SIAM J. SCI. COMPUT.
Vol. 38, No. 5, pp. A3163–A3194

OPTIMAL MODEL MANAGEMENT FOR MULTIFIDELITY MONTE CARLO ESTIMATION*

BENJAMIN PEHERSTORFER†, KAREN WILLCOX†, AND MAX GUNZBURGER‡

Abstract. This work presents an optimal model management strategy that exploits multifidelity surrogate models to accelerate the estimation of statistics of outputs of computationally expensive high-fidelity models. Existing acceleration methods typically exploit a multilevel hierarchy of surrogate models that follow a known rate of error decay and computational costs; however, a general collection of surrogate models, which may include projection-based reduced models, data-fit models, support vector machines, and simplified-physics models, does not necessarily give rise to such a hierarchy. Our multifidelity approach provides a framework to combine an arbitrary number of surrogate models of any type. Instead of relying on error and cost rates, an optimization problem balances the number of model evaluations across the high-fidelity and surrogate models with respect to error and costs. We show that a unique analytic solution of the model management optimization problem exists under mild conditions on the models. Our multifidelity method makes occasional recourse to the high-fidelity model; in doing so it provides an unbiased estimator of the statistics of the high-fidelity model, even in the absence of error bounds and error estimators for the surrogate models. Numerical experiments with linear and nonlinear examples show that speedups by orders of magnitude are obtained compared to Monte Carlo estimation that invokes a single model only.

Key words. multifidelity, surrogate modeling, model reduction, Monte Carlo simulation

AMS subject classifications. 65M22, 65N22

DOI. 10.1137/15M1046472

1. Introduction.

Multilevel techniques have a long and successful history in computational science and engineering, e.g., multigrid for solving systems of equations [8, 25, 9], multilevel discretizations for representing functions [50, 18, 10], and multilevel Monte Carlo and multilevel stochastic collocation for estimating mean solutions of partial differential equations (PDEs) with stochastic parameters [27, 22, 45]. These multilevel techniques typically start with a fine-grid discretization—a high-fidelity model—of the underlying PDE or function. The fine-grid discretization is chosen to guarantee an approximation of the output of interest with the accuracy required by the current problem at hand. Additionally, a hierarchy of coarser discretizations—lower-fidelity surrogate models—is constructed, where a parameter (e.g., mesh width) controls the trade-off between error and computational costs. Changing this parameter gives rise to a multilevel hierarchy of discretizations with known error and cost rates. Multilevel techniques use these error and cost rates to distribute the computational work among the discretizations in the hierarchy, shifting most of the work onto the

*Submitted to the journal's Methods and Algorithms for Scientific Computing section November 2, 2015; accepted for publication (in revised form) July 29, 2016; published electronically October 4, 2016.

http://www.siam.org/journals/sisc/38-5/M104647.html

Funding: The first two authors were supported in part by the AFOSR MURI on multi-information sources of multi-physics systems under award FA9550-15-1-0038, program manager Jean-Luc Cambier, and by the U.S. Department of Energy Applied Mathematics Program, awards DE-FG02-08ER2585 and DE-SC0009297, as part of the DiaMonD Multifaceted Mathematics Integrated Capability Center. The third author was supported by U.S. Department of Energy Office of Science grant DE-SC0009324.

†Department of Aeronautics & Astronautics, MIT, Cambridge, MA 02139 (pehersto@mit.edu, kwillcox@mit.edu).

‡Department of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120 (gunzburg@fsu.edu).

cheap lower-fidelity surrogate models but correcting with a few expensive high-fidelity model outputs to establish accuracy guarantees on the overall result. However, in many situations, we are confronted with richer, more heterogeneous sets of models than just hierarchies of fine- and coarse-grid discretizations. For example, available surrogate models may include projection-based reduced models [44, 43, 24, 3], data-fit interpolation and regression models [21], machine-learning-based support vector machines (SVMs) [16, 48, 11], and simplified-physics models [2, 37]. Distributing work among general surrogate models (i.e., deciding which models to use and when) is challenging because collections of general surrogate models typically do not give rise to a multilevel hierarchy with known error and cost rates.

We present a multifidelity framework to exploit lower-fidelity surrogate models of any type for the acceleration of the Monte Carlo estimation of statistics of outputs of large-scale high-fidelity models. The key ingredient of our multifidelity Monte Carlo (MFMC) method is an optimization problem to distribute the computational work among the models such that the mean squared error (MSE) of the multifidelity estimator is minimized for a given computational budget. Thus, our multifidelity method distributes work using an optimization problem rather than error and cost rates, because such rates are unavailable for general collections of models. We show that the optimization problem has a unique and analytic solution under mild conditions on the models. These conditions, for example, prevent the use of models that are both inaccurate and costly to evaluate. Multilevel techniques guarantee similar conditions on the models by construction, because the model hierarchy is generated by changing the discretization parameter. In contrast, our multifidelity approach targets optimal exploitation of a given set of models and therefore requires explicit conditions that are revealed by analysis of our optimization problem formulation. We prove that there always exists a subset of the given set of models that satisfies these conditions. We develop our methodology in the context of computational models, but our optimization-based multifidelity approach is applicable to information sources beyond models, e.g., experiments, expert opinions, and lookup tables. If each information source has associated a certain fidelity and cost, the solution of our optimization problem determines the optimal number of queries of each information source such that the multifidelity estimator with minimal MSE is obtained.

Several approaches in the literature combine a high-fidelity model with general surrogate models to accelerate computations. Multifidelity optimization uses surrogate models to accelerate convergence to an optimal solution [4, 1, 32, 21]. In statistical inference, two-stage Markov chain Monte Carlo techniques rely on surrogate models to decide whether a proposed sample is further processed by the high-fidelity model, discarded, or used to adapt the surrogate model [14, 19, 17]. We focus here on the estimation of statistics of outputs of models, a setting for which multifidelity methods are often based on control variates. The control variates method provides a framework to derive an estimator with a lower variance than the Monte Carlo estimator by combining samples drawn from the output random variable of interest with samples drawn from a correlated auxiliary random variable [34]. The multilevel Monte Carlo method [27, 22, 15, 46] uses coarse-grid approximations as surrogate models and distributes the work among the models using known error and cost rates. In [6, 5], the control variate method is used to combine the high-fidelity model with a reduced basis model. The focus is on the construction of the reduced basis model with a problem-specific greedy method that tailors the reduced model toward variance reduction. In [36, 37], general surrogate models, including reduced basis models, and other sources of approximate information (e.g., past model evaluations for different

parameters) are exploited in a control variate framework to reduce the computational costs of optimization under uncertainty. The approach in [49] applies the multilevel Monte Carlo method to reduced basis models and approximately solves an optimization problem to derive the number of model evaluations, which requires a posteriori error estimators of the reduced models. StackedMC [47] uses the control variate method and data-fit surrogate models. The surrogate models are constructed with supervised learning techniques. Another body of work combines the high-fidelity model with a surrogate model in the context of the Monte Carlo method with importance sampling [31, 30, 13, 41, 38].

Our MFMC method is based on the multifidelity approach introduced in [36, 37], which considers the use of general surrogate models to accelerate Monte Carlo sampling. That work derives a multifidelity estimator using the control variate method and balances model evaluations between the high-fidelity model and a single surrogate model such that the MSE of the estimator is minimized for a given computational budget. Our MFMC method follows a similar approach but we formulate an optimization problem that explicitly allows an arbitrary number of surrogate models and surrogate models of any type. Our mathematical analysis of the solution of the optimization problem confirms that combining surrogate models of different type, varying approximation quality, and varying costs is often more beneficial than combining accurate surrogate models only. In this sense, surrogate models that inform different aspects of the high-fidelity model are better than surrogate models that are accurate but lack a rich diversity. Our analysis shows that the variance reduction obtained by including an additional surrogate model in the MFMC estimator depends on the new information introduced by the surrogate model compared to the other models in the MFMC estimator, rather than solely on the approximation quality and costs of the surrogate model itself.

Unbiasedness of our MFMC estimator is established by occasional recourse to the high-fidelity model. Thus, our MFMC method does not rely on error bounds of the surrogate models and provides an unbiased estimator independent of the approximation quality of the surrogate models. Our MFMC method is applicable to black-box high-fidelity and surrogate models, i.e., models for which we can evaluate a specified input to obtain the output but for which we do not have access to the model operators (in an assembled form or through their actions on a vector).

This paper is organized as follows. Section 2 describes the problem setup. Section 3 defines the MFMC estimator, derives the optimization problem to balance the number of model evaluations, and provides an interpretation and discussion. Section 4 demonstrates the MFMC estimator on a model that describes the bending of a locally damaged plate and on a model of a tubular reactor that exhibits an oscillatory regime. Runtime savings of up to four orders of magnitude are achieved. Section 5 draws conclusions.

2. Problem setup.

Section 2.1 introduces the high-fidelity model and surrogate models and discusses the Monte Carlo method. Section 2.2 formulates the problem of interest.

2.1. Models.

Let $d \in \mathbb{N}$ and define the input domain $\mathcal{D} \subset \mathbb{R}^d$ and the output domain $\mathcal{Y} \subset \mathbb{R}$. An information source is a function $f : \mathcal{D} \to \mathcal{Y}$ that maps an input $\boldsymbol{z} \in \mathcal{D}$ to an output $y \in \mathcal{Y}$. In this work, all our information sources are computational models that are evaluated at an input $\boldsymbol{z} \in \mathcal{D}$ to obtain an output $y \in \mathcal{Y}$; however, our methodology extends to other information sources such as experiments, expert opinions, and lookup tables, provided these information sources can be evaluated for

any specified input realization.

In the following, we have a high-fidelity model denoted as $f^{(1)} : \mathcal{D} \to \mathcal{Y}$ and (lower-fidelity) surrogate models $f^{(2)}, \ldots, f^{(k)} : \mathcal{D} \to \mathcal{Y}$ with $k \in \mathbb{N}$. We consider the high-fidelity model $f^{(1)}$ as our "truth" model. Note that we use the same input domain for all models $f^{(1)}, \ldots, f^{(k)}$. Note further that we consider scalar-valued models $f^{(1)}, \ldots, f^{(k)}$ only.

The costs of evaluating a model $f^{(i)}$ are $w_i \in \mathbb{R}_+$ for $i = 1, \ldots, k$, where $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$. The costs vector is $\boldsymbol{w} = [w_1, \ldots, w_k]^T \in \mathbb{R}_+^k$, with the set $\mathbb{R}_+^k$ containing k-dimensional vectors with components in $\mathbb{R}_+$. There are no assumptions on the surrogate models. In particular, we explicitly avoid assumptions on the pointwise errors

$$\left| f^{(1)}(\boldsymbol{z}) - f^{(i)}(\boldsymbol{z}) \right|, \qquad \boldsymbol{z} \in \mathcal{D}, i = 2, \ldots, k, \tag{2.1}$$

with respect to the high-fidelity model $f^{(1)}$. Bounds for (2.1) are unnecessary for our methodology, and therefore our methodology is developed independently of the availability of such accuracy guarantees on the surrogate models.

Let Ω be a sample space and $Z : \Omega \to \mathcal{D}$ a random variable with range $\mathcal{D}$. Independent and identically distributed (i.i.d.) realizations of Z are denoted as $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m \in \mathcal{D}$, where $m \in \mathbb{N}$. The variance $\mathrm{Var}[f^{(i)}(Z)]$ of $f^{(i)}(Z)$ is denoted

$$\sigma_i^2 = \mathrm{Var}[f^{(i)}(Z)] \tag{2.2}$$

for $i = 1, \ldots, k$, and the Pearson product-moment correlation coefficient is

$$\rho_{i,j} = \frac{\mathrm{Cov}[f^{(i)}(Z), f^{(j)}(Z)]}{\sigma_i \sigma_j} \tag{2.3}$$

for $i = 1, \ldots, k$ and $j = 1, \ldots, k$. In the following, we ignore models that are uncorrelated to the high-fidelity model, and therefore we have $\rho_{1,i}^2 > 0$ for $i = 1, \ldots, k$. We define $\rho_{i,k+1} = 0$ for $i = 1, \ldots, k$.

The Monte Carlo method draws m i.i.d. realizations $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_m \in \mathcal{D}$ of Z and estimates $\mathbb{E}[f^{(i)}(Z)]$ by

$$\overline{y}_m^{(i)} = \frac{1}{m} \sum_{j=1}^{m} f^{(i)}(\boldsymbol{z}_j) \tag{2.4}$$

for $i = 1, \ldots, k$. The Monte Carlo estimator $\overline{y}_m^{(i)}$ is an unbiased estimator of $\mathbb{E}[f^{(i)}(Z)]$ [42]. If the variance $\sigma_i^2 \in \mathbb{R}$ (i.e., if the variance is finite), then the MSE of the estimator $\overline{y}_m^{(i)}$ with respect to $\mathbb{E}[f^{(i)}(Z)]$ is

$$e(\overline{y}_m^{(i)}) = \mathbb{E}\left[\left(\mathbb{E}[f^{(i)}(Z)] - \overline{y}_m^{(i)} \right)^2 \right] = \frac{\mathrm{Var}[f^{(i)}(Z)]}{m}.$$

The cost of computing the Monte Carlo estimator is

$$c(\overline{y}_m^{(i)}) = w_i m,$$

because the model $f^{(i)}$ is evaluated at m inputs, each with evaluation cost w_i.

2.2. Problem formulation. Our goal is to estimate the expectation

$$s = \mathbb{E}[f^{(1)}(Z)] \tag{2.5}$$

of the high-fidelity model $f^{(1)}$ with realizations of the random variable Z as inputs. We seek an estimator with a computational budget $p \in \mathbb{R}_+$ that optimally exploits the surrogate models $f^{(2)}, \ldots, f^{(k)}$ to achieve a lower MSE than the Monte Carlo estimator with the same computational budget p. Stated differently, we seek a multifidelity estimator that achieves the same MSE as the Monte Carlo estimator but with a lower computational cost. We also seek an estimator that is unbiased with respect to expectation (2.5), even in the absence of accuracy guarantees such as (2.1) on the surrogate models.

3. Multifidelity Monte Carlo. Our MFMC method derives auxiliary random variables from surrogate models and combines them into the unbiased MFMC estimator using the control variate method. An optimization problem distributes the number of model evaluations among the high-fidelity and surrogate models. The optimization problem balances correlation strength and relative computational costs such that the MSE of the estimator is minimized for a given computational budget. We prove that the MFMC estimator is unbiased and derive the condition under which the MFMC estimator has a lower MSE than the Monte Carlo estimator with the same computational budget. Section 3.1 formulates the MFMC estimator and shows that it is unbiased. Sections 3.2 to 3.4 derive the optimization problem to balance the number of model evaluations across the high-fidelity and surrogate models and provide an interpretation and discussion. Sections 3.5 and 3.6 give practical considerations and summarize the MFMC method in Algorithm 2.

3.1. Multifidelity Monte Carlo estimator. Consider the k models $f^{(1)}, \ldots, f^{(k)}$. Let $\boldsymbol{m} = [m_1, \ldots, m_k]^T \in \mathbb{N}^k$ be a vector with integer components $0 < m_1 \leq \ldots \leq m_k$ and let

$$\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{m_k} \in \mathcal{D} \tag{3.1}$$

be m_k i.i.d. realizations of the random variable Z. For $i = 1, \ldots, k$, evaluate model $f^{(i)}$ at the m_i realizations $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{m_i}$ of (3.1) to obtain

$$f^{(i)}(\boldsymbol{z}_1), \ldots, f^{(i)}(\boldsymbol{z}_{m_i}).$$

The component m_i of $\boldsymbol{m}$ is the number of evaluations of model $f^{(i)}$ for $i = 1, \ldots, k$. Derive the Monte Carlo estimate $\overline{y}^{(i)}_{m_i}$ as in (2.4) from the m_i model evaluations $f^{(i)}(\boldsymbol{z}_1), \ldots, f^{(i)}(\boldsymbol{z}_{m_i})$ for $i = 1, \ldots, k$. Additionally, compute the Monte Carlo estimate $\overline{y}^{(i)}_{m_{i-1}}$ from the m_{i-1} model evaluations $f^{(i)}(\boldsymbol{z}_1), \ldots, f^{(i)}(\boldsymbol{z}_{m_{i-1}})$ for $i = 2, \ldots, k$. The Monte Carlo estimator $\overline{y}^{(i)}_{m_{i-1}}$ reuses the first m_{i-1} model evaluations that are used for $\overline{y}^{(i)}_{m_i}$, and therefore the two estimators are dependent. The MFMC estimate $\hat{s}$ of s is then

$$\hat{s} = \overline{y}^{(1)}_{m_1} + \sum_{i=2}^{k} \alpha_i \left(\overline{y}^{(i)}_{m_i} - \overline{y}^{(i)}_{m_{i-1}} \right), \tag{3.2}$$

where $\alpha_2, \ldots, \alpha_k \in \mathbb{R}$ are coefficients that weight the differences $\overline{y}^{(i)}_{m_i} - \overline{y}^{(i)}_{m_{i-1}}$ of the Monte Carlo estimates $\overline{y}^{(i)}_{m_i}$ and $\overline{y}^{(i)}_{m_{i-1}}$ for $i = 2, \ldots, k$. The structure of our MFMC

estimator (3.2) is similar to the structure of multilevel Monte Carlo estimators [15, 46]. Both correct an estimate of high-fidelity quantities with a sum of estimates of differences of lower-fidelity quantities. The distinguishing feature of our MFMC method is the optimal selection of the number of model evaluations $\boldsymbol{m}$ and of the coefficients $\alpha_2, \ldots, \alpha_k$ that is applicable to surrogate models of any type, as introduced in the subsequent sections.

The following lemma shows that the MFMC estimator is an unbiased estimator of s.

LEMMA 3.1. *The MFMC estimator $\hat{s}$ is an unbiased estimator of the expectation s of the high-fidelity model $f^{(1)}$.*

Proof. We have $\boldsymbol{m} \in \mathbb{N}^k$ and $m_1 > 0$ and therefore each model $f^{(1)}, \ldots, f^{(k)}$ is evaluated at least once. The unbiasedness $\mathbb{E}[\hat{s}] = \mathbb{E}[f^{(1)}(Z)]$ follows from the linearity of the expectation. ∎

Because the MFMC estimator $\hat{s}$ is unbiased, the MSE of $\hat{s}$ with respect to s is

$$e(\hat{s}) = \mathbb{E}[(s - \hat{s})^2] = \mathrm{Var}[\hat{s}]. \tag{3.3}$$

The costs of deriving an MFMC estimate $\hat{s}$ are

$$c(\hat{s}) = \sum_{i=1}^{k} w_i m_i = \boldsymbol{w}^T \boldsymbol{m},$$

because to compute for $i = 2, \ldots, k$

$$\overline{y}^{(i)}_{m_i} - \overline{y}^{(i)}_{m_{i-1}}$$

the model $f^{(i)}$ is evaluated at the samples $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{m_i}$, where the first m_{i-1} samples are reused to derive $\overline{y}^{(i)}_{m_{i-1}}$.

3.2. Optimal number of model evaluations. The MFMC estimator defined in (3.2) depends on the number of model evaluations $\boldsymbol{m} \in \mathbb{N}^k$ and on the coefficients $\alpha_2, \ldots, \alpha_k \in \mathbb{R}$. We formulate the selection of the number of model evaluations and of the coefficients as an optimization problem. We first show Lemmas 3.2 and 3.3 before we derive the optimization problem and present its solution in Theorem 3.4.

LEMMA 3.2. *Consider the Monte Carlo estimators $\overline{y}^{(l)}_{m_i}$ and $\overline{y}^{(t)}_{m_j}$ with $1 \leq i, j, l, t \leq k$. We find for the covariance*

$$\mathrm{Cov}\left[\overline{y}^{(l)}_{m_i}, \overline{y}^{(t)}_{m_j}\right] = \frac{1}{\max\{m_i, m_j\}} \rho_{l,t} \sigma_l \sigma_t. \tag{3.4}$$

Proof. We have

$$\mathrm{Cov}\left[\overline{y}^{(l)}_{m_i}, \overline{y}^{(t)}_{m_j}\right] = \frac{1}{m_i m_j} \sum_{i'=1}^{m_i} \sum_{j'=1}^{m_j} \mathrm{Cov}\left[f^{(l)}(Z_{i'}), f^{(t)}(Z_{j'})\right], \tag{3.5}$$

where $Z_1, \ldots, Z_{m_k}$ are random variables that are i.i.d. as the random variable Z. With the definition of the correlation coefficient (2.3), the independence of the random variables $Z_1, \ldots, Z_{m_k}$, and the symmetry of the covariance follows (3.4) from (3.5). ∎

LEMMA 3.3. *The variance* $\mathrm{Var}[\hat{s}]$ *of the MFMC estimator* $\hat{s}$ *is*

$$\mathrm{Var}[\hat{s}] = \frac{\sigma_1^2}{m_1} + \sum_{i=2}^{k} \left(\frac{1}{m_{i-1}} - \frac{1}{m_i} \right) \left(\alpha_i^2 \sigma_i^2 - 2\alpha_i \rho_{1,i} \sigma_1 \sigma_i \right). \tag{3.6}$$

Proof. The variance of a sum of random variables is the sum of their covariances

$$\begin{aligned}
\mathrm{Var}[\hat{s}] &= \mathrm{Var}\left[\overline{y}_{m_1}^{(1)} + \sum_{i=2}^{k} \alpha_i \left(\overline{y}_{m_i}^{(i)} - \overline{y}_{m_{i-1}}^{(i)} \right) \right] \\
&= \mathrm{Var}\left[\overline{y}_{m_1}^{(1)} \right] + \sum_{i=2}^{k} \alpha_i^2 \left(\mathrm{Var}\left[\overline{y}_{m_i}^{(i)} \right] + \mathrm{Var}\left[\overline{y}_{m_{i-1}}^{(i)} \right] \right) \\
&\quad + 2 \sum_{i=2}^{k} \alpha_i \left(\mathrm{Cov}\left[\overline{y}_{m_1}^{(1)}, \overline{y}_{m_i}^{(i)} \right] - \mathrm{Cov}\left[\overline{y}_{m_1}^{(1)}, \overline{y}_{m_{i-1}}^{(i)} \right] \right) \\
&\quad + 2 \sum_{i=2}^{k} \alpha_i \sum_{j=i+1}^{k} \alpha_j \left(\mathrm{Cov}\left[\overline{y}_{m_i}^{(i)}, \overline{y}_{m_j}^{(j)} \right] - \mathrm{Cov}\left[\overline{y}_{m_i}^{(i)}, \overline{y}_{m_{j-1}}^{(j)} \right] \right) \qquad (3.7) \\
&\quad - 2 \sum_{i=2}^{k} \alpha_i \sum_{j=i+1}^{k} \alpha_j \left(\mathrm{Cov}\left[\overline{y}_{m_{i-1}}^{(i)}, \overline{y}_{m_j}^{(j)} \right] - \mathrm{Cov}\left[\overline{y}_{m_{i-1}}^{(i)}, \overline{y}_{m_{j-1}}^{(j)} \right] \right) \qquad (3.8) \\
&\quad - 2 \sum_{i=2}^{k} \alpha_i^2 \, \mathrm{Cov}\left[\overline{y}_{m_i}^{(i)}, \overline{y}_{m_{i-1}}^{(i)} \right].
\end{aligned}$$

With Lemma 3.2 and $m_1 \leq \cdots \leq m_k$, it follows that the covariance terms in (3.7) and (3.8) cancel, which then shows the lemma. ∎

We now formulate the optimization problem to select the number of model evaluations $\boldsymbol{m}$ and the coefficients $\alpha_2, \ldots, \alpha_k$ that minimize the MSE $e(\hat{s})$ of the MFMC estimator $\hat{s}$ with costs $c(\hat{s}) = p$ equal to a computational budget $p \in \mathbb{R}_+$. Since $e(\hat{s}) = \mathrm{Var}[\hat{s}]$, it is sufficient to minimize the variance $\mathrm{Var}[\hat{s}]$ of the MFMC estimator; see section 3.1 and (3.3). We formulate the optimization problem with $\boldsymbol{m} \in \mathbb{R}^k$, rather than in $\mathbb{N}^k$, and round down $\lfloor m_i \rfloor$ to determine the integer number of model evaluations from the real number m_i for $i = 1, \ldots, k$. We do not expect that the rounding significantly impacts the MSE of the MFMC estimator. Usually, the number of model evaluations is $\gg 1$ and therefore changing the number of model evaluations by a fraction less than one is a small change relative to the number of model evaluations. Rounding down ensures that the costs of the MFMC estimator do not exceed the computational budget p. Let therefore $J : \mathbb{R}_+^k \times \mathbb{R}^{k-1} \to \mathbb{R}$ be the objective function with

$$J(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k) = \frac{\sigma_1^2}{m_1} + \sum_{i=2}^{k} \left(\frac{1}{m_{i-1}} - \frac{1}{m_i} \right) \left(\alpha_i^2 \sigma_i^2 - 2\alpha_i \rho_{1,i} \sigma_1 \sigma_i \right), \tag{3.9}$$

and note that $J(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k) = \mathrm{Var}[\hat{s}]$ for $\boldsymbol{m} \in \mathbb{N}^k$ and $0 < m_1 \leq m_2 \leq \cdots \leq m_k$. We define the vector $\boldsymbol{m}^* \in \mathbb{R}^k$ and the coefficients $\alpha_2^*, \ldots, \alpha_k^* \in \mathbb{R}^k$ to be the solution

of the optimization problem

$$\begin{aligned} \underset{\boldsymbol{m}\in\mathbb{R}^k,\alpha_2,\ldots,\alpha_k\in\mathbb{R}}{\arg\min} \quad & J(\boldsymbol{m},\alpha_2,\ldots,\alpha_k) \\ \text{subject to} \quad & m_{i-1}-m_i\leq 0, \quad i=2,\ldots,k, \\ & -m_1\leq 0, \\ & \boldsymbol{w}^T\boldsymbol{m}=p. \end{aligned} \tag{3.10}$$

The inequality constraints $m_{i-1}-m_i\leq 0$ for $i=2,\ldots,k$ and $-m_1\leq 0$ ensure $0\leq m_1\leq\cdots\leq m_k$. Note that we will show $m_1^*>0$ for the optimal m_1^* and therefore $\boldsymbol{m}^*\in\mathbb{R}_+^k$; see Lemma A.1. The equality constraint $\boldsymbol{w}^T\boldsymbol{m}=p$ ensures that the costs of the corresponding MFMC estimator equal the computational budget $p\in\mathbb{R}_+$. Theorem 3.4 provides the global solution of the optimization problem (3.10) under a condition on the costs and the correlation coefficients. Note that in section 3.5 we present an algorithm to select from models $f^{(1)},\ldots,f^{(k)}$ a subset of models that satisfies the conditions required by Theorem 3.4.

THEOREM 3.4. *Let $f^{(1)},\ldots,f^{(k)}$ be k models with ordering*

$$|\rho_{1,1}|>\cdots>|\rho_{1,k}| \tag{3.11}$$

and with costs $w_1,\ldots,w_k$ that satisfy the ratios

$$\frac{w_{i-1}}{w_i}>\frac{\rho_{1,i-1}^2-\rho_{1,i}^2}{\rho_{1,i}^2-\rho_{1,i+1}^2} \tag{3.12}$$

for $i=2,\ldots,k$. Set the coefficients $\alpha_2^*,\ldots,\alpha_k^*$ to

$$\alpha_i^*=\frac{\rho_{1,i}\sigma_1}{\sigma_i} \tag{3.13}$$

for $i=2,\ldots,k$ and the components of $\boldsymbol{r}^*=[r_1^*,\ldots,r_k^*]^T\in\mathbb{R}_+^k$ to

$$r_i^*=\sqrt{\frac{w_1(\rho_{1,i}^2-\rho_{1,i+1}^2)}{w_i(1-\rho_{1,2}^2)}} \tag{3.14}$$

for $i=1,\ldots,k$. Set further the components of $\boldsymbol{m}^*=[m_1^*,\ldots,m_k^*]^T\in\mathbb{R}_+^k$ to $m_i^*=m_1^*r_i^*$ for $i=2,\ldots,k$ and

$$m_1^*=\frac{p}{\boldsymbol{w}^T\boldsymbol{r}^*}, \tag{3.15}$$

where $p\in\mathbb{R}_+$ is the computational budget. The global minimum of (3.10) *is $(\boldsymbol{m}^*,\alpha_2^*,\ldots,\alpha_k^*)$.*

Proof. First note that $(\boldsymbol{m}^*,\alpha_2^*,\ldots,\alpha_k^*)$ satisfies the constraints of the optimization problem (3.10) because of the ordering (3.11) and condition (3.12). Consider a local minimum $(\boldsymbol{m},\alpha_2,\ldots,\alpha_k)$ of (3.10) for which an $i\in\{2,\ldots,k\}$ exists with $m_{i-1}=m_i$. Define $\boldsymbol{l}=[l_1,\ldots,l_q]^T\in\mathbb{N}^q$ with $q\in\mathbb{N}$ such that

$$m_{l_i-1}<m_{l_i}, \qquad i=1,\ldots,q,$$

and

$$m_{l_i}=m_{l_i+1}=\cdots=m_{l_{i+1}-1}, \qquad i=0,\ldots,q,$$

where $l_0 = 1$ and $l_{q+1} = k+1$. Lemma A.1 in Appendix A shows that we have

$$r_{l_i} = \frac{m_{l_i}}{m_1} = \sqrt{\frac{\left(\sum_{j=1}^{l_1-1} w_j\right)}{\left(\sum_{j=l_i}^{l_{i+1}-1} w_j\right)} \frac{(\rho_{1,l_i}^2 - \rho_{1,l_{i+1}}^2)}{(1-\rho_{1,l_1}^2)}} \tag{3.16}$$

for $i = 0, \ldots, q$. Lemma A.3 in Appendix A shows that condition (3.12) on the ratios of the costs and the differences of the squared correlation coefficients leads to

$$\sum_{i=1}^{k} \sqrt{w_i(\rho_{1,i}^2 - \rho_{1,i+1}^2)} < \sum_{i=0}^{q} \sqrt{\left(\sum_{j=l_i}^{l_{i+1}-1} w_j\right)(\rho_{1,l_i}^2 - \rho_{1,l_{i+1}}^2)},$$

from which we obtain

$$\begin{aligned} &\sqrt{\frac{1-\rho_{1,2}^2}{w_1}} \sum_{i=1}^{k} w_i \underbrace{\sqrt{\frac{w_1}{w_i} \frac{(\rho_{1,i}^2 - \rho_{1,i+1}^2)}{(1-\rho_{1,2}^2)}}}_{r_i^*} \\ &\quad < \sqrt{\frac{1-\rho_{1,l_1}^2}{\sum_{t=1}^{l_1-1} w_t}} \sum_{i=0}^{q} \left(\sum_{j=l_i}^{l_{i+1}-1} w_j\right) \underbrace{\sqrt{\frac{\left(\sum_{t=1}^{l_1-1} w_t\right)}{\left(\sum_{t=l_i}^{l_{i+1}-1} w_t\right)} \frac{(\rho_{1,l_i}^2 - \rho_{1,l_{i+1}}^2)}{(1-\rho_{1,l_1}^2)}}}_{r_{l_i}}, \end{aligned} \tag{3.17}$$

where we used $r_1^*, \ldots, r_k^*$ as defined in (3.14) and $r_{l_0}, \ldots, r_{l_q}$ as in (3.16). Lemma A.1 shows $m_1^* = p/(\boldsymbol{w}^T \boldsymbol{r}^*)$ and $m_1 = p/(\boldsymbol{w}^T \boldsymbol{r})$ and therefore (3.17) is equivalent to

$$\sqrt{\frac{(1-\rho_{1,2}^2)p^2}{(m_1^*)^2 w_1}} < \sqrt{\frac{(1-\rho_{1,l_1}^2)p^2}{m_1^2 \sum_{i=1}^{l_1-1} w_i}}.$$

Squaring both sides and dividing by p leads to

$$\frac{\sigma_1^2(1-\rho_{1,2}^2)p}{(m_1^*)^2 w_1} < \frac{\sigma_1^2(1-\rho_{1,l_1}^2)p}{m_1^2 \sum_{i=1}^{l_1-1} w_i}. \tag{3.18}$$

Lemma A.2 in Appendix A shows that the left-hand side of (3.18) is the objective function J evaluated at $\boldsymbol{m}^*$ and $\alpha_2^*, \ldots, \alpha_k^*$ and the right-hand side J evaluated at $\boldsymbol{m}$ and $\alpha_2, \ldots, \alpha_k$. The inequality (3.18) therefore shows that the value of the objective function at $\boldsymbol{m}^*$ and $\alpha_2^*, \ldots, \alpha_k^*$ is smaller than at a local minimum with $\boldsymbol{m}$ and $\alpha_2, \ldots, \alpha_k$ where there exists an $i \in \{2, \ldots, k\}$ with $m_{i-1} = m_i$.

Lemma A.1 shows that only $(\boldsymbol{m}^*, \alpha_2^*, \ldots, \alpha_k^*)$ can be a local minimum with $m_1^* < m_2^* < \cdots < m_k^*$ (strict inequalities) and therefore $(\boldsymbol{m}^*, \alpha_2^*, \ldots, \alpha_k^*)$ is the unique global minimum of (3.10) given ordering (3.11) and condition (3.12). ∎

We round down the components of $\boldsymbol{m}^* = [m_1^*, \ldots, m_k^*]^T \in \mathbb{R}_+^k$ to obtain an integer number of model evaluations and denote the MFMC estimator with $\lfloor \boldsymbol{m}^* \rfloor = [\lfloor m_1^* \rfloor, \ldots, \lfloor m_k^* \rfloor]^T \in \mathbb{N}^k$ and the coefficients $\alpha_2^*, \ldots, \alpha_k^* \in \mathbb{R}$ as $\hat{s}^*$; see above for a discussion on the rounding. In the problem setup, the budget p should be large enough to evaluate the high-fidelity model at least once; otherwise the multifidelity estimator is biased. For the ease of exposition, we treat the components of $\boldsymbol{m}^*$ as integers in the following, because the rounding introduces only constant factors in the discussion.

3.3. Discussion. Let $\hat{s}^*$ be the MFMC estimator with computational budget $p \in \mathbb{R}_+$. The costs $c(\overline{y}_n^{(1)}) = p$ of the Monte Carlo estimator $\overline{y}_n^{(1)}$ with $n = p/w_1$ model evaluations are equal to the costs of the MFMC estimator $\hat{s}^*$ with computational budget p. Corollary 3.5 derives the condition under which it is computationally cheaper to estimate s with the MFMC estimator $\hat{s}^*$ than with the Monte Carlo estimator $\overline{y}_n^{(1)}$.

COROLLARY 3.5. *The MSE $e(\hat{s}^*)$ of the MFMC estimator $\hat{s}^*$ for a given computational budget $p \in \mathbb{R}_+$ is*

$$e(\hat{s}^*) = \frac{\sigma_1^2(1-\rho_{1,2}^2)}{(m_1^*)^2 w_1} p. \tag{3.19}$$

The MSE $e(\hat{s}^*)$ of the MFMC estimator $\hat{s}^*$ is smaller than the MSE of the Monte Carlo estimator $\overline{y}_n^{(1)}$ with $n = p/w_1$ evaluations of $f^{(1)}$, and thus costs $c(\overline{y}_n^{(1)}) = p$, if and only if

$$\sum_{i=1}^{k} \sqrt{\frac{w_i}{w_1}(\rho_{1,i}^2 - \rho_{1,i+1}^2)} < 1. \tag{3.20}$$

Proof. Lemma A.2 shows that the MSE $e(\hat{s}^*)$ of the MFMC estimator $\hat{s}^*$ is (3.19), because the objective J equals $\mathrm{Var}[\hat{s}^*]$ if the components of $\boldsymbol{m}^*$ are integers. The MSE of the Monte Carlo estimator $\overline{y}_n^{(1)}$ is

$$e(\overline{y}_n^{(1)}) = \frac{\sigma_1^2}{n} = \frac{\sigma_1^2}{p} w_1. \tag{3.21}$$

To show the condition (3.20), we rewrite the MSE (3.19) of the MFMC estimator as

$$e(\hat{s}^*) = \frac{\sigma_1^2(1-\rho_{1,2}^2)}{p w_1} \left(\sum_{i=1}^{k} w_i r_i^* \right)^2,$$

where we used $m_1^* = p/(\boldsymbol{w}^T \boldsymbol{r})$; see Theorem 3.4. We obtain with the definition of $\boldsymbol{r}^*$ in Theorem 3.4,

$$e(\hat{s}^*) = \frac{\sigma_1^2}{p} \left(\sum_{i=1}^{k} \sqrt{w_i(\rho_{1,i}^2 - \rho_{1,i+1}^2)} \right)^2. \tag{3.22}$$

We therefore derive from the MSE (3.21) of the Monte Carlo estimator and the MSE (3.22) of the MFMC estimator that $e(\hat{s}^*) < e(\overline{y}_n^{(1)})$ if and only if

$$w_1 > \left(\sum_{i=1}^{k} \sqrt{w_i(\rho_{1,i}^2 - \rho_{1,i+1}^2)} \right)^2,$$

which is equivalent to (3.20). ∎

Note that only the number of model evaluations m_1^* of the high-fidelity model $f^{(1)}$ appears in the definition of the MSE $e(\hat{s}^*)$ but that m_1^* depends through $\boldsymbol{r}^*$ on $m_2^*, \ldots, m_k^*$; see the definition of m_1^* in (3.15).

Consider the condition (3.20) that is sufficient and necessary for the MFMC estimator to achieve a variance reduction compared to the Monte Carlo estimator with the same computational budget. With the MSE $e(\overline{y}_n^{(1)})$ of the Monte Carlo estimator as derived in (3.21) and the MSE $e(\hat{s}^*)$ of the MFMC estimator as derived in (3.22) we obtain the ratio

$$\gamma \equiv \frac{e(\hat{s}^*)}{e\left(\overline{y}_n^{(1)}\right)} = \left(\sum_{i=1}^{k} \sqrt{\frac{w_i}{w_1}(\rho_{1,i}^2 - \rho_{1,i+1}^2)}\right)^2 . \tag{3.23}$$

The ratio (3.23) quantifies the variance reduction achieved by the MFMC estimator. The ratio is in inverse proportion to the variance reduction and therefore the variance of the MFMC estimator is small if the costs $w_1, \dots, w_k$ and the correlation coefficients $\rho_{1,1}, \dots, \rho_{1,k}$ of the models $f^{(1)}, \dots, f^{(k)}$ lead to small terms in the sum in (3.23). Consider a single term $i \in \{1, \dots, k\}$ in the sum in (3.23),

$$\sqrt{\frac{w_i}{w_1}(\rho_{1,i}^2 - \rho_{1,i+1}^2)}.$$

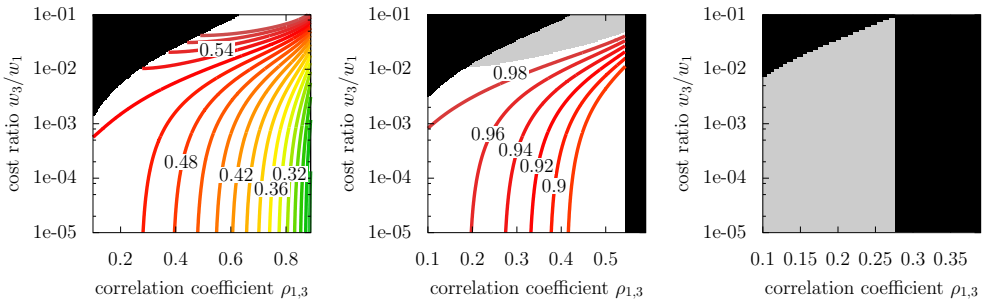
The term shows that the contribution of the model $f^{(i)}$ to the variance reduction is high if the costs w_i are low and the difference of the squared correlation coefficients $\rho_{1,i}^2 - \rho_{1,i+1}^2$ is low. Thus, the contribution of a model is high if the squared correlation coefficient $\rho_{1,i}^2$ of model $f^{(i)}$ is similar to the squared correlation coefficient $\rho_{1,i+1}^2$ of the subsequent model $f^{(i+1)}$. This shows that the contribution of a model cannot be determined by only considering the properties of the model itself but requires taking the properties of other models used in the MFMC estimator into account. Furthermore, the condition (3.20) that the MFMC estimator is computationally cheaper than the Monte Carlo estimator is not a condition on the properties of each model separately but rather is a condition on the properties of all models $f^{(1)}, \dots, f^{(k)}$ together, i.e., on the collective whole of the models.

Corollary 3.5 and the discussion on the variance reduction and the ratio (3.23) of the MSEs show that the correlation $\rho_{1,i}$ between a random variable $f^{(i)}(Z)$ induced by the surrogate model $f^{(i)}$ for $i = 2, \dots, k$ and the random variable $f^{(1)}(Z)$ corresponding to the high-fidelity model replaces the classical, deterministic pointwise error (2.1) that is usually considered to quantify the approximation quality of surrogate models in deterministic settings [43, 23, 24].

3.4. Illustrative examples. Let us consider an example with $k = 3$ models $f^{(1)}, f^{(2)}, f^{(3)}$ to demonstrate the behavior of the MFMC estimator and the interactions among the models. We consider the variance reduction ratio (3.23) as a function of the correlation coefficients $\rho_{1,2}, \rho_{1,3}$ and the costs w_1, w_2, w_3,

$$\gamma(\rho_{1,2}, \rho_{1,3}, w_1, w_2, w_3) = \left(\sqrt{1 - \rho_{1,2}^2} + \sqrt{\frac{w_2}{w_1}(\rho_{1,2}^2 - \rho_{1,3}^2)} + \sqrt{\frac{w_3}{w_1}\rho_{1,3}^2}\right)^2 . \tag{3.24}$$

Note that a lower value $\gamma(\rho_{1,2}, \rho_{1,3}, w_1, w_2, w_3)$ means a higher variance reduction. For illustration, we set $w_1/w_2 = 0.1$ and vary the correlation coefficients $\rho_{1,2}, \rho_{1,3}$ and the costs ratio w_3/w_1. Figure 1(a) shows the contour plot of γ for $\rho_{1,2} = 0.9$, $\rho_{1,3} \in [0.1, 0.9]$, and costs $w_3/w_1 \in [10^{-5}, 10^{-1}]$. The black region indicates where condition (3.12) is violated. The contours show that a model $f^{(3)}$ with a high correlation coefficient $\rho_{1,3}$ leads to a low value of γ and therefore to a high variance



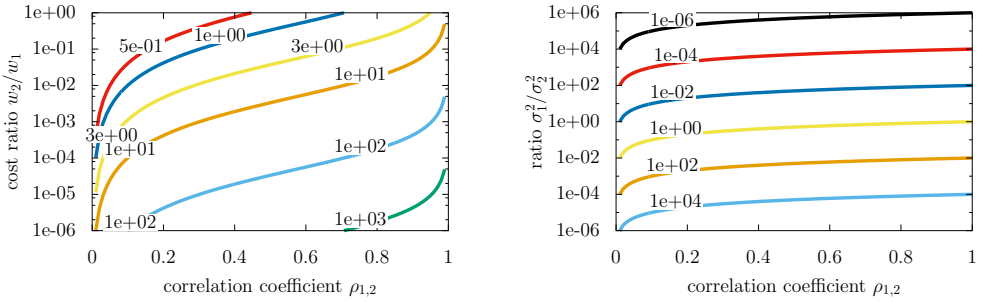
(a) $\rho_{1,2} = 0.9$ (b) $\rho_{1,2} = 0.6$ (c) $\rho_{1,2} = 0.4$

FIG. 1. *The plots show the contours of the variance reduction ratio* (3.23) *of an MFMC estimator with three models* $f^{(1)}, f^{(2)}, f^{(3)}$. *The black region shows where condition* (3.12) *is violated and the gray region where the MFMC estimator leads to a higher variance than the Monte Carlo estimator. The plots confirm that the variance reduction of the MFMC estimator is determined by the properties of the collective whole of the models and not by the properties of each model separately.*

reduction; however, the contours become almost vertical for low cost ratios w_3/w_1, which means that further decreasing the costs w_3 of model $f^{(3)}$ hardly improves the variance reduction if w_3 is already low. This can also be seen in (3.24), where changing the costs w_3 affects the third term only. Thus, if w_3/w_1 is already low, the second term in γ dominates the variance reduction, which is independent of the costs w_3. In Figure 1(b), the correlation coefficient of model $f^{(2)}$ is reduced to $\rho_{1,2} = 0.6$. The gray region shows where γ evaluates to values larger than 1, and therefore where the MFMC estimator has a higher variance than the standard Monte Carlo estimator. Figure 1(b) shows that a high correlation coefficient $\rho_{1,3}$ can violate condition (3.12). Consider now Figure 1(c), where we further reduce the correlation coefficient of the second model $f^{(2)}$ to $\rho_{1,2} = 0.4$. The plot shows that combining $f^{(1)}, f^{(2)}$, and a third model $f^{(3)}$ with $w_3/w_1 \in [10^{-5}, 10^{-1}]$ and $\rho_{1,3} \in [0.1, 0.4]$ into an MFMC estimator either violates condition (3.12) (black region) or leads to a higher variance than the Monte Carlo estimator (gray region). In this situation, it is therefore necessary to change or remove $f^{(2)}$ to obtain an MFMC estimator with a lower variance than the Monte Carlo estimator; see the following section for an approach to handle such a situation.

Let us now consider an example with two models $f^{(1)}, f^{(2)}$ to discuss the sensitivity of the number of model evaluations $\boldsymbol{m}^*$ and the coefficient α_2^* with respect to the correlation coefficients and the costs. Consider Figure 2(a) that shows the contours of the ratio m_2^*/m_1^* as a function of the cost ratio w_2/w_1 and the correlation coefficient $\rho_{1,2}$. The contours indicate that the ratio m_2^*/m_1^* varies smoothly with the correlation coefficient $\rho_{1,2}$ and the cost ratio w_2/w_1. In particular, Figure 2(a) suggests that small perturbations in the correlation coefficient $\rho_{1,2}$ and the costs w_1, w_2 lead to small changes in the number of model evaluations $\boldsymbol{m}^* = [m_1^*, m_2^*]^T$. Similarly, the coefficient α_2^* is in this example insensitive to perturbations in the ratio σ_1/σ_2 and the correlation coefficient $\rho_{1,2}$; see Figure 2(b). We refer to the numerical results in section 4.1 that confirm this behavior.

3.5. Model selection. Theorem 3.4 is applicable to models that are ordered descending with respect to the squared correlation coefficients (3.11) and that satisfy condition (3.12) on the ratios of the costs and correlation coefficients. Even if the



(a) contours of m_2^*/m_1^* (b) contours of α_2^*

FIG. 2. *The plots show that for an example with two models the number of model evaluations* $\boldsymbol{m}^* = [m_1^*, m_2^*]^T$ *and the coefficient* α_2^* *vary smoothly with respect to the costs* w_1, w_2*, the correlation coefficient* $\rho_{1,2}$*, and the variances* σ_1^2, σ_2^2*. In particular, small perturbations in the quantities* $w_1, w_2, \rho_{1,2}, \sigma_1^2, \sigma_2^2$ *lead to small changes in* $\boldsymbol{m}^*$ *and* α_2^**.*

given models $f^{(1)}, \ldots, f^{(k)}$ violate the ordering (3.11) or condition (3.12), there exists a selection of the models $f^{(1)}, \ldots, f^{(k)}$ that can be ordered as in (3.11) and that satisfy condition (3.12), and thus for which an MFMC estimator can be derived. There may even be multiple feasible selections of models for which an MFMC estimator can be constructed, in which case the selection that leads to the MFMC estimator with the lowest variance (i.e., lowest MSE) is of interest.

Given k models $f^{(1)}, \ldots, f^{(k)}$, Algorithm 1 iterates over all subsets $\mathcal{M} \subseteq \{f^{(1)}, \ldots, f^{(k)}\}$ and selects the subset $\mathcal{M}^*$ that leads to the MFMC estimator with the lowest variance. Inputs to Algorithm 1 are the models $f^{(1)}, \ldots, f^{(k)}$, the variance σ_1^2 of the high-fidelity model $f^{(1)}$, the correlation coefficients $\rho_{1,1}, \ldots, \rho_{1,k}$, and the costs $w_1, \ldots, w_k$. Algorithm 1 first orders the models with respect to the squared correlation coefficients. It then initializes the set $\mathcal{M}^* = \{f^{(1)}\}$ with the set $\{f^{(1)}\}$ that contains the high-fidelity model only. The variance of the estimator with computational budget $p = w_1$ that uses the high-fidelity model only (i.e., the Monte Carlo estimator) is $v^* = \sigma_1^2$; see the proof of Corollary 3.5 and (3.21). The algorithm then iterates over all subsets $\mathcal{M} \subseteq \{f^{(1)}, \ldots, f^{(k)}\}$ of which the high-fidelity model $f^{(1)}$ is an element and which satisfy condition (3.12). If the variance v of the MFMC estimator with budget p and with models in the current subset $\mathcal{M}$ is lower than the variance v^*, then the current subset $\mathcal{M}$ is stored in $\mathcal{M}^*$ and the variance $v^* = v$ is set to v. The set $\mathcal{M}^*$ is returned after iterating over all subsets. Note that $\mathcal{M}^*$ cannot be empty because the set $\mathcal{M}^*$ is initialized with $\{f^{(1)}\}$, which satisfies condition (3.12). Note further that a different choice of the computational budget p leads to the same set $\mathcal{M}^*$ because the variance of the MFMC estimator depends linearly on p.

Algorithm 1 does not treat the case where there exists $1 \leq i < j \leq k$ with $\rho_{1,i}^2 = \rho_{1,j}^2$. In such a situation, Algorithm 1 is run for the set of models $\{f^{(1)}, \ldots, f^{(k)}\} \setminus \{f^{(j)}\}$ to obtain $\mathcal{M}_1^*$ and v_1^* and for the set $\{f^{(1)}, \ldots, f^{(k)}\} \setminus \{f^{(i)}\}$ to obtain $\mathcal{M}_2^*$ and v_2^*. If $v_1^* < v_2^*$ then $\mathcal{M}^* = \mathcal{M}_1^*$ and else $\mathcal{M}^* = \mathcal{M}_2^*$.

The computational costs of Algorithm 1 grow exponentially in the number of models k and are bounded by $\mathcal{O}(2^k)$. Note that in many situations fewer than $k = 10$ models are available and therefore the exhaustive search performed by Algorithm 1 is usually computationally feasible and often negligible compared to the computational costs of evaluating models.

Algorithm 1. Model selection.

procedure MSELECT($f^{(1)}, \dots, f^{(k)}, \sigma_1, \rho_{1,1}, \dots, \rho_{1,k}, w_1, \dots, w_k$)
  Ensure $\rho_{1,1}^2 > \rho_{1,2}^2 > \cdots > \rho_{1,k}^2$, reorder if necessary
  Set the computational budget to $p = w_1$
  Initialize $\mathcal{M}^* = \{f^{(1)}\}$ and $v^* = \frac{\sigma_1^2}{p} w_1$
  for $\mathcal{M} \subseteq \{f^{(1)}, \dots, f^{(k)}\}$ **do**
    if $f^{(1)} \notin \mathcal{M}$ **then** continue
    end if
    Set $k' = |\mathcal{M}|$ to the number of elements in $\mathcal{M}$
    Let $i_1, \dots, i_{k'}$ be the indices of the models in $\mathcal{M}$ s.t. $\rho_{1,i_1}^2 > \cdots > \rho_{1,i_{k'}}^2$
    Set $\rho_{1,i_{k'+1}} = 0$
    if condition (3.12) is violated for models in $\mathcal{M}$ **then** continue
    end if
    Compute

$$v = \frac{\sigma_1^2}{p} \left(\sum_{j=1}^{k'} \sqrt{w_{i_j} (\rho_{1,i_j}^2 - \rho_{1,i_{j+1}}^2)} \right)^2$$

    if $v < v^*$ **then**
      $\mathcal{M}^* = \mathcal{M}$
      $v^* = v$
    end if
  end for
  return $\mathcal{M}^*, v^*$
end procedure

3.6. Practical considerations and computational procedure. The optimal vector $\boldsymbol{m}^*$ and the optimal coefficients $\alpha_2^*, \dots, \alpha_k^*$ derived in Theorem 3.4 depend on the variance and the correlation coefficients corresponding to the high-fidelity and the surrogate models. These quantities are usually unavailable and in practice we therefore replace them by their sample estimates. To estimate these quantities, we draw $m' \in \mathbb{N}$ i.i.d. realizations $\boldsymbol{z}_1', \dots, \boldsymbol{z}_{m'}' \in \mathcal{D}$ of the random variable Z and derive the Monte Carlo estimates $\overline{y}_{m'}^{(i)}$ of $\mathbb{E}[f^{(i)}(Z)]$ for $i = 1, \dots, k$. The sample variance is then

$$\bar{\sigma}_i^2 = \frac{1}{m' - 1} \sum_{l=1}^{m'} \left(f^{(i)}(\boldsymbol{z}_l') - \overline{y}_{m'}^{(i)} \right)^2 \tag{3.25}$$

and the sample correlation

$$\bar{\rho}_{1,i} = \frac{1}{\bar{\sigma}_1 \bar{\sigma}_i (m' - 1)} \sum_{l=1}^{m'} \left(f^{(1)}(\boldsymbol{z}_l') - \overline{y}_{m'}^{(1)} \right) \left(f^{(i)}(\boldsymbol{z}_l') - \overline{y}_{m'}^{(i)} \right) \tag{3.26}$$

for $i = 1, \dots, k$. While evaluating the models to obtain the sample variances and sample correlations, we measure the time needed for the model evaluations and derive the costs $w_1, \dots, w_k \in \mathbb{R}_+$. Note that we derive the sample estimates by evaluating the models at the m' realizations $\boldsymbol{z}_1', \dots, \boldsymbol{z}_{m'}'$ and so incur certain costs; however, in many situations, several model outputs are available from previous model runs,

Algorithm 2. Multifidelity Monte Carlo.

procedure MFMC($f^{(1)}, \dots, f^{(k)}, \bar{\sigma}_1, \dots, \bar{\sigma}_k, \bar{\rho}_{1,1}, \dots, \bar{\rho}_{1,k}, w_1, \dots, w_k, p$)
  Ensure $f^{(1)}, \dots, f^{(k)}$ is the result of mSelect defined in Algorithm 1
  Set $\bar{\rho}_{1,k+1} = 0$ and define vector $\bar{\boldsymbol{r}} = [\bar{r}_1, \dots, \bar{r}_k]^T \in \mathbb{R}_+^k$ as

$$\bar{r}_i = \sqrt{\frac{w_1(\bar{\rho}_{1,i}^2 - \bar{\rho}_{1,i+1}^2)}{w_i(1 - \bar{\rho}_{1,2}^2)}}, \qquad i = 1, \dots, k$$

  Select number of model evaluations $\boldsymbol{m} \in \mathbb{R}_+^k$ as

$$\boldsymbol{m} = \left[\frac{p}{\boldsymbol{w}^T\bar{\boldsymbol{r}}}, \bar{r}_2 m_1, \dots, \bar{r}_k m_1\right]^T \in \mathbb{R}_+^k$$

  Round down components of $\boldsymbol{m}$ to obtain integers
  Set coefficients $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_k]^T \in \mathbb{R}^k$ to

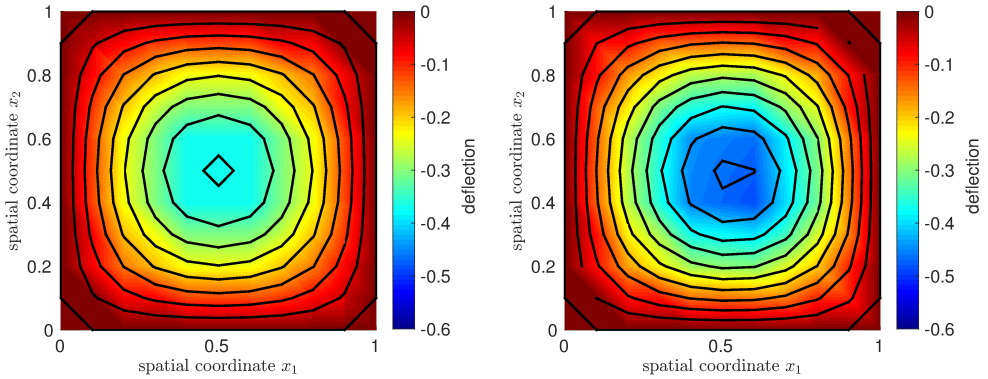
$$\alpha_i = \frac{\bar{\rho}_{1,i}\bar{\sigma}_1}{\bar{\sigma}_i}, \qquad i = 1, \dots, k$$

  Draw $\boldsymbol{z}_1, \dots, \boldsymbol{z}_{m_k} \in \mathcal{D}$ realizations of Z
  Evaluate model $f^{(i)}$ at realizations $\boldsymbol{z}_1, \dots, \boldsymbol{z}_{m_i}$ for $i = 1, \dots, k$
  Compute MFMC estimate $\hat{s}$ as in (3.2)
  return $\hat{s}$
end procedure

which can be reused to derive the sample estimates (e.g., from snapshot data used to create surrogate models). We also note that a small number m' is typically sufficient because small perturbations in the sample estimates have small effects on the number of model evaluations and the coefficients; see section 3.4 and the numerical example in section 4.

Algorithm 2 summarizes the computational procedure for deriving an MFMC estimate. Inputs to Algorithm 2 are the models $f^{(1)}, \dots, f^{(k)}$, the sample variances $\bar{\sigma}_1, \dots, \bar{\sigma}_k$, the sample correlations $\bar{\rho}_{1,1}, \dots, \bar{\rho}_{1,k}$, the costs $w_1, \dots, w_k$, and the computational budget $p \in \mathbb{R}_+$. Algorithm 2 first ensures that the models are in the order derived in Theorem 3.4 and satisfy condition (3.12), and it calls Algorithm 1 if necessary. The vector $\boldsymbol{m}$ and the coefficients $\boldsymbol{\alpha}$ are derived from the sample estimates of the variances and the correlation coefficients. Note that the variances and the correlation coefficients are replaced by their sample estimates and thus the obtained vector $\boldsymbol{m}$ and the coefficients $\boldsymbol{\alpha}$ are in general not the optimal vector $\boldsymbol{m}^*$ and optimal coefficients $\boldsymbol{\alpha}^*$ as derived in Theorem 3.4. Algorithm 2 then draws m_k realizations $\boldsymbol{z}_1, \dots, \boldsymbol{z}_{m_k} \in \mathcal{D}$ of the random variable Z and derives the MFMC estimate $\hat{s}$ as in (3.2), which is then returned. The costs of Algorithm 2 are usually dominated by the costs $\boldsymbol{w}^T\boldsymbol{m}$ of the model evaluations.

4. Numerical experiments. This section demonstrates the MFMC method on three numerical examples. Section 4.1 considers a locally damaged plate and estimates the mean deflection for uncertain properties of the plate. Section 4.2 presents a tubular reactor model that leads to either a steady-state or an oscillatory solution depending on the inputs. We estimate with our MFMC method the expected amplitude of



(a) deflection, no damage (b) deflection, damage up to 20%

FIG. 3. *Plate problem: A local damage at* $\boldsymbol{x}^* = [0.7, 0.4]^T \in \Omega$ *leads to a larger deflection of the plate.*

the oscillation for uncertain inputs. Section 4.3 presents a benchmark example to discuss the MFMC approach in case of surrogate models with significantly different correlation coefficients.

The reported runtime measurements were obtained on compute nodes with Intel Xeon E5-1620 CPUs and 32 GB RAM using a MATLAB implementation.

4.1. Locally damaged plate in bending. Consider the model of a clamped plate in bending [20] with a local damage [40, 39]. The spatial domain is $\Omega = [0, 1]^2 \subset \mathbb{R}^2$ and the inputs $\boldsymbol{z} = [z_1, \ldots, z_4]$ are realizations of the random variable Z with a uniform distribution in the input domain

$$\mathcal{D} = [z_1^{\min}, z_1^{\max}] \times [z_2^{\min}, z_2^{\max}] \times [z_3^{\min}, z_3^{\max}] \times [z_4^{\min}, z_4^{\max}] \subset \mathbb{R}^4$$

with $\boldsymbol{z}^{\min} = [z_1^{\min}, \ldots, z_4^{\min}]^T = [0.05, 1, 0, 0.01]^T \in \mathbb{R}^4$ and $\boldsymbol{z}^{\max} = [z_1^{\max}, \ldots, z_4^{\max}]^T = [0.1, 100, 0.5, 0.1]^T \in \mathbb{R}^4$. The first input z_1 controls the nominal thickness of the plate, the second input z_2 the load, and the third and fourth inputs z_3, z_4 the damage. The damage is a local decrease of the thickness. We define the thickness at position $\boldsymbol{x} \in \Omega$ and input $\boldsymbol{z} \in \mathcal{D}$ as the function $t : \Omega \times \mathcal{D} \to \mathbb{R}$ with

$$t(\boldsymbol{x}; \boldsymbol{z}) = z_1 \left(1 - z_3 \exp\left(-\frac{1}{2z_4^2} \|\boldsymbol{x} - \boldsymbol{x}^*\|_2^2\right)\right),$$

where z_1 is the nominal thickness, z_3, z_4 are the inputs that control the damage, and $\boldsymbol{x}^* = [0.7, 0.4]^T \in \Omega$ is the position of the damage. Figure 3 shows the plate without damage, i.e., $z_3 = 0$, and with a 20% decrease of the thickness, i.e., $z_3 = 0.2$. The output $y \in \mathcal{Y} \subset \mathbb{R}$ of the model is the mean deflection of the plate.

The high-fidelity model $f^{(1)} : \mathcal{D} \to \mathcal{Y}$ of the plate problem is derived with the finite element method as described in [40, 20]. The discretized system of equations is nonlinear in the inputs. The high-fidelity model has 279 degrees of freedom. We are interested in the expected mean deflection $\mathbb{E}[f^{(1)}(Z)]$.

We create five different surrogate models as follows. We evaluate the high-fidelity model $f^{(1)}$ at 1000 realizations of the random variable Z and derive a proper orthogonal decomposition (POD) basis of the corresponding solutions. A reduced model

TABLE 1

The tables summarize the costs and the correlation coefficients of the models used in the plate problem in section 4.1 and in the tubular reactor problem in section 4.2.

	Costs [s]	Corr. coefficient
$f^{(1)}$	4.0894×10^{-1}	1.00000000×10^{0}
$f^{(2)}$	4.9890×10^{-3}	$9.99999983 \times 10^{-1}$
$f^{(3)}$	1.3264×10^{-3}	$9.99999216 \times 10^{-1}$
$f^{(4)}$	2.9550×10^{-4}	$9.99954506 \times 10^{-1}$
$f^{(5)}$	2.2260×10^{-5}	$9.98971009 \times 10^{-1}$
$f^{(6)}$	1.5048×10^{-6}	$9.97555261 \times 10^{-1}$

(a) plate problem

	Costs [s]	Corr. coefficient
$f^{(1)}$	4.4395×10^{1}	1.000000×10^{0}
$f^{(2)}$	6.8409×10^{-1}	9.999882×10^{-1}
$f^{(3)}$	2.9937×10^{-1}	9.999743×10^{-1}
$f^{(4)}$	1.9908×10^{-4}	9.958253×10^{-1}

(b) tubular reactor problem

$f^{(2)}$ with 10 degrees of freedom is derived via Galerkin projection of the discretized system of equations of the high-fidelity model onto the 10-dimensional POD space as described in [40]. Similarly, we define $f^{(3)}$ and $f^{(4)}$ to be reduced models of $f^{(1)}$ with two and five POD basis vectors, respectively. We further derive a data-fit surrogate model $f^{(5)}$ with piecewise linear interpolation over a tensorized grid in $\mathcal{D}$. Let therefore $\phi_i(z) = \max\{1 - |3z - i|, 0\}, i = 0, \ldots, 3$, be basis functions defined at the grid points $i/3, i = 0, \ldots, 3$ in $[0, 1] \subset \mathbb{R}$. The surrogate model $f^{(5)} : \mathcal{D} \to \mathcal{Y}$ is the piecewise linear interpolant defined as

$$f^{(5)}(\boldsymbol{z}) = \sum_{i_1=0}^{3} \sum_{i_2=0}^{3} \sum_{i_3=0}^{3} \sum_{i_4=0}^{3} \beta_{i_1,i_2,i_3,i_4} \prod_{j=1}^{4} \phi_{i_j} \left(\frac{z_j - z_j^{\min}}{z_j^{\max} - z_j^{\min}} \right),$$

where the coefficients β_{i_1,i_2,i_3,i_4} are derived from the interpolation conditions

$$f^{(5)}(\boldsymbol{z}_{i_1,i_2,i_3,i_4}) = f^{(1)}(\boldsymbol{z}_{i_1,i_2,i_3,i_4}), \qquad i_1, i_2, i_3, i_4 = 0, \ldots, 3,$$

at the equidistant grid points

$$\boldsymbol{z}_{i_1,i_2,i_3,i_4} = \begin{bmatrix} i_1/3(z_1^{\max} - z_1^{\min}) + z_1^{\min} \\ i_2/3(z_2^{\max} - z_2^{\min}) + z_2^{\min} \\ i_3/3(z_3^{\max} - z_3^{\min}) + z_3^{\min} \\ i_4/3(z_4^{\max} - z_4^{\min}) + z_4^{\min} \end{bmatrix} \in \mathcal{D}, \qquad i_1, i_2, i_3, i_4 = 0, \ldots, 3,$$

in the domain $\mathcal{D}$. The model $f^{(5)}$ is derived with the "griddedInterpolant" class of MATLAB and the option "linear" [33]. Note that there are other interpolation techniques for deriving surrogate models, e.g., sparse grid techniques [10]. We also derive a regression-based surrogate model $f^{(6)} : \mathcal{D} \to \mathcal{Y}$ with SVMs [16] from 256 realizations of Z and the corresponding high-fidelity model outputs. We use the libsvm implementation [11] with ϵ-SVM (option "-s 3") and radial basis functions (option "-t 2"). We perform a five-fold cross validation to select the kernel bandwidth and the costs parameters. The ϵ in the cost function is set to $\epsilon = 10^{-2}$. Overall, we have the high-fidelity model $f^{(1)}$, the reduced models $f^{(2)}, f^{(3)}, f^{(4)}$, the data-fit surrogate model $f^{(5)}$, and the SVM model $f^{(6)}$.

Let $\overline{y}_n^{(1)}$ be the Monte Carlo estimate of $\mathbb{E}[f^{(1)}(Z)]$ with $n = 10^5$ samples. The sample variances and the sample correlation coefficients for the models $f^{(1)}, \ldots, f^{(6)}$ are computed from 100 realizations of Z; see Table 1(a). Note that the models are ordered descending with respect to the absolute sample correlation coefficients; see

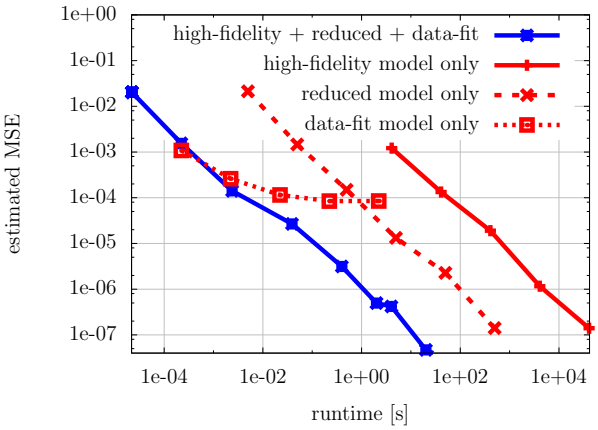


FIG. 4. *Plate model: The MFMC estimator combines the high-fidelity model $f^{(1)}$, the reduced model $f^{(2)}$, and the data-fit model $f^{(5)}$ to estimate the expectation of the deflection of the damaged plate. The runtime to derive an MFMC estimate is four orders of magnitude lower than the runtime to obtain the Monte Carlo estimate of comparable accuracy that uses the high-fidelity model only. Furthermore, the runtime of the MFMC estimator is two orders of magnitude lower than using the Monte Carlo method with the reduced model $f^{(2)}$.*

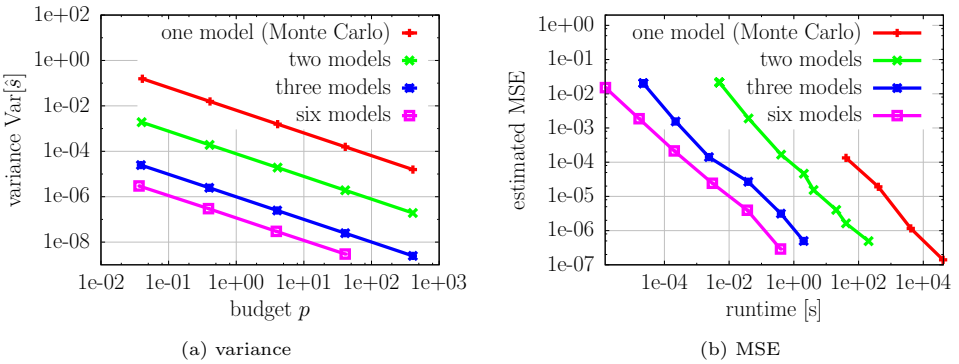
Theorem 3.4. We estimate the MSE of an estimator $\bar{s}$ over 10 runs as

$$\hat{e} = \frac{1}{10} \sum_{i=1}^{10} \left(\overline{y}_n^{(1)} - \bar{s}_i \right)^2, \tag{4.1}$$

where each of the estimates $\bar{s}_1, \ldots, \bar{s}_{10}$ is derived from independent samples. In case of the MFMC estimator $\bar{s} = \hat{s}$, Algorithm 2 is run 10 times to obtain the estimates $\hat{s}_1, \ldots, \hat{s}_{10}$.

Figure 4 reports the estimated MSE (4.1) of the Monte Carlo estimators that use the high-fidelity model $f^{(1)}$, the reduced model $f^{(2)}$, and the data-fit surrogate model $f^{(5)}$. The Monte Carlo estimators that use the reduced model $f^{(2)}$ and the data-fit surrogate model $f^{(5)}$ are biased estimators of expectation $\mathbb{E}[f^{(1)}(Z)]$, which can be seen in Figure 4 in the case of the data-fit surrogate model. Figure 4 compares the estimated MSEs of the Monte Carlo estimators to the estimated MSE of the MFMC estimator that combines the high-fidelity $f^{(1)}$, the reduced $f^{(2)}$, and the data-fit surrogate model $f^{(5)}$. The MFMC estimator achieves a speedup of up to four orders of magnitude compared to the Monte Carlo estimator that uses the high-fidelity model only and a speedup of up to two orders of magnitude compared to using the reduced model only. The results confirm that the MFMC estimator is an unbiased estimator of $\mathbb{E}[f^{(1)}(Z)]$. The MFMC estimator evaluates the data-fit surrogate model, but, in contrast to the Monte Carlo estimator that uses the data-fit surrogate model only, the MFMC estimator balances the model evaluations across all three models such that the low approximation quality of the data-fit surrogate model is compensated and its low computational costs are leveraged.

Figure 5(a) shows that even though the data-fit surrogate model is a poor approximation of the high-fidelity model, the variance of the MFMC estimator is significantly reduced if the data-fit surrogate model is combined with the high-fidelity and the reduced model. This is in agreement with the discussion in section 3.3, which states that the contribution of a surrogate model to the MFMC estimator depends on the



(a) variance (b) MSE

FIG. 5. *Plate model: The plot in* (a) *shows the variance of the MFMC estimator with two (high-fidelity $f^{(1)}$, reduced $f^{(2)}$), three (high-fidelity $f^{(1)}$, reduced $f^{(2)}$, data-fit $f^{(5)}$), and all six models $f^{(1)}, \dots, f^{(6)}$ (high-fidelity $f^{(1)}$, reduced $f^{(2)}, f^{(3)}, f^{(4)}$, data-fit $f^{(5)}$, and SVM $f^{(6)}$). Compared to the Monte Carlo method with the high-fidelity model only, a variance reduction of about four orders of magnitude is achieved. This is similar to the speedup obtained with the MFMC estimator shown in* (b).

properties of the model but also how it relates to the models already present in the MFMC estimator. Figure 5(a) also shows that the variance of the MFMC estimator that uses all six models $f^{(1)}, \dots, f^{(6)}$—the high-fidelity model, three reduced models, the data-fit model, and the SVM model—is only slightly lower than the variance of the MFMC estimator that uses the three models $f^{(1)}, f^{(2)}, f^{(5)}$. This again confirms that the contribution of a surrogate models to the variance reduction depends on how the surrogate model complements the models already present in the MFMC estimator. Note that the variance $\mathrm{Var}[\hat{s}]$ of the MFMC estimator can be estimated without model evaluations from the sample variances and the sample correlation coefficients with (3.6), and thus it is a computationally efficient guide for adding surrogate models to the MFMC estimator. The estimated MSE shown in Figure 5(b) confirms the variance reduction results in Figure 5(a).

Figure 6 reports the relative share of each model in the total number of model evaluations, i.e., in the total number of samples. The shares of the models vary by orders of magnitude between the high-fidelity, the reduced, the data-fit, and the SVM models, reflecting their correlations and costs. Note that the relative shares of the models are independent of the computational budget p, because all components of $\boldsymbol{m}^*$ scale linearly with p; see Theorem 3.4.

We use the sample variances and the sample correlation coefficients to determine the number of model evaluations $\boldsymbol{m}$ and the coefficients $\boldsymbol{\alpha}$. Table 2 compares sample variances and sample correlation coefficients computed from 10, 100, and 1000 samples. The different number of samples leads to different estimates. Note that even though the sample variances $\bar{\sigma}_i, i = 1, \dots, k$, change by a factor of two when increasing the sample size from 10 to 100, the ratios $\bar{\sigma}_1/\bar{\sigma}_i, i = 1, \dots, k$ change only slightly. Since only the ratios $\bar{\sigma}_1/\bar{\sigma}_i, i = 1, \dots, k$, enter the computation of the coefficients $\boldsymbol{\alpha}$, the variations in the sample variances have only a minor effect on the coefficients $\boldsymbol{\alpha}$. This is confirmed by Figure 7, which shows that the perturbations in the sample variances and the sample correlation coefficients have a small effect on the estimated MSE of the MFMC estimator and on the distribution of the work; see also the discussion in section 3.4 and Figure 2.

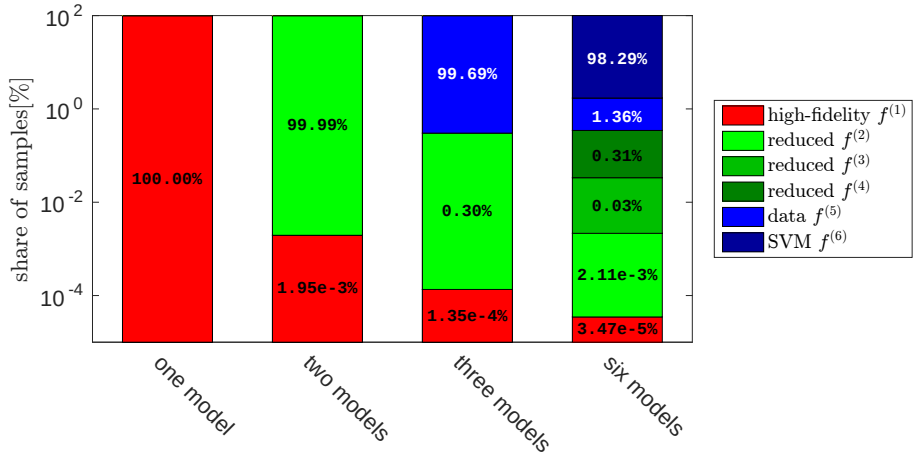


FIG. 6. *Plate model: The plot reports how Algorithm* 2 *distributes the total number of samples across the high-fidelity, the reduced, the data-fit, and the SVM models for the MFMC estimator with two, three, and six models. Note the logarithmic scale on the y-axis.*

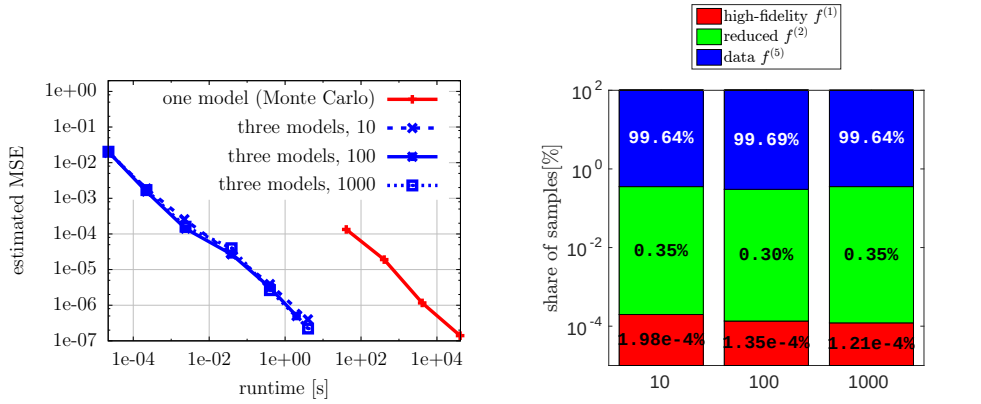
TABLE 2

Plate example: The table compares the sample correlation coefficients and the sample variances computed from 10, 100, *and* 1000 *samples, respectively.*

	Sample correlation coefficients		
	From 10 samples	From 100 samples	From 1000 samples
$f^{(1)}$	1.00000000×10^{0}	1.00000000×10^{0}	1.00000000×10^{0}
$f^{(2)}$	$9.99999963 \times 10^{-1}$	$9.99999983 \times 10^{-1}$	$9.99999986 \times 10^{-1}$
$f^{(3)}$	$9.99997494 \times 10^{-1}$	$9.99999216 \times 10^{-1}$	$9.99999238 \times 10^{-1}$
$f^{(4)}$	$9.99882814 \times 10^{-1}$	$9.99954506 \times 10^{-1}$	$9.99948132 \times 10^{-1}$
$f^{(5)}$	$9.98582674 \times 10^{-1}$	$9.98971009 \times 10^{-1}$	$9.98580437 \times 10^{-1}$
$f^{(6)}$	$9.94953355 \times 10^{-1}$	$9.97555261 \times 10^{-1}$	$9.96563656 \times 10^{-1}$
	Sample variances		
	From 10 samples	From 100 samples	From 1000 samples
$f^{(1)}$	5.322×10^{-3}	1.5511×10^{-2}	1.4968×10^{-2}
$f^{(2)}$	5.322×10^{-3}	1.5510×10^{-2}	1.4967×10^{-2}
$f^{(3)}$	5.338×10^{-3}	1.5531×10^{-2}	1.4975×10^{-2}
$f^{(4)}$	5.306×10^{-3}	1.5535×10^{-2}	1.4946×10^{-2}
$f^{(5)}$	6.921×10^{-3}	1.8963×10^{-2}	1.7429×10^{-2}
$f^{(6)}$	6.051×10^{-3}	1.5566×10^{-2}	1.4298×10^{-2}

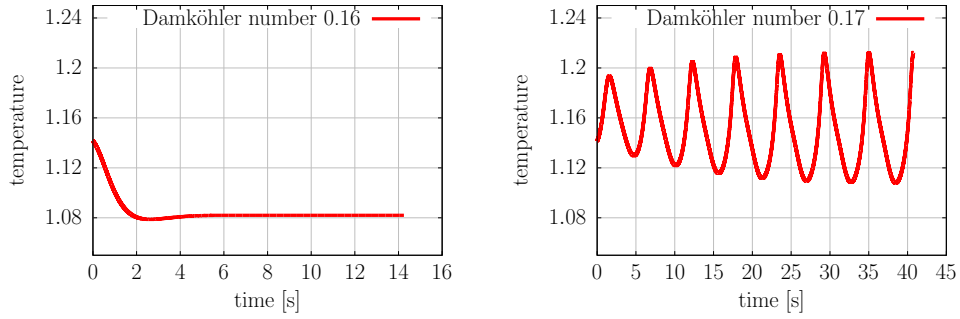
4.2. Limit cycle oscillation in tubular reactor. Consider a one-dimensional nonadiabatic tubular reactor with a single reaction and axial mixing as introduced in [26]. The spatial domain is $\Omega = [0,1] \subset \mathbb{R}$, the time domain is $[0,T] \subset \mathbb{R}$ with $T = 500$s, and the input domain is $\mathcal{D} = \mathbb{R}$. The governing equations are coupled nonlinear time-dependent convection-diffusion-reaction equations

$$\frac{\partial}{\partial t}u^c(x,t;z) = \frac{1}{Pe}\frac{\partial^2}{\partial x^2}u^c(x,t;z) - \frac{\partial}{\partial x}u^\theta(x,t;z) - zg(u^c,u^\theta),$$
$$\frac{\partial}{\partial t}u^\theta(x,t;z) = \frac{1}{Pe}\frac{\partial^2}{\partial x^2}u^\theta(x,t;z) - \frac{\partial}{\partial x}u^\theta(x,t;z) - \beta(u^\theta(x,t;z)-\theta_0) + \epsilon zg(u^c,u^\theta),$$



(a) estimated MSE (b) distribution of work among models

FIG. 7. *Plate model: We use the sample variances and sample correlation coefficients to determine the number of model evaluations* $\boldsymbol{m}$ *and the coefficients* $\boldsymbol{\alpha}$. *The plot in* (a) *compares the estimated MSE of our MFMC estimators (with models* $f^{(1)}, f^{(2)}, f^{(5)}$*) obtained from the sample variances and sample correlation coefficients from* 10, 100, *and* 1000 *samples, respectively; see Table* 2. *The plot indicates that small perturbations in the sample correlation coefficients have a small effect on the MSE of the MFMC estimators. The plot in* (b) *confirms that small perturbations in the sample estimates lead to small changes in the number of model evaluations.*



(a) steady-state regime (b) LCO regime

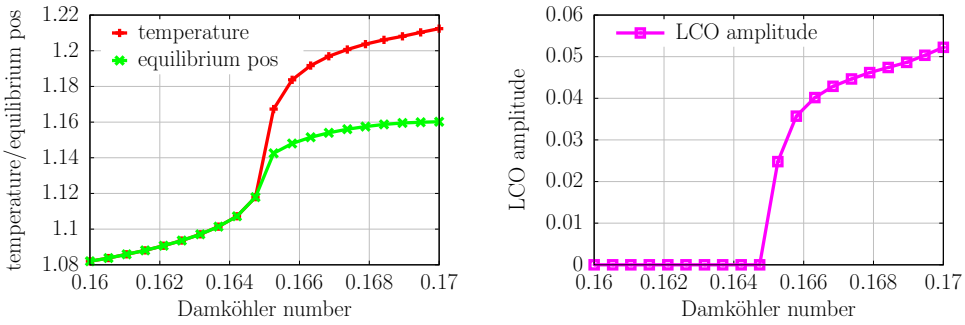
FIG. 8. *Tubular reactor: If the input* $z < z^*$ *is below the critical parameter* $z^* \in \mathcal{D}$, *the temperature and the concentration of the tubular reactor converge to a steady state as shown in* (a) *for the temperature at* $1 \in \Omega$. *If* $z > z^*$, *the tubular reactor enters an LCO as plotted in* (b).

with the concentration $u^c : \Omega \times [0,T] \times \mathcal{D} \to \mathbb{R}$, the temperature $u^\theta : \Omega \times [0,T] \times \mathcal{D} \to \mathbb{R}$, and

$$g(u^c, u^\theta) = u^c \exp\left(\gamma - \frac{\gamma}{u^\theta}\right).$$

The nonlinear function g models an Arrhenius-type nonlinear reaction term. The Péclet number is $Pe = 5$ and $\gamma = 25, \beta = 2.5, \epsilon = 0.5$, and $\theta_0 = 1$ are known constants. We impose Robin boundary conditions at $x = 0$, Neumann boundary conditions at $x = 1$, and the initial condition as in [26].

The input $z \in \mathcal{D}$ is the Damköhler number. The Damköhler number controls the behavior of the reactor. If $z < 0.165$, the concentration and the temperature of the reactor converge to a steady-state solution as shown in Figure 8(a). If $z > 0.165$,



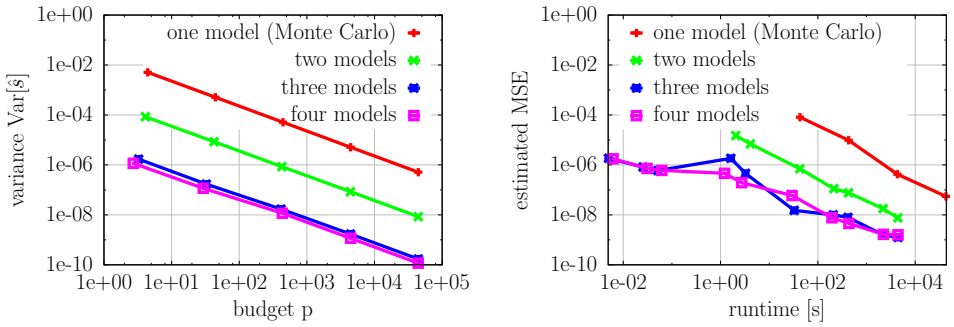
(a) bifurcation diagram (b) LCO amplitude

FIG. 9. *Tubular reactor: The bifurcation diagram in* (a) *visualizes the two regimes of the tubular reactor. The LCO amplitude as shown in* (b) *is the difference between the temperature and the equilibrium position.*

the reactor enters a limit cycle oscillation (LCO) around a nontrivial equilibrium position as shown in Figure 8(b). These two regimes are represented in the bifurcation diagram in Figure 9(a). The LCO amplitude is the amplitude of the oscillation of the temperature at $x = 1$. Figure 9(b) shows the LCO amplitude for inputs in $[0.16, 0.17]$, where for $z < 0.165$ the reactor converges to a steady-state solution and therefore the LCO amplitude is zero. We define the LCO amplitude as the output of the tubular reactor model.

The high-fidelity model $f^{(1)}$ is derived as in [51] and is based on a finite difference discretization of the governing equations on a grid with 101 equidistant grid points in the spatial domain Ω. The high-fidelity model has 198 degrees of freedom. The model is marched forward in time with an explicit fourth-order Runge–Kutta method with time step size 10^{-4}s. The time stepping is stopped if either the solution converged to a steady state or an LCO is detected. In order to derive two reduced models, we evaluate the high-fidelity model at 20 inputs $z_1, \ldots, z_{20} \in [0.16, 0.17]$ that coincide with an equidistant grid in $[0.16, 0.17]$. The concentration and temperature at the grid points in the spatial domain are stored every 0.25 s. A reduced basis is constructed with POD from the stored concentration and temperature. The reduced model $f^{(2)}$ is derived with Galerkin projection of the discretized equations of the high-fidelity model onto the reduced space spanned by the first 10 POD basis vectors. The nonlinear function g is evaluated explicitly—i.e., in each time step, the intermediate reduced solution is projected onto the high-dimensional solution space of the high-fidelity model, the function g is evaluated, and the result is projected back onto the 10-dimensional POD space. The reduced model $f^{(3)}$ is derived as $f^{(2)}$ except that the nonlinear function g is approximated by the discrete empirical interpolation method (DEIM) [12] as described in [51]. The number of DEIM basis vectors and the number of DEIM interpolation points is eight. Similarly to section 4.1, we additionally construct a data-fit surrogate model $f^{(4)}$ with piecewise cubic interpolation from the inputs coinciding with the 10 equidistant grid points in $[0.16, 0.17]$ and the corresponding outputs. The piecewise cubic interpolant is derived with the "interp1" command in MATLAB and option "cubic" [33]. We therefore have four models: the high-fidelity model $f^{(1)}$, two reduced models $f^{(2)}, f^{(3)}$, and the data-fit surrogate model $f^{(4)}$.

We are interested in the expectation of the LCO amplitude $\mathbb{E}[f^{(1)}(Z)]$ if the Damköhler number is uncertain and a realization of the random variable Z with a



(a) variance

(b) estimated MSE (4.1)

FIG. 10. *Tubular reactor: The plot in* (a) *compares the variance of the Monte Carlo estimator with the high-fidelity model $f^{(1)}$ (one model) to the MFMC estimator that uses the high-fidelity model $f^{(1)}$ and the reduced model $f^{(2)}$ (two models) and shows the reduction that is achieved when the data-fit model $f^{(4)}$ is added (three models) and when the reduced model $f^{(3)}$ is added (four models). The MFMC estimator achieves an up to four orders of magnitude higher accuracy than the Monte Carlo estimator; see* (b).

normal distribution with mean 0.167 and standard deviation 0.03. The Monte Carlo estimate $\overline{y}_n^{(1)}$ of $\mathbb{E}[f^{(1)}(Z)]$ is computed from $n = 10^3$ realizations of Z and is used to estimate the MSE as in (4.1). The estimated MSE is computed as in (4.1) over 10 runs. The sample variances and sample correlation coefficients are derived from 100 realizations of Z; see Table 1(b).

Figure 10(a) shows the variance of the MFMC estimators using up to four models. The variance is computed with the sample variances and sample correlation coefficients. Combining the high-fidelity model $f^{(1)}$ with the reduced model $f^{(2)}$ that explicitly evaluates the nonlinear function g and the data-fit surrogate model $f^{(4)}$ leads to a variance reduction of about four orders of magnitude. Adding the reduced model $f^{(3)}$ that approximates the nonlinear function with DEIM improves the variance only slightly. Note that a similar situation was observed in section 4.1 and Figure 5(a). The variance reduction is reflected in Figure 10(b), where the estimated MSE of the MFMC estimator with the high-fidelity $f^{(1)}$, the reduced $f^{(2)}$, and the data-fit surrogate model $f^{(4)}$ achieves an improvement of about four orders of magnitude compared to the Monte Carlo estimator that uses the high-fidelity model only.

4.3. Short column. To demonstrate MFMC on surrogate models with low correlation coefficients, we consider an analytic model of a short column with rectangular cross-sectional area subject to bending and axial force. The analytic model is used in [28, 29, 35] in the context of reliability-based optimal design. We follow [28] and define the high-fidelity model as

$$f^{(1)}(\boldsymbol{z}) = 1 - \frac{4z_4}{z_1 z_2^2 z_3} - \left(\frac{z_5}{z_1 z_2 z_3}\right)^2,$$

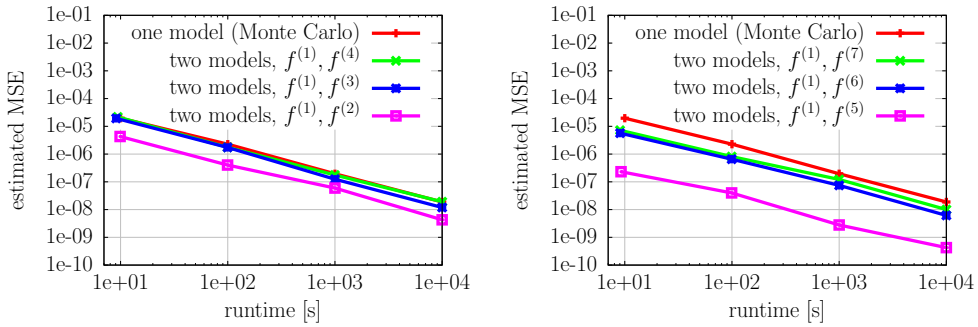
where $\boldsymbol{z} = [z_1, \ldots, z_5]^T \in \mathcal{D}$ with

$$\mathcal{D} = [5, 15] \times [15, 25] \times \mathbb{R}_+ \times \mathbb{R} \times \mathbb{R}.$$

Let Z be the input random variable such that the width z_1 is distributed uniformly in $[5, 15]$, the depth z_2 is distributed uniformly in $[15, 25]$, the yield stress z_3 is log-normally distributed with mean 5 and standard deviation 0.5, the bending moment

TABLE 3
Short column: The table reports the correlation coefficients and the costs of models $f^{(1)}, \ldots, f^{(7)}$.

	$f^{(1)}$	$f^{(2)}$	$f^{(3)}$	$f^{(4)}$	$f^{(5)}$	$f^{(6)}$	$f^{(7)}$
Corr. coeff.	1.000	0.9905	0.8251	0.7183	0.9905	0.8251	0.7183
Costs	1	10^{-1}	10^{-1}	10^{-1}	10^{-5}	10^{-5}	10^{-5}



(a) costs $w_2 = w_3 = w_4 = 10^{-1}$ (b) costs $w_5 = w_6 = w_7 = 10^{-5}$

FIG. 11. *Short column: The plots in* (a) *and* (b) *show that even surrogate models with low correlation coefficients can lead to runtime speedups if the costs are low.*

z_4 is distributed normally with mean 2000 and standard deviation 400, and the axial force z_5 is distributed normally with mean 500 and standard deviation 100 [28, 29]. Since the computational costs of evaluating the high-fidelity model are negligible, we assign the costs $w_1 = 1$ for demonstration purposes.

We derive three functions of $f^{(1)}$ that serve as our surrogate models:

$$f^{(2)}(\boldsymbol{z}) = 1 - \frac{z_4}{z_1 z_2^2 z_3} - \left(\frac{z_5}{z_1 z_2 z_3}\right)^2,$$
$$f^{(3)}(\boldsymbol{z}) = 1 - \frac{z_4}{z_1 z_2^2 z_3} - \left(\frac{z_5(1+z_4)}{z_1 z_2 z_3}\right)^2,$$
$$f^{(4)}(\boldsymbol{z}) = 1 - \frac{z_4}{z_1 z_2^2 z_3} - \left(\frac{z_5(1+z_4)}{z_2 z_3}\right)^2.$$

We assign the costs $w_2 = w_3 = w_4 = 10^{-1}$. We additionally derive the surrogate models $f^{(5)} = f^{(2)}$, $f^{(6)} = f^{(3)}$, and $f^{(7)} = f^{(4)}$ with costs $w_5 = w_6 = w_7 = 10^{-5}$. The sample estimates of the correlation coefficients from 100 realizations of the random variable Z are reported in Table 3. Note that we derived the surrogate models such that we obtain the correlation coefficients $\rho_{1,2} \approx 0.9$, $\rho_{1,3} \approx 0.8$, and $\rho_{1,4} \approx 0.7$.

We are interested in the expected value $\mathbb{E}[f^{(1)}(Z)]$. The Monte Carlo estimate $\overline{y}_n^{(1)}$ of $\mathbb{E}[f^{(1)}(Z)]$ is computed from $n = 10^6$ realizations of Z and is used to estimate the MSE as in (4.1) over 100 runs. First consider the estimated MSE of the MFMC estimator with the models $f^{(1)}, f^{(4)}$ as reported in Figure 11(a). The correlation coefficient ($\rho_{1,4} \approx 0.7$) between $f^{(1)}$ and $f^{(4)}$ is low and the costs w_4 high, and therefore no savings can be obtained compared to the benchmark Monte Carlo estimator that uses the high-fidelity model $f^{(1)}$ only. Models $f^{(3)}$ and $f^{(2)}$ have a larger correlation coefficient than $f^{(4)}$ and therefore lead to larger runtime savings. A similar situation can be seen in Figure 11(b) for the models $f^{(5)}, f^{(6)}$, and $f^{(7)}$, which have the same

correlation coefficients as the models $f^{(2)}$, $f^{(3)}$, and $f^{(4)}$, respectively, but significantly lower costs.

5. Conclusions. The proposed MFMC method leverages computationally cheap surrogate models to reduce the runtime of estimating statistics of expensive high-fidelity models while maintaining unbiasedness of the resulting estimator. An optimization problem with an analytic solution optimally balances the model evaluations across the high-fidelity model and an arbitrary number of surrogate models of any type, including projection-based reduced models, data-fit surrogate models, support vector machines, and simplified models. The MFMC estimator achieved runtime speedups by several orders of magnitude in the presented numerical examples compared to the Monte Carlo estimator that uses the high-fidelity model only.

The benefit of adding a surrogate model to the models available to the MFMC estimator depends on the properties of the surrogate model itself and also on the new information introduced by the surrogate model compared to information provided by the other models already present in the MFMC estimator. The mathematical and numerical results show that combining surrogate models of different types, approximation quality, and costs is often more beneficial than combining accurate surrogate models only. The MFMC estimator is unbiased, independent of the availability of a priori error bounds and a posteriori error estimators for the surrogate models.

Future work includes an extension to models with vector-valued outputs. An ad hoc way to cope with vector-valued outputs is to perform the model management with respect to each component of the output separately. This leads to multiple model management solutions, from which the one with the largest number of high-fidelity model evaluations can be selected. An extension that can directly handle vector-valued models is future work.

Appendix A. Auxiliary lemmas.

LEMMA A.1. *Consider the same setting as in Theorem* 3.4. *Let* $(\boldsymbol{m}, \alpha_2, \dots, \alpha_k)$ *be a local minimum of* (3.10) *with budget* $p \in \mathbb{R}_+$. *Define* $\alpha_1 = 1$ *and* $\alpha_{k+1} = 0$. *Let* $\boldsymbol{l} = [l_1, \dots, l_q]^T \in \mathbb{N}^q$ *be the vector with the components* $l_1, \dots, l_q \in \{2, \dots, k\}$ *that are the* $q \in \mathbb{N}$ *indices with*

$$m_{l_i - 1} < m_{l_i}, \qquad i = 1, \dots, q, \tag{A.1}$$

and

$$m_{l_i} = m_{l_i + 1} = \cdots = m_{l_{i+1} - 1}, \qquad i = 0, \dots, q,$$

and $l_0 = 1$ *and* $l_{q+1} = k + 1$. *Define* $\boldsymbol{r} = [r_1, \dots, r_k]^T \in \mathbb{R}_+^k$ *with the components* $r_i = m_i / m_1$ *with* $i = 1, \dots, k$. *Then, the local minimum* $(\boldsymbol{m}, \alpha_2, \dots, \alpha_k)$ *leads to*

$$r_{l_i} = \frac{m_{l_i}}{m_1} = \sqrt{\frac{\left(\sum_{j=1}^{l_1 - 1} w_j\right)}{\left(\sum_{j=l_i}^{l_{i+1} - 1} w_j\right)} \frac{(\rho_{1,l_i}^2 - \rho_{1,l_{i+1}}^2)}{(1 - \rho_{1,l_1}^2)}} \tag{A.2}$$

with $i = 0, \dots, q$ *and* $m_1 > 0$. *The* q *coefficients* $\alpha_{l_1}, \dots, \alpha_{l_q}$ *are*

$$\alpha_{l_i} = \frac{\rho_{1,l_i} \sigma_1}{\sigma_{l_i}} \tag{A.3}$$

for $i = 1, \dots, q$ *and*

$$m_1 = \frac{p}{\boldsymbol{w}^T \boldsymbol{r}}. \tag{A.4}$$

Proof. A vector $\boldsymbol{l} \in \mathbb{N}^q$ exists and is unique because of the strict inequality in (A.1). If $m_1 = \cdots = m_k$, then $q = 0$. Consider now the Lagrangian

$$\hat{J}(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k, \lambda, \xi, \mu_2, \ldots, \mu_k) = J(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k) + \lambda(\boldsymbol{w}^T\boldsymbol{m} - p) - \xi m_1 + \sum_{i=2}^{k} \mu_i(m_{i-1} - m_i) \tag{A.5}$$

of the optimization problem (3.10) with the Lagrange multipliers $\lambda, \xi, \mu_2, \ldots, \mu_k \in \mathbb{R}$. The partial derivatives of J with respect to $\boldsymbol{m}$ are

$$\begin{bmatrix} \frac{\partial \hat{J}}{\partial m_1} \\ \vdots \\ \frac{\partial \hat{J}}{\partial m_k} \end{bmatrix} = \begin{bmatrix} -\frac{1}{m_1^2}\sigma_1^2 - \frac{1}{m_1^2}(\alpha_2^2\sigma_2^2 - 2\alpha_2\rho_{1,2}\sigma_1\sigma_2) + \lambda w_1 + \mu_2 - \xi \\ \frac{1}{m_2^2}(\alpha_2^2\sigma_2^2 - 2\alpha_2\rho_{1,2}\sigma_1\sigma_2) - \frac{1}{m_2^2}(\alpha_3^2\sigma_3^2 - 2\alpha_3\rho_{1,3}\sigma_1\sigma_3) + \lambda w_2 - \mu_2 + \mu_3 \\ \frac{1}{m_3^2}(\alpha_3^2\sigma_3^2 - 2\alpha_3\rho_{1,3}\sigma_1\sigma_3) - \frac{1}{m_3^2}(\alpha_4^2\sigma_4^2 - 2\alpha_4\rho_{1,4}\sigma_1\sigma_4) + \lambda w_3 - \mu_3 + \mu_4 \\ \frac{1}{m_4^2}(\alpha_4^2\sigma_4^2 - 2\alpha_4\rho_{1,4}\sigma_1\sigma_4) - \frac{1}{m_4^2}(\alpha_5^2\sigma_5^2 - 2\alpha_5\rho_{1,5}\sigma_1\sigma_5) + \lambda w_4 - \mu_4 + \mu_5 \\ \vdots \\ \frac{1}{m_k^2}(\alpha_k^2\sigma_k^2 - 2\alpha_k\rho_{1,k}\sigma_1\sigma_k) + \lambda w_k - \mu_k \end{bmatrix}, \tag{A.6}$$

and with respect to $\alpha_2, \ldots, \alpha_k$

$$\begin{bmatrix} \frac{\partial \hat{J}}{\partial \alpha_2} \\ \vdots \\ \frac{\partial \hat{J}}{\partial \alpha_k} \end{bmatrix} = \begin{bmatrix} \left(\frac{1}{m_1} - \frac{1}{m_2}\right)\left(2\alpha_2\sigma_2^2 - 2\rho_{1,2}\sigma_1\sigma_2\right) \\ \vdots \\ \left(\frac{1}{m_{k-1}} - \frac{1}{m_k}\right)\left(2\alpha_k\sigma_k^2 - 2\rho_{1,k}\sigma_1\sigma_k\right) \end{bmatrix}. \tag{A.7}$$

Since $(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k)$ is a local minimum of (3.10) and since the constraints are affine functions in $\boldsymbol{m}$ and $\alpha_2, \ldots, \alpha_k$, it follows with the Karush–Kuhn–Tucker (KKT) conditions [7] that $\lambda, \xi, \mu_2, \ldots, \mu_k \in \mathbb{R}$ exist with

$$\begin{bmatrix} \frac{\partial \hat{J}}{\partial m_1} & \ldots & \frac{\partial \hat{J}}{\partial m_k} & \frac{\partial \hat{J}}{\partial \alpha_2} & \ldots & \frac{\partial \hat{J}}{\partial \alpha_k} \end{bmatrix} = 0, \tag{A.8}$$

$$m_{i-1} - m_i \leq 0, \qquad i = 2, \ldots, k, \tag{A.9}$$

$$-m_1 \leq 0, \tag{A.10}$$

$$\boldsymbol{w}^T\boldsymbol{m} - p = 0, \tag{A.11}$$

$$\xi, \mu_2, \ldots, \mu_k \geq 0, \tag{A.12}$$

$$\mu_i(m_{i-1} - m_i) = 0, \qquad i = 2, \ldots, k, \tag{A.13}$$

$$-\xi m_1 = 0. \tag{A.14}$$

We use the KKT conditions (A.8)–(A.14) to show that the local minimum $(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k)$ leads to the coefficients $\alpha_{l_1}, \ldots, \alpha_{l_q}$ as in (A.3), to $m_1 > 0$, to $\boldsymbol{r} = [r_1, \ldots, r_k]^T$ as defined in (A.2), and to $m_1 = p/(\boldsymbol{w}^T\boldsymbol{r})$ as in (A.4).

First consider the case where $m_1 = \cdots = m_k$ and therefore $q = 0$. Since $q = 0$, nothing is to show for the coefficients. The vector $\boldsymbol{m}$ satisfies the constraint $\boldsymbol{w}^T\boldsymbol{m} = p$

because $(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k)$ is a local minimum of (3.10). Because $p \in \mathbb{R}_+$ and $0 < w_1, \ldots, w_k$, we have $m_1 > 0$. The definition of $\boldsymbol{r}$ leads to $r_1 = \cdots = r_k = 1$. With the definition $l_0 = 1$ and $l_{q+1} = k + 1$ and the convention $\rho_{1,k+1} = 0$ we have $\boldsymbol{r}$ as in (A.2). We further have $m_1 = p/(\boldsymbol{w}^T\boldsymbol{r})$ because $\boldsymbol{r} = [1, \ldots, 1]^T$.

Consider now the case with $q > 0$, i.e., $m_{l_i-1} < m_{l_i}$ for $i = 1, \ldots, q$. With $m_{l_i-1} < m_{l_i}$ we also obtain $m_{l_i-1} \neq m_{l_i}$ for $i = 1, \ldots, q$. Because $(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k)$ is a local minimum, the partial derivatives of $\hat{J}$ with respect to $\alpha_{l_1}, \ldots, \alpha_{l_q}$ in (A.7) have to be zero (see (A.8)), which leads to the coefficients as in (A.3).

We now show $m_1 > 0$ and $\xi = 0$. Evaluating the objective function J at the local minimum $(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k)$ gives

$$J(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k) = \frac{\sigma_1^2}{m_1} + \sum_{i=1}^{q} \left(\frac{1}{m_{l_i-1}} - \frac{1}{m_{l_i}} \right) (-\rho_{1,l_i}^2 \sigma_1^2),$$

where we used the coefficients as in (A.3). Since $m_1 < m_{l_1} < m_{l_2} < \cdots < m_{l_k}$ per definition of $\boldsymbol{l}$, and since $\sigma_1^2 \neq \rho_{1,l_1}^2 \sigma_1^2$ because $1 = \rho_{1,1}^2 > \rho_{1,l_1}^2$ (see the ordering with respect to the squared correlation coefficients in Theorem 3.4), the objective J converges for $m_1 \to 0$ to ∞^+ from above. Therefore, $m_1 > 0$ and with (A.14) we obtain $\xi = 0$.

The vector $\boldsymbol{m}$ satisfies the constraint $\boldsymbol{w}^T\boldsymbol{m} = p$ because $(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k)$ is a local minimum of (3.10), and therefore $m_1 = p/(\boldsymbol{w}^T\boldsymbol{r})$ as in (A.4). From (A.13) we derive the Lagrange multipliers $\mu_{l_1} = \cdots = \mu_{l_q} = 0$, because $m_{l_i} \neq m_{l_i-1}$ for $i = 1, \ldots, q$.

We now derive the Lagrange multiplier λ. Per definition of $\boldsymbol{l}$ we have $m_1 = \cdots = m_{l_1-1}$. The partial derivative of $\hat{J}$ with respect to m_{l_1-1}, i.e., the component $l_1 - 1$ of the vector of partial derivatives of $\hat{J}$ with respect to $\boldsymbol{m}$ (A.6), leads to

$$\frac{1}{m_1^2}(\alpha_{l_1-1}^2 \sigma_{l_1-1}^2 - 2\alpha_{l_1-1}\rho_{1,l_1-1}\sigma_1\sigma_{l_1-1}) - \frac{1}{m_1^2}(-\rho_{1,l_1}^2\sigma_1^2) + \lambda w_{l_1-1} - \mu_{l_1-1} = 0, \tag{A.15}$$

where we used $\alpha_{l_1} = \rho_{1,l_1}\sigma_1/\sigma_{l_1}$, $\mu_{l_1} = 0$, and $\xi = 0$. Note that $\alpha_1 = 1$ per definition and therefore (A.15) holds also in case $l_1 = 2$. If $l_1 > 2$, we find with the components $1, \ldots, l_1 - 2$ of the vector of partial derivatives of $\hat{J}$ with respect to $\boldsymbol{m}$ (A.6) that

$$\alpha_{l_1-1}^2 \sigma_{l_1-1}^2 - 2\alpha_{l_1-1}\rho_{1,l_1-1}\sigma_1\sigma_{l_1-1} = -\sigma_1^2 + \lambda m_1^2 \sum_{i=1}^{l_1-2} w_i + m_1^2 \mu_{l_1-1}, \tag{A.16}$$

because $m_1 = \cdots = m_{l_1-1}$ and $\mu_2, \ldots, \mu_{l_1-2}$ cancel. We use (A.16) in (A.15) and obtain

$$\frac{1}{m_1^2}\left(-\sigma_1^2 + \lambda m_1^2 \sum_{i=1}^{l_1-2} w_i \right) - \frac{1}{m_1^2}(-\rho_{1,l_1}^2\sigma_1^2) + \lambda w_{l_1-1} = 0,$$

where μ_{l_1-1} canceled, which leads to

$$\lambda = \frac{\sigma_1^2(1 - \rho_{1,l_1}^2)}{m_1^2 \sum_{i=1}^{l_1-1} w_i}. \tag{A.17}$$

We now derive $r_1, \ldots, r_k$. We have per definition $r_1 = \cdots = r_{l_1-1} = 1$ because $m_1 = m_2 = \cdots = m_{l_1-1}$. Consider now $m_{l_i} = \cdots = m_{l_{i+1}-1}$ for $i = 1, \ldots, q$. The components l_i to $l_{i+1} - 1$ of the vector of partial derivatives of $\hat{J}$ with respect to $\boldsymbol{m}$ (A.6) with condition (A.8) of the KKT conditions lead to

$$\frac{1}{m_{l_i}^2}(-\rho_{1,l_i}^2\sigma_1^2) + \lambda \sum_{j=l_i}^{l_{i+1}-1} w_j - \frac{1}{m_{l_i}^2}(-\rho_{1,l_{i+1}}^2\sigma_1^2) = 0, \tag{A.18}$$

where we used $\mu_{l_i} = \mu_{l_{i+1}} = 0$, $\alpha_{l_i} = \rho_{1,l_i}\sigma_1/\sigma_{l_i}$, and $\alpha_{l_{i+1}} = \rho_{1,l_{i+1}}\sigma_1/\sigma_{l_{i+1}}$. Note that $\mu_{l_i+1}, \ldots, \mu_{l_{i+1}-1}$ cancel. Insert λ as derived in (A.17) into (A.18) and obtain

$$\frac{\sigma_1^2}{m_{l_i}^2}(\rho_{1,l_i}^2 - \rho_{1,l_{i+1}}^2) = \frac{\sigma_1^2(1-\rho_{1,l_1}^2)}{m_1^2 \sum_{j=1}^{l_1-1} w_j} \sum_{j=l_i}^{l_{i+1}-1} w_j.$$

This leads to

$$r_{l_i}^2 = \frac{m_{l_i}^2}{m_1^2} = \frac{\left(\sum_{i=1}^{l_1-1} w_i\right)}{\left(\sum_{j=l_i}^{l_{i+1}-1} w_j\right)} \frac{(\rho_{1,l_i}^2 - \rho_{1,l_{i+1}}^2)}{(1-\rho_{1,l_1}^2)}.$$

Note that $1 = \rho_{1,1}^2 > \rho_{1,l_1}^2$ (see Theorem 3.4) and therefore $1 - \rho_{1,l_1}^2 > 0$. The positivity of the components of $\boldsymbol{m}$ given by $m_1 > 0$ and the inequality constraints of the optimization problem, the convention $\rho_{1,k+1} = 0$, and the definition $\alpha_{k+1} = 0$ lead to $\boldsymbol{r}^*$ as given in (A.2). ∎

LEMMA A.2. *Consider the same setting as in Theorem* 3.4. *The value of the objective function J at a local minimum $(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k)$ of* (3.10) *for the budget $p \in \mathbb{R}_+$ is*

$$J(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k) = \frac{\sigma_1^2(1-\rho_{1,l_1}^2)p}{m_1^2 \sum_{i=1}^{l_1-1} w_i}, \tag{A.19}$$

where $q \in \mathbb{N}$ and $\boldsymbol{l} = [l_1, \ldots, l_q]^T \in \mathbb{N}^q$ are defined as in Lemma A.1.

Proof. Use $r_{l_0}, \ldots, r_{l_q}$ and $\alpha_{l_1}, \ldots, \alpha_{l_q}$ as derived in Lemma A.1 and the objective as defined in (3.9) to obtain

$$J(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k) = \frac{\sigma_1^2}{m_1}\left(1 - \sum_{i=1}^{q}\left(\frac{1}{r_{l_{i-1}}} - \frac{1}{r_{l_i}}\right)\rho_{1,l_i}^2\right). \tag{A.20}$$

Note that $r_{l_i} = r_{l_i+1} = \cdots = r_{l_{i+1}-1}$ for $i = 0, \ldots, q$ and thus that the corresponding terms in the sum of the objective J evaluate to 0. Expanding the sum in (A.20) leads to

$$J(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k) = \frac{\sigma_1^2}{m_1}\left(1 - \left(\frac{\rho_{1,l_1}^2}{r_{l_0}} - \sum_{i=1}^{q-1}\frac{(\rho_{1,l_i}^2 - \rho_{1,l_{i+1}}^2)}{r_{l_i}} - \frac{\rho_{1,l_q}^2}{r_{l_q}}\right)\right). \tag{A.21}$$

We obtain for $i = 1, \ldots, q-1$ that

$$\begin{aligned}\frac{(\rho_{1,l_i}^2 - \rho_{1,l_{i+1}}^2)}{r_{l_i}} &= (\rho_{1,l_i}^2 - \rho_{1,l_{i+1}}^2)\underbrace{\sqrt{\frac{\left(\sum_{j=l_i}^{l_{i+1}-1} w_j\right)}{\left(\sum_{j=1}^{l_1-1} w_j\right)}\frac{(1-\rho_{1,l_1}^2)}{(\rho_{1,l_i}^2 - \rho_{1,l_{i+1}^2}^2)}}}_{\frac{1}{r_{l_i}}} \\ &= \frac{(1-\rho_{1,l_1}^2)\left(\sum_{j=l_i}^{l_{i+1}-1} w_j\right)}{\sum_{j=1}^{l_1-1} w_j}\underbrace{\sqrt{\frac{\left(\sum_{j=l_i}^{l_{i+1}-1} w_j\right)}{\left(\sum_{j=1}^{l_1-1} w_j\right)}\frac{(\rho_{1,i}^2 - \rho_{1,i+1}^2)}{(1-\rho_{1,l_1}^2)}}}_{r_{l_i}},\end{aligned}$$

which we use in (A.21) to derive

$$J(\boldsymbol{m}, \alpha_2, \ldots, \alpha_k) = \frac{\sigma_1^2}{m_1}\left(1 - \frac{\rho_{1,l_1}^2}{r_{l_0}} + \left(\frac{1-\rho_{1,l_1}^2}{\sum_{i=1}^{l_1-1} w_i}\right)\sum_{i=1}^{q-1}\sum_{j=l_i}^{l_{i+1}-1} w_j r_{l_i} + \frac{\rho_{1,l_q}^2}{r_{l_q}}\right) = \frac{\sigma_1^2(1-\rho_{1,l_1}^2)}{m_1 \sum_{i=1}^{l_1-1} w_i}\sum_{i=1}^{k} w_i r_i. \tag{A.22}$$

Use $\boldsymbol{w}^T\boldsymbol{r} = p/m_1$ (see Lemma A.1) to obtain (A.19). ∎

LEMMA A.3. *Consider the same setting as in Theorem* 3.4. *In particular, let* $\rho_{1,1}^2 > \cdots > \rho_{1,k}^2 > 0$ *and let the costs* $\boldsymbol{w} = [w_1, \ldots, w_k]^T \in \mathbb{R}_+^k$ *satisfy*

$$\frac{w_{i-1}}{w_i} > \frac{\rho_{1,i-1}^2 - \rho_{1,i}^2}{\rho_{1,i}^2 - \rho_{1,i+1}^2} \tag{A.23}$$

for $i = 2, \ldots, k$. *Let* $q \in \mathbb{N}$ *with* $q < k-1$ *and* $\boldsymbol{l} = [l_1, \ldots, l_q]^T \in \mathbb{N}^q$ *with* $1 < l_1 < \cdots < l_q < k+1$ *and set* $l_0 = 1$ *and* $l_{k+1} = k+1$. *The costs* $w_1, \ldots, w_k$ *and the correlation coefficients* $\rho_{1,1}, \ldots, \rho_{1,k}$ *satisfy the inequality*

$$\sum_{j=1}^{k} \sqrt{w_j(\rho_{1,j}^2 - \rho_{1,j+1}^2)} < \sum_{i=0}^{q} \sqrt{\left(\sum_{j=l_i}^{l_{i+1}-1} w_j\right)\left(\rho_{1,l_i}^2 - \rho_{1,l_{i+1}}^2\right)}. \tag{A.24}$$

Proof. Rewrite (A.24) as

$$\sum_{i=0}^{q}\sum_{j=l_i}^{l_{i+1}-1} \sqrt{w_j(\rho_{1,j}^2 - \rho_{1,j+1}^2)} < \sum_{i=0}^{q} \sqrt{\left(\sum_{j=l_i}^{l_{i+1}-1} w_j\right)\left(\rho_{1,l_i}^2 - \rho_{1,l_{i+1}}^2\right)}. \tag{A.25}$$

There exists an $i \in \{0, \ldots, q\}$ with $l_{i+1} - l_i > 1$ because $q < k-1$. The terms of the outer sum in (A.25) are nonnegative and therefore it is sufficient to show

$$\sum_{j=l_i}^{l_{i+1}-1} \sqrt{w_j(\rho_{1,j}^2 - \rho_{1,j+1}^2)} < \sqrt{\left(\sum_{j=l_i}^{l_{i+1}-1} w_j\right)\left(\rho_{1,l_i}^2 - \rho_{1,l_{i+1}}^2\right)}. \tag{A.26}$$

First consider the case $l_{i+1} - l_i = 2$, where we define for the sake of exposition $a = l_i$, $b = l_i + 1$, and $c = l_{i+1}$. The inequality of the arithmetic and the geometric mean leads to

$$\sqrt{w_a w_b(\rho_{1,a}^2 - \rho_{1,b}^2)(\rho_{1,b}^2 - \rho_{1,c}^2)} < \frac{w_a(\rho_{1,b}^2 - \rho_{1,c}^2) + w_b(\rho_{1,a}^2 - \rho_{1,b}^2)}{2}. \tag{A.27}$$

Note that there is a strict inequality in (A.27) because assumption (A.23) guarantees $w_b(\rho_{1,a}^2 - \rho_{1,b}^2) \neq w_a(\rho_{1,b}^2 - \rho_{1,c}^2)$ through

$$w_b(\rho_{1,a}^2 - \rho_{1,b}^2) < w_a(\rho_{1,b}^2 - \rho_{1,c}^2).$$

Adding the positive quantities $w_a(\rho_{1,a}^2 - \rho_{1,b}^2)$ and $w_b(\rho_{1,b}^2 - \rho_{1,c}^2)$ to both sides of (A.27) results in

$$\left(\sqrt{w_a(\rho_{1,a}^2 - \rho_{1,b}^2)} + \sqrt{w_b(\rho_{1,b}^2 - \rho_{1,c}^2)}\right)^2 < (w_a + w_b)(\rho_{1,a}^2 - \rho_{1,c}^2). \tag{A.28}$$

Taking the square root of (A.28) shows

$$\sqrt{w_a(\rho_{1,a}^2 - \rho_{1,b}^2)} + \sqrt{w_b(\rho_{1,b}^2 - \rho_{1,c}^2)} < \sqrt{(w_a + w_b)(\rho_{1,a}^2 - \rho_{1,c}^2)}, \tag{A.29}$$

which is (A.26) for the case $l_{i+1} - l_i = c - a = 2$.

We now show that the result for $l_{i+1} - l_i = 2$ can be extended to $l_{i+1} - l_i = 3$. Define $a = l_i, b = l_i + 1, c = l_i + 2$, and $d = l_{i+1}$ and use (A.29) to derive

$$\begin{aligned}\sqrt{w_a(\rho_{1,a}^2 - \rho_{1,b}^2)} + \sqrt{w_b(\rho_{1,b}^2 - \rho_{1,c}^2)} + \sqrt{w_c(\rho_{1,c}^2 - \rho_{1,d}^2)} \\ < \sqrt{(w_a + w_b)(\rho_{1,a}^2 - \rho_{1,c}^2)} + \sqrt{w_c(\rho_{1,c}^2 - \rho_{1,d}^2)}.\end{aligned} \tag{A.30}$$

Assumption (A.23) leads to

$$\frac{w_a + w_b}{w_c} = \frac{w_a}{w_c} + \frac{w_b}{w_c} > \frac{w_a}{w_b}\frac{w_b}{w_c} + \frac{w_b}{w_c} > \frac{\rho_{1,a}^2 - \rho_{1,b}^2}{\rho_{1,c}^2 - \rho_{1,d}^2} + \frac{\rho_{1,b}^2 - \rho_{1,c}^2}{\rho_{1,c}^2 - \rho_{1,d}^2} = \frac{\rho_{1,a}^2 - \rho_{1,c}^2}{\rho_{1,c}^2 - \rho_{1,d}^2},$$

and thus we obtain

$$\begin{aligned}\sqrt{w_a(\rho_{1,a}^2 - \rho_{1,b}^2)} + \sqrt{w_b(\rho_{1,b}^2 - \rho_{1,c}^2)} + \sqrt{w_c(\rho_{1,c}^2 - \rho_{1,d}^2)} \\ < \sqrt{(w_a + w_b + w_c)(\rho_{1,a}^2 - \rho_{1,d}^2)}\end{aligned}$$

with the same arguments as (A.29). Induction establishes the inequality (A.26) for the case $l_{i+1} - l_i > 3$. ∎

Acknowledgments. Some of the numerical examples were computed on the computer cluster of the Munich Centre of Advanced Computing. The authors thank Florian Augustin for very helpful discussions and comments.

REFERENCES

[1] N. M. Alexandrov, R. M. Lewis, C. R. Gumbert, L. L. Green, and P. A. Newman, *Approximation and model management in aerodynamic optimization with variable-fidelity models*, J. Aircraft, 38 (2001), pp. 1093–1101, doi:10.2514/2.2877.

[2] D. Allaire and K. Willcox, *A mathematical and computational framework for multifidelity design and analysis with computer models*, Int. J. Uncertain. Quantif., 4 (2014), pp. 1–20, doi:10.1615/Int.J.UncertaintyQuantification.2013004121.

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

[4] A. J. Booker, J. E. Dennis, Jr, P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset, *A rigorous framework for optimization of expensive functions by surrogates*, Struct. Optimization, 17 (1999), pp. 1–13.

[5] S. Boyaval, *A fast Monte-Carlo method with a reduced basis of control variates applied to uncertainty propagation and Bayesian estimation*, Comput. Methods Appl. Mech. Engrg., 241–244 (2012), pp. 190–205.

[6] S. Boyaval and T. Lelièvre, *A variance reduction method for parametrized stochastic differential equations using the reduced basis paradigm*, Commun. Math. Sci., 8 (2010), pp. 735–762.

[7] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2009.
[8] A. Brandt, *Multi-level adaptive solutions to boundary-value problems*, Math. Comp., 31 (1977), pp. 333–390.
[9] W. Briggs, V. E. Henson, and S. F. McCormick, *A Multigrid Tutorial*, SIAM, Philadelphia, 2000.
[10] H.-J. Bungartz and M. Griebel, *Sparse grids*, Acta Numer., 13 (2004), pp. 1–123.
[11] C.-C. Chang and C.-J. Lin, *LIBSVM: A library for support vector machines*, ACM Trans. Intelligent Systems Technology, 2 (2011), pp. 27:1–27:27.
[12] S. Chaturantabut and D. Sorensen, *Nonlinear model reduction via discrete empirical interpolation*, SIAM J. Sci. Comput., 32 (2010), pp. 2737–2764, http://dx.doi.org/10.1137/090766498.
[13] P. Chen and A. Quarteroni, *Accurate and efficient evaluation of failure probability for partial different equations with random input data*, Comput. Methods Appl. Mech. Engrg., 267 (2013), pp. 233–260.
[14] J. A. Christen and C. Fox, *Markov chain Monte Carlo using an approximation*, J. Comput. Graph. Statist., 14 (2005), pp. 795–810.
[15] K. A. Cliffe, M. Giles, R. Scheichl, and A. L. Teckentrup, *Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients*, Comput. Vis. Sci., 14 (2011), pp. 3–15.
[16] C. Cortes and V. Vapnik, *Support-vector networks*, Machine Learning, 20 (1995), pp. 273–297.
[17] T. Cui, Y. M. Marzouk, and K. E. Willcox, *Data-driven model reduction for the bayesian solution of inverse problems*, Internat. J. Numer. Methods Engrg., 102 (2015), pp. 966–990.
[18] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
[19] Y. Efendiev, T. Hou, and W. Luo, *Preconditioning markov chain Monte Carlo simulations using coarse-scale models*, SIAM J. Sci. Comput., 28 (2006), pp. 776–803.
[20] A. Ferreira, *MATLAB Codes for Finite Element Analysis*, Springer, New York, 2008.
[21] A. I. J. Forrester and A. J. Keane, *Recent advances in surrogate-based optimization*, Progr. Aerospace Sci., 45 (2009), pp. 50–79.
[22] M. Giles, *Multi-level Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.
[23] M. A. Grepl, Y. Maday, N. C. Nguyen, and A. T. Patera, *Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations*, ESAIM Math. Model. Numer. Anal., 41 (2007), pp. 575–605, http://dx.doi.org/10.1051/m2an:2007031.
[24] S. Gugercin and A. Antoulas, *A survey of model reduction by balanced truncation and some new results*, Internat. J. Control, 77 (2004), pp. 748–766.
[25] W. Hackbusch, *Multi-Grid Methods and Applications*, Springer, New York, 1985.
[26] R. F. Heinemann and A. B. Poore, *Multiplicity, stability, and oscillatory dynamics of the tubular reactor*, Chem. Engrg. Sci., 36 (1981), pp. 1411–1419.
[27] S. Heinrich, *Multilevel Monte Carlo methods*, in Large-Scale Scientific Computing, S. Margenov, J. Waśniewski, and P. Yalamov, eds., Lecture Notes in Comput. Sci. 2179, Springer, New York, 2001, pp. 58–67.
[28] C. Kirjner-Neto, E. Polak, and A. Der Kiureghian, *Algorithms for reliability-based optimal design*, in Reliability and Optimization of Structural Systems: Proceedings of the Sixth IFIP WG7.5 Working Conference on Reliability and Optimization of Structural Systems 1994, R. Rackwitz, G. Augusti, and A. Borri, eds., Springer, New York, 1995, pp. 144–152.
[29] N. Kuschel and R. Rackwitz, *Two basic problems in reliability-based structural optimization*, Math. Methods Oper. Res., 46 (1997), pp. 309–333.
[30] J. Li, J. Li, and D. Xiu, *An efficient surrogate-based method for computing rare failure probability*, J. Comput. Phys., 230 (2011), pp. 8683–8697.
[31] J. Li and D. Xiu, *Evaluation of failure probability via surrogate models*, J. Comput. Phys., 229 (2010), pp. 8966–8980.
[32] A. March and K. Willcox, *Provably convergent multifidelity optimization algorithm not requiring high-fidelity derivatives*, AIAA J., 50 (2012), pp. 1079–1089, http://arc.aiaa.org/doi/abs/10.2514/1.J051125.
[33] *MATLAB version* 8.6.0 (R2015b), The MathWorks, Inc., Natick, MA, 2015.
[34] B. L. Nelson, *On control variate estimators*, Comput. Oper. Res., 14 (1987), pp. 219–225.
[35] L. Ng, *Multifidelity Approaches for Design Under Uncertainty*, Ph.D. thesis, Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, 2013.
[36] L. Ng and K. Willcox, *Multifidelity approaches for optimization under uncertainty*, Internat. J. Numer. Methods Engrg., 100 (2014), pp. 746–772.
[37] L. Ng and K. Willcox, *Monte-Carlo information-reuse approach to aircraft conceptual design optimization under uncertainty*, J. Aircraft, (2015), pp. 1–12.

[38] B. Peherstorfer, T. Cui, Y. Marzouk, and K. Willcox, *Multifidelity importance sampling*, Comput. Methods Appl. Mech. Engrg., 300 (2016), pp. 490–509.
[39] B. Peherstorfer and K. Willcox, *Detecting and adapting to parameter changes for reduced models of dynamic data-driven application systems*, Procedia Comput. Sci., 51 (2015), pp. 2553–2562.
[40] B. Peherstorfer and K. Willcox, *Dynamic data-driven reduced-order models*, Comput. Methods Appl. Mech. Engrg., 291 (2015), pp. 2141.
[41] B. Peherstorfer and K. Willcox, *Online adaptive model reduction for nonlinear systems via low-rank updates*, SIAM J. Sci. Comput., 37 (2015), pp. A2123–A2150.
[42] C. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, New York, 2004.
[43] G. Rozza, D. Huynh, and A. Patera, *Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations*, Arch. Comput. Methods Engrg., 15 (2007), pp. 1–47.
[44] L. Sirovich, *Turbulence and the dynamics of coherent structures*, Quarter. Appl. Math., 45 (1987), pp. 561–571.
[45] A. L. Teckentrup, P. Jantsch, C. G. Webster, and M. Gunzburger, *A multilevel stochastic collocation method for partial differential equations with random input data*, SIAM/ASA J. Uncertainty Quant., 3 (2015), pp. 1046–1074.
[46] A. L. Teckentrup, R. Scheichl, M. Giles, and E. Ullmann, *Further analysis of multilevel Monte Carlo methods for elliptic PDEs with random coefficients*, Numer. Math., 125 (2013), pp. 569–600.
[47] B. Tracey, D. Wolpert, and J. J. Alonso, *Using supervised learning to improve Monte-Carlo integral estimation*, AIAA J., 51 (2013), pp. 2015–2023.
[48] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
[49] F. Vidal-Codina, N. Nguyen, M. Giles, and J. Peraire, *A model and variance reduction method for computing statistical outputs of stochastic elliptic partial differential equations*, J. Comput. Phys., 297 (2015), pp. 700–720.
[50] H. Yserentant, *On the multi-level splitting of finite element spaces*, Numer. Math., 49 (1986), pp. 379–412, http://dx.doi.org/10.1007/BF01389538.
[51] Y. B. Zhou, *Model reduction for nonlinear dynamical systems with parametric uncertainties*, M.S. thesis, Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, 2012.