

# Estimating posterior quantity of interest expectations in a multilevel scalable framework

Hillary R. Fairbanks<sup>1</sup>  | Sarah Osborn<sup>1</sup>  | Panayot S. Vassilevski<sup>1,2</sup>

<sup>1</sup>Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, California, USA

<sup>2</sup>Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland, Oregon, USA

## Correspondence

Hillary Fairbanks, Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, PO Box 808, L-561, Livermore, CA 94551, USA.  
Email: fairbanks5@llnl.gov

## Funding information

U.S. Department of Energy, Grant/Award Numbers: DE-AC52-07NA27344, DE-NA-0002373

## 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

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

## 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, for example, 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.

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 the 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).<sup>1-3</sup> 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 References 4,5. Other methods include multilevel approaches that utilize coarse grid solvers to accelerate mixing, as in References 6,7 or—in addition to accelerating MCMC mixing—utilize a telescoping sum to perform variance reduction (similar to multilevel Monte Carlo<sup>8-11</sup>) resulting in fewer fine grid samples.<sup>12-14</sup>

An alternative approach to target the posterior is to utilize importance sampling as is done in sequential Monte Carlo.<sup>15,16</sup> 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 References 17-19, where the authors build the sequence of distributions from coarse approximations of the posterior, and in Reference 18 they also employ a variance reduction strategy similar to multilevel Monte Carlo.

Of interest in this work is the ratio estimator<sup>20-23</sup> 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 References 22,23 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;<sup>24</sup> 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 generate both the unknown random field realizations (as developed in References 25,26) 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 article, motivated by the theoretical derivations in Reference 23, we extend their work to solving the mixed Darcy equations using the multilevel, hierarchical sampler method from References 25,26, 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 article is organized into five main sections. In Section 2 we discuss Bayes' law and provide the framework for the ratio estimator as in Reference 23. 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 References 25,26, 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 Reference 23. 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  $(\Omega, \mathcal{A}, \mu_0)$ , with sample space  $\Omega$ ,  $\sigma$ -Algebra  $\mathcal{A}$ , and prior measure  $\mu_0$ , 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  $\mathcal{B} : X \rightarrow V$ , which maps the uncertain input  $u \in X$  to a response. Furthermore, the bounded linear *observational functional* is given by  $\mathcal{H} : V \rightarrow Y$ , 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  $\mathcal{G} := \mathcal{H} \circ \mathcal{B} : X \rightarrow Y$ .

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

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

where  $\eta$  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

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

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  $\mu^y$  and the prior measure  $\mu_0$  of the parameter,<sup>27</sup> and is given by

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

which means that for any measurable function  $\psi$ , we have

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

The *normalization* constant  $Z$  above is given by

$$Z := \int_X \exp[-\Phi(\cdot; y)] d\mu_0(u). \quad (5)$$

### 2.1 | Computation of posterior expectations

We seek the expected value of a QoI,  $Q(u) = \mathcal{Q}[\mathcal{B}(u)]$ ,  $\mathcal{Q} : V \rightarrow \mathbb{R}$ , under the posterior distribution  $\mu^y$ . Typically, we do not have a closed-form expression for the posterior distribution  $\mu^y$ , 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:

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

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 Reference 23, which utilizes a derivation of equation (6.24) in Reference 27.

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, that is, MC and its variants as in Reference 23, 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  $\mathbf{q}$  and pressure  $p$  may be determined by solving

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

with Dirichlet boundary condition  $p = p_D$  on  $\Gamma_D$ , enforced by the right-hand side  $\mathbf{f}$ , and Neumann boundary condition  $\mathbf{q} \cdot \mathbf{n} = 0$  on  $\Gamma_N$ , where  $\Gamma_D$  and  $\Gamma_N$  are nonoverlapping partitions of  $\partial D$ . The scalar QoI  $Q$  will depend on  $\mathbf{q}$  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  $(u, v) = \int_D u v d\mathbf{x}$  for all  $u, v \in W$  and  $\mathbf{R} = H(\text{div}; D) := \{\mathbf{q} \in [L^2(D)]^d \mid \text{div } \mathbf{q} \in L^2(D), \mathbf{q} \cdot \mathbf{n} = 0 \text{ on } \partial D\}$  with inner product  $(\mathbf{q}, \mathbf{s}) = \int_D \mathbf{q} \cdot \mathbf{s} d\mathbf{x}$  for all  $\mathbf{q}, \mathbf{s} \in \mathbf{R}$ .

For the discretization scheme, define the discrete space  $\mathbf{R}_h \subset \mathbf{R}$  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  $(\mathbf{q}_h, p_h) \in \mathbf{R}_h \times W_h$  are calculated via the discretized weak formulation

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

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 Reference 28 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

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

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

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

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<sup>29</sup> from hypre<sup>30</sup> 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,<sup>31</sup> circulant embedding,<sup>24</sup> as well as the stochastic PDE approach.<sup>32</sup> 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, for example, Reference 33, and comparisons with the stochastic PDE are a consideration for future work. In this work we have chosen to form the Gaussian random fields via this stochastic PDE method due to its efficient scalability, as well as the availability of tools to form these realizations on unstructured meshes (see, e.g., References 34,35).

## 4 | PRIOR RANDOM FIELD SAMPLING VIA STOCHASTIC PARTIAL DIFFERENTIAL EQUATION

Following work in Reference 32, a stochastic PDE method may be employed to form Gaussian random field realizations based on the exponential covariance function:

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

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

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

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  $\mathbf{R}_h \subset \mathbf{R}$  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  $(\rho_h, u_h) \in \mathbf{R}_h \times W_h$  from the system

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

with essential boundary conditions  $\rho_h \cdot \mathbf{n} = 0$ . 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  $(\rho_h, \mathbf{s}_h)$ ,  $\Lambda_h$  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  $(\text{div } \mathbf{s}_h, u_h)$ . A realization of the right-hand side of (13) may be calculated, as in Reference 32, using the fact that  $\mathbb{E}[(\mathcal{W}, v_i)] = 0$  and  $\text{cov}((\mathcal{W}, v_i), (\mathcal{W}, v_j)) = (v_i, v_j)$  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

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

Our framework allows for different preconditioners to solve this system<sup>35</sup>; and within this framework, we utilize the hybridization AMG method of References 36,37, 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.<sup>29</sup>

**Remark 1** (Artificial boundary effects). As discussed in Reference 32 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 Reference 26. In Reference 26, 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  $\mathbf{R}_\ell$  and  $W_\ell$  on an unstructured triangulation  $\mathcal{T}_\ell$  of size  $h_\ell$ , 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  $\ell$  refer to different levels of mesh widths (characteristic diameter of agglomerated elements) with  $h_0 < h_1 < \dots < h_L$ , for example,  $h_0$  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 AMGe, so that we are able to construct operator-dependent coarse spaces for  $H(\text{div})$  problems with guaranteed approximation properties on general, unstructured grids; see References 38–41 for technical details.

Consider the linear system in (14) where a subscript  $\ell$  indicates mesh size  $h_\ell$ . 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 Reference 25. Define the interpolation operators  $P_u : W_{\ell+1} \rightarrow W_\ell$  and  $P_\rho : \mathbf{R}_{\ell+1} \rightarrow \mathbf{R}_\ell$  constructed using techniques from AMGe.<sup>38–40</sup> First, a level  $\ell$  realization of  $\xi_\ell(\omega^{(i)}) \sim \mathcal{N}(0, I)$  is generated, and second the coarse system is solved

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

using the hybridization and AMG technique as discussed in Section 4.1. Third, to solve for corresponding fine-level  $U_\ell$ , the interpolated coarse solution,

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

is applied as an initial guess fine-level  $\ell$  solver. Work showing that both level  $\ell$  and level  $\ell + 1$  solutions produce Gaussian random fields is found in proposition 3.1 of Reference 25.



## 5 | MONTE CARLO AND MULTILEVEL MONTE CARLO FOR POSTERIOR QOI MEAN ESTIMATION

In this section we briefly review key concepts of MC and MLMC methods<sup>8-11</sup> for estimating the posterior mean  $\mathbb{E}_{\mu^*}[Q(u)] = R/Z$  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  $Q_h(u_h) = Q_h[B_h(u_h)]$  and

$$\Phi_h(u_h; y) = \frac{1}{2\sigma_\eta^2} \|y - \mathcal{H}_h(B_h(u_h))\|^2, \quad (16)$$

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

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

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  $f_h(u_h) : X_h \rightarrow \mathbb{R}$ , where the formulation and notation will directly apply to both  $R_h$  and  $Z_h$ .

### 5.1 | Standard Monte Carlo methods

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

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

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

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

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  $\varepsilon^2$ , 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, for example, variance reduction methods, where by reducing the variance of  $f_h$ , fewer forward simulations are required, thus reducing the overall cost. To this end, 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  $f_L, \dots, f_1$  of the QoI  $f_h$ , where  $f_\ell := f_{h_\ell}$  for  $\ell := L, \dots, 1$  denotes the QoI obtained from the level  $\ell$  solver and  $f_h := f_0$ , then we have

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

where  $Y_\ell = f_\ell - f_{\ell+1}$  for  $\ell = 0, \dots, L-1$ , 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  $Y_\ell$ :

$$\hat{Y}_{\ell, N_\ell} = \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} (f_\ell(u_\ell^{(i)}) - f_{\ell+1}(u_{\ell+1}^{(i)})), \quad (21)$$

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

$$\hat{f}_h^{ML, \mu_0} = \sum_{\ell=0}^L \hat{Y}_{\ell, N_\ell}. \quad (22)$$

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

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

The cost reduction lies in the sampling error of MLMC; compared with the sampling error of MC (in (19)), the MLMC variance associated with the finest-level QoI, level  $\ell = 0$ , is  $\mathbb{V}_{\mu_0}[Y_0]$ , that is, the variance of the difference between  $f_0$  and  $f_1$ . As  $h \rightarrow 0$ ,  $Y_\ell \rightarrow 0$ , 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  $N_\ell$  on finer levels. Based on work in Reference 9 it can be shown that the optimal number of samples to obtain a desired sampling error of  $\varepsilon^2/2$  is given by

$$N_\ell \gtrsim \frac{2}{\varepsilon^2} \left[ \sum_{k=0}^L \sqrt{\mathbb{V}_{\mu_0}[Y_k] C_k} \right] \sqrt{\frac{\mathbb{V}_{\mu_0}[Y_\ell]}{C_\ell}}, \quad (24)$$

where  $C_\ell$  is the cost per sample of  $Y_\ell$ . In practice, the variance for each level correction  $Y_\ell$  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, for example, References 10,11. Let  $M_\ell$  be the spatial degrees of freedom (DOFs) for the mesh on level  $\ell$  (i.e., the number of unknowns in the linear solver). Briefly stated, theorem 1 of Reference 10 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.  $C_\ell \lesssim M_\ell^\gamma$ ,

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

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

and

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

where  $C(\hat{f}_h^{ML, \mu_0})$  is the total cost of performing MLMC to calculate  $\hat{f}_h^{ML, \mu_0}$ . More concisely, given a value of  $\varepsilon$ , and values  $\alpha, \beta$ , and  $\gamma$ , 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  $\alpha, \beta$ , and  $\gamma$ , to determine whether or not a value of  $\varepsilon$  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

$$\hat{Q}_{h,N}^{MC,\mu^y} := \frac{\hat{R}_{h,N}^{MC,\mu_0}}{\hat{Z}_{h,N}^{MC,\mu_0}}, \quad (25)$$

and the MLMC posterior estimator to be

$$\hat{Q}_h^{ML,\mu^y} := \frac{\hat{R}_h^{ML,\mu_0}}{\hat{Z}_h^{ML,\mu_0}}. \quad (26)$$

Using  $\hat{R}_h$  and  $\hat{Z}_h$  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 Reference 23, where the authors show

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

For bounded  $\hat{R}_h/\hat{Z}_h$  and  $Z$  away from zero, this indicates that the MSE of the posterior mean scales as the sum of the MSE of  $\hat{R}_h$  and the MSE of  $\hat{Z}_h$ ; in particular, MC and MLMC theory can be directly applied to the individual errors  $\mathbb{E}[(R - \hat{R}_h)^2]$  and  $\mathbb{E}[(Z - \hat{Z}_h)^2]$  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 Reference 42 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  $N_\ell$ .

## 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 stochastic PDE 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 Reference 43, to test the estimator in a large-scale setting.

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

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

with  $\mathbf{q}$  a solution of (7). To form the ratio estimate, we utilize independent simulations of  $R_\ell$  and  $Z_\ell$ ; both which require us to simulate the forward model to obtain velocity and pressure solutions  $(\mathbf{q}_\ell, p_\ell)$ . A sample of  $Z_\ell$  requires the calculation of the pressure  $p_\ell$  to estimate  $\Phi_\ell$  (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  $R_\ell$  requires, in addition, the calculation of  $Q_\ell$  (a discrete estimation of (28)). Algorithm 1 outlines the different steps taken to form scalable samples of  $R_\ell$  or  $Z_\ell$ .

Numerical simulations—of both the stochastic PDE and forward Darcy solvers—were performed using tools developed in ParELAG,<sup>34</sup> 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 stochastic PDE and forward Darcy problems. ParELAGMC,<sup>35</sup> a parallel element agglomeration MLMC library, was applied to generate all MLMC results, with images generated with GLVis.<sup>44</sup> All timing results were executed on the Quartz cluster at Lawrence Livermore National Laboratory, consisting of 2688 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 Reference 45), 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  $R_\ell$  and  $Z_\ell$  on each level. In step (ii), a realization of  $u_\ell$  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  $k_\ell$  on the four different levels. For step (iii) of Algorithm 1, the Darcy solver is completed with boundary conditions of

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

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 with single-level estimates for  $R_\ell$  and  $Z_\ell$ . 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  $\alpha$ . 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  $N_\ell$  (see (24)). The variance for both these correction terms decays with a slope of  $\beta \approx 0.6$ . These values of  $\alpha$  and  $\beta$  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.

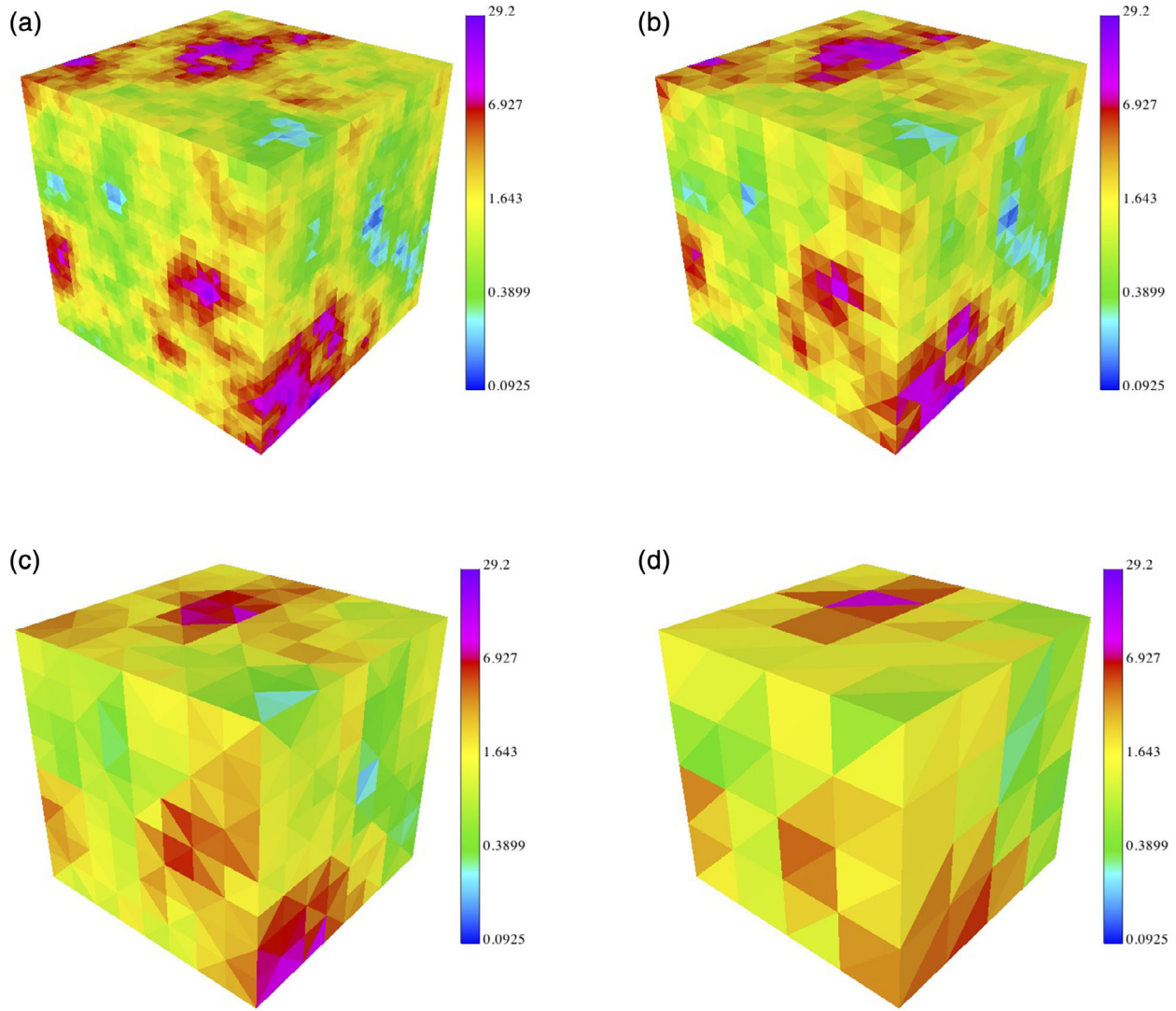
While the previous figure indicates that MLMC may be applied to the estimates of  $R$  and  $Z$ , the same does not necessarily hold for the QoI  $Q$ . Since  $Q$  is the ratio of the  $R$  and  $Z$  estimates, it must be confirmed that the MSE of  $Q$  scales according to the MSE of both  $R$  and  $Z$  estimates. To do this we approximate the discretization error and sampling error

---

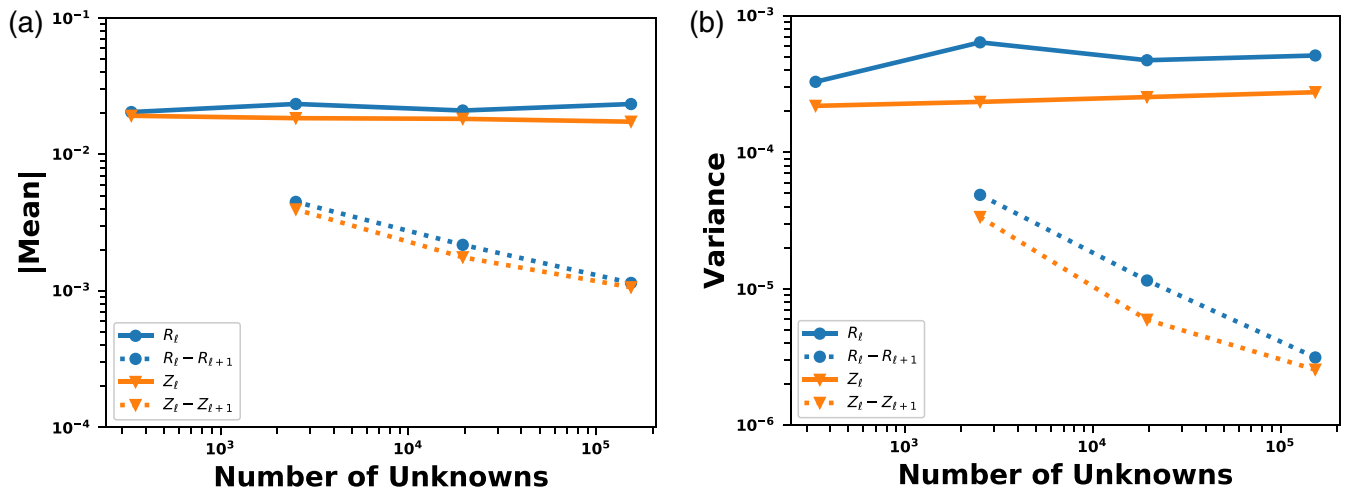
### Algorithm 1. Algorithm for simulating $R_\ell$ or $Z_\ell$

---

- (i) Sample  $\xi_\ell(\omega) \sim \mathcal{N}(0, I)$
  - (ii) Form Gaussian random field realization  $u_\ell$  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  $R_\ell$ , calculate  $Q_\ell$  via (28); otherwise skip
  - (vi) Form a realization of  $R_\ell$  or  $Z_\ell$  via (17)
-



**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  $\ell = 3$



**FIGURE 2** (a) Absolute value of sample mean and (b) sample variance values for  $R_\ell$  and  $Z_\ell$  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

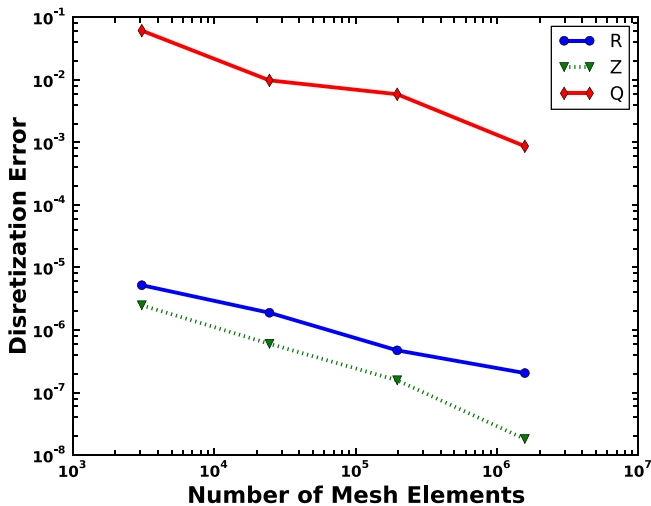
of  $R$ ,  $Z$ , and  $Q$ . Figure 3 displays the decay of  $R_\ell - R_{\ell+1}$ ,  $Z_\ell - Z_{\ell+1}$ , and  $\hat{Q}_\ell - \hat{Q}_{\ell+1}$  estimates using the average of 500 samples. The decay lines indicate that as the discretization error of  $Q$  decays with the error decay of  $R$  and  $Z$ .

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 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  $C_{MLMC} = NC_0$ , resulting in

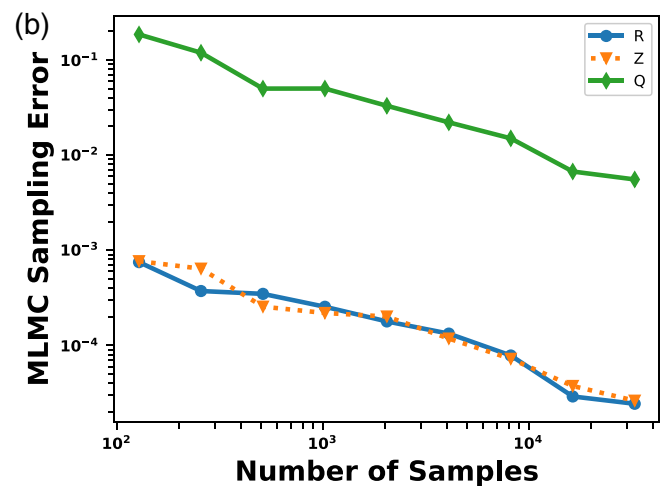
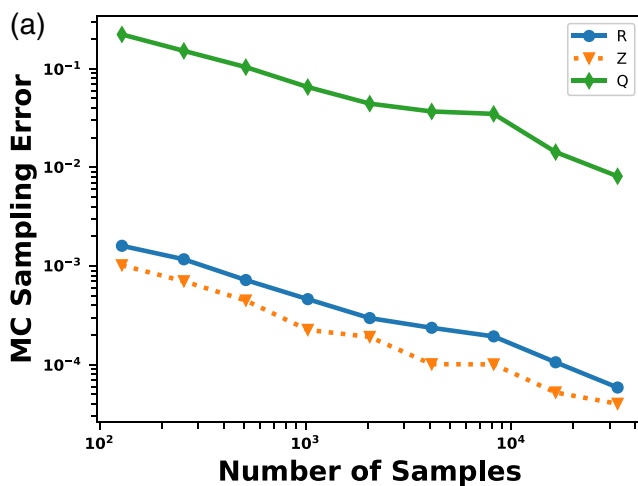
$$\tilde{N}_\ell = \frac{NC_0}{\sum_{k=0}^L \sqrt{V_k C_k}} \sqrt{\frac{V_\ell}{C_\ell}}, \quad (29)$$

on each level. Figure 4(b) provides the sampling error of a two-level MLMC as a function of  $N$ , so that it is comparable with MC.

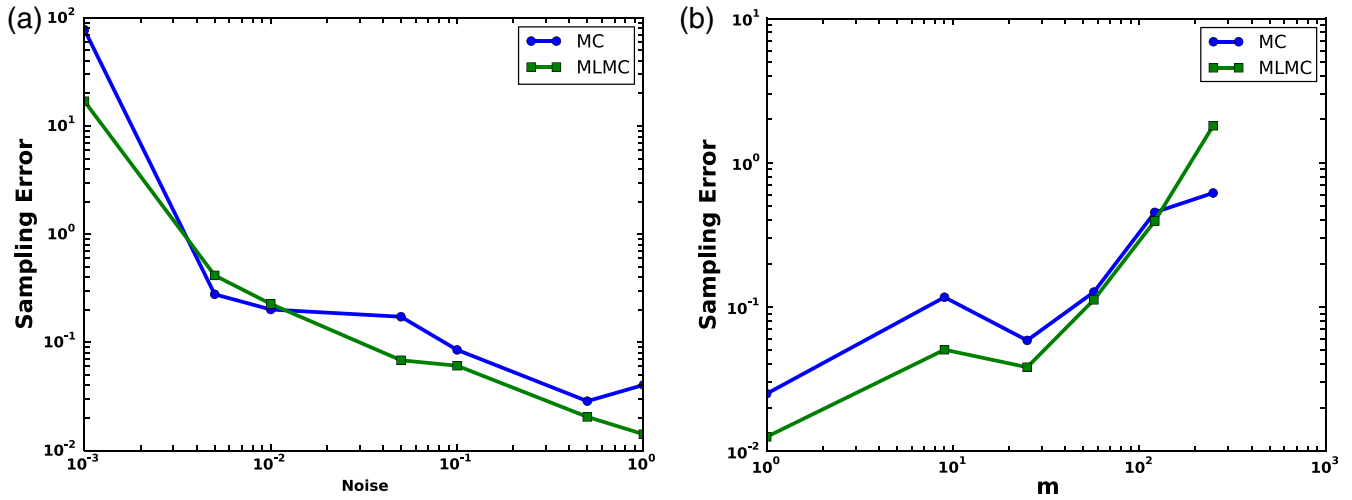
The next set of results compares the effect of noise and number of observations on the sampling error for ratio estimates of MC and MLMC, again treating  $h_2$  as the fine-level mesh size. Figure 5(a) displays the estimated sampling error as a function of the observation noise variance  $\sigma_\eta^2$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



**FIGURE 3** Decay of the discretization error for  $Q_\ell$ ,  $R_\ell$ , and  $Z_\ell$  for  $N = 500$  samples on each level. The discretization error convergence rate of the ratio estimator  $Q_\ell$  is approximately equal to the convergence rates of individual prior estimators  $R_\ell$  and  $Z_\ell$



**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)

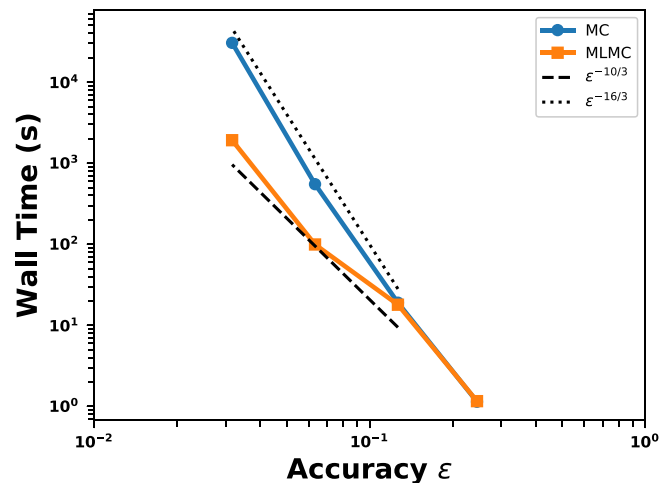


**FIGURE 5** (a) Comparison of MC and MLMC sampling errors, with fixed  $N = 1048$  and  $m = 9$ , as a function of noise parameter (sample average of error for 10 runs). (b) Comparison of MC and MLMC sampling errors, with fixed  $N = 1048$  and  $\sigma_\eta^2 = 0.1$ , as a function of noise parameter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

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.

The final result in Figure 6 compares the total wall time of MC and MLMC for various values of  $\epsilon$ . 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, that is, each data point, was completed on four MPI processors on Quartz. For  $\epsilon = 0.25$ , MLMC requires one level (to meet the discretization error estimate), thus MC and MLMC have the same costs. For each smaller value of  $\epsilon$ , an additional level is incorporated into MLMC so that the smallest value  $\epsilon$  is associated with the four-level MLMC. Furthermore, as the value of  $\epsilon$  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  $\epsilon$  (as discussed in Section 5). These values come from the  $\beta < \gamma$  scenario, where the MLMC cost increases as  $\epsilon^{-2-(\gamma-\beta)/\alpha}$ . 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  $\epsilon^{-10/3}$ . For single-level MC, the cost grows as  $\epsilon^{-2-\gamma/\alpha} = \epsilon^{-16/3}$ .



**FIGURE 6** Computational cost of performing MC versus MLMC for various values of the root MSE accuracy  $\epsilon$ . Both MC and MLMC runs completed on four processors



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

## 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.”<sup>43</sup> We consider the performance of computing the MLMC ratio estimator using a weak scaling perspective, that is, increasing the fine-level problem size proportionally to the processor count, for three different MLMC computations.

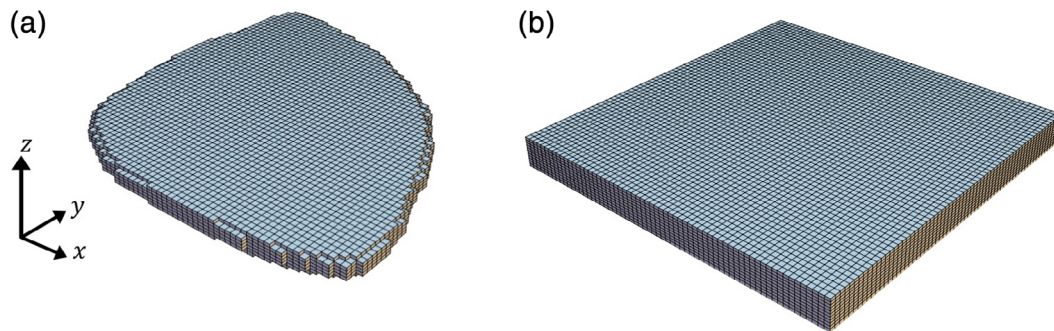
The Egg model is a 3D synthetic reservoir model containing an ensemble of synthetic realizations that contain channels of high permeability to represent subsurface river patterns. With regard to the discretization, the irregular Egg domain is contained by a  $480 \text{ m} \times 480 \text{ m} \times 28 \text{ m}$  bounding box. For the stochastic PDE solver (step (ii) of Algorithm 1), the Egg model domain is embedded within a  $512 \text{ m} \times 512 \text{ m} \times 44 \text{ m}$  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  $8 \text{ m} \times 8 \text{ m} \times 4 \text{ m}$ . For finer levels, the mesh of Figure 7 is uniformly refined by a factor of two in each direction.

To generate realizations of  $R_\ell$  and  $Z_\ell$ , we following Algorithm 1. In step (ii), realizations of  $u_\ell$  are formed with correlation length  $\lambda = 10 \text{ m}$ . Dissimilar to the unit cube example, after calculating  $u_\ell$  in step (ii), the permeability field is formed using data from Reference 43. 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  $k_\ell^{-1}$  on three different levels. These images represent solutions calculated on the coarsest mesh, as 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 with boundary conditions of

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

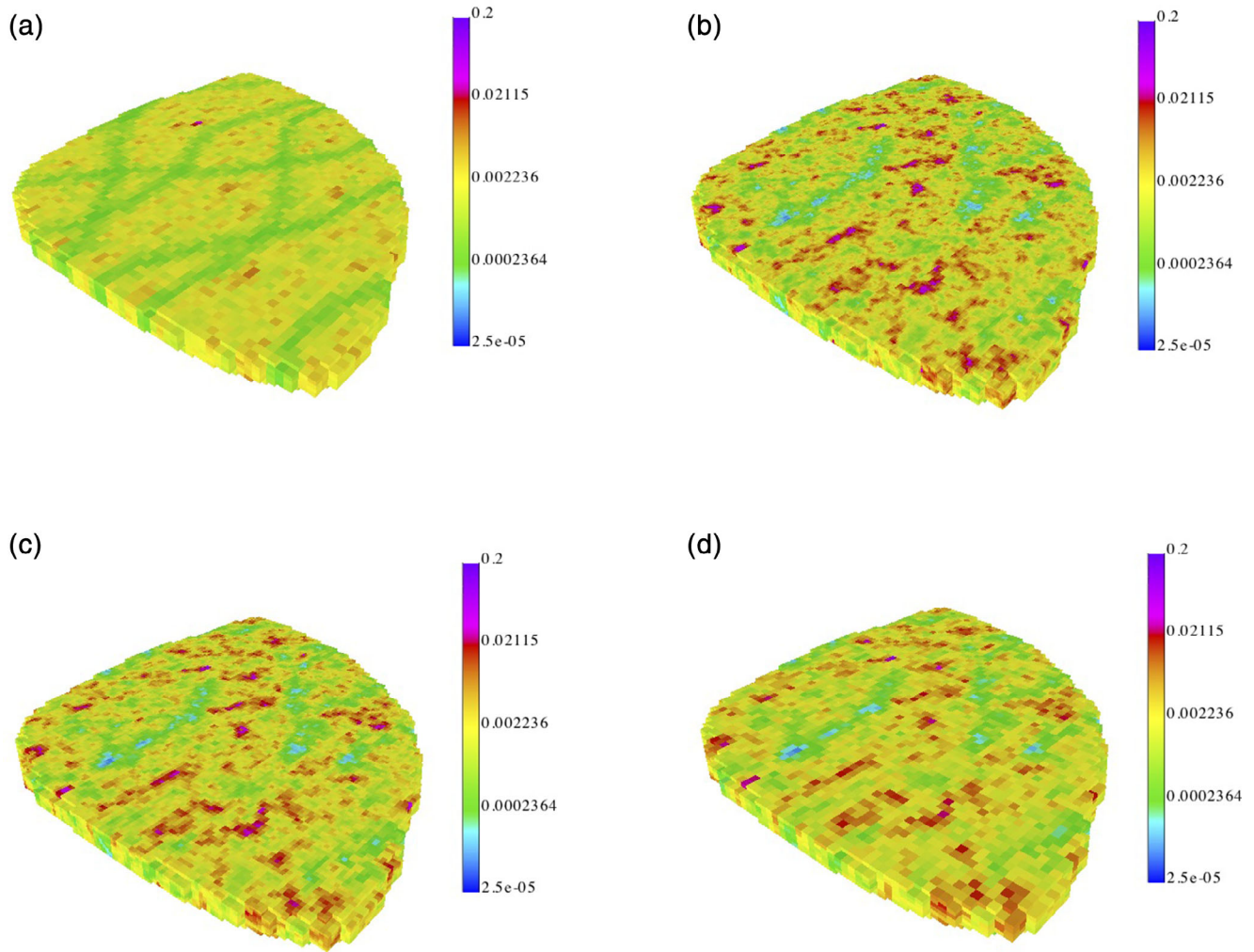
The reference observational data—localized average pressure—are extracted about  $m = 12$  locations in the domain at  $z = 12$  and  $(x, y)$  coordinates provided in Table 1. We consider observational noise with  $\sigma_\eta^2 = 1.0$ .

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  $7.6\text{e}+07$  elements (corresponding to the largest problem in the weak scaling results below), with five levels using 2304 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  $\varepsilon$  for the remaining results.



**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





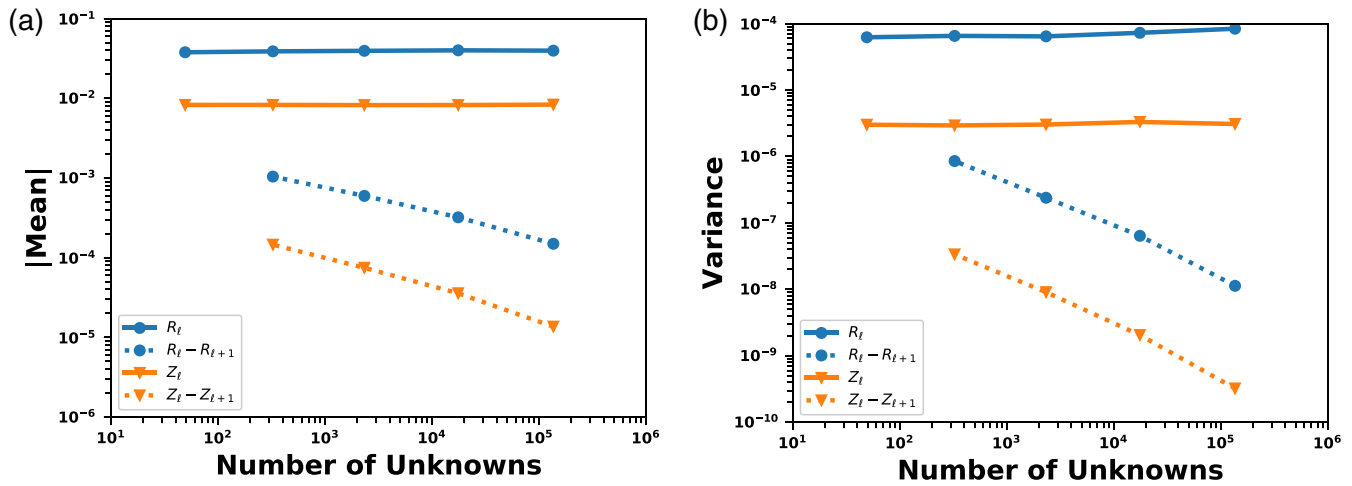
**FIGURE 8** (a) Inverse of permeability field 18 from Reference 43. (b)–(d) A single realization of the coefficient  $k_e^{-1}$  on levels  $\ell = 0, 1$ , and 2 on the smallest problem with 36 processors

**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)

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  $1.39\text{e}+05$  DOFs, on the fine level. Then the target MSE value  $\epsilon^2$  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  $\text{NP} = 36$ , 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.

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,



**FIGURE 9** (a) Absolute value of sample mean and (b) sample variance values for  $R_\ell$  and  $Z_\ell$  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

**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	2817	61	470	1746	28,199	
2304	3.0488e+08	2.50e-06	16,824	111	813	2764	6892	102,931

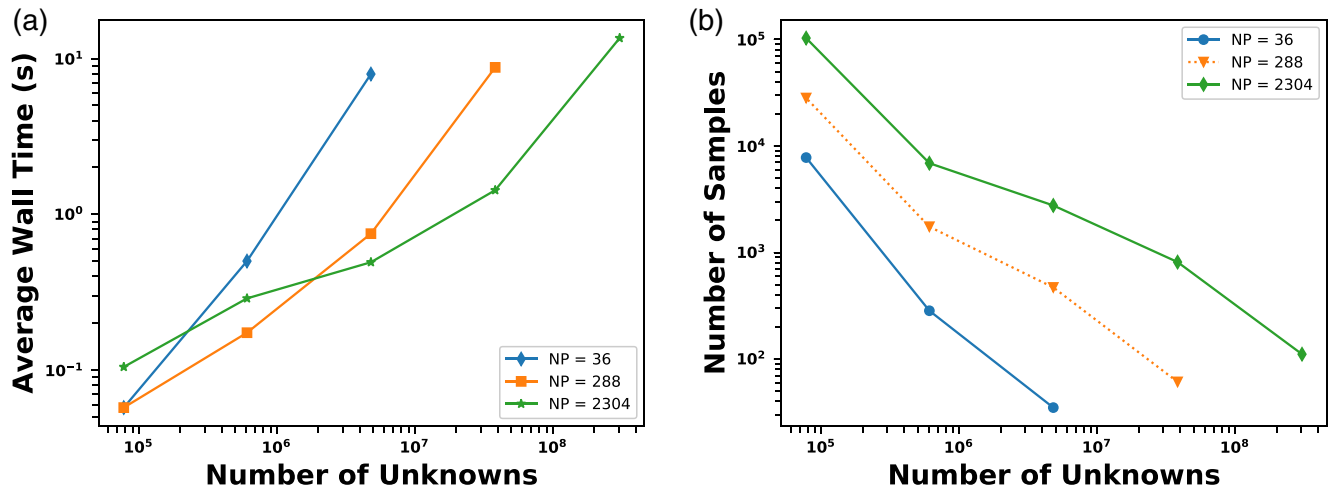
NP	$\ell = 0$	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$
36	7.95	0.501	0.0576		
288	8.77	0.752	0.174	0.0576	
2304	13.5	1.43	0.493	0.289	0.105

Note: Levels  $\ell = 0, 1, 2, 3$  provide weak scaling results, where each column maintains the same number of DOFs per processor.

**TABLE 3** Average wall time (s) per simulation, separated into the three problem sizes (NP), on each level

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, e.g., 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 Reference 46. 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. Reference 47).

Figure 10(a) displays the average computational time per sample (averaged over the combined cost for R and Z) compared with the number of global unknowns in the forward problem. For optimal scaling, this cost will grow linearly, and

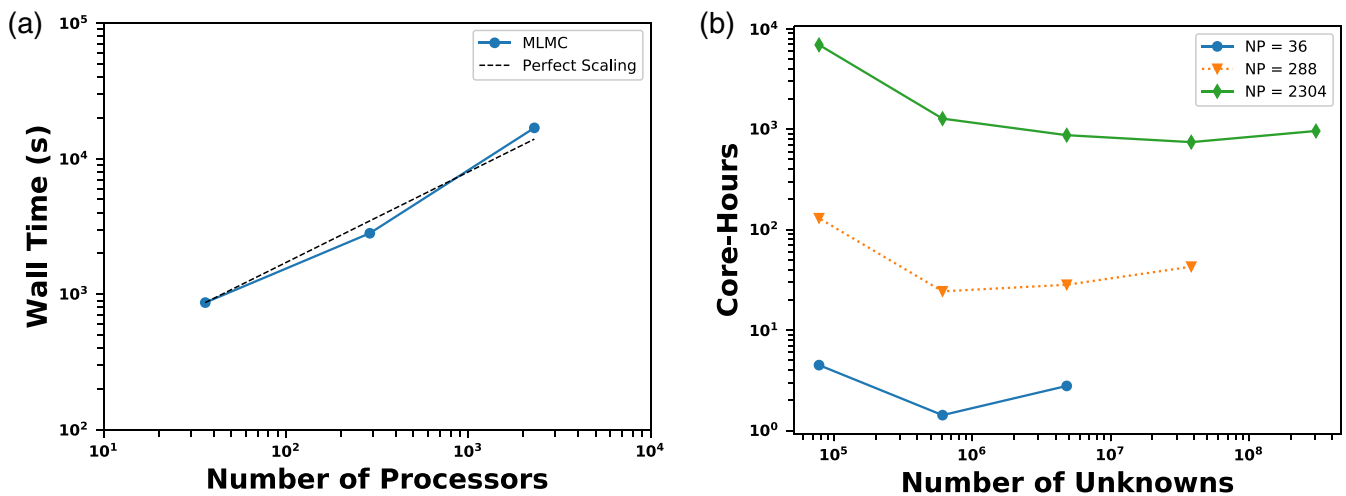


**FIGURE 10** (a) Average wall time per forward simulation on each level, for each of the three problem sizes. (b) Total number of simulations  $N_\ell$  (for combined  $R$  and  $Z$ ) on each level

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 regard to the adaptive MLMC results, Figure 10(b) provides the number of samples  $N_\ell$  (for combined  $R$  and  $Z$ ) for each level. For a fixed number of processors, the number of samples increases for larger  $\ell$ , 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  $\varepsilon^2$ , and thus will need more samples to achieve the associated target sampling error.

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 with the  $\varepsilon$ -adjusted perfect scaling. Here the  $\varepsilon$ -adjusted cost growth takes into account that as we increase the problem size, we are also decreasing  $\varepsilon$ ; 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



**FIGURE 11** (a) Total wall time for the full MLMC algorithm across the different problem sizes. Here perfect scaling takes the change in  $\varepsilon$ —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

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., Reference 46). 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 Reference 47.

## 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 References 23,25,26 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,<sup>43</sup> 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 increased 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., Reference 47).

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 Reference 23). 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 Reference 13 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.

## ACKNOWLEDGMENTS

This work was initiated while Hillary R. Fairbanks was an intern at Lawrence Livermore National Laboratory funded partially by the U.S. Department of Energy's National Nuclear Security Administration under the Predictive Science Academic Alliance Program II at Stanford University, Grant DE-NA-0002373. 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). This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, 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. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

## CONFLICTS OF INTEREST

Panayot Vassilevski is an editor of the journal.

## ORCID

Hillary R. Fairbanks  <https://orcid.org/0000-0003-3021-0298>

Sarah Osborn  <https://orcid.org/0000-0003-3256-3337>

## REFERENCES

- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys*. 1953;21(6):1087–92.
- Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*. 1970;57(1):97–109.
- Gilks WR, Richardson S, Spiegelhalter D. Markov chain Monte Carlo in practice. Hoboken, NJ: Chapman & Hall/CRC Press; 1995.
- Martin J, Wilcox LC, Burstedde C, Ghattas O. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM J Sci Comput*. 2012;34(3):A1460–87.
- Petra N, Martin J, Stadler G, Ghattas O. A computational framework for infinite-dimensional Bayesian inverse problems, Part II: stochastic Newton MCMC with application to ice sheet flow inverse problems. *SIAM J Sci Comput*. 2014;36(4):A1525–55.
- Higdon D, Lee H, Bi Z. A Bayesian approach to characterizing uncertainty in inverse problems using coarse and fine-scale information. *IEEE Trans Signal Process*. 2002;50(2):389–99.
- Efendiev Y, Hou T, Luo W. Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. *SIAM J Sci Comput*. 2006;28(2):776–803.
- Heinrich S. Multilevel Monte Carlo methods. Large-scale scientific computing. New York, NY: Springer; 2001. p. 58–67.
- Giles MB. Multilevel Monte Carlo path simulation. *Oper Res*. 2008;56(3):607–17.
- Cliffe KA, Giles MB, Scheichl R, Teckentrup AL. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Comput Vis Sci*. 2011;14(1):3–15.
- Teckentrup AL, Scheichl R, Giles MB, Ullmann E. Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients. *Numer Math*. 2013;125(3):569–600.
- Hoang VH, Schwab C, Stuart AM. Complexity analysis of accelerated MCMC methods for Bayesian inversion. *Inverse Probl*. 2013;29(8):085010.
- Dodwell TJ, Ketelsen C, Scheichl R, Teckentrup AL. A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow. *SIAM/ASA J Uncertain Quantif*. 2015;3(1):1075–108.
- Hoang VH, Quek JH, Schwab C. Analysis of a multilevel Markov chain Monte Carlo finite element method for Bayesian inversion of log-normal diffusions. *Inverse Probl*. 2020;36(3):035021. <https://dx.doi.org/10.1088/1361-6420/ab2a1e>.
- Doucet A, Freitas ND, Gordon N. An introduction to sequential Monte Carlo methods. *Sequential Monte Carlo methods in practice*. New York, NY: Springer; 2001. p. 3–14.
- Moral PD, Doucet A, Jasra A. Sequential Monte Carlo samplers. *J R Stat Soc Ser B Stat Methodol*. 2006;68(3):411–36.
- Koutsourelakis PS. A multi-resolution, non-parametric, Bayesian framework for identification of spatially-varying model parameters. *J Comput Phys*. 2009;228(17):6184–211.
- Beskos A, Jasra A, Law K, Tempone R, Zhou Y. Multilevel sequential Monte Carlo samplers. *Stoch Process Appl*. 2017;127(5):1417–40.
- Latz J, Papaioannou I, Ullmann E. Multilevel sequential<sup>2</sup> Monte Carlo for Bayesian inverse problems. *J Comput Phys*. 2018;368:154–78.
- Schillings C, Schwab C. Sparse, adaptive Smolyak quadratures for Bayesian inverse problems. *Inverse Probl*. 2013;29(6):065011.
- Dick J, Gantner RN, Gia QTL, Schwab C. Higher order quasi-Monte Carlo integration for Bayesian estimation; 2016. arXiv preprint arXiv:160207363.
- Dick J, Gantner RN, Gia QTL, C Schwab C. Multilevel higher-order quasi-Monte Carlo Bayesian estimation. *Math Models Methods Appl Sci*. 2017;27(05):953–95.
- Scheichl R, Stuart AM, Teckentrup AL. Quasi-Monte Carlo and multilevel Monte Carlo methods for computing posterior expectations in elliptic inverse problems. *SIAM/ASA J Uncertain Quantif*. 2017;5(1):493–518.
- Graham IG, Kuo FY, Nuyens D, Scheichl R, Sloan IH. Analysis of circulant embedding methods for sampling stationary random fields. *SIAM J Numer Anal*. 2018;56(3):1871–95.
- Osborn S, Vassilevski PS, Villa U. A multilevel, hierarchical sampling technique for spatially correlated random fields. *SIAM J Sci Comput*. 2017;39(5):S543–62.
- Osborn S, Zulian P, Benson T, Villa U, Krause R, Vassilevski PS. Scalable hierarchical PDE sampler for generating spatially correlated random fields using nonmatching meshes. *Numer Linear Algebra Appl*. 2018;25(3):e2146.
- Stuart AM. Inverse problems: a Bayesian perspective. *Acta Numerica*. 2010;19:451–559.
- Graham IG, Scheichl R, Ullmann E. Mixed finite element analysis of lognormal diffusion and multilevel Monte Carlo methods. *Stoch Partial Differ Equ*. 2016;4(1):41–75.
- Henson VE, Yang UM. BoomerAMG: a parallel algebraic multigrid solver and preconditioner. *Appl Numer Math*. 2000;41:155–77.
- HYPRE: high performance pre-conditioners; <http://www.llnl.gov/CASC/hypre/>.
- Ghanem RG, Spanos PD. Stochastic finite elements: a spectral approach. New York, NY: Springer; 1991.
- Lindgren F, Rue H, Lindström J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J Royal Stat Soc Ser B Stat Methodol*. 2011;73(4):423–98.
- Cai D, Chow E, Erlandson L, Saad Y, Xi Y. SMASH: structured matrix approximation by separation and hierarchy. *Numer Linear Algebra Appl*. 2018;25(6):e2204.
- ParELAG: element-agglomeration algebraic multi-grid and upscaling library, version 2.0; 2015. <http://github.com/LLNL/parelag>.
- ParELAGMC: parallel element agglomeration multilevel Monte Carlo library; 2018. <http://github.com/LLNL/parelagmc>.
- Lee CS, Vassilevski PS. Parallel solver for  $\mathbf{H}(\text{div})$  problems using hybridization and AMG. *Domain decomposition methods in science and engineeringXXIII*. Cham: Springer; 2017. p. 69–80.



37. Dobrev V, Kolev T, Lee CS, Tomov V, Vassilevski PS. Algebraic hybridization and static condensation with application to scalable H(div) preconditioning. *SIAM J Sci Comput.* 2019;41(3):B425–47.
38. Lashuk I, Vassilevski PS. The construction of the coarse de Rham complexes with improved approximation properties. *Comput Meth in Appl Math.* 2014;14(2):257–303.
39. Lashuk IV, Vassilevski PS. Element agglomeration coarse Raviart–Thomas spaces with improved approximation properties. *Numer Linear Alg Appl.* 2012;19(2):414–26.
40. Pasciak JE, Vassilevski PS. Exact de Rham sequences of spaces defined on macro-elements in two and three spatial dimensions. *SIAM J Sci Comput.* 2008;30(5):2427–46.
41. Kalchev DZ, Lee CS, Villa U, Efendiev Y, Vassilevski PS. Upscaling of mixed finite element discretization problems by the spectral AMGe method. *SIAM J Sci Comput.* 2016;38(5):A2912–33.
42. Ullmann E, Papaioannou I. Multilevel estimation of rare events. *SIAM/ASA J Uncertain Quantif.* 2015;3(1):922–53.
43. Jansen JD, Fonseca RM, Kahrobaei S, Siraj MM, Essen GMV, den Hof PMJV. The egg model—a geological ensemble for reservoir simulation. *Geosci Data J.* 2014;1(2):192–5.
44. GLVis: OpenGL finite element visualization tool; [glvis.org](http://glvis.org).
45. Kaipio J, Somersalo E. Statistical inverse problems: discretization, model reduction and inverse crimes. *J Comput Appl Math.* 2007;198(2):493–504.
46. Gahvari H, Baker AH, Schulz M, Yang UM, Jordan KE, Gropp W. Modeling the performance of an algebraic multigrid cycle on HPC platforms. *Proceedings of the International Conference on Supercomputing*, Tucson, AZ; 2011. p. 172–181.
47. Drzisga D, Gmeiner B, Rude U, Scheichl R, Wohlmuth B. Scheduling massively parallel multigrid for multilevel Monte Carlo methods. *SIAM J Sci Comput.* 2017;39(5):S873–97.

**How to cite this article:** Fairbanks HR, Osborn S, Vassilevski PS. Estimating posterior quantity of interest expectations in a multilevel scalable framework. *Numer Linear Algebra Appl.* 2021;28:e2352. <https://doi.org/10.1002/nla.2352>