SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY - INFORMATICS

TECHNICAL UNIVERSITY OF MUNICH

Master's Thesis in Informatics

Reduced-dimension Context-aware Multi-fidelity Monte Carlo Sampling

Julia Konrad

SCHOOL OF COMPUTATION, INFORMATION AND TECHNOLOGY - INFORMATICS

TECHNICAL UNIVERSITY OF MUNICH

Master's Thesis in Informatics

Reduced-dimension Context-aware Multi-fidelity Monte Carlo Sampling

Dimensionsreduziertes kontextsensitives Multi-fidelity Monte Carlo Sampling

Author:	Julia Konrad
Supervisor:	Univ.-Prof. Dr. Hans-Joachim Bungartz
Advisors:	Dr. rer. nat. Ionuț-Gabriel Farcaș, Dr. rer. nat. Tobias Neckel, Hon.-Prof. Dr. Frank Jenko
Submission Date:	January 15, 2023

I confirm that this master's thesis is my own work and I have documented all sources and material used.

Garching, January 15, 2023 Julia Konrad

Acknowledgments

First of all, I would like to thank Prof. Hans-Joachim Bungartz for his support and the opportunity to write this thesis at the Chair of Scientific Computing.

Secondly, a big thank you goes to my advisor Dr. Ionuț-Gabriel Farcaș for his guidance, advice, and patience, as well as for always making time to answer questions and discuss with me. I am very grateful for his mentorship over the past few years. Thank you so much.

I would also like to thank Dr. Tobias Neckel for his support regarding this thesis as well as other related projects I had the opportunity to be a part of during my studies at TUM.

Moreover, I would like to express my sincere gratitude towards Prof. Frank Jenko for taking an interest in my work and being supportive throughout it.

Finally, I would like to thank my family for all the moral support and encouragement over the last year.

Abstract

Multi-fidelity Monte Carlo sampling has proven to be an efficient method for quantifying uncertainty in applications with a large number of stochastic input parameters and computationally expensive models. The method consists of evaluating low-fidelity models in addition to the given high-fidelity model in order to speed up the computation of high-fidelity model statistics. In the regular multi-fidelity Monte Carlo sampling approach, low-fidelity models are static and cannot be changed. Context-aware multi-fidelity Monte Carlo sampling takes into account that e.g., data-driven low-fidelity models can be improved using evaluations of the high-fidelity model. This method trades off refining the low-fidelity models and sampling both types of models. In this thesis, we use sensitivity information to construct low-fidelity models that depend only on subsets of important input parameters in addition to a full-dimensional low-fidelity model. We explore the potential of such reduced-dimension low-fidelity models to further reduce the mean squared error of context-aware multi-fidelity Monte Carlo estimators. To this end, our method is used to perform uncertainty quantification in a scenario from plasma micro-turbulence simulation that models the suppression of turbulence by energetic particles, and for which quantifying uncertainty can be challenging using traditional approaches. In our experiments, the context-aware Monte Carlo algorithm with reduced-dimension low-fidelity models provided a speedup of two orders of magnitude as compared to standard Monte Carlo sampling, which corresponds to a reduction of the computational effort from 48 days to to six hours on 240 cores on parallel supercomputers.

Contents

1 Introduction

1.1 Motivation

Research towards the goal of the practical realization of nuclear fusion power plants as a viable energy source has been ongoing for multiple decades. Harnessing fusion power would supply clean and virtually unlimited electricity, thereby significantly reducing greenhouse gas emissions and paving the way towards global carbon neutrality. Most approaches to the design of fusion reactors make use of magnetically confined plasmas, in which the fusion reactions take place. A plasma is a hot, ionized gas, which, due to the electrically charged particles it contains, can be influenced by strong magnetic fields. Devices used for this purpose are for example tokamas or stellarators. The aim is to create a "burning", i.e., self-sustaining, plasma, in which the energy released by the fusion reactions heats the surrounding plasma enough in order to sustain the fusion process.

A significant challenge in maintaining this state for an extended period of time is the occurrence of micro-turbulence due to intrinsically steep temperature and density gradients. The resulting turbulent transport decreases the energy confinement time, which must, however, reach a certain value to enable the plasma to sustain itself. Therefore, a key step towards making nuclear fusion an attainable source of energy is to further our understanding of the turbulent transport and its prevention.

To this end, the predictive numerical simulation of the physical processes taking place in fusion plasmas plays an important role [14, 4, 31]. As for most real-world applications, the input parameters to the models used for such simulations are, however, inherently subject to uncertainty, e.g., due to measurement errors or even incomplete knowledge. For this reason, it is paramount that proper uncertainty quantification (UQ), see, e.g., [36], is conducted. In plasma micro-turbulence analysis, this is a challenging task since there is typically a large number of parameters that affect the turbulent flows. In addition to this, the underlying simulations are usually computationally expensive, which renders the task even more difficult.

In cases such as this, the number of suitable methods for performing UQ is restricted by the so-called curse of dimensionality, see, e.g., [5, 15]. The term describes the fact that the required computational effort grows exponentially with the number of input parameters if a full-grid approach is used to explore the space of uncertain inputs. This

fact does, however, not apply to sampling-based methods, for which the computational effort needed to increase the precision does not depend on the stochastic dimensionality of the problem.

The simplest method of this kind is standard Monte Carlo (MC) sampling [22], in which the underlying model is evaluated for (pseudo-)random samples that are subsequently used to compute a quantity such as an estimator for the mean or variance of the model output, or even sensitivity information, see [35] for the latter. In view of the characteristic slow convergence of the mean squared error (MSE) of MC estimators with regards to the sampling budget, this approach is typically not feasible, especially if the costs of evaluating the model are large.

Multi-fidelity Monte Carlo (MFMC) sampling [29, 28, 26, 32] is based on the control variate technique [25] and speeds up this sampling process by employing a hierarchy of so-called high- and low-fidelity models. Both kinds of models compute the same output of interest, however, the low-fidelity models achieve this with smaller computational costs than the high-fidelity model and are, in return, also less accurate. Here, the high-fidelity model computes our output of interest, while low-fidelity models provide a computationally less expensive approximation of the high-fidelity model. Such low-fidelity models can, for example, be data-driven, projection-based, or use reduced physics. Different types of low-fidelity models are discussed in [30].

The context-aware MFMC (CA-MFMC) sampling algorithm of [9, 11] takes into account the fact that, when low-fidelity models are not readily available for the use in MFMC, they must be especially constructed. To this end, they are often refined or trained using evaluations from the high-fidelity model, which means that an additional cost is incurred in order to obtain these samples. In this process, it is not immediately clear how far a low-fidelity model should be refined since increased accuracy usually also leads to increased evaluation costs, which in turn negatively affects the amount of samples that can be evaluated within a given computational budget. CA-MFMC addresses these points by balancing the costs of improving data-driven low-fidelity models with the high- and low-fidelity evaluation costs in the subsequent sampling process. The algorithm finds an optimal trade-off by computing the number of high-fidelity evaluations that should be used to train the low-fidelity models in order to minimize the MSE of the resulting estimators. When the low-fidelity models are refined accordingly, they are context-aware in the sense that they are constructed specifically to be used together with the high-fidelity model to reduce the variance of the MFMC estimators. The CA-MFMC approach has recently been shown to be very promising for making UQ at a large scale feasible in real-world applications, not least in the simulation of plasma turbulence [9, 11].

The goal of this thesis is to further study the potential of CA-MFMC for the efficient quantification of uncertainty in plasma turbulence simulations. Our key aim is

to explore the role that reduced-dimension low-fidelity models can play in further speeding up the CA-MFMC computations. To this end, we make use of that fact that in the simulation of many real-world applications, only a subset of the input parameters affects the output of the corresponding model significantly and construct low-fidelity models that depend only on such subsets.

1.2 Related Work

The CA-MFMC algorithm of [9, 11] generalizes a context-aware bi-fidelity MFMC sampling method that was introduced in [27]. This bi-fidelity approach considers the construction of a single, data-driven context-aware low-fidelity model for which the accuracy and evaluation costs are described by algebraic rate depending on the number of samples used to train the model. The work [41] analyses a bi-fidelity case, in which the context-aware low-fidelity models are obtained by a polynomial chaos approach. In [9, 11], the results from [27] are extended to models with more general rates. Additionally, a method of sequentially introducing low-fidelity models into the algorithm such that they are all context-aware is presented. The trade-off between refining and sampling low-fidelity models is also discussed in [1] in the context of importance sampling.

Regular, or static, MFMC has also been successfully used to perform UQ in plasma physics problems [9, 19]. Both works employed the gyrokinetic simulation code Gene [18, 16] as a high-fidelity model. The work [20] used MFMC to speed up computations of the energy confinement time in the design of stellarators.

A sensitivity-driven dimension-adaptive sparse grid interpolation approach of [10, 9] has also previously been applied to quantify uncertainty in plasma turbulence simulations. In [10], it was used to compute the expectation and variance of the corresponding Gene output, and, in particular, to perform sensitivity analysis. A parameter scan that allowed a subsequent dimensionality reduction of the stochastic input space was conducted in [13]. The work [12] employed the sparse grid algorithm to efficiently quantify uncertainty and compute sensitivity information in a nonlinear, and therefore very computationally expensive, Gene setup.

In [19], static MFMC was used to perform UQ in a plasma micro-turbulence scenario with 14 uncertain inputs and Gene as the high-fidelity model. Full- and reduced-dimension data-driven low-fidelity models were constructed based on sensitivity information in the form of Sobol' indices [37], which were obtained from the same dimension-adaptive sparse grid procedure as mentioned above. These computationally cheaper models were shown to be greatly effective in reducing the computational effort of computing MFMC mean estimators compared to standard MC. The low-fidelity models employed in that work were all obtained in a static manner, i.e., each one was

refined once using a fixed number of interpolation points, without regard to the other models in the hierarchy.

1.3 Our Contribution

In this thesis, we extend the static MFMC framework of [19] such that the employed low-fidelity models are additionally context-aware. We use the CA-MFMC algorithm of [9, 11] in conjunction with the sensitivity-driven interpolation approach of [10, 9] to sequentially introduce full- and reduced-dimension low-fidelity models into the hierarchy. In this way, for every added low-fidelity model, sensitivity information is used to decrease the stochastic dimensionality appropriately, which reduces the evaluation costs of the model but has only small effects on the accuracy. The obtained model is then refined using an optimal number of high-fidelity samples that minimizes the MSE of the resulting MFMC estimator.

We apply the proposed methodology to the numerical simulation of plasma micro-turbulence in order to study its efficiency in the quantification of uncertainty. To this end, we consider a realistic scenario that investigates the effect of energetic particles on the suppression of micro-turbulence. As a high-fidelity model, we employ the plasma turbulence code GENE to compute the growth rates of plasma micro-instabilities that lead to turbulent transport. In our experiments, we compute the MSEs of CA-MFMC estimators using full- and reduced-dimension models and compare them to static MFMC and standard MC estimators. From analytical computations of the expected MSEs, we find that, already for small computational budgets, CA-MFMC estimators are more accurate than the static MFMC estimators that we consider in this work. We confirm this for selected subsets of low-fidelity models and a given budget, where the CA-MFMC estimators are by orders of magnitude more efficient than standard MC. For larger budgets, our computations show that the speedup achieved by CA-MFMC estimators using certain combinations of low-fidelity models amounts to around two orders of magnitude, thereby providing a significant reduction of the computational costs.

1.4 Outline of This Thesis

The remainder of this document is organized as follows. Chapter 2 summarizes the mathematical modeling of plasma micro-turbulence and the standard MC and MFMC algorithms used in uncertainty propagation. Based on these two algorithms, Chapter 3 introduces the CA-MFMC approach and discusses the construction of context-aware reduced-dimension low-fidelity models. Chapter 4 presents details on the gyrokinetic code GENE, which we use as a high-fidelity model in our experiments. Further, the

chapter outlines the sensitivity-driven dimension-adaptive sparse grid interpolation technique that we use to construct full- and reduced-dimension low-fidelity models. Our numerical results are presented in Chapter 5. Therein, we apply CA-MFMC with reduced-dimension models to a scenario from plasma micro-turbulence analysis with 14 uncertain inputs, in which the suppression of turbulence by fast ions is modeled. Finally, we present a summary of the work conducted in this thesis and give an outlook on future research in Chapter 6.

2 Theoretical Background

This chapter provides a summary of standard MFMC sampling and of the real-world application considered in this work. Section 2.1 describes the application background in more detail and gives an overview of the modeling of plasma micro-turbulence using the so-called gyrokinetic approach. In Section 2.2 we summarize the standard MC and MFMC sampling algorithms on which the CA-MFMC approach employed in this work is based.

2.1 Numerical Simulation of Plasma Micro-Turbulence

In the following, we summarize the ideas behind the gyrokinetic modeling of plasma turbulence and outline its mathematical description using the gyrokinetic Maxwell-Vlasov system. We refer the reader to [4, 16, 31] for more details on the gyrokinetic approach.

2.1.1 The Gyrokinetic Approach

Gyrokinetic theory was first developed in the 1980s [16] and is a popular approach for describing plasma micro-turbulence. In gyrokinetics, the plasma is modeled in a five-dimensional phase space (three positions, two velocities). This is in contrast to the classical kinetic model in six dimensions (three positions, three velocities), which provides an appropriate description of the setting, however, computing a solution is not feasible with current-day computational resources [9].

The gyrokinetic approach makes use of that fact that, in strongly magnetized plasmas, the gyrofrequency of the charged particles, i.e., the frequency of their movement around the magnetic field lines, takes place at a much larger time scale than the turbulence. Gyrokinetics removes this high-frequency gyromotion from the model. This means that the exact position of the particle in the orbit around the field line is not resolved and particles are effectively modeled as moving rings in the magnetic field. Therefore, the phase space is reduced from six to five dimensions solely by omitting information that is not relevant to assessing the micro-turbulence and a computationally feasible description of the setting is obtained.

2.1.2 Gyrokinetic Maxwell-Vlasov Equations

The summary and notation presented in this section closely follows [9, 19]. In gyrokinetics, the evolution of the plasma over time is modeled by a distribution function $\pi_s(t, \mathbf{R}, v_\parallel, \mu_m)$ for each of the particle species s, i.e., a function which describes the number of particles at a time t for given five-dimensional gyrocenter coordinates. These coordinates consist of the position $\mathbf{R} = (R_x, R_y, R_z)$ of the gyrocenter in real space, the component $v_\parallel$ of its velocity parallel to the magnetic field lines and its magnetic moment $\mu_m := m_s v_\perp^2 / 2B$. Here, m_s denotes the mass of the particle species s, $v_\perp$ is the velocity component perpendicular to the magnetic field lines, and B describes the magnetic field. We assume that this distribution function can be decomposed as

$$\pi_s(t, \mathbf{R}, v_\parallel, \mu_m) = \pi_{s,0}(\mathbf{R}, v_\parallel, \mu_m) + \pi_{s,1}(t, \mathbf{R}, v_\parallel, \mu_m),$$

where only the non-static part $\pi_{s,1}$ depends on t in addition to the gyrocenter coordinates. For each particle species, the evolution of the distribution function is governed respectively by a nonlinear partial differential equation: the gyrokinetic Vlasov equation with collisions. It reads

$$\dot{\pi}_s + \dot{\mathbf{R}} \cdot \nabla \pi_s + \dot{v}_\parallel \frac{\partial \pi_s}{\partial v_\parallel} = C(\pi_s, \pi_{s'}), \tag{2.1}$$

where $C(\pi_s, \pi_{s'})$ is a collision operator. The time evolution of $\mathbf{R}$ and $v_\parallel$ is described by

$$\dot{\mathbf{R}} = v_\parallel b_0 + \frac{B_0}{B^*_{0,\parallel}} (\mathbf{v}_E + \mathbf{v}_{\nabla B} + \mathbf{v}_c), \quad \text{and} \tag{2.2}$$

$$\dot{v}_\parallel = -\frac{\dot{\mathbf{R}}}{m_s v_\parallel} \left(q_s \nabla \bar{\Phi}_1 + \frac{q_s}{c} \frac{\partial \bar{A}_{1,\parallel}}{\partial t} \mathbf{b}_0 + \mu_m \nabla (B_0 + \bar{B}_{1,\parallel}) \right), \tag{2.3}$$

where B_0 is the modulus of the magnetic field vector $\mathbf{B}_0$ and $\mathbf{b}_0 := \frac{\mathbf{B}_0}{B_0}$ is its normalized unit vector, $B^*_{0,\parallel} := \mathbf{b}_0 \mathbf{B}^*_0$ denotes the parallel component of $\mathbf{B}^*_0 := \mathbf{B}_0 + \nabla \times \left(\mathbf{B}_0 v_\parallel / \Omega_s \right)$ with $\Omega_s := q_s B_0 / m_s c$ being the gyrofrequency of the particle species s with charge q_s. Here, c is the speed of light. Furthermore, $\bar{\Phi}_1$ describes the gyroaveraged perturbed part of the electrostatic potential Φ, $\bar{A}_{1,\parallel}$ is the gyroaveraged modulus of the parallel component of the perturbed vector potential, and $\bar{B}_{1,\parallel}$ is the gyroaveraged modulus of the parallel component of the magnetic field perturbations. Additionally, $\mathbf{v}_E = \frac{c}{B_0^2} \mathbf{B}_0 \times \nabla \bar{\xi}_1$ describes the generalized $\mathbf{E} \times \mathbf{B}$ drift velocity, where $\bar{\xi}_1$ is the gyroaveraged modified potential $\bar{\xi}_1 = \bar{\Phi}_1 - \frac{v_\parallel}{c} \bar{A}_{1,\parallel}$, $\mathbf{v}_{\nabla B} = \frac{\mu_m c}{q_s B_0^2} \mathbf{B}_0 \times \nabla B_0$ is the gradient-B drift velocity, and $\mathbf{v}_c = \frac{v_\parallel^2}{\Omega_s} (\nabla \times \mathbf{b}_0)_\perp$ denotes the curvature drift velocity.

Additionally, Maxwell's equations are needed for a complete description of the plasma over time. Firstly, Φ_1 can be computed from the Poisson equation

$$-\nabla^2 \Phi_1 = 4\pi \sum_s q_s n_{1,s}, \tag{2.4}$$

and secondly, using Ampère's law, the quantities $A_{1,\parallel}$ and $B_{1,\parallel}$ are defined by

$$-\nabla^2 A_{1,\parallel} = \frac{4\pi}{c} \sum_s n_{1s} q_s u_{1s,\parallel} \quad \text{and} \quad (\nabla \times \mathbf{B}_1)_\perp = \frac{4\pi}{c} \sum_s n_{1s} \mathbf{u}_{1s,\perp}, \tag{2.5}$$

where, for particle species s, n_{1s} is the 0th space moment, $u_{1s,\parallel}$ is the 1st order velocity moment in $v_\parallel$ of π_s, and $\mathbf{u}_{1s,\perp}$ denotes the 1st order velocity moment in μ_m of π_s. Together with the gyrokinetic Vlasov equation (2.1), Maxwell's equations (2.4) and (2.5) define the system of equations required for a gyrokinetic model of fusion plasmas. We give specifics on the gyrokinetic simulation code employed in this work in Section 4.1.

2.2 Monte Carlo Sampling in Uncertainty Propagation

In forward UQ, or uncertainty propagation, the goal is to quantify the effect of uncertainties in the input parameters on the output of the underlying model. In this section, we summarize the mathematical basis of uncertainty propagation required for this work and describe how standard MC and MFMC are used to compute mean and variance estimators of the high-fidelity model output.

2.2.1 Preliminaries

We consider a high-fidelity model $f^{(0)} : \mathcal{X} \to \mathcal{Y}$, which maps from an input domain $\mathcal{X} \subset \mathbb{R}^d$ with dimensionality $d \in \mathbb{N}$ to the output domain $\mathcal{Y}$. In the application considered in this work, the output is scalar, i.e., $\mathcal{Y} \subset \mathbb{R}$. Uncertain inputs are modeled by a random vector $\mathbf{\Theta} : \Omega \to \mathcal{X}$ with probability space $(\Omega, \mathcal{A}, \mathcal{P})$, where Ω is the sample space, $\mathcal{A}$ is a σ-algebra, and $\mathcal{P} : \mathcal{A} \to [0, 1]$ is the corresponding probability measure. We denote a realization of the random vector $\mathbf{\Theta}$ by $\boldsymbol{\theta} := [\theta_1, \ldots, \theta_d]^T$ and its probability density function (pdf) by π. In this work, we assume that the input space has a product structure, i.e.,

$$\mathcal{X} = \bigotimes_{i=1}^{d} \mathcal{X}_i,$$

where $\mathcal{X}_i \subseteq \mathbb{R}, i = 1, \ldots, d$. Similarly, we make the assumption for the pdf of $\boldsymbol{\Theta}$ that

$$\pi(\boldsymbol{\theta}) = \prod_{i=1}^{d} \pi_i(\theta_i),$$

where $\pi_i, i = 1, \ldots, d$, is the marginal pdf corresponding to θ_i. In particular, this implies that the uncertain parameters $\theta_1, \ldots, \theta_d$ must be independent. Moreover, the mean or expectation of $f^{(0)}(\boldsymbol{\Theta})$ is

$$\mu_0 \coloneqq \mathbb{E}[f^{(0)}(\boldsymbol{\Theta})] = \int_{\mathcal{X}} f^{(0)}(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}. \tag{2.6}$$

Finally, we want to note that MC sampling of any type is always an outer-loop application [29]. By this, we mean that any involved models are treated as black boxes, where the actual implementation is irrelevant to the problem at hand and only the input-output mapping provided by the model is of interest. The models are evaluated for samples of the input pdf and the obtained evaluations are then used by the outer-loop application to compute some overall output of interest. In uncertainty propagation, these quantities are the statistics of the high-fidelity model output $f^{(0)}(\boldsymbol{\Theta})$, such as mean and variance. In this way, given a sample of the input distribution, the problem becomes deterministic in that one or multiple models are evaluated for fixed inputs. A schematic of uncertainty propagation as an outer-loop application is depicted in Figure 2.1.

2.2.2 Standard Monte Carlo Sampling

In order to compute unbiased statistics of the high-fidelity model output using standard MC sampling [22], the high-fidelity model $f^{(0)}$ is evaluated for a number of samples $m \in \mathbb{N}$ of the input probability distribution. The obtained model evaluations are then used to compute an estimator for the mean of the high-fidelity model output as

$$\hat{E}^{MC} \coloneqq \bar{E}_m^{(0)} = \frac{1}{m} \sum_{i=1}^{m} f^{(0)}(\boldsymbol{\theta}_i) \tag{2.7}$$

where $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$ are m independent and identically distributed (i.i.d.) samples of the input random vector $\boldsymbol{\Theta}$.

For unbiased estimators, the mean squared error (MSE) is equal to its variance [21]. Let $w_0 > 0$ be the evaluation costs of the high-fidelity model $f^{(0)}$. Then, for a

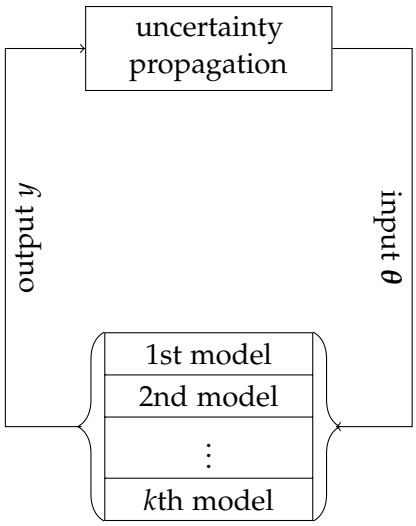

Figure 2.1: In uncertainty propagation, one or multiple models are evaluated for samples of the input random vector. The obtained evaluations are subsequently used to compute the quantity of interest, such as mean and variance of the high-fidelity model output (source: [30]).

computational budget p, the MSE of the standard MC mean estimator is

$$\mathrm{MSE}[\bar{E}_m^{(0)}] = \frac{\sigma_0^2}{m} = \frac{\sigma_0^2}{p} w_0. \tag{2.8}$$

Since $f^{(0)}$ is evaluated at m samples to compute $\bar{E}_m^{(0)}$ and $\overline{\mathrm{Var}}_m^{(0)}$, the cost of computing the standard MC estimator (2.7) is $m \times w_0$. Considering that the convergence of the MSE (2.8) is in $\mathcal{O}(p^{-1})$, these costs are often prohibitive to obtaining an accurate result using standard MC since an unfeasibly large number of – usually computationally expensive – high-fidelity model evaluations may be necessary. Note that, in the sense of traditional model reduction, standard MC sampling could also be applied using a reduced model with lower evaluation costs in place of the high-fidelity model; this would however introduce a bias to the resulting estimators and therefore, require a reduced model with a high accuracy guarantee, which may not always be available.

2.2.3 Multi-Fidelity Monte Carlo Sampling

We have discussed that increasing the number of MC samples m in order to obtain more accurate standard MC estimators is not always possible, especially if the high-fidelity model has large evaluation costs. If m cannot be increased, a different option for

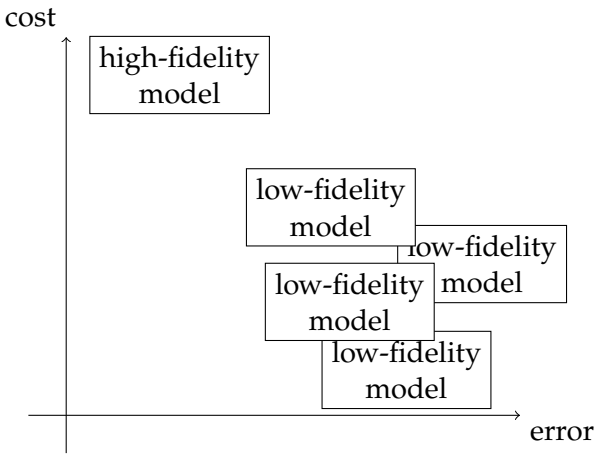

Figure 2.2: Relationship between high- and low-fidelity models. The high-fidelity model is characterized by its small error and high evaluation costs, while low-fidelity models typically exhibit higher errors but are also less expensive to evaluate (source: [30]).

reducing the MSE of standard MC estimators is to decrease the numerator of (2.8), i.e., the term σ_0^2. Using control variates [25], this can be achieved by including in the computation of the estimators additional random variables that are correlated with the estimated quantity, i.e., in our case, with the high-fidelity mean. In MFMC sampling [29, 26, 32], standard MC estimators for the mean of the low-fidelity models are used for this purpose.

Therefore, assume that in addition to the high-fidelity model $f^{(0)}$, we have k low-fidelity models $f^{(1)}, \dots, f^{(k)}$. Compared to the high-fidelity model, these typically are much cheaper to evaluate, but are also less accurate. This trade-off is visualized in Figure 2.2. The smaller evaluation costs are leveraged in MFMC, where they allow for more model evaluations within a given computational budget. At the same time, the accuracy and unbiasedness of the resulting estimators is still guaranteed by keeping the high-fidelity model in the loop.

The low-fidelity models in MFMC are characterized in terms of their evaluation costs $w_j, j = 1, \dots, k$, and their Pearson correlation coefficients ρ_j, $j = 1, \dots, k$, with respect to the high-fidelity model $f^{(0)}$. The correlation coefficient quantifies the linear correlation between the two models and is defined as

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

where $\sigma_j^2 := \mathrm{Var}[f^{(j)}(\boldsymbol{\Theta})]$, $j = 1, \ldots, k$ is the variance of the low-fidelity model $f^{(j)}$. The evaluation costs, correlation coefficients as well as the variances of the low-fidelity models act as inputs to the MFMC algorithm and are used to assign the amount of evaluations performed by each of the models. When they are not known beforehand, they can be estimated numerically using standard MC sampling. An estimator for the correlation coefficient ρ_j is given by

$$\bar{\rho}_j = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{f^{(0)}(\boldsymbol{\theta}_i) - \bar{E}_n^{(0)}}{\sigma_0} \right) \left(\frac{f^{(j)}(\boldsymbol{\theta}_i) - \bar{E}_n^{(j)}}{\sigma_j} \right), \quad j = 1, \ldots, k. \tag{2.9}$$

MFMC estimators are computed as follows. Define the number of model evaluations of each of the models $f^{(0)}, f^{(1)}, \ldots, f^{(k)}$ as

$$1 \leq m_0 \leq m_1 \leq \ldots \leq m_k.$$

Here, we need to assure that the computational budget is large enough such that at least one high-fidelity evaluation is possible. Otherwise, the resulting estimators will be biased as discussed above. Draw m_k i.i.d. samples $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{m_k}$ of $\boldsymbol{\Theta}$ according to the input pdf and evaluate model $f^{(j)}, j = 0, 1, \ldots, k$, at the first m_j samples, i.e., compute

$$f^{(j)}(\boldsymbol{\theta}_1), \ldots, f^{(j)}(\boldsymbol{\theta}_{m_j}), \quad j = 0, 1, \ldots, k.$$

From these evaluations, compute for the high- and low-fidelity models the standard MC estimators

$$\bar{E}_{m_j}^{(j)} = \frac{1}{m_j} \sum_{i=1}^{m_j} f^{(j)}(\boldsymbol{\theta}_i), \quad j = 0, 1, \ldots, k. \tag{2.10}$$

For the low-fidelity models, compute the estimators

$$\bar{E}_{m_{j-1}}^{(j)} = \frac{1}{m_{j-1}} \sum_{i=1}^{m_{j-1}} f^{(j)}(\boldsymbol{\theta}_i), \quad j = 0, 1, \ldots, k, \tag{2.11}$$

which are obtained from a subset of the model evaluations used in (2.10), which makes the two estimators in (2.10) and (2.11) dependent. The unbiased MFMC mean estimator is computed as

$$\hat{E}^{\mathrm{MFMC}} := \bar{E}_{m_0}^{(0)} + \sum_{j=1}^{k} \alpha_j (\bar{E}_{m_j}^{(j)} - \bar{E}_{m_{j-1}}^{((j)}), \tag{2.12}$$

where we use coefficients $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$. These coefficients as well as the number of model evaluations $m_0, m_1, \ldots, m_k$ are chosen such that the MSE of the resulting mean

estimator is minimized. This is achieved as follows. The low-fidelity models must be ordered according to their correlation coefficients, i.e.,

$$1 =: \rho_0 > |\rho_1| > \ldots > |\rho_k|, \tag{2.13}$$

and must additionally satisfy

$$\frac{w_{j-1}}{w_j} > \frac{\rho_{j-1}^2 - \rho_j^2}{\rho_j^2 - \rho_{j+1}^2}, \quad j = 1, \ldots k,$$

where $\rho_{k+1} = 0$. Then, the coefficients $\alpha_1^*, \ldots, \alpha_k^*$ that minimize the MSE are $\alpha_j^* = \rho_j$. Further, define

$$r_j^* = \sqrt{\frac{w_0(\rho_j^2 - \rho_{j+1}^2)}{w_j(1 - \rho_j^2)}}, \quad j = 0, 1, \ldots, k,$$

and let $\mathbf{r}^* := [r_0, r_1^*, \ldots, r_k^*]$ and $\mathbf{w} := [w_0, w_1, \ldots, w_k]$. The optimal number of model evaluations $m_0^*, m_1^* \ldots, m_k^*$ are computed as

$$m_0^* = \frac{p}{\mathbf{w}^T \mathbf{r}^*}, \quad m_j^* = r_j^* m_0^*, \quad j = 1, \ldots, k.$$

With this definition, we generally have that $m_j^* \in \mathbb{R}, j = 1, \ldots, k$. In order to obtain an integer number of model evaluations, we round down the computed values, i.e., we use $\lfloor m_j^* \rfloor$, $j = 0, 1, \ldots, k$, in the MFMC algorithm to ensure that the prescribed computational budget is not exceeded.

In this work, we focus on estimating the mean of the high-fidelity output. We note that the variance can be estimated in the same way as the mean by replacing the standard MC mean estimators by ones for the variance. In this case, the values given above for $m_0^*, m_1^*, \ldots, m_k^*$ and $\alpha_1^*, \ldots, \alpha_k^*$ do not minimize the MSE of the variance estimator; in practice, they do, however, provide accurate results and can be used to estimate the variance as well. We refer to [32] for a more in-depth discussion.

The MSE of the MFMC mean estimator $\hat{E}^{\text{MFMC}}$ which uses the optimal coefficients $\alpha_1^*, \ldots, \alpha_k^*$ and number of model evaluations $m_0^*, m_1^*, \ldots, m_k^*$ is

$$\text{MSE}[\hat{E}^{\text{MFMC}}] = \frac{\sigma_0^2}{p} \left(\sum_{j=1}^{k} \sqrt{w_j(\rho_j^2 - \rho_{j+1}^2)} \right)^2. \tag{2.14}$$

For a given budget p, this MSE is smaller than MSE of the standard MC mean estimator

$\hat{E}^{\mathrm{MC}}$ from (2.8) if and only if

$$\frac{\mathrm{MSE}[\hat{E}^{\mathrm{MFMC}}]}{\mathrm{MSE}[\hat{E}^{\mathrm{MC}}]} = \left(\sum_{j=1}^{k} \sqrt{\frac{w_j}{w_0}(\rho_j^2 - \rho_{j+1}^2)} \right)^2 < 1. \tag{2.15}$$

Given a set of low-fidelity models alongside the high-fidelity model, the model selection algorithm in [29] uses (2.15) to select a subset of low-fidelity models that, when used in MFMC, will lead to the MFMC mean estimator with the smallest possible MSE for the given set of models. Note that in this procedure, the low-fidelity models are assumed to exist with fixed correlation coefficients and evaluation costs and at this point, are not refined or adapted further.

The costs of computing the MFMC estimators (2.12) and (**??**) are $\sum_{i=0}^{k} m_i w_i$ since the whole computational budget is used on sampling the models. Here, the construction of the low-fidelity models in not taken into account, which can however be highly relevant for, e.g., data-driven models which require a high number of high-fidelity evaluations for their refinement. These aspects are addressed by the CA-MFMC algorithm, which is summarized in the following chapter.

3 Context-Aware Multi-Fidelity Monte Carlo Sampling

This chapter presents the CA-MFMC sampling algorithm of [9, 11], which will be extensively used in our numerical results. Section 3.1 introduces notation and formalizes the relation between the number of training samples and the accuracy and evaluation costs of low-fidelity models. In Section 3.2, we first describe CA-MFMC in the bi-fidelity case, i.e., if only one low-fidelity model is employed, and then summarize the use of multiple low-fidelity models with the algorithm. In our presentation, we follow mainly the notation used in [11].

3.1 Assumptions

Firstly, note that in the description of CA-MFMC, we assume w.l.o.g. that the evaluation costs of the models are normalized w.r.t. to the high-fidelity evaluation costs w_0 and formulate the algorithm in terms of $\bar{w}_0 := 1$ and $\bar{w}_j := \frac{w_j}{w_0}$, $j = 1, \dots, k$. In the same manner, we normalize the available computational budget p w.r.t. w_0 and view it as the number of high-fidelity evaluations that can be obtained within the actual budget p, i.e., we use $\bar{p} := \lfloor \frac{p}{w_0} \rfloor$, where we round down in order not to exceed p with any of the following computations.

Further, in CA-MFMC, we use low-fidelity models which depend on the number of high-fidelity evaluations that are used to improve their approximation quality. These high-fidelity samples can for example describe interpolation points or training pairs for neural networks. In contrast to static MFMC, where the correlation coefficient and evaluation costs of a low-fidelity model are fixed quantities, they are now determined by the number of high-fidelity samples used to refine the model.

We denote as $f_{n_j}^{(j)}$ the jth low-fidelity model in the hierarchy which is constructed using n_j high-fidelity evaluations. This low-fidelity model has correlation coefficient $\rho_j(n_j)$ w.r.t. the high-fidelity model $f^{(0)}$ and (normalized) evaluation costs $\bar{w}_j(n_j)$, which both depend on n_j. Assumption 1 formalizes the dependence of the correlation coefficient on the number of high-fidelity samples n_j.

Assumption 1 *For every low-fidelity model $f_{n_j}^{(j)}$, $j = 1, \dots, k$, the accuracy $1 - \rho_j^2(n_j)$ can be bounded from above as*

$$1 - \rho_j^2(n_j) \leq c_{a,j} r_{a,j}(n_j) \tag{3.1}$$

where $c_{a,j} \in \mathbb{R}^+$, and $r_{a,j} : (0, \infty) \to (0, \infty)$ is decreasing and at least twice continuously differentiable.

A similar assumption is made for the evaluation costs of the low-fidelity models in Assumption 2.

Assumption 2 *For every low-fidelity model $f_{n_j}^{(j)}$, $j = 1, \dots, k$, the normalized evaluation costs $\bar{w}_j(n_j)$ can be bounded from above as*

$$\bar{w}_j(n_j) \leq c_{c,j} r_{c,j}(n_j) \tag{3.2}$$

where $c_{c,j} \in \mathbb{R}^+$, and $r_{c,j} : (0, \infty) \to (0, \infty)$ is increasing and at least twice continuously differentiable.

The accuracy and cost rates from Assumptions 1 and 2 must be estimated numerically from pilot runs via regression if they are not available analytically (as they would be, e.g., if the low-fidelity model interpolates the solution of an elliptic partial differential equation [9]). If there are at least two evaluations of the high- and low-fidelity model available, those can be used in importance sampling [2] to compute the rates.

Remark 1 *High- and low-fidelity evaluations used to compute correlation coefficients and evaluations costs in the regression step can be re-used in the subsequent MFMC computations. This was shown in [19], where model evaluations computed to estimate correlation coefficients and evaluation costs for regular MFMC were also used in the sampling process. A comparison to MFMC with an independent set of samples showed that the bias introduced by this re-use was negligible. Alternatively, the costs incurred by this preprocessing step can also be taken into account in CA-MFMC by subtracting them from the total computational budget.*

3.2 Context-Aware Multi-Fidelity Sampling Algorithm

In this section, we summarize the CA-MFMC algorithm and show how to refine the low-fidelity models such that every model is designed explicitly to be used together with the previous models in the hierarchy and so, is context-aware. Once the low-fidelity models have been constructed accordingly, MFMC sampling as discussed in Section 2.2.3 can be applied.

3.2.1 One Low-Fidelity Model

We first study the bi-fidelity setting, in which only one low-fidelity model is used together with the high-fidelity model. The goal of CA-MFMC is to determine how many high-fidelity evaluations should be employed in the refinement of this low-fidelity model in order for the resulting MFMC estimator to be as accurate as possible.

To this end, we examine the MSE (2.14) of the bi-fidelity CA-MFMC mean estimator $\hat{E}_{n_1}^{\text{CA-MFMC}}$, in which n_1 high-fidelity training samples are used to refine the low-fidelity model. Since the correlation coefficient and evaluation costs of $f_{n_1}^{(1)}$ now depend on n_1, this is true for the MSE as well. In order to optimally balance the training of the low-fidelity model with the cost of the model evaluations for MFMC, CA-MFMC must additionally take n_1 into account when considering the computational budget available for MFMC sampling. After $f_{n_1}^{(1)}$ has been refined using n_1 high-fidelity evaluations, the remaining budget is $\bar{p} - n_1$. The corresponding MSE expressed in terms of the normalized evaluation costs reads

$$\text{MSE}[\hat{E}_{n_1}^{\text{CA-MFMC}}] = \frac{\sigma_0^2}{\bar{p} - n_1} \left(\sqrt{1 - \rho_1^2(n_1)} + \sqrt{\bar{w}_1(n_1)\, \rho_1^2(n_1)} \right). \tag{3.3}$$

In [11], an upper bound on (3.3), which depends on the accuracy and cost rates of $f_{n_1}^{(1)}$ defined in Assumptions 1 and 2, is derived as

$$\text{MSE}[\hat{E}_{n_1}^{\text{CA-MFMC}}] \leq \frac{2\sigma_0^2}{\bar{p} - n_1} (c_{a,1} r_{a,1}(n_1) + c_{c,1} r_{c,1}(n_1)) \tag{3.4}$$

The explicit dependence of (3.4) on the rates, and therefore also n_1, allows us to perform a minimization of the upper bound on the MSE in order to compute an optimal number of high-fidelity training samples n_1^*. Thus, we solve

$$\begin{aligned} \underset{n_1 \in \mathbb{N}}{\arg\min} \quad & \frac{1}{\bar{p} - n_1} (c_{a,1} r_{a,1}(n_1) + c_{c,1} r_{c,1}(n_1)) \\ \text{s.t.} \quad & 1 \leq n_1 \leq \bar{p} - 1. \end{aligned} \tag{3.5}$$

We restrict n_1 to be less or equal to $\bar{p} - 1$ in order to guarantee that at least one high-fidelity evaluation is possible in MFMC sampling, meaning the resulting MFMC estimator will be unbiased. Since (3.5) constitutes an integer program, we relax it and find a solution $\tilde{n}_1^* \in [1, \bar{p} - 1]$, which we then round up to $n_1^* := \lceil \tilde{n}_1^* \rceil \in \mathbb{N}$.

Theorem 1 presents a condition derived in [11] under which the solution of (3.5) is a unique global minimizer.

Theorem 1 *If the functions $r_{a,1}$ and $r_{c,1}$ satisfy Assumptions 1 and 2, and if additionally*

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

holds for all $n_1 \in [1, \bar{p} - 1]$, then the objective function in (3.5) *has a unique global minimum with minimizer $n_1^* \in [1, \bar{p} - 1]$.*

It follows that as long as the prerequisites to Theorem 1 are fulfilled, there exists an optimal trade-off between refining the low-fidelity model and using the remaining computational budget for MFMC sampling.

Remark 2 *The condition* (3.6) *holds if $r_{a,1}$ is strictly convex and $r_{c,1}$ is convex.*

Another property of the solution to the minimization problem (3.5) is that the resulting optimal number n_1^* of high-fidelity evaluations used to construct $f_{n_1}^{(1)}$ is bounded from above for $p \to \infty$. We state this in the following theorem, the proof of which can be found in [11].

Theorem 2 *If Theorem 1 applies with* (3.6) *satisfied for all $n_1 \in (0, \infty)$, and if there exists $\bar{n}_1 \in (0, \infty)$ such that*

$$c_{a,1} r'_{a,1}(\bar{n}_1) + c_{c,1} r'_{c,1}(\bar{n}_1) = 0, \tag{3.7}$$

then $\bar{n}_1$ is unique and the minimizer n_1^* of the objective function in (3.5) *is bounded from above by $n_1^* \leq \max\{1, \bar{n}_1\}$ independent of the computational budget.*

This means that even for an infinite budget, the optimal number of samples used to train the low-fidelity model is finite. Therefore, from a certain computational budget onwards, refining the low-fidelity model any further only leads to an increase in the MSE of the MFMC mean estimator and the remaining budget should instead be used to sample the models. Together with the fact that, according to Theorem 1, there exists a fixed optimal n_1^* for a given budget, this shows an important contrast to standard model reduction techniques. Therein, the goal is typically to find reduced models that approximate the high-fidelity model as closely as possible. In MFMC, however, the increased evaluation costs of a very accurate low-fidelity model can make such a model not beneficial for computing estimators that are as accurate as possible. Instead, the low-fidelity trained on n_1^* high-fidelity samples is context-aware in that it is constructed specifically to be used together with the high-fidelity model in order to minimize the MSE of MFMC estimators.

Finally, we discuss the rate with which the MSE for CA-MFMC converges. For the regular MFMC mean estimator (2.12), the MSE (2.14) decays with a rate in $\mathcal{O}(p^{-1})$. For CA-MFMC with one low-fidelity model and algebraic accuracy and cost rates, it was shown in [27] that if $\bar{w}_1 \ll 1$, the convergence of the MSE is faster than $\mathcal{O}(p^{-1})$

in a pre-asymptotic regime, meaning for small computational budgets p. A proof is given in [27] for the limit case $\bar{w}_1 = 0$, where the convergence is in $\mathcal{O}(p^{-1-\alpha_1})$ for $r_{a,1}(n_1) = n_1^{-\alpha_1}$, $\alpha_1 > 0$. On the other hand, once the limit described in Theorem 2 is reached, the convergence can also be slower than $\mathcal{O}(p^{-1})$ if n_1 is chosen larger than the upper bound that exists for n_1^* and $p \to \infty$.

3.2.2 Multiple Low-Fidelity Models

We now discuss the extension of the CA-MFMC algorithm to MFMC with $k > 1$ low-fidelity models. The task here is to find the optimal number of high-fidelity evaluations $n_1^*, \dots, n_k^*$ that should be used to construct the low-fidelity models $f_{n_1}^{(1)}, \dots, f_{n_k}^{(k)}$ in order to minimize the MSE of the corresponding MFMC mean estimator and thus, to achieve an optimal balance between refining and sampling the low-fidelity models.

The MSE of a CA-MFMC mean estimator $\hat{E}_{n_1,\dots,n_k}^{\text{CA-MFMC}}$ which uses multiple low-fidelity models can be bounded as

$$\text{MSE}[\hat{E}_{n_1,\dots,n_k}^{\text{CA-MFMC}}] \leq \frac{(k+1)\sigma_0^2}{\bar{p}_{k-1} - n_k}(\kappa_{k-1} + \bar{w}_{k-1}(n_{k-1})c_{a,k}r_{a,k}(n_k) + c_{c,k}r_{c,k}(n_k)), \tag{3.8}$$

where $\bar{p}_{k-1}$ is the number of remaining high-fidelity evaluations after the first $k-1$ low-fidelity models have been refined, i.e., $\bar{p}_{k-1} = \bar{p} - \sum_{j=1}^{k-1} n_j$ and $\bar{p}_0 := \bar{p}$, and with $\bar{w}_0(n_0) = 1$, we have

$$\kappa_{k-1} = \sum_{j=0}^{k-2} w_j(n_j)(1 - \rho_{j+1}^2(n_{j+1})), \quad \kappa_0 = 0. \tag{3.9}$$

For $k = 1$, the MSE (3.8) corresponds to the MSE (3.3) in the bi-fidelity case. When we have $k > 1$, the MSE depends on the correlation coefficients $\rho_1(n_1), \dots, \rho_{k-1}(n_{k-1})$ and evaluations costs $\bar{w}_1(n_1), \dots, \bar{w}_{k-1}(n_{k-1})$ of the first $k-1$ low-fidelity models, which are determined by the number of training samples $n_1, \dots, n_{k-1}$. The CA-MFMC algorithm uses this fact to iteratively add the low-fidelity models to the hierarchy. In the first iteration, the first low-fidelity model is added as described in Section 3.2.1. In every iteration $j > 1$, we add the low-fidelity model $f_{n_j}^{(j)}$ and compute the corresponding optimal n_j^*. Since at this point, $n_1^*, \dots, n_{j-1}^*$ have already been computed, they are used to obtain the correlation coefficients and evaluation costs of the previous models using the accuracy and cost rate functions in Assumptions 1 and 2. Those values can then be inserted into the MSE (3.8). In this way, the minimization problem solved to compute

n_j^* is again one-dimensional:

$$\begin{aligned} \underset{n_j \in \mathbb{N}}{\arg\min} \quad & \frac{1}{\bar{p}_{j-1} - n_j} (\kappa_{j-1} + \bar{w}_{j-1}(n_{j-1}) c_{a,j} r_{a,j}(n_j) + c_{c,j} r_{c,j}(n_j)) \\ \text{s.t.} \quad & 1 \leq n_j \leq \bar{p}_{j-1} - 1. \end{aligned} \tag{3.10}$$

We again relax (3.10) and find a solution $\tilde{n}_j^* \in \mathbb{R}$, which is then up to obtain $n_j^* := \lceil \tilde{n}_j^* \rceil$. For subsequent iterations, it is assumed that the previous $n_1^*, \ldots, n_j^*$ are integers.

It was shown in [11] that in iteration j, a unique global minimum to the one-dimensional optimization problem (3.10) exists under similar circumstances as in the case with a single low-fidelity model:

Theorem 3 *If the functions $r_{a,j}$ and $r_{c,j}$ satisfy Assumptions 1 and 2, and if additionally, in iteration j,*

$$\bar{w}_{j-1}(n_{j-1}) c_{a,j} r_{a,j}''(n_j) + c_{c,j} r_{c,j}''(n_j) > 0 \tag{3.11}$$

holds for all $n_j \in [1, \bar{p}_{j-1} - 1]$, then the objective function in (3.10) *has a unique global minimum with minimizer $n_j^* \in [1, \bar{p}_{j-1} - 1]$.*

In the same manner, the $n_1^*, \ldots, n_k^*$ are bounded from above for $p \to \infty$ as well, as the following theorem shows.

Theorem 4 *If, in iteration j, Theorem 3 applies with* (3.11) *satisfied for all $n_j \in (0, \infty)$, and if there exists $\bar{n}_j \in (0, \infty)$ such that*

$$\bar{w}_{j-1}(n_{j-1}) c_{a,j} r_{a,j}'(\bar{n}_j) + c_{c,j} r_{c,j}'(\bar{n}_j) = 0, \tag{3.12}$$

then there exists an upper bound $n_j^* \leq \bar{n}_j^*$ with $\bar{n}_j^* \in (0, \infty)$ independent of the computational budget.

Moreover, Remark 2 can be applied iteratively to give a condition under which the previous two theorems hold.

Remark 3 *Since the low-fidelity models are added sequentially, the objective function in each iteration $j > 1$ only differs by constant factors from the case with one low-fidelity model. Therefore, Theorems 3 and 4 also hold if $r_{a,j}$ is strictly convex and $r_{a,j}$ is convex.*

As in static MFMC, the low-fidelity models must be ordered according to their correlation coefficients as in (2.13). If this condition is not satisfied by the $\rho_j(n_j^*), j = 1, \ldots, k$, obtained by sequentially adding the low-fidelity models to the CA-MFMC algorithm, we adjust the order in which the low-fidelity models are added. This changes the order in which the n_j^* are computed and will therefore result in changed correlation coefficients for the individual models as well.

3.3 Reduced-Dimension Context-Aware Low-Fidelity Models

We now introduce the methodological novelty of this work, which is the construction of reduced-dimension context-aware low-fidelity models for MFMC. To the best of our knowledge, such models have not yet been considered in other works.

In order to be able reduce the stochastic dimensionality in a meaningful way, knowledge about the input sensitivities is required. In MFMC, this can often mean first constructing a full-dimensional low-fidelity model and using the information obtained from that process to determine the most important stochastic parameters. Following that, reduced-dimension models depending only on subsets of important parameters can be constructed.

We build on the work in [19], where the described approach was used to find reduced-dimension low-fidelity models, which were subsequently used in static MFMC sampling. The idea can be taken one step further by considering reduced-dimension low-fidelity models which are also context-aware. Finding reduced-dimension models for MFMC is a sequential process, where the dimensionality reduction can only take place after some initial computations, i.e., after the low-fidelity model which depends on all d inputs has been found. Afterwards, reduced-dimension models can be obtained successively, where in each step, the dimensionality is reduced further. This process can be carried out for as long as the obtained models are still useful for variance reduction in MFMC, meaning as long as only insignificant input parameters are disregarded such that the resulting model is still a good approximation of the high-fidelity model.

In the same manner, the CA-MFMC algorithm sequentially constructs and refines low-fidelity models based on the properties of the previous models in the hierarchy. Therefore, it is straightforward to connect the two approaches and make the low-fidelity models with reduced dimensionality also context-aware. That is, we first find the full-dimensional low-fidelity model in a context-aware manner and use it to determine sensitivity information about the inputs, based on which we then sequentially add reduced-dimension low-fidelity models using the CA-MFMC algorithm.

4 High- and Low-Fidelity Models for Plasma Micro-Turbulence Analysis

In this chapter, we summarize the high- and low-fidelity models used in the CA-MFMC algorithm in our numerical experiments. Section 4.1 outlines the gyrokinetic code GENE that constitutes our high-fidelity model for plasma micro-turbulence simulation. The construction of data-driven low-fidelity models via a so-called sensitivity-driven dimension-adaptive sparse grid interpolation approach is described in Section 4.2, where we also summarize how reduced-dimension models of this type can be obtained.

4.1 High-Fidelity Model: The Gyrokinetic Simulation Code GENE

The high-fidelity model we use to simulate micro-turbulence in magnetized fusion plasmas is the gyrokinetic code GENE [18, 16]. We gave a short outline of gyrokinetic theory and the equations used to model plasma turbulence in Section 2.1. GENE is an Eulerian code that solves the five-dimensional gyrokinetic Maxwell-Vlasov system presented in Equations (2.1), (2.4) and (2.5) on a fixed grid using finite difference, finite volume and spectral methods.

In this work, we use GENE to perform local, or flux-tube, simulations. This means that the simulation domain is a small box that is aligned to the magnetic field lines as opposed to radially global, or full-torus, simulations, which can be orders of magnitude more computationally expensive. Further, we study linear simulations of plasma turbulence. In linear simulations, the parts of the gyrokinetic Vlasov equation (2.1) which are nonlinear in the 5D gyrokinetic coordinates $(\mathbf{R}, v_{\parallel}, \mu_m)$ are omitted. Writing the gyrokinetic equation for particle species s as

$$\dot{\pi}_s = \mathbb{O}_{\text{lin}}(\pi_s) + \mathbb{O}_{\text{nonlin}}(\pi_s),$$

with operators $\mathbb{O}_{\text{lin}}$ and $\mathbb{O}_{\text{nonlin}}$ comprising the linear and nonlinear terms of the gyrokinetic Vlasov equation respectively, the (discretized) equation solved in linear GENE simulations is

$$\dot{\boldsymbol{\pi}}_s = O_{\text{lin}} \boldsymbol{\pi}_s,$$

where $\dot{\boldsymbol{\pi}}$ is a vector containing the linearized values of the distribution function of particle species s on a discretization grid, and O_{lin} is the discretized version of the linear operator $\mathbb{O}_{\text{lin}}$, i.e., O_{lin} is a matrix.

Such linear simulations can be used to gain valuable insights regarding general trends or parameter dependencies. The much more complex numerical simulations of nonlinear turbulent transport, which are necessary to make reliable quantitative predictions, will be considered in future research. These simulations represent chaotic processes in space and time, in which a large number of degrees of freedom are usually strongly coupled in a highly nonlinear fashion, and are computationally much more expensive at, e.g., $\mathcal{O}(10^4)$ CPU hours on supercomputers or even more.

The output of interest for our simulations is the growth rate of the dominant eigenmode, which is given by the real part γ_1 of the dominant eigenvalue, i.e., by the largest real part of an eigenvalue, of the linear operator O_{lin}. Therefore, to summarize, the high-fidelity model is described by

$$f^{(0)}(\boldsymbol{\theta}) = \gamma_1(\boldsymbol{\theta}),$$

where γ_1 is computed in linear flux-tube simulations using the gyrokinetic simulation code GENE.

4.2 A Data-Driven Low-Fidelity Model

In this work, we use data-driven low-fidelity models based on the sensitivity-driven dimension-adaptive sparse grid interpolation algorithm from [9, 10]. Section 4.2.1 summarizes sparse grid interpolation and establishes notation used in Section 4.2.2, in which the dimension-adaptive refinement algorithm is detailed. We outline the procedure by which we obtain reduced-dimension low-fidelity models of the same type in Section 4.2.3. The notation in this section follows the references cited above as well as the summary of this algorithm that is given in [19].

4.2.1 Interpolation on Sparse Grids

The low-fidelity model constructed using the approach presented in this section is an interpolant of the high-fidelity model on a sparse grid. We refer to [5] for more details on sparse grids. Such an approximation of the high-fidelity model $f^{(0)}$ reads

$$\mathcal{U}_{\mathcal{L}}^{d}[f^{(0)}] = \sum_{\boldsymbol{\ell}\in\mathcal{L}} \Delta_{\boldsymbol{\ell}}^{d}[f^{(0)}], \tag{4.1}$$

where $\mathcal{L} \subset \mathbb{N}^d$ is a finite set of multiindices $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_d) \in \mathbb{N}^d$, which describes the refinement of the grid, and the terms $\Delta^d_{\boldsymbol{\ell}}[f^{(0)}]$ are the so-called hierarchical surpluses

$$\Delta^d_{\boldsymbol{\ell}}[f^{(0)}] = \sum_{\mathbf{z} \in \{0,1\}^d} (-1)^{|\mathbf{z}|_1} \mathcal{U}^d_{\boldsymbol{\ell}-\mathbf{z}}[f^{(0)}], \tag{4.2}$$

where $|\mathbf{z}|_1 = \sum_{i=1}^d z_i$. The multidimensional operators $\mathcal{U}^d_{\boldsymbol{\ell}-\mathbf{z}}$ in (4.2) are tensorizations of one-dimensional operators $\mathcal{U}^i_{\ell_i - z_i}$:

$$\mathcal{U}^d_{\boldsymbol{\ell}-\mathbf{z}}[f^{(0)}] = \left(\bigotimes_{i=1}^d \mathcal{U}^i_{\ell_i - z_i} \right) [f^{(0)}]. \tag{4.3}$$

To construct the interpolation low-fidelity models, univariate Lagrange interpolation is used to define the one-dimensional operators $\mathcal{U}^i_{\ell_i}$ as

$$\mathcal{U}^i_{\ell_i} : C^0(\mathcal{X}_i) \to \mathbb{P}_{P_{\ell_i}}, \quad \mathcal{U}^i_{\ell_i}[g] := \sum_{n=1}^{\ell_i} g(\theta_n) L_n(\theta), \tag{4.4}$$

where $\mathbb{P}_{P_{\ell_i}}$ is the space of univariate polynomials of degree P_{ℓ_i}, $g : \mathcal{X}_i \to \mathbb{R}$ is a univariate function, and $\{L_n(\theta)\}_{n=1}^{\ell_i}$ are the Lagrange polynomials of degree $n-1$ satisfying $L_n(\theta_m) = \delta_{nm}$ for Kronecker's delta function δ_{nm}. To improve the numerical stability of Lagrange interpolation, (4.4) is implemented using the barycentric formula [3].

As interpolation nodes $\{\theta_n\}_{n=1}^{\ell_i}$, we employ weighted (L)-Leja points [24], which are computed as

$$\theta_1 = \arg\max_{\theta \in \mathcal{X}_i} \sqrt{\pi_i(\theta)} \tag{4.5}$$

$$\theta_n = \arg\max_{\theta \in \mathcal{X}_i} \sqrt{\pi_i(\theta)} \prod_{m=1}^{n-1} |(\theta - \theta_m)|, \quad n = 2, \ldots, \ell_i. \tag{4.6}$$

We note that the (L)-Leja points are nested and so, increasing the refinement level ℓ_i simply means adding new points to the set of interpolation nodes used on the previous level. This property is of use in the dimension-adaptive construction of the low-fidelity model, where, in refining the grid, all (computationally expensive) high-fidelity model evaluations performed for previous points can be re-used. Moreover, only one new point is added per level, i.e., only one additional high-fidelity evaluation must be carried out when the refinement level is increased.

In summary, the sparse grid approximation (4.3) can be written as

$$\mathcal{U}_{\boldsymbol{\ell}}^{d}[f^{(0)}](\boldsymbol{\theta}) = \sum_{\mathbf{p}\in\mathcal{P}_{\boldsymbol{\ell}}} f^{(0)}(\boldsymbol{\theta}_{\mathbf{p}}) L_{\mathbf{p}}(\boldsymbol{\theta}) \tag{4.7}$$

where $\boldsymbol{\theta}_{\mathbf{p}} = (\theta_{p_1}, \ldots, \theta_{p_d})$ is the multivariate (L)-Leja point corresponding to $\mathbf{p}$, $L_{\mathbf{p}}(\boldsymbol{\theta}) \coloneqq \prod_{i=1}^{d} L_{p_i}(\theta_i)$ is a multivariate Lagrange polynomial, and $\mathcal{P}_{\boldsymbol{\ell}} \coloneqq \{\mathbf{p} \in \mathbb{N}^d : \mathbf{0} \leq \mathbf{p} \leq (\ell_1 - 1, \ldots, \ell_d - 1)\}$.

4.2.2 Sensitivity-Driven Dimension-Adaptivity

In the previous section, we have outlined the construction of a sparse grid interpolant of the high-fidelity model. What is left to define such that the model can be fully computed is the set of multiindices $\mathcal{L}$. Evaluating the resulting sparse grid approximation (4.1) means computing the hierarchical surpluses (4.2) for each $\boldsymbol{\ell} \in \mathcal{L}$. In order to find a low-fidelity model for MFMC that is cheap to evaluate, the aim is to keep $\mathcal{L}$ as small as possible, while still guaranteeing that the model is an accurate approximation of the high-fidelity model. This goal can be achieved by using the sensitivity-driven dimension-adaptive algorithm of [9, 10] to compute $\mathcal{L}$, which we describe in the following.

In general dimension-adaptivity [15, 17], a refinement indicator $\epsilon(\boldsymbol{\ell})$ is computed for each subspace $\boldsymbol{\ell}$ that is a candidate for further refinement and based on $\epsilon(\boldsymbol{\ell})$, the next subspace to be refined is selected. In the dimension-adaptive algorithm employed here, $\epsilon(\boldsymbol{\ell})$ is based on sensitivity information in the form of Sobol' indices [37, 38], which give information about the importance of the individual input parameters to the approximated high-fidelity model output. They can be computed from the spectral coefficients $c_{\mathbf{p}}$ used in the pseudo-spectral representation [40] of the multivariate interpolation operators $\mathcal{U}_{\boldsymbol{\ell}}^{d}$ that can be written as

$$\mathcal{U}_{\boldsymbol{\ell}}^{d}[f^{(0)}](\boldsymbol{\theta}) \overset{(4.7)}{=} \sum_{\mathbf{p}\in\mathcal{P}_{\boldsymbol{\ell}}} f^{(0)}(\boldsymbol{\theta}_{\mathbf{p}}) L_{\mathbf{p}}(\boldsymbol{\theta}) = \sum_{\mathbf{p}\in\mathcal{P}_{\boldsymbol{\ell}}} c_{\mathbf{p}} \Phi_{\mathbf{p}}(\boldsymbol{\theta}), \tag{4.8}$$

where $\Phi_{\mathbf{p}}(\boldsymbol{\theta}) \coloneqq \prod_{i=1}^{d} \Phi_i(\theta_i)$ are orthonormal polynomials w.r.t. the pdf π. The spectral coefficients $c_{\mathbf{p}}$ are obtained by solving the linear system of equations

$$\sum_{\mathbf{p}\in\mathcal{P}_{\boldsymbol{\ell}}} c_{\mathbf{p}} \Phi_{\mathbf{p}}(\boldsymbol{\theta}_k) = \mathcal{U}_{\boldsymbol{\ell}}^{d}[f^{(0)}](\boldsymbol{\theta}_k)$$

for all (L)-Leja points $\boldsymbol{\theta}_k$ corresponding to the multiindex $\boldsymbol{\ell}$. Using the spectral coeffi-

cients, the surpluses (4.2) can be re-written as

$$\Delta_{\boldsymbol{\ell}}^{d}[f^{(0)}](\boldsymbol{\theta}) = \sum_{\mathbf{p}\in\mathcal{P}_{\boldsymbol{\ell}}} \Delta c_{\mathbf{p}} \Phi_{\mathbf{p}}(\boldsymbol{\ell}), \quad \Delta c_{\mathbf{p}} = \sum_{\mathbf{z}\in\{0,1\}^d} (-1)^{|\mathbf{z}|_1} c_{\mathbf{p}-\mathbf{z}}, \quad \Delta c_{\mathbf{0}} \coloneqq c_{\mathbf{0}}. \tag{4.9}$$

Further, it was shown in [9, 10] that the following holds:

$$\left\| \Delta_{\boldsymbol{\ell}}^{d}[f^{(0)}] \right\|_{L^2}^2 = \sum_{\mathbf{p}\in\mathcal{P}_{\boldsymbol{\ell}}} \Delta c_{\mathbf{p}}^2 = \Delta c_{\mathbf{0}}^2 + \sum_{\mathbf{p}\in\mathcal{P}_{\boldsymbol{\ell}}\setminus\{\mathbf{0}\}} \Delta c_{\mathbf{p}}^2 = (\mathbb{E}_{\boldsymbol{\ell}}[f^{(0)}])^2 + \Delta\mathrm{Var}_{\boldsymbol{\ell}}[f^{(0)}], \tag{4.10}$$

where $\mathbb{E}_{\boldsymbol{\ell}}[f^{(0)}])$ is the expectation surplus corresponding to the subspace given by $\boldsymbol{\ell}$, and $\Delta\mathrm{Var}_{\boldsymbol{\ell}}[f^{(0)}]$ is the associated variance surplus. Next, according to the equivalence between pseudo-spectral projection and Sobol' decompositions [38], the variance surplus $\Delta\mathrm{Var}_{\boldsymbol{\ell}}[f^{(0)}]$ can be further decomposed into surpluses $\Delta\mathrm{Var}_{\boldsymbol{\ell}}^{i}[f^{(0)}]$ due to the individual stochastic parameters as well as the surplus $\Delta\mathrm{Var}_{\boldsymbol{\ell}}^{\mathrm{inter}}[f^{(0)}]$ given by all of their possible interactions:

$$\Delta\mathrm{Var}_{\boldsymbol{\ell}}[f^{(0)}] = \sum_{i=1}^{d} \Delta\mathrm{Var}_{\boldsymbol{\ell}}^{i}[f^{(0)}] + \Delta\mathrm{Var}_{\boldsymbol{\ell}}^{\mathrm{inter}}[f^{(0)}]. \tag{4.11}$$

The individual variance surpluses can be computed as

$$\Delta\mathrm{Var}_{\boldsymbol{\ell}}^{i}[f^{(0)}] \coloneqq \sum_{\mathbf{p}\in\mathcal{J}_i} \Delta c_{\mathbf{p}}^2, \quad \mathcal{J}_i \coloneqq \{\mathbf{p}\in\mathcal{P}_{\boldsymbol{\ell}} : \mathbf{p}_i \neq 0 \wedge \mathbf{p}_j = 0, \forall j \neq i\}, \tag{4.12}$$

$$\Delta\mathrm{Var}_{\boldsymbol{\ell}}^{\mathrm{inter}}[f^{(0)}] \coloneqq \sum_{\mathbf{p}\in\mathcal{J}_{\mathrm{inter}}} \Delta c_{\mathbf{p}}^2, \quad \mathcal{J}_{\mathrm{inter}} \coloneqq \{\mathbf{p}\in\mathcal{P}_{\boldsymbol{\ell}} : |\mathbf{p}|_0 \geq 1\}, \tag{4.13}$$

for $i = 1, \ldots, d$, where $|\mathbf{p}|_0$ denotes the number of non-zero entries in $\mathbf{p}$. Here, we know that the terms in (4.12) and (4.13) are unnormalized Sobol' indices, which quantify the local contribution to the high-fidelity model output of each uncertain input in the given subspace as well as the contribution of the interaction of all the parameters. Based on this information, the sensitivity-driven approach refines the grid only in directions that are indicated as important to the model output. In this way, the algorithm exploits the fact that in the simulation of real-world scenarios, the inputs are often anisotropically coupled, as well as that the intrinsic stochastic dimensionality is often smaller than the actual number of input parameters, i.e., only a subset of the uncertain inputs is in fact important to the simulation. This is achieved in the following way.

The refinement indicator is defined as $\epsilon(\boldsymbol{\ell}) \coloneqq s_{\boldsymbol{\ell}}$, where $s_{\boldsymbol{\ell}}$ is the so-called sensitivity score. In each refinement step, $s_{\boldsymbol{\ell}}$ is computed as follows. We start with $s_{\boldsymbol{\ell}} = 0$ and define tolerances $\boldsymbol{\tau} \coloneqq (\tau_1, \ldots, \tau_d, \tau_{d+1})$, which act as inputs to the algorithm. Here,

$\tau_1, \ldots, \tau_d$ describe a threshold for the directional variance in each of the d stochastic dimensions, and τ_{d+1} describes a threshold for the variance due to the interaction between the directions. For a given multiindex $\boldsymbol{\ell}$, we compute the variance surpluses, or unnormalized Sobol' indices, (4.12) and (4.13) as described above and compare them with the respective entry in $\boldsymbol{\tau}$. If the surplus is larger than the corresponding tolerance, $s_{\boldsymbol{\ell}}$ is increased by 1. Therefore, $s_{\boldsymbol{\ell}}$ takes values between 0 and $d+1$, and reflects the contribution of the corresponding subspace. Once all the sensitivity scores are computed, the subspace corresponding to the $\boldsymbol{\ell}$ with the highest score is refined. In refining the multiindex $\boldsymbol{\ell}$, we add to the set of multiindices $\mathcal{L}$ the forward neighbors $\boldsymbol{\ell} + \mathbf{e}_i, i = 1, \ldots, d$ of $\boldsymbol{\ell}$, where $\mathbf{e}_i$ is the ith unit vector, such that the resulting set $\mathcal{L}$ remains downward closed, i.e., for any $\boldsymbol{\ell} \in \mathcal{L}$, we also have that $\boldsymbol{\ell} - \mathbf{e}_i \in \mathcal{L}, i = 1, \ldots d$. The refinement of the grid is finished if (i) for all of the possible subspace to be refined, none of the tolerances in $\boldsymbol{\tau}$ are exceeded, (ii) if a maximum refinement level is reached for at least one of the dimensions, or (iii) if there are no multiindices left to refine.

4.2.3 Finding Reduced-Dimension Low-Fidelity Models

In Section 3.3, we discussed the context-aware construction of reduced-dimension low-fidelity models for the use in MFMC. We described how, in order to reduce the stochastic dimensionality of a problem, we need to know which input parameters are the most relevant for the high-fidelity model output. Given that the refinement approach described in the current section computes sensitivity information for all the stochastic inputs, it provides exactly the information we need in order to select the important parameters. Therefore, the sensitivity-driven algorithm is a very suitable method for additionally constructing reduced-dimension low-fidelity models. We describe how this can be achieved.

After a sparse grid approximation that depends on all d input parameters has been found using the sensitivity-driven procedure, we can obtain total Sobol' indices [38] of the inputs from the data produced by the algorithm. These can be computed without any additional high- or low-fidelity evaluations. Let $\{c_{\mathbf{p}}\}_{\mathbf{p}=\mathbf{0}}^{\mathbf{P}_{\mathcal{L}}}$ be the coefficients of pseudo-spectral representation (4.8) of the interpolant at the end of the refinement procedure and let $\mathbf{P}_{\mathcal{L}}$ be the multivariate degree of the resulting polynomial. Then the total Sobol' indices $S_i^T, i = 1, \ldots, d$, are computed as

$$S_i^T = \frac{\sum_{\mathbf{p} \in \mathcal{K}_i} c_{\mathbf{p}}^2}{\mathrm{Var}[\mathcal{U}_{\boldsymbol{\ell}}^d[f^{(0)}]]}, \quad \mathrm{Var}[\mathcal{U}_{\mathcal{L}}^d[f^{(0)}]] = \sum_{\mathbf{p} \neq \mathbf{0}} c_{\mathbf{p}}^2, \tag{4.14}$$

where $\mathcal{K}_i = \{0 < \mathbf{p} < \mathbf{P}_{\mathcal{L}} : \mathbf{p}_i \neq \mathbf{0}\}$. These Sobol' indices give a measure of the global importance of the d uncertain inputs parameters to the high-fidelity output. The

index S_i^T incorporates information about the individual contribution of parameter i as well as its interaction with the other parameters. For this reason, the total Sobol' indices are well suited to determine which stochastic parameters are overall the most relevant to the high-fidelity model computations and to construct low-fidelity models which depend only on these important inputs. The procedure for obtaining reduced-dimension low-fidelity models based on the sensitivity-driven interpolation algorithm is therefore as follows.

First, we apply the sensitivity-driven approach to the full set of parameters to obtain a full-dimensional low-fidelity model, which allows us to compute the total Sobol' indices for all uncertain inputs according to (4.14). We then choose a set of input parameters with the largest total Sobol' indices on which the reduced-dimension model will depend. This selection can be made, e.g., by discarding all inputs whose Sobol' indices are smaller than some threshold. A reduced-dimension low-fidelity model is constructed by fixing all unimportant parameters to their deterministic mean value and re-running the sensitivity-driven approach for the remaining set of parameters to compute a sparse grid interpolant with a smaller dimensionality.

Consequently, the resulting reduced-dimension low-fidelity model will require fewer grid points for its construction since only directions that are important for the high-fidelity output will be explored. Therefore, fewer multiindices will be added to the set $\mathcal{L}$. This also means that fewer high-fidelity evaluations must be obtained in the process. Because $\mathcal{L}$ is smaller, the resulting model will be less expensive to evaluate than a full-dimensional model. At the same time, the accuracy of the reduced-dimension model can be expected to be close to that of the full-dimensional model because only parameters that do not contribute significantly are omitted.

5 Numerical Results

In this chapter, we present our numerical results obtained by using the CA-MFMC algorithm with reduced-dimension low-fidelity models to quantify uncertainty in a scenario from plasma micro-turbulence analysis with 14 uncertain input parameters. Section 5.1 summarizes the considered plasma physics setup and describes relevant details regarding the employed high-fidelity model. In Section 5.2, we construct low-fidelity models of full- and reduced stochastic dimensionality using the sensitivity-driven algorithm of [10, 9] and estimate their accuracy and cost rates. The results of the experiments performed using CA-MFMC sampling and our full- and reduced-dimension low-fidelity models are presented in Section 5.3.

5.1 Turbulence Suppression by Energetic Particles: Setup and High-Fidelity Model

One of the avenues of research that has recently been shown to be highly promising regarding the suppression of micro-turbulence in magnetically confined fusion plasmas is the effect of fast ions on the instabilities. Such energetic particles are generated by auxiliary heating schemes, and are, for example, fast deuterium created by Neutral Beam Injection or Helium-3 obtained by Ion Cyclotron Resonance Heating. It is known that the main cause of plasma micro-instabilities is often the so-called ion temperature gradient (ITG) driven turbulence. Numerically, it has been shown, see, e.g., [7, 8, 6, 33], that the mechanism by which fast ions interact with the other particles can significantly reduce ITG driven micro-turbulence and thereby increase the plasma confinement time and energy output. The effect of the fast particles on the micro-instabilities is attributed to a wave-fast ion resonance interaction. A decrease of the turbulence can be observed when the frequency of the ITG instability is similar to the magnetic-drift frequency of the introduced energetic particles. When this is the case, the free energy can be redistributed and micro-turbulence is reduced. For more details on this process, see, e.g., [7]. Studying the stabilizing effect of the energetic particles in numerical simulations is thus of high practical relevance to the research into the prevention of plasma micro-turbulence, and therefore, to the realization of nuclear fusion power as a whole. Generally, there is a large number of parameters that characterize the fast ions

and other particle species, as well as the magnetic geometry of the setup, which could all affect the quality of the achieved turbulence suppression. Naturally, such quantities are all subject to uncertainties. Therefore, performing UQ is a key step in the research and simulation of these processes.

In this work, we consider a scenario with 14 uncertain input parameters that models the effect of fast deuterium and fast Helium-3 particles on the micro-turbulence. This setting is inspired by experiments conducted in [7, 8] for a configuration similar to a realistic plasma discharge in the JET fusion experiment. The same physics setup was studied in [19]. Therein, the regular MFMC algorithm was applied to quantify uncertainty in this scenario. In addition to a low-fidelity model of full dimension, the work used reduced-dimension low-fidelity models to reduce the computational effort needed to obtain accurate MFMC estimates. This approach led to a significant decrease of the MSE of MFMC estimators by multiple orders of magnitude. In that work, however, the full- and reduced-dimension low-fidelity models were constructed in a static manner, i.e., they were constructed using a set number of interpolation points. Given these already fixed models, the MFMC algorithm was used to find the best possible allocation of model evaluations among the high- and low-fidelity models such that the resulting estimators were as accurate as possible. With the numerical experiments carried out in this thesis, we extend the work from [19] by performing MFMC with data-driven low-fidelity models of full- and reduced-dimension, which are additionally context-aware. In that, we construct models that are designed in an optimal manner to be used together with the other models in the hierarchy, the goal being to further decrease the MSE of such estimators compared to the non-context-aware versions. We also mention the work [13], in which this application scenario was also studied. Therein, to begin with, a setup with 21 uncertain inputs was considered. Sensitivity analysis revealed the importance of the individual stochastic parameters, based on which the stochastic input space was reduced to the same 14 inputs we treat as uncertain in our experiments.

As a high-fidelity model, we employ the gyrokinetic code GENE [18, 16] for modeling plasma micro-turbulence and conduct linear simulations in the flux-tube limit, see Section 4.1 for a discussion. To compute the output of interest, i.e., the linear growth rate γ_1 of the micro-instabilities, we use GENE's initial value solver.

We consider 14 uncertain input parameters. Twelve of these stochastic inputs are the temperatures, densities, and the respective logarithmic gradients of the involved particles. In total, we consider five different kinds of particles. Fast deuterium and Helium-3 constitute the energetic particles. Additionally, we have deuterium ions and electrons as the two other main species. As the fifth type of particles, we consider Carbon impurities, whose simulation parameters are kept fixed to their deterministic mean values. The two remaining stochastic inputs characterize the magnetic geometry:

$\boldsymbol{\theta}$	**parameter**	**symbol**	**left bound**	**right bound**
θ_1	safety factor	q	1.3230	2.1705
θ_2	magnetic shear	$\hat{s}$	0.3920	0.6533
θ_3	ion log temperature gradient	ω_{T_i}	3.4230	5.7050
θ_4	ion log density gradient	ω_{n_i}	0.0047	0.0078
θ_5	ion temperature	T_i	0.7500	1.2500
θ_6	fastD log temperature gradient	ω_{T_D}	0.7742	1.2903
θ_7	fastD log density gradient	ω_{n_D}	3.5412	5.9022
θ_8	fastD density	n_D	0.0450	0.0750
θ_9	fastD temperature	T_D	7.2500	12.2500
θ_{10}	He-3 log temperature gradient	$\omega_{T_{^3\mathrm{He}}}$	5.5543	9.2573
θ_{11}	He-3 log density gradient	$\omega_{n_{^3\mathrm{He}}}$	0.3770	0.6283
θ_{12}	He-3 density	$n_{^3\mathrm{He}}$	0.0525	0.0875
θ_{13}	He-3 temperature	$T_{^3\mathrm{He}}$	9.0000	15.0000
θ_{14}	electron log temperature gradient	ω_{T_e}	1.6695	2.7825

Table 5.1: Summary of the 14 uncertain input parameters to our application scenario from plasma micro-turbulence analysis. All inputs are modeled with a uniform probability distribution using the respective left and right bounds given in the table.

the safety factor q and the magnetic shear $\hat{s}$. All uncertain parameters are modeled as independent uniform random variables with 25% bounds around a nominal value obtained from expert opinion and measurements. The parameters with their respective bounds are listed in Table 5.1. The density of the deuterium ions and the electron logarithmic density gradient are computed in terms of the densities and respective gradients of the other particle species such that quasi-neutrality, a central plasma characteristic, is fulfilled. In our experiments, we consider a fixed perpendicular wave number $k_y\rho_s = 0.5$, which is known to be the most unstable since it exhibits the highest linear ITG growth rate [7]. Additionally, an analytical Miller equilibrium [23] is used to describe the magnetic geometry, while inter- and intra-species collisions are modeled by a linearized Landau-Boltzmann operator.

The GENE runs are performed on the Linux cluster CoolMUC-2 at Leibniz Supercomputing Centre[1], where 240 cores on nine Intel Xeon E5-2690 v3 nodes are used per simulation. The number of points in each dimension of the grid used to discretize the five-dimensional gyrokinetic phase space is

$$N_x = 21, \quad N_y = 1, \quad N_z = 24, \quad N_{v_\parallel} = 32, \quad N_{\mu_m} = 16,$$

[1] `https://doku.lrz.de/display/PUBLIC/Linux+Cluster`

which means that, in total, we have $N_x \times N_y \times N_z \times N_{v_\parallel} \times N_{\mu_m} = 258,048$ degrees of freedom. All simulations were performed using double precision arithmetics. For this setup, the average single-core runtime for one high-fidelity simulation is estimated from 50 standard MC samples as 11,574.8697 seconds. The variance of the high-fidelity model is estimated from the same samples as $\sigma_0^2 = 0.004379$.

5.2 Full- and Reduced-Dimension Low-Fidelity Models for CA-MFMC

In this section, we construct full- and reduced-dimension low-fidelity models for the described plasma physics scenario that are subsequently used for CA-MFMC. We employ the sensitivity-driven dimension-adaptive sparse grid interpolation algorithm of [10, 9], which was summarized in Section 4.2.

In our experiments, we compare two different kinds of low-fidelity evaluation costs that arise from different ways in which the model evaluations are computed. Firstly, we consider a simple evaluation operation in which the hierarchical surpluses of the sparse grid construction are computed strictly according to (4.2). The same evaluation function was used to obtain the numerical results in [19]. Secondly, we apply a newly implemented, faster evaluation procedure, where the terms $(-1)^{|\mathbf{z}|_1}$, i.e., the signs of the terms in the summation in (4.2), are computed and stored in advance. Therefore, they do not have to be computed every time a model evaluation is obtained, which results in significant savings in the runtime of the evaluation operation. For all low-fidelity models, we estimate cost rates using both types of evaluation functions and use both in the CA-MFMC sampling algorithm, which allows us to study the impact of changes in the cost rate on CA-MFMC. We use $\bar{w}_{j,s}$ with subscript s (”slow”) to refer to the evaluation costs of low-fidelity model $f_{n_j}^{(j)}$ when the first procedure is used, and employ $\bar{w}_{j,f}$ with subscript f (”fast”) to denote costs using the new implementation. Similarly, if the context requires the evaluation procedure to be specified, $n_{j,s}$ and $n_{j,f}$ are used to refer to the respective number of high-fidelity training samples from which the model $f_{n_j}^{(j)}$ is constructed. All low-fidelity evaluation costs reported in this chapter were measured on a laptop computer with an Intel Core i5-8250U CPU at 1.60GHz.

5.2.1 Estimation of Accuracy and Cost Rates

In order for the CA-MFMC to be able to find the optimal number of high-fidelity evaluations that should be used to construct the low-fidelity models, estimates of the accuracy and cost rates of the models as defined in Assumptions 1 and 2 are needed. We summarize the procedure we use to obtain these rates in the following.

Firstly, we know that parametrized high-fidelity models generally have a finite number of bounded derivatives [9]. This implies that interpolation approximations of such models have algebraic accuracy rates [39]. Additionally, we have discussed in Section 4.2.2 how the evaluation costs of the sparse grid interpolation low-fidelity models are determined by the number of multi-indices in $\mathcal{L}$ as the hierarchical surpluses are evaluated for each $\boldsymbol{\ell} \in \mathcal{L}$. Whenever a multi-index is added to $\mathcal{L}$, one new (L)-Leja point is added as well. Therefore, the number of interpolation nodes is the same as the number of multi-indices in $\mathcal{L}$. For this reason, the sensitivity-driven low-fidelity models also have algebraic cost rates, which, due to the facts mentioned above, can even expected to be linear in the number of grid points. We can thus write the accuracy and cost rates of model $f_{n_j}^{(j)}$, $j = 1, \dots, k$, as

$$c_{a,j} r_{a,j}(n_j) = c_{a,j} n_j^{-\alpha_j} \quad \text{and} \quad c_{c,j} r_{c,j}(n_j) = c_{c,j} n_j^{\beta_j} \tag{5.1}$$

with constants $c_{a,j}, c_{c,j} > 0$ and $\alpha_j, \beta_j > 0$.

Remark 4 *The algebraic accuracy rate given in* (5.1) *is strictly convex for all $n_j \in [1, \bar{p} - 1]$ and $\bar{p} > 0$. In addition, if $\beta_j > 1$ holds, the algebraic cost rate in* (5.1) *is strictly convex for the same n_j as well. In this case, as stated in Remarks 2 and 3, the objective functions that are minimized in CA-MFMC in order to find the optimal number of high-fidelity training samples n_j^* all have a unique global minimum according to Theorems 1 and 3. Additionally, the minimizer n_j^* is bounded from above independent of the computational budget.*

In our application, the values $c_{a,j}, c_{c,j}, \alpha_j$ and β_j must be estimated numerically. To this end, for all low-fidelity models considered in this work, we apply the sensitivity-driven algorithm using tolerances $\boldsymbol{\tau} = 10^{-12} \cdot \mathbf{1}_{15}$ and a maximum refinement level $L_{\max} = 20$. During the execution of the refinement procedure, we use 50 standard MC samples to estimate the correlation coefficient ρ_j, according to (2.9), and the two different types of evaluation costs $\bar{w}_{j,s}$ and $\bar{w}_{j,f}$ of the sparse grid interpolant at different values of n_j. Here, we choose these at intervals of 25 grid points. Since estimates for the evaluation costs can be obtained using only (cheap) low-fidelity evaluations, unless otherwise stated, we compute the costs as an average of 10 independent estimates of $\bar{w}_{j,s}$ and $\bar{w}_{j,f}$, respectively, to ensure that the resulting cost rates are accurate. As stated in Remark 1, the pilot samples used in this pre-processing step can be re-used as MFMC model evaluations without significantly biasing the sampling results.

We note that in each refinement step of the dimension-adaptive algorithm, the number of newly explored subspaces can be as large as the stochastic dimensionality d. In practice, as the algorithm progresses and the sparse grid is refined further, the amount of added multi-indices per refinement step is, however, generally much smaller than that. This is due to the fact that the algorithm adapts to the structure of the

underlying problem and only explores the important dimensions further. Despite this fine granularity, not every value of n_j generally defines a valid sparse grid obtained by the algorithm. Therefore, whenever the number of grid points n_j is not feasible, we choose the next largest, valid value for n_j.

Based on the obtained correlation coefficients and evaluation costs for different values of n_j, we perform numerical regression in order to estimate the constants in (5.1). We write the problem in the form of a linear regression, see, e.g., [34], and use the least-squares procedure `lstsq`[2] provided by the `scipy.linalg` module for `python` to compute a solution.

5.2.2 Full-Dimension Low-Fidelity Model

We begin by considering a low-fidelity model $f_{n_1}^{(1)}$ that depends on all 14 uncertain inputs. We apply the sensitivity-driven refinement procedure and estimate the accuracy and cost rates of this full-dimensional model via regression as described above. The results are depicted in Figure 5.1. It can be observed that the algebraic accuracy and cost function do indeed approximate the rates well. We show the cost rate for the slower, simple evaluation procedure of the sparse grid interpolant as well as the faster implementation. We note that, for the first case, the evaluation costs are not estimated as averages of multiple standard MC estimates and only for smaller values of n_1. Because of the high dimensionality and the orders of magnitude difference in runtime of the two evaluation operations, computing more estimates would require a large amount of computational effort. Nevertheless, we note that these (slow) evaluation costs behave almost perfectly in a linear fashion with $\beta_1 = 0.9948$, as we had expected, which is why we assume the estimated rate to be accurate. Similarly to the slower implementation of the evaluation function, the cost rate using the faster procedure behaves in a way that is close to linear, however, here, the estimated rate is slightly larger at $\beta_1 = 1.3073$. The constants estimated for the accuracy and both cost rates are summarized in Table 5.2. From these computations, we also observe that in the 14-dimensional setting, the new implementation of the evaluation function is more efficient by about three orders of magnitude, which presents significant savings.

5.2.3 Reduced-Dimension Low-Fidelity Models

We now apply the method we described in Section 3.3 to find reduced-dimension low-fidelity models for CA-MFMC. Based on the low-fidelity model with full stochastic dimension and the high-fidelity evaluations used to construct it, we compute sensitivity information to inform the dimensionality reduction, see Section 4.2.3.

[2] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.linalg.lstsq.html`

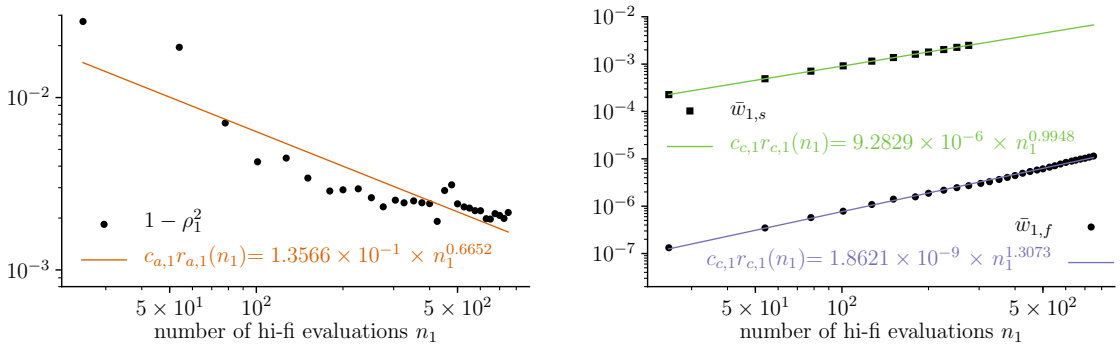

Figure 5.1: Estimated accuracy (left) and cost rates (right) for the 14-dimensional low-fidelity model $f_{n_1}^{(1)}$.

full-dimension low-fidelity model $\mathbf{f_{n_1}^{(1)}}$ (14D)		
rate	**estimated constants**	
accuracy $1-\rho_1^2$	$c_{a,1} = 1.3566 \times 10^{-1}$	$\alpha_1 = 0.6652$
evaluation costs $\bar{w}_{1,s}$	$c_{c,1} = 9.2829 \times 10^{-6}$	$\beta_1 = 0.9948$
evaluation costs $\bar{w}_{1,f}$	$c_{c,1} = 1.8621 \times 10^{-9}$	$\beta_1 = 1.3073$

Table 5.2: Estimated constants of the accuracy and cost rates for the 14-dimensional low-fidelity model $f_{n_1}^{(1)}$.

We use (4.14) to obtain the total Sobol' indices of the 14 input parameters from the spectral coefficients of the sparse grid interpolation constructed from $n_1 = 101$ high-fidelity samples. The high-fidelity evaluations used here were obtained when the the accuracy and cost rates were estimated, meaning there are no additional high-fidelity costs to these computations. The results are depicted in Figure 5.2. It can be observed that especially the logarithmic ion temperature gradient ω_{T_i} is an important parameter in this setup with total Sobol' index $S_3^T = 0.5759$. Following that, the density $n_{^3He}$ of the Helium-3 particles and their logarithmic temperature gradient $\omega_{T_{^3He}}$ exhibit the next highest sensitivity indices with $S_{12}^T = 0.1721$ and $S_{10}^T = 0.1085$. As discussed in [19], this is consistent with our plasma physics setup, where the micro-turbulence is mainly driven by the ion temperature gradient and Helium-3 is one of the energetic particle species introduced with the goal of reducing the turbulence. For the five parameters $\theta_2, \theta_4, \theta_6, \theta_7$, and θ_{11}, the computed Sobol' indices are all smaller than 5×10^{-4}, which suggests that these inputs are not important to the high-fidelity model output. Additionally, the total Sobol' indices for $\theta_8, \theta_9, \theta_{13}$ all fall below 0.01. This means

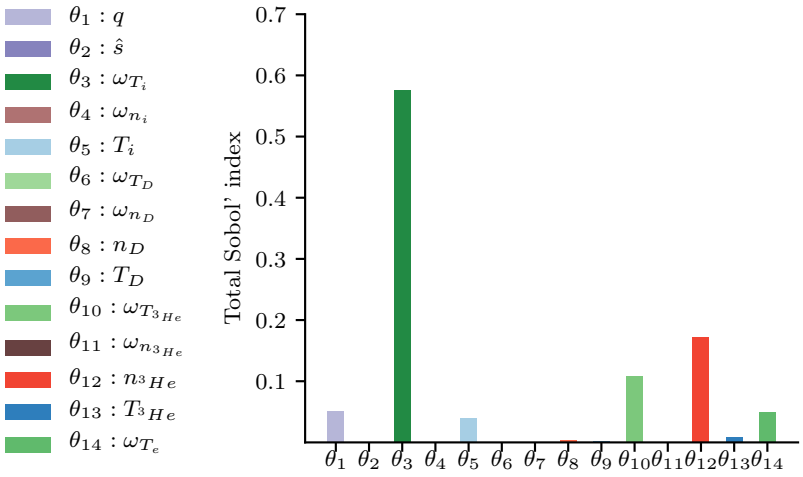

Figure 5.2: Total Sobol' indices of the 14 stochastic input parameters obtained from sensitivity-driven dimension-adaptive sparse grid interpolation using $n_1 = 101$ grid points.

$f^{(j)}$	d_j	$\boldsymbol{\theta}_j$
$f^{(0)}$	14	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}\}$
$f^{(1)}_{n_1}$	14	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}\}$
$f^{(2)}_{n_2}$	9	$\{\theta_1, \theta_3, \theta_5, \theta_8, \theta_9, \theta_{10}, \theta_{12}, \theta_{13}, \theta_{14}\}$
$f^{(3)}_{n_3}$	6	$\{\theta_1, \theta_3, \theta_5, \theta_{10}, \theta_{12}, \theta_{14}\}$

Table 5.3: Summary of the high- and low-fidelity models with their input dimensionality d_j and input parameters $\boldsymbol{\theta}_j$.

that in total, there are at least eight parameters that do no contribute significantly. The obtained sensitivity information is used to find reduced-dimension low-fidelity models. Based on the computed total Sobol' indices, we follow [19], where the input space was first reduced to the most important nine, and in a second step, to the most important six parameters. Table 5.3 summarizes the high- and low-fidelity models we use for MFMC with their respective input parameters.

We fix the parameters that are not important to the mean values of their probability distribution as given in Table 5.1. The dimension-adaptive algorithm is then applied to find a low-fidelity model with nine and six parameters, respectively, and the accuracy and cost rates are estimated as for the full-dimension low-fidelity model. The results of the regression are shown in Figures 5.3 and 5.4, and the estimated rate constants are summarized in Table 5.4.

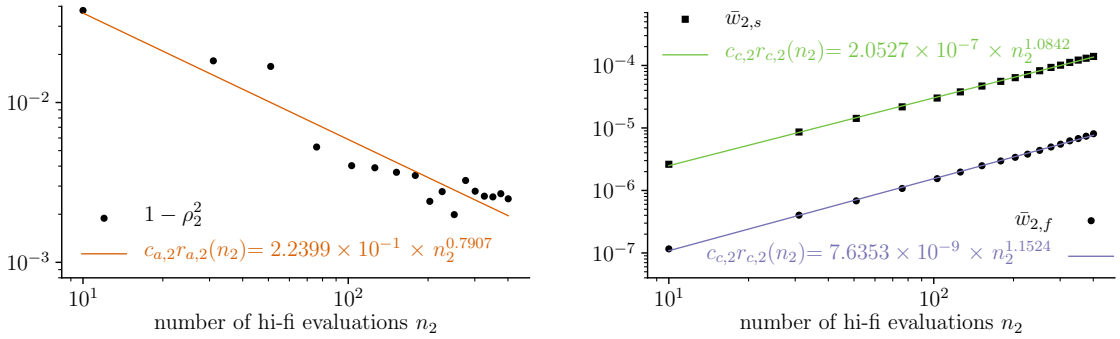

Figure 5.3: Estimated accuracy (left) and cost rates (right) for the nine-dimensional low-fidelity model $f_{n_2}^{(2)}$.

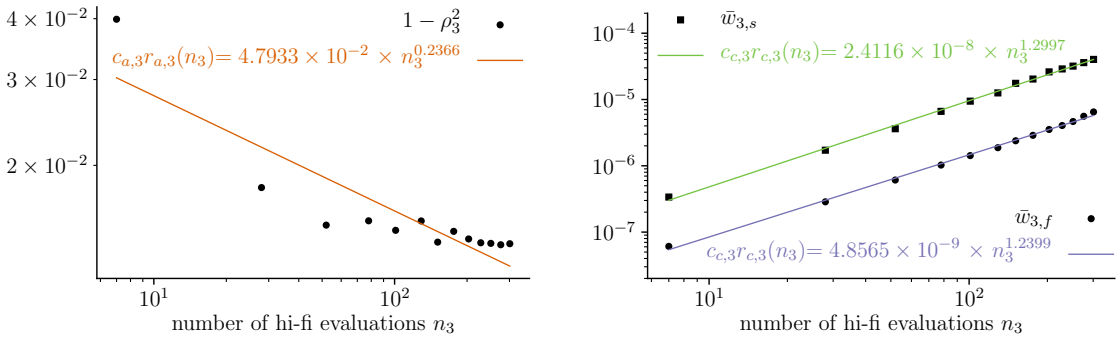

Figure 5.4: Estimated accuracy (left) and cost rates (right) for the six-dimensional low-fidelity model $f_{n_3}^{(3)}$.

We observe that for both reduced-dimension models, the cost rates can be described well by algebraic, here even almost linear, functions. We point out that the difference in evaluation costs between the two evaluation methods lessens with the dimensionality of the low-fidelity model. This is because the number of terms in (4.2), and therefore, the number of signs that have to be computed in the slower evaluation procedure decreases exponentially with the number of dimensions d. Furthermore, we note that, for the accuracy rate of the six-dimensional low-fidelity model, the exponent α_3 is estimated to be only 0.2366, whereas for the 14- and nine-dimensional models, we had $\alpha_1 = 0.6652$ and $\alpha_2 = 0.7907$. Additionally, towards the end of the examined range of values for the number of interpolation points n_3, the accuracy of the six-dimensional model starts

reduced-dimension low-fidelity model $\mathbf{f}_{\mathbf{n}_2}^{(2)}$ (9D)		
rate	**estimated constants**	
accuracy $1-\rho_2^2$	$c_{a,2} = 2.2399 \times 10^{-1}$	$\alpha_2 = 0.7907$
evaluation costs $\bar{w}_{2,s}$	$c_{c,2} = 2.0527 \times 10^{-7}$	$\beta_2 = 1.0842$
evaluation costs $\bar{w}_{2,f}$	$c_{c,2} = 7.6353 \times 10^{-9}$	$\beta_2 = 1.1524$

reduced-dimension low-fidelity model $\mathbf{f}_{\mathbf{n}_3}^{(3)}$ (6D)		
rate	**estimated constants**	
accuracy $1-\rho_3^2$	$c_{a,3} = 4.7933 \times 10^{-2}$	$\alpha_3 = 0.2366$
evaluation costs $\bar{w}_{3,s}$	$c_{c,3} = 2.4116 \times 10^{-8}$	$\beta_3 = 1.2997$
evaluation costs $\bar{w}_{3,f}$	$c_{c,3} = 4.8565 \times 10^{-9}$	$\beta_3 = 1.2399$

Table 5.4: Estimated constants of the accuracy and cost rates for the 9- and 6-dimensional reduced-dimension low-fidelity models $f_{n_2}^{(2)}$ and $f_{n_3}^{(3)}$.

to stagnate and not decrease further. We attribute both of these observations to the dimensionality reduction, where only six out of 14 inputs remained. Even though we have only omitted inputs with small total Sobol' indices, those values were nonetheless greater than zero. Therefore, reducing the input dimensionality still means that some information is lost and the model will not be as accurate as a full-dimensional model. Consequently, increasing the number of grid points of the interpolation low-fidelity model cannot replace the effect of the uncertainty introduced by the omitted inputs, which was not captured when parameters were fixed to their mean values.

5.3 Context-Aware Multi-Fidelity Monte Carlo Sampling

Now, we apply the full- and reduced-dimension low-fidelity models obtained in the previous section together with the high-fidelity model GENE to perform forward UQ in our plasma micro-turbulence scenario. We use the CA-MFMC algorithm to obtain estimators for the mean of the linear growth rate and compute the MSE w.r.t. a reference solution. Additionally, we compare the results of CA-MFMC with those of standard MC and MFMC with static low-fidelity models.

5.3.1 Analytical MSE

In order to provide an overview of the accuracy of CA-MFMC in the plamsa physics application under consideration, we compute the analytical MSE (2.14) of CA-MFMC mean estimators. We consider subsets of the three available low-fidelity models and

computational budgets $p \in [10^5, 10^{11}]$ seconds, where we include the larger budgets in this range in order to show the behavior of the convergence rates of the CA-MFMC estimators dependent on the budgets.

Recall that applying the CA-MFMC algorithm means, given a computational budget, finding the optimal number of high-fidelity evaluations that should be used to refine the involved low-fidelity models. This is achieved by minimizing the upper bounds (3.3) and (3.8) on the MSE of the MFMC mean estimator w.r.t. the number of training samples, see the optimization problems (3.5) and (3.10). We solve this numerical minimization using the `minimize`[3] function in the `scipy.optimize` library for `python`. Therein, we employ the `trust-constr` method to perform constrained optimization using a trust-region algorithm in order to find an accurate minimizer. With the obtained n_j^*, $j = 1, \dots, k$, for each low-fidelity model in the respective model combinations, we use Assumptions 1 and 2 to estimate the correlation coefficient and evaluation costs of the model constructed using n_j^* interpolation points. These can be used to compute the analytical MSE (2.14). Since the optimal number of high-fidelity samples used for refining a low-fidelity model is dependent on the computational budget, this process must be carried out for each budget separately.

As discussed in Remark 4, an optimal n_j^* exists for algebraic accuracy and cost rates as defined in (5.1). The n_j^* is unique and bounded when the accuracy rate is strictly convex and the cost rate is convex, which is true if $\beta_j > 1$. This condition holds for the cost rates of all our full- and reduced-dimension low-fidelity models, except for the cost rate w.r.t. the slower evaluation costs $\bar{w}_{1,s}$ of the full-dimensional model $f_{n_1}^{(1)}$. For this configuration, we have $\beta_1 = 0.9948 < 1$, see Table 5.1. Since the corresponding cost rate is therefore concave, it is not clear whether Theorems 1 and 3 hold. In our setup, the 14-dimensional low-fidelity model is the most accurate, implying that it always needs to be added first to the multi-fidelity hierarchy. Therefore, it is sufficient to check whether (3.6) in Theorem 1 applies for the considered computational budgets $p \in [10^5, 10^{11}]$. If this is the case, then the solution $n_{1,s}^*$ of the minimization problem (3.5) exists and is a global minimizer of the objective function. The condition (3.6) reads

$$u(n_1) := c_{a,1} r''_{a,1}(n_1) + c_{a,1} r''_{c,1}(n_1) > 0, \tag{5.2}$$

where, in our setting, $c_{a,1} r_{c,1}(n_1)$ describes the cost rate w.r.t. $\bar{w}_{1,s}$. We verify whether (5.2) holds for all computational budgets $p \in [10^5, 10^{11}]$ seconds. To this end, it is sufficient to examine the values of $u(n_1)$ for the largest considered budget of 10^{11} seconds since $u(n_1)$ does not depend on the budget. For $p = 10^{11}$ seconds, the corresponding possible number of high-fidelity evaluations is $n_1 \in [1, 8.6394 \times 10^6]$. If

[3] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html`

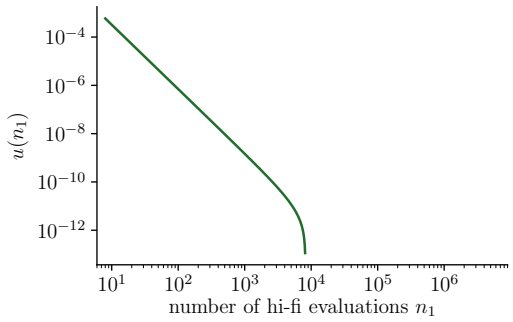

Figure 5.5: The function $u(n_1)$ defined in (5.2) for values of n_1 corresponding to budgets $p \in [10^5, 10^{11}]$. For $n_1 > 8200$, $u(n_1)$ becomes negative.

the condition (5.2) holds for all values of n_1 in this range, then Theorem 1 applies to computational budgets smaller than this as well. In Figure 5.5, we plot the function $u(n_1)$ for the considered n_1. We notice that $u(n_1)$ is only positive for n_1 up to around 8200, which corresponds to a budget of 9.4×10^7 seconds. This means that, using Theorem 1, we can only guarantee the existence of a unique global minimum of the objective function for budgets up to this size. However, condition (3.6) does not necessarily need to hold for the minimum in the considered range of budgets to exist. For our setting, we show this numerically. In Figure 5.6, we plot the objective function for budgets $p \in \{10^8, 10^9, 10^{10}, 10^{11}\}$. It can be observed all the depicted functions are clearly (stricly) convex and that a unique minimum exists. Based on these results, we expect the objective to be convex for all budgets $p \in [9.4 \times 10^7, 10^{11}]$ seconds. Therefore, we can assume that, for the budgets we consider here, a unique $n_{1,s}^*$ also exists when CA-MFMC is performed using the 14-dimensional low-fidelity model and the slower evaluation operation.

In addition to CA-MFMC mean estimators with all possible combinations of our low-fidelity models $f_{n_1}^{(1)}$, $f_{n_2}^{(2)}$, and $f_{n_3}^{(3)}$, we also consider standard MC and static MFMC mean estimators. We perform static MFMC using (i) a full-dimensional sparse grid low-fidelity model with a fixed, small number of interpolation points, and (ii) the same model with a larger number of grid points. The two considered models and their correlation coefficients, evaluation costs and variances are summarized in Table 5.5. To ensure a fair comparison to CA-MFMC, we take into account the fixed number of high-fidelity samples used for constructing the static low-fidelity models and compute the analytical MSE for regular MFMC for a budget $p - n_1 \times w_0$. For standard MC, the whole budget is used for model evaluations as usual.

We depict the analytical MSE for the considered computational budgets in the top

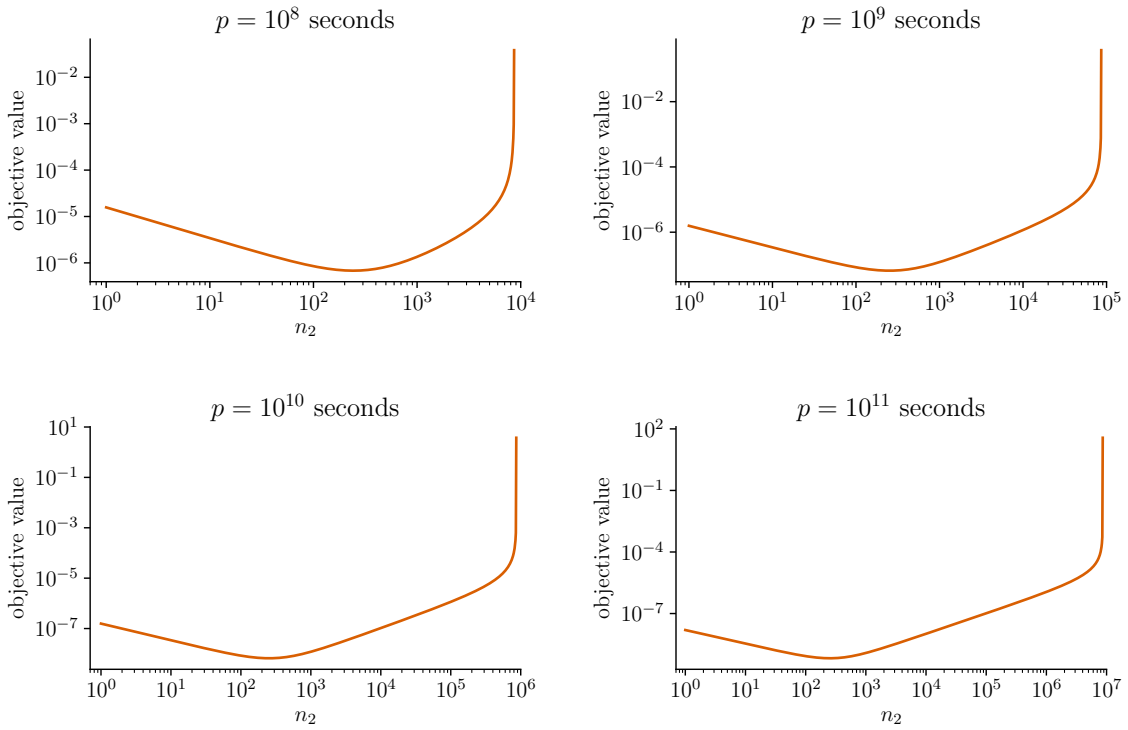

Figure 5.6: The objective function from (3.5) for computational budgets $p \in \{10^8, 10^9, 10^{10}, 10^{11}\}$ and the full-dimensional low-fidelity model $f_{n_1}^{(1)}$ with the slower evaluation operation. For all depicted budgets, the objective is convex.

n_1	ρ_1	$w_{1,s}$	$w_{1,f}$	σ_1^2
25	0.9860	2.6277	0.0014	0.0045
400	0.9860	44.3595	0.0578	0.0048

Table 5.5: Correlation coefficients ρ_1, (unnormalized) slow and fast evaluation costs $w_{1,s}$ and $w_{1,f}$, and variances σ_1^2 of the considered static, full-dimensional low-fidelity models constructed from n_1 high-fidelity evaluations.

plots in Figure 5.7 using the evaluation costs given by the slower evaluation function, and in Figure 5.8 for the faster evaluation procedure. In the bottom plot in both figures, we show again for easier readability the MSE values for the budget $p = 10^{10}$ seconds, as an example budget for which CA-MFMC has reached the asymptotic regime, meaning the numbers n_j^*, $j = 1, \ldots, k$, do not increase with the budgets anymore, and

instead the MSE decreases linearly with p. Firstly, for both cost rates, we observe that the convergence of the CA-MFMC mean estimator is faster than $\mathcal{O}(p^{-1})$ for smaller computational budgets. For the slower evaluation procedure, the MSE stars to converge with $\mathcal{O}(p^{-1})$ at about $p = 10^8$, whereas for the faster method, this regime begins at around $p = 5 \times 10^8$. Secondly, the figures show that, as expected, all MFMC and CA-MFMC estimators are multiple orders of magnitude more efficient than standard MC estimators. The static MFMC estimators with both $n_1 = 25$ and 400 continuously provide the least accurate results for all budgets. The MSEs of CA-MFMC estimators in which the number the number of high-fidelity training samples for $f_{n_1}^{(1)}$ is determined in a context-aware manner are at least one order of magnitude smaller than if n_1 is chosen as one of the two fixed values considered here. The fact that this difference in accuracy stems only from choosing a different number of high-fidelity training samples while the same computational budget is used underlines again the benefits of applying the CA-MFMC approach instead of statically constructing the low-fidelity models.

Comparing the analytical MSEs of estimators using the slow vs. estimators using the faster evaluation procedure, we observe that, of course, the faster method leads to MFMC and CA-MFMC estimators with smaller MSEs. In the asymptotic limit, this difference in accuracy amounts to up to around one order of magnitude for some of the model combinations used in CA-MFMC, e.g., using only the 14-dimensional model, where this has the most significant impact.

Furthermore, we can use these theoretical results to find the subsets of low-fidelity models that lead to CA-MFMC estimators with the smallest MSEs, and which should therefore be used when performing CA-MFMC sampling in practice. For the small computational budgets considered here, the different accuracy and cost rates of the models, and with that, the different convergence rates of the MSEs, mean that the most efficient combination of models is dependent on the budget. We can observe this, for example, in Figure 5.8, where the CA-MFMC estimator using $f_{n_1}^{(1)}$ and $f_{n_2}^{(2)}$ switches from being the one with the largest MSE for budget $p = 10^5$ to one of the most efficient estimators for larger budgets. We examine the accuracy of selected CA-MFMC estimators for one of these pre-asymptotic budgets in more detail in the following section. We now discuss the most efficient combinations for larger budgets, i.e., once all estimators converge with $\mathcal{O}(p^{-1})$. For the slower evaluation case, an accurate estimate is provided by CA-MFMC using only the nine-dimensional low-fidelity model $f_{n_2}^{(2)}$, which is only improved by adding the six-dimensional model $f_{n_3}^{(3)}$ to the hierarchy as well. Other combinations of the 14-, nine- and six-dimensional models follow closely behind CA-MFMC with only $f_{n_2}^{(2)}$. CA-MFMC using only the full-dimensional model $f_{n_1}^{(1)}$ is less accurate than those combinations. When the improved evaluation procedure is used, CA-MFMC using both reduced-dimension low-fidelity models still provides

the most accurate estimator, followed by the other combinations of 14-, nine- and six-dimensional models. Here, the MSEs of CA-MFMC with either $f_{n_1}^{(1)}$ or $f_{n_2}^{(2)}$ are closer than for the first evaluation function, with the nine-dimensional model still leading to a more accurate estimate.

5.3.2 Selection of Low-Fidelity Models for CA-MFMC

Based on the theoretical analysis in the previous section, we select a number of model combinations that we use to perform CA-MFMC in practice. As a computational budget, we choose $p = 5 \times 10^6$, which means that the computed MSEs are still in the pre-asymptotic regime, where the convergence is faster than $\mathcal{O}(p^{-1})$. This budget corresponds to $\lfloor (5 \times 10^6)/w_0 \rfloor = 431$ high-fidelity evaluations.

Firstly, we choose to perform CA-MFMC with only the nine-dimensional low-fidelity model $f_{n_2}^{(2)}$ in addition to the high-fidelity model since we know that this selection leads to small MSEs for both evaluation procedures. For comparison, we also compute CA-MFMC estimators using only the full-dimensional model $f_{n_1}^{(1)}$, for which we expect larger MSEs compared to estimators with only $f_{n_2}^{(2)}$. According to Figures 5.7 and 5.8 this difference will be more pronounced when the slower evaluation procedure is used. Additionally, only with regards to the more expensive evaluation costs $\bar{w}_{j,s}$, we perform CA-MFMC with both reduced-dimension models $f_{n_2}^{(2)}$ and $f_{n_3}^{(3)}$, where the reduction in MSE compared to using only $f_{n_3}^{(2)}$ is more distinct than with the faster evaluation operation.

We first compute CA-MFMC estimators using only the 14-dimensional low-fidelity model. For the budget $p = 5 \times 10^6$, we showed numerically in Section 5.3.1 using Theorem 1 that even though the cost rate w.r.t. $\bar{w}_{1,s}$ is concave and Remark 4 does not apply, a unique global minimizer $n_{1,s}^*$ of the objective function used to find the number of high-fidelity training samples still exists. By performing the minimization of the objective using the `scipy.optimize.minimize` procedure as described in the previous section, we obtain $n_{1,s}^* = 122$ and $n_{1,f}^* = 173$. In practice, we use 127 and 180 points, respectively, since the computed $n_{1,s}^*$ and $n_{1,f}^*$ do not correspond to a valid sparse grid obtained by the sensitivity-driven algorithm. We adjust the budget used for sampling the models accordingly. In the left plot in Figure 5.9, we depict the convergence of $n_{1,s}^*$ and $n_{1,f}^*$ with regards to the computational budget. It can be observed that, for both kinds of evaluation costs, the number of high-fidelity training samples is bounded from above, albeit the upper bound is significantly smaller when the low-fidelity evaluation are more expensive. Here, the obtained bounds were $\bar{n}_{1,s} = 254$ and $\bar{n}_{1,f} = 6874$. These numbers underline the trade-off computed by the CA-MFMC algorithm. When fewer low-fidelity samples can be obtained in a given budget due to higher evaluation costs,

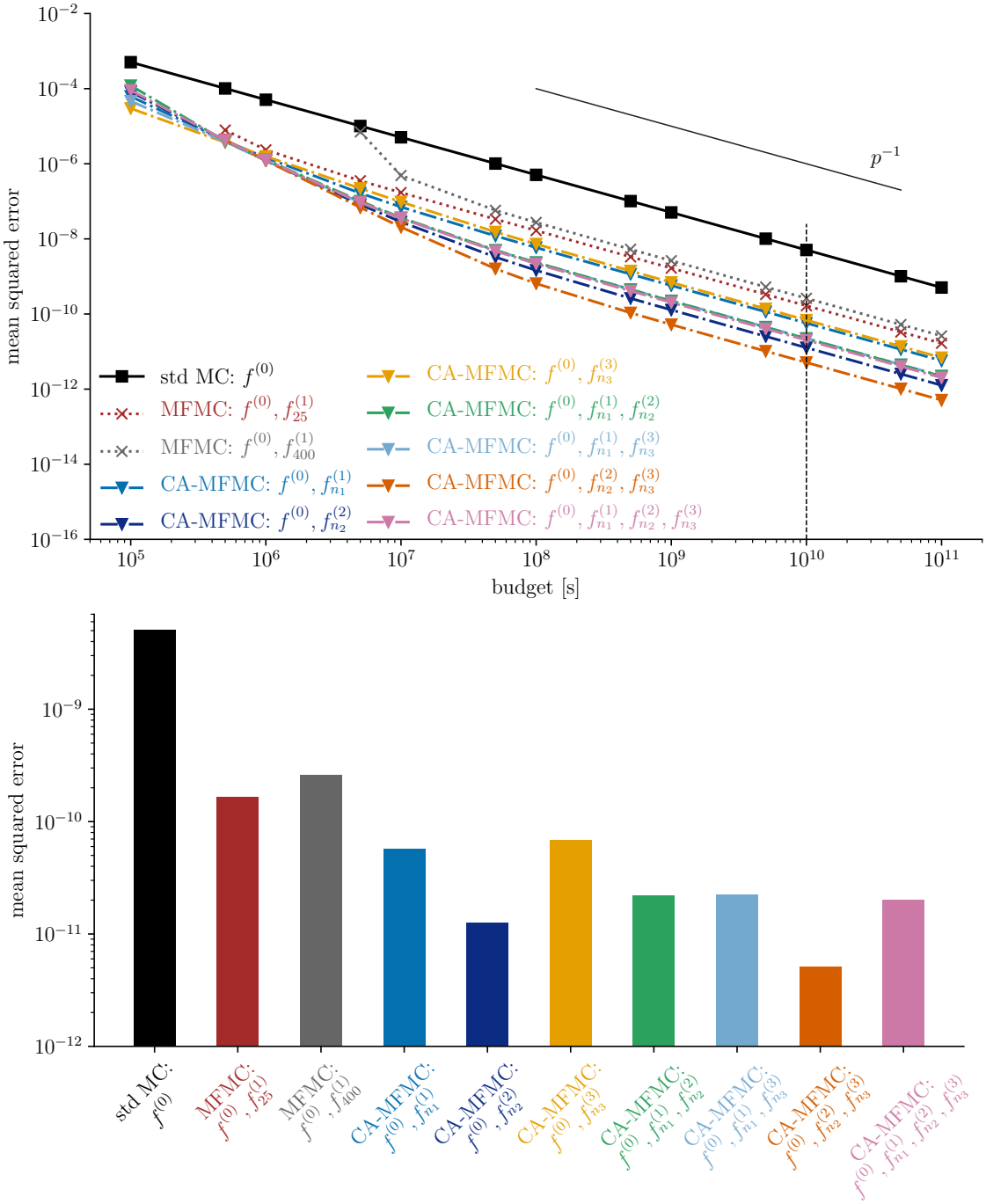

Figure 5.7: Analytical MSE of different combinations of low-fidelity models in CA-MFMC, MFMC with static full-dimensional models, and standard MC for the **slower** evaluation procedure. The bottom plot shows the MSE values for $p = 10^{10}$.

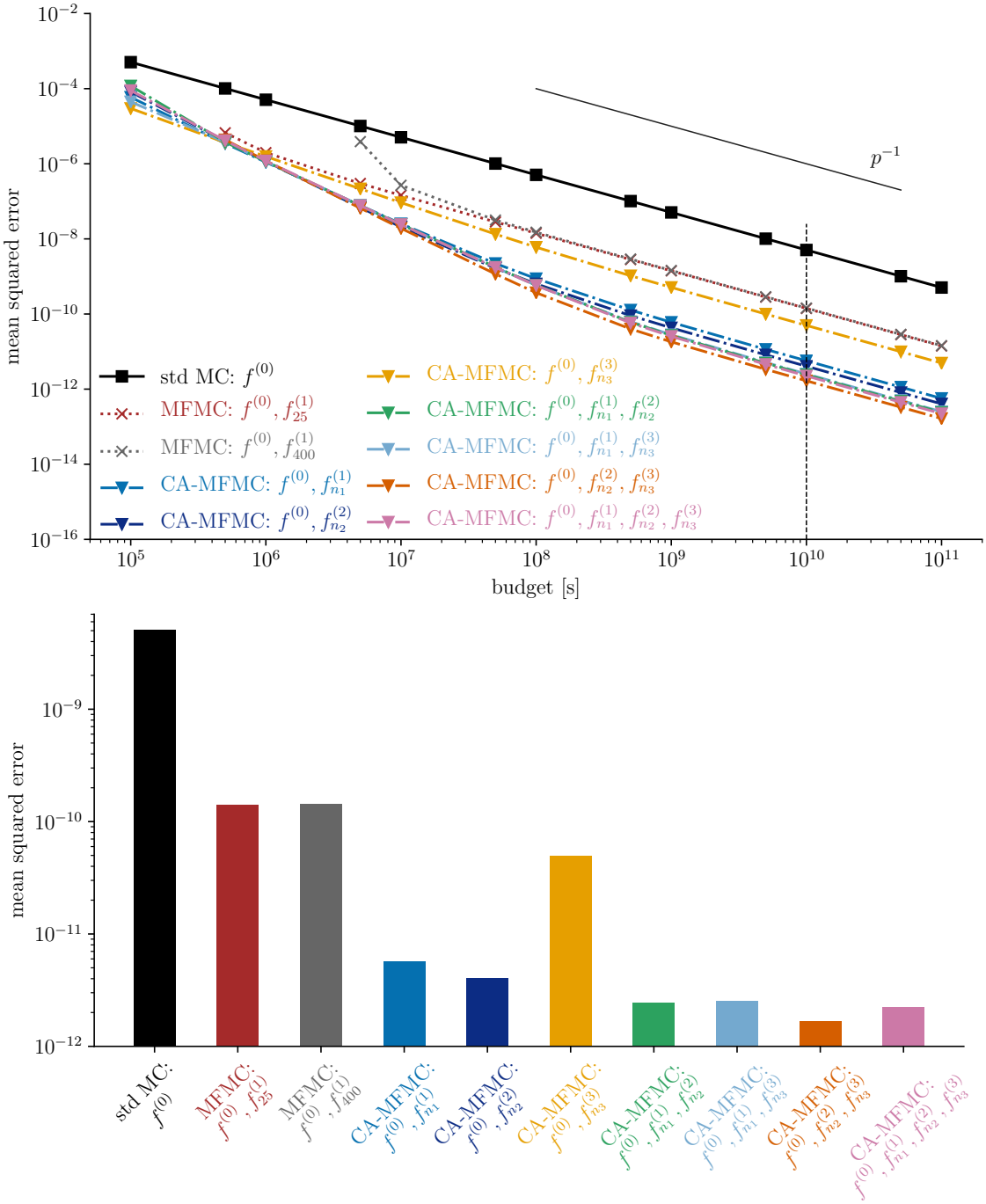

Figure 5.8: Analytical MSE of different combinations of low-fidelity models in CA-MFMC, MFMC with static full-dimensional models, and standard MC for the **faster** evaluation procedure. The bottom plot shows the MSE values for $p = 10^{10}$.

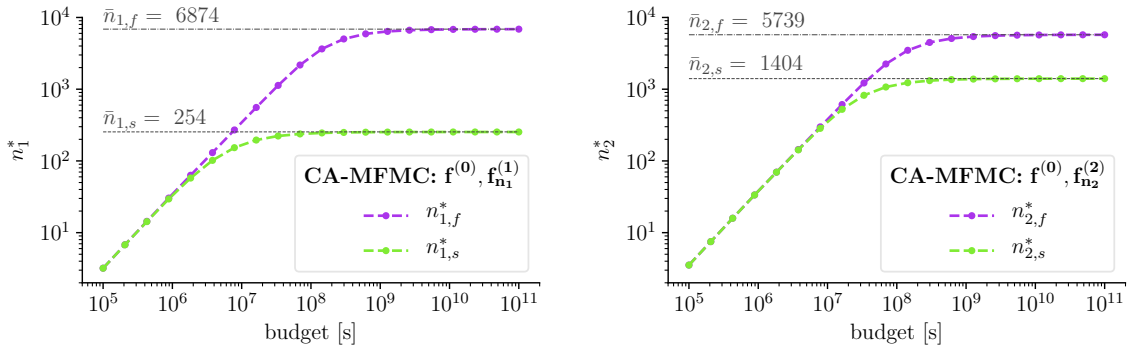

Figure 5.9: Optimal number of high-fidelity evaluations used to refine the low-fidelity models $f^{(1)}_{n_1}$ (left) and $f^{(2)}_{n_2}$ (right) when either are used as the only low-fidelity model in CA-MFMC.

less effort is spent on improving the model such that a sufficient number of high- and low-fidelity evaluations can be acquired in the sampling step. Note that, for the slow evaluation operation, the number of high-fidelity samples $n_1 = 400$, which we chose to construct one of the static low-fidelity models in the previous section, was larger than the upper bound $n^*_{1,s}$ computed here. This shows an instance where too much effort was spent on refining the low-fidelity model, leading to a ratio of (high) accuracy to evaluation costs that did not achieve as high of a variance reduction in MFMC as a smaller number $n^*_{1,s}$ could have.

Next, we perform CA-MFMC using the nine-dimensional model $f^{(2)}_{n_2}$. For the considered budget, the algorithm computes $n^*_{2,s} = 187$ and $n_{2,f} = 191$, where, in practice, we use 188 and 192 points to construct the sparse grids. The right plot in Figure 5.9 shows the convergence of $n^*_{2,s}$ and $n^*_{2,f}$ for CA-MFMC with only the nine-dimensional low-fidelity model. Similarly to the 14-dimensional case, the upper bound $\bar{n}_{2,f}$ using the less expensive evaluation operation is much larger than for the more expensive ones with $\bar{n}_{2,f} = 5739$ and $\bar{n}_{2,s} = 1404$.

For both CA-MFMC with only either $f^{(1)}_{n_1}$ or $f^{(2)}_{n_2}$, the optimal number of refinement samples is similar for both types of evaluation costs when the computational budgets are small. This is due to the large high-fidelity evaluations costs, which mean that, for example, a budget of 10^5 seconds corresponds only to $\lfloor 8.63 \rfloor = 8$ high-fidelity evaluations, 10^6 seconds to $\lfloor 86.39 \rfloor = 86$ evaluations and so on. Therefore, for small budgets, the number of high-fidelity training samples cannot be large, regardless of the low-fidelity evaluation costs, since the remaining budget must allow for enough evaluations of the low-fidelity model and especially, the expensive high-fidelity model.

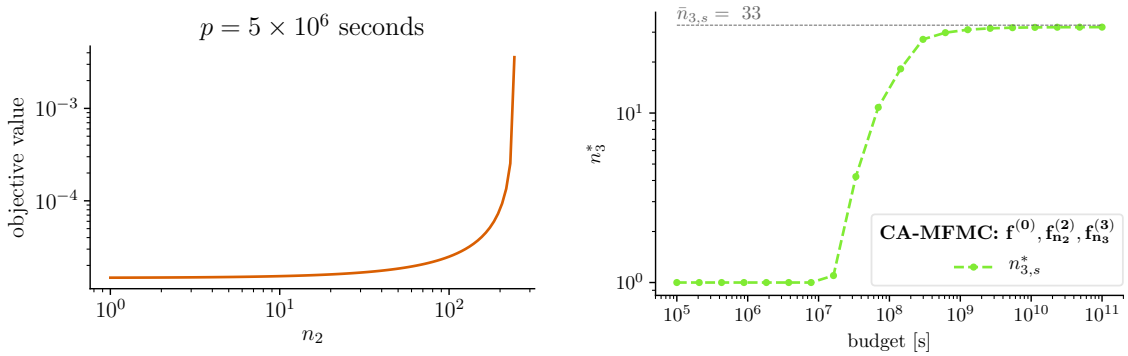

Figure 5.10: Left: The objective function that is minimized to find $n_{3,s}^*$ after $n_{2,s}^*$ has been computed. Right: Optimal number of high-fidelity evaluations $n_{3,s}^*$ used to refine the low-fidelity model $f_{n_3}^{(3)}$ when it is sequentially introduced into the CA-MFMC algorithm after $f_{n_2}^{(2)}$.

Thirdly, we examine the effects of adding the six-dimensional low-fidelity model $f_{n_3}^{(3)}$ to the CA-MFMC algorithm using only $f_{n_2}^{(2)}$ with the slower evaluation method. When $f_{n_3}^{(3)}$ is introduced as the second low-fidelity model, we find that, for small budgets, the real-valued minimizer of the objective function in (3.10) lies between 1 and 2, i.e., in practice $n_{3,s}^* = 2$ would be used. We plot the objective function for $p = 5 \times 10^6$ seconds in the left plot in Figure 5.10. This indicates that the remaining budget after $f_{n_2}^{(2)}$ has been constructed is not large enough to refine $f_{n_3}^{(3)}$ further and allow for a sufficient number of MFMC samples to be evaluated by all models. The right plot in Figure 5.10 shows the convergence of $n_{3,s}^*$ with regards to the computational budget. We observe that $n_{3,s}^*$ only starts increasing for budgets around $p = 2 \times 10^7$ seconds. The upper bound on $n_{3,s}^*$ is comparatively small as well at $\bar{n}_{3,s} = 33$. This can be explained by the slow convergence of the corresponding accuracy rate, meaning the model will gain relatively little in accuracy, while the evaluation costs keep increasing roughly linearly with the number of points n_2, see Figure 5.4. However, even for a small number of training samples, the six-dimensional model still contributes to variance reduction in MFMC if the computational budget is large enough. This demonstrates the fact that, in MFMC, low-fidelity models need not necessarily be very accurate as long as the ratio of evaluation costs to accuracy is beneficial to the algorithm. As observed in Figure 5.7, the CA-MFMC estimator using the nine- and six-dimensional low-fidelity models analytically still presents a small improvement compared to CA-MFMC with $f_{n_3}^{(2)}$ alone. For this reason, we also perform CA-MFMC with both models for the budget

$p = 5 \times 10^6$, where $n^*_{3,s} = 2$. Since the sensitivity-driven approach does not allow for a sparse grid using only two points, we use the next largest grid, which contains seven points.

5.3.3 Estimated MSE for Fixed Computational Budget

We compute CA-MFMC mean estimators using the three subsets of low-fidelity models described in the previous section for a computational budget of $p = 5 \times 10^6$ seconds. For comparison, we also compute standard MC estimators. We obtain a reference result for the mean value of the high-fidelity output by applying CA-MFMC for a budget $p_{\text{ref}} = 10^7$ seconds, where the nine-dimensional $f^{(2)}_{n_2}$ is used as a low-fidelity model together with the fast evaluation procedure. We know from Figure 5.8 that this will lead to an accurate estimator. Here, we obtained $n^*_{2,f} = 380$. In practice, we use 385 training samples to construct the model. The obtained reference mean is $\mu_{\text{ref}} = 0.1972$. We compute 10 independent replicates of the considered standard MC and CA-MFMC estimators. The MSE is estimated as

$$\text{MSE}[\hat{E}^{(\cdot)}] = \frac{1}{10} \sum_{i=1}^{10} (\hat{\mu}_{\text{ref}} - \hat{E}^{(\cdot)}_i)^2, \tag{5.3}$$

where $\hat{E}^{(\cdot)}$ and $\hat{E}^{(\cdot)}_i$ are either standard MC or CA-MFMC mean estimators.

The MSEs of the computed estimators for $p = 5 \times 10^6$ are shown in the right plot in Figure 5.11 with the analytical MSEs, which were shown previously in Figures 5.7 and 5.8, on the left. The computed MSEs confirm the predictions and the trends observed in the results of the analytical computations. As expected, using more computationally expensive low-fidelity models given by the slower evaluation procedure leads to larger MSEs. The difference in MSE is larger when only the 14-dimensional model is employed in CA-MFMC compared to using the nine-dimensional model, which can be explained by the larger reduction of the evaluation costs when the faster implementation is used, see the cost rates in Figures 5.1 and 5.3. CA-MFMC using either the 14- or nine-dimensional low-fidelity model and the fast evaluation procedure is more accurate than standard MC by at least two orders of magnitude, with $f^{(2)}_{n_2}$ as the low-fidelity model leading to the more accurate estimator. For CA-MFMC using the nine- and six-dimensional low-fidelity models and the slow evaluation procedure, the MSE of the computed estimators is slightly larger than what would be expected given the analytical results. Besides the fact that the number $n^*_{3,s}$ that was computed by the CA-MFMC algorithm could not be used exactly, we attribute this to one main factor. Considering the results of the regression in Figure 5.4 (left) used to determine the accuracy rate of $f^{(3)}_{n_3}$, we see that the accuracy of the model at $n_3 = 7$ presents, to some extent, an

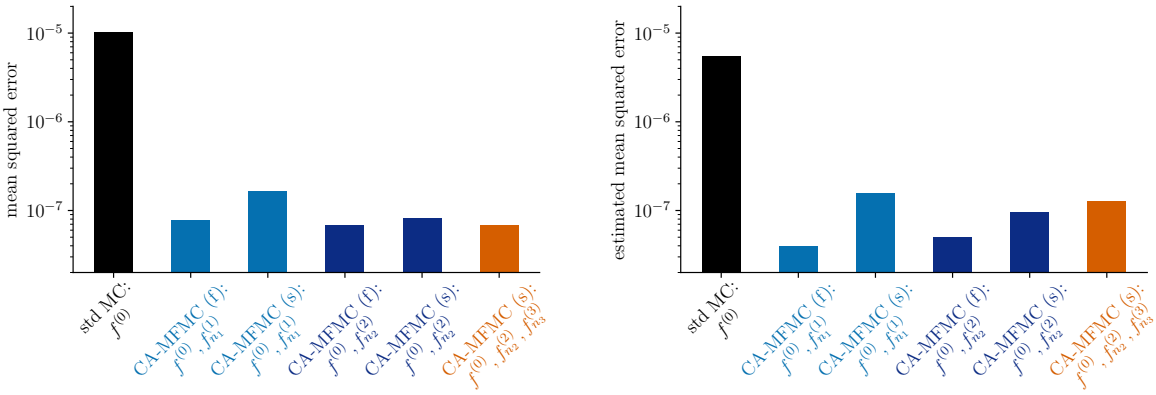

Figure 5.11: Analytical (left) and computed MSE (right) of standard MC and CA-MFMC estimators for different model combinations and $p = 5 \times 10^6$.

outlier. The accuracy of the model at this value of n_3 is not approximated well by the computed accuracy rate. In fact, the model is less accurate in reality than the accuracy rate suggests. The CA-MFMC algorithm therefore assumes the low-fidelity model to be higher correlated with the high-fidelity model than it actually is, explaining why the MSE of the corresponding estimator is larger than the analytical value would imply.

For the numerical results we obtained, we observe the following speedup. Out of the estimators we computed here, we consider the most accurate one, which is the CA-MFMC estimator using the nine-dimensional low-fidelity model and the faster evaluation procedure. The computational budget $p = 5 \times 10^6$ corresponds to around 5.7 hours of runtime on 240 cores on the CoolMUC-2 Linux cluster. To obtain a standard MC estimator of the same MSE on 240 cores, a computational budget of 48 days would be needed. These numbers illustrate the potential of the CA-MFMC algorithm as a method for tackling the propagation of uncertainty in settings with expensive high-fidelity models, where, as shown here, standard MC may not be feasible approach.

Finally, we discuss the split of the computational budget in CA-MFMC between offline and online costs, i.e., between improving the low-fidelity models and evaluating high- and low-fidelity models for samples of the input distribution. We depict the distribution of the budget between those two steps in Figure 5.12 for the example of the CA-MFMC estimator using the 14-dimensional low-fidelity model. It can be observed for both evaluation procedures that in the pre-asymptotic regime, the percentage of the budget used for either refining or sampling the models decreases only slightly, which is consistent with the fact that, here, n_1^* grows with the budget. Once the convergence of the CA-MFMC estimators reaches $\mathcal{O}(p^{-1})$, the share of the budget that is used

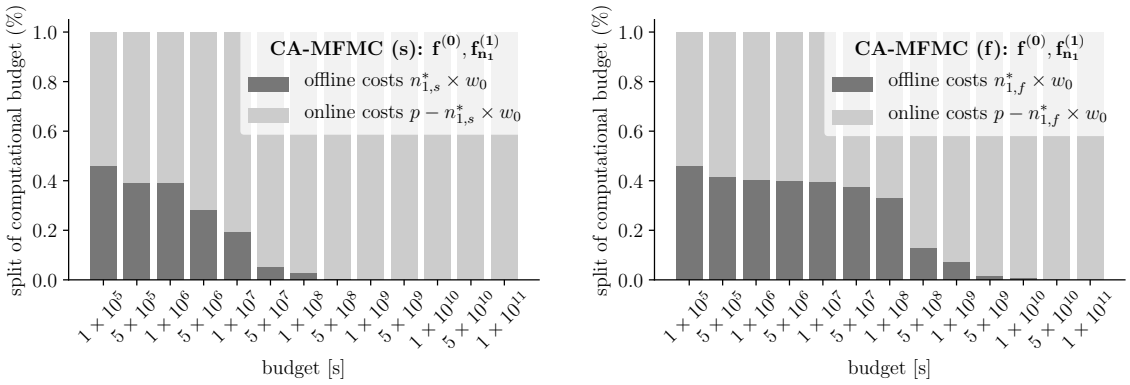

Figure 5.12: CA-MFMC with the 14-dimensional low-fidelity model: Distribution of the computational budgets between refining the low-fidelity model and sampling the high- and low-fidelity model for the slow (left) and fast evaluation functions (right).

in the offline stage decreases until almost all of the computational effort is invested in sampling the high- and low-fidelity model. For example, at $p = 10^{11}$ using fast evaluation operation, only 0.0007% of the budget is used for refining the low-fidelity model.

6 Conclusion

In this thesis, we considered the propagation of uncertainty in plasma micro-turbulence simulations, which play an important part in nuclear fusion power research. Performing UQ in such scenarios can often be challenging due to the high stochastic dimensionality and large computational costs of evaluating the underlying models.

We expanded on the work of [19], in which static MFMC with full- and reduced-dimension low-fidelity models obtained from sensitivity-driven dimension-adaptive sparse grid interpolation was applied to quantify uncertainty in plasma physics problems. We formulated an extension of this methodology by combining the construction of data-driven low-fidelity models of reduced input dimensionality with the application of the CA-MFMC sampling algorithm. Using this approach, context-aware full- and reduced-dimension low-fidelity models are obtained, which are refined using an optimal number of high-fidelity evaluations such that the resulting MFMC estimators are as accurate as possible. At the same time, further speedup is gained by leveraging the low evaluation costs of the reduced-dimension models.

We demonstrated the potential of our method by applying it to a plasma physics scenario with 14 uncertain inputs that simulates the suppression of micro-turbulence by fast particles such as deuterium and Helium-3. As a high-fidelity model, the gyrokinetic plasma turbulence code GENE was employed. We used the dimension-adaptive sparse grid algorithm to first construct a low-fidelity model that depends on all 14 inputs. The sensitivity information revealed by this process was employed to reduce the stochastic dimensionality of further models, first down to nine, and then to six parameters. We computed accuracy and cost rates for all three models via regression from pilot runs and used them in the CA-MFMC algorithm to sequentially find the optimal number of grid points used to construct the sparse grid interpolation models. Moreover, we examined the analytically computed MSE of MFMC mean estimators using different subsets of the low-fidelity models. This information was employed to select subsets of low-fidelity models that lead to accurate estimators, for which we showed the convergence and boundedness of the optimal number of training samples w.r.t. to the computational budget. The effect of the cost rate on CA-MFMC was studied by computing estimators for low-fidelity models using both a slower and a faster evaluation procedure. Furthermore, we applied the CA-MFMC algorithm in practice using the selected model combinations and computed the MSE for a budget of 5×10^6

seconds, which is still part of the pre-asymptotic regime. The most efficient of these estimators achieved a speedup of about two orders of magnitude, which corresponds to a runtime reduction from 48 days on 240 cores using standard MC down to less than six hours using CA-MFMC with the nine-dimensional low-fidelity model. A significant reduction of the analytical MSE was observed in the asymptotic regime w.r.t. the budget, i.e., when the convergence of CA-MFMC estimators has decreased to $\mathcal{O}(p^{-1})$. Once this limit is reached, the most efficient CA-MFMC estimator considered here is around two orders of magnitude more accurate than the examined static MFMC estimators. These results showed the potential of using context-aware reduced-dimension models in MFMC for performing UQ in settings with a large number of input parameters and computationally expensive high-fidelity models.

We conclude with a direction for future research. In the current work, we studied linear gyrokinetic simulations of plasma micro-turbulence. High-fidelity computations in which the nonlinear parts of the gyrokinetic equations are taken into account as well are typically multiple orders of magnitude more computationally expensive, posing a very significant increase of the already high evaluation costs in the realistic linear scenario we considered here. Nonlinear simulations are, however, of great importance since they additionally provide quantitative results as opposed to the more qualitative predictions obtained from linear runs. Based on the promising results we obtained in this work, CA-MFMC with reduced-dimension data-driven low-fidelity models could be a useful tool in the current efforts to find ways for efficient UQ in nonlinear gyrokinetic plasma turbulence simulations as well as other large-scale real-world problems. Examining the power of the framework presented in this thesis for such an application is part of future work.

List of Figures

List of Tables

Bibliography

[1] T. Alsup and B. Peherstorfer, *Context-aware surrogate modeling for balancing approximation and sampling costs in multi-fidelity importance sampling and bayesian inverse problems*, SIAM/ASA Journal on Uncertainty Quantification, (2022).

[2] S. Amaral, D. L. Allaire, and K. Willcox, *Optimal L_2-norm empirical importance weights for the change of probability measure*, Stat. Comput., 27 (2017), pp. 625–643, https://doi.org/10.1007/s11222-016-9644-3, https://doi.org/10.1007/s11222-016-9644-3.

[3] J.-P. Berrut and L. N. Trefethen, *Barycentric lagrange interpolation*, SIAM Review, 46 (2004), pp. 501–517.

[4] A. Brizard and T. S. Hahm, *Foundations of nonlinear gyrokinetic theory*, Reviews of Modern Physics, 79 (2007), p. 421.

[5] H.-J. Bungartz and M. Griebel, *Sparse grids*, Acta Numerica, 13 (2004), p. 147–269, https://doi.org/10.1017/S0962492904000182.

[6] A. Di Siena, R. Bilato, T. Görler, A. B. n. Navarro, E. Poli, V. Bobkov, D. Jarema, E. Fable, C. Angioni, Y. O. Kazakov, R. Ochoukov, P. Schneider, M. Weiland, F. Jenko, and the ASDEX Upgrade Team, *New High-Confinement Regime with Fast Ions in the Core of Fusion Plasmas*, Phys. Rev. Lett., 127 (2021), p. 025002, https://doi.org/10.1103/PhysRevLett.127.025002, https://link.aps.org/doi/10.1103/PhysRevLett.127.025002.

[7] A. Di Siena, T. Görler, H. Doerk, E. Poli, and R. Bilato, *Fast-ion stabilization of tokamak plasma turbulence*, Nuclear Fusion, 58 (2018), p. 054002, https://doi.org/10.1088/1741-4326/aaaf26, https://doi.org/10.1088/1741-4326/aaaf26.

[8] A. Di Siena, T. Görler, E. Poli, R. Bilato, H. Doerk, and A. Zocco, *Resonant interaction of energetic ions with bulk-ion plasma micro-turbulence*, Physics of Plasmas, 26 (2019), p. 052504, https://doi.org/10.1063/1.5087203.

[9] I.-G. Farcaș, *Context-aware Model Hierarchies for Higher-dimensional Uncertainty Quantification*, dissertation, Technische Universität München, München, 2020.

[10] I.-G. Farcaș, T. Görler, H.-J. Bungartz, F. Jenko, and T. Neckel, *Sensitivity-driven adaptive sparse stochastic approximations in plasma microinstability analysis*, J. Comput. Phys., 410 (2020), p. 109394, https://doi.org/10.1016/j.jcp.2020.109394.

[11] I.-G. Farcaș, B. Peherstorfer, T. Neckel, F. Jenko, and H.-J. Bungartz, *Context-aware learning of hierarchies of low-fidelity models for multi-fidelity uncertainty quantification*, 2022, https://doi.org/10.48550/ARXIV.2211.10835, https://arxiv.org/abs/2211.10835.

[12] I.-G. Farcaș, G. Merlo, and F. Jenko, *A general framework for quantifying uncertainty at scale*, Communications Engineering, 1 (2022), pp. 1–11.

[13] I.-G. Farcaș, A. D. Siena, and F. Jenko, *Turbulence suppression by energetic particles: a sensitivity-driven dimension-adaptive sparse grid framework for discharge optimization*, Nuclear Fusion, 61 (2021), p. 056004, https://doi.org/10.1088/1741-4326/abecc8.

[14] X. Garbet, Y. Idomura, L. Villard, and T. Watanabe, *Gyrokinetic simulations of turbulent transport*, Nuclear Fusion, 50 (2010), p. 043002.

[15] T. Gerstner and M. Griebel, *Dimension-adaptive tensor-product quadrature*, Computing, 71 (2003), p. 65–87, https://doi.org/10.1007/s00607-003-0015-5, https://doi.org/10.1007/s00607-003-0015-5.

[16] T. Görler, X. Lapillonne, S. Brunner, T. Dannert, F. Jenko, F. Merz, and D. Told, *The global version of the gyrokinetic turbulence code GENE*, Journal of Computational Physics, 230 (2012), pp. 7053–7071, https://doi.org/10.1016/j.jcp.2011.05.034.

[17] M. Hegland, *Adaptive sparse grids*, ANZIAM J., 44 (2001), https://doi.org/10.21914/anziamj.v44i0.685.

[18] F. Jenko, W. Dorland, M. Kotschenreuther, and B. Rogers, *Electron temperature gradient driven turbulence*, Physics of Plasmas, 7 (2000), pp. 1904–1910.

[19] J. Konrad, I.-G. Farcaș, B. Peherstorfer, A. Di Siena, F. Jenko, T. Neckel, and H.-J. Bungartz, *Data-driven low-fidelity models for multi-fidelity Monte Carlo sampling in plasma micro-turbulence analysis*, Journal of Computational Physics, 451 (2022), p. 110898, https://doi.org/https://doi.org/10.1016/j.jcp.2021.110898.

[20] F. Law, A. J. Cerfon, and B. Peherstorfer, *Accelerating the estimation of collisionless energetic particle confinement statistics in stellarators using multifidelity Monte Carlo*, Nuclear Fusion, (2021), http://iopscience.iop.org/article/10.1088/1741-4326/ac4777.

[21] C. Lemieux, *Monte Carlo and Quasi-Monte Carlo Sampling*, Springer, New York, NY, USA, 2009, https://doi.org/10.1007/978-0-387-78165-5.

[22] N. Metropolis and S. Ulam, *The Monte Carlo Method*, Journal of the American Statistical Association, 44 (1949), pp. 335–341.

[23] R. Miller, M. Chu, J. Greene, Y. Lin-Liu, and R. Waltz, *Noncircular, finite aspect ratio, local equilibrium model*, Physics of Plasmas, 5 (1998), pp. 973–978.

[24] A. Narayan and J. D. Jakeman, *Adaptive Leja Sparse Grid Constructions for Stochastic Collocation and High-Dimensional Approximation*, SIAM Journal on Scientific Computing, 36 (2014), pp. A2952–A2983, https://doi.org/10.1137/140966368, https://doi.org/10.1137/140966368, https://arxiv.org/abs/https://doi.org/10.1137/140966368.

[25] B. L. Nelson, *On control variate estimators*, Computers & Operations Research, 14 (1987), pp. 219–225, https://doi.org/10.1016/0305-0548(87)90024-4.

[26] L. W. T. Ng and K. E. Willcox, *Multifidelity approaches for optimization under uncertainty*, International Journal for Numerical Methods in Engineering, 100 (2014), pp. 746–772, https://doi.org/https://doi.org/10.1002/nme.4761, https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.4761, https://arxiv.org/abs/https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.4761.

[27] B. Peherstorfer, *Multifidelity Monte Carlo Estimation with Adaptive Low-Fidelity Models*, SIAM/ASA Journal on Uncertainty Quantification, 7 (2019), pp. 579–603, https://doi.org/10.1137/17M1159208.

[28] B. Peherstorfer, M. Gunzburger, and K. Willcox, *Convergence analysis of multifidelity monte carlo estimation*, Numerische Mathematik, 139 (2018), pp. 683–707.

[29] B. Peherstorfer, K. Willcox, and M. Gunzburger, *Optimal model management for multifidelity Monte Carlo estimation*, SIAM Journal on Scientific Computing, 38 (2016), pp. A3163–A3194, https://doi.org/10.1137/15M1046472.

[30] B. Peherstorfer, K. Willcox, and M. Gunzburger, *Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization*, SIAM Review, 60 (2018), pp. 550–591, https://doi.org/10.1137/16M1082469.

[31] M. J. Püschel, *Electromagnetic Effects in Gyrokinetic Simulations of Plasma Turbulence*, PhD thesis, Westfälische Wilhelms-Universität Münster, 2009.

[32] E. Qian, B. Peherstorfer, D. O'Malley, V. V. Vesselinov, and K. Willcox, *Multifidelity Monte Carlo Estimation of Variance and Sensitivity Indices*, SIAM/ASA Journal on Uncertainty Quantification, 6 (2018), pp. 683–706, https://doi.org/10.1137/17M1151006.

[33] M. Romanelli, A. Zocco, F. Crisanti, and J.-E. Contributors, *Fast ion stabilization of the ion temperature gradient driven modes in the Joint European Torus hybrid-scenario plasmas: a trigger mechanism for internal transport barrier formation*, Plasma Physics and Controlled Fusion, 52 (2010), p. 045007, https://doi.org/10.1088/0741-3335/52/4/045007, https://dx.doi.org/10.1088/0741-3335/52/4/045007.

[34] T. A. Runkler, *Data Analytics*, Springer Vieweg, Wiesbaden, 2020.

[35] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global sensitivity analysis: the primer*, John Wiley & Sons, 2008.

[36] R. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, Computational Science and Engineering, Society for Industrial and Applied Mathematics, 2013, https://books.google.de/books?id=4c1GAgAAQBAJ.

[37] I. M. Sobol, *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates*, Mathematics and Computers in Simulation, 55 (2001), pp. 271–280, https://doi.org/https://doi.org/10.1016/S0378-4754(00)00270-6, https://www.sciencedirect.com/science/article/pii/S0378475400002706. The Second IMACS Seminar on Monte Carlo Methods.

[38] B. Sudret, *Global sensitivity analysis using polynomial chaos expansions*, Reliability Engineering & System Safety, 93 (2008), pp. 964–979, https://doi.org/https://doi.org/10.1016/j.ress.2007.04.002, https://www.sciencedirect.com/science/article/pii/S0951832007001329. Bayesian Networks in Dependability.

[39] L. N. Trefethen, *Approximation Theory and Approximation Practice, Extended Edition*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2019, https://doi.org/10.1137/1.9781611975949.

[40] D. Xiu, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, Princeton University Press, 2010, http://www.jstor.org/stable/j.ctv7h0skv (accessed 2023-01-02).

[41] H. Yang, Y. Fujii, K. W. Wang, and A. A. Gorodetsky, *Control variate polynomial chaos: Optimal fusion of sampling and surrogates for multifidelity uncertainty quantification*, 2022, https://doi.org/10.48550/ARXIV.2201.10745, https://arxiv.org/abs/2201.10745.