayesian inverse problems, part ii: Stochastic newton MCMC with application to ice sheet flow inverse problems. *SIAM Journal on Scientific Computing*, 36(4):A1525–A1555, 2014.
[43] R. Scheichl, A.M. Stuart, and A.L. Teckentrup. Quasi-Monte Carlo and multilevel Monte Carlo methods for computing posterior expectations in elliptic inverse problems. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):493–518, 2017.
[44] C. Schillings and C. Schwab. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Problems*, 29(6):065011, 2013.
[45] A.M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.
[46] A.L. Teckentrup, R. Scheichl, M.B. Giles, and E. Ullmann. Further analysis of multilevel Monte Carlo methods for elliptic pdes with random coefficients. *Numerische Mathematik*, 125(3):569–600, 2013.
[47] E. Ullmann and I. Papaioannou. Multilevel estimation of rare events. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):922–953, 2015.