Context-aware learning of hierarchies of low-fidelity models for multi-fidelity uncertainty quantification

Ionuț-Gabriel Farcaș[a,*], Benjamin Peherstorfer[b], Tobias Neckel[c], Frank Jenko[d,a,c], Hans-Joachim Bungartz[c]

[a] *Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin TX, USA*
[b] *Courant Institute of Mathematical Sciences, New York University, New York NY, USA*
[c] *Department of Informatics, Technical University of Munich, Garching, Germany*
[d] *Max Planck Institute for Plasma Physics, Garching, Germany*

Dataset link: https://github.com/ionutfarcas/context-aware-mfmc

Abstract

Multi-fidelity Monte Carlo methods leverage low-fidelity and surrogate models for variance reduction to make tractable uncertainty quantification even when numerically simulating the physical systems of interest with high-fidelity models is computationally expensive. This work proposes a context-aware multi-fidelity Monte Carlo method that optimally balances the costs of training low-fidelity models with the costs of Monte Carlo sampling. It generalizes the previously developed context-aware bi-fidelity Monte Carlo method to hierarchies of multiple models and to more general types of low-fidelity models. When training low-fidelity models, the proposed approach takes into account the context in which the learned low-fidelity models will be used, namely for variance reduction in Monte Carlo estimation, which allows it to find optimal trade-offs between training and sampling to minimize upper bounds of the mean-squared errors of the estimators for given computational budgets. This is in stark contrast to traditional surrogate modeling and model reduction techniques that construct low-fidelity models with the primary goal of approximating well the high-fidelity model outputs and typically ignore the context in which the learned models will be used in upstream tasks. The proposed context-aware multi-fidelity Monte Carlo method applies to hierarchies of a wide range of types of low-fidelity models such as sparse-grid and deep-network models. Numerical experiments with the gyrokinetic simulation code GENE show speedups of up to two orders of magnitude compared to standard estimators when quantifying uncertainties in small-scale fluctuations in confined plasma in fusion reactors. This corresponds to a runtime reduction from 72 days to four hours on one node of the Lonestar6 supercomputer at the Texas Advanced Computing Center.
© 2023 Elsevier B.V. All rights reserved.

* Corresponding author.
 E-mail addresses: ionut.farcas@austin.utexas.edu (I.-G. Farcaș), pehersto@cims.nyu.edu (B. Peherstorfer), neckel@in.tum.de (T. Neckel), frank.jenko@ipp.mpg.de (F. Jenko), bungartz@in.tum.de (H.-J. Bungartz).

1. Introduction

Uncertainty quantification is an essential building block for achieving predictive numerical simulations of physical systems. To make accurate Monte Carlo estimation of uncertainties tractable even when simulations of the physical system of interest are computationally expensive, multi-fidelity methods rely on low-fidelity or surrogate models: these models are leveraged for variance reduction to achieve speedups and the high-fidelity models are occasionally evaluated to guarantee unbiasedness; see [1] for a survey on multi-fidelity methods. However, if low-fidelity models are not readily available, then they need to be constructed and trained first, which can incur additional computational costs and require additional evaluations of the high-fidelity models to generate training data.

In this work, we build on the context-aware bi-fidelity Monte Carlo method introduced in [2] and propose the context-aware multi-fidelity Monte Carlo (CA-MFMC) method that trades off the costs of training hierarchies of multiple low-fidelity models with the costs of Monte Carlo sampling to obtain multi-fidelity estimators that minimize upper bounds of the mean-squared errors for given computational budgets. The proposed approach is context-aware [2–6] in the sense that the low-fidelity models are trained to maximize variance reduction in Monte Carlo estimation. This means that the context in which the learned models will be used, namely for variance reduction in Monte Carlo estimation, is taken into account during training, which distinguishes it from traditional surrogate modeling and model reduction techniques that construct low-fidelity models with the primary goal of approximating well the high-fidelity model outputs and typically ignore the context in which the learned models are ultimately used [7,8]. Our proposed CA-MFMC method can be combined with a wide range of types of data-fit low-fidelity models such as sparse-grid-based models and deep-network models. We show that CA-MFMC achieves speedups of up to two orders of magnitude compared to single-fidelity Monte Carlo and standard multi-fidelity estimators when quantifying uncertainties in a plasma micro-turbulence scenario from the ASDEX Upgrade Experiment.[1]

We build on multi-fidelity Monte Carlo (MFMC) estimators [1,9–15] that leverage a given hierarchy of low-fidelity models for variance reduction. Several works have employed MFMC estimators to quantify uncertainties in plasma micro-turbulence simulations [4,16] and for estimating statistics in collisionless energetic particle confinement in stellarators [17]. See also [18] for other techniques for uncertainty quantification in plasma simulations. None of these methods trade off the training of low-fidelity models with sampling, however. The works [19–21] formulate generalized control variate techniques for multi-fidelity uncertainty propagation. There is a wide range of other multi-fidelity techniques that are based on other concepts than control variates and aim to find surrogate models from multi-fidelity information sources such as the collocation approach introduced in [22,23] and the multi-fidelity networks proposed in [24,25]; see also [26–29]. There have been extensions to the multilevel Monte Carlo method [30–32] that consider spatially-adaptive mesh refinements and optimal mesh hierarchies [33–35]. However, in that line of work, the low-fidelity models are coarse-grid approximations and thus the costs of constructing low-fidelity models are typically considered to be negligible and therefore are ignored. In contrast, we consider more general types of low-fidelity models such as data-fit models that incur training costs.

The first work that considered trading off training low-fidelity models and multi-fidelity Monte Carlo estimation is [2], which studies the bi-fidelity setting in which the low-fidelity model has algebraic accuracy and cost rates. In [36] a similar trade off is considered in the bi-fidelity case but with more general cost and error rates and a specific focus on polynomial-chaos-based surrogate models. We go far beyond by introducing the CA-MFMC estimator based on context-aware learning that learns hierarchies of more than one low-fidelity model for variance reduction. We first extend the work on a single low-fidelity model with algebraic cost and error rates in [2] to apply to low-fidelity models with more general rates. Key is that the corresponding bounds can be nested, which motivates the sequential training of low-fidelity models to obtain a hierarchy. In the proposed sequential training approach to fit hierarchies of low-fidelity models for CA-MFMC estimators, each step leads to an optimal trade off between training and sampling. This leads to a context-aware learning approach because the models are learned such that the CA-MFMC estimator achieves a low mean-squared error, rather than being trained to accurately approximate high-fidelity model outputs as in traditional model reduction.

To demonstrate the performance of the proposed CA-MFMC estimator, we apply it to quantifying uncertainties in plasma micro-turbulence simulations motivated by the ITER experiment.[2] The goal of the ITER experiment is to

---

[1] https://www.ipp.mpg.de/16195/asdex
[2] https://www.iter.org

create a self-sustained plasma in magnetic confinement devices. A physics obstacle are the above-mentioned small-scale fluctuations which cause energy loss rates despite sophisticated plasma confinements via strong and shaped magnetic fields. Building on the gyrokinetic simulation code GENE [37], we show that the proposed estimator achieves speedups of up to two orders of magnitude compared to single-fidelity Monte Carlo and standard multi-fidelity estimators, which translates into a runtime reduction from 72 days to about four hours on one node of the Lonestar6 supercomputer at the Texas Advanced Computing Center.[3]

The remainder of this paper is organized as follows. Section 2 introduces the notation and summarizes the traditional MFMC algorithm [12,14] and the bi-fidelity context-aware algorithm formulated in [2]. Section 3 introduces our context-aware learning approach for low-fidelity models with general cost/error rates and multiple low-fidelity models for multi-fidelity sampling methods. Section 4 presents numerical results in two scenarios: a heat conduction problem defined on a two-dimensional spatial domain with nine uncertain parameters and a realistic plasma micro-turbulence scenario with 12 uncertain inputs, for which one realization of the uncertain inputs requires a total runtime of about 411 seconds on 32 cores. The code and data to reproduce our numerical results are available at https://github.com/ionutfarcas/context-aware-mfmc.

2. Preliminaries

This section reviews traditional MFMC estimators [12,14] and the bi-fidelity context-aware MFMC estimator [2] that uses a high-fidelity model and a single low-fidelity model only.

2.1. Static multi-fidelity Monte Carlo estimation

Let $f^{(0)} : \mathcal{X} \to \mathcal{Y}$ represent the input–output response of a potentially expensive-to-evaluate computational model. The input domain is $\mathcal{X} \subset \mathbb{R}^d$ and the output is scalar with output domain $\mathcal{Y} \subset \mathbb{R}$. We consider the situation where the input $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_d]^T$ is a realization of a random variable $\boldsymbol{\Theta} = [\Theta_1, \Theta_2, \ldots, \Theta_d]^T$ so that the output $f^{(0)}(\boldsymbol{\theta})$ becomes a realization of a random variable $f^{(0)}(\boldsymbol{\Theta})$ too. We denote the probability density function of $\boldsymbol{\Theta}$ as π. Our goal is to estimate the expected value of $f^{(0)}(\boldsymbol{\Theta})$

$$\mu_0 = \mathbb{E}[f^{(0)}(\boldsymbol{\Theta})] = \int_{\mathcal{X}} f^{(0)}(\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}.$$

The MFMC estimator introduced in [12,14] combines a hierarchy of $k + 1$ models, consisting of a high-fidelity model $f^{(0)}$ and k low-fidelity models $f^{(1)}, f^{(2)}, \ldots, f^{(k)}$. The accuracy of the low-fidelity models is measured by their Pearson correlation coefficient with respect to $f^{(0)}$,

$$\rho_j = \frac{\mathrm{Cov}[f^{(0)}, f^{(j)}]}{\sigma_0 \sigma_j} \in [-1, 1], \quad j = 1, 2, \ldots, k,$$

where σ_j^2 denotes the variance of the output random variable $f^{(j)}(\boldsymbol{\Theta})$ for $j = 0, 1, \ldots, k$ and $\mathrm{Cov}[\cdot, \cdot]$ is the covariance. In the following, we set $\rho_{k+1} = 0$. The evaluation costs of the models are $0 < w_1, w_2, \ldots, w_k$, respectively. We normalize the cost of evaluating the high-fidelity model such that $w_0 = 1$, without loss of generality. The models are assumed to satisfy the following ordering

$$1 = |\rho_0| > |\rho_1| > \cdots > |\rho_k|, \quad \frac{w_{j-1}}{w_j} > \frac{\rho_{j-1}^2 - \rho_j^2}{\rho_j^2 - \rho_{j+1}^2}, \quad j = 1, 2, \ldots, k. \tag{2.1}$$

Let now $m_j \in \mathbb{N}$ for $j = 0, 1, \ldots, k$ denote the number of evaluations of model $f^{(j)}$, which satisfy

$$0 \leq m_0 \leq m_1 \leq \cdots \leq m_k$$

due to (2.1). Consider m_k independent and identically distributed (i.i.d.) samples drawn from the input density π:

$$\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{m_k}.$$

To obtain the MFMC estimator, model $f^{(j)}$ is evaluated at the first m_j samples $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_{m_j}$ to obtain output random variables

$$f^{(j)}(\boldsymbol{\theta}_1), f^{(j)}(\boldsymbol{\theta}_2), \ldots, f^{(j)}(\boldsymbol{\theta}_{m_j}), \tag{2.2}$$

---

[3] https://www.tacc.utexas.edu/systems/lonestar6

for $j = 0, 1, \ldots, k$. From (2.2), derive the following standard MC estimators for $j = 1, 2, \ldots, k$:

$$\hat{E}_{m_j}^{(j)} = \frac{1}{m_j} \sum_{i=1}^{m_j} f^{(j)}(\boldsymbol{\theta}_i), \quad \hat{E}_{m_{j-1}}^{(j)} = \frac{1}{m_{j-1}} \sum_{i=1}^{m_{j-1}} f^{(j)}(\boldsymbol{\theta}_i)$$

as well as

$$\hat{E}_{m_0}^{(0)} = \frac{1}{m_0} f^{(0)}(\boldsymbol{\theta}_i).$$

The MFMC estimator is

$$\hat{E}^{\text{MFMC}} = \hat{E}_{m_0}^{(0)} + \sum_{j=1}^{k} \alpha_j (\hat{E}_{m_j}^{(j)} - \hat{E}_{m_{j-1}}^{(j)})$$

with the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}$. The total computational cost of the MFMC estimator is therefore

$$p = \sum_{j=0}^{k} m_j w_j.$$

Fixing p and then selecting $m_0^*, m_1^*, \ldots, m_k^*$ model evaluations and $\alpha_1^*, \alpha_2^*, \ldots, \alpha_k^*$ coefficients such that the variance of the MFMC estimator $\hat{E}^{\text{MFMC}}$ is minimized, gives the MSE

$$\text{MSE}[\hat{E}^{\text{MFMC}}] = \frac{\sigma_0^2}{p} \left(\sum_{j=0}^{k} \sqrt{w_j(\rho_j^2 - \rho_{j+1}^2)} \right)^2. \tag{2.3}$$

The optimal $m_0^*, m_1^*, \ldots, m_k^*$ model evaluations and $\alpha_1^*, \alpha_2^*, \ldots, \alpha_k^*$ coefficients are available in closed form for a given budget p; see [14] for details.

The MFMC estimator leverages the cheaper-to-evaluate low-fidelity models with the aim to achieve a lower MSE than a standard MC estimator with the same costs. That is, given m i.i.d. samples from $\boldsymbol{\Theta}$, the standard MC mean estimator is

$$\hat{E}_m^{(0)} = \frac{1}{m} \sum_{i=1}^{m} f^{(0)}(\boldsymbol{\theta}_i). \tag{2.4}$$

The computational cost of the Monte Carlo estimator (2.4) of $\mathbb{E}[f^{(0)}(\boldsymbol{\Theta})]$ is $p = m w_0 = m$ because the function $f^{(0)}$ is evaluated at m realizations and the cost of each evaluation of $f^{(0)}$ is $w_0 = 1$. The MSE of the standard MC estimator $\hat{E}_n^{(0)}$ reads

$$\text{MSE}[\hat{E}_m^{(0)}] = \frac{\sigma_0^2}{p} w_0 = \frac{\sigma_0^2}{p}. \tag{2.5}$$

2.2. Context-aware bi-fidelity Monte Carlo estimator with algebraic accuracy and cost rates

The work [2] introduces a context-aware bi-fidelity MFMC approach that trades off the training costs of constructing a low-fidelity model $f_n^{(1)}$ and sampling. Following [2], the low-fidelity model $f_n^{(1)}$ is obtained via a training process such as data-fit models and reduced models. Correspondingly, the subscript n refers to the number of high-fidelity evaluations that are needed to train $f_n^{(1)}$. For example, in case of obtaining $f_n^{(1)}$ via training a neural network, the subscript n refers to the number of training input–output pairs obtained with the high-fidelity model $f^{(0)}$. This also means that the correlation coefficient $\rho_1(n)$ between $f^{(0)}(\boldsymbol{\Theta})$ and $f_n^{(1)}(\boldsymbol{\Theta})$ as well as the evaluation cost $w_1(n)$ of $f_n^{(1)}$ depend on n.

The dependency of the correlation and costs on n are described as follows: to trade off training low-fidelity model and sampling, the work [2] makes the assumption that the correlation coefficient between the high-fidelity output random variable $f^{(0)}(\boldsymbol{\Theta})$ and low-fidelity random variable $f_n^{(1)}(\boldsymbol{\Theta})$ is bounded with respect to n by

$$1 - \rho_1^2(n) \le c_1 n^{-\alpha},$$

where $c_1, \alpha > 0$ are constants. The cost $w_1(n)$ of evaluating the low-fidelity model is bounded by

$$w_1(n) \leq c_2 n^\beta$$

with constants $c_2, \beta > 0$.

A budget p corresponds to p high-fidelity model evaluations because we have $w_0 = 1$. Thus, if n high-fidelity model samples are used for training $f_n^{(1)}$, then a budget of $p-n$ is left for sampling. The corresponding context-aware bi-fidelity estimator is

$$\hat{E}_n^{\text{CA-MFMC}} = \hat{E}_{m_0^*}^{(0)} + \alpha_1^* \left(\hat{E}_{m_1^*}^{(1)} - \hat{E}_{m_0^*}^{(1)} \right)$$

where m_0^*, m_1^* and α_1^* are chosen to minimize the MSE of $\hat{E}_n^{\text{CA-MFMC}}$ for a given p and n. Consequently, the MSE of $\hat{E}_n^{\text{CA-MFMC}}$ depends on p and n

$$\text{MSE}(\hat{E}_n^{\text{CA-MFMC}}) = \frac{\sigma_0^2}{p - n} \left(\sqrt{1 - \rho_1^2(n)} + \sqrt{w_1(n)\rho_1^2(n)} \right)^2,$$

which is in contrast to the MFMC estimator $\hat{E}^{\text{MFMC}}$ for which the MSE (2.3) depends on p only and is independent of potential training costs of constructing low-fidelity models.

If the budget p is fixed, n can be chosen by minimizing the upper bound of the MSE

$$\text{MSE}(\hat{E}_n^{\text{CA-MFMC}}) \leq \frac{2\sigma_0^2}{p - n} \left(c_1 n^{-\alpha} + c_2 n^\beta \right). \tag{2.6}$$

The work [2] shows that there exists a unique n^* that minimizes (2.6) with respect to n for a given p under certain assumptions; however, no closed form expression of n^* is available and thus n^* needs to be found numerically. The optimum n^* is bounded independently of the budget p; see [2] for details.

3. Context-aware multi-fidelity Monte Carlo estimation

This section presents the novelty of this paper. We introduce context-aware learning for MFMC for low-fidelity models with general cost/error rates and hierarchies of multiple – more than one – low-fidelity models.

3.1. Setup with multiple models

We now consider multiple low-fidelity models $f_{n_1}^{(1)}, f_{n_2}^{(2)}, \ldots, f_{n_k}^{(k)}$ that take $n_1, n_2, \ldots, n_k$ evaluations, respectively, of the high-fidelity model to be trained. We use the convention $n_0 = 0$ in the following. We refer to $n_1, n_2, \ldots, n_k$ as the number of training samples to fit the low-fidelity models. We then define the CA-MFMC estimator as

$$\hat{E}_{n_1, n_2, \ldots, n_k}^{\text{CA-MFMC}} = \hat{E}_{m_0^*}^{(0)} + \sum_{j=1}^k \alpha_j^* \left(\hat{E}_{m_j^*}^{(j)} - \hat{E}_{m_{j-1}^*}^{(j)} \right),$$

where $m_0^*, m_1^*, \ldots, m_k^*$ and $\alpha_1^*, \alpha_2^*, \ldots, \alpha_k^*$ minimize the MSE of $\hat{E}_{n_1, n_2, \ldots, n_k}^{\text{CA-MFMC}}$ for given $n_1, n_2, \ldots, n_k$ and budget p. The MSE of the CA-MFMC estimator is

$$\text{MSE}[\hat{E}_{n_1, n_2, \ldots, n_k}^{\text{CA-MFMC}}] = \frac{\sigma_0^2}{p - \sum_{j=1}^k n_j} \left(\sum_{j=0}^k \sqrt{w_j(n_j)(\rho_j^2(n_j) - \rho_{j+1}^2(n_{j+1}))} \right)^2, \tag{3.1}$$

where now the MSE depends on the $n_1, n_2, \ldots, n_k$ training samples used for constructing the low-fidelity models. We are interested in choosing $n_1, n_2, \ldots, n_k$ to minimize an upper bound of the MSE (3.1) so that the training costs given by $n_1, n_2, \ldots, n_k$ and the costs of taking samples from the high- and the low-fidelity output random variables are balanced.

We make the following assumptions on the accuracy and cost behavior of the low-fidelity models with respect to the number of training samples $n_1, n_2, \ldots, n_k$:

Assumption 1. For all $j = 1, 2, \ldots, k$, there exists a constant $c_{a,j} > 0$ and a positive, with respect to n_j decreasing, at least twice continuously differentiable function $r_{a,j}(n_j) : (0, \infty) \to (0, \infty)$ such that

$$1 - \rho_j^2(n_j) \le c_{a,j} r_{a,j}(n_j), \quad j = 1, 2, \ldots, k.$$

Assumption 2. For all $j = 1, 2, \ldots, k$, there exists a constant $c_{c,j} > 0$ and a positive, with respect to n_j increasing, at least twice continuously differentiable function $r_{c,j}(n_j) : (0, \infty) \to (0, \infty)$ such that the evaluation costs are bounded as

$$w_j(n_j) \le c_{c,j} r_{c,j}(n_j), \quad j = 1, 2, \ldots, k.$$

Assumptions 1 and 2 hold for a wide range of typical error and cost rates: the functions $r_{a,j}(n) = n^{-\alpha}, \alpha > 0$ and $r_{a,j}(n) = e^{-\alpha n}, \alpha > 0$ satisfy Assumption 1 for $n \in (0, \infty)$. The functions $r_{c,j}(n) = n^{\alpha}, \alpha > 1$ and $r_{c,j}(n) = e^{\alpha n}, \alpha > 0$ satisfy Assumption 2 for $n \in (0, \infty)$. Additionally, the error and cost rates given here are strictly convex.

Remark 1. The setup that we consider consists of k low-fidelity models $f_{n_1}^{(1)}, f_{n_2}^{(2)}, \ldots, f_{n_k}^{(k)}$ that are trained on data obtained from $n_1, n_2, \ldots, n_k$ high-fidelity model evaluations, respectively. Depending on the model type, training data can be reused: Data obtained from the n_i high-fidelity model evaluations to train model $f_{n_i}^{(i)}$ can be helpful to train model $f_{n_j}^{(j)}$ with $j \ne i$. For example, training data generated to construct a sensitivity-driven sparse grid interpolant [4,38] can be reused to construct a projection-based reduced model with proper orthogonal decomposition. However, whether training data can be reused or not depends on the model type. It can also depend on other constraints and details of the setup. Thus, to keep the methodology general, we ignore such problem-dependent specifics in the following.

3.2. Context-aware multi-fidelity Monte Carlo sampling with one low-fidelity model

Let us first consider only one low-fidelity model, which means that $k = 1$. The following is novel compared to what was introduced in [2] because here we allow more general error and cost rates than [2]; cf. Section 2.2 where [2] is reviewed.

We obtain the following upper bound of the MSE of the CA-MFMC estimator (3.1),

$$
\begin{aligned}
\mathrm{MSE}[\hat{E}_{n_1}^{\mathrm{CA-MFMC}}] &= \frac{\sigma_0^2}{p - n_1} \left(\sqrt{1 - \rho_1^2(n_1)} + \sqrt{w_1(n_1)\rho_1^2(n_1)} \right)^2 \\
&= \frac{\sigma_0^2}{p - n_1} \left(1 - \rho_1^2(n_1) + w_1(n_1)\rho_1^2(n_1) + 2\sqrt{1 - \rho_1^2(n_1)}\sqrt{w_1(n_1)\rho_1^2(n_1)} \right) \\
&\le \frac{\sigma_0^2}{p - n_1} \left(2(1 - \rho_1^2(n_1)) + 2w_1(n_1)\rho_1^2(n_1) \right) & (3.2) \\
&\le \frac{2\sigma_0^2}{p - n_1} \left(1 - \rho_1^2(n_1) + w_1(n_1) \right) & (3.3) \\
&\le \frac{2\sigma_0^2}{p - n_1} \left(c_{a,1} r_{a,1}(n_1) + c_{c,1} r_{c,1}(n_1) \right), & (3.4)
\end{aligned}
$$

where inequality (3.2) is due to the means inequality $\forall a, b \in \mathbb{R}_+ \Rightarrow 2\sqrt{a}\sqrt{b} \le a + b$, inequality (3.3) results from $\rho_1^2 \le 1$ and (3.4) is due to Assumptions 1 and 2.

The objective that we want to minimize with respect to n_1 is

$$u_1 : [1, p - 1] \to (0, \infty), n_1 \mapsto \frac{1}{p - n_1} \left(c_{a,1} r_{a,1}(n_1) + c_{c,1} r_{c,1}(n_1) \right). \tag{3.5}$$

The following proposition clarifies when a unique global minimum exists.

Proposition 1. *Consider function u_1 defined in* (3.5) *with $r_{a,1}$ and $r_{c,1}$ satisfying Assumptions 1 and 2, respectively. Additionally assume that*

$$c_{a,1} r_{a,1}''(n_1) + c_{c,1} r_{c,1}''(n_1) > 0 \tag{3.6}$$

holds for all $n_1 \in [1, p-1]$, where $r''_{a,1}$ and $r''_{c,1}$ are the second derivatives of $r_{a,1}$ and $r_{c,1}$, respectively. Then, the function u_1 has a unique global minimum $n_1^* \in [1, p-1]$.

Proof. Define the function $h_1(n_1) = c_{a,1}r_{a,1}(n_1) + c_{c,1}r_{c,1}(n_1)$ and notice that

$$h''_1(n_1) = c_{a,1}r''_{a,1}(n_1) + c_{c,1}r''_{c,1}(n_1) > 0, \qquad n_1 \in [1, p-1] \tag{3.7}$$

due to (3.6), which means that h_1 is strictly convex in $[1, p-1]$. We now consider u_1 and its first derivative

$$u'_1(n_1) = \frac{(p-n_1)h'_1(n_1) + h_1(n_1)}{(p-n_1)^2}. \tag{3.8}$$

Case 1: If $u'_1(n_1) > 0$ for all $n_1 \in [1, p-1]$, then u_1 is strictly increasing and the minimum of u_1 in $[1, p-1]$ is taken at the left boundary $n_1^* = 1$, which is the global unique minimum.

Case 2: Analogously to Case 1, if $u'_1(n_1) < 0$ for all $n_1 \in [1, p-1]$, then u_1 is strictly decreasing and the unique minimum is $p-1$.

Case 3: There is a stationary point $n_1^* \in [1, p-1]$ such that $u'_1(n_1^*) = 0$. At a stationary point n_1^* of u_1, the second derivative,

$$u''_1(n_1) = \frac{h''_1(n_1)}{p-n_1} + \frac{2}{p-n_1}u'_1(n_1),$$

is positive $u''_1(n_1^*) > 0$ because of (3.7) and $u'_1(n_1^*) = 0$. Thus, together with the fact that function u_1 is twice continuously differentiable and univariate, we have that u''_1 is positive in a neighborhood about all stationary points, because the inequality $h''_1(n_1^*) > 0$ is strict. Thus, in the interior $(1, p-1)$, there can only be minima, which also means that at the boundaries of the interval $[1, p-1]$ can only be maxima. This means that there exists only one minimum, which shows uniqueness. $\square$

Remark 2. Assumption (3.6) in Proposition 1 is satisfied if $r_{a,1}$ is strictly convex and $r_{c,1}$ is convex, which holds for the examples given in Section 3.1.

We will next show that the minimizer of (3.5) is also bounded from above, independent of the computational budget, p, which implies that after a finite number of samples, all high-fidelity model samples should be used in the Monte Carlo estimator rather than being used to improve the low-fidelity model.

Proposition 2. *Let Proposition 1 apply with condition (3.6) satisfied for all $n_1 \in (0, \infty)$. Additionally, assume that there exists an $\bar{n}_1 \in (0, \infty)$ such that*

$$c_{a,1}r'_{a,1}(\bar{n}_1) + c_{c,1}r'_{c,1}(\bar{n}_1) = 0 \tag{3.9}$$

holds. Then $\bar{n}_1$ is unique and the minimum n_1^* of u_1 derived in Proposition 1 is bounded from above independently of the budget p as

$$n_1^* \leq \max\{1, \bar{n}_1\}. \tag{3.10}$$

Proof. Consider h_1 introduced in the proof of Proposition 2. There is an $\bar{n}_1 \in (0, \infty)$ with $h'_1(\bar{n}_1) = 0$ as given by (3.9). The stationary point $\bar{n}_1$ is unique because h'_1 is strictly increasing because we assumed that (3.6) holds in $(0, \infty)$. Furthermore, the stationary point $\bar{n}_1$ is independent of p because h_1 does not depend on p.

Case 1: If $\bar{n}_1 < 1$, then this implies that $h'_1 > 0$ (strictly increasing because $h''_1 > 0$) for all $[1, p-1]$ for all $p \geq 1$ and thus with u'_1 given in (3.8) we obtain $u'_1 > 0$ in $[1, p-1]$. Thus, $n_1^* = 1$ and (3.10) is a bound independent of p.

Case 2: Let now $\bar{n}_1 \geq 1$. First, consider budgets $p > \bar{n}_1 + 1$ so that $\bar{n}_1 \in [1, p-1)$. Because $h'_1(\bar{n}_1) = 0$ with $\bar{n}_1 \in [1, p-1)$ and h'_1 strictly increasing, we have that $h'_1 > 0$ in $(\bar{n}_1, p-1]$ and thus $u'_1 > 0$ in $(\bar{n}_1, p-1]$, which means u_1 cannot be strictly decreasing on all of $[1, p-1]$. Thus, the minimum of u_1 in $[1, p-1]$ is either at a stationary point of u_1 or at 1. If it is a stationary point, then $h'_1(n_1^*) < 0$ has to hold to obtain $u'_1(n_1^*) = 0$ and we obtain $h'_1(n_1^*) < 0 = h'_1(\bar{n}_1)$, which implies $n_1^* \leq \bar{n}_1$ because h'_1 is strictly increasing and thus (3.10) is a p-independent bound of n_1^*. If it $n_1^* = 1$, then $1 = n_1^* \leq \bar{n}_1$ still holds because $\bar{n}_1 \geq 1$. Second, consider budgets $p \leq \bar{n}_1 + 1$ so that $\bar{n}_1 \geq p-1$. Then, it holds $n_1^* \leq p-1 \leq \bar{n}_1$, which again leads to the bound (3.10).

This shows that depending on the properties of h_1, the maximum of 1 and $\bar{n}_1$ is an upper bound of n_1^*. $\square$

3.3. Context-aware multi-fidelity Monte Carlo sampling with two or more low-fidelity models

The second novelty of our proposed context-aware approach is that we can consider more than just one low-fidelity model. In the following, we introduce a sequential training approach to fit hierarchies of low-fidelity models for the CA-MFMC estimator, where in each step the optimal trade off between training and sampling is achieved.

We first show an upper bound of the MSE (3.1) in terms of accuracy and cost rates for the case with $k > 1$ low-fidelity models:

Lemma 1. *Let Assumptions 1 and 2 hold. Then, the MSE (3.1) of the CA-MFMC estimator $\hat{E}^{\text{CA-MFMC}}_{n_1,n_2,\ldots,n_k}$ can be bounded as*

$$\text{MSE}[\hat{E}^{\text{CA-MFMC}}_{n_1,n_2,\ldots,n_k}] \leq \frac{(k+1)\sigma_0^2}{p_{k-1} - n_k} \left(\kappa_{k-1} + \hat{c}_{a,k} r_{a,k}(n_k) + c_{c,k} r_{c,k}(n_k) \right), \tag{3.11}$$

with $\hat{c}_{a,1} = c_{a,1}, p_0 = p, \kappa_0 = 0$ and

$$p_{k-1} = p - \sum_{j=1}^{k-1} n_j, \quad \kappa_{k-1} = c_{a,1} r_{a,1}(n_1) + \sum_{j=1}^{k-2} c_{c,j} r_{c,j}(n_j) c_{a,j+1} r_{a,j+1}(n_{j+1}), \quad \hat{c}_{a,k} = c_{c,k-1} r_{c,k-1}(n_{k-1}) c_{a,k}. \tag{3.12}$$

Proof. With the sums of squares inequality,

$$\left(\sum_{j=0}^{k} a_j \right)^2 \leq (k+1) \sum_{j=0}^{k} a_j^2, \qquad a_0, a_1, \ldots, a_k \in \mathbb{R},$$

we obtain

$$\text{MSE}[\hat{E}^{\text{CA-MFMC}}_{n_1,n_2,\ldots,n_k}] = \frac{\sigma_0^2}{p - \sum_{j=1}^{k} n_j} \left(\sum_{j=0}^{k} \sqrt{w_j(n_j) \left(\rho_j^2(n_j) - \rho_{j+1}^2(n_{j+1}) \right)} \right)^2$$

$$\leq \frac{\sigma_0^2 (k+1)}{p - \sum_{j=1}^{k} n_j} \sum_{j=0}^{k} w_j(n_j) (\rho_j^2(n_j) - \rho_{j+1}^2(n_{j+1}))$$

$$= \frac{\sigma_0^2 (k+1)}{p - \sum_{j=1}^{k} n_j} \left(\sum_{j=0}^{k-2} w_j(n_j) (\rho_j^2(n_j) - \rho_{j+1}^2(n_{j+1})) + \right.$$

$$\left. w_{k-1}(n_{k-1}) (\rho_{k-1}^2(n_{k-1}) - \rho_k^2(n_k)) + w_k(n_k) \rho_k^2(n_k) \right),$$

where we used the convention that $\rho_{k+1} = 0$ in the last equation. With $w_0 = 1$, $\rho_0 = 1$ and $\rho_j^2 \leq 1$ for $j = 1, 2, \ldots, k$, and p_{k-1}, κ_{k-1} defined in (3.12), the bound (3.11) follows with Assumptions 1 and 2. $\square$

Bound (3.11) in Lemma 1 consists of the component κ_{k-1} and the factor $\hat{c}_{a,k}$ that both depend on $n_1, n_2, \ldots, n_{k-1}$ only as well as the terms $r_{a,k}(n_k)$ and $r_{c,k}(n_k)$ that depend on n_k only. This decomposition of the bound motivates a sequential approach of adding low-fidelity models to the CA-MFMC estimator. At iteration $j = 1, 2, \ldots, k$, we consider the objective

$$u_j(n_j; n_1, n_2, \ldots, n_{j-1}) : [1, p_{j-1} - 1] \to (0, \infty), \qquad n_j \mapsto \frac{1}{p_{j-1} - n_j} \left(\kappa_{j-1} + \hat{c}_{a,j} r_{a,j}(n_j) + c_{c,j} r_{c,j}(n_j) \right). \tag{3.13}$$

For $j = 1$, with the convention $\kappa_0 = 0$ and $\hat{c}_{a,1} = c_{a,1}$, we obtain the objective defined in (3.5). For $j > 1$, we obtain objectives u_j in (3.13) that depend on $n_1, n_2, \ldots, n_{j-1}$. For given $n_1, n_2, \ldots, n_{j-1}$, there is a global, unique minimizer of u_j in $[1, p_{j-1} - 1]$ as the following proposition shows.

Proposition 3. *Let Lemma 1 apply. For $j = 1$, let Proposition 1 apply and let n_1^* be the corresponding unique global minimum. Iterate now over $j = 2, 3, \ldots, k$ and consider the objective $u_j(n_j; n_1^*, n_2^*, \ldots, n_{j-1}^*)$ defined in (3.13) at iteration j. If it holds that*

$$\hat{c}_{a,j} r_{a,j}''(n_j) + c_{c,j} r_{c,j}''(n_j) > 0, \tag{3.14}$$

then there exists a unique minimizer $n_j^* \in [1, p_{j-1} - 1]$ of u_j.

Proof. Define the function

$$h_j(n_j) = \kappa_{j-1} + \hat{c}_{a,j} r_{a,j}(n_j) + c_{c,j} r_{c,j}(n_j) \tag{3.15}$$

and notice that

$$h_j''(n_j) = \hat{c}_{a,j} r_{a,j}''(n_j) + c_{c,j} r_{c,j}''(n_j) > 0$$

for $j \geq 2$ because (3.14) is assumed. Thus, the same arguments as in the proof of Proposition 1 apply and thus unique minimizers $n_2^*, n_3^*, \ldots, n_k^*$ exist. $\square$

The next proposition shows the minimizers of (3.13) for $j = 1, 2, \ldots, k$ are bounded from above, independent of the computational budget p.

Proposition 4. *Let Proposition 3 apply, with the condition (3.14) holding on $(0, \infty)$. Furthermore, analogously to Proposition 2, let there exist $\bar{n}_j \in (0, \infty)$ so that*

$$\hat{c}_{a,j} r_{a,j}'(\bar{n}_j) + c_{c,j} r_{c,j}'(\bar{n}_j) = 0$$

holds for all $j = 1, 2, \ldots, k$. Then, for $j = 1, 2, \ldots, k$, there exist bounds $\bar{n}_1^*, \bar{n}_2^*, \ldots, \bar{n}_k^*$ that are independent of p so that $n_j^* \leq \bar{n}_j^*$ for $j = 1, 2, \ldots, k$.

Proof. Proof via induction and applying Proposition 2 sequentially to Proposition 3. $\square$

Propositions 1–4 highlights that even though using more than n_j^* high-fidelity model evaluations as training data to construct the jth low-fidelity model can lead to a more accurate low-fidelity model in terms of correlation coefficient, it also increases its evaluation costs, which ultimately leads to an increase of the upper bound of the MSE of the MFMC estimator (3.11) for a fixed budget and thus a poorer estimator. This shows that it is beneficial in multi-fidelity methods to trade off accuracy and evaluation costs of the low-fidelity models.

Remark 3. Increasing the model hierarchy by adding the jth context-aware low-fidelity model $f_{n_j}^{(j)}$ must ensure that the MFMC ordering (2.1) is satisfied. An indication that the ordering is wrong can be obtained by estimating the correlation coefficient via Assumption 1 as $\rho_j^2 \approx 1 - c_{a,j} r_{a,j}(n_j^*)$ and evaluation costs via Assumption 1 as $w_j \approx c_{c,j} r_{c,j}(n_j^*)$ based on n_j^*. If (2.1) is not satisfied, the order of the low-fidelity models in the multi-fidelity hierarchy must be changed accordingly.

4. Numerical experiments and discussion

We now present numerical results. In Section 4.1, we consider a heat conduction problem given by a parametric elliptic partial differential equation (PDE) with nine uncertain parameters, defined on a two-dimensional spatial domain. Our goal in this experiment is to draw initial insights about the proposed context-aware learning algorithm. To this end, we will employ two heterogeneous low-fidelity models: an accurate low-fidelity model that is computationally more expensive than a second, less accurate, low-fidelity model. Section 4.2 considers plasma micro-turbulence simulations that depend on 12 uncertain inputs. In this example, we consider three low-fidelity models: a coarse-grid model and two non-intrusive data-driven low-fidelity models based on sparse grid approximations and deep neural network regression.

4.1. Heat conduction in a two-dimensional spatial domain

All numerical experiments in this section were performed using an eight core Intel i7-10510U CPU and 16 GB of RAM.

Table 1
Thermal block: Estimated accuracy and evaluation cost rates and their constants for the RB low-fidelity model.

Rate	Estimated rate constants
Accuracy: $c_{a,1}e^{-\alpha_1 n_1}$	$c_{a,1} = 0.6312,\ \alpha_1 = 0.5754$
Evaluation cost: $c_{c,1}n_1^{\beta_1}$	$c_{c,1} = 9.6233 \times 10^{-5},\ \beta_1 = 1.0704$

4.1.1. Setup

The thermal block problem [8] is defined on a two-dimensional spatial domain $\Omega = (0,1)^2 = \bigcup_{i=1}^d \Omega_i$, divided into $d = B_1 \times B_2$ non-overlapping vertical and horizontal square blocks, Ω_i with $i = 1, 2, \ldots, d$. The mathematical model is given by the parametric elliptic PDE

$$
\begin{aligned}
-\operatorname{div} k(x, y, \boldsymbol{\theta})\nabla u(x, y) &= 0 \text{ in } \Omega, \\
u(x, y) &= 0 \text{ on } \Gamma_D, \\
k(x, y, \boldsymbol{\theta})\nabla u(x, y) \cdot \boldsymbol{n} &= j \text{ on } \Gamma_{N,j}, \quad j = 0, 1,
\end{aligned}
\tag{4.1}
$$

where Γ_D is a Dirichlet boundary, $\Gamma_{N,0}$ and $\Gamma_{N,1}$ are Neumann boundaries, and

$$
k(x, y, \boldsymbol{\theta}) = \sum_{i=1}^d \theta_i \chi_{\Omega_i}(x, y)
$$

is the piece-wise constant heat conductivity field parametrized by $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_d] \subset \mathbb{R}^d$, where χ_{Ω_i} is the indicator function of Ω_i. We set $B_1 = B_2 = 3$ and thus $d = 9$. Here, $k(x, y, \boldsymbol{\theta})$ is parametrized by $d = 9$ uniformly distributed random parameters in $[1, 10]^9$. The output of interest is the mean heat flow at the Neumann boundary $\Gamma_{N,1}$ given by

$$
\mathbb{E}[u(\boldsymbol{\theta})] = \int_{\Gamma_{N,1}} u(x, y, \boldsymbol{\theta})\mathrm{d}x\mathrm{d}y.
\tag{4.2}
$$

This setup is a slight modification of the problem considered by Patera and Rozza in [8].

The high-fidelity model is a finite element discretization of (4.1) consisting of 7,200 triangular elements with mesh width $h = \sqrt{2}/60$, provided by the RBMATLAB.[4] The high-fidelity model evaluation cost is 0.1150 s. Moreover, its variance, estimated using 100 MC samples, is $\sigma_0^2 \approx 0.0018$.

4.1.2. A low-fidelity model based on the reduced basis method

The reduced-basis (RB) low-fidelity model $f_{n_1}^{(1)}$ is constructed using a greedy strategy, as described, for example, in [39]. We employ the implementation provided by RBMATLAB. It has been shown that greedy RB low-fidelity models have exponential accuracy rates for problems similar to the thermal block [39], which is what we use when fitting the error decay. Once the reduced basis is found in the offline stage, evaluating the low-fidelity model online entails solving a dense linear system of size equal to the number of reduced bases to find the reduced basis coefficients. Therefore, we model the evaluation cost rate as algebraic in the reduced-model dimension.

The constants in the rate functions are estimated via regression from pilot runs. To estimate the constants in the exponential accuracy rate, we use 100 high- and low-fidelity evaluations. For estimating the constants in the evaluation cost rate, we average the runtimes of the low-fidelity model constructed using increasing values of n_1, evaluated at 1,000 MC samples. We perform these runtime measurements 50 times and average the results. The estimated rates are visualized in Fig. 1 and the constants are shown in Table 1.

Remark 4. When the accuracy and cost rates in Assumptions 1 and 2 are unavailable analytically, one typically estimates them numerically in a pilot study, as in our experiments. The sensitivity of the performance of MFMC to errors in the estimates of the correlation coefficients and costs is studied in detail in [14], where the finding is that typically crude estimates of correlation coefficients and costs are sufficient. In particular, errors in the estimates of the correlation coefficients and costs have no effect on the bias of MFMC estimators, and the same holds for the

---

[4] https://www.morepas.org/software/rbmatlab/
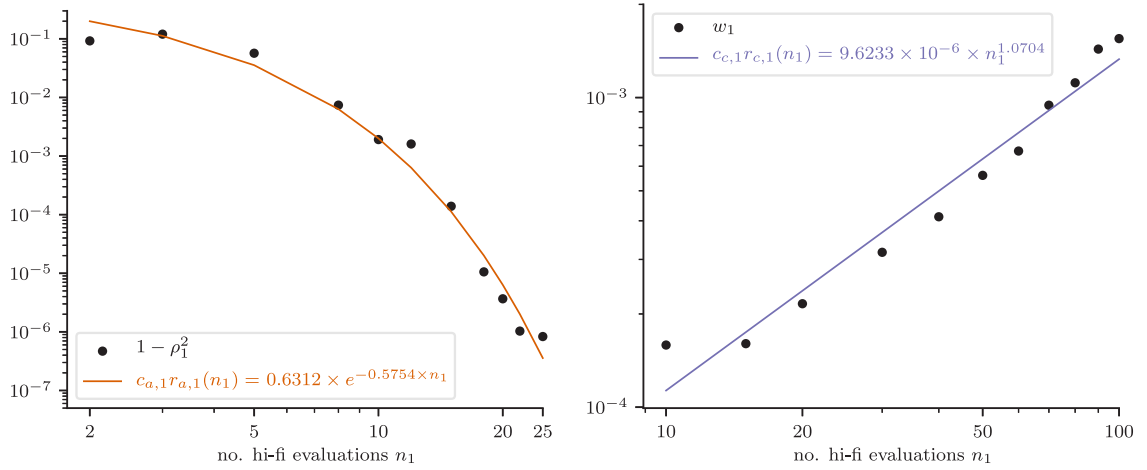
Fig. 1. Thermal block: Estimated accuracy (left) and evaluation cost rates (right) for the RB low-fidelity model.
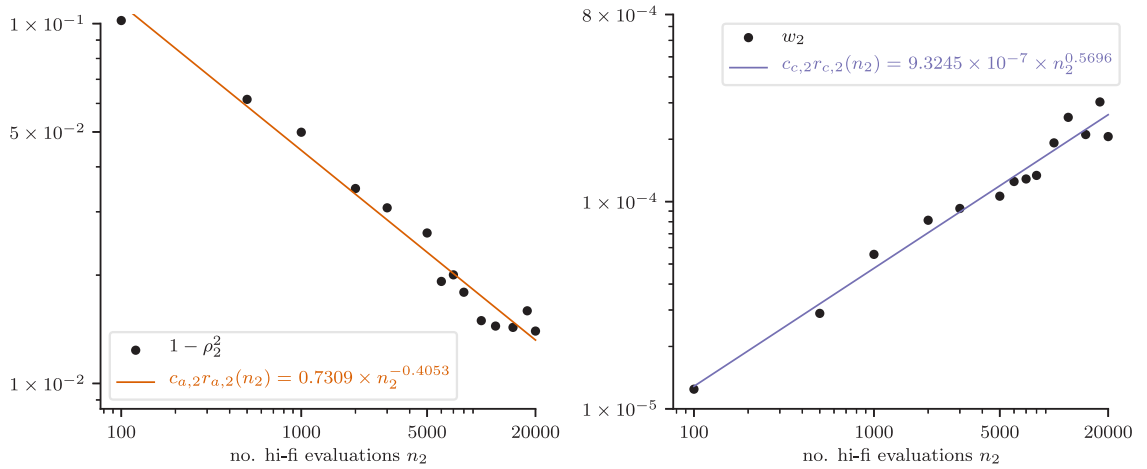
Fig. 2. Thermal block: Estimated accuracy (left) and evaluation cost rates (right) for the ε-SVR low-fidelity model.

CA-MFMC estimators considered in this work. Furthermore, samples used in pilot studies can be reused in MFMC estimators. While reusing samples can introduce a bias, the bias is typically negligible in practice [16].

4.1.3. A data-fit low-fidelity model

The second low-fidelity model $f_{n_2}^{(2)}$ that we consider in the thermal block example is an ε-support vector regression (ε-SVR) model. Our numerical implementation is based on LIBSVM [40]. The training data consists of pairs of pseudo-random realizations of the nine-dimensional input and the corresponding high-fidelity model evaluations. In our experiments, we used $\varepsilon = 10^{-2}$. We model the accuracy and evaluation cost rates as algebraic. To estimate the constants in the accuracy rate, we use 100 high- and low-fidelity evaluations. The constants in the evaluation cost rate are estimated by averaging the runtimes of the low-fidelity model evaluated at 1,000 MC samples. We perform these runtime measurements 50 times and average the results. The estimated rates are visualized in Fig. 2 and the constants are shown in Table 2. Note that the ε-SVR low-fidelity model is less accurate but cheaper to evaluate than the RB low-fidelity model.

Table 2

Thermal block: Accuracy and evaluation cost rates and their constants for the ε-SVR low-fidelity model.

Rate	Estimated rate constants
Accuracy: $c_{a,2} n_2^{-\alpha_2}$	$c_{a,2} = 0.7309$, $\alpha_2 = 0.4053$
Evaluation cost: $c_{c,2} n_2^{\beta_2}$	$c_{c,2} = 9.3245 \times 10^{-7}$, $\beta_2 = 0.5696$
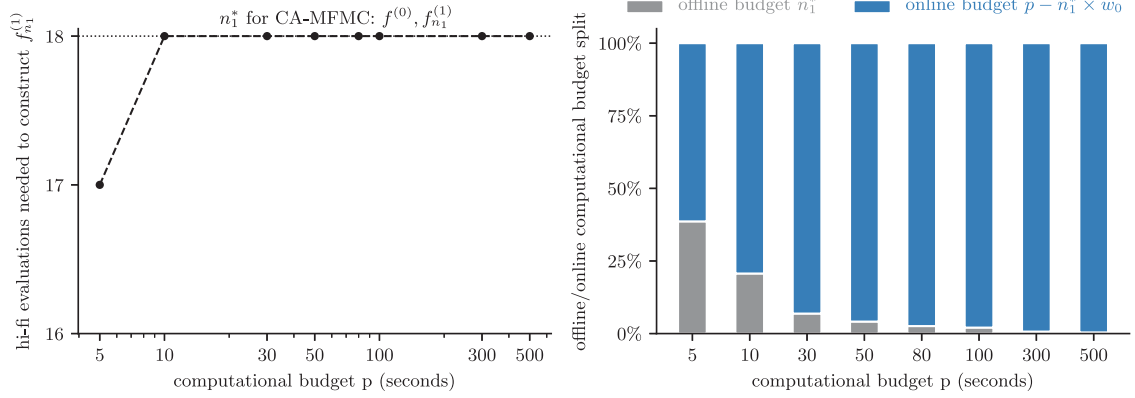
Fig. 3. Thermal block: Left, optimal number of high-fidelity training samples for training the RB low-fidelity model for the context-aware estimator. Right, visualization of how the context-aware estimator splits the computational budget between training (offline) and sampling (online).

4.1.4. Context-aware multi-fidelity Monte Carlo with RB low-fidelity model

We first consider only the RB low-fidelity model and budgets $p \in \{5, 10, 30, 50, 80, 100, 300, 500\}$ seconds. We compare, in terms of MSE, our context-aware estimator with standard MC sampling and MFMC in which the RB low-fidelity model is constructed using a fixed, a priori chosen number of basis independent of the budget p. We compute the MSEs from N replicates as

$$e_{\text{MSE}}(\hat{E}^{(\cdot)}) = \frac{1}{N} \sum_{n=1}^{N} \left(\hat{\mu}_{\text{ref}} - \hat{E}^{(\cdot)} \right)^2, \tag{4.3}$$

where $\hat{\mu}_{\text{ref}}$ represents the reference mean estimator and $\hat{E}^{(\cdot)}$ is either the CA-MFMC, standard MFMC or the MC estimator. The reference result $\hat{\mu}_{\text{ref}}$ was obtained using the context-aware multi-fidelity estimator in which the two low-fidelity models, RB and ε-SVR, were sequentially added with a computational budget $p_{\text{ref}} = 10^3$ seconds (cf. Section 4.1.5). To distinguish between the employed estimators, we use "*std. MC:* $f^{(0)}$" to denote the standard MC estimator, "*MFMC:* $f^{(0)}, f^{(1)}$" to refer to the MFMC estimator in which the RB low-fidelity model is statically constructed independent of the budget p, and "*CA-MFMC:* $f^{(0)}, f^{(1)}_{n_1}$" to refer to the context-aware multi-fidelity estimator with the RB low-fidelity model.

Based on the estimated rate parameters in Table 1, both the accuracy and cost rates of the RB low-fidelity model are strictly convex for all $n_1 \in [1, p-1]$ and any budget $p > 0$, and hence Propositions 1 and 2 apply. Therefore the optimal, context-aware number of high-fidelity model evaluations for constructing the RB low-fidelity model, n_1^*, exists, is unique and bounded with respect to the budget. The optimal number of training samples n_1^* for increasing budgets is shown in Fig. 3, on the left. Notice that $n_1^* \leq 18$ and this upper bound is attained for budgets $p \geq 10$ seconds, which is because the RB low-fidelity models is accurate already with few basis vectors. Fig. 3, right, depicts the split of the total computational budget between constructing the RB low-fidelity model (offline phase, using n_1^* high-fidelity model evaluations) and multi-fidelity sampling (online phase, using the remaining budget). Because n_1^* is bounded, the percentage of the budget allocated to constructing the RB model decreases with p.

Fig. 4 compares the MSEs of regular MC, static MFMC with RB low-fidelity models of fixed dimension 2, 8, 50, and CA-MFMC estimators. The left plot reports the analytical MSEs derived using Eqs. (2.3), (2.5) and (3.1), which can be evaluated with the constants reported in Tables 1 and 3.

Table 3
Thermal block: Correlations, evaluation costs and variances of the RB low-fidelity model $f^{(1)}$ with dimensions 2, 8, 50, respectively.

| n_1	ρ_1	w_1	σ_1^2
2	0.9277	1.0915×10^{-4}	9.4347×10^{-4}
8	0.9939	1.9791×10^{-4}	0.0015
50	0.9999	6.7539×10^{-4}	0.0018
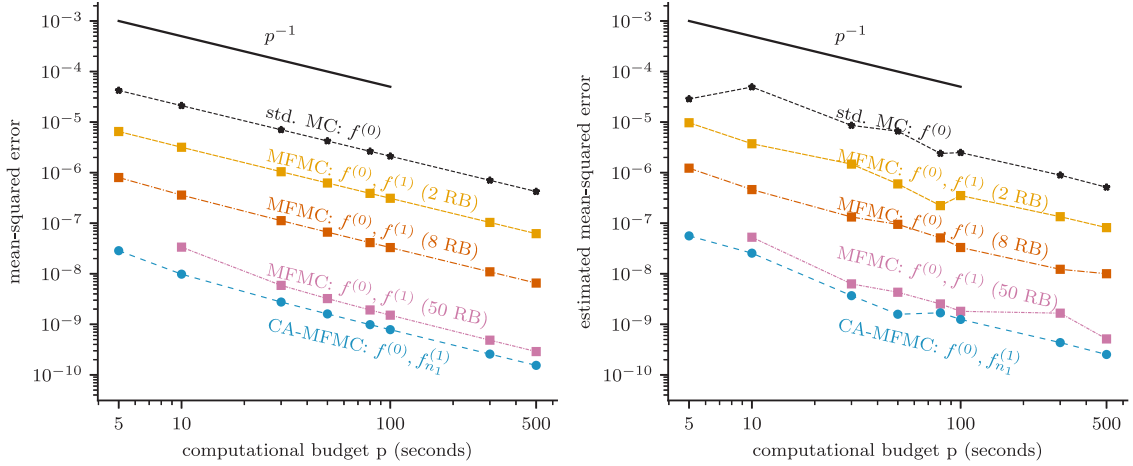
Fig. 4. Thermal block: Analytical MSE is shown left and estimated MSE is shown right. The context-aware MFMC estimator CA-MFMC achieves the lowest MSE by trading off the costs of training the low-fidelity model and sampling. Even though we see that the static MFMC with an RB model of dimension 50 performs well here, it is unclear before the error has been computed what dimension of the RB low-fidelity model is a good choice. Furthermore, the static estimator with RB model of dimension 50 requires at least 50 high-fidelity model evaluations to start with, whereas the CA-MFMC adaptively chooses the number of high-fidelity model evaluations n_1^* for training the low-fidelity model based on the available budget.

Computing the analytical MSEs requires no numerical simulations of the forward model as long as the constants in Tables 1 and 3 are available. The right plot shows the MSEs obtained numerically by taking $N = 50$ replicates of the estimators. All multi-fidelity estimators give a lower MSE than standard MC sampling by at least half an order of magnitude, even when using only two dimensions in the static RB low-fidelity model. Overall, the proposed context-aware multi-fidelity estimator gives the lowest MSE of all employed estimators. It is about four orders of magnitude more accurate than standard MC sampling in terms of MSE. Our context-aware estimator is also more accurate than the MFMC estimators with static models because 2 and 8 are too few basis vectors and 50 are too many basis vectors that make the low-fidelity model unnecessarily expensive to evaluate. In contrast, our context-aware estimator optimally trades off training and sampling and so achieves the lowest MSE of all estimators, which agrees with Propositions 1 and 2. Even though the static MFMC estimator with $n_1 = 50$ basis vectors is close to the CA-MFMC estimator, it is unclear how to choose the dimension in static MFMC a priori and if, e.g., 50 is a good choice, whereas the proposed context-aware approach provides the optimal n_1^* with respect to an upper bound of the MSE for a given budget p; cf. Section 3.2 and Fig. 3, left.

4.1.5. Context-aware multi-fidelity Monte Carlo with RB and data-fit low-fidelity models

We now sequentially add, as detailed in Section 3.3, the second low-fidelity model that is based on ε-SVR. We consider budgets of $p \in \{5, 10, 30, 50\}$ seconds. The abbreviation "*CA-MFMC:* $f^{(0)}, f_{n_1}^{(1)}, f_{n_2}^{(2)}$" refers to the CA-MFMC estimator in which the ε-SVR low-fidelity model $f_{n_2}^{(2)}$ is sequentially added after the RB low-fidelity model $f_{n_1}^{(1)}$.

As it can be seen from Table 2, the algebraic cost rate of the ε-SVR low-fidelity model is concave since $0 < \beta_2 < 1$, which implies that we need to verify whether the assumptions in Propositions 3 hold true. With definition (3.15), we consider the function $h_2(n_2) = \kappa_1 + \hat{c}_{a,2} r_{a,2}(n_2) + c_{c,2} r_{c,2}(n_2)$ with $\kappa_1 = c_{a,1} r_{a,1}(n_1^*)$ and
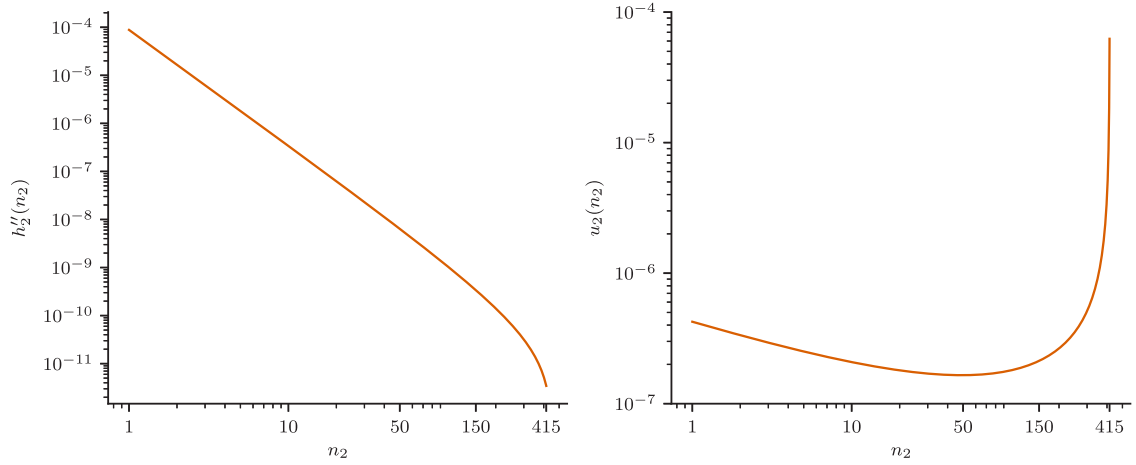
Fig. 5. Thermal block: The plot on the left demonstrates that (3.14) holds for budget $p = 50$ seconds, which corresponds to $n_2 \leq 415$. The plot on the right shows the objective u_2 defined in (3.13), which has a unique, global minimizer in $[1, 415]$.

$\hat{c}_{a,2} = c_{c,1} r_{c,1}(n_1^*) c_{a,2}$ and check if it satisfies condition (3.14). Since κ_1 and $\hat{c}_{a,2}$ are constants for a fixed n_1^*, it is hence sufficient to verify (3.14) for the largest considered budget, $p = 50$ seconds. For this budget, we have $n_1^* = 18$ (cf. Fig. 3, left) and $\kappa_1 = 2.0061 \times 10^{-5}$. Moreover, $n_2 \in [1, \lfloor p/w0 \rfloor - n_1^* - 1] = [1, 415]$. The left plot in Fig. 5 shows that $h_2''(n_2) > 0$ for all $n_2 \in [1, 415]$ which implies that (3.14) is satisfied for all budgets $0 < p \leq 50$. Therefore the number of high-fidelity model evaluations used to construct the ε-SVR low-fidelity model n_2^* exists and is unique with respect to the budget $p - n_1^* \times w_0$ that is left after adding the context-aware RB low-fidelity model. Fig. 5, right, depicts the objective $u_2(n_2)$ defined in (3.13).

Fig. 6 compares the MSE of the estimators. In the left plot, we show the analytical MSEs computed using Eqs. (2.3), (2.5) and (3.1) and the constants reported in Tables 1–3, without requiring any new numerical simulations. The MSEs reported in the right plot were computed using $N = 50$ replicates. Sequentially adding the ε-SVR low-fidelity model to the context-aware estimator further decreases the MSE. This shows that even though the slowly decreasing accuracy rate of the ε-SVR low-fidelity model indicates that it would necessitate a large training set to be accurate in single-fidelity settings for predicting high-fidelity model outputs, training it using at most $n_2^* = 50$ high-fidelity training samples in our context-aware learning approach is sufficient to achieve variance reduction and thus a more accurate multi-fidelity estimator. This indicates that even low-fidelity models with poor predictive capabilities can be useful for multi-fidelity methods as long as they capture the trend of the high-fidelity model.

Fig. 7 shows the distribution of samples among the models for budget $p = 50$ seconds. All multi-fidelity estimators allocate more than 92% of the total number of samples to low-fidelity models. Only a small percentage of number of samples correspond to the high-fidelity model, which explains the speedups.

4.2. Plasma micro-turbulence simulation in the ASDEX upgrade experiment

We now consider uncertainty quantification in a large-scale scenario concerning the simulation of plasma micro-turbulence in magnetic confinement devices. These simulations are practically relevant for efforts such as the international ITER experiment. The goal of ITER is to demonstrate that nuclear fusion, the physical process that powers the stars, such as the Sun, can be replicated in magnetic confinement devices, which will amount to a milestone on the way towards a future fusion power plant. A physics obstacle is turbulent transport driven by small-scale fluctuations which causes energy loss rates despite sophisticated plasma confinements via strong and shaped magnetic fields. Quantifying, predicting, and controlling turbulent transport is therefore essential for designing optimized fusion power plants. Key parameters that characterize the underlying micro-instabilities, such as the temperature and density gradients of the plasma species, as well as further parameters such as those characterizing the magnetic equilibrium are affected by uncertainty due to, e.g., measurement errors. Therefore, a proper quantification of uncertainty plays a crucial role in these simulations. The number of uncertain inputs can
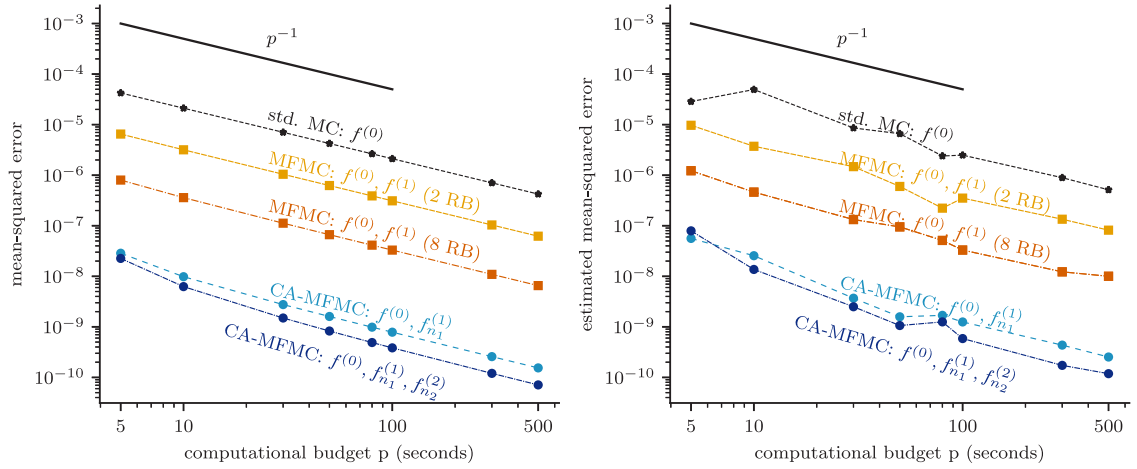
Fig. 6. Thermal block: The left plot shows the analytic MSE and the right plot showed the estimated MSE. When adding the second model $f_{n_2}^{(2)}$, the MSE of the CA-MFMC estimator is further reduced compared to the CA-MFMC estimator with only one low-fidelity model.
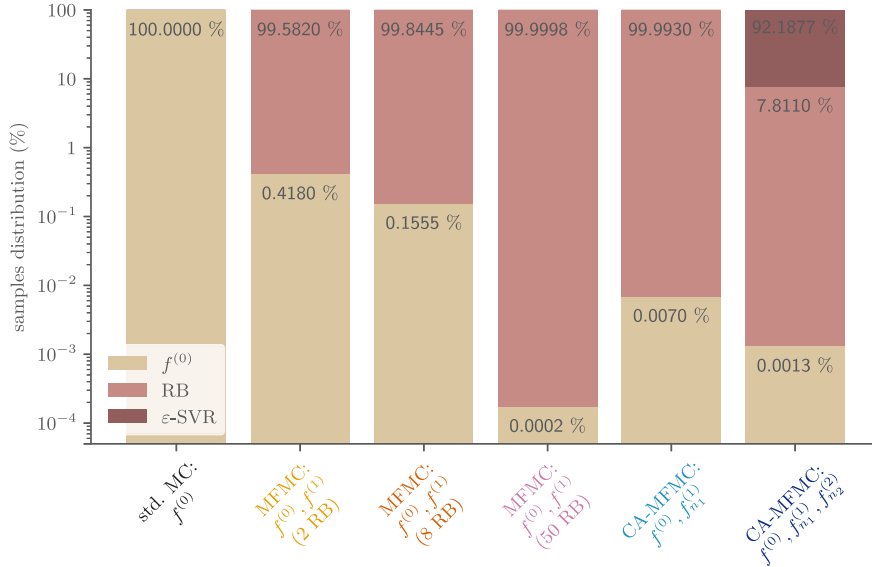
Fig. 7. Thermal block: Sample distribution across models for budget $p = 50$ seconds.

nevertheless amount to a few tens or more, prohibiting the use of brute-force approaches. This together with the large computational costs of the underlying high-fidelity models make multi-fidelity methods especially relevant for accelerating or even enabling uncertainty quantification in plasma micro-turbulence simulations.

4.2.1. Setup of the experiment

We focus on linear (in the phase space variables) gyrokinetic simulations, which are used to characterize the underlying micro-instabilities. Gyrokinetic models reduce the dimension of the phase space of classical kinetic models from six (three positions, three velocities) to five (three positions, two velocities), which leads to the gyrokinetic Vlasov–Maxwell system of nonlinear integro-differential equations; we refer to [41] and the references therein for details.

We consider here a parameter set from a discharge from the ASDEX Upgrade Experiment, which is similar to the parameter set used in [42] for a validation study of gyrokinetic simulations. The experiment considered

Table 4
ASDEX Upgrade Experiment: The 12 uniformly distributed uncertain inputs.

θ	Parameter name	Symbol	Left bound	Right bound
θ_1	Plasma β	β	0.4889×10^{-3}	0.5975×10^{-3}
θ_2	Collision frequency	ν_c	0.6412×10^{-2}	0.8676×10^{-2}
θ_2	Ion and electron density gradient	ω_n	1.1563	1.9272
θ_4	Ion temperature gradient	ω_{T_i}	2.0966	3.4943
θ_5	Temperature ratio	T_i/T_e	0.6142	0.6788
θ_6	Electron temperature gradient	ω_{T_e}	4.0403	6.7338
θ_7	Effective ion charge	Z_{eff}	1.2800	1.9200
θ_8	Safety factor	q	2.1708	2.3993
θ_9	Magnetic shear	$\hat{s}$	1.9927	2.4356
θ_{10}	Elongation	k	1.2815	1.4165
θ_{11}	Elongation gradient	s_k	0.2081	0.2544
θ_{12}	Triangularity	δ	0.0528	0.0583

here is characterized by two particle species, (deuterium) ions and electrons. Moreover, the magnetic geometry is described by an analytical Miller equilibrium [43]. We consider both electromagnetic effects and collisions. Collisions are modeled by a linearized Landau–Boltzmann operator. The total number of uncertain inputs is 12, which are modeled as given in Table 4.

We perform experiments at one bi-normal wave-number $k_y \rho_s = 0.8$. The output of interest is the growth rate (spectral abscissa) of the dominant micro-instability mode, which corresponds to the maximum real part over the spectrum. For this parameter set, the dominant micro-instability mode at $k_y \rho_s = 0.8$ and at the nominal values of the input parameters is electron-temperature-gradient-driven, characterized by a negative frequency (imaginary part), while the first subdominant mode is ion-temperature-gradient-driven, characterized by a positive frequency. However, for certain values of the ion temperature and density gradients within the bounds showed in Table 4, the first subdominant ion mode can transition to become the dominant mode. This mode transition can be sharp and in turn lead to sharp transitions or even discontinuities in the growth rate. To avoid the potentially detrimental effects of this discontinuity, in our experiments, we determine the growth rate of the underlying electron-driven micro-instability mode. To this end, we compute the first two micro-instability eigenmodes and then select the growth rate that has a corresponding negative frequency.

4.2.2. High-fidelity model

The high-fidelity model is provided by the gyrokinetic simulation code GENE [37] in the flux-tube limit [44]. This assumes periodic boundary conditions over the radial x and bi-normal ky directions. To discretize the five dimensional gyrokinetic phase-space domain, we use 15 Fourier modes in the radial direction and 24 points along the field line, in the z direction. The bi-normal coordinate ky is decoupled in linear flux-tube simulations and hence only one such mode is used here. In velocity space, we employ 48 equidistant and symmetric parallel velocity grid points and 16 Gauss–Laguerre distributed magnetic moment points. This results in a total of $15 \times 1 \times 24 \times 48 \times 16 = 276,480$ degrees of freedom in the 5D phase space.

The average high-fidelity model runtime is 410.9941 s on 32 cores on one node on the Lonestar6 supercomputer at the Texas Advanced Computing Center; cf. Section 1. One such node comprises two AMD EPYC 7763 64-Core Processors for a total of 128 cores per socket, and 256 GB of RAM. The high-fidelity variance, estimated using 20 MC samples, is $\sigma_0^2 \approx 0.0279$.

4.2.3. A coarse-grid low-fidelity model

We first consider a coarse-grid (CG) low-fidelity model obtained by coarsening the grid in z direction (20 points), $v_{||}$ direction (20 points) and μ direction (8 Gauss–Laguerre points), for a total of $48,000 = 15 \times 20 \times 20 \times 8$ degrees of freedom, which corresponds to $5.76 \times$ fewer degrees of freedom than the high-fidelity model. We note that this represents still a useful resolution for this experiment while further considerable coarsening of the underlying mesh could lead to significant loss of accuracy or potentially even unphysical solutions. The average evaluation cost of relative to the high-fidelity model is $w_{\text{CG}} = 0.0849$ on 16 cores on one node on Lonestar6. The estimates of the variance and correlation coefficient, approximated using 20 MC high- and low-fidelity model evaluations,
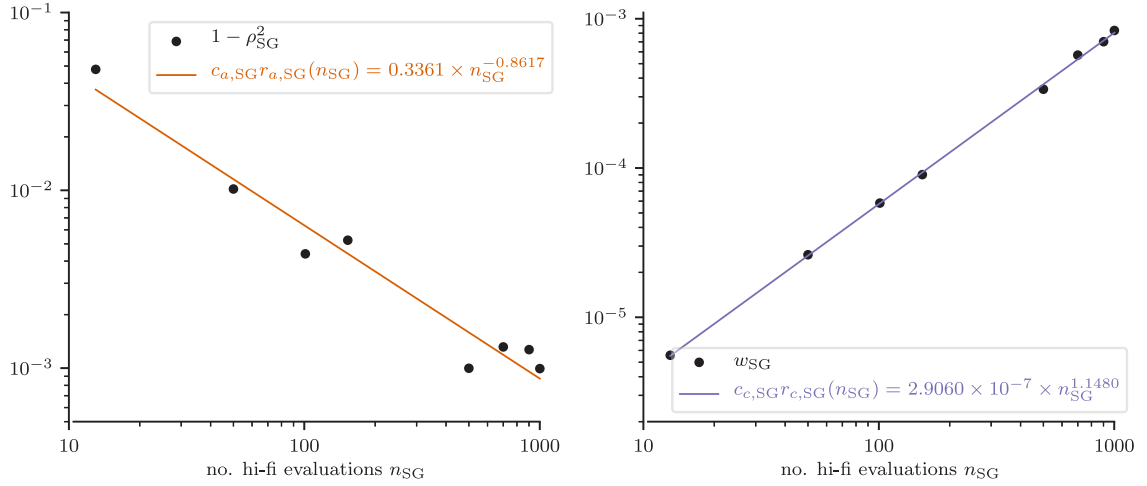
Fig. 8. ASDEX Upgrade Experiment: Estimated accuracy (left) and evaluation cost rates (right) for the SG low-fidelity model.

Table 5
ASDEX Upgrade Experiment: Estimated accuracy and evaluation cost rates, and their constants for the SG low-fidelity model.

Rate	Estimated rate constants
Accuracy: $c_{a,\mathrm{SG}} n_{\mathrm{SG}}^{-\alpha_{\mathrm{SG}}}$	$c_{a,\mathrm{SG}} = 0.3361,\ \alpha_{\mathrm{SG}} = 0.8617$
Evaluation cost: $c_{c,\mathrm{SG}} n_{\mathrm{SG}}^{\beta_{\mathrm{SG}}}$	$c_{c,\mathrm{SG}} = 2.9060 \times 10^{-7},\ \beta_{\mathrm{SG}} = 1.1480$

are $\sigma_{\mathrm{CG}}^2 \approx 0.0256$ and $\rho_{\mathrm{CG}} \approx 0.9990$, respectively. Because this model requires no training, and thus no high-fidelity model evaluations to generate training data, we can set the rate functions constant as $r_{a,\mathrm{CG}} \equiv 1 - \rho_{\mathrm{CG}}^2$ and $r_{c,\mathrm{CG}} \equiv w_{\mathrm{CG}}$, as well as $c_{a,\mathrm{CG}} = 1$ and $c_{c,\mathrm{CG}} = 1$.

Single-fidelity settings require fine discretizations for obtaining results with desired accuracy. Here, we want to show that when coarse-grid low-fidelity models are correlated with the high-fidelity model, their lower computational cost can be leveraged for speeding up multi-fidelity sampling methods. This is particularly relevant in plasma physics simulations in which usually only a limited number of high-fidelity simulations can be performed for training data-fit low-fidelity models. Also, we want to show that our algorithm has a wide scope in the sense that a wide variety of low-fidelity models can be used.

4.2.4. A sensitivity-driven dimension-adaptive sparse grid low-fidelity model

We now additionally consider a data-driven low-fidelity model based on sensitivity-driven dimension-adaptive interpolation on sparse grids (SG) [4,38], with code available.[5] SG low-fidelity models have been successfully used for uncertainty quantification in plasma micro-instability simulations [4,38,45], including computationally expensive nonlinear simulations [18]. The adaptive procedure to construct the SG low-fidelity model is terminated when the underlying sparse grid reaches cardinality n_{SG}^*, which corresponds to n_{SG}^* high-fidelity training samples, where n_{SG}^* is determined as discussed in Section 3. We note that even though the interpolation points used to construct the SG low-fidelity model have fine granularity (see [4,38]), in cases where we cannot have exactly n_{SG}^* sparse-grid points, we take the closest, feasible number of grid points to n_{SG}^*.

We use 20 high- and low-fidelity model evaluations to asses the accuracy rate of the SG low-fidelity model; see also [16]. To estimate the evaluation cost rate, we average the runtimes of the low-fidelity model evaluated at 1,000 MC samples. As can be seen in Fig. 8 and Table 5, accuracy and evaluation cost can be modeled well with an algebraic rate.

---

[5] https://github.com/ionutfarcas/sensitivity-driven-sparse-grid-approx
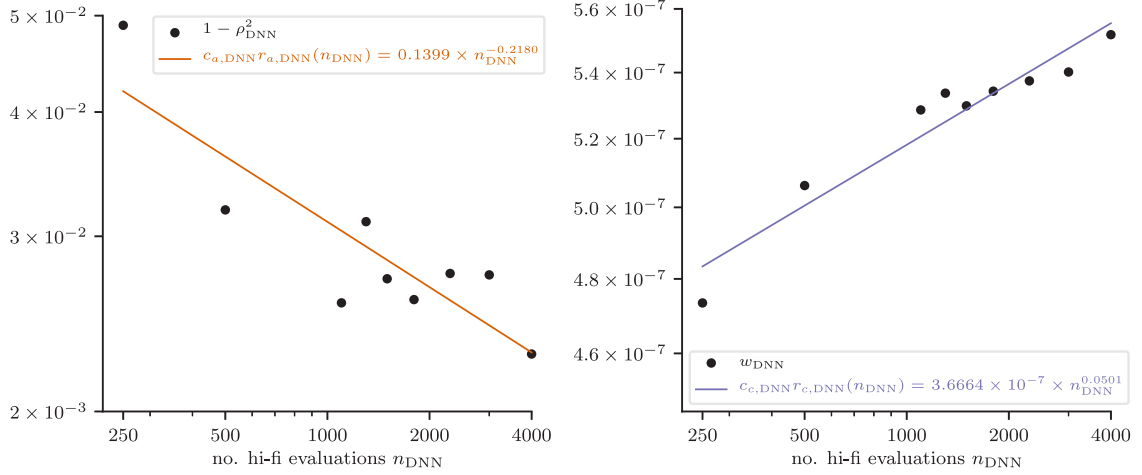
Fig. 9. ASDEX Upgrade Experiment: Estimated accuracy (left) and evaluation cost rates (right) for the DNN low-fidelity model.

Table 6
ASDEX Upgrade Experiment: Estimated accuracy and evaluation cost rates, and their constants for the DNN low-fidelity model.

| Rate	Estimated rate constants
Accuracy: $c_{a,\mathrm{DNN}} n_{\mathrm{DNN}}^{-\alpha_{\mathrm{DNN}}}$	$c_{a,\mathrm{DNN}} = 0.1399$, $\alpha_{\mathrm{DNN}} = 0.2180$
Evaluation cost: $c_{c,\mathrm{DNN}} n_{\mathrm{DNN}}^{\beta_{\mathrm{DNN}}}$	$c_{c,\mathrm{DNN}} = 3.6664 \times 10^{-7}$, $\beta_{\mathrm{DNN}} = 0.0501$

4.2.5. A deep learning low-fidelity model

A second data-driven low-fidelity model is given by a fully connected, feed-forward three-layer deep neural network (DNN) with ReLU activation functions that is trained on high-fidelity data. The number of nodes of the hidden layers is $\lfloor 0.05 \times n_{\mathrm{DNN}} \rfloor$ and thus depends on the number of training samples n_{DNN}. Training is done over 20 epochs using an Adam optimizer in which the loss function is the MSE. The accuracy rate is estimated from 20 high- and low-fidelity model evaluations and the cost rate from 1,000 low-fidelity model evaluations; see Fig. 9 and Table 6. Notice that within the considered training size, the accuracy rate decreases rather slowly, at an algebraic rate of only about 0.2. This indicates that a large training size is required for obtaining accurate approximations should the DNN model be used to replace the high-fidelity model. In addition, the evaluation cost rate increases very slowly, which is because the network has few layers only and the employed libraries are highly optimized.

4.2.6. Context-aware multi-fidelity estimation

We consider the computational budget $p = 5 \times 10^5$ seconds, which corresponds to 1,216 high-fidelity model evaluations and is the limit of the computational resources available to us for this project. To determine which combinations of models lead to CA-MFMC estimators with low MSEs for a budget p, we plot the analytic MSEs in Fig. 10. The CA-MFMC estimator with the high-fidelity model $f^{(0)}$, the coarse-grid low-fidelity model $f^{(1)} = f^{(\mathrm{CG})}$, and either the SG low-fidelity model $f_{n_2}^{(2)} = f_{n_2}^{(\mathrm{SG})}$ or the DNN low-fidelity model $f_{n_2}^{(2)} = f_{n_2}^{(\mathrm{DNN})}$ lead to lower MSEs than the CA-MFMC estimator with models $f^{(0)}, f_{n_1}^{(\mathrm{SG})}, f_{n_2}^{(\mathrm{DNN})}$, which omits the coarse-grid model $f^{(\mathrm{CG})}$. We therefore consider the CA-MFMC estimators with $f^{(0)}, f^{(1)} = f^{(\mathrm{CG})}$ and either $f_{n_2}^{(2)} = f_{n_2}^{(\mathrm{SG})}$ or $f_{n_2}^{(2)} = f_{n_2}^{(\mathrm{DNN})}$ (cf. Section 3.3). Note that all multi-fidelity estimators achieve a lower MSE than the standard Monte Carlo estimator that uses the high-fidelity model alone. This experiment shows that the analytic MSE can be used to select model combinations.

Based on the estimated constants in Table 5, the accuracy and cost rates of the SG low-fidelity model are both strictly convex on $[1, p-1]$ and any budget $p > 0$, and therefore Propositions 3 and 4 apply. Hence, the optimal, context-aware number of high-fidelity model evaluations for constructing the SG low-fidelity model exists, is unique
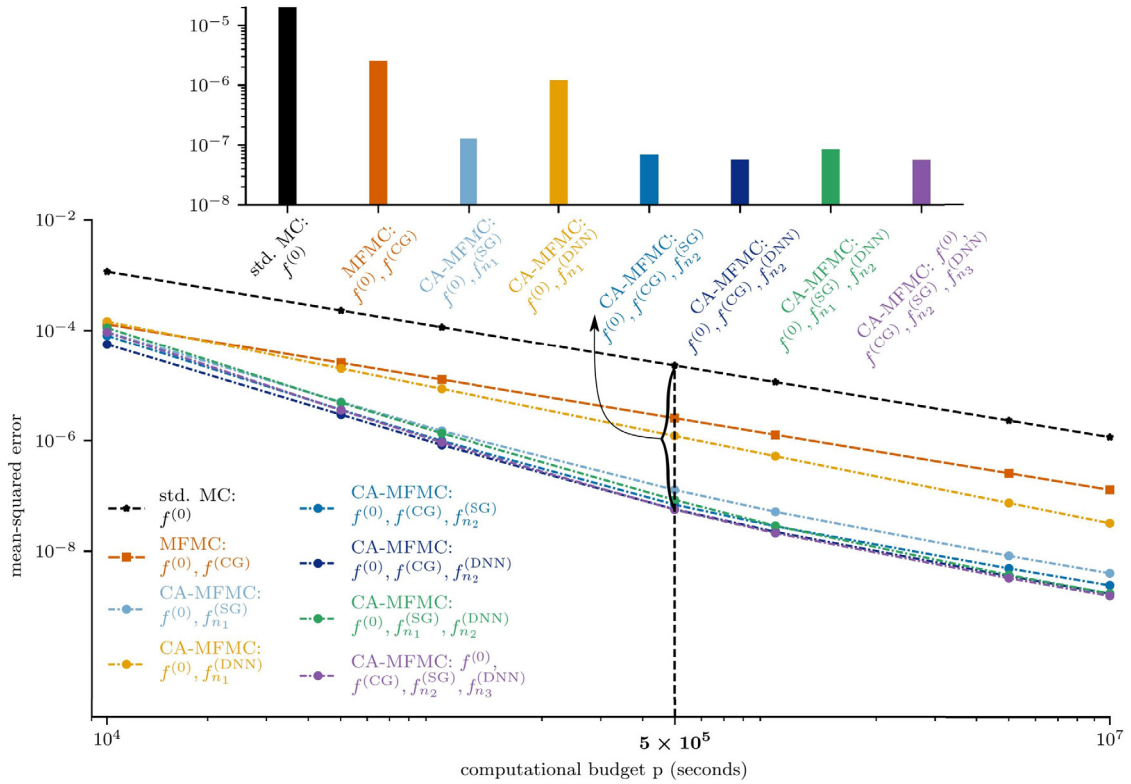
Fig. 10. ASDEX Upgrade Experiment: Combinations of models can be selected based on the analytical MSE, which can be computed without additional model evaluations as long as the accuracy and cost rates are available. In this example, the CA-MFMC estimator with the high-fidelity model $f^{(0)}$, the coarse-grid low-fidelity model $f^{(1)} = f^{(CG)}$ and either the SG low-fidelity model $f_{n_2}^{(2)} = f_{n_2}^{(SG)}$ or the DNN low-fidelity model $f_{n_2}^{(2)} = f_{n_2}^{(DNN)}$ have the highest (best) cost/error ratio, which indicates a good model selection.

and bounded with respect to the budget. Let us now consider the DNN low-fidelity model when used in the estimator with models $f^{(0)}, f^{(1)} = f^{CG}, f_{n_2}^{(2)} = f_{n_2}^{(2)}$. We see from Table 6 that the algebraic cost rate of the DNN low-fidelity model is concave since $0 < \beta_{DNN} < 1$, which means that we need to verify whether the assumptions of Proposition 3 hold. To this end, we verify whether the function $h_2(n_2) = \kappa_1 + \hat{c}_{a,2}r_{a,2}(n_2) + c_{c,2}r_{c,2}(n_2)$ satisfies condition (3.14), where $c_{c,2} = c_{c,DNN}, r_{a,2} \equiv r_{a,DNN}$, and $r_{c,2} \equiv r_{c,DNN}$. When computing κ_1 and $\hat{c}_{a,2}$, recall from Section 4.2.3 that model $f^{(1)} = f^{(CG)}$ requires no training and therefore has constant costs $w_1 = w_{CG}$ and constant correlation coefficient $\rho_1 = \rho_{CG}$. This means that the rate functions $r_{a,1} \equiv 1 - \rho_1^2$ and $r_{c,1} \equiv w_1$ are constant with $c_{a,1} = 1$ and $c_{c,1} = 1$ and thus $\kappa_1 = 1 - \rho_1^2$ and $\hat{c}_{a,2} = c_{a,DNN}$. Because κ_1 and $\hat{c}_{a,2}$ are fixed when optimizing for n_2, the function h_2 is independent of the budget p and thus it suffices to verify this condition for budget $p = 10^7$ seconds, the largest budget considered in Fig. 10, for which $n_2 \in [1; 24, 330]$. The left plot in Fig. 11 shows that $h_2''(n_2) > 0$ for all $n_2 \in [1; 24, 330]$ which implies that (3.14) is satisfied for all budgets $0 < p \le 10^7$, including the budget considered in our computations, $p = 5 \times 10^5$ seconds. Therefore Propositions 3 applies and the optimal, context-aware number of high-fidelity model evaluations for constructing the DNN low-fidelity model, n_2^*, exists and is unique. The right plot in Fig. 11 depicts the objective u_2 defined in (3.13), which has a unique global minimizer.

In the left plot of Fig. 12, we show the number of training samples for the CA-MFMC estimators with models $f^{(0)}, f_{n_1}^{(1)} = f_{n_1}^{(SG)}$ and $f^{(0)}, f^{(1)} = f^{(CG)}, f_{n_2}^{(2)} = f_{n_2}^{(SG)}$. In the right plot of Fig. 12, the number of training samples n_2^* for the DNN low-fidelity model $f_{n_2}^{(2)} = f_{n_2}^{(DNN)}$ used for the CA-MFMC estimator with models $f^{(0)}, f^{(1)} = f^{(CG)}, f_{n_2}^{(2)} = f_{n_2}^{(DNN)}$ is shown. Training the DNN low-fidelity model with only $n_2^* = 159$ samples in case of our budget $p = 5 \times 10^5$ is sufficient to obtain a low-fidelity model that is useful for variance reduction in the CA-MFMC estimator together with the high-fidelity model, even though the low-fidelity model alone is insufficient to provide accurate predictions in single-fidelity settings.
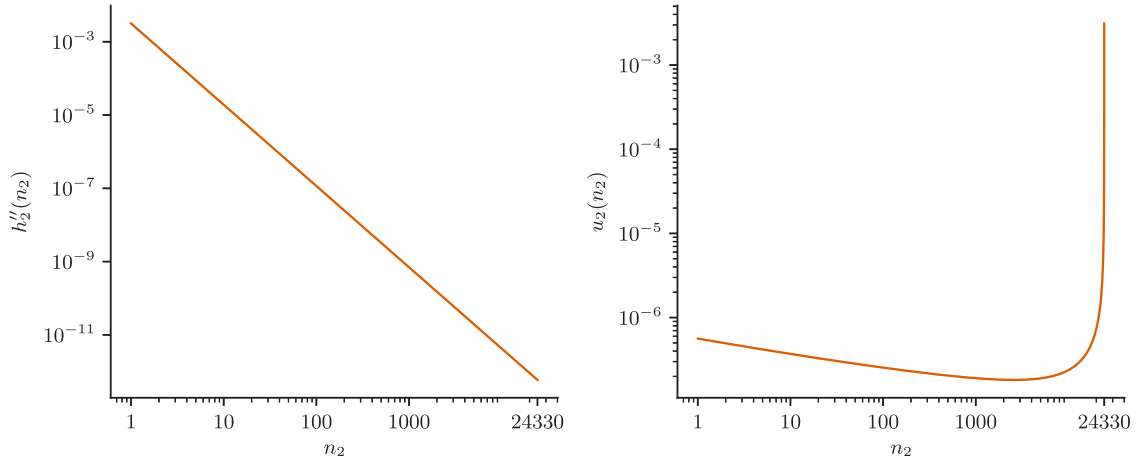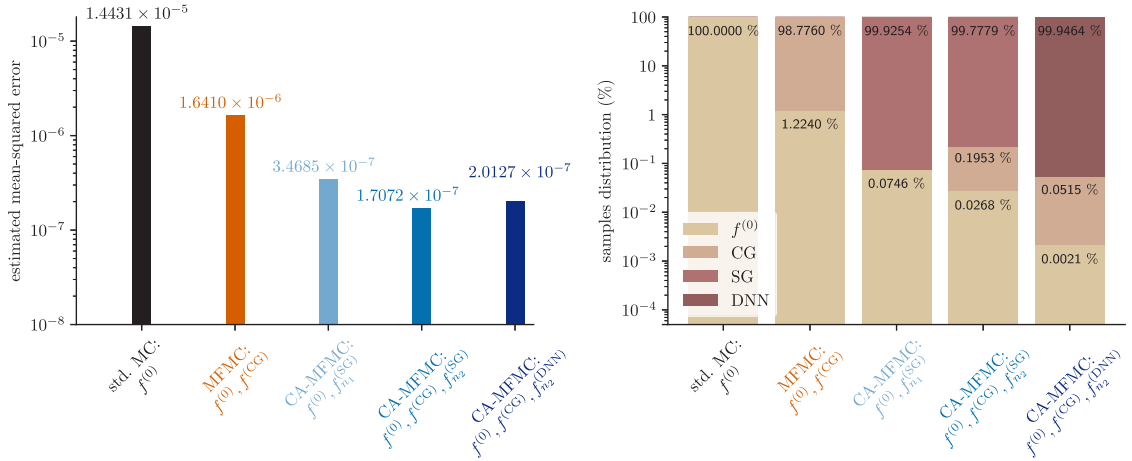
Fig. 11. ASDEX Upgrade Experiment: On the left, we verify that $h_2'' > 0$ satisfies (3.14) up to budget $p = 10^7$ seconds. On the right, we plot the objective u_2 defined in (3.13), which has a unique global minimizer.
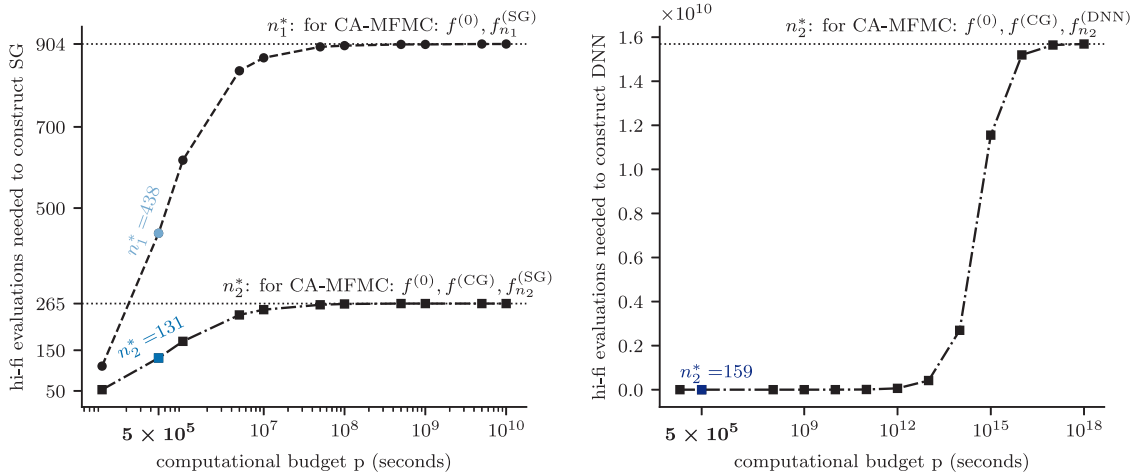
Fig. 12. ASDEX Upgrade Experiment: The number of high-fidelity training samples selected by the proposed context-aware approach for obtaining a multi-fidelity estimator with highest (best) cost/error ratio.

Consider now the left plot of Fig. 13 that compares the MSEs of the CA-MFMC estimators to static MFMC and regular MC. The reference solution was obtained via the MFMC estimator depending on the coarse-grid and DNN low-fidelity models for a budget of $p = 3 \times 10^6$ seconds using $N = 5$ replicates; here we needed $n_2^* = 847$ high-fidelity evaluations to construct the DNN low-fidelity model. The average reference value of the expected growth rate of the dominant electron temperature gradient-driven micro-instability mode is $\hat{\mu}_{\mathrm{ref}} = 0.3967$. The context-aware estimators are up to two orders of magnitude more accurate than the standard MC estimator. The roughly two orders of magnitude speedup of the CA-MFMC estimators translate into a runtime reduction from 72 days to roughly four hours on 32 cores on one node of the Lonestar6 supercomputer. We also observe that MFMC with the high-fidelity and coarse-grid model is about one order of magnitude more accurate in terms of the MSE than the standard MC estimator, showing that coarse-grid plasma physics models can be useful in multi-fidelity settings. The trend that we observe here numerically are in alignment with the analytic MSEs shown in Fig. 10.

The right plot of Fig. 13 shows how the number of samples are distributed among the models, noting that the static MC and MFMC estimators use the entire computational budget for sampling only, whereas the context-aware estimators use the budget that remains after the low-fidelity models are constructed. All multi-fidelity estimators allocate more than 98% of the total number of samples to low-fidelity models, which explains the speedups.

Fig. 13. ASDEX Upgrade Experiment results for budget $p = 5 \times 10^5$ seconds: The left plots show that the CA-MFMC estimators achieve about two orders of magnitude speedup, which translates into a runtime reduction from 72 days to roughly four hours on one node of the Lonestar6 supercomputer. The right plot shows the sample distribution across models in the employed estimators. The static MC and MFMC estimators use the entire computational budget for sampling only, whereas the context-aware estimators use the remaining budget after the low-fidelity models are constructed.

However, there is still a small percentage corresponding to the high-fidelity model, which is desirable in practical applications to achieve unbiasedness of the estimators.

5. Conclusions

The proposed context-aware learning approach for variance reduction trades off training hierarchies of low-fidelity models with Monte Carlo estimation. By leveraging the trade-off between either adding a high-fidelity model sample to the training set for constructing low-fidelity models versus using the high-fidelity model sample in the Monte Carlo estimator, our analysis shows that the quasi-optimal number of training samples from the high-fidelity model is bounded independent of the computational budget and thus that after a finite number of samples, all high-fidelity model samples should be used in the Monte Carlo estimator rather than being used to improve the low-fidelity model. This means that low-fidelity models can be over-trained for multi-fidelity uncertainty quantification. We have demonstrated the over-training numerically in a thermal block scenario, in which a reduced basis low-fidelity model constructed from 50 high-fidelity training samples led to a multi-fidelity estimator with a poorer cost/error ratio than the proposed context-aware estimator that determined 18 to be the maximum quasi-optimal number of training samples. This has clear implications on training data-fit and machine-learning-based models too, which can require large data sets for achieving high prediction accuracy but are sufficient to be trained with few data points when they are merely used for variance reduction together with the high-fidelity model in multi-fidelity uncertainty quantification, as we have shown in a numerical experiment with plasma micro-turbulence simulations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code and data to reproduce our numerical results are available at https://github.com/ionutfarcas/context-aware-mfmc.

Acknowledgments

I.-G.F. and F.J. were partially supported by the Exascale Computing Project, USA (No. 17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration. B.P. acknowledges support from the Air Force Office of Scientific Research (AFOSR), USA award FA9550-21-1-0222 (Dr. Fariba Fahroo). We thank Tobias Goerler for useful discussions and insights about the considered plasma micro-turbulence simulation scenario. We also gratefully acknowledge the compute and data resources provided by the Texas Advanced Computing Center at The University of Texas at Austin (https://www.tacc.utexas.edu/).

References

[1] B. Peherstorfer, K. Willcox, M. Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization, SIAM Rev. 60 (3) (2018) 550–591.

[2] B. Peherstorfer, Multifidelity Monte Carlo estimation with adaptive low-fidelity models, SIAM/ASA J. Uncertain. Quantif. 7 (2) (2019) 579–603.

[3] T. Alsup, B. Peherstorfer, Context-aware surrogate modeling for balancing approximation and sampling costs in multi-fidelity importance sampling and Bayesian inverse problems, 2020, arXiv 2010.11708.

[4] I.G. Farcas, Context-Aware Model Hierarchies for Higher-Dimensional Uncertainty Quantification (Ph.D. thesis), Technical University of Munich, 2020.

[5] S. Werner, B. Peherstorfer, Context-aware controller inference for stabilizing dynamical systems from scarce data, 2022, arXiv 2207.11049.

[6] N. Shyamkumar, S. Gugercin, B. Peherstorfer, Towards context-aware learning for control: Balancing stability and model-learning error, in: 2022 American Control Conference, ACC, 2022, pp. 4808–4813.

[7] P. Benner, S. Gugercin, K. Willcox, A survey of projection-based model reduction methods for parametric dynamical systems, SIAM Rev. 57 (4) (2015) 483–531.

[8] G. Rozza, D. Huynh, A. Patera, Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations, Arch. Comput. Methods Eng. 15 (3) (2007) 1–47.

[9] A. Gruber, M. Gunzburger, L. Ju, R. Lan, Z. Wang, Multifidelity Monte Carlo estimation for efficient uncertainty quantification in climate-related modeling, EGUsphere 2022 (2022) 1–27.

[10] A. Gruber, M. Gunzburger, L. Ju, Z. Wang, A multifidelity Monte Carlo method for realistic computational budgets, 2022, arXiv 2206.07572.

[11] P. Khodabakhshi, K.E. Willcox, M. Gunzburger, A multifidelity method for a nonlocal diffusion model, Appl. Math. Lett. 121 (2021) 107361.

[12] L. Ng, K. Willcox, Multifidelity approaches for optimization under uncertainty, Internat. J. Numer. Methods Engrg. 100 (10) (2014) 746–772.

[13] B. Peherstorfer, M. Gunzburger, K. Willcox, Convergence analysis of multifidelity Monte Carlo estimation, Numer. Math. 139 (3) (2018) 683–707.

[14] B. Peherstorfer, K. Willcox, M. Gunzburger, Optimal model management for multifidelity Monte Carlo estimation, SIAM J. Sci. Comput. 38 (5) (2016) A3163–A3194.

[15] E. Qian, B. Peherstorfer, D. O'Malley, V. Vesselinov, K. Willcox, Multifidelity Monte Carlo estimation of variance and sensitivity indices, SIAM/ASA J. Uncertain. Quantif. 6 (2) (2018) 683–706.

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

[17] F. Law, A. Cerfon, B. Peherstorfer, Accelerating the estimation of collisionless energetic particle confinement statistics in stellarators using multifidelity Monte Carlo, Nucl. Fusion 62 (7) (2022) 076019.

[18] I.G. Farcaş, G. Merlo, F. Jenko, A general framework for quantifying uncertainty at scale, Commun. Eng. 1 (1) (2022) 43.

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

[20] D. Schaden, E. Ullmann, On multilevel best linear unbiased estimators, SIAM/ASA J. Uncertain. Quantif. 8 (2) (2020) 601–635.

[21] D. Schaden, E. Ullmann, Asymptotic analysis of multilevel best linear unbiased estimators, SIAM/ASA J. Uncertain. Quantif. 9 (3) (2021) 953–978.

[22] J. Hampton, H.R. Fairbanks, A. Narayan, A. Doostan, Practical error bounds for a non-intrusive bi-fidelity approach to parametric/stochastic model reduction, J. Comput. Phys. 368 (2018) 315–332.

[23] A. Narayan, C. Gittelson, D. Xiu, A stochastic collocation algorithm with multifidelity models, SIAM J. Sci. Comput. 36 (2) (2014) A495–A521.

[24] A.A. Gorodetsky, J.D. Jakeman, G. Geraci, MFNets: data efficient all-at-once learning of multifidelity surrogates as directed networks of information sources, Comput. Mech. 68 (4) (2021) 741–758.

[25] A.A. Gorodetsky, J.D. Jakeman, G. Geraci, M.S. Eldred, MFNets: Multi-fidelity data-driven networks for Bayesian learning and prediction, Int. J. Uncertain. Quantif. 10 (6) (2020) 595–622.

[26] S. De, J. Britton, M. Reynolds, R. Skinner, K. Jansen, A. Doostan, On transfer learning of neural networks using bi-fidelity data for uncertainty propagation, Int. J. Uncertain. Quantif. 10 (6) (2020) 543–573.

[27] F. Newberry, J. Hampton, K. Jansen, A. Doostan, Bi-fidelity reduced polynomial chaos expansion for uncertainty quantification, Comput. Mech. 69 (2) (2022) 405–424.

[28] M. Razi, R.M. Kirby, A. Narayan, Fast predictive multi-fidelity prediction with models of quantized fidelity levels, J. Comput. Phys. 376 (2019) 992–1008.

[29] Y. Xu, V. Keshavarzzadeh, R.M. Kirby, A. Narayan, A bandit-learning approach to multifidelity approximation, SIAM J. Sci. Comput. 44 (1) (2022) A150–A175.

[30] K.A. Cliffe, M.B. Giles, R. Scheichl, A.L. Teckentrup, Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients, Comput. Vis. Sci. 14 (1) (2011) 3–15.

[31] M.B. Giles, Multilevel Monte Carlo path simulation, Oper. Res. 56 (3) (2008) 607–617.

[32] A.L. Teckentrup, R. Scheichl, M.B. Giles, E. Ullmann, Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients, Numer. Math. 125 (3) (2013) 569–600.

[33] N. Collier, A.L. Haji-Ali, F. Nobile, E. von Schwerin, R. Tempone, A continuation multilevel Monte Carlo algorithm, BIT Numer. Math. 55 (2) (2015) 399–432.

[34] M. Eigel, C. Merdon, J. Neumann, An adaptive multilevel Monte Carlo method with stochastic bounds for quantities of interest with uncertain data, SIAM/ASA J. Uncertain. Quantif. 4 (1) (2016) 1219–1245.

[35] A.L. Haji-Ali, F. Nobile, E. von Schwerin, R. Tempone, Optimization of mesh hierarchies in multilevel Monte Carlo samplers, Stoch. Partial. Differ. Equ. Anal. Comput. 4 (1) (2016) 76–112.

[36] H. Yang, Y. Fujii, K.W. Wang, A.A. Gorodetsky, Control variate polynomial chaos: Optimal fusion of sampling and surrogates for multifidelity uncertainty quantification, 2022.

[37] F. Jenko, W. Dorland, M. Kotschenreuther, B.N. Rogers, Electron temperature gradient driven turbulence, Phys. Plasmas 7 (5) (2000) 1904–1910.

[38] I.G. Farcaş, T. Görler, H.J. Bungartz, F. Jenko, T. Neckel, Sensitivity-driven adaptive sparse stochastic approximations in plasma microinstability analysis, J. Comput. Phys. 410 (2020) 109394.

[39] P. Binev, A. Cohen, W. Dahmen, R. DeVore, G. Petrova, P. Wojtaszczyk, Convergence rates for greedy algorithms in reduced basis methods, SIAM J. Math. Anal. 43 (3) (2011) 1457–1472.

[40] C.C. Chang, C.J. Lin, LIBSVM: A library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 27:1–27:27.

[41] A. Brizard, T. Hahm, Foundations of nonlinear gyrokinetic theory, Rev. Modern Phys. 79 (2007).

[42] S.J. Freethy, T. Görler, A.J. Creely, G.D. Conway, S.S. Denk, T. Happel, C. Koenen, P. Hennequin, A.E. White, Validation of gyrokinetic simulations with measurements of electron temperature fluctuations and density-temperature phase angles on ASDEX upgrade, Phys. Plasmas 25 (5) (2018) 055903.

[43] R. Miller, M. Chu, J. Greene, Y. Lin-Liu, R. Waltz, Noncircular, finite aspect ratio, local equilibrium model, Phys. Plasmas 5 (1998) 979.

[44] T. Dannert, F. Jenko, Gyrokinetic simulation of collisionless trapped-electron mode turbulence, Phys. Plasmas 12 (7) (2005) 072309.

[45] I.G. Farcaş, A.D. Siena, F. Jenko, Turbulence suppression by energetic particles: a sensitivity-driven dimension-adaptive sparse grid framework for discharge optimization, Nucl. Fusion 61 (5) (2021) 056004.