The Multilevel Monte Carlo Method for Simulations of Turbulent Flows

QINGSHAN CHEN
Department of Mathematical Sciences, Clemson University, Clemson, South Carolina

Ju Ming
School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, China

(Manuscript received 16 February 2018, in final form 12 June 2018)

ABSTRACT

In this paper, the application of the multilevel Monte Carlo (MLMC) method to numerical simulations of turbulent flows with uncertain parameters is investigated. Several strategies for setting up the MLMC method are presented, and the advantages and disadvantages of each strategy are also discussed. A numerical experiment is carried out using an idealized model for the Antarctic Circumpolar Current (ACC) with uncertain, small-scale bottom topographic features. It is demonstrated that unlike the pointwise solutions, the averaged volume transports are correlated across grid resolutions, and the MLMC method can increase simulation efficiency without losing accuracy in uncertainty assessments.

1. Introduction

The Monte Carlo (MC) method is a numerical technique for ensemble simulations, and it was invented by Metropolis and Ulam (1949) at the Los Alamos National Laboratory in the 1940s to simulate the interactions among a large number of neutron particles. The MC method has a constant convergence rate of 1/2 with regard to the number of samples, which is considered low. But unlike other popular ensemble techniques (e.g., polynomial chaos, collocation, and sparse grid), the MC is not subject to the curse of dimensionality, a term referring to the situation in which the volume of the sample space increases exponentially with the dimension and, thus, makes any fixed dataset increasingly sparse and statistically insignificant.

When the MC is applied to the complex systems described by differential equations with uncertainties (e.g., incomplete knowledge and/or inaccurate data), the total computational cost of the ensemble simulation will grow polynomially in terms of the computational cost for a single simulation. In other words, as one moves onto higher resolutions, the total computational cost for an ensemble simulation using the MC technique grows much faster than the computational cost of an individual simulation. To mitigate such rapid growth, many alternatives have been proposed, such as the quasi-Monte Carlo method (Dick et al. 2013; Kuo et al. 2012; Niederreiter 1993), the variance reduction method (Hammersley and Handscomb 1964), and the importance sampling and stratified sampling methods (Bugallo et al. 2015; Liu 2001).

Among these ameliorated methods, the multilevel Monte Carlo (MLMC) method has attracted much attention for its promising potential in reducing computational time for uncertainty quantification (UQ) problems (e.g., Heinrich 2001; Kebaier 2005; Giles 2008 and references therein). Similar to the multigrid method for iteratively solving large linear deterministic systems (Wesseling 1992), the MLMC algorithm utilizes a hierarchy of resolutions, instead of one, and uses the results of cheap low-resolution simulations to improve the results of high-resolution simulations. Higher efficiency can be achieved when the deduction in the variance in the samples from one level to the next can offset the increase in the computational expenses.

Not long after its introduction, it was realized that the MLMC method can also be applied to partial differential equations. Barth et al. (2011) couple the MLMC method with the finite element method (FEM) to solve stochastic elliptic equations, and rigorous error analysis is given. Mishra et al. (2012a,b, 2016) and Mishra and Schwab (2012) couple the MLMC method with the finite volume method for hyperbolic systems.
Kornhuber et al. (2014) apply the MLMC with FEM to study stochastic elliptic variational inequalities. Li et al. (2016) couple the MLMC with the weak Galerkin method to study the elliptic equations. For a survey of the MLMC and the literature on its applications, see Giles (2015).

The present work explores the effectiveness of the MLMC method as an ensemble simulator of turbulent geophysical flows. Simulations of this sort are different from simulations of steady-state problems or laminar flows in at least two ways, and these differences present new challenges for the MLMC method. First, for turbulent flows, pointwise behaviors are no longer relevant. After the initial spinup period, the differences between solutions on two different meshes are just differences between noises, even though all the other settings are kept the same. Thus, the usual notion of error convergence (i.e., pointwise error under certain norms) no longer applies. Because of the unreliable nature of the pointwise behavior of the solutions, the purpose of turbulent simulations is often focused on computing certain aggregated quantities of interests (QoI), such as the mean sea surface temperature (SST) over a region. So far, the pointwise error convergence has been a key assumption in the analysis of the MLMC method (Barth et al. 2011). One goal of this work is to adapt the analysis of the MLMC to QoI. The analysis will be based on a few assumptions about the simulations. Both the assumptions and the conclusions will be examined in a numerical experiment involving the Antarctic Circumpolar Current (ACC).

The second difference between simulations of laminar or steady-state flows and simulations of turbulent flows is that the latter often make use of turbulence closures, such as the eddy viscosity of Boussinesq, the anticipated potential vorticity method (Sadourny and Basdevant 1985), the Gent–McWilliams closure (Gent and McWilliams 1990; Gent et al. 1995), and the LANS-α model (Holm 1999). These closures often need to be adjusted according to the level of mesh resolutions. This means that at different resolutions, a different discrete model is being used. This is a dramatic departure from the scenario with steady-state or laminar flows, where the discrete model is kept the same, and only grid resolution varies. But this departure does not automatically invalidate the MLMC for turbulent flows. Eddy closures are implemented to prevent instability and to improve qualitative large-scale behavior of the solution, and the accuracy in the prediction of the QoI will likely still be aligned with the grid resolutions. For this reason, we expect that the prediction will improve or worsen as the mesh refines or coarsens. Based on this premise, we expect that the MLMC will still be effective for turbulent flows. We point out that this premise is unproven and will be examined during the numerical tests.

The objectives of this paper are twofold. First, we present a rigorous analysis of the MLMC method as it is applied to ensemble simulations of turbulent flows, with the goal of computing certain QoIs. How to set up a MLMC ensemble simulation (e.g., how many samples to use at each level) depends on many factors, such as the accuracy of the numerical scheme and the users’ error tolerance. Under this analysis, we present four strategies for setting up the MLMC simulation, and error and cost estimates are derived for each strategy. These strategies are practical, simple to use, and lead to samplings that are close to being optimal. Second, we test the overall effectiveness of the MLMC method and the performance of each of these strategies using a re-entrant jet model mimicking the ACC. The MLMC method is employed to compute the mean volume transport of the current, given the presumably incomplete knowledge concerning the bottom topography. The model used here is highly idealized, and thus the result from this model is not expected to be an accurate representation of the reality. The model is utilized as a controlled but physically relevant setting to test the effectiveness of the MLMC method on turbulent flows. The rest of the paper is organized as follows. In section 2, we briefly review the classical MC method and detail the analysis of the MLMC method under various strategies. In section 3, we apply the MLMC method to the ACC simulation and examine the effectiveness of the method under different strategies. The paper concludes with some remarks in section 4.

2. The Monte Carlo and the multilevel Monte Carlo methods

We designate the QoI to be calculated by $U$, which can be, for example, volume transport or mean SST.

We denote the number of levels of grid resolution by $L$, the spatial resolution at each level by $r_l, 1 \leq l \leq L$, and the highest resolution by $r = r_1$. We assume that $r_l = 2r_{l-1} = 2^{l-1}r_1 = 2^{l-1}r$. (1)

We assume that at each level, the computational cost is proportional to the total number of spatial–temporal degrees of freedom (DOF) $N_l$. For simplicity, in what follows, we identify the computational cost with $N_l$. We further assume that the total number of degrees of freedom is proportional to $r_l^{-3}$:

$$N_l = C_2 r_l^{-3}, \quad N = N_1 = C_1 r_1^{-3},$$ (2)
where $C_1$ is a constant. The cubic relation between $N_s$ and $r_s$ is tailored toward models of large-scale geophysical flows, where the vertical resolution is often held fixed, and the time step size varies linearly according to the horizontal resolution. From (1) and (2), it is derived that

$$\frac{N}{N_s} = 8^{-(l-1)}. \quad (3)$$

\[\text{a. The Monte Carlo method}\]

The classical MC method is applied on ensemble simulations at a fixed resolution $r$. The numerical approximation of the QoI $U$ at this resolution is denoted by $U_r$, and the computational cost of each individual simulation by $N$, which is related to the grid resolution through (2). We denote each realization by a superscript $m$, as in $U^m_r$ and $U^m$, $1 \leq m \leq M$. The ensemble mean by the MC method (hereafter MC mean) is defined as

$$E_M[U_r] = \frac{1}{M} \sum_{m=1}^{M} U^m_r. \quad (4)$$

We now examine the difference between the sample mean and the expectation $E[U]$ of the true solution $U$. We use the standard notations for the $\sigma$-finite probability space $(\Omega, \mathcal{F}, P)$, where the sample space $\Omega$ is a set of all possible outcomes, $\mathcal{F}$ is a $\sigma$-algebra of events, and $P : \mathcal{F} \to [0, 1]$ is a probability measure. The difference between the sample mean and the expectation can be decomposed into two terms:

$$\|E[U] - E_M[U_r]\|_{L^2(\Omega, P)} \leq \frac{\delta}{\sqrt{M}} + \epsilon, \quad (5)$$

where $\delta$ is the standard deviation in the true solution, and $\epsilon$ is the $L^2$ norm of the error at the resolution $r$ that assumes the form

$$\epsilon = \|U - U_r\|_{L^2(\Omega, P)} = C_2 r^\alpha, \quad (6)$$

where $C_2$ is a constant independent of the grid resolution.

The first term on the right-hand side represents the discretization error of the probability space, and the second term represents the discretization error of the temporal–spatial space.

For a given resolution $r$, the spatial–temporal discretization error is fixed. The sample size $M$ should be chosen so that the probability space discretization error is on the same order as the temporal–spatial discretization error. Thus, we set

$$M = \frac{\delta^2}{\epsilon^2}. \quad (7)$$

Given this choice of $M$, the combined errors in the MC mean can be given:

$$\|E[U] - E_M[U_r]\|_{L^2(\Omega, P)} \leq 2\epsilon. \quad (8)$$

The total computational cost for the Monte Carlo method $\mathcal{C}_MC$ can also be calculated, using (2), (6), and (7):

$$\mathcal{C}_MC = N \times M = \frac{\delta^2}{C_2^{2/3}} C_2 N^{1 + (2/3)\alpha}. \quad (9)$$

The total computational cost for the ensemble simulation using the classical MC method grows polynomially in terms of $N$, with a degree of $1 + (2/3)\alpha$. Also as expected, larger deviation $\delta$ in the true solution would demand more computational resource.

\[\text{b. The multilevel Monte Carlo method}\]

We denote the numerical approximation of $U$ at each level by $U_l$, $1 \leq l \leq L$, and each realization of $U_l$ by $U^{m}_l$, $1 \leq m \leq M_l$.

We note that the numerical approximation $U_l$ at the lowest level (highest resolution) can be decomposed as

$$U_l = \sum_{i=1}^{L-1} (U_i - U_{i+1}) + U_L. \quad (10)$$

Then, clearly,

$$E[U_r] = \sum_{i=1}^{L-1} E[U_i - U_{i+1}] + E[U_L]. \quad (11)$$

In practice, the mean is approximated by MC mean, and as it has been shown above, the accuracy of such approximation is determined by two competing factors: the variance in the random variable $\delta$ and the sample size $M$. A larger variance requires a larger sample size. The success of the MLMC method is built on the hypothesis that the variance of the difference between two solutions at successive levels is much smaller than the variance of each individual solution and thus requires a much smaller sample size. We now define the $L$-level sample mean of $U$ $E^L[U]$ as

$$E^L[U] = \sum_{i=1}^{L-1} E_{M_i}[U_i - U_{i+1}] + E_{M_L}[U_L], \quad (12)$$

where $M_i$ represents the sample size, and the sample mean $E_{M_i}$ at each level is defined in the same way as in (4). The relation between $M_i$ and the total sample size $M_l$ at each level is as follows:
\[
\begin{aligned}
\begin{cases}
M_1 &= M_1', \\
M_l &= M_{l-1} + M_l',
\end{cases} & 2 \leq l \leq L.
\end{aligned}
\] (13)

We now examine the theoretical mean \( E[U] \) and the \( L \)-level sample mean \( E^L[U] \).

\[
\|E[U] - E^L[U]\|_{\mathcal{L}^2(\Omega; P)} \leq \|E[U] - E[U_1]\|_{\mathcal{L}^2(\Omega; P)}
+ \|E[U_1] - E^L[U]\|_{\mathcal{L}^2(\Omega; P)}.
\] (14)

We note that

\[
\|E[U] - E[U_1]\|_{\mathcal{L}^2(\Omega; P)} \leq E\left[\|U - U_1\|_{\mathcal{L}^2(\Omega; P)}\right]
= \|U - U_1\|_{\mathcal{L}^2(\Omega; P)}.
\]

By the standing assumption (6), we obtain an estimate of the first term on the right-hand side of (14):

\[
\|E[U] - E[U_1]\|_{\mathcal{L}^2(\Omega; P)} \leq \epsilon = C_2 \times r_1^a.
\] (15)

For the second term on the right-hand side of (14), using the relation (11), the definition (12), and the central limit theorem, we find that

\[
\|E[U_1] - E^L[U]\|_{\mathcal{L}^2(\Omega; P)} \leq \sum_{l=1}^{L-1} \frac{\delta[U_l - U_{l+1}]}{\sqrt{M_l}} + \frac{\delta[U]}{\sqrt{M_L}}.
\]

For \( 1 \leq l \leq L - 1 \), from the standard definition of variance, we deduce that

\[
\delta[U_l - U_{l+1}] \leq E\left[\|U_l - U_{l+1}\|^2\right]
\leq 2(E[|U_l - U|^2] + E[|U_{l+1} - U|^2]).
\]

Again, thanks to the standing assumption (6),

\[
\delta[U_l - U_{l+1}] \leq 2C_2^a(1 + 4^a) \times r_1^a.
\] (16)

For the variance at the lowest resolution \( \delta[U_L] \), using the identity \( U_L = E[U_L] = U_L = U + U - E[U] + E[U] - E[U_L] \) and the assumption (6), we can show that

\[
\delta[U_L] \leq 2C_2^a r_1^a + \delta[U].
\]

Combining the last three estimates, we obtain

\[
\|E[U_1] - E^L[U]\|_{\mathcal{L}^2(\Omega; P)} \leq C_2 \sqrt{2(1 + 4^a)} \sum_{l=1}^{L-1} \frac{r_l^a}{\sqrt{M_l}}
+ \frac{2C_2 r_1^a}{\sqrt{M_L}} + \frac{\delta[U]}{\sqrt{M_L}}.
\]

Assuming that \( \alpha \geq 0 \), which should be true for all practically useful numerical schemes, we may bring the second term on the right-hand side into the summation to obtain

\[
\|E[U_1] - E^L[U]\|_{\mathcal{L}^2(\Omega; P)} \leq C_2 \sqrt{2(1 + 4^a)} \sum_{l=1}^{L} \frac{r_l^a}{\sqrt{M_l}} + \frac{\delta[U]}{\sqrt{M_L}}.
\] (17)

Combining (14), (15), and (17) leads us to

\[
\|E[U] - E^L[U]\|_{\mathcal{L}^2(\Omega; P)} \leq e + C_2 \sqrt{2(1 + 4^a)} \times \sum_{l=1}^{L} \frac{r_l^a}{\sqrt{M_l}} + \frac{\delta[U]}{\sqrt{M_L}}.
\] (18)

This estimate shows that the error in the \( L \)-level sample mean of the quantity \( U \) can be decomposed into three components: the temporal–spatial discretization error (first term), the error for using a multilevel structure (second term), and the probability space discretization error (third term).

So far, the sample size at each level \( M_l \) is yet to be determined. Determining the sample size will be a delicate balancing act between controlling the computational cost and controlling the error. \( \text{Giles} \ (2015) \) uses the Lagrangian multiplier technique and derives an optimal sampling, given a target accuracy of the estimation. However, in order to be able to calculate the optimal sampling, one needs to have perfect knowledge about various parameters, such as the intrinsic variance of the QoI and the accuracy of the numerical scheme, which are most likely missing in real-world applications.

In this work, we take a different approach. We forego the goal of reaching the “optimal” sampling and instead aim to develop several practical strategies that are attuned, and can be readily applied, to large-scale geophysical flows. In our strategies, the target accuracy is no longer a variable; we simply assume that the users want to achieve an accuracy that is commensurate with the highest resolution that they can afford, which is often a key constraint imposed on the users by the availability of the resources.

The potential of the MLMC method lies in the fact that within the probability space discretization error (third term), the standard deviation of the analytical solution is divided by the sample size at the highest level (lowest resolution), where the computational cost for an individual simulation is the lowest. We should also note that the first term, the temporal–spatial discretization error, is not affected by the sample size at any level. A general principle for determining the sample size is to make sure the probability space discretization error...
(third term) and each term in the summation (second term) is roughly on the order of the temporal–spatial discretization error (first term) or smaller. Thus, the sample size at each level is determined by the desired error distribution across the levels of resolutions. Here, we explore several different strategies for determining the sample size at each level.

1) STRATEGY 1

Our first strategy is to choose the sample size for each level so that each term in the summation of (18) is equal to the temporal–spatial discretization error. Hence, we set

\[ e = C_2^2 \varepsilon_i^\alpha = C_2 \sqrt{2(1 + 4^\alpha)} \frac{r_f^d}{\sqrt{M_i}}, \]

which leads to

\[ M_i = 2(1 + 4^\alpha) \cdot 2^{2\alpha(l-1)}, \quad 1 \leq i \leq L. \]

The number of \( L \) is determined by requiring that the sample size at the lowest resolution \( M_L \) be sufficiently large to make the probability discretization error be on the same order as the temporal–spatial discretization error as well:

\[ \frac{\delta[U]}{\sqrt{M_L}} = e, \]

from which we deduce [also using (19)] that

\[ L = 1 + \frac{2 \log \delta[U] - 2 \log e - \log(2(1 + 4^\alpha))}{2^\alpha \log 2}. \]

Using (15) and (2), we find that

\[ L = \frac{2 \log \delta[U]}{\alpha \log 2} + \frac{\log N}{3 \log 2} + 1 - \frac{\log C_1}{3 \log 2} \]

\[ - \frac{2 \log C_2 + \log(2(1 + 4^\alpha))}{2^\alpha \log 2}. \]

The expression on the right-hand side indicates that generally, a larger variance in the analytical solution requires more levels. Under the same order of convergence (\( \alpha \)) and the same constant coefficients \( C_1 \) and \( C_2 \), a higher number of degrees of freedom (\( N \), or, in other words, a finer mesh) also requires more levels.

Based on this strategy, the total error in the \( L \)-level sample mean is

\[ \| \mathbf{E}[U] - \mathbf{E}^L[U] \|_{\ell_p} \leq (L + 2)e. \]

We denote the computational cost under this strategy as \( \mathcal{C}_{\text{MLMC}} \). Using relation (13) and formula (19), we can calculate \( \mathcal{C}_{\text{MLMC}} \) as follows:

\[ \mathcal{C}_{\text{MLMC}} = \sum_{i=1}^{L} M_i N_i \leq \frac{9(1 + 4^\alpha)^2}{4} N \sum_{i=1}^{L} 2^{2\alpha-3}(i-1). \]

We note that the right-hand side of the above inequality is a geometric series. If \( \alpha < 3/2 \), then the right-hand side converges to a finite number as \( L \) tends to infinity, with the limit depending on the convergence rate \( \alpha \) only. Thus, in this case, the computational cost grows linearly as \( N \) increases:

\[ \mathcal{C}_{\text{MLMC}} = O(N). \]

If \( \alpha = 3/2 \), then the right-hand side can be simply written as

\[ \mathcal{C}_{\text{MLMC}} \leq \frac{9(1 + 4^\alpha)^2}{4} N L. \]

With \( L \) as given in (22), we conclude that

\[ \mathcal{C}_{\text{MLMC}} = O\{N(\log \delta[U] + \log N)\}. \]

We note that the case where \( \alpha = 3/2 \) exactly is rare in practice. But the result obtained here, together with the result for the \( \alpha < 3/2 \), indicates that with larger \( \alpha \), the computational cost will increase faster as \( N \) increases.

Finally, if \( \alpha > 3/2 \), then one can apply the summation formula for a finite geometric series to obtain

\[ \mathcal{C}_{\text{MLMC}} = \frac{9(1 + 4^\alpha)^2}{4} N \times \frac{2^{2\alpha-3}L - 1}{2^{2\alpha-3} - 1}. \]

Upon substituting the expression (22) for \( L \) in the above, we obtain

\[ \mathcal{C}_{\text{MLMC}} = O\left(\delta[U]^{\frac{2\alpha-1}{\alpha}} \times N^{1 + \frac{2\alpha-3}{\alpha}}\right). \]

In this case, the computational cost grows polynomially in \( N \) and \( \delta[U] \), similar to the situation with the classical Monte Carlo method [see (9)], but the exponents on both \( \delta[U] \) and \( N \) are lower in the case here, indicating that even if the convergence rate \( \alpha \) is greater than 3/2, there are still potential savings in computational time by choosing the MLMC method.

The downside of this strategy is that the error depends on the number of levels, which could be large. In the following strategies, we amplify \( M_i \) by certain factors so that the summation in (18) actually converges, even as the number of levels goes to infinity, so that the final error is actually independent of the number of levels taken.
2) STRATEGY 2

Under this strategy, we require the error term in the summation on the right-hand side of (18) to decrease exponentially as the level number \( l \) goes up; that is, we set

\[
\frac{\sqrt{2(1 + 4^a)} \times r_1^a}{\sqrt{M_l}} = \left( \frac{1}{2} \right)^{l-1} r_1^a,
\]

which leads to

\[
M_l = 2(1 + 4^a) \times 4^{(l-1)(a+1)}.
\] (27)

The number of levels required is determined as under the previous strategy and is found to be

\[
L = \frac{2}{(a + 1) \log 4} \log \delta[U] + \frac{2a}{3(a + 1) \log 4} \log N + 1
\]

\[
- \frac{2a}{3} \log C_1 + 2 \log C_2 + \log(1 + 4^a)
\]

\[
\frac{(a + 1) \log 4}{(\alpha + 1) \log 4}.
\] (28)

Under this strategy, the error in the \( L \)-level mean is independent of the number of levels for

\[
\|E[U] - E^L[U]\|_{\mathcal{L}_2(\Omega)} \leq \epsilon \left( 1 + \sum_{l=1}^L \left( \frac{1}{2} \right)^{l-1} + 1 \right) \leq 4\epsilon.
\] (29)

We denote the computational cost under this strategy by \( \mathcal{C}_{\text{MLMC2}} \). It is calculated as follows:

\[
\mathcal{C}_{\text{MLMC2}} = \sum_{l=1}^L M_l N_l = 2(1 + 4^a)(1 + 4^{-(a+1)})N
\]

\[
\times \prod_{l=1}^L 2^{(l-1)(2a-1)}.
\]

It is clear from the expression above that the critical convergence rate under this strategy is 1/2. How the total computational cost grows as the computational cost \( N \) for a single high-resolution simulation increases depends on whether the actual convergence rate is \( \alpha < 1/2 \), \( = 1/2 \), or \( > 1/2 \). Following a procedure similar to the one above, we find that if \( \alpha < 1/2 \), the computational cost \( \mathcal{C}_{\text{MLMC2}} \) grows linearly in \( N \):

\[
\mathcal{C}_{\text{MLMC2}} \leq \frac{2(1 + 4^a)(1 + 4^{-(a+1)})}{1 - 2^{2a-1}} N = O(N).
\] (30)

If \( \alpha = 1/2 \), then

\[
\mathcal{C}_{\text{MLMC2}} \leq O(N \log \delta[U] + \log N).
\] (31)

Finally, if \( \alpha > 1/2 \),

\[
\mathcal{C}_{\text{MLMC2}} = C_3 \frac{2(1 + 4^a)(1 - 4^{-(\alpha+1)})}{2^{(2\alpha-1)} - 1} \delta[U]^{\frac{3}{2\alpha-1}}
\]

\[
\times N^{1+[\alpha(2\alpha-1)(3\alpha+1)]}.
\]

(32)

Similar to the conventional MC method, the computational cost for the MLMC method under the current scenario \( \alpha > 1/2 \) still grows polynomially as \( N \) increases, but at a lower degree, for it is trivial to verify that

\[
\frac{\alpha(2\alpha-1)}{3\alpha+1} \leq \frac{2}{3}.
\]

The impact of the variance in the analytical solution on the computational cost is also lower, for it is obvious that

\[
\frac{2\alpha - 1}{\alpha + 1} \leq 2.
\]

3) STRATEGY 3

This strategy chooses a sample size so that each term on the right-hand side of (18), including the individual terms in the summation, contributes equally to the total error and then amplifies the sample size by a level-dependent factor to ensure convergence. With \( \sigma > 0 \) being a positive parameter, we let

\[
M_l = 2(1 + 4^a)(L - l + 1)^{(l+\sigma)} \times 2^{2a(l-1)}.
\] (33)

It is clear that this strategy leads to a much larger sample size at the finest resolution [cf. (27) or (19)], which goes against the main idea of the MLMC method of running a small sample at the highest resolution. As a consequence, little or no savings in computational cost is expected for this strategy. We will examine the performance of this strategy in the part of the numerical test. Here, we only list (without derivation details) the number of levels, the sample size at each level, and the estimates on the error and computational cost.

The number of levels under this strategy is found to be

\[
L = \log \frac{\delta[U]}{\log 4} - \log N + \log \frac{C_1}{3 \log 2} - \frac{2 \log C_2 + \log(1 + 4^a)}{2a \log 2}.
\] (34)
\[ \|E[U] - E^L[U]\|_{\mathcal{L}^2(\Omega;P)} \leq e \left( 3 + \int_1^\infty l^{-(1+\sigma)} \, dl \right) \leq \left( 3 + \frac{1}{\sigma} \right) e. \] (35)

Here, the parameter \( \sigma \) is the same as the one in (33).

As under the previous two strategies, the estimates on the computational cost depend on the convergence rate of the numerical scheme, and the critical convergence rate is again found to be 3/2, as under strategy 1. We denote the computational cost under the current strategy by \( \mathcal{C}_{\text{MLMC3}} \). If \( \alpha < 3/2 \), then

\[ \mathcal{C}_{\text{MLMC3}} = \mathcal{O} \left( N \times (\log \delta[U] + \log N)^{2(1+\sigma)} \right). \] (36)

If \( \alpha = 3/2 \), then

\[ \mathcal{C}_{\text{MLMC3}} = \mathcal{O} \left( N \times (\log \delta[U] + \log N)^{2\sigma+3} \right). \] (37)

If \( \alpha > 3/2 \), then

\[ \mathcal{C}_{\text{MLMC3}} = \mathcal{O} \left( (\log \delta[U] + \log N)^{2(1+\sigma)} \right. \\
\times \left. \delta[U]^{\frac{4\sigma}{3\sigma - 3}} \times N^{1 + \frac{1}{3\sigma - 3}} \right). \] (38)

4) STRATEGY 4

Strategy 4 is similar to strategy 3, but the sample sizes are amplified at higher levels (lower resolutions). We set

\[ M_l = 2(1 + 4^\alpha) \times 2^{(1+\sigma)l} \times 2^{2\alpha(l-1)}. \] (39)

To determine the number of levels \( L \), we require \( M_L \) to satisfy the relation (20):

\[ L^{2(1+\sigma)} \times 2^{2\alpha(L-1)} = \frac{\delta[U]^2}{e^2 \times 2(1 + 4^\sigma)}. \] (40)

The number of levels cannot be solved for explicitly from (40). But it is clear that it is smaller than that of strategy 3. This strategy leads to the same total error in the \( L \)-level sample mean:

\[ \|E[U] - E^L[U]\|_{\mathcal{L}^2(\Omega;P)} \leq \left( 3 + \frac{1}{\sigma} \right) e. \] (41)

We denote the computational cost under this strategy by \( \mathcal{C}_{\text{MLMC4}} \):

\[ \mathcal{C}_{\text{MLMC4}} = \sum_{l=1}^L M_l \times N_l \leq \frac{9(1 + 4^\alpha)}{4} \times N \times \sum_{l=1}^L 2^{(1+\sigma)l} \times 2^{2(2\alpha-3)(l-1)}. \]

It is clear from the expression above that the critical convergence rate for this strategy is 3/2, the same as those for strategies 1 and 2. Following a similar procedure laid out before, one finds that if \( \alpha < 3/2 \), then

\[ \mathcal{C}_{\text{MLMC4}} = \mathcal{O} \left( N \times (\log \delta[U] + \log N)^{2\sigma+3} \right). \] (42)

If \( \alpha = 3/2 \), then

\[ \mathcal{C}_{\text{MLMC4}} = \mathcal{O} \left( N \times (\log \delta[U] + \log N)^{2(1+\sigma)} \right) \times \delta[U]^{\frac{4\sigma}{3\sigma - 3}} \times N^{1 + \frac{1}{3\sigma - 3}} \}. \] (43)

If \( \alpha > 3/2 \), then

\[ \mathcal{C}_{\text{MLMC4}} = \mathcal{O} \left( N \times (\log \delta[U] + \log N)^{2(1+\sigma)} \right) \times \delta[U]^{\frac{4\sigma}{3\sigma - 3}} \times N^{1 + \frac{1}{3\sigma - 3}} \}. \] (44)

For easy comparisons, the computational cost for each strategy and the cost for the classical MC method are summarized in Table 1. All strategies, except strategy 3, experience three types of cost growth, depending on the convergence rate \( \alpha \): linear, quasi linear, and polynomial. When the convergence rate is high, the computational cost for all strategies grows polynomially, similar to the situation of the classical MC method. But the degrees of the polynomials are lower, compared with the classical MC method, offering potential savings in computing times. Strategy 3 appears less appealing in that it lacks linear growth for the computational cost, obviously because the sample size at the lowest level (highest resolution) is amplified.

c. Estimates

Under each of the strategies discussed above, the calculation of the number of levels, the sample size at each level, and the computational cost all depend on a few key parameters:

- Parameter \( \delta[U] \), the standard deviation in the true solution.
- Parameter \( \alpha \), the convergence rate of the numerical scheme regarding the QoI.
- Parameter \( e \), the \( L^2 \) norm of the error in the first approximation \( U_l \).

Determining the true values of these parameters touches upon several fundamental mathematical and numerical issues that, in many cases involving real-world applications, are completely open. For example, for many nonlinear systems (e.g., the three-dimensional Navier–Stokes equations governing fluid flows), the existence and uniqueness of a global solution are still open questions. Similarly, the numerical analysis to determine the convergence rate of numerical schemes for
nonlinear systems is very challenging or even not possible. We leave these theoretical issues to future endeavors. In the current work, we explore approaches to estimate these parameters from the discrete simulation data.

The standard deviation \( \delta [U] \) in the true solution can be approximated by the unbiased sample variance (Kenney and Keeping 1951):

\[
\delta [U] = \frac{1}{M_i - 1} \sum_{m=1}^{M_i} (U_{i,m}^n - \mathbb{E}_{M_i}[U_i])^2. \tag{45}
\]

The convergence rate \( \alpha \) cannot be calculated directly using the \( L^2 \) norm of the error in \( U_i \) and the relation (6), since the true solution \( U \) is not available. Instead, we use the standard deviation of the difference between solutions at two consecutive levels: \( \delta [U_i - U_{i+1}] \). Instead of the coefficient \( \sqrt{2(1 + 4\rho)}C_3 \) on the right-hand side of (16), we assume that there exists another constant \( C_3 \), such that

\[
\delta [U_i - U_{i+1}] = C_3 \rho_i^\alpha. \tag{46}
\]

The computation of \( \alpha \) will not be affected by the value of \( C_3 \), since

\[
\frac{\delta [U_1 - U_2]}{\delta [U_2 - U_3]} = \left( \frac{r_1}{r_2} \right) = \left( \frac{1}{2} \right)^\alpha. \tag{47}
\]

Of course, in actual calculations, the standard deviation on the left-hand side of (46) will be replaced by the square root of the unbiased sample variance [formula (45)].

The \( L^2 \) norm of the error in the first approximation \( U_1 \) cannot be calculated directly from (6) either, due to the lack of the true solution \( U \). Instead, using (16) and the convergence rate just computed, we can obtain an estimate of \( e \):

\[
e = \frac{\delta [U_1 - U_2]}{\sqrt{2(1 + 4\rho)}}. \tag{48}
\]

### 3. Numerical experiments with ACC

The ACC is a circular current surrounding the Antarctic continent. It is the primary channel through which the world’s oceans (Atlantic, Indian, and Pacific) communicate. Thanks to the predominantly westerly wind in that region, the current flows from west to east. The ACC is the strongest current in the world, volume-wise. It is estimated that the volume transport is about 135 Sv (1 Sv = \( 10^6 \) m\(^3\) s\(^{-1}\)) through the Drake Passage (Gent et al. 2001; Hughes et al. 1999; Warren et al. 1996), which is about 135 times the total volume transport of all the rivers in the world. The above estimate is a time average; the actual volume transport oscillates on seasonal and interdecadal scales.

Here, we demonstrate how the MLMC method can be combined with an ocean circulation model to quantify the volume transport of the ACC. To stay focused on the methodology that is being explored here, we sharply reduce the physics of this problem while still retaining its essential features. The fluid domain is a re-entrant rectangle that is 2000 km long and 1733 km wide, the same size used in Chen et al. (2016); see also McWilliams and Chow (1981). The flow is governed by a three-layer isopycnal model, which reads...
The prognostic variables $h$, $u$, and $\sigma$ denote the layer thickness, horizontal velocity, and some tracer, respectively. The diagnostic variables $q$, $\phi$, and $K$ denote the potential vorticity, Montgomery potential, and kinetic energy, respectively, and they are defined as

\[
q_i = \frac{\nabla \times u_i + f}{h_i}, \quad i = 1, 2, 3, \\
K_i = \frac{1}{2} |u_i|^2, \quad i = 1, 2, 3, \\
\phi_1 = p_0 + \rho_i g(h_1 + h_2 + h_3 + b), \\
\phi_2 = \phi_1 + (p_2 - p_1) g(h_2 + h_3 + b), \\
\phi_3 = \phi_2 + (p_3 - p_2) g(h_3 + b),
\]

with $p_0$ denoting the surface pressure and $b$ the bottom topography; $D_i$ denotes the horizontal viscous diffusion, which usually takes the form of harmonic or biharmonic diffusion. The external forcing term $F_i$ for each layer is specified as follows:

\[
F_i = \begin{cases} 
\frac{\tau}{\rho_i h_i} (\text{wind stress}), & i = 1, \\
0, & i = 2, \\
-\mathbf{d} (\text{bottom drag}), & i = 3.
\end{cases}
\]

The model is made up of three isopycnal layers with mean layer thickness of 500, 1250, and 3250 m and with densities of 1010, 1013, and 1016 kg m$^{-3}$. The system is forced by a zonal wind stress on the top layer with the form

\[
\tau = \tau_0 \sin^2 \left( \frac{\pi y}{L_y} \right),
\]

where $\tau_0 = 0.1$ N m$^{-2}$. In the context of incomplete knowledge about the ocean floor, we assume that the bottom topography of the domain is largely flat with small but random features:

\[
b = \sum_{k,j=1}^{20} \frac{H}{k^2 + F} \left[ a_{kl}(\omega) \cos \left( \frac{2\pi k x}{L_x} \right) + b_{kl}(\omega) \sin \left( \frac{2\pi k x}{L_x} \right) \right] \sin \left( \frac{\pi y}{L_y} \right),
\]

where $a_{kl}(\omega)$ and $b_{kl}(\omega)$ are random variables. Thus, the bottom is controlled by 578 random parameters. One sample of the topography is shown in Fig. 1. Adding random noises using truncated Fourier series is very common in the literature (see Treguier and McWilliams 1990). This approach allows one to add random perturbations at different scales without making the data discontinuous, which would cause trouble for the numerical schemes.

The numerical simulations are conducted using the Model Prediction across Scales (MPAS) isopycnal ocean model (Ringler et al. 2013). MPAS implements a C-grid finite difference/finite volume scheme that is detailed in Thuburn et al. (2009) and Ringler et al. (2010). MPAS utilizes arbitrarily unstructured Delaunay-Voronoi tessellations (Du et al. 1999, 2003). For this experiment, we have four levels of resolutions available: 10, 20, 40, and 80 km. To account for the effect of the unresolved eddies, the biharmonic hyperviscosity is used. The viscosity parameters are chosen to minimize the diffusive effect while still ensuring a stable simulation. For the aforementioned resolutions, the viscosity parameters are $10^9$, $10^{10}$, $10^{11}$, and $10^{12}$ m$^2$ s$^{-1}$, respectively. At the coarsest resolution (80 km), the Gent–McWilliams (GM) closure (Gent and McWilliams 1990; Gent et al. 1995) is turned on, with a constant parameter $400$ m$^2$ s$^{-1}$ to account for the cross-channel transport and to prevent the top fluid layer thickness from thinning to zero. GM is not used in any other
higher-resolution simulations. The configurations for each mesh resolution are summarized in Table 2. Each simulation is run for 40 years to spin up the current. The output data are saved every 10 days for the next 10 years.

There is not a universally accepted definition for turbulence (Frisch 1995; Tennekes and Lumley 1972). The current work adopts a broad view on this concept and characterizes turbulent flows by the rapid mixings and interactions among a wide range of motion scales. The interior of large-scale geophysical flows has Reynolds numbers on the order of $10^{20}$. Thus, the large-scale geophysical flows are turbulent in nature, and mesoscale and submesoscale eddy activities are important parts of the ocean dynamics (McWilliams 1985; Danabasoglu et al. 1994; Lévy et al. 2001; Holland 2010; Brannigan et al. 2015). For turbulent flows, the pointwise instantaneous behavior of the flow is not reliable anymore. But it is expected that observing the flow long enough can reveal reliable and useful statistics about the flow. Figure 2 shows the snapshots of the relative vorticity field in year 40 for different mesh resolutions with the same bottom topography profile. The highest resolution, 10 km (Fig. 2a), depicts a scene of rapid mixing by a wide range of mesoscale and submesoscale eddies, which are considered part of the turbulence (see, e.g., Lévy et al. 2012). As the mesh gets coarser, the level of eddy activities decreases. The comparison also makes

| Eddy closures | Spatial DOF | Time step (s) | Processes |
|---------------|-------------|--------------|-----------|
| 10 km         | Hyperviscosity | 480 000     | 45        | 64        |
| 20 km         | Hyperviscosity | 120 000     | 90        | 16        |
| 40 km         | Hyperviscosity | 30 000      | 180       | 4         |
| 80 km         | Hyper. + GM   | 7500        | 360       | 1         |

Fig. 2. The snapshots of the vorticity field at year 40, computed with the random bottom topography sample 1.
it clear that these flows are largely independent of each other, for there appears to be no correlation among the basic flow patterns of these simulations, other than the fact that they are all west–east flows driven by a common wind stress. However, a comparison of the volume transport by these simulations over a common set of topographic profiles tells a different and reassuring story. In Fig. 3, each curve represents results on one mesh resolution. On any particular topography profile, the results from different resolutions do not match exactly. But the curves across all 20 samples, especially those for 10, 20, and 40 km, follow a similar pattern. The agreement of the patterns of the curves indicates that a great deal of information in the curve for the highest resolutions is actually available in the curves of the lower resolutions, and this validates the multilevel method that we are pursuing here.

To set up the MLMC simulations under the various strategies proposed before, three key parameters are needed: the standard deviation $\delta[U]$ in the true solution, the error $e$ in the finest solutions, and the convergence rate $\alpha$. Fully determining these key parameters requires the true solution $U$, which is not available in any practical applications. But it can be easily estimated (see section 2c). Using data from the 20-km simulations, we compute the MC mean and the standard deviation according to the formulas (4) and (45). To probe the sensitivity of these estimates to the sample sizes, we compute the quantities with several independent sample sets with varying sizes, and the results are shown in Fig. 4 (left panel). Based on this figure, we take

$$\delta[U] \approx 7.36 \times 10^7 \text{ m}^3 \text{ s}^{-1}.$$  

Using the formula (47) and data from 10-, 20-, 40-, and 80-km simulations (Fig. 4, right panel), we also find an approximation to the convergence rate $\alpha$:

$$\alpha \approx 1.07.$$  

This convergence appears slow, but it is not unexpected for long-term simulations of turbulent flows. The underlying numerical scheme, namely, a C-grid finite volume scheme, has been found to be accurate of orders $1 \sim 2$ for laminar flows (Ringler et al. 2010; see also Chen et al. 2013). Finally, the error $e$ in the finest solutions is calculated using the formula (48) and data from the 10- and 20-km simulations:

$$e \approx 9.60 \times 10^6 \text{ m}^3 \text{ s}^{-1}.$$  

Using these estimated parameters and the formulas set forth under various strategies proposed in the previous

---

**Fig. 3.** The changes of volume transport across a subset of the sample space.

**Fig. 4.** Sample mean and variance and the convergence rate. The mean and variance are estimated using samples from the 20-km simulations.
Table 3. The setup of the MLMC under different strategies. The sample sizes at level 4 (80 km) are not calculated because data at this resolution are not needed in computing the ensemble mean under any of the strategies.

| Strategy   | $L$ | $m_1$ | $m_2$ | $m_3$ | $m_4$ | Computing load |
|------------|-----|-------|-------|-------|-------|----------------|
| Classical MC | —   | 59    |       |       |       | 59             |
| Strategy 1  | 3   | 11    | 48    | 210   |       | 22.4           |
| Strategy 2  | 2   | 11    | 191   |       |       | 36.3           |
| Strategy 3  | 3   | 876   | 763   | 210   |       | 1096.1         |
| Strategy 4  | 2   | 11    | 763   |       |       | 107            |

In section, we calculate the number of levels, the sample size at each level, and the computational load for each strategy for the multilevel method. The sample size and computational load for the classical Monte Carlo method are also calculated. The computational load for one single simulation at the highest resolution (lowest level) is considered one unit. The computational loads for various ensemble strategies under considerations here are calculated based on this unit. The issues of, for example, efficiency and overhead are neglected. For example, the classical Monte Carlo method requires 59 simulations at the highest resolution; therefore, its computational load is 59. The results are listed in Table 3. Several striking features are present in the results. First of all, under all the strategies, the numbers of required levels are low (2 or 3). This can be attributed to the fact that the errors in the finest solutions are high, compared to the variance in the true solution. Second, the sample sizes at the lowest level (highest resolution) are identical for strategies 1, 2, and 4 (11 for all three). This is no coincidence. A careful examination of the formulas (19), (27), and (39) reveal that at the lowest level $l = 1$, the sample sizes for strategies 1, 2, and 4 are identical and depend on the convergence rate $\alpha$ only. Thus, regardless of the actual highest resolution used, the sample sizes for this model at the lowest level will remain the same (=11) and identical for all three strategies. Finally, for strategy 3, the sample size at the highest resolution is too high and results in a computational load even higher than that of the classical Monte Carlo method. The reason is that this strategy requires an error distribution that is low at the highest resolution and high at the lowest resolution. In terms of computational load, strategy 1 is the optimal choice, followed by strategy 2. Both strategies are better than the classical Monte Carlo method. Strategy 4 is actually more costly than the classical method, due to the bloated sample size at the second level.

We proceed to calculate the estimates and the corresponding errors using the classical MC method and the MLMC method under strategies 1 and 2. The results from the classical MC can serve as a reference, since by the analysis of section 2, this error should be the smallest. Strategies 3 and 4 are not used due to the sheer sizes of their computational loads. The classical MC and both strategies 1 and 2 produce similar estimates of $\sim 3.58 \times 10^8$ m$^3$ s$^{-1}$ for the volume transport (second column of Table 4). The error for each method is listed in the third column. The result of the classical MC method has an error of about 5.4%. The error for strategy 1 is the largest, about 13.5%. This is expected because the error for this strategy depends on the number of levels [see (23)], which is higher than that for strategy 2.

The primary advantage of MLMC is efficiency. The first indicator for efficiency is, of course, the computational load that each method will incur, listed in Table 3. These numbers are the theoretical computational loads for the ensemble methods, and other practical factors, such as the computational overhead and parallelization, are not considered. Here, we examine the actual efficiency for each method. First, we look at the total CPU hours used by each method (fourth column of Table 4). The MLMC with strategy 1 uses the least amount of CPU time, 45 917 CPU hours, a savings of 71%, compared with the MC method. Strategy 2 uses 68 457 CPU hours, a savings of 67%.

The savings of MLMC strategies in the actual CPU times are largely in line with the savings in the computational loads (Table 3) and slightly more dramatic. This is due to the increased CPU efficiency with the MLMC methods. It is well known that MC methods are easy to parallelize and therefore are highly scalable on supercomputers. The MLMC has the potential to increase the efficiency over the classical MC even further.

Table 4. Comparison between the classical MC and the MLMC under strategies 1 and 2. The efficiency is calculated as total DOF/total CPU time. Total DOF is calculated as spatial DOF $\times$ time steps $\times$ number of samples.

|          | Estimated volume transport (m$^3$ s$^{-1}$) | Estimated error (m$^3$ s$^{-1}$) | CPU time (h) | CPU efficiency [DOF/CPU (s)] |
|----------|------------------------------------------|---------------------------------|--------------|-------------------------------|
| Classical MC | $3.58 \times 10^8$ | $1.92 \times 10^7$ | 158 446 | 343 172 |
| Strategy 1  | $3.55 \times 10^8$ | $4.80 \times 10^7$ | 45 917 | 449 719 |
| Strategy 2  | $3.57 \times 10^8$ | $3.84 \times 10^7$ | 68 457 | 488 013 |
by running more small-sized simulations and fewer large simulations. Because of the large sizes of the computations in this project, we are not able to perform an actual scalability analysis, which involves running the experiment with different numbers of total available processes. However, we can indirectly examine the issue of scalability by comparing the efficiency for each CPU core for the methods considered here (last column of Table 4). Compared with the classical MC, strategy 1 increases the CPU efficiency by 31%, and strategy 2 increases the CPU efficiency even more, by 42%.

4. Discussion

The success of the MLMC method relies on a crucial assumption, namely, that a lot of the information contained in high-resolution simulations is also available from low-resolution simulations under identical or similar model configurations. The higher the correlation, the better the MLMC method will work. For steady-state or laminar flows, especially when the simulations are backed up by rigorous error estimates, the correlation between high- and low-resolution solutions is high and quantifiable, and the MLMC method works very well (see references cited in introduction). For long-term simulations of turbulent flows, the situation is different. It is known that the pointwise behaviors of high- and low-resolution solutions of turbulent flows are uncorrelated (Fig. 2). In most engineering or geophysical applications, pointwise behaviors of turbulent flows are of little interest. What is important are certain aggregated quantities, such as mean SST. The questions of whether these aggregated quantities are correlated across different resolutions and whether the MLMC method can be used to save computation times then naturally arise. Through an experiment with the Antarctic Circumpolar Current, the present work gives affirmative answers to both of these questions. The conclusions drawn in this work cannot be generalized universally to all turbulent flows because, after all, there are no universal theories for turbulent flows yet. But it is reasonable to expect that the same results should hold in similar situations. Specifically, the MLMC method can be effective in saving computation times when the QoI demonstrates a certain level of correlation across different resolutions.

Another objective of this paper is to explore how the MLMC simulation can be set up. Four different strategies are presented, based on the desired error distributions. For all strategies discussed, the higher the convergence rate, the faster the computational cost will grow. This sounds counterintuitive. Here, the focus is on how fast the total computational cost will grow in terms of the computational cost of a single high-resolution simulation (e.g., linearly or quadratically). Of course, for the same highest resolution, a higher convergence rate will eventually lead to more accurate results, and the associated higher computational cost is a price paid for this higher accuracy.

The four strategies discussed in this work correspond to different distribution profiles of the errors across the levels. Strategy 1 corresponds to a constant distribution profile across all levels, and strategy 2 corresponds to an exponential profile. Strategies 3 and 4 also result from constant distribution profiles (similar to strategy 1), but the sample sizes are amplified by a level-dependent power function either at high resolutions (strategy 3) or at low resolutions (strategy 4). Among the four strategies, strategy 3, which inflates the sample sizes at high resolutions, seems to be of little use because of the unreasonably high cost. Strategy 1 is the most natural choice, but it may lead to a bigger margin of error if the number of levels is high. In that situation, strategies 2 and 4 can be used. In our experiment with the ACC, the highest resolution has 40 000 grid points, and strategy 2 outperforms strategy 4 with a lower computational cost. But it should be kept in mind that even at a very modest convergence rate ($\alpha < 1/2$), the total computational cost for an ensemble simulation under strategy 2 grows polynomially with respect to the computational cost for a single high-resolution simulation. Therefore, it is conceivable that as higher resolutions are taken into use, strategy 4 will eventually outperform strategy 2.

The sampling issue was treated as an optimization problem in Giles (2015): find the sampling that achieves the given accuracy goal with the smallest computational cost. However, in real applications, such as simulations of turbulent flows, many parameters (e.g., the variance in the quantity of interest and the convergence rate of the numerical scheme) are not available, and finding the truly “optimal” sampling is not possible. The current work instead aims to identify a few simple and practical strategies that allow the users to set up an MLMC ensemble simulation. Then, it is natural to ask how our sampling strategies resemble or differ from the optimal sampling of Giles (2015), assuming that all the necessary parameters are available. The answer is that our strategies are actually very close to those of Giles (2015). The key parameters in the analysis of Giles (2015) are $\alpha$ (convergence rate of the numerical scheme), $\beta$ (convergence rate in the variance), and $\gamma$ (growth rate of the computational cost of individual simulation). For most applications, one can assume that $\beta = 2\alpha$; in our study, $\gamma$ is taken to be 3, which is appropriate for large-scale
geophysical flows. Giles (2015) identifies three scenarios concerning the growth of the overall computational cost of the MLMC, corresponding to whether $\beta > \gamma$, $\beta = \gamma$, or $\beta < \gamma$. The dividing case of $\beta = \gamma$ corresponds to the critical convergence rate of $\alpha = 3/2$ that we identify under strategies 1, 3, and 4. Strategy 2 of this study assumes not available in Giles (2015).

While our strategies might not be optimal, they excel in practicality by providing simple formulas for calculating the number of levels and the constant of proportionality for the sample size at each level, which are not available in Giles (2015).

Acknowledgments. This work was in part supported by Simons Foundation (319070 to Qingshan Chen) and NSF of China (91330104 to Ju Ming).

REFERENCES
Barth, A., C. Schwab, and N. Zollinger, 2011: Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients. Numer. Math., 119, 123–161, https://doi.org/10.1007/s00211-011-0377-0.

Brannigan, L., D. P. Marshall, A. Naveira-Garabato, and A. J. George Nurser, 2015: The seasonal cycle of submesoscale flows. Ocean Modell., 92, 69–84, https://doi.org/10.1016/j.ocemod.2015.05.002.

Bugallo, M. F., L. Martino, and J. Corander, 2015: Adaptive importance sampling in signal processing. Digital Signal Process., 47, 36–49, https://doi.org/10.1016/j.dsp.2015.05.014.

Chen, Q., T. Ringler, and M. Gunzburger, 2013: A co-volume scheme for the rotating shallow water equations on conforming non-orthogonal grids. J. Comput. Phys., 240, 174–197, https://doi.org/10.1016/j.jcp.2013.01.003.

——, ——, and P. R. Gent, 2016: Extending a potential vorticity transport eddy closure to include a spatially-varying coefficient. Comput. Math. Appl., 71, 2206–2217, https://doi.org/10.1016/j.camwa.2015.12.041.

Danabasoglu, G., J. McWilliams, and P. Gent, 1994: The role of mesoscale tracer transports in the global ocean circulation. Science, 264, 1123–1126, https://doi.org/10.1126/science.264.5162.1123.

Dick, J., F. Y. Kuo, and I. H. Sloan, 2013: High-dimensional integration: The quasi-Monte Carlo way. Acta Numer., 22, 133–288, https://doi.org/10.1017/S0962492913000044.

Du, Q., V. Faber, and M. Gunzburger, 1999: Centroidal Voronoi tessellations: Applications and algorithms. SIAM Rev., 41, 637–676, https://doi.org/10.1137/00031614599352836.

——, M. D. Gunzburger, and L. Ju, 2003: Constrained centroidal Voronoi tessellations for surfaces. SIAM J. Sci. Comput., 24, 1488–1506, https://doi.org/10.1137/S1064827501391576.
McWilliams, J. C., 1985: Submesoscale, coherent vortices in the ocean. Rev. Geophys., 23, 165–182, https://doi.org/10.1029/RG023i002p00165.

——, and J. H. S. Chow, 1981: Equilibrium geostrophic turbulence I: A reference solution in a b-plane channel. J. Phys. Oceanogr., 11, 921–949, https://doi.org/10.1175/1520-0485(1981)011<0921:EGTIAR>2.0.CO;2.

Metropolis, N., and S. Ulam, 1949: The Monte Carlo method. J. Amer. Stat. Assoc., 44, 335–341, https://doi.org/10.1080/01621459.1949.10483310.

Mishra, S., and C. Schwab, 2012: Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data. Math. Comput., 81, 1979–2018, https://doi.org/10.1090/S0025-5718-2012-02574-9.

——, ——, and J. Sukys, 2012a: Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions. J. Comput. Phys., 231, 3365–3388, https://doi.org/10.1016/j.jcp.2012.01.011.

——, ——, and ——, 2012b: Multilevel Monte Carlo finite volume methods for shallow water equations with uncertain topography in multi-dimensions. J. Comput. Phys., 231, 3365–3388, 2012-02574-9.

——, ——, and ——, 2016: Multi-level Monte Carlo finite volume methods for uncertainty quantification of acoustic wave propagation in random heterogeneous layered medium. J. Comput. Phys., 312, 192–217, https://doi.org/10.1016/j.jcp.2016.02.014.

Niederreiter, H., 1993: Random Number Generation and Quasi-Monte Carlo Methods. SIAM, 241 pp.

Ringler, T. D., J. Thuburn, J. B. Klemp, and W. C. Skamarock, 2010: A unified approach to energy conservation and potential vorticity dynamics for arbitrarily-structured C-grids. J. Comput. Phys., 229, 3065–3090, https://doi.org/10.1016/j.jcp.2009.12.007.

——, M. Petersen, R. L. Higdon, D. Jacobsen, P. W. Jones, and M. Maltrud, 2013: A multi-resolution approach to global ocean modeling. Ocean Modell., 69, 211–232, https://doi.org/10.1016/j.ocemod.2013.04.010.

Sadourny, R., and C. Basdevant, 1985: Parameterization of subgrid scale barotropic and baroclinic eddies in quasi-geostrophic models: Anticipated potential vorticity method. J. Atmos. Sci., 42, 1353–1363, https://doi.org/10.1175/1520-0469(1985)042<1353:POSSBA>2.0.CO;2.

Tennekes, H., and J. L. Lumley, 1972: A First Course in Turbulence. MIT Press, 300 pp.

Thuburn, J., T. Ringler, W. Skamarock, and J. Klemp, 2009: Numerical representation of geostrophic modes on arbitrarily structured C-grids. J. Comput. Phys., 228, 8321–8335, https://doi.org/10.1016/j.jcp.2009.08.006.

Treguier, A., and J. McWilliams, 1990: Topographic influences on wind-driven, stratified flow in a b-plane channel: An idealized model for the Antarctic Circumpolar Current. J. Phys. Oceanogr., 20, 321–343, https://doi.org/10.1175/1520-0485(1990)020<0321:TIOWDC>2.0.CO;2.

Warren, B. A., J. H. LaCasce, and P. E. Robbins, 1996: On the obscurantist physics of “form drag” in theorizing about the circumpolar current. J. Phys. Oceanogr., 26, 2297–2301, https://doi.org/10.1175/1520-0485(1996)026<2297:OTOPOD>2.0.CO;2.

Wesseling, P., 1992: An Introduction to Multigrid Methods. John Wiley & Sons, 284 pp.