Multifidelity Quasi-Monte Carlo Methods

Nicole Aretz, Michael B. Giles, Frances Y. Kuo, Dirk Nuyens, Karen Willcox

August 2026

Abstract

This paper introduces the *multifidelity Quasi-Monte Carlo* (MFQMC) method and develops its theoretical foundation. Multifidelity methods leverage the correlation of a fixed high-fidelity model with cheaper low-fidelity models to improve the accuracy (respectively, cost) of estimating high-fidelity statistics, compared to single-fidelity approaches for the same computational budget (respectively, target accuracy). The MFQMC method replaces the sample points in a multifidelity Monte Carlo (MFMC) estimator (Ng and Willcox, 2014) by *randomly shifted embedded lattice rules* (Cools, Kuo, and Nuyens, 2006) to obtain an unbiased estimator of the high-fidelity expectation. For a root mean-square error demand of $\varepsilon > 0$, the MFQMC estimator has a sampling cost of nearly $\mathcal{O}(\varepsilon^{-1})$, which is provably better than the corresponding cost of $\mathcal{O}(\varepsilon^{-2})$ for MFMC. The paper proposes a model selection algorithm using pilot samples to choose a combination of models which minimizes the theoretical error bound and total cost, based on the theoretically derived optimal sample size for each model in the multifidelity ensemble. The MFQMC algorithm then adaptively increases the number of samples for the chosen models until a desired error or budget threshold is met. We compare the performance of the MC, QMC, MFMC, and MFQMC methods in estimating the expected ten-year change in the ice volume above floatation of Pine Island Glacier under uncertainties in the basal friction field.

1 Introduction

In this paper we develop a theoretical foundation for *multifidelity Quasi-Monte Carlo* (MFQMC) methods. We formulate an MFQMC algorithm and demonstrate its success in reducing the computational cost of uncertainty quantification (UQ) for the problem of estimating the expected change in ice volume above floatation of the Pine Island Glacier in West Antarctica over a ten-year period.

For many scientific and engineering problems, quantifying uncertainty is of critical practical importance yet the cost of high-fidelity simulations is prohibitively high for UQ methods, which typically require hundreds to thousands of simulations. Low-fidelity models, for example, computed with coarser meshes or using simplified physics, are cheaper but can be inaccurate. *Multifidelity Monte Carlo* (MFMC) methods [50, 52, 53, 54] combine an ensemble of models with varying fidelities in an estimator of the high-fidelity expectation that has, in comparison to using one single high-fidelity model, reduced computational cost to achieve a given error threshold, or reduced error for a given computational budget. MFMC has been shown to achieve orders of magnitude speedup in UQ tasks for a broad

[1]Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, USA (nicole.aretz@austin.utexas.edu, kwillcox@utexas.edu);
Mathematical Institute, University of Oxford, UK (mike.giles@maths.ox.ac.uk);
School of Mathematics and Statistics, UNSW Sydney, Australia (f.kuo@unsw.edu.au);
Department of Computer Science, KU Leuven, Belgium (dirk.nuyens@kuleuven.be).

range of scientific and engineering problems, such as aerospace applications [14, 50, 51], Earth and climate systems [2, 26, 36], and plasma micro-turbulence analysis [39]. MFMC also provides an underpinning for multifidelity approaches to active subspace and covariance estimation [45, 46, 47].

MFMC shares some common features with *multilevel Monte Carlo* (MLMC) methods [20, 21], which have also been used to accelerate UQ in many application areas, such as random PDEs [6, 9, 58], dynamical systems [1, 29], and engineering and finance [4, 13, 35]. It is important to recognize that MFMC differs from MLMC both philosophically and algorithmically in a number of key ways that lead to different theoretical results and change the frame of reference in which the computational gains of each method are assessed. Firstly, the *multifidelity models are not necessarily nested nor ordered* in an obvious way; they are driven by what models or codes are readily available. On the other hand, the levels in MLMC typically arise from a single model using a carefully chosen sequence of refinement in resolution, e.g., increasing dimensionality or reducing finite element meshwidth. Secondly, *in MFMC the sample points are deliberately nested and reused*, whereas in MLMC the sample points are typically independent across levels. This leads to a different structure in the estimators and their mean-square errors. Thirdly, in MLMC the ground truth is usually the limit of a converging sequence and there is a truncation error referred to as the bias, whereas in MFMC the *highest fidelity model is considered to be the ground truth with no bias*. MLMC analysis then usually targets the convergence rate of the cost-versus-error balance relative to standard MC on a sequence of finer levels, while MFMC analysis targets the improvement in the implied constant relative to standard MC on the given highest fidelity model.

Other related methods include multi-index Monte Carlo (MIMC) methods [28] which can be described as applying multilevel methods to several refinement aspects at the same time, yielding a similar structure to sparse grid techniques [5]; and also approximate control variates (ACV) [24] and multilevel best linear unbiased estimators (MLBLUE) [10, 57] which, like MLMC and MFMC, follow the control variate viewpoint to find the best scaling of estimators.

With standard (single-level, single fidelity) MC methods, to achieve a root mean-square accuracy of ε requires $N = \mathcal{O}(\varepsilon^{-2})$ samples, and so the cost is proportional to ε^{-2}. However, the total cost is also proportional to the cost of a single sample on that level. When the limit of a converging sequence is the ground truth (i.e., the multilevel perspective), reducing the overall error also requires switching to a finer level of resolution which increases the cost of each sample. As a result, the total cost is often $\mathcal{O}(\varepsilon^{-3})$ or worse. In this setting, under favorable conditions MLMC restores the $\mathcal{O}(\varepsilon^{-2})$ total cost with most samples being computed with $\mathcal{O}(1)$ cost, independent of the target accuracy ε. Similarly, MFMC (with the highest fidelity model being the ground truth) has the same $\mathcal{O}(\varepsilon^{-2})$ total cost with most of the samples being obtained using the cheapest lowest-fidelity model. In this paper, we go beyond these MLMC and MFMC improvements to achieve a provably better convergence rate: the main contribution of this paper is to provide theory for *multifidelity Quasi-Monte Carlo* (MFQMC) methods, to yield a provably better rate close to $\mathcal{O}(\varepsilon^{-1})$.

Quasi-Monte Carlo (QMC) methods [11, 12] are designed to be better than MC for high-dimensional integration and approximation problems. QMC has been applied successfully, with full theoretical justification, to compute expected values of quantities of interest involving parametric PDEs in uncertainty quantification (UQ). Recent applications of QMC in UQ include, e.g., tumor growth modeling [16], wave scattering [25], optimal control [27], random domains [30]; see [41] for a survey of earlier theoretical works. Beyond computing expected values, QMC has also been used for function approximation, including, e.g., probability density estimation [18], kernel interpolation [37], and

deep neural network training [38]. QMC has long been used to improve the convergence rates of multilevel methods in mathematical finance [23] and for parametric PDEs, e.g., [15, 19, 43, 44]. QMC has also been combined with multivariate decomposition methods (MDM) [17, 42], which are also cost-saving methods and are particularly suited for infinite-dimensional problems requiring efficient truncation strategies [48, 49]. Control variates with QMC have been considered in [3, 33]. Randomized QMC is often used in practice to ensure that the estimators are unbiased and also to allow for practical error estimation. Common QMC randomization strategies include shifting, digital-shifting, and scrambling [31].

In this paper, we put the powerful theory of QMC's improved convergence rates together with the practical computational gains of multifidelity methods, providing theory and an algorithm for MFQMC. Due to the need for sample reuse, an essential property we will need in MFQMC is the embedding property of a good QMC sequence, namely, that every initial block of the sequence is of good quality. In this paper we will make use of *randomly shifted embedded lattice rules* [7] constructed to have this good property. While this good embedding property may hold for many other QMC sequences, lattice rules have another essential property that is needed for our theory in this paper: the lattice points form a group under addition modulo the integers, meaning that the sum or difference of any two lattice points is another lattice point. This is the essence of our key result in Lemma 3.1 below. An analogous result can be obtained for randomly digitally-shifted digital nets [11]. Sample reuse with QMC has been considered previously in [55], but without theoretical justification.

The structure of the paper is as follows. In §2 we provide background on randomized QMC and review key results for MFMC, setting the scene to present our theory for MFQMC in §3. Then in §4 we carry out numerical experiments, first for a simple test problem and then for a simulation of Pine Island Glacier. Finally in §5 we give concluding remarks.

2 Background

We consider the integral over the s-dimensional unit cube

$$I(f) := \int_{[0,1]^s} f(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y},$$

where the dimensionality s is large. This can be equivalently formulated over the translated unit cube $[-\frac{1}{2}, \frac{1}{2}]^s$, or any translated and rescaled bounded domain. We use $\boldsymbol{y} = (y_1, \ldots, y_s) \in [0,1]^s$ as the integration variable over the parametric domain, reserving $\boldsymbol{x}$ for the spatial variable for the PDEs. An integral over the unbounded domain $\mathbb{R}^s$ can be mapped into the unit cube with an appropriate change of variables.

2.1 Monte Carlo method

A simple Monte Carlo (MC) estimator for $I(f)$ is

$$Q^{\mathrm{MC}}(f) = \frac{1}{N} \sum_{i=0}^{N-1} f(\boldsymbol{t}_i), \tag{2.1}$$

with i.i.d. uniform random samples $\{\boldsymbol{t}_i : i = 0, \ldots, N-1\} \subset [0,1]^s$. The MC estimator is unbiased, i.e., $\mathbb{E}[Q^{\mathrm{MC}}(f)] = I(f)$, and its mean-square error (or variance) satisfies

$$\mathrm{mse}(Q^{\mathrm{MC}}(f)) = \mathbb{V}[Q^{\mathrm{MC}}(f)] = \mathbb{E}[\,|Q^{\mathrm{MC}}(f) - I(f)|^2\,] = \frac{\sigma^2(f)}{N}, \tag{2.2}$$

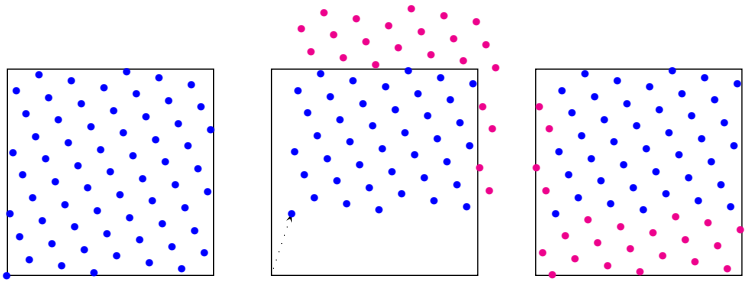

Figure 1: Applying a $(0.1, 0.3)$-shift to a 64-point lattice rule in two dimensions. Left: original lattice. Middle: moving all points by $(0.1, 0.3)$. Right: wrapping the points back inside the unit cube. Figure from [11, Figure 2.3].

where $\sigma^2(f) := I(f^2) - [I(f)]^2$ is the variance of the function f. In practice, an estimator for the mean-square error is given by

$$\operatorname{mse}(Q^{\mathrm{MC}}(f)) \approx \frac{1}{N\,(N-1)} \sum_{i=0}^{N-1} \big(f(\boldsymbol{t}_i) - Q^{\mathrm{MC}}(f)\big)^2. \tag{2.3}$$

2.2 Randomized Quasi-Monte Carlo method

Quasi-Monte Carlo (QMC) rules take the same form as (2.1) but the points are deterministically chosen to be "better" than random. A *shifted* QMC rule [11, 12] that estimates $I(f)$ takes the form

$$Q^{\mathrm{QMC}}(f) = Q^{\mathrm{QMC}}(f, \boldsymbol{\Delta}) = \frac{1}{N} \sum_{i=0}^{N-1} f(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta} \rrbracket), \tag{2.4}$$

where $\boldsymbol{\Delta} \in [0,1]^s$ is the shift, and we use double brackets $\llbracket \cdot \rrbracket$ around a vector to indicate that we take the fractional part of each component in the vector, or more precisely, the double brackets mean the operation "modulo 1". For example, $\llbracket (1.7, -0.4) \rrbracket = (0.7, 0.6)$. Figure 1 illustrates the process of shifting a QMC rule with $N = 64$ points in $s = 2$ dimensions.

QMC error is often expressed in terms of its *worst case error* in a normed function space $\mathcal{H}$. For the shifted QMC rule (2.4) this is defined by

$$e^{\mathrm{wor}}_{\mathcal{H}}(\{\boldsymbol{t}_i\}_{i=0}^{N-1}, \boldsymbol{\Delta}) := \sup_{f \in \mathcal{H},\, \|f\|_{\mathcal{H}} \le 1} |Q^{\mathrm{QMC}}(f, \boldsymbol{\Delta}) - I(f)|.$$

Due to linearity, it follows that for any integrand $f \in \mathcal{H}$ we have

$$|Q^{\mathrm{QMC}}(f, \boldsymbol{\Delta}) - I(f)| \le e^{\mathrm{wor}}_{\mathcal{H}}(\{\boldsymbol{t}_i\}_{i=0}^{N-1}, \boldsymbol{\Delta})\, \|f\|_{\mathcal{H}}. \tag{2.5}$$

In other words, the QMC error is bounded by its worst case error (depending on the points) times the norm of the integrand. This generalizes the classical Koksma–Hlawka inequality for low-discrepancy sequences. Explicit expressions for the squared worst case error are known for reproducing kernel Hilbert spaces. The convergence rate for the worst case error will depend on the pairing of a particular QMC rule with an appropriate function space. See [11] for a detailed survey on QMC theory.

A *randomly shifted* QMC estimator, in which $\boldsymbol{\Delta}$ is uniformly distributed in $[0,1]^s$, is unbiased, since

$$\mathbb{E}_{\boldsymbol{\Delta}}[Q^{\mathrm{QMC}}(f,\boldsymbol{\Delta})] = \frac{1}{N}\sum_{i=0}^{N-1}\int_{[0,1]^s} f(\llbracket \boldsymbol{t}_i+\boldsymbol{\Delta}\rrbracket)\,\mathrm{d}\boldsymbol{\Delta} = \frac{1}{N}\sum_{i=0}^{N-1}\int_{[0,1]^s} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = I(f),$$

where the subscript in $\mathbb{E}_{\boldsymbol{\Delta}}$ highlights that the expected value is taken with respect to the random shift $\boldsymbol{\Delta}\in[0,1]^s$ rather than the deterministic points $\{\boldsymbol{t}_i\}$. Later we will use the notations $\mathbb{V}_{\boldsymbol{\Delta}}$ and $\mathrm{Cov}_{\boldsymbol{\Delta}}$ analogously.

For completeness, we show how to do the change of variables involving $\llbracket\cdot\rrbracket$ in one dimension. We substitute $y=t+\Delta$ for $\Delta<1-t$ and $y=t+\Delta-1$ for $\Delta\ge 1-t$, to obtain

$$\begin{aligned}\int_0^1 f(\llbracket t+\Delta\rrbracket)\,\mathrm{d}\Delta &= \int_0^{1-t} f(t+\Delta)\,\mathrm{d}\Delta + \int_{1-t}^1 f(t+\Delta-1)\,\mathrm{d}\Delta\\ &= \int_t^1 f(y)\,\mathrm{d}y + \int_0^t f(y)\,\mathrm{d}y = I(f).\end{aligned} \tag{2.6}$$

The mean-square error (or variance) of a randomly shifted QMC rule satisfies

$$\begin{aligned}\mathrm{mse}(Q^{\mathrm{QMC}}(f)) &= \mathbb{E}_{\boldsymbol{\Delta}}[\,|Q^{\mathrm{QMC}}(f,\boldsymbol{\Delta})-I(f)|^2\,]\\ &\le \mathbb{E}_{\boldsymbol{\Delta}}[\,e_{\mathcal{H}}^{\mathrm{wor}}(\{\boldsymbol{t}_i\}_{i=0}^{N-1},\boldsymbol{\Delta})^2\,\|f\|_{\mathcal{H}}^2\,] = e_{\mathcal{H}}^{\mathrm{sh}}(\{\boldsymbol{t}_i\}_{i=0}^{N-1})^2\,\|f\|_{\mathcal{H}}^2 \le \frac{c_\mu^2\,\|f\|_{\mathcal{H}}^2}{N^{2\mu}},\end{aligned} \tag{2.7}$$

where we made use of (2.5) and the *"shift-averaged" worst case error*

$$e_{\mathcal{H}}^{\mathrm{sh}}(\{\boldsymbol{t}_i\}_{i=0}^{N-1}) := \sqrt{\mathbb{E}_{\boldsymbol{\Delta}}[\,e_{\mathcal{H}}^{\mathrm{wor}}(\{\boldsymbol{t}_i\}_{i=0}^{N-1},\boldsymbol{\Delta})^2\,]} \le \frac{c_\mu}{N^\mu}.$$

For the purposes of this paper it suffices to say that we know how to construct "good" *randomly shifted lattice rules* (see the next subsection) in *first-order mixed Sobolev space* such that the shift-averaged worst case error is bounded of the form $\frac{c_\mu}{N^\mu}$, with the convergence rate $\mu\approx 1$. Here the "constant" $c_\mu^2\,\|f\|_{\mathcal{H}}^2$ can be bounded independently of the dimension s, with appropriately chosen weight parameters for the function space, provided that f has the appropriate ordering of the variables in decaying importance, see [11]. This is a clear improvement over MC.

A deterministic QMC rule is biased. QMC with a single random shift provides an unbiased estimator of the integral and a superior convergence rate to MC, but a single random shift does not allow for practical error estimation. In practice, we use a small number of independent random shifts $\boldsymbol{\Delta}_1,\ldots,\boldsymbol{\Delta}_R$ and take the average

$$Q_R^{\mathrm{QMC}}(f) := \frac{1}{R}\sum_{k=1}^R Q^{\mathrm{QMC}}(f,\boldsymbol{\Delta}_k)$$

as the final approximation to $I(f)$. The corresponding mean-square error satisfies

$$\mathrm{mse}(Q_R^{\mathrm{QMC}}(f)) \le \frac{1}{R}\,\frac{c_\mu^2\,\|f\|_{\mathcal{H}}^2}{N^{2\mu}}.$$

A practical estimate for the mean-square error is given by

$$\mathrm{mse}(Q_R^{\mathrm{QMC}}(f)) \approx \frac{1}{R\,(R-1)}\sum_{k=1}^R \big(Q^{\mathrm{QMC}}(f,\boldsymbol{\Delta}_k)-Q_R^{\mathrm{QMC}}(f)\big)^2. \tag{2.8}$$

Since we have a higher convergence rate with respect to N than with respect to R, we keep R fixed (e.g., $R = 16$) and increase N as appropriate in practice.

As a fair comparison between MC and RQMC in terms of the number of function evaluations, we should take in practice $N^{\mathrm{MC}} = R\,N^{\mathrm{QMC}}$.

For simplicity, in the following theoretical discussions we consider first $R = 1$.

2.3 Embedded lattice rules

The points of a (rank-1) *lattice rule* with N points are given by

$$\boldsymbol{t}_i = \left[\!\!\left[\frac{i\,\boldsymbol{z}}{N}\right]\!\!\right] = \frac{i\boldsymbol{z}}{N} \bmod 1, \qquad i = 0, \ldots, N-1, \tag{2.9}$$

where $\boldsymbol{z} \in \mathbb{Z}^s$ is the *generating vector* which determines the quality of a lattice rule. Typically, a choice of $\boldsymbol{z}$ is "good" only for a fixed value of N.

In an *embedded lattice rule* [7, 32, 34], a single generating vector $\boldsymbol{z}$ is designed to be "good" for a range of values of $N = 2^m$ for $m_{\min} \le m \le m_{\max}$. The limits $m_{\min}$ and $m_{\max}$ can be specified according to our needs in practice. In such a scenario, the points are treated as a sequence ordered according to

$$\boldsymbol{t}_i = [\![\phi_{\mathrm{rad}}(i)\,\boldsymbol{z}]\!], \qquad i = 0, 1, 2, \ldots,$$

where the *radical inverse function* $\phi_{\mathrm{rad}}(i)$ is defined as follows: if $i = (\cdots i_2 i_1 i_0)_2$ is the binary representation of the index i, then $\phi_{\mathrm{rad}}(i) = (0.i_0 i_1 i_2 \cdots)_2$ flips the bits around the binary point.

If we only want to work with complete lattice rules with $N_\ell = 2^{m_\ell}$ points (they do not need to be consecutive powers of 2), then we do not need the radical inverse function to generate the points. Suppose we already have the N points of the lattice rule from (2.9). To double the number of points, we can keep the existing N points and then generate N new points using the formula (2.9) with N replaced by $2N$ and take only odd indices i.

2.4 Multifidelity Monte Carlo method

Suppose now that f_0 is the ground-truth model. We assume that there are surrogate models $f_1, f_2, \ldots$ of possibly decreasing fidelity and cost (the ordering is desired but not essential). *Multifidelity Monte Carlo* (MFMC) methods can be viewed as iterated control variates to estimate $I(f_0)$

$$Q^{\mathrm{MFMC}}(f_0) := \frac{1}{N_0}\sum_{i=0}^{N_0-1} f_0(\boldsymbol{t}_i) + \sum_{\ell=1}^{L} \alpha_\ell \left(\frac{1}{N_\ell}\sum_{i=0}^{N_\ell-1} f_\ell(\boldsymbol{t}_i) - \frac{1}{N_{\ell-1}}\sum_{i=0}^{N_{\ell-1}-1} f_\ell(\boldsymbol{t}_i)\right), \tag{2.10}$$

where for $L \ge 1$ we have real coefficients $\{\alpha_\ell\}_{\ell=1}^{L}$, *one single sequence of points* $\{\boldsymbol{t}_i\}_{i\ge 0}$ in all terms, and a strictly increasing sequence $1 \le N_0 < N_1 < N_2 < \cdots$.

The MFMC estimator (2.10) is unbiased, since

$$\begin{aligned}\mathbb{E}[Q^{\mathrm{MFMC}}(f_0)] &= \int_{[0,1]^s} \cdots \int_{[0,1]^s} Q^{\mathrm{MFMC}}(f_0)\,\mathrm{d}\boldsymbol{t}_0 \cdots \mathrm{d}\boldsymbol{t}_{N_L-1} \\ &= \frac{1}{N_0}\sum_{i=0}^{N_0-1} I(f_0) + \sum_{\ell=1}^{L} \alpha_\ell \left(\frac{1}{N_\ell}\sum_{i=0}^{N_\ell-1} I(f_\ell) - \frac{1}{N_{\ell-1}}\sum_{i=0}^{N_{\ell-1}-1} I(f_\ell)\right) = I(f_0).\end{aligned}$$

The following theorem states the mean-square error (or variance) which requires accounting for the sample reuse across the levels; it is originally from [53, Lemma 3.3].

Theorem 2.1 (MFMC variance reduction) *The mean-square error or variance of the MFMC estimator* (2.10) *with* $1 \le N_0 < N_1 < N_2 < \cdots$ *is*

$$\begin{aligned}\operatorname{mse}(Q^{\mathrm{MFMC}}(f_0)) &= \mathbb{V}[Q^{\mathrm{MFMC}}(f_0)] \\ &= \frac{\sigma^2(f_0)}{N_0} + \sum_{\ell=1}^{L} \Big(\frac{1}{N_{\ell-1}} - \frac{1}{N_\ell}\Big)\big(\alpha_\ell^2\,\sigma^2(f_\ell) - 2\,\alpha_\ell\,\sigma(f_\ell, f_0)\big), \end{aligned} \tag{2.11}$$

which is minimized by taking

$$\alpha_\ell^* = \frac{\sigma(f_\ell, f_0)}{\sigma^2(f_\ell)} \quad \textit{for all } \ell = 1, \dots, L, \tag{2.12}$$

giving the minimum

$$\operatorname{mse}(Q_*^{\mathrm{MFMC}}(f_0)) = \frac{\sigma^2(f_0)}{N_0} - \sum_{\ell=1}^{L} \Big(\frac{1}{N_{\ell-1}} - \frac{1}{N_\ell}\Big)\frac{\sigma^2(f_\ell, f_0)}{\sigma^2(f_\ell)}, \tag{2.13}$$

where $\sigma^2(f_\ell) := I(f_\ell^2) - I(f_\ell)^2$ *is the variance of* f_ℓ*, and* $\sigma(f_\ell, f_0) := I(f_\ell\, f_0) - I(f_\ell)\, I(f_0)$ *is the covariance between* f_ℓ *and* f_0*.*

To pave the way for our later MFQMC analysis, we provide a complete proof of Theorem 2.1. The proof relies on the following magic lemma on MC sample reuse, which originally appeared in [53, Lemma 3.2].

Lemma 2.2 (MC sample reuse) *Let* $F, G \in L_2([0,1]^s)$*. For a single sequence of independent uniform random points* $\{\boldsymbol{t}_i\}_{i\ge 0} \subset [0,1]^s$*, let* $Q_N(F) := \frac{1}{N}\sum_{i=0}^{N-1} F(\boldsymbol{t}_i)$*. Then for all* $N, N' \ge 1$*, with* $N_{\max} := \max(N, N')$ *and* $N_{\min} := \min(N, N')$*, we have*

$$\operatorname{Cov}[Q_N(F), Q_{N'}(G)] = \frac{\sigma(F,G)}{N_{\max}}, \quad \sigma(F,G) := I(F\,G) - I(F)\,I(G).$$

The key observation is that the covariance depends only on $N_{\max}$*, i.e.,*

$$\operatorname{Cov}[Q_N(F), Q_{N'}(G)] = \operatorname{Cov}[Q_{N_{\max}}(F), Q_{N_{\max}}(G)]. \tag{2.14}$$

Proof. We have

$$\begin{aligned}\operatorname{Cov}[Q_N(F), Q_{N'}(G)] &= \frac{1}{N}\frac{1}{N'}\sum_{i=0}^{N-1}\sum_{i'=0}^{N'-1} \operatorname{Cov}[F(\boldsymbol{t}_i), G(\boldsymbol{t}_{i'})] \\ &= \frac{1}{N}\frac{1}{N'}\sum_{i=0}^{N_{\min}-1} \big(\mathbb{E}[F(\boldsymbol{t}_i)\,G(\boldsymbol{t}_i)] - I(F)\,I(G)\big) = \frac{\sigma(F,G)}{N_{\max}},\end{aligned}$$

where the $i \ne i'$ terms vanish because the points $\boldsymbol{t}_i$ and $\boldsymbol{t}_{i'}$ are independent so $\mathbb{E}[F(\boldsymbol{t}_i)\,G(\boldsymbol{t}_{i'})] = I(F)\,I(G)$. □

Since each coefficient α_ℓ scales f_ℓ in (2.10), we can write MFMC succinctly as

$$Q^{\mathrm{MFMC}}(f_0) = \sum_{\ell=0}^{L} \big(Q_{N_\ell}(g_\ell) - Q_{N_{\ell-1}}(g_\ell)\big), \quad g_\ell := \alpha_\ell\, f_\ell, \tag{2.15}$$

with $\alpha_0 := 1$, $Q_{N_{-1}} := 0$, $N_{-1} := 0$ and $1/N_{-1} := 0$.

Proof of Theorem 2.1. The mean-square error of (2.15) can be written as

$$\begin{aligned}\operatorname{mse}(Q^{\mathrm{MFMC}}(f_0)) &= \sum_{\ell=0}^{L}\sum_{\ell'=0}^{L}\operatorname{Cov}[Q_{N_\ell}(g_\ell) - Q_{N_{\ell-1}}(g_\ell), Q_{N_{\ell'}}(g_{\ell'}) - Q_{N_{\ell'-1}}(g_{\ell'})]\\ &= \sum_{\ell=0}^{L}\mathbb{V}[Q_{N_\ell}(g_\ell) - Q_{N_{\ell-1}}(g_\ell)]\\ &\quad + 2\sum_{\ell'=0}^{L}\sum_{\ell=\ell'+1}^{L}\Big(\operatorname{Cov}[Q_{N_\ell}(g_\ell), Q_{N_{\ell'}}(g_{\ell'})] - \operatorname{Cov}[Q_{N_\ell}(g_\ell), Q_{N_{\ell'-1}}(g_{\ell'})]\\ &\qquad - \operatorname{Cov}[Q_{N_{\ell-1}}(g_\ell), Q_{N_{\ell'}}(g_{\ell'})] + \operatorname{Cov}[Q_{N_{\ell-1}}(g_\ell), Q_{N_{\ell'-1}}(g_{\ell'})]\Big). \qquad (2.16)\end{aligned}$$

Applying (2.14) to (2.16), we conclude that all $\ell' \neq 0$ terms in the double sum cancel out, giving

$$\begin{aligned}\operatorname{mse}(Q^{\mathrm{MFMC}}(f_0)) &= \mathbb{V}[Q_{N_0}(g_0)] + \sum_{\ell=1}^{L}\mathbb{V}[Q_{N_\ell}(g_\ell) - Q_{N_{\ell-1}}(g_\ell)]\\ &\quad + 2\sum_{\ell=1}^{L}\Big(\operatorname{Cov}[Q_{N_\ell}(g_\ell), Q_{N_0}(g_0)] - \operatorname{Cov}[Q_{N_{\ell-1}}(g_\ell), Q_{N_0}(g_0)]\Big), \qquad (2.17)\end{aligned}$$

where we can use (2.14) again to write for $\ell \geq 1$,

$$\begin{aligned}\mathbb{V}[Q_{N_\ell}(g_\ell) - Q_{N_{\ell-1}}(g_\ell)] &= \mathbb{V}[Q_{N_\ell}(g_\ell)] - 2\operatorname{Cov}[Q_{N_\ell}(g_\ell), Q_{N_{\ell-1}}(g_\ell)] + \mathbb{V}[Q_{N_{\ell-1}}(g_\ell)]\\ &= \mathbb{V}[Q_{N_{\ell-1}}(g_\ell)] - \mathbb{V}[Q_{N_\ell}(g_\ell)]. \qquad (2.18)\end{aligned}$$

Finally we apply (2.2) to (2.17)–(2.18) to obtain

$$\begin{aligned}\operatorname{mse}(Q^{\mathrm{MFMC}}(f_0)) &= \frac{\sigma^2(g_0)}{N_0} + \sum_{\ell=1}^{L}\Big(\frac{\sigma^2(g_\ell)}{N_{\ell-1}} - \frac{\sigma^2(g_\ell)}{N_\ell}\Big) + 2\sum_{\ell=1}^{L}\Big(\frac{\sigma(g_\ell, g_0)}{N_\ell} - \frac{\sigma(g_\ell, g_0)}{N_{\ell-1}}\Big)\\ &= \frac{\sigma^2(g_0)}{N_0} + \sum_{\ell=1}^{L}\Big(\frac{1}{N_{\ell-1}} - \frac{1}{N_\ell}\Big)\big(\sigma^2(g_\ell) - 2\,\sigma(g_\ell, g_0)\big).\end{aligned}$$

Substituting $g_0 = f_0$ and $g_\ell = \alpha_\ell f_\ell$ yields (2.11). The optimal choice α_ℓ^* follows easily from (2.11), since $\alpha^2 A - 2\alpha B$ is minimized at $\alpha = B/A$ when $A > 0$. □

The mean-square error of the MFMC estimator (2.10) can be estimated using the sample variance if $N_{\ell+1} - N_\ell \geq 2$ for $\ell = 0, \ldots, L-1$. To see this, we regroup the sums in (2.10) in terms of independent samples $\boldsymbol{t}_i$ as

$$Q^{\mathrm{MFMC}}(f_0) = \sum_{\ell=0}^{L}\sum_{i=N_{\ell-1}}^{N_\ell - 1} h_\ell(\boldsymbol{t}_i), \quad h_\ell(\boldsymbol{t}) := \frac{g_\ell(\boldsymbol{t})}{N_\ell} + \sum_{k=\ell+1}^{L}\Big(\frac{1}{N_k} - \frac{1}{N_{k-1}}\Big)g_k(\boldsymbol{t}),$$

with $g_\ell := \alpha_\ell\, f_\ell$ and $\alpha_0 := 1$. Then we may use an unbiased estimate for each $\mathbb{V}[h_\ell]$ as follows

$$\begin{aligned}\operatorname{mse}(Q^{\mathrm{MFMC}}(f_0)) &= \sum_{\ell=0}^{L}(N_\ell - N_{\ell-1})\mathbb{V}[h_\ell]\\ &\approx \sum_{\ell=0}^{L}\frac{N_\ell - N_{\ell-1}}{N_\ell - N_{\ell-1} - 1}\sum_{i=N_{\ell-1}}^{N_\ell-1}\Bigg(h_\ell(\boldsymbol{t}_i) - \underbrace{\frac{1}{N_\ell - N_{\ell-1}}\sum_{i'=N_{\ell-1}}^{N_\ell-1} h_\ell(\boldsymbol{t}_{i'})}_{\approx \mathbb{E}[h_\ell]}\Bigg)^2. \qquad (2.19)\end{aligned}$$

3 Multifidelity QMC method

Consider a single QMC sequence $\{\boldsymbol{t}_i\}_{i\ge0}$ and one random shift $\boldsymbol{\Delta}$. With $g_\ell := \alpha_\ell\, f_\ell$, $\alpha_0 := 1$, and $1 \le N_0 < N_1 < N_2 < \cdots$, we define the MFQMC method by

$$\begin{aligned} Q^{\mathrm{MFQMC}}(f_0) &:= \frac{1}{N_0}\sum_{i=0}^{N_0-1} g_0(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}\rrbracket) \\ &\quad + \sum_{\ell=1}^{L}\Big(\frac{1}{N_\ell}\sum_{i=0}^{N_\ell-1} g_\ell(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}\rrbracket) - \frac{1}{N_{\ell-1}}\sum_{i=0}^{N_{\ell-1}-1} g_\ell(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}\rrbracket)\Big) \\ &=: \sum_{\ell=0}^{L}\big(Q_{N_\ell}(g_\ell) - Q_{N_{\ell-1}}(g_\ell)\big), \end{aligned} \tag{3.1}$$

where we now define $Q_N(F) := \frac{1}{N}\sum_{i=0}^{N-1} F(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}\rrbracket)$ for any $F \in L_2([0,1]^s)$ and $N \ge 1$, with $Q_{N_{-1}} := 0$, $N_{-1} := 0$ and $1/N_{-1} := 0$. The succinct MFQMC expression (3.1) is deliberately written to match (2.15) for MFMC, but now the randomness arises from the shift $\boldsymbol{\Delta}$ and not the points $\{\boldsymbol{t}_i\}$.

Recall that a randomly shifted lattice rule is unbiased, see (2.6). The MFQMC estimator (3.1) is therefore unbiased, with

$$\mathbb{E}_{\boldsymbol{\Delta}}[Q^{\mathrm{MFQMC}}(f_0)] = \int_{[0,1]^s} Q^{\mathrm{MFQMC}}(f_0)\,\mathrm{d}\boldsymbol{\Delta} = I(g_0) + \sum_{\ell=1}^{L}\big(I(g_\ell) - I(g_\ell)\big) = I(g_0).$$

3.1 Variance of MFQMC

For any QMC sequence $\{\boldsymbol{t}_i\}_{i\ge0}$, the mean-square error (or variance) of the MFQMC estimator (3.1) with respect to the random shift $\boldsymbol{\Delta}$ can also be expanded in the form (2.16). Interestingly, if we use *randomly shifted embedded lattice rules* in MFQMC, then we have the following magic lemma for sample reuse, analogous to Lemma 2.2 for MFMC. Hence we can express the mean-square error of MFQMC also in the form (2.17)–(2.18).

Lemma 3.1 (QMC sample reuse) *Let $F, G \in L^2([0,1]^s)$. For a randomly shifted embedded lattice sequence with points $\{\boldsymbol{t}_i\}_{i\ge0} \subset [0,1]^s$ (where the initial 2^m points for every $m \ge 0$ form a complete lattice rule), let $Q_N(F) := \frac{1}{N}\sum_{i=0}^{N-1} F(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}\rrbracket)$. Then for all $N = 2^m$, $N' = 2^{m'}$ with $m, m' \ge 0$ and $N_{\max} := \max(N, N')$, we have*

$$\mathrm{Cov}_{\boldsymbol{\Delta}}[Q_N(F), Q_{N'}(G)] = \frac{1}{N_{\max}}\sum_{i=0}^{N_{\max}-1} (F \star G)(\boldsymbol{t}_i) - I(F)\, I(G), \tag{3.2}$$

where we defined a correlation operator

$$(F \star G)(\boldsymbol{x}) := \int_{[0,1]^s} F(\llbracket \boldsymbol{x} + \boldsymbol{y}\rrbracket)\, G(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}.$$

Since $I(F \star G) = I(F)\, I(G)$, the covariance (3.2) is the cubature error for $F \star G$.

The key observation is that the covariance depends only on $N_{\max}$, i.e.,

$$\mathrm{Cov}_{\boldsymbol{\Delta}}[Q_N(F), Q_{N'}(G)] = \mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_{\max}}(F), Q_{N_{\max}}(G)]. \tag{3.3}$$

Proof. We have

$$\begin{aligned}\mathrm{Cov}_{\boldsymbol{\Delta}}[Q_N(F), Q_{N'}(G)] &= \frac{1}{N}\frac{1}{N'}\sum_{i=0}^{N-1}\sum_{i'=0}^{N'-1}\mathrm{Cov}_{\boldsymbol{\Delta}}[F(\llbracket \boldsymbol{t}_i+\boldsymbol{\Delta}\rrbracket), G(\llbracket \boldsymbol{t}_{i'}+\boldsymbol{\Delta}\rrbracket)]\\ &= \frac{1}{N}\frac{1}{N'}\sum_{i=0}^{N-1}\sum_{i'=0}^{N'-1}\int_{[0,1]^s} F(\llbracket \boldsymbol{t}_i+\boldsymbol{\Delta}\rrbracket)\, G(\llbracket \boldsymbol{t}_{i'}+\boldsymbol{\Delta}\rrbracket)\,\mathrm{d}\boldsymbol{\Delta} - I(F)\, I(G).\end{aligned}$$

Using a change of variables, we can write

$$\int_{[0,1]^s} F(\llbracket \boldsymbol{t}_i+\boldsymbol{\Delta}\rrbracket)\, G(\llbracket \boldsymbol{t}_{i'}+\boldsymbol{\Delta}\rrbracket)\,\mathrm{d}\boldsymbol{\Delta} = \int_{[0,1]^s} F(\llbracket \boldsymbol{t}_i-\boldsymbol{t}_{i'}+\boldsymbol{y}\rrbracket)\, G(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}.$$

In one dimension, this works as follows

$$\begin{aligned}&\int_0^1 F(\llbracket t_i+\Delta\rrbracket)\, G(\llbracket t_{i'}+\Delta\rrbracket)\,\mathrm{d}\Delta\\ &= \int_0^{1-t_{i'}} F(\llbracket t_i+\Delta\rrbracket)\, G(t_{i'}+\Delta)\,\mathrm{d}\Delta + \int_{1-t_{i'}}^1 F(\llbracket t_i+\Delta\rrbracket)\, G(t_{i'}+\Delta-1)\,\mathrm{d}\Delta\\ &= \int_{t_{i'}}^1 F(\llbracket t_i+y-t_{i'}\rrbracket)\, G(y)\,\mathrm{d}y + \int_0^{t_{i'}} F(\llbracket t_i+y+1-t_{i'}\rrbracket)\, G(y)\,\mathrm{d}y\\ &= \int_0^1 F(\llbracket t_i-t_{i'}+y\rrbracket)\, G(y)\,\mathrm{d}y.\end{aligned}$$

Thus, with our definition of the correlation operator $F\star G$, we can write

$$\mathrm{Cov}_{\boldsymbol{\Delta}}[Q_N(F), Q_{N'}(G)] = \frac{1}{N}\frac{1}{N'}\sum_{i=0}^{N-1}\sum_{i'=0}^{N'-1}(F\star G)(\llbracket \boldsymbol{t}_i-\boldsymbol{t}_{i'}\rrbracket) - I(F)\, I(G). \tag{3.4}$$

The significance is that the terms in (3.4) now depend on the difference $\llbracket \boldsymbol{t}_i-\boldsymbol{t}_{i'}\rrbracket$ rather than on the two points separately.

Assume without loss of generality that $N'\le N$. Then for any fixed index $i'<N'$, we claim that the following two point sets are the same:

$$\begin{aligned}P_1 &:= \{\,\llbracket \boldsymbol{t}_i-\boldsymbol{t}_{i'}\rrbracket \,:\, i=0,\dots,N-1\},\\ P_2 &:= \{\quad \boldsymbol{t}_i \qquad\;\; :\, i=0,\dots,N-1\}.\end{aligned}$$

Note that P_2 is the lattice point set with N points, while P_1 is obtained by shifting (modulo 1) all points in P_2 by the lattice point $\boldsymbol{t}_{i'}$. Since $\boldsymbol{t}_{i'}\in P_2$, the group property of lattice rules ensures that $P_1=P_2$. Hence we can reduce the double sum in (3.4) into a single sum as follows:

$$\frac{1}{N}\frac{1}{N'}\sum_{i=0}^{N-1}\sum_{i'=0}^{N'-1}(F\star G)(\llbracket \boldsymbol{t}_i-\boldsymbol{t}_{i'}\rrbracket) = \frac{1}{N}\frac{1}{N'}\sum_{i'=0}^{N'-1}\sum_{i=0}^{N-1}(F\star G)(\boldsymbol{t}_i) = \frac{1}{N}\sum_{i=0}^{N-1}(F\star G)(\boldsymbol{t}_i),$$

which yields (3.2). We have also

$$\begin{aligned}I(F\star G) &= \int_{[0,1]^s}\int_{[0,1]^s} F(\llbracket \boldsymbol{x}+\boldsymbol{y}\rrbracket)\, G(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{x}\\ &= \int_{[0,1]^s}\int_{[0,1]^s} F(\boldsymbol{x}')\,\mathrm{d}\boldsymbol{x}'\, G(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = I(F)\, I(G).\end{aligned}$$

Thus (3.2) is the cubature error for the integrand $F\star G$. □

While the covariance (3.2) is explicitly the cubature error for the integrand $F \star G$, rather than working to establish a norm bound on $F \star G$, it is easier to obtain a bound on the covariance in terms of the original functions F and G. Using (2.7) and (3.3), we can bound a general covariance term as follows:

$$\begin{aligned}
|\operatorname{Cov}_{\boldsymbol{\Delta}}[Q_N(F), Q_{N'}(G)]| &= |\operatorname{Cov}_{\boldsymbol{\Delta}}[Q_{N_{\max}}(F), Q_{N_{\max}}(G)]| \\
&\le \sqrt{\mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_{\max}}(F)]}\,\sqrt{\mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_{\max}}(G)]} \\
&= \sqrt{\mathbb{E}_{\boldsymbol{\Delta}}[\,|Q_{N_{\max}}(F) - I(F)|^2\,]}\,\sqrt{\mathbb{E}_{\boldsymbol{\Delta}}[\,|Q_{N_{\max}}(G) - I(G)|^2\,]} \\
&\le e^{\mathrm{sh}}_{\mathcal{H}}(\{\boldsymbol{t}_i\}_{i=0}^{N_{\max}-1})^2\,\|F\|_{\mathcal{H}}\,\|G\|_{\mathcal{H}} \\
&\le \frac{c_\mu^2\,\|F\|_{\mathcal{H}}\,\|G\|_{\mathcal{H}}}{N_{\max}^{2\mu}},
\end{aligned} \tag{3.5}$$

where $\mu \approx 1$ under an appropriate function space setting [11].

Recall that the key property (3.3) for MFQMC is exactly the same as (2.14) for MFMC, and thus the mean-square error (or variance) of MFQMC can also be written as (2.17)–(2.18), as we state in (3.6) below.

A crude upper bound can be obtained using the triangle inequality, with the difference of two covariances in (2.17) bounded as $|a-b| \le |a| + |b|$ followed by (3.5), giving

$$\operatorname{mse}(Q^{\mathrm{MFQMC}}(f_0)) \le \frac{c_\mu^2\,\|g_0\|_{\mathcal{H}}^2}{N_0^{2\mu}} + \sum_{\ell=1}^{L} \frac{c_\mu^2\,\|g_\ell\|_{\mathcal{H}}^2}{N_{\ell-1}^{2\mu}} + 4\sum_{\ell=1}^{L} \frac{c_\mu^2\,\|g_\ell\|_{\mathcal{H}}\,\|g_0\|_{\mathcal{H}}}{N_{\ell-1}^{2\mu}}.$$

This bound loses all cancellations between covariance terms and overestimates the constant. In the theorem below we regroup the covariance terms to capture cancellations between models and improve the constant.

Theorem 3.2 (MFQMC variance decomposition) *For the MFQMC estimator* (3.1) *given by a randomly shifted embedded lattice sequence, with* $N_\ell = 2^{m_\ell}$ *and* $1 \le N_0 < N_1 < N_2 < \cdots$, *its mean-square error or variance is*

$$\begin{aligned}
\operatorname{mse}(Q^{\mathrm{MFQMC}}(f_0)) = \mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_0}(f_0)] &+ \sum_{\ell=1}^{L} \alpha_\ell^2\,\mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_\ell}(f_\ell) - Q_{N_{\ell-1}}(f_\ell)] \\
&+ 2\sum_{\ell=1}^{L} \alpha_\ell\,\operatorname{Cov}_{\boldsymbol{\Delta}}[Q_{N_\ell}(f_\ell) - Q_{N_{\ell-1}}(f_\ell), Q_{N_0}(f_0)],
\end{aligned} \tag{3.6}$$

which is minimized by taking

$$\alpha_\ell^* = \frac{-\operatorname{Cov}_{\boldsymbol{\Delta}}[Q_{N_\ell}(f_\ell) - Q_{N_{\ell-1}}(f_\ell), Q_{N_0}(f_0)]}{\mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_\ell}(f_\ell) - Q_{N_{\ell-1}}(f_\ell)]} \quad \text{for all } \ell = 1, \ldots, L. \tag{3.7}$$

The mean-square error (3.6) *can also be written as*

$$\begin{aligned}
\operatorname{mse}(Q^{\mathrm{MFQMC}}(f_0)) &= \sum_{\ell=0}^{L} E_\ell, \\
E_\ell &:= \operatorname{Cov}_{\boldsymbol{\Delta}}[Q_{N_\ell}(g_\ell - g_{\ell+1}), Q_{N_\ell}(2g_0 - g_\ell - g_{\ell+1})],
\end{aligned} \tag{3.8}$$

with $g_{L+1} := 0$, $g_\ell := \alpha_\ell\, f_\ell$, $\alpha_0 := 1$, $\alpha_{L+1} := 0$, *and it satisfies (cf.* (2.7) *with* $\mu \approx 1$*)*

$$\begin{aligned}
\operatorname{mse}(Q^{\mathrm{MFQMC}}(f_0)) &\le \sum_{\ell=0}^{L} |E_\ell| \le \sum_{\ell=0}^{L} \frac{W_\ell}{N_\ell^{2\mu}}, \\
W_\ell &:= c_\mu^2\,\|g_\ell - g_{\ell+1}\|_{\mathcal{H}}\,\|(g_0 - g_\ell) + (g_0 - g_{\ell+1})\|_{\mathcal{H}}.
\end{aligned} \tag{3.9}$$

Proof. The expression (3.6) follows immediately by combining the two covariance terms in (2.17) and substituting $g_\ell = \alpha_\ell f_\ell$. The optimal α_ℓ^* follows directly from (2.17), since $\alpha^2 A + 2\alpha B$ is minimized at $\alpha = -B/A$ when $A > 0$.

Starting again from (2.17)–(2.18), we can write

$$\begin{aligned}\mathrm{mse}(Q^{\mathrm{MFQMC}}(f_0)) = \mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_0}(g_0)] + \sum_{\ell=1}^{L} \Big(\mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_{\ell-1}}(g_\ell)] - \mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_\ell}(g_\ell)]\Big) \\ + 2\sum_{\ell=1}^{L} \Big(\mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_\ell}(g_\ell), Q_{N_\ell}(g_0)] - \mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_{\ell-1}}(g_\ell), Q_{N_{\ell-1}}(g_0)]\Big),\end{aligned}$$

where we used (3.3) to replace $Q_{N_0}(g_0)$ by $Q_{N_\ell}(g_0)$ and $Q_{N_{\ell-1}}(g_0)$ inside the two covariance terms. Using linearity of Q_N and bilinearity of covariance, we combine terms with respect to Q_{N_ℓ} and $Q_{N_{\ell-1}}$

$$\begin{aligned}\mathrm{mse}(Q^{\mathrm{MFQMC}}(f_0)) = \mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_0}(g_0)] \\ + \sum_{\ell=1}^{L} \mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_\ell}(g_\ell), Q_{N_\ell}(2g_0 - g_\ell)] - \sum_{\ell=1}^{L} \mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_{\ell-1}}(g_\ell), Q_{N_{\ell-1}}(2g_0 - g_\ell)].\end{aligned}$$

Next we reindex the second sum and separate out terms involving Q_{N_0} and Q_{N_L}

$$\begin{aligned}\mathrm{mse}(Q^{\mathrm{MFQMC}}(f_0)) = \mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_0}(g_0)] \\ + \mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_L}(g_L), Q_{N_L}(2g_0 - g_L)] + \sum_{\ell=1}^{L-1} \mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_\ell}(g_\ell), Q_{N_\ell}(2g_0 - g_\ell)] \\ - \mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_0}(g_1), Q_{N_0}(2g_0 - g_1)] - \sum_{\ell=1}^{L-1} \mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_\ell}(g_{\ell+1}), Q_{N_\ell}(2g_0 - g_{\ell+1})].\end{aligned}$$

Finally we combine terms according to Q_{N_ℓ} by factorization

$$\begin{aligned}\mathrm{mse}(Q^{\mathrm{MFQMC}}(f_0)) = \mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_0}(g_0 - g_1), Q_{N_0}(g_0 - g_1)] \\ + \sum_{\ell=1}^{L-1} \mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_\ell}(g_\ell - g_{\ell+1}), Q_{N_\ell}(2g_0 - g_\ell - g_{\ell+1})] \\ + \mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_L}(g_L), Q_{N_L}(2g_0 - g_L)],\end{aligned}$$

which yields (3.8). Applying (3.5) to (3.8) yields the final upper bound. □

3.2 MFQMC with multiple random shifts

So far the MFQMC estimator (3.1) has one random shift $\boldsymbol{\Delta}$. In practice, we will take R independent random shifts $\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_R$ to allow for error estimation. With a slight abuse of notation, we write

$$Q_R^{\mathrm{MFQMC}}(f_0) := \frac{1}{R}\sum_{k=1}^{R} Q^{\mathrm{MFQMC}}(f_0, \boldsymbol{\Delta}_k). \tag{3.10}$$

Its mean-square error is $1/R$ of the mean-square error with a single shift in (3.8), i.e.,

$$\mathrm{mse}(Q_R^{\mathrm{MFQMC}}(f_0)) = \frac{1}{R}\sum_{\ell=0}^{L} E_\ell, \quad E_\ell := \mathrm{Cov}_{\boldsymbol{\Delta}}[Q_{N_\ell}(g_\ell - g_{\ell+1}), Q_{N_\ell}(2g_0 - g_\ell - g_{\ell+1})],$$

which can be estimated in practice by

$$\mathrm{mse}(Q_R^{\mathrm{MFQMC}}(f_0)) \approx \frac{1}{R(R-1)} \sum_{k=1}^{R} \big(Q^{\mathrm{MFQMC}}(f_0, \boldsymbol{\Delta}_k) - Q_R^{\mathrm{MFQMC}}(f_0)\big)^2. \tag{3.11}$$

Summing up estimates of E_ℓ or $|E_\ell|$ and then dividing by R gives an alternative estimate for the mean-square error of MFQMC.

Our theory for MFQMC hinges upon the property (3.3) that the covariance is determined by $N_{\max}$ independently of $N_{\min}$. This allows us in theory to freely replace the smaller of N and N' in the covariance by any value up or down between $N_{\max}$ and 1. To compute the covariance for E_ℓ in practice, we would need to replace the second Q_{N_ℓ} by Q_{N_0}, since it is expensive to evaluate the highest fidelity model f_0.

We computed (3.3) numerically for two 1-dimensional test functions F, G and a sequence of sample sizes for N, N' using a machine-precision quadrature rule to compute the covariance, and verified that the covariance is indeed determined by $N_{\max}$ regardless of $N_{\min}$. However, when we estimated the covariance using samples over the random shifts, i.e.,

$$\mathrm{Cov}_{\boldsymbol{\Delta}}[Q_N(F), Q_{N'}(G)] \approx \frac{1}{R-1} \sum_{k=1}^{R} (F_k - \overline{F})\,(G_k - \overline{G}), \tag{3.12}$$

where $F_k := \frac{1}{N} \sum_{i=0}^{N-1} F(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}_k \rrbracket)$, $\overline{F} := \frac{1}{R} \sum_{k=1}^{R} F_k$, $G_k := \frac{1}{N'} \sum_{i=0}^{N'-1} G(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}_k \rrbracket)$, and $\overline{G} := \frac{1}{R} \sum_{k=1}^{R} G_k$, we obtained severely inaccurate estimates that change depending on the ratio $N_{\max}/N_{\min}$, even with R in the thousands. This is due to the high variance of the covariance estimator.

Instead of using (3.12) to estimate E_ℓ in our algorithm, we will use the upper bound (cf. (3.5))

$$|\,\mathrm{Cov}_{\boldsymbol{\Delta}}[Q_N(F), Q_{N'}(G)]\,| \le \sqrt{\mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_{\max}}(F)]}\,\sqrt{\mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_{\max}}(G)]},$$

where the variances can be estimated by

$$\mathbb{V}_{\boldsymbol{\Delta}}[Q_N(F)] \approx \frac{1}{R-1} \sum_{k=1}^{R} (F_k - \overline{F})^2,$$

and then rescaled based on (3.5) to give

$$\mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_{\max}}(F)] \approx \mathbb{V}_{\boldsymbol{\Delta}}[Q_N(F)] \left(\frac{N}{N_{\max}}\right)^{2\mu}.$$

Thus in practice we can estimate $|E_\ell| \approx W_\ell / N_\ell^{2\mu}$, where

$$W_\ell \approx \sqrt{\mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_\ell}(g_\ell - g_{\ell+1})]}\,\sqrt{\mathbb{V}_{\boldsymbol{\Delta}}[Q_{N_0}(2g_0 - g_\ell - g_{\ell+1})]}\,(N_\ell\, N_0)^\mu. \tag{3.13}$$

3.3 MFQMC cost-versus-error analysis

We now derive the theoretically optimal sample sizes N_ℓ as real numbers (rather than powers of 2 or integers) to help guide our practical algorithm.

Theorem 3.3 (MFQMC optimal sampling) *The MFQMC estimator* (3.10) *has*

$$\mathrm{cost}(Q_R^{\mathrm{MFQMC}}(f_0)) = \sum_{\ell=0}^{L} R\, N_\ell\, C_\ell \quad \textit{and} \quad \mathrm{mse}(Q_R^{\mathrm{MFQMC}}(f_0)) \le \sum_{\ell=0}^{L} \frac{W_\ell}{R\, N_\ell^{2\mu}},$$

where C_ℓ is the cost for one single evaluation of model f_ℓ, and W_ℓ is the level-ℓ constant in the mean-square error bound (3.9)*. Given a desired mean-square error demand ε^2 or a maximum total budget $\text{₿} > 0$, the optimal sample sizes are*

$$N_\ell := N_0 \Big(\frac{C_0}{W_0}\frac{W_\ell}{C_\ell}\Big)^{\frac{1}{2\mu+1}} \quad \textit{for all} \quad \ell = 1, \dots, L, \tag{3.14}$$

$$N_0 := \begin{cases} \big(\frac{W_0}{C_0}\big)^{\frac{1}{2\mu+1}} S^{\frac{1}{2\mu}}\, R^{-\frac{1}{2\mu}}\, \varepsilon^{-\frac{1}{\mu}} & \textit{for desired accuracy } \varepsilon^2, \textit{ or} \\ \big(\frac{W_0}{C_0}\big)^{\frac{1}{2\mu+1}} S^{-1}\, R^{-1}\, \text{₿} & \textit{for maximum budget } \text{₿}, \end{cases} \tag{3.15}$$

$$S := \sum_{\ell=0}^{L} W_\ell^{\frac{1}{2\mu+1}} C_\ell^{\frac{2\mu}{2\mu+1}}. \tag{3.16}$$

These choices lead to

$$\begin{cases} \mathrm{cost}(Q_R^{\mathrm{MFQMC}}(f_0)) = S^{\frac{2\mu+1}{2\mu}}\, R^{\frac{2\mu-1}{2\mu}}\, \varepsilon^{-\frac{1}{\mu}} & \textit{for desired accuracy } \varepsilon^2, \textit{ or} \\ \mathrm{mse}(Q_R^{\mathrm{MFQMC}}(f_0)) = S^{2\mu+1}\, R^{2\mu-1}\, \text{₿}^{-2\mu} & \textit{for maximum budget } \text{₿}. \end{cases} \tag{3.17}$$

Proof. To find the optimal values of $\{N_\ell\}$ to achieve a mean-square error demand ε^2, we use the Lagrange multiplier argument:

$$\mathcal{L}_{\mathrm{accuracy}}(N_0, \dots, N_L, \lambda) := \sum_{\ell=0}^{L} R\, N_\ell\, C_\ell + \lambda \Big(\sum_{\ell=0}^{L} \frac{W_\ell}{R\, N_\ell^{2\mu}} - \varepsilon^2 \Big),$$

$$\frac{\partial \mathcal{L}_{\mathrm{accuracy}}}{\partial N_\ell} = R\, C_\ell - \frac{2\,\mu\,\lambda\, W_\ell}{R\, N_\ell^{2\mu+1}} = 0, \qquad \frac{\partial \mathcal{L}_{\mathrm{accuracy}}}{\partial \lambda} = \sum_{\ell=0}^{L} \frac{W_\ell}{R\, N_\ell^{2\mu}} - \varepsilon^2 = 0. \tag{3.18}$$

The first condition in (3.18) gives $N_\ell^{2\mu+1} C_\ell / W_\ell = 2\,\mu\,\lambda / R^2$ for all $\ell = 0, \dots, L$, and thus $N_\ell^{2\mu+1} C_\ell / W_\ell = N_0^{2\mu+1} C_0 / W_0$ for all $\ell = 1, \dots, L$, yielding the formula for N_ℓ in (3.14). This choice of N_ℓ leads to

$$\sum_{\ell=0}^{L} \frac{W_\ell}{R\, N_\ell^{2\mu}} = \sum_{\ell=0}^{L} \frac{W_\ell}{R\, N_0^{2\mu} (\frac{C_0}{W_0}\frac{W_\ell}{C_\ell})^{\frac{2\mu}{2\mu+1}}} = \frac{1}{R\, N_0^{2\mu}} \Big(\frac{W_0}{C_0}\Big)^{\frac{2\mu}{2\mu+1}} S, \tag{3.19}$$

$$\sum_{\ell=0}^{L} R\, N_\ell\, C_\ell = \sum_{\ell=0}^{L} R\, N_0 \Big(\frac{C_0}{W_0}\frac{W_\ell}{C_\ell}\Big)^{\frac{1}{2\mu+1}} C_\ell = R\, N_0 \Big(\frac{C_0}{W_0}\Big)^{\frac{1}{2\mu+1}} S, \tag{3.20}$$

with S defined by (3.16). Setting (3.19) to be equal to ε^2 (i.e., satisfying the second condition in (3.18)) gives the first formula for N_0 in (3.15). Plugging this formula for N_0 into (3.20) then yields the final cost expression for the case of a desired accuracy ε^2.

Alternatively, to work with a maximum budget ₿ we consider instead

$$\mathcal{L}_{\mathrm{budget}}(N_0, \dots, N_L, \lambda) := \sum_{\ell=0}^{L} \frac{W_\ell}{R\, N_\ell^{2\mu}} + \lambda \Big(\sum_{\ell=0}^{L} R\, N_\ell\, C_\ell - \text{₿} \Big),$$

$$\frac{\partial \mathcal{L}_{\mathrm{budget}}}{\partial N_\ell} = -\frac{2\,\mu\, W_\ell}{R\, N_\ell^{2\mu+1}} + \lambda\, R\, C_\ell = 0, \qquad \frac{\partial \mathcal{L}_{\mathrm{budget}}}{\partial \lambda} = \sum_{\ell=0}^{L} R\, N_\ell\, C_\ell - \text{₿} = 0. \tag{3.21}$$

The first condition in (3.21) gives $N_\ell^{2\mu+1} C_\ell / W_\ell = 2\,\mu / (\lambda\, R^2)$ for all $\ell = 0, \dots, L$. Thus we obtain the same formula for N_ℓ in (3.14), as well as (3.19) and (3.20). Setting (3.20) to be equal to ₿ (i.e., satisfying the second condition in (3.21)) gives the second formula for N_0 in (3.15). Plugging this formula for N_0 into (3.19) then yields the final mean-square error bound for the case of a maximum budget ₿. □

We see from the theorem that the optimal cost for a given desired accuracy, and the optimal mean-square error for a given maximum cost, are both increasing functions of the number of random shifts R when the anticipated convergence rate satisfies $\mu > 1/2$. So we want to keep R as small as possible as long as it gives reasonable variance estimates in practice. Moreover, since the optimal cost and optimal error are both increasing functions of the scaling factor S, but otherwise independent of the selected models, in practice we should select models to minimize S.

3.4 MFQMC algorithm

Our MFQMC method is a combination of Algorithm 1, which takes care of model selection, and Algorithm 2, which successively increases the number of samples until a desired accuracy or maximum budget is reached.

Algorithm 1 (MFQMC model selection) *Input*: High-fidelity model $f_0 \equiv \widetilde{f}_0$, other models $\widetilde{f}_1, \dots, \widetilde{f}_{L_{\max}}$, $L_{\max} \geq 1$, number of random shifts $R \geq 2$, number of pilot samples $N_* = 2^{m_*} \geq 1$, dimension $s \geq 1$, extensible QMC sequence $\{\boldsymbol{t}_i\}_{i\geq 0} \subset [0,1]^s$, anticipated rate of convergence $\mu > 0$.

1. Generate independent random shifts $\boldsymbol{\Delta}_1, \dots, \boldsymbol{\Delta}_R \in [0,1]^s$.
2. Evaluate and store for $\ell = 0, \dots, L_{\max}$ and $k = 1, \dots, R$:

$$\begin{aligned}
\widetilde{A}_{\ell,m_*,k} &:= \tfrac{1}{N_*} \textstyle\sum_{i=0}^{N_*-1} \widetilde{f}_\ell(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}_k \rrbracket), \\
\widetilde{\sigma}^2_{\ell,k} &:= \tfrac{1}{N_*} \textstyle\sum_{i=0}^{N_*-1} \widetilde{f}^2_\ell(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}_k \rrbracket) - \widetilde{A}^2_{\ell,m_*,k}, \\
\widetilde{\sigma}_{0,\ell,k} &:= \tfrac{1}{N_*} \textstyle\sum_{i=0}^{N_*-1} \widetilde{f}_0(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}_k \rrbracket)\, \widetilde{f}_\ell(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}_k \rrbracket) - \widetilde{A}_{0,m_*,k}\, \widetilde{A}_{\ell,m_*,k}, \\
\widetilde{\alpha}_\ell &:= \tfrac{1}{R} \textstyle\sum_{k=1}^{R} \frac{\widetilde{\sigma}_{0,\ell,k}}{\widetilde{\sigma}^2_{\ell,k}}, \qquad \widetilde{\rho}_{0,\ell} := \tfrac{1}{R} \textstyle\sum_{k=1}^{R} \frac{\widetilde{\sigma}_{0,\ell,k}}{\widetilde{\sigma}_{0,k}\, \widetilde{\sigma}_{\ell,k}},
\end{aligned}$$

   and record average cost $\widetilde{C}_\ell$ for one single evaluation of $\widetilde{f}_\ell$.
3. Rescale averages by coefficients $\widetilde{A}_{\ell,m_*,k} \leftarrow \widetilde{\alpha}_\ell\, \widetilde{A}_{\ell,m_*,k}$ for all ℓ and k.
4. Sort and reorder models according to $|\widetilde{\rho}_{0,1}| \geq |\widetilde{\rho}_{0,2}| \geq \cdots$.
5. *Model Selection Loop*: For every combination $\mathfrak{u} = \{j_1, \dots, j_L\} \subseteq \{1, \dots, L_{\max}\}$ of models $(f_0, f_1, \dots, f_L) = (\widetilde{f}_0, \widetilde{f}_{j_1}, \dots, \widetilde{f}_{j_L})$:
   a. Identify $A_{\ell,m_*,k} \equiv \widetilde{A}_{j_\ell,m_*,k}$, $C_\ell \equiv \widetilde{C}_{j_\ell}$, $A_{L+1,m_*,k} \equiv 0$.
   b. With $\overline{A}_{\ell,m_*} := \frac{1}{R} \sum_{k=1}^{R} A_{\ell,m_*,k}$, compute

$$\begin{aligned}
W_\ell := 2^{(2m_*)\mu} &\sqrt{\tfrac{1}{R-1} \textstyle\sum_{k=1}^{R} [(A_{\ell,m_*,k} - \overline{A}_{\ell,m_*}) - (A_{\ell+1,m_*,k} - \overline{A}_{\ell+1,m_*})]^2} \\
&\times \sqrt{\tfrac{1}{R-1} \textstyle\sum_{k=1}^{R} [2(A_{0,m_*,k} - \overline{A}_{0,m_*}) - (A_{\ell,m_*,k} - \overline{A}_{\ell,m_*}) - (A_{\ell+1,m_*,k} - \overline{A}_{\ell+1,m_*})]^2},
\end{aligned}$$

   c. If $\frac{W_\ell}{C_\ell} \geq \frac{W_{\ell+1}}{C_{\ell+1}}$ for *any* $\ell = 0, \dots, L-1$, then set $S_{\mathfrak{u}} := \infty$. Otherwise, estimate the anticipated mean-square error and cost scaling $S_{\mathfrak{u}} := \sum_{\ell=0}^{L} W_\ell^{\frac{1}{2\mu+1}}\, C_\ell^{\frac{2\mu}{2\mu+1}}$.
6. *Return*: $\mathfrak{u}_* := \operatorname{argmin}_{\mathfrak{u} \subseteq \{1,\dots,L_{\max}\}} S_{\mathfrak{u}}$.

Algorithm 2 (MFQMC algorithm) *Input*: Preselected models $\{f_\ell\}_{\ell=0}^{L}$ from Algorithm 1, $L \geq 1$, number of random shifts $R \geq 2$, number of pilot samples $N_* = 2^{m_*} \geq 1$, dimension $s \geq 1$, extensible QMC sequence $\{\boldsymbol{t}_i\}_{i\geq 0} \subset [0,1]^s$, anticipated rate of convergence

$\mu > 0$, accuracy demand $\varepsilon^2 \in (0,1)$, maximum budget $\ddot{\mathfrak{B}} > 0$, as well as previously stored control variate coefficients $\{\alpha_\ell\}_{\ell=0}^L$, model costs $\{C_\ell\}_{\ell=0}^L$, random shifts $\{\boldsymbol{\Delta}_k\}_{k=1}^R \subset [0,1]^s$, and averages $A_{\ell,m_*,k} := \frac{1}{N_*}\sum_{i=0}^{N_*-1} \alpha_\ell\, f_\ell(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}_k \rrbracket)$.

Initialize: $N_\ell = 2^{m_\ell}$ with $m_\ell = m_* + \ell$ for $\ell = 0, \dots, L$.

MFQMC Loop:

1. Extend averages $A_{\ell,m,k} := \frac{1}{2}A_{\ell,m-1,k} + \frac{1}{2^m}\sum_{i=2^{m-1}}^{2^m-1} \alpha_\ell\, f_\ell(\llbracket \boldsymbol{t}_i + \boldsymbol{\Delta}_k \rrbracket)$ for $m = m_* + 1, \dots, m_\ell$ as required. Identify $A_{L+1,m,k} \equiv 0$.
2. Compute MFQMC estimator
$$\texttt{MFQMC} := \tfrac{1}{R}\textstyle\sum_{k=1}^R Q_k, \quad Q_k := A_{0,m_0,k} + \sum_{\ell=1}^L (A_{\ell,m_\ell,k} - A_{\ell,m_{\ell-1},k}),$$
$$\texttt{MSE} := \tfrac{1}{R(R-1)}\textstyle\sum_{k=1}^R (Q_k - \texttt{MFQMC})^2, \quad \texttt{COST} := \sum_{\ell=0}^L R\, N_\ell\, C_\ell.$$
3. If $\texttt{MSE} \le \varepsilon^2$ or $\texttt{COST} > \ddot{\mathfrak{B}}$ then *Return*: $(\texttt{MFQMC}, \texttt{MSE}, \texttt{COST})$.
4. With $\overline{A}_{\ell,m} := \frac{1}{R}\sum_{k=1}^R A_{\ell,m,k}$, estimate error contribution for $\ell = 0, \dots, L$:
$$|E_\ell| := 2^{(m_0-m_\ell)\mu}\sqrt{\tfrac{1}{R-1}\textstyle\sum_{k=1}^R [(A_{\ell,m_\ell,k} - \overline{A}_{\ell,m_\ell}) - (A_{\ell+1,m_\ell,k} - \overline{A}_{\ell+1,m_\ell})]^2}$$
$$\times\sqrt{\tfrac{1}{R-1}\textstyle\sum_{k=1}^R [2(A_{0,m_0,k} - \overline{A}_{0,m_0}) - (A_{\ell,m_0,k} - \overline{A}_{\ell,m_0}) - (A_{\ell+1,m_0,k} - \overline{A}_{\ell+1,m_0})]^2}.$$
5. For each $\ell = 0, \dots, L$, estimate the cost-benefit ratio if N_ℓ is doubled:
$$\mathrm{CBR}_\ell := \frac{|E_\ell| + \sum_{j=\ell+1}^L \delta_j\, |E_j|}{N_\ell\, C_\ell + \sum_{j=\ell+1}^L \delta_j\, N_j\, C_j},$$
where $\delta_j := 1$ if N_j also needs to be doubled to ensure that $N_\ell < N_{\ell+1} < \cdots < N_L$, otherwise $\delta_j := 0$.
6. Set $\ell^* := \mathrm{argmax}_{0 \le \ell \le L} \mathrm{CBR}_\ell$. Double N_{ℓ^*} and subsequent values of N_j accordingly. Then return to Step 1.

In Algorithm 1, we use a tilde to denote quantities associated with all available models, whereas quantities without a tilde refer to a specific combination of models. At the beginning, we compute N_* pilot samples for all available models to estimate their cost and covariances, and follow the MFMC strategy in [53, Theorem 3.4] to set control variate coefficients $\alpha_\ell := \sigma(f_0, f_\ell)/\sigma^2(f_\ell)$ and sort all models according to decreasing correlations with the highest fidelity model $\rho(f_0, f_\ell) := \sigma(f_0, f_\ell)/(\sigma(f_0)\,\sigma(f_\ell))$. In Algorithm 1 we use the estimated control variate coefficients and correlations for each shift and then take the average.

For each of the $2^{L_{\max}}$ possible combinations of models, we estimate $\{W_\ell\}$ using (3.13) with $N_\ell = N_0 = N_*$, and eliminate any combination if $\frac{W_\ell}{C_\ell} < \frac{W_{\ell+1}}{C_{\ell+1}}$ is violated for some ℓ; this condition is needed to ensure that the sequence $\{N_\ell\}$ from (3.14) is strictly increasing. (While we need $\{N_\ell\}$ to be a strictly increasing sequence for the MFQMC algorithm, from the point of view of estimating the constants $\{W_\ell\}$ it is fine to use the same value N_* to estimate different W_ℓ.) Next, we estimate the scaling factor S in (3.16) and select the combination of models for which this scaling factor is smallest.

The computation of α_ℓ in Algorithm 1 is based on the optimal value for MFMC in (2.12) instead of the optimal value for MFQMC in (3.7). For MFQMC the formula (3.7) depends on the values of N_ℓ and $N_{\ell-1}$, but we cannot get numerically accurate estimates of the covariance using (3.12). Such inaccuracies are acceptable as rough indicators in the algorithm (e.g., to decide which N_ℓ to double) but are not acceptable as final control variate coefficients. More work is needed on how to reliably use the optimal α_ℓ^* for MFQMC.

Once model selection is completed by Algorithm 1, in the main MFQMC loop in Algorithm 2 we compute the MFQMC estimator and estimate the mean-square error both from the outside (i.e., computing `MSE` in Step 2) and on the inside (i.e., estimating each error contribution $|E_\ell|$ in Step 4 and then computing $\frac{1}{R}\sum_{\ell=0}^{L}|E_\ell|$). We estimate $\{W_\ell\}$ using (3.13) again but now with the correct values for N_ℓ, N_0, and then estimate $|E_\ell| \approx W_\ell/N_\ell^{2\mu}$. Then we compute the *cost-benefit ratio* CBR_ℓ if N_ℓ is doubled. This follows the approach in [23], which is also used in [28] and justified based on the classical knapsack optimization which gives a Pareto-optimal solution. We will also need to double some subsequent N_j to ensure that the sequence remains increasing.

The numerator of CBR_ℓ is the reduction in the mean-square error. If N_ℓ is doubled, then the reduction in mean-square error theoretically is $E_\ell/R - E_\ell/(2^{2\mu}R) = (1 - 2^{-2\mu})\,E_\ell/R$. Each doubling of a subsequent N_j will contribute a reduction of $(1-2^{-2\mu})\,E_j/R$. This explains the numerator of CBR_ℓ, where the common factor $(1-2^{-2\mu})/R$ is dropped because it is a constant and will not affect the optimization.

The denominator of CBR_ℓ is the increase in the cost. If N_ℓ is doubled, then the increase in cost is $2R\,N_\ell\,C_\ell - R\,N_\ell\,C_\ell = R\,N_\ell\,C_\ell$. Each doubling of a subsequent N_j will contribute an increase of $R\,N_j\,C_j$. This explains the denominator of CBR_ℓ, again with the constant factor R dropped.

We choose the index ℓ with the largest cost-benefit ratio, and double N_ℓ and subsequent N_j as appropriate. Then we update averages and return to compute the new MFQMC estimator, together with updated estimates for the mean-square error and cost. This process is repeated until a desired accuracy is achieved or the maximum budget is reached.

At every iteration of the MFQMC loop, we expect the estimate of $\frac{1}{R}\sum_{\ell=0}^{L}|E_\ell| \approx \frac{1}{R}\sum_{\ell=0}^{L}W_\ell/N_\ell^{2\mu}$ to follow closely the same convergence rate as the mean-square error estimate `MSE`, just with a larger constant.

4 Numerical experiments

We compare the performance of MC, MFMC, QMC, and MFQMC for a simple illustrative problem and a comprehensive Antarctica Pine Island Glacier simulation.

4.1 Simple illustrative problem

For $\boldsymbol{y} \in [0,1]^s$ with $s = 128$ and $L_{\max} = 7$, we define the functions $f_0 \equiv \widetilde{f}_0$,

$$\widetilde{f}_\ell(\boldsymbol{y}) := \prod_{j=1}^{2^{7-\ell}} \exp\Big(\frac{y_j}{j^2}\Big) = \exp\Big(\sum_{j=1}^{2^{7-\ell}} \frac{y_j}{j^2}\Big), \quad \ell = 0, \dots, L_{\max},$$

and assign the costs $\widetilde{C}_\ell := 2^{-\ell}$ proportionally to the number of floating-point operations required to evaluate $\widetilde{f}_\ell$. For this simple illustrative problem we know the exact value of the integral $I(f_0) = \prod_{j=1}^{2^7} j^2(\exp(j^{-2}) - 1)$.

Figure 2 shows the sample root mean-square errors (2.3), (2.8), (2.19), (3.11) computed in 100 independent runs of MC, MFMC, QMC, and MFQMC for a computational budget of $\mathfrak{B} = 10^6$. The setup, sampling cost, root mean-square error convergence rate (least-squares fit) and relative error reduction factor (compared to MC) for each method are summarized in Table 1. For QMC and MFQMC we use $R = 8$ i.i.d. random shifts of an embedded lattice sequence $\{\boldsymbol{t}_i\}_{i\geq 0}$ generated following [7] with function space "weights" tailored to a spectral decay of $\mathcal{O}(j^{-2})$. MFMC [53, Algorithm 1] with $N_* = 8$ pilot samples has a total pilot sampling cost of $N_* \sum_{\ell=0}^{L_{\max}} \widetilde{C}_\ell = 15.9375$. MFQMC with $N_* = 4$ pilot samples has a total pilot sampling cost of $R\,N_* \sum_{\ell=0}^{L_{\max}} \widetilde{C}_\ell = 63.75$. For each of the 100

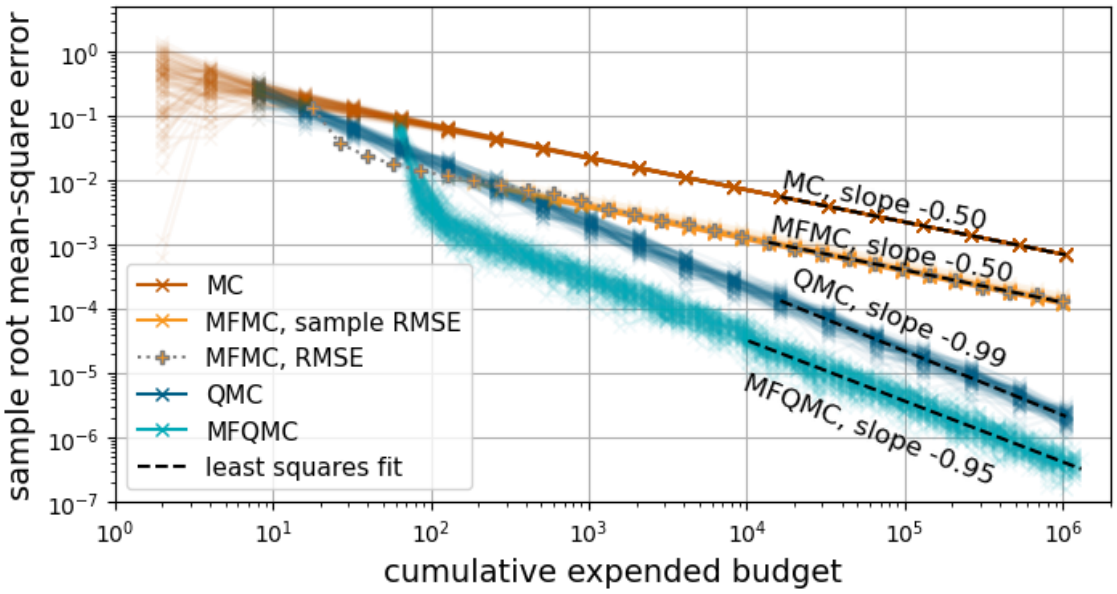

Figure 2: Sample root mean-square error for 100 independent runs of MC, MFMC, QMC, and MFQMC. Because the MFMC sample variance formula (2.19) can only be evaluated for sample sizes that are at least two apart, the plot includes the MFMC population root mean-square error $(\frac{1}{100}\sum_{r=1}^{100}(Q_r^{\mathrm{MFMC}}(f_0)-I(f_0))^2)^{1/2}$ over the 100 independent runs $Q_1^{\mathrm{MFMC}}(f_0),\dots,Q_{100}^{\mathrm{MFMC}}(f_0)$ of the MFMC estimator.

method	setup	sampling cost	rate	reduction
MC		$N\,C_0$	-0.50	1.0
MFMC	$N_* = 8$	$N_*\sum_{\ell=0}^{L_{\max}}\widetilde{C}_\ell + \sum_{\ell=0}^{L}(N_\ell - N_*)\,C_\ell$	-0.50	5.3
QMC	$R = 8$	$R\,N\,C_0$	-0.99	316.5
MFQMC	$R = 8, N_* = 4$	$R\,N_*\sum_{\ell=0}^{L_{\max}}\widetilde{C}_\ell + R\sum_{\ell=0}^{L}(N_\ell - N_*)\,C_\ell$	-0.95	1690.8

Table 1: Comparison between MC, MFMC, QMC, and MFQMC given total budget $\text{฿} = 10^6$.

independent runs, L denotes the number of models selected by Algorithm 1 and $N_\ell \geq N_*$ for $\ell = 0,\dots,L$ is the sample size for f_ℓ chosen by Algorithm 2 under reindexing of the selected models. At total budget $\text{฿} = 10^6$, the sample root mean-square error of MFQMC is reduced by a factor of 1690.8 compared to MC, and it converges at the rate $\text{฿}^{-0.95}$.

Figure 3 compares the sample root mean-square error obtained in 100 independent runs of MFQMC Algorithms 1 and 2 for $R = 2, 4, 8, 16$ shifts. The least-squares fits show that, as explained in Section 3.3, for the same computational cost, MFQMC performs worse when R is large. However, for small R, the spread of the sample root mean-square errors is relatively large, spanning up to two orders of magnitude for $R = 2$. A large spread indicates that the variance estimate cannot be trusted in practice.

4.2 Pine Island Glacier

We model the change in ice volume above floatation of the Pine Island Glacier in West Antarctica over a ten-year period subject to uncertainties in its basal friction field β. The complete modeling setup and model selection process are described in the Supplementary

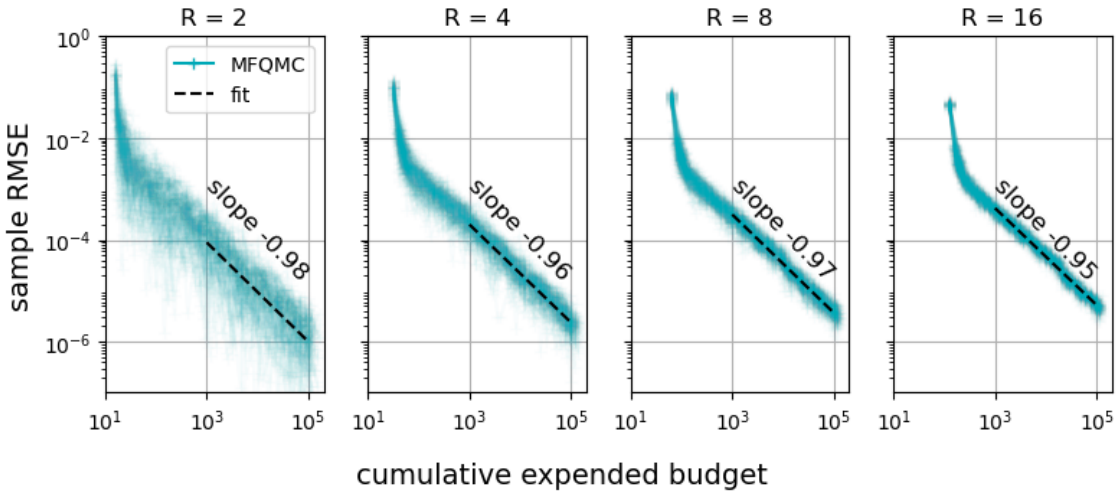

Figure 3: Sample root mean-square error (RMSE) computed in 100 independent runs of MFQMC with $R = 2, 4, 8, 16$ random shifts and $N_* = 4$ pilot samples. The least-squares fit is computed for sampling cost between 10^3 and 10^5.

ℓ	CPUh	t_f	dofs/step	velocity	temperature
0	27.219	10	170,160	3D, HO, t-dep.	3D, t-dep.
1	0.211	10	31,905	2D, SSA, t-dep.	2D, const.
2	0.087	4	31,905	2D, SSA, t-dep.	2D, const.
3	0.006	2	10,635	2D, SSA, const.	2D, const.

Table 2: Overview of models selected by Algorithm 1 for the Pine Island Glacier example, listing average computational cost (CPUh) over pilot samples, the final time (t_f) and degrees of freedom (dofs) per time step of the finite element solver, and how the velocity and temperature fields are modeled (dimension, governing equation, time-dependence). See the Supplementary Material for more details.

Material; here, we only give a brief overview.

The basal friction field β describes the sliding of the ice along any underlying bedrock, determining the velocity with which the ice flows into the ocean. We model the uncertainty in the basal friction field β with a log-normal distribution, i.e., $\ln \beta \sim \mathcal{N}(\ln \beta_0, \mathcal{C})$. The mean $\ln \beta_0$ and the covariance operator $\mathcal{C}$ were tailored to satellite information of Pine Island Glacier's surface velocity and melt. We perform a Karhunen–Loève expansion truncated at dimension $s := 2048$, thus

$$\ln \beta(\boldsymbol{x}; \boldsymbol{y}) \approx \ln \beta_0(\boldsymbol{x}) + \sum_{j=1}^{s} \Phi^{-1}(y_j) \sqrt{\lambda_j}\, \varphi_j(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega_{2\mathrm{D}}, \quad \boldsymbol{y} \in [0,1]^s,$$

where $\Omega_{2\mathrm{D}} \subset \mathbb{R}^2$ is the 2D horizontal extent of the modeled ice, $\boldsymbol{y}$ is uniformly distributed on $[0,1]^s$, Φ^{-1} is the inverse cumulative distribution function for the standard normal $\mathcal{N}(0,1)$, and (λ_j, φ_j) are the eigenpairs of $\mathcal{C}$. For QMC and MFQMC we use $R := 8$ random shifts of an embedded lattice sequence $\{\boldsymbol{t}_i\}_{i \geq 0} \subset [0,1]^s$. The realization of the friction field at the ith lattice point with the kth shift is denoted by $\beta(\boldsymbol{x}; \boldsymbol{y}^{i,k})$ with $\boldsymbol{y}^{i,k} := [\![\boldsymbol{t}_i + \boldsymbol{\Delta}_k]\!]$.

The ice volume above floatation $V_a(t)$ is defined as the volume of the glacier's ice at time $t \in [0, 10]$ that would contribute to sea level rise when melted. The difference $V_a(10) - V_a(0)$

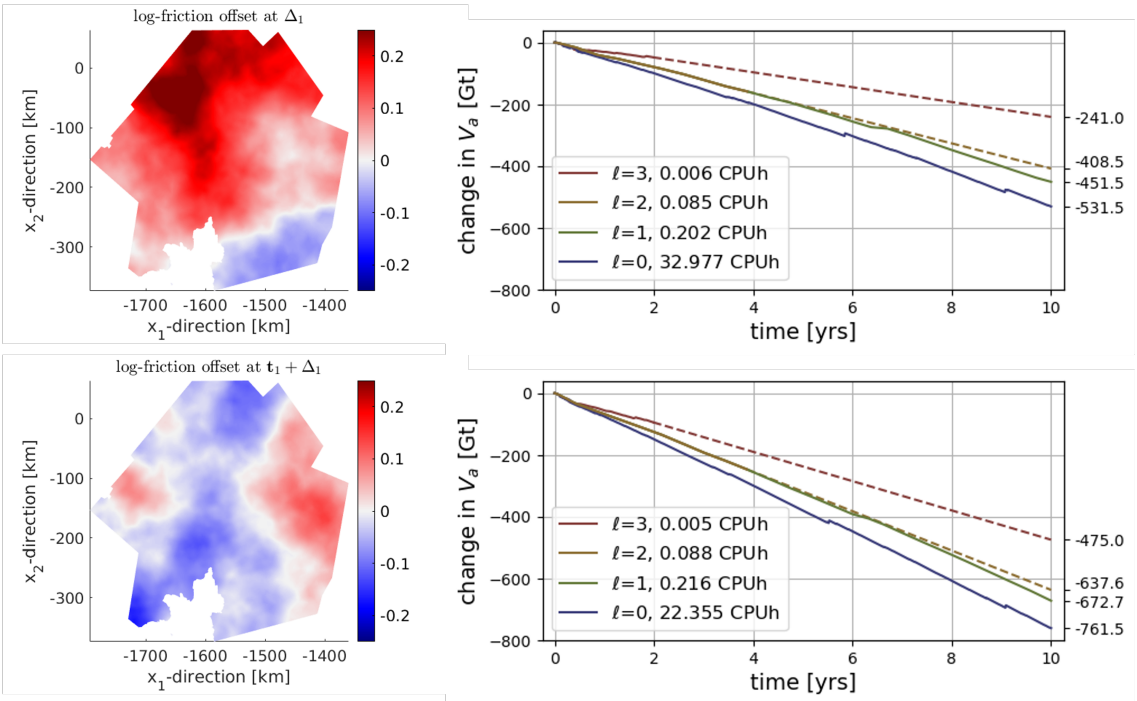

Figure 4: Left: Log-friction offset $\ln \beta(\boldsymbol{x}; \boldsymbol{y}) - \ln \beta_0(\boldsymbol{x})$ for $\boldsymbol{y}^{0,1} = \boldsymbol{\Delta}_1$ (top) and $\boldsymbol{y}^{1,1} = [\![\boldsymbol{t}_1 + \boldsymbol{\Delta}_1]\!]$ (bottom). Right: induced change $V_a^{(\ell)}(t; \boldsymbol{y}) - V_a^{(\ell)}(0; \boldsymbol{y})$ in ice volume above floatation for models $\ell = 0, \ldots, 3$ over time $t \in [0, 10]$. Dashed lines are used in the extrapolation intervals of models $\ell = 2$ and $\ell = 3$.

measures the volume of sea level relevant ice that the glacier has gained or lost over the ten-year period, for example through melting. The ice volume above floatation $V_a(t)$ is computed from the 2D ice thickness field $h(t)$ at any time $t \in [0, 10]$. The ice thickness field follows an advection equation, which is linked to the uncertain basal friction field β through the advective ice velocity. Our chosen high- and low-fidelity models differ in their governing equations of the ice velocity, their treatment of ice temperature, and their simulated time period; in particular, the high-fidelity model computes the velocity fields as 3D variables whereas each low-fidelity model only approximates their 2D depth-average. The models are discretized with linear finite elements with 10,635 degrees of freedom for 2D, and 53,175 degrees of freedom for 3D variables. An overview of the physical and numerical setup of the high-fidelity model and our $L := 3$ selected low-fidelity models is given in Table 2; we refer to the Supplementary Material for details on their exact modeling setups and on the model selection process with Algorithm 1.

Denoting with $V_a^{(\ell)}(t; \boldsymbol{y})$ the ice volume above floatation at time $t \in [0, 10]$ computed by model $\ell \in \{0, 1, 2, 3\}$ for the friction field $\beta(\boldsymbol{x}; \boldsymbol{y})$, Figure 4 (right) shows the change $V_a^{(\ell)}(t; \boldsymbol{y}) - V_a^{(\ell)}(0; \boldsymbol{y})$ as a function of time $t \in [0, 10]$ for the log-friction field offset $\ln \beta(\boldsymbol{x}; \boldsymbol{y}) - \ln \beta_0(\boldsymbol{x})$ at $\boldsymbol{y}^{0,1} = [\![\boldsymbol{t}_0 + \boldsymbol{\Delta}_1]\!] = \boldsymbol{\Delta}_1$ (top) and $\boldsymbol{y}^{1,1} = [\![\boldsymbol{t}_1 + \boldsymbol{\Delta}_1]\!]$ (bottom). Note that the initial value $V_a^{(\ell)}(0; \boldsymbol{y}) = V_a^{(0)}(0; \boldsymbol{0})$ is the same for all models $\ell \in \{0, 1, 2, 3\}$ and all friction fields β. We see that all three low-fidelity models $\ell \in \{1, 2, 3\}$ underestimate the change in V_a obtained by the high-fidelity simulation $\ell = 0$, however, they do capture the general effect that $\boldsymbol{y}^{1,1}$ causes stronger changes in V_a than $\boldsymbol{y}^{0,1}$.

We define $f_\ell(\boldsymbol{y}) := V_a^{(\ell)}(10; \boldsymbol{y}) - V_a^{(\ell)}(0; \boldsymbol{y})$ as the change in $V_a^{(\ell)}(t; \boldsymbol{y})$ computed by each model $\ell \in \{0, 1, 2, 3\}$ after ten years. We run MC with $N = 2^7$ and QMC with $RN = 8N = 2^7$ samples of the high-fidelity model f_0 and record the sample mean-square errors. We

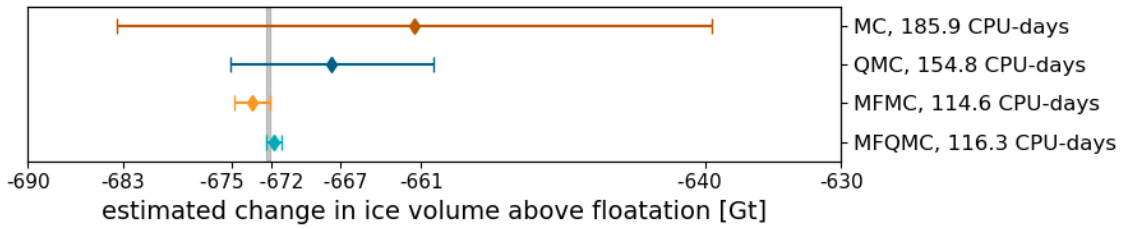

Figure 5: 95% confidence intervals computed with MC, MFMC, QMC, and MFQMC, and their overlap (shaded gray).

run MFMC [53] and MFQMC Algorithm 2 using all four models $\ell \in \{0, \dots, 3\}$, stopping when a budget of $\ddot{\mathrm{B}} = 120$ CPU-days would be exceeded. To save overall computational cost, QMC and MFQMC have the same $R = 8$ random shifts, and these shifts are reused as the first 8 samples for MC and MFMC.

Figure 5 shows the 95% confidence intervals $[Q^\diamond(f_0) - 1.96\sqrt{\mathrm{mse}(Q^\diamond(f_0))}, Q^\diamond(f_0) + 1.96\sqrt{\mathrm{mse}(Q^\diamond(f_0))}]$ obtained for each method $\diamond \in \{\mathrm{MC}, \mathrm{QMC}, \mathrm{MFMC}, \mathrm{MFQMC}\}$ with $\mathrm{mse}(Q^\diamond(f_0))$ being the sample variance computed via (2.3), (2.8), (2.19), and (3.11). The overlap $[-672.41, -672.17]$ in Gt of all four confidence intervals is shaded gray. We clearly see the increased certainty of the multifidelity methods, reducing the width of the 95% confidence intervals from ± 21.97 Gt and ± 7.48 Gt with MC and QMC to ± 1.24 Gt and ± 0.56 Gt with MFMC and MFQMC while also using fewer computational resources (114.6 CPU-days and 116.3 CPU-days with MFMC and MFQMC compared to 185.9 CPU-days with MC and 154.8 CPU-days with QMC).

Figure 6 shows $Q^\diamond(f_0)$ (left) and $\sqrt{\mathrm{mse}(Q^\diamond(f_0))}$ (right) obtained for each method $\diamond \in \{\mathrm{MC}, \mathrm{QMC}, \mathrm{MFMC}, \mathrm{MFQMC}\}$, plotted against the cumulative expended computational budget. A least-squares fit reveals approximate decay rates of -0.51 for MC, -0.89 for QMC, -0.60 for MFMC, and -0.93 for MFQMC. The observed rate of -0.60 for MFMC is likely an artifact of the pilot sampling cost; we expect it to converge toward the theoretical MC convergence rate -0.5 if more samples are taken. The step-function shape of the curve for the sample root mean-square error with MFQMC is likely a consequence of the step-wise doubling of sample sizes in Algorithm 2.

The sample sizes for MC and QMC increase by a factor of 2 for each data point in Figure 6, starting with 4 samples for MC, and 8 shifts for QMC. The sample sizes for MFMC and MFQMC are more involved, and are shown in Figure 7 plotted against the expended computational budget. For MFMC we imposed fixed sampling ratios of $N_1/N_0 = 38.0$, $N_2/N_0 = 71.9$, $N_3/N_0 = 4118.4$ from which we only deviated through rounding and when enforcing $N_0 \geq 8$ to ensure that all pilot samples were reused in the MFMC estimator. As a consequence, the sample sizes for MFMC in Figure 7 are continuous up to rounding to integers. For MFQMC the sample sizes are powers of two, and doubled one at a time by Algorithm 2 based on the cost-benefit ratio for each level; the sample sizes are hence step functions. Comparing the MFMC and the MFQMC sample sizes, we see that MFMC prioritizes large sample sizes for the low-fidelity models, while MFQMC increases the number of high-fidelity samples more readily. At their final data points at similar computational cost, MFMC has taken 7.7 times more samples of the lowest-fidelity model $\ell = 3$ than MFQMC (compare $126,618$ samples to $16,384$) but less than half as many high-fidelity samples (30 compared to 64).

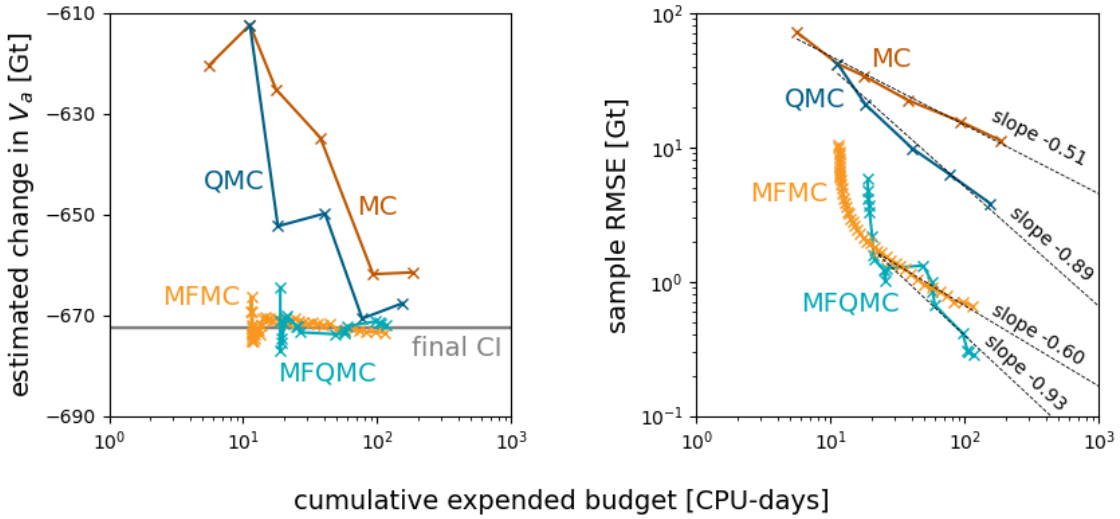

Figure 6: Estimated expected change $V_a(10) - V_a(0)$ (left) and corresponding sample root mean-square error (right) with MC, MFMC, QMC, and MFQMC over expended sampling cost in CPU-days.

5 Concluding remarks

In this work we introduced the MFQMC method, combining the sampling structure of the MFMC method from [50] with randomly shifted embedded lattice rules. The enabling result of our work is Lemma 3.1, which states that the covariance between any two QMC estimators (using the same lattice rule and shifts) depends only on the larger but not the smaller number of QMC points in the two estimators. The computational cost required for the MFQMC estimator to achieve a target root mean-square error ε scales almost as $\mathcal{O}(\varepsilon^{-1})$ (see Theorem 3.3), provably improving the rate $\mathcal{O}(\varepsilon^{-2})$ of MFMC (Theorem 2.1) and with a smaller implied constant than with QMC sampling if paired with appropriate model selection (Algorithm 1). In a numerical example estimating the expected ten-year loss of ice volume above floatation of the Pine Island Glacier under uncertainties in the basal friction field, MFQMC achieved a 39.4-fold and 13.4-fold reduction in the width of the 95% confidence interval compared to the MC and QMC methods at 62.5% of the MC cost and 75.1% of the QMC cost.

In Theorem 3.3 we proved theoretically optimal sample size ratios for distributing any computational budget among the available models for minimal mean-square error of the MFQMC estimator. However, these optimal sampling ratios are incompatible with choosing the number of QMC points as powers of two, which ensures that we always have complete lattice point sets in Lemma 3.1. In future work we will investigate how to lift this requirement to further improve the performance of the MFQMC method in practice.

Acknowledgment

Nicole Aretz and Karen Willcox acknowledge financial support from the Department of Energy grant DE-SC002317, the Air Force Office of Scientific Research MURI grant FA9550-24-1-0327, and the National Science Foundation grant #2103942. Frances Kuo acknowledges financial support from the Australian Research Council Discovery Project (DP240100769). Dirk Nuyens acknowledges support from the KU Leuven research fund under C1 grant 3E230463. Mike Giles, Frances Kuo and Dirk Nuyens acknowledge support

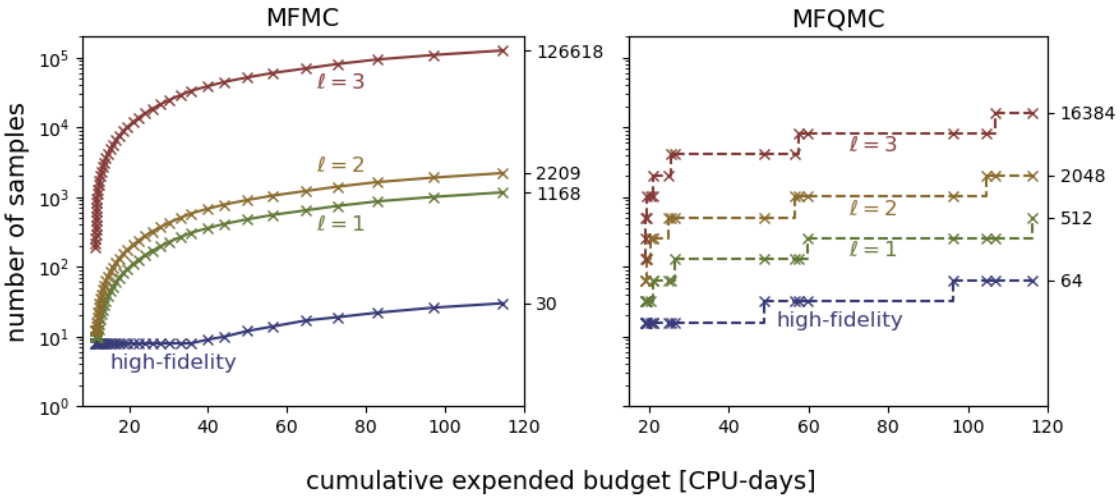

Figure 7: Sample sizes chosen by MFMC (left) and MFQMC (right). For MFQMC, sample sizes include the random shifts.

from the J. Tinsley Oden Faculty Fellowship Research Program.

References

[1] D. F. Anderson and D. J. Higham, *Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics*, Multiscale Model. Simul., 10, 146–179 (2012).

[2] N. Aretz, M. Gunzburger, M. Morlighem, and K. Willcox, *Multifidelity uncertainty quantification for ice sheet simulations*, Comp. Geosci., 29, 5 (2025).

[3] H. Ben-Ameur, P. L'Ecuyer, and C. Lemieux, *Combination of general antithetic transformations and control variables*, Math. Oper. Res., 29, 946–960 (2004).

[4] K. Bujok, B. M. Hambly, and C. Reisinger, *Multilevel simulation of functionals of Bernoulli random variables with application to basket credit derivatives*, Methodol. Comput. Appl. Probab., 17, 579–604 (2015).

[5] H.-J. Bungartz and M. Griebel, *Sparse grids*, Acta Numerica, 13, 147–269 (2004).

[6] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup, *Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients*, Comput. Visual. Sc., 14, 3–15 (2011).

[7] R. Cools, F. Y. Kuo, and D. Nuyens, *Constructing embedded lattice rules for multivariate integration*, SIAM J. Sci. Comput., 28, 2162–2188 (2006).

[8] R. Cools, F. Y. Kuo, D. Nuyens, and I. H. Sloan, *Fast component-by-component construction of lattice algorithms for multivariate approximation with POD and SPOD weights*, Math. Comp., 90, 787–812 (2021).

[9] M. Croci, M. B. Giles, M. E. Rognes, and P. E. Farrell, *Efficient white noise sampling and coupling for multilevel Monte Carlo with nonnested meshes*, SIAM/ASA J. Uncertain. Quantif., 6, 1630–1655 (2018).

[10] M. Croci, K. E. Willcox, and S. J. Wright, *Multi-output multilevel best linear unbiased estimators via semidefinite programming*, Comput. Methods. Appl. Mech. Engrg., 413, 116130 (2023).

[11] J. Dick, F. Y. Kuo, and I. H. Sloan, *High-dimensional integration: The quasi-Monte Carlo way*, Acta Numer., 22, 133–288 (2013).

[12] J. Dick, P. Kritzer, and F. Pillichshammer, *Lattice rules: Numerical Integration, Approximation, and Discrepancy*, Springer Series in Computational Mathematics, 58 (2022).

[13] T. J. Dodwell, S. Kynaston, R. Butler, R. T. Haftka, N. H. Kim, and R. Scheichl, *Multilevel Monte Carlo simulations of composite structures with uncertain manufacturing defects*, Probabilist. Eng. Mech., 63, 103116 (2021).

[14] G. Geraci, M. S. Eldred, and G. Iaccarino, *A multifidelity multilevel Monte Carlo method for uncertainty propagation in aerospace applications*, in: 19th AIAA non-deterministic approaches conference, 1951 (2017).

[15] A. D. Gilbert, M. B. Giles, F. Y. Kuo, I. H. Sloan, and A. Srikumar, *Multilevel lattice-based kernel approximation for elliptic PDE with random input*, in review (arXiv:2504.15810).

[16] A. D. Gilbert, F. Y. Kuo, D. Nuyens, G. Pash, I. H. Sloan, and K. E. Willcox, *Quasi-Monte Carlo methods for uncertainty quantification of tumor growth modeled by a parametric semi-linear parabolic reaction-diffusion equation*, to appear in SIAM/ASA J. Uncertain. Quantif. (arXiv:2509.25753).

[17] A. D. Gilbert, F. Y. Kuo, D. Nuyens, and G. W. Wasilkowski, *Efficient implementations of the Multivariate Decomposition Method for approximating infinite-variate integrals*, SIAM J. Sci. Comput., 40, A3240–A3266 (2018).

[18] A. D. Gilbert, F. Y. Kuo, and I. H. Sloan, *Analysis of preintegration followed by quasi-Monte Carlo integration for distribution functions and densities*, SIAM J. Numer. Anal., 61, 135–166 (2023).

[19] A. D. Gilbert and R. Scheichl, *Multilevel quasi-Monte Carlo for random elliptic eigenvalue problems I: regularity and error analysis*, IMA J. Numer. Anal., 44, 466–503 (2024).

[20] M. B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56, 607–617 (2008).

[21] M. B. Giles, *Multilevel Monte Carlo methods*, Acta Numer., 24, 259–328 (2015).

[22] M. B. Giles, F. Y. Kuo, and I. H. Sloan, *Combining sparse grids, multilevel MC and QMC for elliptic PDEs with random coefficients*, in: Monte Carlo and Quasi-Monte Carlo Methods (A. B. Owen and P. W. Glynn, eds.), Springer-Verlag, 265–281 (2018).

[23] M. B. Giles and B. J. Waterhouse, *Multilevel quasi-Monte Carlo path simulation*, Radon Series Comp. Appl. Math., 8, 1–18 (2009).

[24] A. A. Gorodetsky, G. Geraci, M. S. Eldred, and J. D. Jakeman, *A generalized approximate control variate framework for multifidelity uncertainty quantification*, J. Comput. Phys., 408, 109257 (2020).

[25] I. G. Graham, F. Y. Kuo, D. Nuyens, I. H. Sloan, and E. A. Spence, *Quasi-Monte Carlo methods for uncertainty quantification of wave propagation and scattering problems modelled by the Helmholtz equation*, published online in IMA J. Numer. Anal. (2026) `https://doi.org/10.1093/imanum/draf134`.

[26] A. Gruber, M. Gunzburger, L. Ju, R. Lan, and Z. Wang, *Multifidelity Monte Carlo estimation for efficient uncertainty quantification in climate-related modeling*, Geosci. Model Dev., 16, 1213–1229 (2023).

[27] P. A. Guth, V. Kaarnioja, F. Y. Kuo, C. Schillings, and I. H. Sloan, *Parabolic PDE-constrained optimal control under uncertainty with entropic risk measure using quasi-Monte Carlo integration*, Numer. Math., 156, 565–608 (2024).

[28] A. L. Haji-Ali, F. Nobile, and R. Tempone, *Multi-index Monte Carlo: When sparsity meets sampling*, Numer. Math., 132, 767–806 (2016).

[29] A. L. Haji-Ali and R. Tempone, *Multilevel and multi-index Monte Carlo methods for the McKean–Vlasov equation*, Statistics and Computing, 28, 923–935 (2018).

[30] H. Hakula, H. Harbrecht, V. Kaarnioja, F. Y. Kuo, and I. H. Sloan, *Uncertainty quantification for random domains using periodic random variables*, Numer. Math., 156, 273–317 (2024).

[31] F. J. Hickernell and H. S. Hong, *Quasi-Monte Carlo methods and their randomisations*, in: Applied Probability: AMS/IP Studies in Advanced Mathematics, 26 (R. Chan, Y.-K. Kwok, D. Yao and Q. Zhang, eds), AMS, 59–77 (2002).

[32] F. J. Hickernell, H. S. Hong, P. L'Ecuyer, and C. Lemieux, *Extensible lattice sequences for quasi-Monte Carlo quadrature*, SIAM J. Sci. Computing, 22, 1117–1138 (2000).

[33] F. J. Hickernell, C. Lemieux, and A. B. Owen, *Control variates for quasi-Monte Carlo*, Statistical Science, 20, 1–31 (2005).

[34] F. J. Hickernell and H. Niederreiter, *The existence of good extensible rank-1 lattices*, J. Complexity, 19, 286–300 (2003).

[35] A. S. N. Huda and R. Živanović, *Improving distribution system reliability calculation efficiency using multilevel Monte Carlo method*, Int. Trans. Electr. Energy Syst., 27, e2333 (2017).

[36] J. D. Jakeman, M. Perego, D. T. Seidl, T. A. Hartland, T. R. Hillebrand, M. J. Hoffman, and S. F. Price, *An evaluation of multifidelity methods for quantifying uncertainty in projections of ice-sheet mass change*, Earth Syst. Dyn., 16, 513–544 (2025).

[37] V. Kaarnioja, Y. Kazashi, F. Y. Kuo, F. Nobile, and I. H. Sloan, *Fast approximation by periodic kernel-based lattice-point interpolation with application in uncertainty quantification*, Numer. Math., 150, 33–77 (2022).

[38] A. Keller, F. Y. Kuo, D. Nuyens, and I. H. Sloan, *Regularity and tailored regularization of deep neural networks, with application to parametric PDEs in uncertainty quantification*, published online in Math. Comp. (2026) `https://doi.org/10.1090/mcom/4220`.

[39] J. Konrad, I.-G. Farcaş, B. Peherstorfer, A. Di Siena, F. Jenko, T. Neckel, and H.-J. Bungartz, *Data-driven low-fidelity models for multi-fidelity Monte Carlo sampling in plasma micro-turbulence analysis*, J. Comput. Phys., 451, 110898 (2022).

[40] F. Y. Kuo, W. Mo, and D. Nuyens, *Constructing embedded lattice-based algorithms for multivariate function approximation with a composite number of points*, Constr. Approx., 61, 81–113 (2025).

[41] F. Y. Kuo and D. Nuyens, *Application of quasi-Monte Carlo methods to elliptic PDEs with random diffusion coefficients – a survey of analysis and implementation*, Found. Comput. Math., 16, 1631–1696 (2016).

[42] F. Y. Kuo, D. Nuyens, L. Plaskota, I. H. Sloan, and G. W. Wasilkowski, *Infinite-dimensional integration and the multivariate decomposition method*, J. Comput. App. Math., 326, 217–234 (2017).

[43] F. Y. Kuo, R. Scheichl, Ch. Schwab, I. H. Sloan, and E. Ullmann, *Multilevel quasi-Monte Carlo methods for lognormal diffusion problems*, Math. Comp., 86, 2827–2860 (2017).

[44] F. Y. Kuo, Ch. Schwab, and I. H. Sloan, *Multilevel quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients*, Found. Comput. Math., 15, 411–449 (2015).

[45] R. R. Lam, O. Zahm, Y. M. Marzouk, and K. E. Willcox, *Multifidelity dimension reduction via active subspaces*, SIAM J. Sci. Comput., 42, A929–A956 (2020).

[46] A. Maurais, T. Alsup, B. Peherstorfer, and Y. Marzouk, *Multi-fidelity covariance estimation in the log-euclidean geometry*, International Conference on Machine Learning, PMLR, 24214–24235 (2023).

[47] A. Maurais, T. Alsup, B. Peherstorfer, and Y. M. Marzouk, *Multifidelity covariance estimation via regression on the manifold of symmetric positive definite matrices*, SIAM J. Math. Data Sci., 7, 189–223 (2025).

[48] D. T. P. Nguyen and D. Nuyens, *MDFEM: Multivariate decomposition finite element method for elliptic PDEs with uniform random diffusion coefficients using higher-order QMC and FEM*, Numer. Math., 148, 633–669 (2021).

[49] D. T. P. Nguyen and D. Nuyens, *MDFEM: Multivariate decomposition finite element method for elliptic PDEs with lognormal diffusion coefficients using higher-order QMC and FEM*, ESAIM: M2AN, 55, 1461–1505 (2021).

[50] L. W. T. Ng and K. E. Willcox, *Multifidelity approaches for optimization under uncertainty*, Int. J. Numer. Methods Eng., 100, 746–772 (2014).

[51] B. Peherstorfer, P. S. Beran, and K. E. Willcox, *Multifidelity Monte Carlo estimation for large-scale uncertainty propagation*, in 2018 AIAA Non-Deterministic Approaches Conference, 1660 (2018).

[52] B. Peherstorfer, M. Gunzburger, and K. Willcox, *Convergence analysis of multifidelity Monte Carlo estimation*, Numer. Math., 139, 683–707 (2018).

[53] B. Peherstorfer, K. Willcox, and M. Gunzburger, *Optimal model management for multifidelity Monte Carlo estimation*, SIAM J. Sci. Comput., 38, A3163–A3194 (2016).

[54] B. Peherstorfer, K. Willcox, and M. Gunzburger, *Survey of multifidelity methods in uncertainty propagation, inference, and optimization*, SIAM Rev., 60, 550–591 (2018).

[55] P. Robbe, D. Nuyens, and S. Vandewalle, *Recycling samples in the multigrid multilevel (Quasi-)Monte Carlo method*, SIAM J. Sci. Comput., 41, S37–S60 (2019).

[56] P. Robbe, D. Nuyens, and S. Vandewalle, *Enhanced multi-index Monte Carlo by means of multiple semicoarsened multigrid for anisotropic diffusion problems*, Numer. Linear Algebra Appl., e2281 (2019).

[57] D. Schaden and E. Ullmann, *On multilevel best linear unbiased estimators*, SIAM/ASA J. Uncertain. Quantif., 8, 601–635 (2020).

[58] A. L. Teckentrup, R. Scheichl, M. B. Giles, and E. Ullmann, *Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients*, Numer. Math., 125, 569–600 (2013).

A Pine Island Glacier modeling setup

A.1 Quantity of interest and uncertainty

We denote with $\Omega_{2D} \subset \mathbb{R}^2$ (in meters) the horizontal extent of Pine Island Glacier, and with $b : \Omega_{2D} \to \mathbb{R}$ (in meters above sea level) its bedrock topography. Given an ice thickness field $h : \Omega_{2D} \to \mathbb{R}_{\geq 0}$ (in meters), the altitude of the ice base is $h_b := \max\{b, -\frac{\rho_{\text{ice}}}{\rho_{\text{water}}} h\}$ (in meters above sea level) with ice density $\rho_{\text{ice}} = 917\text{kg/m}^3$ and ocean water density $\rho_{\text{water}} = 1023\text{kg/m}^3$. We define $\Omega_g(h) := \{(x_1, x_2) \in \Omega_{2D} : h_b(x_1, x_2) = b(x_1, x_2)\}$ as the area in which the ice is grounded on the bedrock; the remaining ice in $\Omega_f(h) := \Omega_{2D} \setminus \Omega_g(h)$ floats in the ocean. Our quantity of interest is the (grounded) ice volume above floatation

$$V_a(t) := \frac{\rho_{\text{ice}}}{10^{12}} \int_{\boldsymbol{x} \in \Omega_g(h(t))} h(t)(\boldsymbol{x}) + \min\{b(\boldsymbol{x}), 0\} \, \frac{\rho_{\text{water}}}{\rho_{\text{ice}}} \, \mathrm{d}\boldsymbol{x} \qquad \text{(in Gt)}, \tag{A.1}$$

which is the mass of the ice that, when melted, contributes to sea level rise.

Starting at initial condition $h(0) = h_0 : \Omega_{2D} \to \mathbb{R}_{\geq 0}$, the ice thickness field $h(t)$ changes for time $t \geq 0$ following the advection equation

$$\frac{\mathrm{d}h}{\mathrm{d}t} = -\nabla \cdot h \, \overline{\boldsymbol{v}}(t) - m_b(h) + m_s \qquad \text{in } \Omega_{2D}, \tag{A.2}$$

where $\overline{\boldsymbol{v}}(t) : \Omega_{2D} \to \mathbb{R}^2$ (in m/yr) is a depth-averaged ice velocity field, $m_s : \Omega_{2D} \to \mathbb{R}$ (in m/yr) is the surface ice accumulation, and $m_b(h) := 50 \, \mathbb{1}_{\Omega_f(h)} : \Omega_{2D} \to \{0, 50\}$ with indicator function $\mathbb{1}_{\Omega_f(h)}$ applies a (moderate) basal melt rate of 50 m/yr to floating ice.

A major driver of the ice velocity—and hence the depth-averaged ice velocity $\overline{\boldsymbol{v}}$ in (A.2)—is the basal friction field, which we denote with $\beta : \Omega_{2D} \to \mathbb{R}_{\geq 0}$ (in $s^{1/2}m^{-1/2}$). In practice β is only indirectly observable through measurements of the surface ice velocity and consequently considered one of the largest uncertainties in ice sheet simulations. Following [2, 17], we model β with a log-normal distribution, $\ln \beta \sim \mathcal{N}(\ln \beta_0, \mathcal{C})$ with mean $\ln \beta_0 : \Omega_{2D} \to \mathbb{R}$ and a double-Laplacian Matérn covariance operator $\mathcal{C} := (-c_1\Delta + c_2 I)^{-2}$. Here, the mean $\ln \beta_0$ was inferred in an inverse problem from satellite measurements [12, 13] of Pine Island Glacier's 2007–2009 surface ice velocity; the scaling coefficients $c_1, c_2 > 0$ of the covariance operator were tuned so that the standard deviation of $V_a(10) - V_a(0)$ for our pilot samples has the same order of magnitude as the standard deviation reported for Pine Island Glacier in [15, Figure 7b].

We use a randomized double-pass algorithm [17] in the $L^2(\Omega_{2D})$ inner product to compute the leading eigenpairs $\{(\lambda_j, \varphi_j)\}_{j \geq 1} \subset \mathbb{R}_{\geq 0} \times L^2(\Omega_{2D})$ of $\mathcal{C}$, truncating at dimension $s = 2048$. The eigenvalue decay and six exemplary eigenmodes are shown in Figure A.1; asymptotically, the eigenvalue decay follows the rate $\mathcal{O}(j^{-2})$. Tailored to this decay rate, we construct an embedded lattice sequence $\{\boldsymbol{t}_i\}_{i \geq 0}$ following [4]. We can then sample from the Karhunen–Loève expansion of $\mathcal{C}$ via

$$\ln \beta(\boldsymbol{x}; \boldsymbol{y}) \approx \ln \beta_0(\boldsymbol{x}) + \sum_{j=1}^{s} \Phi^{-1}(y_j) \, \sqrt{\lambda_j} \, \varphi_j(\boldsymbol{x}),$$

for all $\boldsymbol{x} \in \Omega_{2D}$, where $\Phi^{-1} : (0, 1) \to \mathbb{R}$ is the inverse cumulative distribution function for the standard normal distribution and $\boldsymbol{y} \in [0, 1]^s$ is uniformly distributed. Figure A.2 shows the mean log-friction field $\ln \beta_0$ and the offsets $\ln \beta(\boldsymbol{x}; \boldsymbol{y}) - \ln \beta_0(\boldsymbol{x})$ at the exemplary QMC points $\boldsymbol{y} \in \{\boldsymbol{t}_2, \boldsymbol{t}_4, \boldsymbol{t}_5, \boldsymbol{t}_8, \boldsymbol{t}_9, \boldsymbol{t}_{10}, \boldsymbol{t}_{11}\}$.

Table A.1 lists the datasets used in our modeling setup.

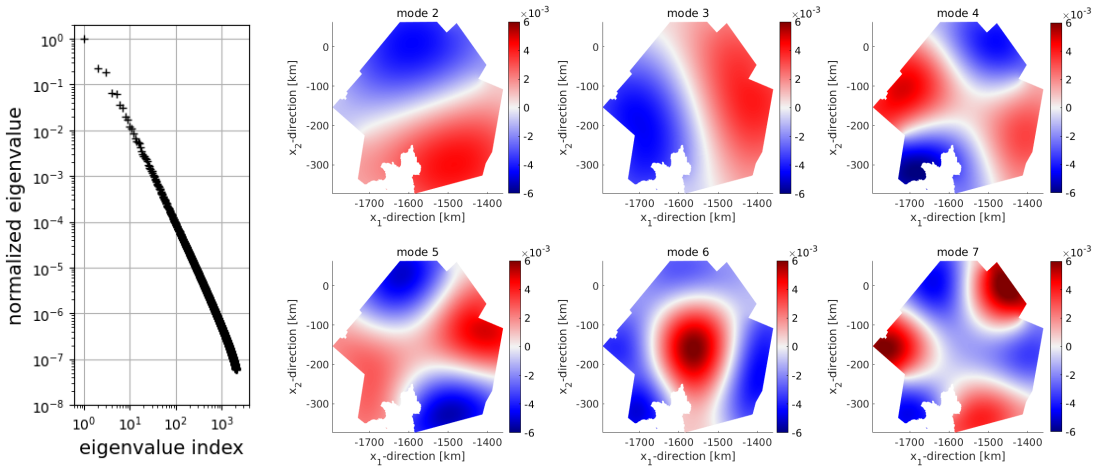

Figure A.1: Eigenvalue decay (left) and eigenmodes (right) of the covariance operator $\mathcal{C}$ in the log-normal distribution of the basal friction field.

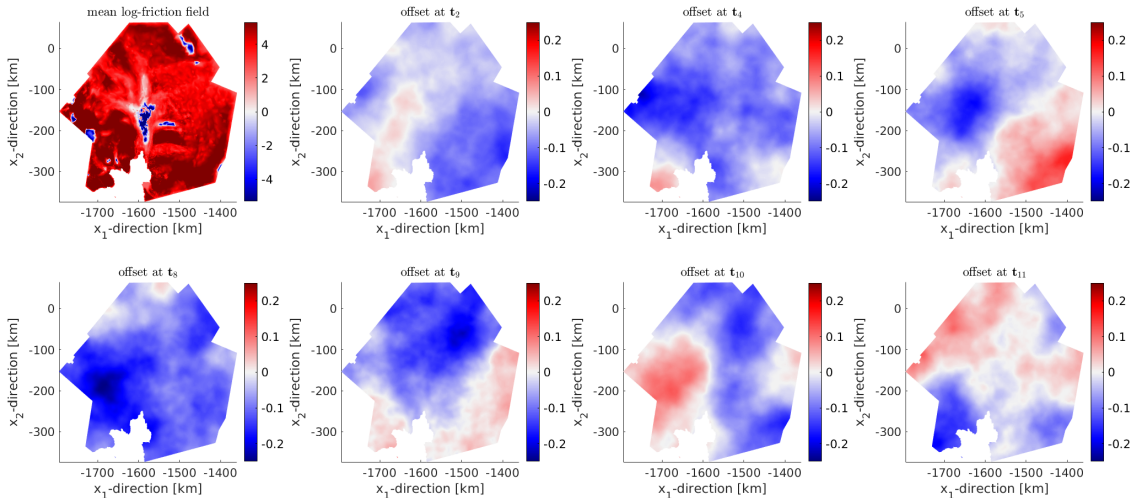

Figure A.2: Mean log-friction field $\ln \beta_0$ (top left) and offsets $\ln \beta(\boldsymbol{x}; \boldsymbol{y}) - \ln \beta_0(\boldsymbol{x})$ at 7 representative points $\boldsymbol{y}$.

Dataset	Use in our Pine Island Glacier example
[14]	horizontal ice extent Ω_{2D} and resolution for discretization
[10]	bedrock topography b
[1, 7]	surface altitude used to compute initial ice thickness h_0
[12, 13]	surface velocity used to infer the mean basal friction field β_0
[15]	annual melt rates used to tune covariance operator $\mathcal{C}$
[16]	surface mass balance m_s
[3]	Dirichlet boundary condition for ice temperature T
[9]	geothermal heat flux, basal Neumann boundary condition for T
[11, p.97]	ice temperature to ice rigidity conversion $B(T)$

Table A.1: References and datasets used in the modeling setup of our Pine Island Glacier example.

Type	CPUh	Dofs	$V_a(10)$	governing equations
HO-based	27.219	170,160	(A.1)	(A.2), (A.4), (A.5)
SSA-based	0.211	31,905	(A.1)	(A.2), (A.6)
fixed-velocity	0.030	10,635	(A.1)	(A.2), (A.6) at $t = 0$
extrapolation	base$\times t_f/10$	base	(A.7)	same as base model

Table A.2: Modeling setup overview for low-fidelity models. Computational cost is for ten years of simulations, degrees of freedom (dofs) are per time step. For any extrapolation-type low-fidelity model, the computational cost and degrees of freedom depend on the respective base model.

A.2 Model types

In the model selection in Section A.3 below, we choose between $L_{\max} = 29$ low-fidelity models. These low-fidelity models are combinations of three different model types—which we call "HO-based", "SSA-based", and "fixed-velocity" below—and ten different simulation endpoints $t_f \in \{1, \dots, 10\}$, with any $t_f < 10$ paired with an "extrapolation" technique. The high-fidelity model is the HO-based simulation until the final time $t = 10$. A summary of our models is given in Table A.2.

HO-based models. We consider two physical models for obtaining the depth-averaged ice velocity $\overline{\boldsymbol{v}}(t)$ in (A.2). First, in our high-fidelity model and HO-based low-fidelity models, we use the so-called higher-order (HO) equations in which the velocities v_1, v_2 in x_1- and x_2-direction are characterized as 3-dimensional fields on

$$\Omega_{3\mathrm{D}}(h) := \{(x_1, x_2, x_3) \in \Omega_{2\mathrm{D}} \times \mathbb{R} : \, h_b(x_1, x_2) \le x_3 \le h_b(x_1, x_2) + h(x_1, x_2)\}. \tag{A.3}$$

Specifically, for any given ice thickness field h, the velocities $v_1, v_2 : \Omega_{3\mathrm{D}}(h) \to \mathbb{R}$ are characterized by the system of PDEs

$$\nabla \cdot (2\,\mu\,\dot{\varepsilon}_{\mathrm{HO},i}) = \rho_{\mathrm{ice}}\, g\, \frac{\partial(h_b + h)}{\partial x_i} \quad \text{in } \Omega_{3\mathrm{D}}(h), \qquad i \in \{1, 2\}, \tag{A.4}$$

for all $(x_1, x_2, x_3) \in \Omega_{3\mathrm{D}}(h)$, where $g = 9.81$ m/s^2 is the gravitational acceleration,

$$\dot{\varepsilon}_{\mathrm{HO},1} = \begin{pmatrix} 2\frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} \\ \frac{1}{2}\frac{\partial v_1}{\partial x_2} + \frac{1}{2}\frac{\partial v_2}{\partial x_1} \\ \frac{1}{2}\frac{\partial v_1}{\partial x_3} \end{pmatrix}, \qquad \dot{\varepsilon}_{\mathrm{HO},2} = \begin{pmatrix} \frac{1}{2}\frac{\partial v_1}{\partial x_2} + \frac{1}{2}\frac{\partial v_2}{\partial x_1} \\ \frac{\partial v_1}{\partial x_1} + 2\frac{\partial v_2}{\partial x_2} \\ \frac{1}{2}\frac{\partial v_2}{\partial x_3} \end{pmatrix}$$

are the strain rates, and μ is the effective ice viscosity modeled by Glen's flow law $\mu = \frac{B(T)}{\sqrt{2}} \|(\dot{\varepsilon}_{\mathrm{HO},1}, \dot{\varepsilon}_{\mathrm{HO},2})\|^{-2/3}$ with temperature-dependent ice rigidity $B(T)$. The ice temperature T is modeled by the heat equation,

$$\frac{\partial T}{\partial t} = \frac{k_{\mathrm{ice}}}{\rho_{\mathrm{ice}} c_{\mathrm{ice}}} \Delta T - \boldsymbol{v} \cdot \nabla T + \frac{\mathrm{trace}(\sigma \cdot \dot{\varepsilon})}{\rho_{\mathrm{ice}}\, c_{\mathrm{ice}}} \qquad \text{in } \Omega_{3\mathrm{D}}(h), \tag{A.5}$$

with constant thermal conductivity $k_{\mathrm{ice}} = 2.4$ W/(m K), heat capacity $c_{\mathrm{ice}} = 2093$ J/(kg K), Cauchy stress tensor σ, strain rate tensor $\dot{\varepsilon}$, and constant Dirichlet boundary conditions at the ice-air interface and Neumann boundary conditions otherwise. The velocity vector $\boldsymbol{v} := (v_1, v_2, v_3) : \Omega_{3\mathrm{D}}(h) \to \mathbb{R}^3$ includes the vertical velocity field $v_3 : \Omega_{3\mathrm{D}}(h) \to \mathbb{R}$, which is recovered from v_1, v_2 through the incompressibility condition $\mathrm{div}(\boldsymbol{v}) = 0$. The basal friction field β enters the HO PDE system (A.4) in the sliding boundary conditions

$$2\,\mu\,\dot{\varepsilon}_{\mathrm{HO},i} \cdot \mathbf{n} = \begin{cases} -\beta^2\, P_{\mathrm{eff}}\, v_i & \text{if } (x_1, x_2) \in \Omega_g(h), \\ 0 & \text{if } (x_1, x_2) \in \Omega_f(h), \end{cases} \qquad i \in \{1, 2\},$$

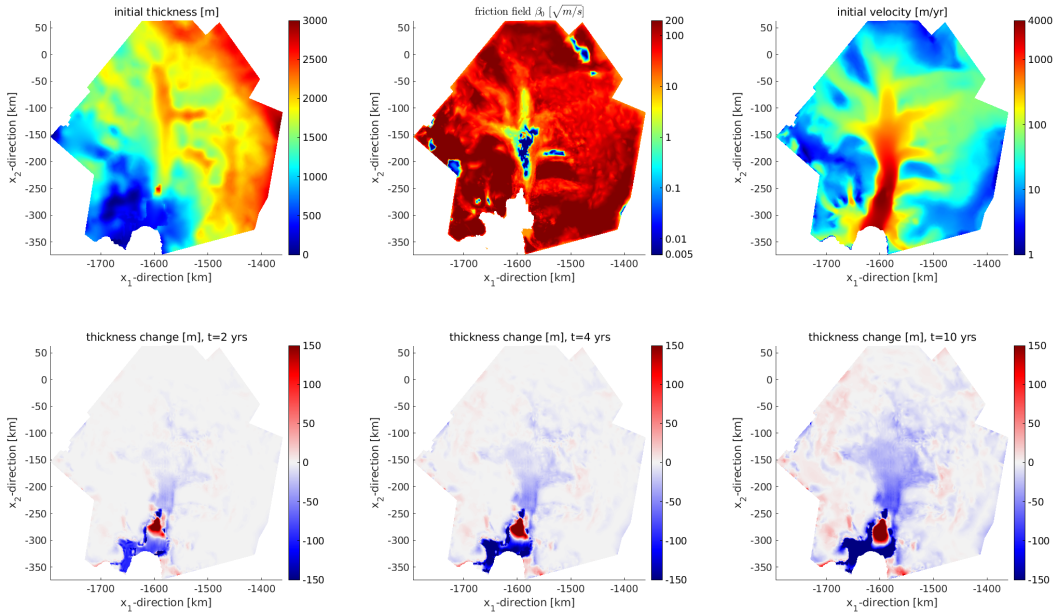

Figure A.3: Initial condition and ice thickness change at parametric mean. Top: Initial ice thickness h_0 in meters (left), friction field β_0 in $\sqrt{s/m}$ on $\Omega_g(h_0)$ (middle), and initial surface ice velocity in meters per year (right). Bottom: Change $h(t) - h_0$ in ice thickness at time $t = 2$ (left), $t = 4$ (middle), and $t = 10$ (right).

for all points (x_1, x_2, x_3) on the basal boundary of $\Omega_{3D}(h)$, with $\mathbf{n}$ denoting the outward pointing unit normal and with $P_{\text{eff}} := \rho_{\text{ice}}\, g\, h + \rho_{\text{water}}\, g\, \min\{b, 0\}$ being the effective pressure. At the remaining boundaries of $\Omega_{3D}(h)$ we apply water- and air-pressure at the ice-water and ice-air interfaces, and zero-Neumann boundary conditions on the remaining ice-ice interfaces. Given $v_1, v_2 : \Omega_{3D}(h) \to \mathbb{R}$, the depth-averaged ice velocity $\overline{\boldsymbol{v}} = (\overline{v}_1, \overline{v}_2)$ in the ice thickness equation (A.2) is obtained via

$$\overline{v}_i(x_1, x_2) := \frac{1}{h(x_1, x_2)} \int_0^{h(x_1,x_2)} v_i(x_1, x_2, h_b(x_1, x_2) + z)\, \mathrm{d}z, \qquad i \in \{1, 2\},$$

for all $(x_1, x_2) \in \Omega_{2D}$. Figure A.3 shows the initial ice thickness h_0, and the mean friction field β_0 with corresponding surface ice velocity and induced changes in the ice thickness computed by solving equations (A.2), (A.4), (A.5).

SSA-based low-fidelity models. As a low-fidelity physical approximation of the HO system of equations (A.4), we use the Shallow-Shelf Approximation (SSA) which models $\overline{\boldsymbol{v}} = (\overline{v}_1, \overline{v}_2) : \Omega_{2D} \to \mathbb{R}^2$ directly on the 2-dimensional domain Ω_{2D} by neglecting vertical shear stresses. Specifically, we solve the system of PDEs

$$\nabla \cdot (2\, \bar{\mu}\, h\, \dot{\varepsilon}_{\text{SSA},i}) - \mathbb{1}_{\Omega_g(h)}\, \beta^2\, P_{\text{eff}}\, \overline{v}_i = \rho_{\text{ice}}\, g\, \frac{\partial(h_b + h)}{\partial x_i} \quad \text{in } \Omega_{2D}, \quad i \in \{1, 2\}, \tag{A.6}$$

with $\mathbb{1}_{\Omega_g(h)} : \Omega_{2D} \to \{0, 1\}$ being the indicator function for the grounded ice domain $\Omega_g(h)$ and with the strain rates

$$\dot{\varepsilon}_{\text{SSA},1} = \begin{pmatrix} 2\frac{\partial \overline{v}_1}{\partial x_1} + \frac{\partial \overline{v}_2}{\partial x_2} \\ \frac{1}{2}\frac{\partial \overline{v}_1}{\partial x_2} + \frac{1}{2}\frac{\partial \overline{v}_2}{\partial x_1} \end{pmatrix}, \qquad \dot{\varepsilon}_{\text{SSA},2} = \begin{pmatrix} \frac{1}{2}\frac{\partial \overline{v}_1}{\partial x_2} + \frac{1}{2}\frac{\partial \overline{v}_2}{\partial x_1} \\ \frac{\partial \overline{v}_1}{\partial x_1} + 2\frac{\partial \overline{v}_2}{\partial x_2} \end{pmatrix},$$

and Neumann boundary conditions applying water- and air-pressure on the ice-ocean and ice-air interfaces of $\partial\Omega_{2D}$. The mean ice viscosity field $\bar{\mu} : \Omega_{2D} \to \mathbb{R}$ is constant in time, such that (A.2) and (A.6) can be solved together without needing to compute a temperature field; in total, using the SSA to approximate the depth-averaged ice velocity simplifies the HO-based high-fidelity model from three 3D-variables (temperature and horizontal ice velocity) and one 2D-variable (ice thickness) to three 2D-variables. However, because the SSA neglects vertical shear stresses, it can fail to approximate ice movement in high-friction areas where there is a high discrepancy between basal and surface ice movement. We refer to [6] for a detailed description of the HO equations and the SSA, and a discussion of their differences.

We discretize the ice thickness equation (A.2), the HO equations (A.4), the heat equation (A.5), and the SSA (A.6) with linear finite elements. The mesh resolution varies between 250 m and 10 km, similar to [14]. The (discrete) 2D variables (ice thickness and depth-averaged ice velocity) have 10,635 degrees of freedom each; the 3D variables (ice velocity of the HO equations, ice temperature) have 53,175 degrees of freedom. We use the Ice Sheet and Sea Level System Model (ISSM, [8]) to solve the governing equations for each sample $\ln \beta(\boldsymbol{x}; \boldsymbol{y})$ of the log-friction field and to compute the ice volume above floatation $V_a(t)$ for $t \in [0, 10]$. We use adaptive time stepping adhering to a CFL condition. On 4 cores, our high-fidelity model (ice thickness equation (A.2), HO equations (A.4), heat equation (A.5)) takes, on average, 27.219 CPUh on a dual-socket server equipped with two 64-core AMD EPYC 7H12 processors, totaling 128 physical cores (256 hardware threads) with a base clock speed of 2.60 GHz, and 2.0 TiB of RAM. On the same machine but on a single core, solving the ice thickness equation (A.2) with depth-averaged ice velocity computed with the SSA (A.6) takes, on average, 0.211 CPUh.

Fixed-velocity low-fidelity models. For fixed-velocity low-fidelity models, we follow the approach in [5] and compute the depth-averaged ice velocity $\overline{\boldsymbol{v}}$ for a given basal friction field only once at time $t = 0$ with the SSA (A.6). In the solve of (A.2) for the ice thickness $h(t)$, $t \in [0, 10]$, the depth-averaged ice velocity $\overline{\boldsymbol{v}}$ is then kept fixed, reducing the SSA-based low-fidelity system (A.2), (A.6) from three 2D variables to one 2D variable per time step and its average runtime from 0.211 to 0.03 CPUh. A comparison of the change $V_a(t) - V_a(0)$, $t \in [0, 10]$, computed with the high-fidelity model (A.2), (A.4), (A.5), the SSA-based system (A.2), (A.6), and the fixed-velocity approximation of (A.2) is shown in Figure A.4. Because the ice velocity only changes slowly in time, fixed-velocity low-fidelity models offer good approximations of SSA-based system (A.2), (A.6) for time t close to zero, but increasingly deviate as t increases.

Extrapolation-based low-fidelity models. For our final type of low-fidelity model we follow the extrapolation approach from [2]: For a given basal friction field β, we first run a base model—which may be any of the introduced model types (HO-based, SSA-based, or fixed-velocity)—until an early stopping point $t_f \in \{1, \ldots, 9\}$. We then compute the average annual change $(V_a^{\text{base}}(t_f) - V_a(0))/t_f$ in ice volume above floatation, and extrapolate it linearly for any $t > t_f$ via

$$V_a^{\text{extrap}}(t) := V_a(0) + \frac{t}{t_f}(V_a^{\text{base}}(t_f) - V_a(0)), \tag{A.7}$$

where $V_a^{\text{base}}(t)$, $t \in [0, t_f]$, is the ice volume above floatation computed with the chosen base model. The extrapolation approach reduces the runtime of its base model approximately to $t_f/10$ times that of the corresponding ten-year simulation.

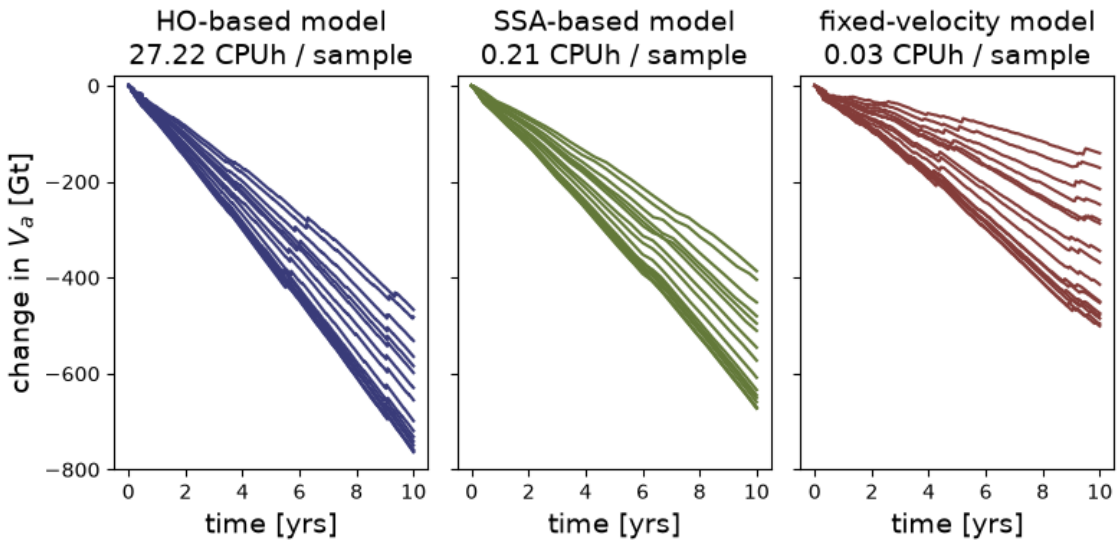

Figure A.4: Change $V_a(t) - V_a(0)$ in ice volume above floatation (in Gt) for $t \in [0, 10]$ with $V_a(t)$ computed with the HO-based model (left), the SSA-based model (middle), and the fixed-velocity model (right) for the 16 pilot samples.

ℓ	CPUh	Dofs	Model type	t_f	$V_a(t)$	Governing equations
0	27.219	170,160	HO-based	10	(A.1)	(A.2), (A.4), (A.5)
1	0.211	31,905	SSA-based	10	(A.1)	(A.2), (A.6)
2	0.087	31,905	extrap (SSA-based)	4	(A.7)	(A.2), (A.6)
3	0.006	10,635	extrap (fixed-velocity)	2	(A.7)	(A.2), (A.6) at $t = 0$

Table A.3: Overview on average CPU-time over pilot samples, degrees of freedom (dofs) per time step, and combination of governing equations of the high-fidelity model $\ell = 0$ and the selected low-fidelity models $\ell \in \{1, 2, 3\}$.

A.3 Model selection

With the three base model types (HO, SSA, and fixed-velocity), the nine possible extrapolation points $t_f \in \{1, \ldots, 9\}$ years and with the high-fidelity model being the HO-based simulation of $V_a(t)$ until $t = 10$ years, we have $L = 29$ low-fidelity models in total. We draw $R = 8$ random shifts, and compute the ten-year change $V_a(10) - V_a(0)$ in ice volume above floatation with each model for each shift and the first $N_* = 2 = 2^1$ QMC points, for a total of 16 pilot samples per model. Our pilot sampling cost is 439.365 CPUh. For the extrapolation-type low-fidelity models we reuse outputs of $V_a^{\text{base}}(t)$ of the ten-year simulation of the respective base model; consequently we do not count the extrapolation-type low-fidelity models towards our pilot sampling cost. For the model selection, we set the cost of the extrapolation-type models to be their base model cost times the ratio of number of time steps needed to reach $t \geq t_f$ compared to $t = 10$ years.

We run the model selection Algorithm 1 with the additional restriction to choose at most six low-fidelity models. With this restriction, the model selection loop (step 5) in Algorithm 1 remains reasonable (88 seconds) despite the large number $L = 29$ of low-fidelity model candidates. Algorithm 1 chooses three low-fidelity models: First, for $\ell = 1$, the SSA-based, ten-year simulation of $V_a(t)$ through the system (A.2), (A.6); second, for $\ell = 2$, the extrapolation (A.7) beyond $t_f = 4$ years using (A.2), (A.6) as base model; and third, for $\ell = 3$, the extrapolation (A.7) beyond $t_f = 2$ years with the fixed-velocity base model. An overview of our final model ensemble is provided in Table A.3.

References

[1] J. Bamber, J. Gomez-Dans, and J. Griggs, *A new 1 km digital elevation model of the Antarctic derived from combined satellite radar and laser data - Part 1: Data and methods*, The Cryosphere, 3 (2009), pp. 101–111, https://doi.org/10.5194/tc-3-101-2009.

[2] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler, *A computational framework for infinite-dimensional Bayesian inverse problems Part I: The linearized case, with application to global seismic inversion*, SIAM Journal on Scientific Computing, 35 (2013), pp. A2494–A2523.

[3] J. C. Comiso, *Variability and trends in Antarctic surface temperatures from in situ and satellite infrared measurements*, Journal of Climate, 13 (2000), pp. 1674–1696.

[4] R. Cools, F. Y. Kuo, and D. Nuyens, *Constructing embedded lattice rules for multivariate integration*, SIAM Journal on Scientific Computing, 28 (2006), pp. 2162–2188.

[5] L. Gkimisis, N. Aretz, M. Tezzele, T. Richter, P. Benner, and K. E. Willcox, *Non-intrusive reduced-order modeling for dynamical systems with spatially localized features*, Computer Methods in Applied Mechanics and Engineering, 444 (2025), p. 118115.

[6] R. Greve and H. Blatter, *Dynamics of ice sheets and glaciers*, Springer Science & Business Media, 2009.

[7] J. Griggs and J. Bamber, *A new 1 km digital elevation model of Antarctica derived from combined radar and laser data - Part 2: Validation and error estimates*, The Cryosphere, 3 (2009), pp. 113–123, https://doi.org/10.5194/tc-3-113-2009.

[8] E. Larour, H. Seroussi, M. Morlighem, and E. Rignot, *Continental scale, high order, high spatial resolution, ice sheet modeling using the Ice Sheet System Model (ISSM)*, Journal of Geophysical Research: Earth Surface, 117 (2012).

[9] C. F. Maule, M. E. Purucker, N. Olsen, and K. Mosegaard, *Heat flux anomalies in Antarctica revealed by satellite magnetic data*, Science, 309 (2005), pp. 464–467.

[10] F. Nitsche, S. Jacobs, R. Larter, and K. Gohl, *Bathymetry of the Amundsen Sea continental shelf: Implications for geology, oceanography, and glaciology*, Geochemistry, Geophysics, Geosystems, 8 (2007), https://doi.org/10.1029/2007GC001694.

[11] W. S. B. Paterson, *The Physics of Glaciers*, Pergamon Press, Oxford, London, New York, 3rd ed., 1994.

[12] E. Rignot, J. Mouginot, and B. Scheuchl, *Ice Flow of the Antarctic Ice Sheet*, Science, 333 (2011), pp. 1427–1430, https://doi.org/10.1126/science.1208336.

[13] E. Rignot, J. Mouginot, and B. Scheuchl, *MEaSUREs InSAR-Based Antarctica Ice Velocity Map, Version 1*, 2011, https://doi.org/10.5067/MEASURES/CRYOSPHERE/nsidc-0484.001.

[14] H. Seroussi, M. Morlighem, E. Rignot, J. Mouginot, E. Larour, M. Schodlok, and A. Khazendar, *Sensitivity of the dynamics of Pine Island Glacier, West Antarctica, to climate forcing for the next 50 years*, The Cryosphere, 8 (2014), pp. 1699–1710, https://doi.org/10.5194/tc-8-1699-2014.

[15] H. Seroussi, V. Verjans, S. Nowicki, A. J. Payne, H. Goelzer, W. H. Lipscomb, A. Abe-Ouchi, C. Agosta, T. Albrecht, X. Asay-Davis, et al., *Insights into the vulnerability of Antarctic glaciers from the ISMIP6 ice sheet model ensemble and associated uncertainty*, The Cryosphere, 17 (2023), pp. 5197–5217.

[16] D. G. Vaughan, J. L. Bamber, M. Giovinetto, J. Russell, and A. P. R. Cooper, *Reassessment of net surface mass balance in Antarctica*, Journal of Climate, 12 (1999), pp. 933–946, https://doi.org/10.1175/1520-0442(1999)012<0933:RONSMB>2.0.CO;2.

[17] U. Villa, N. Petra, and O. Ghattas, *HIPPYlib: An Extensible Software Framework for Large-Scale Inverse Problems Governed by PDEs: Part I: Deterministic Inversion and Linearized Bayesian Inference*, ACM Trans. Math. Softw., 47 (2021), https://doi.org/10.1145/3428447.