

# A multifidelity control variate approach for the multilevel Monte Carlo technique

By G. Geraci, M. Eldred<sup>†</sup> AND G. Iaccarino

## 1. Motivation and objectives

Numerical studies of any physical phenomenon or engineering device cannot leave aside the quantification of uncertainties on the simulations' outputs. Therefore, Uncertainty Quantification (UQ) has emerged in recent years as a fundamental tool for the scientific computing community. The propagation of uncertainties for real-case applications remains challenging due to the presence of discontinuous responses, high-gradients or in general high non-linearities. Very often all these features are associated with high dimensionality of the parameter space. In such situations it appears very difficult to efficiently apply established deterministic approaches for UQ as, for instance, stochastic collocation or generalized Polynomial Chaos approaches (Le Maître & Knio 2010). Despite the improvements obtained in recent years through the application of different techniques to mitigate the curse of dimensionality associated with these algorithms, as for instance in (Doostan & Owhadi 2011; Tryoen *et al.* 2012), in high-dimensional spaces their applicability remains limited. More recently, statistical sampling methods, which historically were among the first ones to be proposed, have gained popularity. The seminal Monte Carlo (MC) strategy has been generalized to multilevel approaches leading to the so-called Multilevel MC method (MLMC). The MLMC method was first introduced in the context of numerical quadrature (Heinrich 2001) and subsequently generalized by Giles to enhance the MC path simulation for stochastic differential equations in finance (Giles 2008). More recently, MLMC has been applied to a variety of problems ranging from elliptic, parabolic and hyperbolic problems (Cliffe *et al.* 2011; Barth *et al.* 2011; Mishra *et al.* 2012). The MLMC method has been demonstrated to be superior to MC in all possible scenarios, and it is a good candidate for many applications which actually are intractable by other UQ methods. Despite all these advantages, the number of simulations required remains high, even if they are more favorable compared to other techniques. The goal of this brief is to show a first effort to reduce the overall computational burden through the use of a variety of hierarchies of models. In particular, if it is possible to introduce a low-cost computational model, albeit less accurate, it is virtually conceivable to draw a large number of its realizations to decrease the variability of the MLMC estimator. To present this idea the remaining of the brief is organized as follows. In Section 2 the MC method is briefly introduced and the notation used throughout the paper is established. Section 3 is devoted to the introduction of the MLMC method which is extended to the multilevel multifidelity (MLMF) algorithm in Section 4. Numerical results for a parabolic partial differential equation describing the evolution of a temperature profile along a rod, making comparisons between MC, MLMC and MLMF, are reported in Section 5. Finally, conclusions and perspectives are drawn in Section 6.

<sup>†</sup> Sandia National Laboratories, Albuquerque NM

## 2. Monte Carlo and control variate

Monte Carlo is definitely the most popular algorithm for stochastic simulations. The reasons of its use are its extreme flexibility and its simplicity, flexibility and the provable convergence behavior that is independent of the number of input uncertainties. These aspects often make MC the only feasible way to handle realistic stochastic simulations. A quantity of interest  $Q : \Xi \rightarrow \mathbb{R}$ , represented as a random variable (RV), can be introduced as a function of a random vector  $\boldsymbol{\xi} \in \Xi \subset \mathbb{R}^d$ . The goal of any MC simulation is computing statistics for  $Q$ . In this work we focus on the computation of the expected value  $\mathbb{E}[Q]$  for  $Q = Q(\boldsymbol{\xi})$ ; however, a similar derivation is possible for other statistical moments and other properties of  $Q$ . The MC estimator  $\hat{Q}_N^{MC}$  for  $\mathbb{E}[Q]$  is defined as follows

$$\hat{Q}_N^{MC} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N Q^{(i)}, \quad (2.1)$$

where  $Q^{(i)} \stackrel{\text{def}}{=} Q(\boldsymbol{\xi}^{(i)})$  and  $N$  is used to indicate the number of realizations of the model. Some important properties of an estimator can be defined by introducing some definitions:

- Error (as function of the number of simulations  $N$ ):  $\varepsilon_N = \mathbb{E}[Q] - \hat{Q}_N^{MC}$ ;
- Bias is the distance with respect to the true statistic:  $\text{Bias}[\hat{Q}_N^{MC}] = \mathbb{E}[\varepsilon_N]$ ;
- Thus, the root-mean-square error follows:  $\text{RMSE}[\hat{Q}_N^{MC}] = \sqrt{\mathbb{E}[\varepsilon_N^2]}$ .

For large  $N$  it is trivial to demonstrate that the MC estimator is unbiased, i.e., its bias is zero and its convergence to the true statistics is  $\mathcal{O}(N^{-1/2})$ . Moreover, since each sequence of realizations is different, another crucial property of any estimator is its own variance. For  $\hat{Q}_N^{MC}$  it is

$$\text{Var}(\hat{Q}_N^{MC}) = \text{Var}\left(\frac{1}{N} \sum_{i=1}^N Q^{(i)}\right) = \frac{1}{N^2} \sum_{i=1}^N \text{Var}(Q) = \frac{\text{Var}(Q)}{N}. \quad (2.2)$$

Furthermore, by means of the central limit theorem it is possible to show that the error  $\varepsilon_N / \sqrt{\text{Var}(\hat{Q}_N^{MC})}$  is proportional to a standardized normal distributed RV  $\mathcal{N}(0, 1)$ . As a consequence, it is possible to define a confidence interval for the MC estimator which has an amplitude proportional to the standard deviation of the estimator. Without any loss of generality, in the following the 99.7% confidence interval is employed as  $\hat{Q}_N^{MC} \pm 3\text{Var}^{1/2}(\hat{Q}_N^{MC})$ . Indeed, the variance of the estimator plays a fundamental role in the quality of the numerical results, and their usefulness, for a MC simulation. The reduction of the variance (for a fixed computational cost) is a very effective way of improving the quality of the MC prediction. A closer inspection of Eq. (2.2) indicates that only an increase in the number of simulations  $N$  will reduce the overall variance, since  $\text{Var}(Q)$  is an intrinsic property of the model under analysis. However, more sophisticated techniques have been proposed to accelerate the convergence of a MC simulation. For instance, an incomplete list of these techniques can include stratified sampling, importance sampling, Latin hypercube, deterministic Sobol' sequences and control variate (see Liu, J.S. 2001 for more details). The work proposed in this brief is related to the control variate approach, which, in turn, is based on the idea of replacing the RV  $Q$  with a different one which has the same expected value, but with a smaller variance. The goal is to reduce the numerator in Eq. (2.2), and hence the value of the variance without requiring a larger number of simulations. In a practical setting, the control variate makes use of an auxiliary functional

$G = G(\boldsymbol{\xi})$  for which the expected value is known. Indeed, the alternative estimator can be defined as

$$\hat{Q}_N^{MCCV} = \hat{Q}_N^{MC} - \beta \left( \hat{G}_N^{MC} - \mathbb{E}[G] \right). \quad (2.3)$$

The MC control variate estimator  $\hat{Q}_N^{MCCV}$  is unbiased (irrespective of the value of the parameter  $\beta \in \mathbb{R}$ ), but its variance now has a more complex dependence not only on the  $\text{Var}(Q)$ , but also on  $\text{Var}(G)$  and the covariance between  $Q$  and  $G$  since

$$\text{Var} \left( \hat{Q}_N^{MCCV} \right) = \frac{1}{N} \left( \text{Var}(Q) + \beta^2 \text{Var}(G) - 2\beta \text{Cov}(Q, G) \right). \quad (2.4)$$

The parameter  $\beta$  can be used to minimize the value of the variance  $\text{Var} \left( \hat{Q}_N^{MCCV} \right)$

requiring  $\frac{d \text{Var} \left( \hat{Q}_N^{MCCV} \right)}{d\beta} = 0$ . Simple manipulations lead to  $\beta = \frac{\text{Cov}(Q, G)}{\text{Var}(G)}$ , obtaining

$$\text{Var} \left( \hat{Q}_N^{MCCV} \right) = \text{Var} \left( \hat{Q}_N^{MC} \right) (1 - \rho^2), \quad (2.5)$$

for the estimator. Therefore, the overall variance of the estimator  $\hat{Q}_N^{MCCV}$  is proportional to the variance of the standard MC estimator  $\hat{Q}_N^{MC}$  through a factor  $1 - \rho^2$  where

$\frac{\text{Cov}(Q, G)}{\text{Var}^{1/2}(Q)\text{Var}^{1/2}(G)}$  is the Pearson correlation coefficient between  $Q$  and  $G$ . Since  $0 < \rho^2 < 1$ , the variance  $\text{Var} \left( \hat{Q}_N^{MCCV} \right)$  is always less than the corresponding  $\text{Var} \left( \hat{Q}_N^{MC} \right)$ .

### 3. Multilevel MC technique

In the previous section the MC has been introduced as referring to a scalar functional  $Q$  for which one aims to compute statistics. However, if this quantity of interest (QoI) is governed by a generic partial differential equation (PDE), as happens in many real applications, the dependence on the physical  $\boldsymbol{x} \in \Omega \subset \mathbb{R}^n$  and/or temporal  $t \in T \subset \mathbb{R}^+$  coordinates should be included, hence  $Q = Q(\boldsymbol{x}, \boldsymbol{\xi}, t)$ . A finite spatial/temporal resolution is always employed to solve numerically a PDE, implying the presence of a discretization error in addition to the stochastic error  $\varepsilon_N$ . In the remainder of this brief, the term discretization is used with reference to the spatial tessellation. For a generic tessellation with  $M$  degrees-of-freedom (DOFs), the PDE solution of  $Q$  is referred to as  $Q_M$ . Since  $Q_M \rightarrow Q$  for  $M \rightarrow \infty$ , then  $\mathbb{E}[Q_M] \rightarrow \mathbb{E}[Q]$  for  $M \rightarrow \infty$  with a prescribed order of convergence. Applying the MC estimator Eq. (2.1) to  $Q_M$  will reflect the presence of a finite spatial resolution in addition to a finite sampling. Indeed, the MC estimator for  $Q_M$ ,

$$\hat{Q}_{M,N}^{MC} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N Q_M^{(i)} \quad (3.1)$$

has a mean square error MSE which is

$$\mathbb{E} \left[ \left( \hat{Q}_{M,N}^{MC} - \mathbb{E}[Q] \right)^2 \right] = N^{-1} \text{Var}(Q_M) + \left( \mathbb{E}[Q_M - Q] \right)^2, \quad (3.2)$$

where the first term represents the variance of the estimator, whereas  $\left( \mathbb{E}[Q_M - Q] \right)^2$  is the bias introduced by the spatial discretization. The two contributions appear to be independent of each other; an accurate result of a MC simulation can be obtained by only

drawing the required  $N$  number of simulations of  $Q_M(\boldsymbol{\xi})$  at a fine enough resolution  $M$ . Since the numerical cost of a PDE is related to the number of DOFs of the tessellation, the total cost of a MC simulation for a PDE can easily become intractable for complex multiphysics computationally intensive applications. The multilevel idea has been introduced as an extension of the control variate approach when an additional (with respect to the stochastic space) discretization space is present. The basic idea, borrowed from the multigrid approach, is to replace the evaluation of the statistics of  $Q_M$  with a sequence of evaluations at coarser levels. If it is possible to define a sequence of discretization levels†  $\{M_\ell : \ell = 0, \dots, L\}$  with  $M_0 < M_1 < \dots < M_L \stackrel{\text{def}}{=} M$ , the expected value  $\mathbb{E}[Q_M]$  can be decomposed, exploiting the linearity of the expected value operator as

$$\mathbb{E}[Q_M] = \mathbb{E}[Q_{M_0}] + \sum_{\ell=1}^L \mathbb{E}[Q_{M_\ell} - Q_{M_{\ell-1}}]. \quad (3.3)$$

If the difference function  $Y_\ell$  is defined according to

$$Y_\ell = \begin{cases} Q_{M_0} & \text{if } \ell = 0 \\ Q_{M_\ell} - Q_{M_{\ell-1}} & \text{if } 0 < \ell \leq L, \end{cases} \quad (3.4)$$

the expected value  $\mathbb{E}[Q_M] = \sum_{\ell=0}^L \mathbb{E}[Y_\ell]$ . A multilevel MC estimator is obtained when a MC estimator is written at each level for  $Y_\ell$ , obtaining the following multilevel MC estimator  $\hat{Q}_M^{\text{ML}}$

$$\hat{Q}_M^{\text{ML}} \stackrel{\text{def}}{=} \sum_{\ell=0}^L \hat{Y}_{\ell, N_\ell}^{\text{MC}} = \sum_{\ell=0}^L \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} Y_\ell^{(i)}. \quad (3.5)$$

Since it is unbiased, the advantage of using this estimator instead of the MC one is evident once its MSE is derived

$$\begin{aligned} \mathbb{E} \left[ (\hat{Q}_M^{\text{ML}} - \mathbb{E}[Q])^2 \right] &= \sum_{\ell=0}^L \text{Var} \left( \hat{Y}_{\ell, N_\ell}^{\text{MC}} \right) + (\mathbb{E}[Q_M - Q])^2 \\ &= \sum_{\ell=0}^L N_\ell^{-1} \text{Var}(Y_\ell) + (\mathbb{E}[Q_M - Q])^2. \end{aligned} \quad (3.6)$$

The bias term  $(\mathbb{E}[Q_M - Q])^2$  corresponds to that obtained by using the MC estimator, while the part related to the sampling contribution is now divided over the whole set of levels. Since  $Q_M \rightarrow Q$ , then the difference function  $Y_\ell \rightarrow 0$ . Indeed, the corresponding number of samples  $N_\ell$  required to decrease the variance associated with the  $\ell$ th level is expected to decrease with  $\ell$ . Because the computational cost is proportional to the number of DOFs, reducing the number of samples to draw for high levels lowers the total computational cost. Moreover, achieving higher accuracy is possible by adding more levels. The MLMC algorithm is based on the repartition of the numerical cost between different levels, thus a key aspect is played by the choice of  $N_\ell$ . If the computational cost at each level is  $\mathcal{C}_\ell$ , the total cost of the MLMC algorithm is

$$\mathcal{C}(\hat{Q}_M^{\text{ML}}) = \sum_{\ell=0}^L N_\ell \mathcal{C}_\ell. \quad (3.7)$$

† In the following, this kind of discretization, spatial and/or temporal, is referred to as being geometrical.

The computational cost at each level (for  $\ell > 0$ ) is the sum of the solution of the cost of evaluating  $Q_{M_\ell}^{(i)}$  and  $Q_{M_{\ell-1}}^{(i)}$  since this cost is the effective cost of evaluating a single realization for  $Y_\ell^{(i)}$ . Fixing an accuracy equal to  $\varepsilon^2/2$  for the MSE related to the sampling part (the remaining  $\varepsilon^2/2$  is associated with the deterministic bias), the minimization problem is formulated as a constraint optimization by using of a Lagrange multiplier  $\lambda$

$$f(N_\ell, \lambda) = \sum_{\ell=0}^L N_\ell C_\ell + \lambda \left( \sum_{\ell=0}^L N_\ell^{-1} \text{Var}(Y_\ell) - \varepsilon^2/2 \right). \quad (3.8)$$

The result of the minimization is

$$N_\ell = \frac{2}{\varepsilon^2} \left[ \sum_{k=0}^L (\text{Var}(Y_k) C_k)^{1/2} \right] \sqrt{\frac{\text{Var}(Y_\ell)}{C_\ell}}. \quad (3.9)$$

A practical implementation of the overall MLMC algorithm is reported in Appendix A as Algorithm 1.

#### 4. A multifidelity approach for MLMC

The MLMC can be considered as a recursive control variate technique since it aims at reducing the variance of the target function as an effective way of limiting the sampling at high resolution. Nonetheless, at each level the difference function  $Y_\ell$  can itself be the object of an additional control variate. Despite the generality of the idea, in this brief the analysis is reduced to cases in which it is possible to formulate the following assertions:

- the model of reference is called high-fidelity (HF) and it is represented by the one producing the evaluations of  $Q_M$  (hereafter  $Q_M^{\text{HF}}$ ) in the previous section;
- since the computational cost of HF is high, only a limited number of runs can be afforded for this model;
- a second model low-fidelity (LF) is defined by trading accuracy with computational cost. It is assumed to be inexpensive so that a high number of realizations can be taken;
- for HF and LF it is possible to independently provide discretization levels such that a MLMC can potentially be performed;
- for the sake of simplicity, the number of geometrical levels is equal for HF and LF and is identified as  $L_{\text{HF}}$ .

The pivotal idea in the multilevel multifidelity approach proposed in this brief is to use the HF model to control the part of the MSE subject to the deterministic bias, while the low computational cost of the LF is exploited to run a very large number of realizations to obtain a low variance estimator. This goal can be accomplished if the following estimator (based on the MLMC approach described in the previous section) is formulated

$$\mathbb{E}[Q_M^{\text{HF}}] = \sum_{l=0}^{L_{\text{HF}}} \mathbb{E}[Y_\ell^{\text{HF}}] = \sum_{l=0}^{L_{\text{HF}}} \hat{Y}_\ell^{\text{HF}} = \sum_{l=0}^{L_{\text{HF}}} Y_\ell^{\text{HF},*}, \quad (4.1)$$

where

$$Y_\ell^{\text{HF},*} = \hat{Y}_\ell^{\text{HF}} + \alpha_\ell \left( \hat{Y}_\ell^{\text{LF}} - \mathbb{E}[Y_\ell^{\text{LF}}] \right). \quad (4.2)$$

The last equality in Eq. (4.1) is possible because the estimator  $Y_\ell^{\text{HF},*}$  is unbiased with respect to  $\hat{Y}_\ell^{\text{HF}}$ , hence with respect to the true value  $\mathbb{E}[Y_\ell^{\text{HF}}]$ . The control variate is obtained by means of the LF realizations of the model for which the expected value can

be computed in two different ways:  $\hat{Y}_\ell^{\text{LF}}$  and  $\mathbb{E}[Y_\ell^{\text{LF}}]$ . In particular, a MC estimator is employed for each term ( $\hat{Y}_\ell^{\text{HF}}$ ,  $\hat{Y}_\ell^{\text{LF}}$  and  $\mathbb{E}[Y_\ell^{\text{LF}}]$ ) such that the estimation of  $\hat{Y}_\ell^{\text{LF}}$  and  $\mathbb{E}[Y_\ell^{\text{LF}}]$  are different. The number of LF realizations is chosen to be equal to the sum of an affordable number of HF realizations,  $N_\ell^{\text{HF}}$ , and an additional and independent set of realizations  $\Delta_\ell^{\text{LF}}$ , hence  $N_\ell^{\text{LF}} = N_\ell^{\text{HF}} + \Delta_\ell^{\text{LF}}$ . The set  $\Delta_\ell^{\text{LF}}$  is written as proportional to  $N_\ell^{\text{HF}}$  by means of a parameter  $r_\ell \in \mathbb{R}_0^+$

$$N_\ell^{\text{LF}} = N_\ell^{\text{HF}} + \Delta_\ell^{\text{LF}} = N_\ell^{\text{HF}} + r_\ell N_\ell^{\text{HF}} = N_\ell^{\text{HF}}(1 + r_\ell). \quad (4.3)$$

The set of samples  $\Delta_\ell^{\text{LF}}$  is independent of  $N_\ell^{\text{HF}}$  and, for clarity of exposure, is denoted in the following by the index  $j$ . The MC estimators for  $\hat{Y}_\ell^{\text{HF}}$ ,  $\hat{Y}_\ell^{\text{LF}}$  and  $\mathbb{E}[Y_\ell^{\text{LF}}]$  are

$$\begin{aligned} \hat{Y}_\ell^{\text{HF}} &\stackrel{\text{def}}{=} \frac{1}{N_\ell^{\text{HF}}} \sum_{i=1}^{N_\ell^{\text{HF}}} Y_\ell^{\text{HF},(i)}, \\ \hat{Y}_\ell^{\text{LF}} &\stackrel{\text{def}}{=} \frac{1}{N_\ell^{\text{HF}}} \sum_{i=1}^{N_\ell^{\text{HF}}} Y_\ell^{\text{LF},(i)}, \\ \hat{\mathbb{E}}[Y_\ell^{\text{LF}}] &\stackrel{\text{def}}{=} \frac{1}{N_\ell^{\text{LF}}} \sum_{i=1}^{N_\ell^{\text{LF}}} Y_\ell^{\text{LF},(i)} = \frac{1}{N_\ell^{\text{HF}}(1+r_\ell)} \sum_{i=1}^{N_\ell^{\text{HF}}} Y_\ell^{\text{LF},(i)} + \frac{1}{N_\ell^{\text{HF}}(1+r_\ell)} \sum_{j=1}^{r_\ell N_\ell^{\text{HF}}} Y_\ell^{\text{LF},(j)}, \end{aligned} \quad (4.4)$$

which lead to the following multilevel multifidelity (MLMF) estimator  $\hat{Q}_M^{\text{MLMF}}$

$$\begin{aligned} \hat{Q}_M^{\text{MLMF}} &= \sum_{l=0}^{L_{\text{HF}}} \left( \frac{1}{N_\ell^{\text{HF}}} \sum_{i=1}^{N_\ell^{\text{HF}}} Y_\ell^{\text{HF},(i)} + \frac{\alpha_\ell r_\ell}{N_\ell^{\text{HF}}(1+r_\ell)} \sum_{i=1}^{N_\ell^{\text{HF}}} Y_\ell^{\text{LF},(i)} \right. \\ &\quad \left. - \frac{\alpha_\ell}{N_\ell^{\text{HF}}(1+r_\ell)} \sum_{j=1}^{r_\ell N_\ell^{\text{HF}}} Y_\ell^{\text{LF},(j)} \right), \end{aligned} \quad (4.5)$$

since each term is obtained by means of an unbiased MC estimator, the overall estimator is unbiased. Its variance is obtained by considering the contributions of each term with the addition of a covariance term between the HF and LF estimators defined over the set denoted by the index  $i$

$$\begin{aligned} \text{Var}(\hat{Q}_M^{\text{MLMF}}) &= \sum_{l=0}^{L_{\text{HF}}} \left( \frac{1}{N_\ell^{\text{HF}}} \text{Var}(Y_\ell^{\text{HF}}) + \frac{\alpha_\ell^2 r_\ell^2}{(1+r_\ell)^2 N_\ell^{\text{HF}}} \text{Var}(Y_\ell^{\text{HF}}) \right. \\ &\quad \left. + \frac{\alpha_\ell^2 r_\ell}{(1+r_\ell)^2 N_\ell^{\text{HF}}} \text{Var}(Y_\ell^{\text{HF}}) \right. \\ &\quad \left. + 2 \frac{\alpha_\ell r_\ell}{(N_\ell^{\text{HF}})^2 (1+r_\ell)} \text{Cov} \left( \sum_{i=1}^{N_\ell^{\text{HF}}} Y_\ell^{\text{HF},(i)}, \sum_{i=1}^{N_\ell^{\text{HF}}} Y_\ell^{\text{LF},(i)} \right) \right) \\ &= \sum_{l=0}^{L_{\text{HF}}} \left( \frac{1}{N_\ell^{\text{HF}}} \text{Var}(Y_\ell^{\text{HF}}) + \frac{\alpha_\ell^2 r_\ell}{(1+r_\ell) N_\ell^{\text{HF}}} \text{Var}(Y_\ell^{\text{HF}}) \right. \\ &\quad \left. + 2 \frac{\alpha_\ell r_\ell}{(1+r_\ell) N_\ell^{\text{HF}}} \rho_{HL} \text{Var}^{1/2}(Y_\ell^{\text{HF}}) \text{Var}^{1/2}(Y_\ell^{\text{LF}}) \right), \end{aligned} \quad (4.6)$$

in which the following properties are employed (assuming  $a, b, c$  and  $d$  random variables and  $\alpha, \beta, \gamma$  and  $\delta$  are constant)

$$\begin{aligned} \mathbb{V}ar(\alpha a + \beta b + \gamma c) &= \alpha^2 \mathbb{V}ar(a) + \beta^2 \mathbb{V}ar(b) + \gamma^2 \mathbb{V}ar(c) \\ &\quad + 2\alpha\beta \text{Cov}(a, b) + 2\alpha\gamma \text{Cov}(a, c) + 2\beta\gamma \text{Cov}(b, c) \end{aligned} \quad (4.7)$$

and

$$\text{Cov}(\alpha a + \beta b, \gamma c + \delta d) = \alpha\gamma \text{Cov}(a, c) + \alpha\delta \text{Cov}(a, d) + \beta\gamma \text{Cov}(b, c) + \beta\delta \text{Cov}(b, d). \quad (4.8)$$

The treatment of the covariance term requires more attention since it includes summations of LF and HF realizations which are dependent two by two

$$\begin{aligned} \text{Cov}\left(\sum_{i=1}^{N_\ell^{\text{HF}}} Y_\ell^{\text{HF},(i)}, \sum_{i=1}^{N_\ell^{\text{LF}}} Y_\ell^{\text{LF},(i)}\right) &= \text{Cov}\left(Y_\ell^{\text{HF}} + \sum_{i=2}^{N_\ell^{\text{HF}}} Y_\ell^{\text{HF},(i)}, Y_\ell^{\text{LF}} + \sum_{i=2}^{N_\ell^{\text{LF}}} Y_\ell^{\text{LF},(i)}\right) \\ &= \text{Cov}(Y_\ell^{\text{HF}}, Y_\ell^{\text{LF}}) + \text{Cov}\left(\sum_{i=2}^{N_\ell^{\text{HF}}} Y_\ell^{\text{HF},(i)}, \sum_{i=2}^{N_\ell^{\text{LF}}} Y_\ell^{\text{LF},(i)}\right) \\ &= N_\ell^{\text{HF}} \text{Cov}(Y_\ell^{\text{HF}}, Y_\ell^{\text{LF}}) \\ &= N_\ell^{\text{HF}} \rho_{HL} \mathbb{V}ar^{1/2}(Y_\ell^{\text{HF}}) \mathbb{V}ar^{1/2}(Y_\ell^{\text{LF}}). \end{aligned} \quad (4.9)$$

The Pearson correlation coefficient between the HF and LF models is indicated by  $\rho_{HL}$  in the previous equations. Assuming the vector  $r_\ell$  as a parameter, level by level the variance is minimized, mimicking the standard control variate approach, and thus requiring  $\frac{d \mathbb{V}ar(Y_\ell^{\text{HF},\star})}{d \alpha_\ell} = 0$  obtaining the optimal coefficient  $\alpha_\ell = -\rho_{HL} \frac{\mathbb{V}ar^{1/2}(Y_\ell^{\text{HF}})}{\mathbb{V}ar^{1/2}(Y_\ell^{\text{LF}})}$ . By making use of the optimal coefficient  $\alpha_\ell$ , it is possible to show that the variance  $\mathbb{V}ar(Y_\ell^{\text{HF},\star})$  is proportional to the variance  $\mathbb{V}ar(Y_\ell^{\text{HF}})$  through a factor  $\Lambda_\ell(r_\ell)$ , which is an explicit function of the ratio  $r_\ell$ . The final expression for the overall variance  $\mathbb{V}ar(\hat{Q}_M^{\text{MLMF}})$  is

$$\begin{aligned} \mathbb{V}ar(\hat{Q}_M^{\text{MLMF}}) &= \sum_{l=0}^{L_{\text{HF}}} \frac{1}{N_\ell^{\text{HF}}} \mathbb{V}ar(Y_\ell^{\text{HF}}) \Lambda_\ell(r_\ell) \quad \text{where} \\ \Lambda_\ell(r_\ell) &= \left(1 - \frac{r_\ell}{1 + r_\ell} \rho_{HL}^2\right). \end{aligned} \quad (4.10)$$

Note that  $\Lambda_\ell(r_\ell)$  represents a penalty with respect to the classical control variate approach presented in Section 2, which stems from the need to evaluate the unknown function  $\mathbb{E}[Y_\ell^{\text{LF}}]$ . However, the ratio  $r_\ell/(r_\ell + 1)$  is dependent on the additional number of LF evaluations  $\Delta_\ell^{\text{LF}}$ , hence it is fair to assume that it can be made very close to the unity by choosing an affordable large  $r_\ell$ , i.e.,  $\Delta_\ell^{\text{LF}} \gg N_\ell^{\text{HF}}$ .

A practical implementation of the MLMF algorithm is the following:

1. The geometrical MLMC is performed on the low-fidelity  $\rightarrow N_\ell^{\text{LF}}$ ;
2. A maximum cost for HF is decided according to the available resources  $\rightarrow r_\ell$ ;
3. The level dependent number of HF simulations is evaluated as  $N_\ell^{\text{HF}} = N_\ell^{\text{LF}}/(1 + r_\ell)$ ;
4.  $\mathbb{E}[Y_\ell^{\text{HF}}]$  and  $\mathbb{V}ar(Y_\ell^{\text{HF}})$  are evaluated with  $N_\ell^{\text{HF}}$  realizations;

5.  $\rho_{HL}$  and  $\alpha_\ell = \alpha_\ell(\rho_{HL})$  are evaluated<sup>†</sup>

$$\rho_{HL} = \frac{\sum_{i=1}^{N_\ell^{\text{HF}}} \left( Y_\ell^{\text{HF},(i)} - \mathbb{E}[Y_\ell^{\text{HF}}] \right) \left( Y_\ell^{\text{LF},(i)} - \mathbb{E}[Y_\ell^{\text{LF}}] \right)}{\sqrt{\sum_{i=1}^{N_\ell^{\text{HF}}} \left( Y_\ell^{\text{HF},(i)} - \mathbb{E}[Y_\ell^{\text{HF}}] \right)^2 \sum_{i=1}^{N_\ell^{\text{LF}}} \left( Y_\ell^{\text{LF},(i)} - \mathbb{E}[Y_\ell^{\text{LF}}] \right)^2}};$$

6 the estimator for the mean is obtained via Eq. (4.6);

7 a corresponding confidence interval is computed as  $\hat{Q}_M^{\text{MLMF}} \pm 3 \text{Var}^{1/2}(\hat{Q}_M^{\text{MLMF}})$ , where  $\text{Var}^{1/2}(\hat{Q}_M^{\text{MLMF}})$  is obtained by using Eq. (4.10).

## 5. Numerical results

In this section two different kinds of numerical results are presented. The MLMF algorithm is compared to the classical MLMC using the same set of samples, i.e., the same number of realizations and same realizations. This comparison is useful to demonstrate that the reduced variance of the new estimator comes from the estimator itself and not from the choice of the specific realizations. However, in this case the MLMC would derive no advantage from the optimization of its number of samples. For that reason, it is referred to in the following as incomplete MLMC (iMLMC). In addition, the numerical results involving a comparison in terms of number of realizations are reported. In this latter case, the number of realizations for the MLMC method is optimized on each level following section 3. The numerical results presented in this section are based on the following partial differential equation describing the temperature evolution  $u$  in a rod  $\Omega = [0, 1]$  with uncertain initial condition and diffusivity  $\alpha$

$$\begin{cases} \frac{\partial u(x, \boldsymbol{\xi}, t)}{\partial t} - \alpha(\boldsymbol{\xi}) \frac{\partial^2 u(x, \boldsymbol{\xi}, t)}{\partial x^2} = 0, & \alpha > 0, x \in [0, L] = \Omega \subset \mathbb{R} \\ u(x, \boldsymbol{\xi}, 0) = u_0(x, \boldsymbol{\xi}), & t \in [0, t_F] \text{ and } \boldsymbol{\xi} \in \Xi \subset \mathbb{R}^d \\ u(x, \boldsymbol{\xi}, t)|_{\partial\Omega} = 0. \end{cases} \quad (5.1)$$

To obtain an analytical solution, the following initial condition (with a prescribed spectral content) is chosen

$$u_0(x, \boldsymbol{\xi}) = \mathcal{G}(\boldsymbol{\xi})\mathcal{F}_1(x) + \mathcal{I}(\boldsymbol{\xi})\mathcal{F}_2(x), \quad \text{where}$$

$$\mathcal{F}_1(x) = \sin(\pi x),$$

$$\mathcal{F}_2(x) = \sin(2\pi x) + \sin(3\pi x) + 50(\sin(9\pi x) + \sin(21\pi x)),$$

$$\mathcal{I}(\boldsymbol{\xi}) = \frac{7}{2} \left[ \sin(\xi_1) + 7 \sin^2(\xi_2) + \frac{1}{10} \xi_3^4 \sin(\xi_1) \right],$$

$$\mathcal{G}(\boldsymbol{\xi}) = \mathcal{G}(\xi_5, \xi_6, \xi_7) = 50 \frac{|4\xi_5 - 2| + a_i}{1 + a_i} \frac{|4\xi_6 - 2| + a_i}{1 + a_i} \frac{|4\xi_7 - 2| + a_i}{1 + a_i}, \quad a_i = -1/2. \quad (5.2)$$

The problem definition is complemented by specifying seven independent RVs:

- $\xi_i \sim \mathcal{U}(-\pi, \pi)$ , for  $i = 1, 2, 3$
- $\alpha = \xi_4 \sim \mathcal{U}(\alpha_{\min}, \alpha_{\max})$
- $\xi_i \sim \mathcal{U}(0, 1)$ , for  $i = 5, 6, 7$ .

<sup>†</sup> For this step,  $\mathbb{E}[Y_\ell^{\text{HF}}]$  can be obtained by  $N_\ell^{\text{LF}}$  realizations.

	$N_x$				
# modes	$\ell = 0$	$\ell = 1$	$\ell = 2$	$\ell = 3$	
<b>LF</b>	3	5	15	30	60
<b>HF</b>	21	30	60	100	200

TABLE 1. # of modes and number of points  $N_x$  for all the geometrical levels and fidelity levels

A QoI is defined as the integral of the temperature along the rod, for which the expected value is the desired statistic. Without showing all the details for the sake of brevity, the following exact condition can be obtained

$$\mathbb{E}[Q(\boldsymbol{\xi}, t_F)] = \int_{\Xi} \int_{\Omega} u(x, \boldsymbol{\xi}, t) dx p(\boldsymbol{\xi}) d\boldsymbol{\xi} = \frac{100}{\pi} H_1 + \frac{98}{12\pi} H_3 + 50 \frac{98}{36\pi} H_9 + 50 \frac{98}{84\pi} H_{21}$$

$$H_n = \frac{1}{t_F n^2 \pi^2 \Delta\alpha} [\exp(-n^2 \pi^2 \alpha_{\max} t_F) (1 - \exp(n^2 \pi^2 t_F \Delta\alpha))],$$
(5.3)

where  $\Delta\alpha = \alpha_{\max} - \alpha_{\min}$ .

Numerically, the solution is computed according to the (truncated) Fourier expansion

$$\begin{cases} u(x, \boldsymbol{\xi}, t) = \sum_{n=1}^{\#\text{modes}} A_n(\boldsymbol{\xi}) \sin\left(\frac{n\pi}{L}x\right) \exp\left(-\alpha(\boldsymbol{\xi})\left(\frac{n\pi}{L}\right)^2 t\right) \\ A_n(\boldsymbol{\xi}) = \frac{\langle u_0(x, \boldsymbol{\xi}), \sin\left(\frac{n\pi}{L}x\right) \rangle}{\langle \sin\left(\frac{n\pi}{L}x\right), \sin\left(\frac{n\pi}{L}x\right) \rangle} = \frac{2}{L} \int_{\Omega} u_0(x, \boldsymbol{\xi}) \sin\left(\frac{n\pi}{L}x\right) dx. \end{cases}$$
(5.4)

Thus, for each realization, the following procedure is followed:

- a deterministic uniform mesh containing  $N_x$  points is generated over the physical space  $\Omega$ ;
- a single realization for  $\boldsymbol{\xi}^{(i)}$  is drawn;
- a trapezoidal quadrature rule is employed to compute all the terms  $A_n(\boldsymbol{\xi}^{(i)})$  and  $\sin(n\pi x)$  using their evaluations on the  $N_x$  mesh points for all  $n \leq \#\text{modes}$ .

The modal expansion Eq. (5.4) enables us to obtain models with different fidelity, while the need for a quadrature rule over the physical space still permits distinct geometrical levels to be defined. For the following numerical results, the LF and HF parametrizations are defined as listed in Table 1.

In Figure 1, the results for a MC run of the LF and HF models are reported. As is evident, the LF is chosen to be biased with respect to the HF model so that the higher number of its realizations can improve the estimator through variance reduction only. The exact solution for HF is 41.98, whereas is only 33.15 for LF; the discrepancy between the two models accounts for 21% and it retains entirely the high-frequency content of the solution.

The MLMF and the iMLMC are compared in Figure 2. These results are obtained using  $r_\ell = [50, 5, 3, 3]$  and 20 pilot runs, per level, for the LF. The number of evaluations is computed in terms of equivalent numerical cost: the cost of each realization is evaluated as the product of the number of modes times the number of points of the deterministic

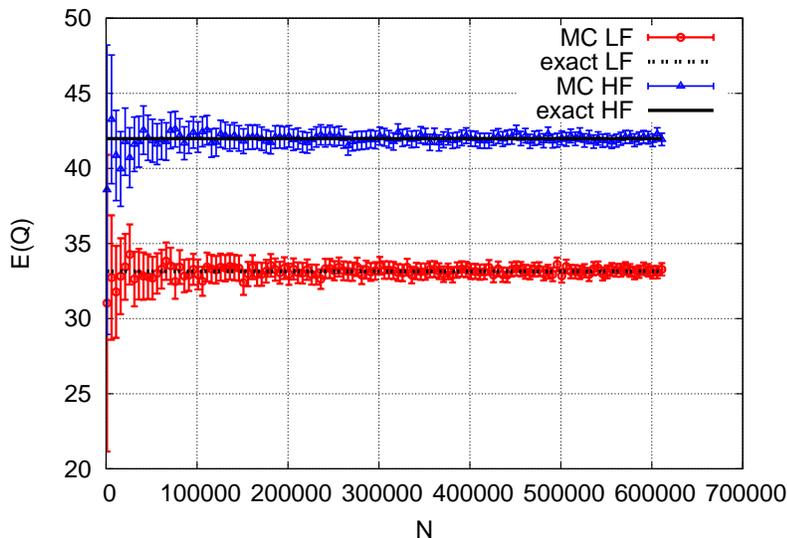


FIGURE 1. Monte Carlo convergence and its 99.7% confidence interval for LF and HF. The analytical solution is also reported as reference.

mesh. The total cost of each algorithm is computed as the sum of all the costs from the different realizations and successively is normalized by the unitary cost at the finest resolution level. This procedure is introduced to make comparable multilevel algorithms based on different geometrical levels. The left of Figure 2 shows how both numerical methods are converging to the exact solution. The variance reduction provided by the MLMF estimator is such that a narrow confidence interval can be obtained. Moreover, the accuracy  $\varepsilon$  can be computed as a direct measure of the quality of the prediction, and it is reported on the right of Figure 2. The accuracy is increased by roughly a half order of magnitude for the MLMF with respect to iMLMC. This reduction translates into a reduction of one order of magnitude of the number of realizations needed to achieve the same accuracy. The number of simulations reported is relative only to the HF cost, while the LF is considered to have no cost. In a real case scenario, the ratio between the two cost is high enough to legitimately consider the cost of a LF simulation negligible with respect to a HF one.

The comparison between the MLMF and MLMC with the same number of realizations is reported in Figure 3. On the left, the confidence interval as a function of the number of realizations is reported and it shows the superiority of MLMF, which attains a more reliable prediction even if both methods converge to the exact solution. On the right of Figure 3, the accuracy  $\varepsilon$  is reported and, for reference, the value of the corresponding accuracy obtained by means of a MC run is also shown. The MLMC results are superior to the MC method, reducing the error by a half an order of magnitude. MLMF further reduces the magnitude of the error by a half an order when compared to MLMC.

The application of the MLMC on the LF model, with  $\varepsilon = 1.5$ , leads to the following vector of realizations  $N_\ell^{\text{LF}} = [22161, 159, 20, 20]$ . Choosing the ratio  $r_\ell = [50, 5, 3, 3]$ , the algorithm produces the outputs reported in Table 2.

The correlation  $\rho_{HL}$  between LH and HF is the key parameter which guides the algorithm to achieve a variance reduction (reported in the %red column of Table 2). Note

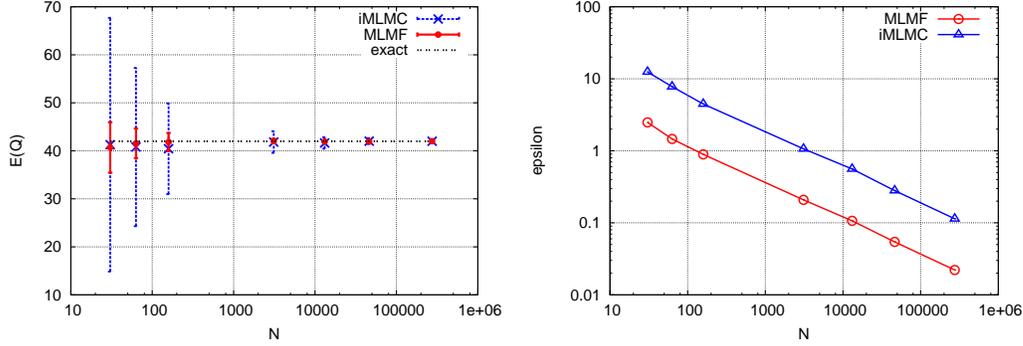


FIGURE 2. Comparison between MLMF and iMLMC. On the left the expected values and their 99.7% confidence intervals are reported, while on the right the accuracy  $\epsilon$  is shown as a function of the number of realizations.

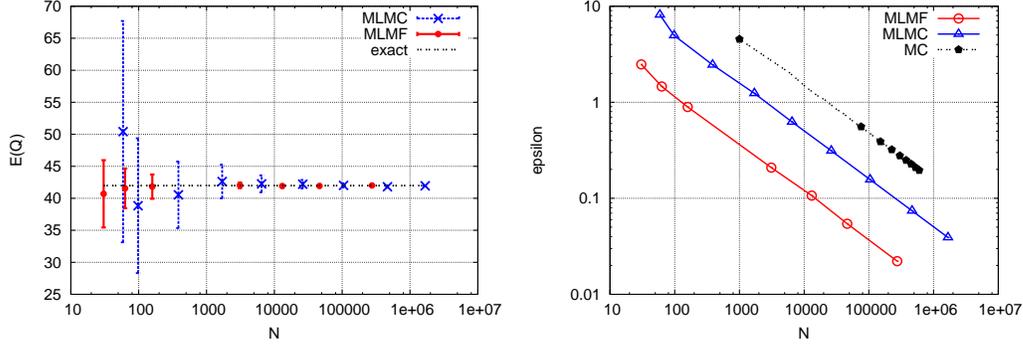


FIGURE 3. Comparison between MLMF and MLMC. On the left the expected values and their confidence intervals are reported, while on the right the accuracy  $\epsilon$  is shown as a function of the number of realizations (the accuracy for MC is also reported as a reference).

---

$N_\ell^{\text{HF}}$	$\rho_{HL}$	$\mathbb{E}[Y_\ell^{\text{HF}}]$	$\mathbb{E}[Y_\ell^{\text{HF}}]_{\text{corr}}$	$\text{Var}(Y_\ell^{\text{HF}})$	$\text{Var}(Y_\ell^{\text{HF}})_{\text{corr}}$	%red
434	0.9917	0.3607E+02	0.5423E+01	0.2870E+02	0.2767E+02	0.9641E+02
26	0.2875	0.7398E+00	0.2897E-01	0.1365E+00	0.9442E-02	0.6916E+01
5	-0.1391	0.2416E+00	0.1259E-01	0.1000E-01	0.1452E-03	0.1452E+01
5	0.8732	0.2859E-01	-0.9045E-02	0.8649E-04	0.4946E-04	0.5719E+02

---

TABLE 2. Outputs obtained executing the MLMF algorithm with  $r_\ell = [50, 5, 3, 3]$ . For both the expected value and the variance, the values obtained by employing only HF realizations ( $\mathbb{E}[Y_\ell^{\text{HF}}]$  and  $\text{Var}(Y_\ell^{\text{HF}})$ ) are reported in comparison to those obtained by the control variate part ( $\mathbb{E}[Y_\ell^{\text{HF}}]_{\text{corr}}$  and  $\text{Var}(Y_\ell^{\text{HF}})_{\text{corr}}$ ). Finally, the variance reduction is quantified in the last column.

also that the effect of the correction terms<sup>†</sup> on both the expected value and the variance. In particular, the variance is mainly reduced in correspondence with the coarsest level, while the reduction is much lower for finer levels. This behavior does not affect

<sup>†</sup> The correction term includes the part of the estimator related to the LF control variate, see Eq. (4.1).

the efficiency of the overall algorithm because, according to the design of the multilevel strategies, the original function is replaced in order to diminish the variance contained in the finer levels. In this specific case, the difference of the variance  $\text{Var}(Y_\ell^{\text{HF}})$  between the first and second levels is more than two orders of magnitude.

## 6. Summary and perspectives

A multilevel multifidelity estimator for forward propagation of uncertainties is proposed. The multilevel MC estimator is modified in order to benefit from the realizations obtained by a less computationally expensive, less accurate, model. In particular, the advantages stem from the possibility of estimating the desired statistics for the quantity of interest with a tighter confidence interval. Indeed, the amplitude of the confidence interval, being related to the variance of the estimator itself, is reduced by means of a low-fidelity model as the control variate on each level. Some differences exist with respect to the more classical control variate approach in which the expected value of the control variate is known. In this brief, the control variate statistics, i.e., the statistics of the low-fidelity model, are not known and they are recovered by means of a MC estimator. The difference between the set of realizations used to compute the statistics for the high- and low-fidelity model outputs is the key element to guarantee the variance reduction. The set of realizations is chosen with a partial overlap between the high- and low-fidelity models allowing variance reduction through the correlation between the two models. A numerical example dealing with the quantification of the stochastic behavior of a rod subject to uncertain diffusivity and initial temperature profile is reported. The case is formulated in order to obtain a fully analytical solution for the statistics that can also be used for verification purposes. The proposed MLMF algorithm is shown, for this case, to be superior to both MLMC and MC. Current research directions are focused on the development of an optimization procedure aimed at balancing the computational effort between the low- and high-fidelity models and taking advantage of their correlation when it is present. In principle, after the first iteration, which corresponds to the MLMF in the present form, additional iterations can be carried out to distribute the computational burden towards the LF if it is well correlated, while retaining a relatively large number of HF realizations whenever the correlation between the models is low.

## Acknowledgments

This investigation was funded by the Advanced Simulation and Computing (ASC) program of the US Department of Energy's National Nuclear Security Administration via the PSAAP-II Center at Stanford. Gianluca Geraci gratefully acknowledges financial support through PSAAP II during an extended visit at the Sandia National Laboratories, Albuquerque, where part of this work originated. The authors also gratefully acknowledge Prof. Alireza Doostan and Hillary Fairbanks, University of Colorado at Boulder, for their insightful discussions on UQ multilevel strategies.

**Appendix A. A possible MLMC practical implementation**

---

**Algorithm 1:** Practical implementation of a MLMC algorithm.

---

```

 $\Delta_\ell = N_w$  and  $N_\ell = 0$ 
while  $\sum_{\ell=0}^L (\Delta_\ell) > 0$  do
  for  $\ell = 0 : L$  do
    if  $\Delta_\ell > 0$  then
      Generate  $\Delta_\ell$  samples
      Evaluate  $Q_\ell$  and  $Q_{\ell-1}$ 
       $Y_\ell = Y_{\ell-1} + (Q_\ell - Q_{\ell-1})$  and  $Y_\ell^2 = Y_{\ell-1}^2 + (Q_\ell - Q_{\ell-1})^2$ 
       $N_\ell = N_{\ell-1} + \Delta_\ell$ 
       $\mathbb{E}[Y_\ell] = Y_\ell / N_\ell$  and  $\text{Var}(Y_\ell) = Y_\ell^2 / N_\ell - (\mathbb{E}[Y_\ell])^2$ 
    end
  end
   $N_\ell = \frac{2}{\varepsilon^2} \left[ \sum_{k=0}^L (\text{Var}(Y_k) C_k)^{1/2} \right] \sqrt{\frac{\text{Var}(Y_\ell)}{C_\ell}}$ 
   $\Delta_\ell$  is updated
end
 $\mathbb{E}[\hat{Q}_M^{ML}] = \sum_{\ell=0}^L \mathbb{E}[Y_\ell]$ 

```

---

## REFERENCES

- HEINRICH, S. 2001 Multilevel Monte Carlo methods *Large-Scale Scientific Computing, Third International Conference LSSC 2001, Lecture Notes in Computer Science*, Springer-Verlag, 58–67.
- MISHRA, S. & SCHWAB, C. 2012 Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data. *Math. Comput.* **81**(280), 1979–2018.
- POETTE, B., DESPRÉS, B. & LUCOR, D. 2009 Uncertainty quantification for systems of conservation laws. *J. Comput. Phys.* **228**, 2443–2467.
- DOOSTAN, A., & OWHADI, H. 2011 A non-adapted sparse approximation of PDEs with stochastic inputs. *J. Comput. Phys.* **228**(12), 4332–4345.
- BARTH, A., SCHWAB, CH. & ZOLLINGER, N. 2011 Multi-level Monte Carlo finite element method for elliptic PDEs with stochastic coefficients. *Numer. Math.* **119**(1), 123–161.
- CLIFFE, K.A., GILES, M.B., SCHEICHL, R. & TECKENTRUP, A.L. 2011 Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Comput. Vis. Sci.* **14**(1), 3–15.
- GILES, M.B. 2008 Multilevel Monte Carlo path simulation. *Oper. Res.* **56**, 607–617.
- LE MAÎTRE, O. & KNIIO, O. Spectral Methods for Uncertainty Quantification With Applications to Computational Fluid Dynamics, Springer Verlag, 2010.
- TRYOEN, J., LE MAÎTRE, O. & ERN, A. 2012 Adaptive anisotropic spectral stochastic methods for uncertain scalar conservation laws. *SIAM J. Sci. Comput.*, **34**(5), A2459–A2481.
- LIU, J.S. 2001 Monte Carlo strategies in scientific computing, Springer Verlag, 2001.