

Data-driven low-fidelity models for multi-fidelity Monte Carlo sampling in plasma micro-turbulence analysis

Julia Konrad^{a,1}, Ionuț-Gabriel Farcaș^{b,1}, Benjamin Peherstorfer^c, Alessandro Di Siena^b,
Frank Jenko^{a,b,d}, Tobias Neckel^a, Hans-Joachim Bungartz^a

^a*Department of Informatics, Technical University of Munich*

^b*Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin*

^c*Courant Institute of Mathematical Sciences, New York University*

^d*Max Planck Institute for Plasma Physics*

Abstract

The linear micro-instabilities driving turbulent transport in magnetized fusion plasmas (as well as the respective nonlinear saturation mechanisms) are known to be sensitive with respect to various physical parameters characterizing the background plasma and the magnetic equilibrium. Therefore, uncertainty quantification is essential for achieving predictive numerical simulations of plasma turbulence. However, the high computational costs of the required gyrokinetic simulations and the large number of parameters render standard Monte Carlo techniques intractable. To address this problem, we propose a multi-fidelity Monte Carlo approach in which we employ data-driven low-fidelity models that exploit the structure of the underlying problem such as low intrinsic dimension and anisotropic coupling of the stochastic inputs. The low-fidelity models are efficiently constructed via sensitivity-driven dimension-adaptive sparse grid interpolation using both the full set of uncertain inputs and subsets comprising only selected, important parameters. We illustrate the power of this method by applying it to two plasma turbulence problems with up to 14 stochastic parameters, demonstrating that it is up to four orders of magnitude more efficient than standard Monte Carlo methods measured in single-core performance, which translates into a runtime reduction from around eight days to one hour on 240 cores on parallel machines.

Keywords: multi-fidelity Monte Carlo sampling, plasma micro-turbulence, sensitivity-driven adaptivity, reduced-dimension low-fidelity models

2000 MSC: 62P35, 65C05, 65Z05, 68U99

Email addresses: `julia.konrad@tum.de` (Julia Konrad), `ionut.farcas@austin.utexas.edu` (Ionuț-Gabriel Farcaș), `pehersto@cims.nyu.edu` (Benjamin Peherstorfer), `alessandro.disiena@austin.utexas.edu` (Alessandro Di Siena), `frank.jenko@ipp.mpg.de` (Frank Jenko), `neckel@in.tum.de` (Tobias Neckel), `bungartz@in.tum.de` (Hans-Joachim Bungartz)

¹These authors contributed equally

1. Introduction

The goal of creating “burning” plasmas in a toroidal magnetic device represents a milestone on the way towards the practical realization of fusion power on Earth. Efforts are currently underway to take this important step in the international ITER experiment, with operation projected to begin within the next few years. One of the key physics problems in this context is the understanding, prediction, and control of plasma turbulence. Turbulent fluctuations of various plasma quantities induce, in particular, high levels of cross-field heat transport, thus determining the so-called energy confinement time. The latter needs to exceed a certain threshold (according to the Lawson criterion) to achieve a burning plasma state. The linear micro-instabilities driving this turbulent transport (as well as the respective nonlinear saturation mechanisms) are known to be very sensitive with respect to various physical parameters characterizing the background plasma and the magnetic equilibrium. Moreover, experimental measurements of these parameters are, of course, subject to uncertainties. Therefore, the development of a predictive capability calls for a systematic quantification of uncertainties. However, the high computational costs of numerical simulations and the potentially large number of stochastic inputs, parameters, and coefficients make uncertainty quantification challenging for these problems [25].

1.1. Towards quantifying uncertainty in plasma turbulence problems

To overcome the challenges of quantifying uncertainty in real-world problems such as plasma micro-turbulence simulations, we present a framework based on multi-fidelity Monte Carlo (MFMC) sampling [34, 38, 39]. MFMC reduces the mean-squared error of standard Monte Carlo (MC) estimators by making use of control variates [33]. Specifically, in MFMC, the control variates are standard MC estimators given by low-fidelity models, which are computationally cheap approximations of the underlying high-fidelity model. Examples of low-fidelity models include reduced-physics models, e.g., coarse-grid approximations, data-fit low-fidelity models, such as interpolation, spectral projection or regression, or projection-based reduced models, including the proper orthogonal decomposition or the reduced basis method; see [40] and the references therein for more details.

Control variates have been used to reduce the error of MC estimators in kinetic models such as Boltzmann’s equation in [13]. The work [13] was generalized in [14], where multiple control variates are employed. In addition, in [28], MFMC was employed in a benchmark scenario prominent in plasma micro-turbulence simulations [15]. In [17, 37], a context-aware version of MFMC is formulated, in which data-driven low-fidelity models are explicitly constructed to being used together with the high-fidelity model; see also [1].

Besides sampling-based methods, stochastic collocation and quasi-Monte Carlo have also been used to address the challenges of uncertainty in plasma physics. In [29], generalized polynomial chaos and quasi-Monte Carlo methods were employed in multiscale plasma fusion simulations. Moreover, [45] used nonintrusive stochastic collocation methods based on non-uniform interpolation sequences in a drift-wave turbulence study from a linear plasma experiment. In addition, adaptive sparse grids [5] were considered in [47] to compute the runaway probability of electrons.

1.2. Novelty and contribution of proposed data-driven multi-fidelity approach for uncertainty propagation in plasma turbulence problems

In this work, we employ the well established plasma turbulence code GENE [27] which solves the gyrokinetic Vlasov-Maxwell coupled system of partial differential equations in 5D state space, focusing on calculating the linear growth rates of the dominant micro-instabilities. For constructing low-fidelity models, we primarily build on approximations with adaptive sparse grids that have been shown to be effective in quantifying uncertainties in complex applications: In [16, 17], a sensitivity-driven dimension-adaptive sparse grid approximation method was introduced to efficiently compute quantities of interest such as expectation, variance and Sobol’ indices for global sensitivity analysis [43]. The sensitivity-driven approach exploits the fact that typically in most real-world problems (i) the intrinsic stochastic dimension is lower than the ambient dimension and (ii) the uncertain inputs are anisotropically coupled. In [16, 17], the sensitivity-driven approach was applied to two plasma micro-turbulence test cases, including a scenario with 12 uncertain inputs based on an ASDEX Upgrade discharge [20]. The same approach was recently used in a study concerning turbulence stabilization via energetic particles with 21 stochastic parameters [19].

In our MFMC approach, we leverage the sensitivity-driven sparse-grid models as low-fidelity models. Furthermore, we exploit sensitivity information to reduce the stochastic dimension. This has also been studied in [18] in the context of multilevel stochastic collocation for fluid-structure interaction simulations. In this paper, we rely on the feature that the construction of sparse-grid models reveals detailed information about the sensitivity of each stochastic input and use this information to additionally construct low-fidelity models with lower input dimensions, i.e., models in which only few of the inputs are varied while the remaining inputs are fixed to their mean value.

To summarize, the present paper is novel in at least two ways: from a methodological perspective, we employ data-driven, structure-exploiting low-fidelity models in MFMC with the goal to make MFMC feasible in computationally expensive, real-world problems for which standard sampling methods are unfeasible. From an application scenario point of view, we employ our MFMC approach in large-scale plasma micro-instability analysis, a problem of high practical relevance.

1.3. Outline of this paper

This paper is organized as follows. In Section 2, we explain the notation and summarize the foundations of multi-fidelity uncertainty propagation. In Section 3, we describe the hierarchy of models used in this paper. Our numerical results in two plasma micro-turbulence scenarios are presented in Section 4. The first scenario is a modified benchmark in which we consider either three or eight uncertain parameters (Section 4.1). The second scenario discussed in Section 4.2 is concerned with turbulence stabilization via energetic particles with 14 uncertain inputs, for which one realization of the stochastic inputs requires 240 cores and a total runtime of about 3.2 hours. We close with conclusions in Section 5.

2. Multi-fidelity uncertainty propagation

In this section, we introduce notation and provide a brief summary of the mathematical foundations of multi-fidelity uncertainty propagation with MFMC.

2.1. Uncertainty propagation

Let $f^{(0)} : \mathcal{X} \rightarrow \mathcal{Y}$ be a bounded and measurable function with respect to the Lebesgue measure and the Borel σ -algebra on \mathbb{R} . The domain $\mathcal{X} \subset \mathbb{R}^d$ is the set of d -dimensional inputs $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_d]^T$ and the domain $\mathcal{Y} \subset \mathbb{R}$ is the range of scalar-valued outputs $y = f^{(0)}(\boldsymbol{\theta})$. In the following, we consider the situation where $f^{(0)}$ is computationally expensive to evaluate because evaluating $f^{(0)}$ entails large-scale, high-fidelity numerical simulations. Denote the computational costs of evaluating $f^{(0)}$ at an input $\boldsymbol{\theta} \in \mathcal{X}$ as $w_0 \gg 0$.

Let now $\boldsymbol{\Theta} = [\Theta_1, \dots, \Theta_d]^T$ be a random vector with image \mathcal{X} and with probability density function π . The goal of uncertainty propagation is to estimate quantities of interest such as the expected value of the output random variable $f^{(0)}(\boldsymbol{\Theta})$

$$\mu_0 = \mathbb{E}[f^{(0)}(\boldsymbol{\Theta})] = \int_{\mathcal{X}} f^{(0)}(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (2.1)$$

and its variance

$$\sigma_0^2 = \text{Var}[f^{(0)}(\boldsymbol{\Theta})] = \mathbb{E}[(f^{(0)}(\boldsymbol{\Theta}))^2] - \mathbb{E}[f^{(0)}(\boldsymbol{\Theta})]^2. \quad (2.2)$$

MC estimators of the expected value and variance of $f^{(0)}(\boldsymbol{\Theta})$ are

$$\hat{E}_n^{(0)} := \frac{1}{n} \sum_{i=1}^n f^{(0)}(\boldsymbol{\theta}_i), \quad \widehat{\text{Var}}_n^{(0)} := \frac{1}{n-1} \sum_{i=1}^n (f^{(0)}(\boldsymbol{\theta}_i) - \hat{E}_n^{(0)})^2, \quad (2.3)$$

respectively, where $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_n$ are n independent and identically distributed (i.i.d.) samples from $\boldsymbol{\Theta}$. The costs of the MC estimator $\hat{E}_n^{(0)}$ of $\mathbb{E}[f^{(0)}(\boldsymbol{\Theta})]$ are $p = nw_0$ because $f^{(0)}$ is evaluated at n realizations. The mean-squared error (MSE) of the estimator $\hat{E}_n^{(0)}$ is

$$\text{MSE}[\hat{E}_n^{(0)}] = \frac{\sigma_0^2}{p} w_0. \quad (2.4)$$

Because we consider the situation in which the evaluation costs of $f^{(0)}$ are large, the slow decay of the MSE in (2.4) with respect to the budget p makes the MC estimator $\hat{E}_n^{(0)}$ computationally prohibitive. For example, in Section 4.2, one plasma micro-turbulence simulation requires, on average, a w_0 of about 3.2 hours on 240 cores in total. A large number of such simulations is not feasible.

2.2. Multi-fidelity Monte Carlo sampling

The MFMC method proposed in [34, 39] exploits that in many problems in science and engineering we either have available or can construct low-fidelity models that approximate $f^{(0)}$ with lower computational costs than w_0 . Let us assume we have available k low-fidelity models $f^{(1)}, f^{(2)}, \dots, f^{(k)}$. MFMC exploits the lower computational cost of the low-fidelity

models to obtain an estimator that has lower costs than a standard MC estimator with the same MSE. MFMC guarantees unbiasedness even if the approximation error of the low-fidelity models cannot be bounded rigorously. The quality of low-fidelity models for MFMC is ascertained via their Pearson correlation coefficient with respect to $f^{(0)}$,

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

where σ_j^2 denotes the variance of model j for $j = 0, 1, \dots, k$ and Cov is the covariance.

Let $m_j \in \mathbb{N}$ for $j = 0, 1, \dots, k$ denote the number of evaluations of model $f^{(j)}$ such that

$$1 \leq m_0 < m_1 < \dots < m_k. \quad (2.5)$$

That is, we ensure that the high-fidelity model is evaluated at least once. Consider m_k i.i.d. realizations drawn from the input density, $\pi(\boldsymbol{\theta})$:

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

Model $f^{(j)}$ for $j = 0, 1, \dots, k$ is evaluated at the first m_j samples $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_{m_j}$ to obtain

$$f^{(j)}(\boldsymbol{\theta}_1), f^{(j)}(\boldsymbol{\theta}_2), \dots, f^{(j)}(\boldsymbol{\theta}_{m_j}). \quad (2.6)$$

We compute the following standard MC estimators for expectation and variance:

$$\hat{E}_{m_j}^{(j)} = \frac{1}{m_j} \sum_{m=1}^{m_j} f^{(j)}(\boldsymbol{\theta}_m), \quad \widehat{\text{Var}}_{m_j}^{(j)} = \frac{1}{m_j - 1} \sum_{m=1}^{m_j} (f^{(j)}(\boldsymbol{\theta}_m) - \hat{E}_{m_j}^{(j)})^2. \quad (2.7)$$

Additionally, for the low-fidelity models $f^{(j)}$ with $j = 1, 2, \dots, k$ we compute the estimators

$$\hat{E}_{m_{j-1}}^{(j)} = \frac{1}{m_{j-1}} \sum_{m=1}^{m_{j-1}} f^{(j)}(\boldsymbol{\theta}_m), \quad \widehat{\text{Var}}_{m_{j-1}}^{(j)} = \frac{1}{m_{j-1} - 1} \sum_{m=1}^{m_{j-1}} (f^{(j)}(\boldsymbol{\theta}_m) - \hat{E}_{m_{j-1}}^{(j)})^2. \quad (2.8)$$

Note that the estimators in (2.7) use the full set of evaluations of model $f^{(j)}$, while the estimators in (2.8) use the first m_{j-1} evaluations. This makes the estimators in (2.8) dependent on the estimators in (2.7).

The MFMC estimator for the expectation of the high-fidelity model is given by

$$\hat{E}^{\text{MFMC}} := \hat{E}_{m_0}^{(0)} + \sum_{j=1}^k \alpha_j (\hat{E}_{m_j}^{(j)} - \hat{E}_{m_{j-1}}^{(j)}) \quad (2.9)$$

and the estimator for the variance is

$$\widehat{\text{Var}}^{\text{MFMC}} := \widehat{\text{Var}}_{m_0}^{(0)} + \sum_{j=1}^k \alpha_j (\widehat{\text{Var}}_{m_j}^{(j)} - \widehat{\text{Var}}_{m_{j-1}}^{(j)}) \quad (2.10)$$

where $\alpha_1, \alpha_2, \dots, \alpha_k \in \mathbb{R}$.

Let w_j denote the runtime of the low-fidelity model $f^{(j)}$ for $j = 1, 2, \dots, [w_0, w_1, \dots, w_k]^T$. Given a budget $p := \sum_{j=0}^k w_j m_j$, in [39, Theorem 3.4] it is shown that if

$$1 =: \rho_0 > \rho_1 > \rho_2 > \dots > \rho_k \text{ and } \frac{w_{j-1}}{w_j} > \frac{\rho_{j-1}^2 - \rho_j^2}{\rho_j^2 - \rho_{j+1}^2}, \quad j = 1, 2, \dots, k, \quad (2.11)$$

then the MSE of the MFMC mean estimator (2.9) is minimized given the budget p , and

$$\alpha_j^* = \frac{\rho_j \sigma_1}{\sigma_j}, \quad m_0^* = \frac{p}{\mathbf{w}^T \mathbf{r}^*}, \quad m_j^* = r_j^* m_0^*, \quad j = 1, 2, \dots, k, \quad (2.12)$$

where $\mathbf{r}^* := [r_0^*, r_1^*, \dots, r_k^*]^T$, $r_j^* = \sqrt{w_0(\rho_j^2 - \rho_{j+1}^2)/w_j(1 - \rho_1^2)}$. Using the optimal number of model evaluations $m_0^*, m_1^*, \dots, m_k^*$ and coefficients $\alpha_1^*, \alpha_2^*, \dots, \alpha_k^*$, the MSE of \hat{E}^{MFMC} is

$$\text{MSE}[\hat{E}^{\text{MFMC}}] = \frac{\sigma_0^2}{p} \left(\sum_{j=0}^k \sqrt{w_j(\rho_j^2 - \rho_{j+1}^2)} \right)^2. \quad (2.13)$$

If we compare the above MSE with the MSE of the standard MC mean estimator (2.4), it holds that $\text{MSE}[\hat{E}^{\text{MFMC}}] < \text{MSE}[\hat{E}^{\text{MC}}]$ if and only if

$$\gamma := \left(\sum_{j=0}^k \sqrt{\frac{w_j}{w_0}(\rho_j^2 - \rho_{j+1}^2)} \right)^2 < 1. \quad (2.14)$$

The above formula gives a means to determine which low-fidelity models lead to an MFMC estimator with a smaller MSE than the MSE of the standard MC estimator for which the same budget, p , was used. In [39, Algorithm 1], a model selection algorithm is proposed, which, given a set of low-fidelity models, selects the subset that leads to the MFMC estimator with the lowest MSE.

In the specific situation of two models and that the expected value of the surrogate-model output $f^{(1)}(\theta)$ is known in closed form, it can be used in (2.9) directly instead of the estimator $\hat{E}_{m_1}^{(1)}$. This specific situation is investigated with control functionals in [36] and in the context of MFMC in the asymptotic limit in [37]. However, for typical surrogate models, a closed form of the expected value is unavailable and one has to rely on the estimator $\hat{E}_{m_1}^{(1)}$.

We point out that the standard MFMC algorithm of [39] is for the estimation of the mean of the high-fidelity model. That is, given a budget p and assuming that (2.11) holds true, the minimization of the MSE of the mean estimator leads to the optimal parameters in (2.12). These parameters are not necessarily optimal for estimating the variance of the high-fidelity model with the MFMC estimator (2.10). Nevertheless, in [41], it was shown that using the mean-optimal parameters for estimating the variance too leads to a comparable MSE as when directly minimizing the MSE of the variance estimator, which is significantly more challenging since the variance is a nonlinear operator. Thus, the realizations of the model outputs can be used to compute both mean and variance estimators.

Remark 1. The ingredients in MFMC are the correlation coefficients, ρ_j , and evaluation costs, w_j , of the models in the hierarchy. In some problems the quantities are readily available from physical and theoretical insights. If the quantities are unavailable, then they can be estimated from typically a small number of pilot samples [39]. We will demonstrate with our numerical examples in Section 4.2 that the pilot samples can be re-used in the MFMC estimator. Re-using the pilot samples can introduce a bias; however, the mean-squared error of the estimator is typically dominated by the limited number of samples rather than the bias due to re-using pilot samples.

3. Plasma micro-turbulence models: high-fidelity and data-driven low-fidelity models

In this section, we present the hierarchy of models used in our numerical experiments. In Section 3.1, we summarize the gyrokinetic theory for plasma micro-turbulence simulations, which is the basis for the high-fidelity model employed in this work. The main methodology used to construct data-driven low-fidelity models via sensitivity-driven sparse grids is presented in Section 3.2. For a broader overview, we also consider a second type of data-driven low-fidelity models based on deep neural networks, which we summarize in Section 3.3.

3.1. High-fidelity model

We build on gyrokinetic theory [4] to assess plasma micro-turbulence, which reduces the dimension of the state space of classical kinetic models from six (three positions, three velocities) to five (three positions, two velocities). At the same time, a number of small—and irrelevant (for the specific scenario at hand)—space-time scales are removed from the problem, leading to savings of several orders of magnitude in terms of computational costs. Throughout this paper the gyrokinetic equations are solved numerically with the simulation code GENE [27]. In the gyrokinetic code GENE each particle species, s , is characterized by a distribution function, which is split into a static background (π_0) – often assumed a Maxwellian distribution function – and a fluctuating part (π_1)

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

where t is time, $\mathbf{R} = (R_x, R_y, R_z)$ describes the position of the gyrocenter in real space, v_{\parallel} denotes the velocity parallel to the background magnetic field, and $\mu_m := m_s v_{\perp}^2 / 2B$ is the magnetic moment, where m_s denotes the mass of species s , v_{\perp} is the perpendicular component of the velocity and B denotes the magnetic field. Here, it is assumed that $\pi_1/\pi_0 \sim \epsilon_{\delta} \ll 1$. The dynamics of the perturbed part of the distribution function $\pi_{1,s}$ are modelled by the first order (in the expansion parameter ϵ_{δ}) gyrokinetic Vlasov equation with collisions:

$$\begin{aligned} C(\pi_s, \pi_{s'}) = & \dot{\pi}_{1,s} + \mathbf{v}_{1,c} \cdot \left(\nabla \pi_{0,s} - \frac{\mu_m}{m_s v_{\parallel}} \frac{\partial \pi_{0,s}}{\partial v_{\parallel}} \nabla B_0 \right) + \left(v_{\parallel} \hat{\mathbf{b}}_0 + \mathbf{v}_{1,c} \right) \cdot \\ & \left(\nabla \pi_{1,s} + \frac{1}{m_s v_{\parallel}} (q_s \mathbf{E}_1 - \mu_m \nabla B_{1,\parallel}) \frac{\partial \pi_{0,s}}{\partial v_{\parallel}} + \frac{1}{m_s v_{\parallel}} (q_s \mathbf{E}_1 - \mu_m \nabla (B_0 + B_{1,\parallel})) \frac{\partial \pi_{1,s}}{\partial v_{\parallel}} \right), \end{aligned} \quad (3.1)$$

where $C(\pi_s, \pi_{s'})$ is a collision operator, B_0 the background magnetic field, \mathbf{E}_1 the perturbed part of the electric field and $\mathbf{v}_{1,c}$ the drift velocity due to the fields. To self-consistently evolve the perturbed part of the distribution function $\pi_{1,s}$, we also have to advance three scalar quantities characterizing the electromagnetic fields by solving Maxwell's equations: the electrostatic potential, Φ_1 , the parallel component of the vector potential, $A_{1,\parallel}$, which is linked to the perpendicular magnetic field perturbations, and the parallel component of the magnetic field perturbations, $B_{1,\parallel}$, i.e.,

$$\begin{aligned} \text{Poisson's equation: } \quad \nabla \cdot \mathbf{E}_1 &= -\nabla^2 \Phi_1 = 4\pi \sum_s q_s n_{1s}, \\ \text{Ampère's law for } A_{1,\parallel}: \quad -\nabla^2 A_{1,\parallel} &= \frac{4\pi}{c} \sum_s n_{1s} q_s u_{1s,\parallel}, \\ \text{Ampère's law for } B_{1,\parallel}: \quad (\nabla \times \mathbf{B}_1)_\perp &= \frac{4\pi}{c} \sum_s n_{1s} \mathbf{u}_{1s,\perp}, \end{aligned} \tag{3.2}$$

where q_s is the charge, c is the speed of light, and n_{1s} is the 0th space moment, $u_{1s,\parallel}$ the 1st order velocity moment in v_\parallel of π_s and $\mathbf{u}_{1s,\perp}$ the 1st order velocity moment in μ_m of π_s . The velocity integrals of π_1 are computed in the particle coordinates, while the Vlasov equation is written in the gyrocenter coordinates [7]. The system of equations (3.1)-(3.2) defines the partial differential Vlasov-Maxwell equations solved in GENE.

In GENE, the gyrokinetic Vlasov-Maxwell system of nonlinear PDEs is solved using the method of lines: the spatial operators are discretized using either finite differences on equidistant points or spectral (Fourier) decompositions, while the resulting system of ODEs is efficiently integrated using explicit methods, such as fourth-order Runge-Kutta.

In this work, we are interested in linear (in phase space variables $\mathbf{R}, v_\parallel, \mu_m$) gyrokinetic simulations, which can be used to characterize the underlying micro-instabilities. The equations employed in linear simulations are obtained from the gyrokinetic Vlasov-Maxwell system (3.1)-(3.2) by simply neglecting all nonlinear terms. In these simulations, GENE is run in the so-called flux-tube limit, which allows having periodic boundary conditions over the radial (x) direction [7]. The generic form of linear gyrokinetic simulations reads:

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

where $\boldsymbol{\pi}_s$ is a vector holding the discretized five-dimensional distribution function of species s . The discrete form of the operator \mathbb{O}_{lin} is a matrix, O_{lin} , which leads to

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

To obtain the above linear equation, GENE employs Fourier decompositions in the radial (x) direction and equidistant points along the field line, i.e., in the z direction; in flux-tube simulations, only one point is used in the bi-normal (ky) direction. In velocity space, equidistant points are applied for the parallel velocity, while Gauss-Laguerre quadrature points are used to discretize the magnetic moment.

The high-fidelity model used in this work computes the growth rate or spectral abscissa, i.e., the maximum real part over the spectrum, which we denote by γ_1 :

$$f^{(0)}(\boldsymbol{\Theta}) = \gamma_1(\boldsymbol{\Theta}) \quad (3.3)$$

which characterizes the dominant micro-instability mode. The entries in $\boldsymbol{\Theta}$ are parameters characterizing the particle species, such as temperature, density and their (logarithmic) gradients, parameters characterizing the magnetic geometry or parameters associated to electromagnetic effects, if electrons are used in the gyrokinetic equations etc. In this work, $\boldsymbol{\Theta}$ has up to 14 components, including the safety factor, q , which describes the relationship of the number of toroidal turns of a magnetic field line to number of poloidal turns; and its derivative, the magnetic shear, \hat{s} ; β , the ratio of kinetic to magnetic pressure inside the plasma; the normalized collision frequency, ν_e ; and parameters characterizing the underlying species, such as their density, n_s , and its negative logarithmic gradient, ω_{n_s} , and their temperature, T_s , and its negative logarithmic gradient, ω_{T_s} . For more details about the gyrokinetic equations solved by GENE, we refer the reader to [23] and the references therein.

3.2. Low-fidelity model: Sensitivity-driven dimension-adaptive sparse grid interpolation

In this section, we summarize the sensitivity-driven dimension-adaptive sparse grid interpolation approach formulated in [16, 17], which is our main methodology used to create data-driven low-fidelity models for MFMC. Our notation is similar to [16, 17].

3.2.1. Interpolation on sparse grids

The sparse grid interpolation [5] low-fidelity model reads

$$\mathcal{U}_{\mathcal{L}}^d[f^{(0)}] = \sum_{\boldsymbol{\ell} \in \mathcal{L}} \Delta_{\boldsymbol{\ell}}^d[f^{(0)}], \quad (3.4)$$

where $\boldsymbol{\ell} = (\ell_1, \ell_2, \dots, \ell_d) \in \mathbb{N}^d$ denotes a multiindex, $\mathcal{L} \subset \mathbb{N}^d$ is a multiindex set, and

$$\Delta_{\boldsymbol{\ell}}^d[f^{(0)}] = \sum_{\mathbf{z} \in \{0,1\}^d} (-1)^{|\mathbf{z}|_1} \mathcal{U}_{\boldsymbol{\ell}-\mathbf{z}}^d[f^{(0)}] \quad (3.5)$$

are the so-called hierarchical surpluses, where $|\mathbf{z}|_1 := \sum_{i=1}^d z_i$. The surpluses are computed from full-grid operators, $\mathcal{U}_{\boldsymbol{\ell}}^d$, which are tensorizations of one-dimensional approximations:

$$\mathcal{U}_{\boldsymbol{\ell}}^d[f^{(0)}] = \left(\bigotimes_{i=1}^d \mathcal{U}_{\ell_i}^i \right) [f^{(0)}]. \quad (3.6)$$

Therefore, the sparse grid interpolation low-fidelity model (3.4) is a linear combination of hierarchical surpluses (3.5), which are computed from tensorizations (3.6) of one-dimensional operators, $\mathcal{U}_{\ell_i}^i$. To ensure that (3.6) can be computed, we need the following two assumptions: the image, \mathcal{X} , of the stochastic inputs needs to have a product structure:

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

which implies that the input density, π , needs to have a product structure as well:

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

where \mathcal{X}_i is the image of the density π_i associated with input θ_i . The latter assumptions means that the stochastic inputs need to be independent. This can be relaxed if a (possibly nonlinear) transformation is used, e.g., a transport map [30].

In this work, $\mathcal{U}_{\ell_i}^i$ are interpolation operators based on Lagrange polynomials (see, e.g., [2, 6, 16, 17, 35]), constructed using weighted (L)-Leja points [26, 32]. Let $g : \mathcal{X}_i \rightarrow \mathbb{R}$ be a univariate function and let $\mathbb{P}_{P_{\ell_i}}$ be the space of univariate polynomials of degree $P_{\ell_i} \in \mathbb{N}$ for $i = 1, 2, \dots, d$. Univariate Lagrange interpolation is defined as:

$$\mathcal{U}_{\ell_i}^i : C^0(\mathcal{X}_i) \rightarrow \mathbb{P}_{P_{\ell_i}}, \quad \mathcal{U}_{\ell_i}^i[g] := \sum_{n=1}^{\ell_i} g(\theta_n) L_n(\theta), \quad (3.7)$$

where $\{\theta_n\}_{n=1}^{\ell_i}$ are weighted (L)-Leja points computed w.r.t. the density π_i :

$$\begin{aligned} \theta_1 &= \operatorname{argmax}_{\theta \in \mathcal{X}_i} \pi_i(\theta) \\ \theta_n &= \operatorname{argmax}_{\theta \in \mathcal{X}_i} \pi_i(\theta) \prod_{m=1}^{n-1} |(\theta - \theta_m)|, \quad n = 2, 3, \dots, \ell_i, \end{aligned}$$

and $\{L_n(\theta)\}_{n=1}^{\ell_i}$ are Lagrange polynomials of degree $n - 1$ satisfying the interpolation condition $L_n(\theta_m) = \delta_{nm}$, where δ_{nm} is Kronecker's delta function. For improved numerical stability, we implement (3.7) in terms of the barycentric formula (see, e.g., [2]). It follows that the multivariate interpolation operator (3.6) is determined as

$$\mathcal{U}_{\boldsymbol{\ell}}^d[f^{(0)}] = \sum_{\mathbf{p} \in \mathcal{P}_{\boldsymbol{\ell}}} f^{(0)}(\boldsymbol{\theta}_{\mathbf{p}}) L_{\mathbf{p}}(\boldsymbol{\theta}), \quad (3.8)$$

where $L_{\mathbf{p}}(\boldsymbol{\theta}) := \prod_{i=1}^d L_{p_i}(\theta_i)$ and $\mathcal{P}_{\boldsymbol{\ell}} := \{\mathbf{p} \in \mathbb{N}^d : \mathbf{0} \leq \mathbf{p} \leq \mathbf{P}_{\boldsymbol{\ell}} := (\ell_1 - 1, \ell_2 - 1, \dots, \ell_d - 1)\}$.

3.2.2. Sensitivity-driven dimension-adaptivity

To fully define (3.4), we need to specify the multiindex set, \mathcal{L} . \mathcal{L} is critical for the computational efficiency of constructing the sparse grid approximation: the larger the cardinality of \mathcal{L} is, the larger the cost of finding the approximation. To reduce this cost, we employ sensitivity-driven dimension-adaptivity.

An important ingredient of dimension-adaptive algorithms [21, 24] is a refinement indicator, $\varepsilon(\boldsymbol{\ell})$, which is used to decide which multiindices to refine in an adaptive step. In the sensitivity-driven procedure of [16, 17], $\varepsilon(\boldsymbol{\ell})$ is computed in terms of unnormalized Sobol' indices [43]: we first find the equivalent spectral projection representation of the multivariate interpolation operators (3.8):

$$\mathcal{U}_{\boldsymbol{\ell}}^d[f^{(0)}] = \sum_{\mathbf{p} \in \mathcal{P}_{\boldsymbol{\ell}}} f^{(0)}(\boldsymbol{\theta}_{\mathbf{p}}) L_{\mathbf{p}}(\boldsymbol{\theta}) = \sum_{\mathbf{p} \in \mathcal{P}_{\boldsymbol{\ell}}} c_{\mathbf{p}} \Phi_{\mathbf{p}}(\boldsymbol{\theta}), \quad (3.9)$$

where $\Phi_{\mathbf{p}}(\boldsymbol{\theta}) := \prod_{i=1}^d \Phi_i(\theta_i)$ are orthonormal polynomial w.r.t. π and $c_{\mathbf{p}}$ are the spectral coefficients. For example, if π is the uniform distribution, $\Phi_{\mathbf{p}}(\boldsymbol{\theta})$ are the Legendre polynomials. To determine the spectral coefficients $c_{\mathbf{p}}$, we simply solve a 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 weighted (L)-Leja points $\boldsymbol{\theta}_k$ associated to the multiindex $\boldsymbol{\ell}$ (see [16, 17]). Next, we rewrite the hierarchical interpolation surpluses (3.5) in terms of spectral projections:

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

It follows that

$$\|\Delta_{\boldsymbol{\ell}}^d[f^{(0)}]\|_{L^2}^2 = \sum_{\mathbf{p} \in \mathcal{P}_{\boldsymbol{\ell}}} \Delta c_{\mathbf{p}}^2 = \Delta c_{\mathbf{0}}^2 + \sum_{\mathbf{p} \in \mathcal{P}_{\boldsymbol{\ell}} \setminus \{\mathbf{0}\}} \Delta c_{\mathbf{p}}^2 = (\mathbb{E}_{\boldsymbol{\ell}}[f^{(0)}])^2 + \Delta \text{Var}_{\boldsymbol{\ell}}[f^{(0)}]. \quad (3.10)$$

The first equality in (3.10) is due to Parseval's identity, whereas the last equality follows from the orthonormality of the spectral projection basis [46]. From (3.10), we therefore obtain the variance corresponding to each subspace.

The variance represents global information in the underlying subspace and it therefore does not provide any information about the individual parameters, their interaction, or which of these are stochastically important. To obtain such information, we decompose $\Delta \text{Var}_{\boldsymbol{\ell}}[f^{(0)}]$ further by exploiting the equivalence between spectral projection and Sobol' decompositions [44]:

$$\Delta \text{Var}_{\boldsymbol{\ell}}[f^{(0)}] = \sum_{i=1}^d \Delta \text{Var}_{\boldsymbol{\ell}}^i[f^{(0)}] + \Delta \text{Var}_{\boldsymbol{\ell}}^{\text{inter}}[f^{(0)}], \quad (3.11)$$

where

$$\Delta \text{Var}_{\boldsymbol{\ell}}^i[f^{(0)}] := \sum_{\mathbf{p} \in \mathcal{J}_i} \Delta c_{\mathbf{p}}^2, \quad \mathcal{J}_i := \{\mathbf{p} \in \mathcal{P}_{\boldsymbol{\ell}} : \mathbf{p}_i \neq 0 \wedge \mathbf{p}_j = 0, \forall j \neq i\}, \quad (3.12)$$

$$\Delta \text{Var}_{\boldsymbol{\ell}}^{\text{inter}}[f^{(0)}] := \sum_{\mathbf{p} \in \mathcal{J}_{\text{inter}}} \Delta c_{\mathbf{p}}^2, \quad \mathcal{J}_{\text{inter}} := \{\mathbf{p} \in \mathcal{P}_{\boldsymbol{\ell}} : |\mathbf{p}|_0 \geq 1\}, \quad (3.13)$$

where $|\mathbf{p}|_0$ denotes the number of non-zero entries in \mathbf{p} . Thus, we can decompose the variance in each subspace associated to a multiindex $\boldsymbol{\ell}$ into variances associated to each individual parameters (3.12) and the variance involving all interactions between the d inputs (3.13). Moreover, from [44], we know that the terms in (3.11) are unnormalized Sobol' indices [43], which ascertain the stochastic importance of each input and of their interaction. Thus, the decomposition (3.11) allows us to have detailed information about the individual inputs as well as their interaction.

We define the sensitivity-driven refinement indicator as $\varepsilon(\boldsymbol{\ell}) := s_{\boldsymbol{\ell}}$, where $s_{\boldsymbol{\ell}}$ is an integer (for each subspace) called the sensitivity index [16, 17]. We compute $s_{\boldsymbol{\ell}}$ as follows. Initially, $s_{\boldsymbol{\ell}} = 0$. Based on user-defined $d + 1$ tolerances $\boldsymbol{\tau} := (\tau_1, \tau_2, \dots, \tau_d, \tau_{d+1})$ – one for each individual direction (d in total) and the last one for all interactions – we compare each term

in (3.11) with the prescribed tolerances. If the tolerance is not exceeded, we increase s_ℓ by one. Thus, s_ℓ takes values between zero and $d + 1$. Denote by \mathbf{k} the multiindex associated with the subspace having the largest sensitivity score. In each refinement step, this subspace is refined by adding all forward neighbors $\mathbf{k} + \mathbf{e}_i$, $i = 1, 2, \dots, d$ of \mathbf{k} , where \mathbf{e}_i denotes the i th unit vector in \mathbb{R}^d , which keeps the multiindex set downward closed, i.e., without holes [21]. If two or more sensitivity indices are equal, we select the multiindex with the largest variance $\Delta \text{Var}_\ell[f^{(0)}]$. The algorithm stops when either (i) all tolerances $\boldsymbol{\tau}$ are reached, (ii) a prescribed maximum level, L_{\max} , is reached, or (iii) there are no more multiindices to be refined. We note that the summarized refinement procedure is inherently sequential. In each refinement step, at most d new grid points are added, i.e., at most d simulations can be performed in parallel per refinement step. Nevertheless, for functions that are smooth, the method is usually efficient, hence compensating for the restricted outer-loop parallelism. In addition, for problems for which a significant amount of resources is needed for a single simulation, performing multiple such simulations in parallel might not be feasible.

The sensitivity-driven dimension-adaptive sparse grid algorithm was originally designed for single-fidelity uncertainty propagation settings [16] with the goal of estimating the expectation, variance or Sobol' indices for sensitivity analysis of outputs of interest. Nevertheless, since the algorithm is based on interpolation, it implicitly provides a low-fidelity approximation of the high-fidelity model as well. Therefore, in this work we employ the sensitivity-driven algorithm to create structure-exploiting low-fidelity models in multi-fidelity settings. Moreover, since the algorithm provides detailed information about the stochastic parameters' sensitivities, we exploit this information to additionally create reduced-dimension low-fidelity models by fixing the unimportant stochastic parameters to a deterministic value, e.g., their expectation. We illustrate the potential of the sensitivity-driven algorithm to creating full- and reduced-dimension low-fidelity models in the following example.

3.2.3. Illustrative example

Consider a high-fidelity model $f^{(0)} : [0, 1]^8 \rightarrow \mathbb{R}$ that depends on eight uniformly distributed random variables $\theta_1, \theta_2, \dots, \theta_8$:

$$f^{(0)}(\boldsymbol{\theta}) = 1 + \cos(\pi + \theta_1 + 0.55\theta_2 + 0.05\theta_3 + 0.8\theta_4 + 0.02\theta_5 + 0.001\theta_6 + 0.1\theta_7 + 0.0005\theta_8). \quad (3.14)$$

Initially, we use the sensitivity-driven algorithm to construct an eight-dimensional low-fidelity model using tolerances $\boldsymbol{\tau} = 10^{-8} \cdot \mathbf{1}_9$. To exploit the algorithm to construct reduced-dimension low-fidelity models, we ascertain the sensitivity of the eight stochastic parameters provided by the 8D low-fidelity model. The eight total Sobol' indices read

$$\begin{aligned} \hat{S}_1^T &= 0.5179, & \hat{S}_2^T &= 0.1542, & \hat{S}_3^T &= 0.0012, & \hat{S}_4^T &= 0.3288, \\ \hat{S}_5^T &= 0.0002, & \hat{S}_6^T &= 5.8239 \times 10^{-7}, & \hat{S}_7^T &= 0.0050, & \hat{S}_8^T &= 1.4559 \times 10^{-7}. \end{aligned}$$

Thus, we see that θ_1 is the most important parameter, θ_4 the second most important and so on. Note that the sensitivity of the eight inputs is reflected by their weights too.

Based on the information provided by the total Sobol' indices, we construct reduced models with stochastic dimension reduced from seven down to four. We approximate the

Pearson correlation coefficient between the original $f^{(0)}$ with all eight stochastic inputs (3.14) and all sensitivity-driven interpolants using 10,000 MC samples. The results are presented in Table 1. In the second column, we show which subset of inputs was used to find the low-fidelity model, in the third column, we see the estimated correlation coefficient, and in the last column, we show the number of high-fidelity model evaluations to find the sparse grid low-fidelity model. We see that the eight-dimensional low-fidelity model is very accurate, having a correlation coefficient of 0.9999, and is also relatively cheap to construct, requiring only 146 evaluations of $f^{(0)}$. However, when the two most unimportant parameters, θ_6 and θ_9 , are neglected, we see that the correlation coefficient virtually does not change. Furthermore, even if we reduce the stochastic dimension down to three, i.e., the inputs are only θ_1 , θ_2 and θ_4 , the correlation coefficient is still close to one, i.e., 0.9960, while the cost of finding the low-fidelity model decreases to only 49 model evaluations.

j	θ	ρ_j	# high-fidelity evaluations
1	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8\}$	0.9999	146
2	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7\}$	0.9999	137
3	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_7\}$	0.9999	129
4	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_7\}$	0.9998	103
5	$\{\theta_1, \theta_2, \theta_4, \theta_7\}$	0.9991	74
6	$\{\theta_1, \theta_2, \theta_4\}$	0.9960	49

Table 1: Inputs to the low-fidelity models obtained by the dimension-adaptive sparse grid interpolation low-fidelity model and their properties.

We remark that reduced-dimension sparse grid low-fidelity models can be directly obtained from the full-dimension sparse grid low-fidelity model by fixing the unimportant inputs to a deterministic value. However, in this case the evaluation costs of both low-fidelity models will be the same, thus restricting the potential of reduced-dimension low-fidelity models in MFMC; recall that the important ingredients in MFMC are both the correlation coefficient and the evaluation cost of low-fidelity models (see Section 2.2). To this end, in this work we construct reduced-dimension low-fidelity model explicitly from high-fidelity evaluations, as we did in this section. In this way, we obtain low-fidelity model with lower evaluation times than the runtime of the full-dimension low-fidelity model.

3.3. Low-fidelity model: Deep neural network regression

We also consider a data-driven low-fidelity model based on a feed-forward deep neural network, which is trained by solving a regression problem for a set of given training samples; see, e.g., [3, 22].

This model is used in Section 4.1.1, in which we consider three stochastic parameters $\theta_1, \theta_2, \theta_3$ and approximate the corresponding scalar-valued GENE output $\gamma_1(\theta)$. We use a fully-connected network with two hidden layers of three neurons each and ReLU activation functions. This architecture is depicted in Figure 1. We train the model using an RMSprop optimizer with a learning rate of $\alpha = 0.001$ that minimizes the MSE of the neural network output compared to the given training data. As training samples, we make use of existing

data that we have available from previous numerical experiments. We construct and train the neural network using the Keras API on Tensorflow for `python`.

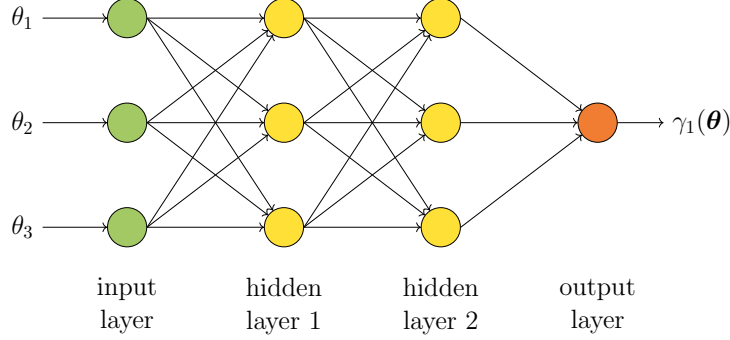


Figure 1: The structure of the neural network used as a low-fidelity model of GENE.

4. Numerical results

In this section, we employ the proposed data-driven multi-fidelity framework for uncertainty propagation in two plasma micro-turbulence analysis scenarios to estimate the expectation and variance of the growth rate of dominant eigenmodes, γ_1 (see Section 3.1). In Section 4.1, we consider a modified Cyclone Base Case benchmark scenario to draw our initial conclusions about the potential of our MFMC approach to quantify uncertainty in plasma micro-turbulence analysis. First, in Section 4.1.1, we consider the three uncertain parameters that are usually the most important for plasma micro-turbulence simulations (the log density and temperature gradients of the species). For a more comprehensive overview of this scenario, we extend the number of uncertain inputs to eight in Section 4.1.2. In Section 4.2, we study turbulence suppression by energetic particles with 14 uncertain parameters.

4.1. Modified Cyclone Base Case

We first consider a modified version of the Cyclone Base Case scenario. In the original Cyclone Base Case, which was first presented in [15], only one particle species is considered, i.e., deuterium ions. In our modified version, we have two particle species: deuterium ions and electrons [16, 17]. Furthermore, the magnetic geometry is described by the analytical Miller equilibrium [31] instead of the typically considered, simpler $\hat{s} - \alpha$ model. We also consider collisions modelled by a linearized Landau operator. In this way our setup more closely resembles realistic plasma micro-turbulence analysis problems.

Setup. To discretize the 5D gyrokinetic state space, we employ $184,320 := 15 \times 24 \times 1 \times 32 \times 16$ degrees of freedom in total: we use 15 Fourier modes in the radial (x) direction and 24 points along the field line, in the z direction; recall that in flux-tube simulations, only one point is used in the bi-normal (ky) direction. In velocity space, we employ 32 equidistant symmetric parallel velocity grid points and 16 Gauss-Laguerre distributed magnetic moment points.

The high-fidelity GENE simulations are performed using 32 cores on two Intel Xeon E5-2697 nodes of the CoolMUC-2 Linux cluster.² The low-fidelity evaluations were performed on a laptop computer with an Intel Core i5-8250U CPU running at 1.60 GHz. All calculations were performed in double precision arithmetic.

4.1.1. Case with three uncertain inputs

First, we consider three uncertain parameters: the ion and electron logarithmic temperature gradients, ω_{T_i} and ω_{T_e} , as well as their logarithmic density gradient, ω_n . Note that to ensure the quasi-neutrality assumption in plasma physics, the density gradients need to be equal. It is known that changes in these parameters are causing the underlying micro-instability [15] and therefore these three inputs are of interest with respect to uncertainty propagation. We summarize the setup employed in our experiments in Table 2. The parameters are modelled as uniform random variables with bounds of $\pm 25\%$ around their nominal value, which is their typical value used in typical plasma micro-turbulence numerical studies.

	parameter	symbol	probability distribution
θ_1	ion/electron log density gradient	ω_n	$\mathcal{U}(1.6650, 2.7750)$
θ_2	ion log temperature gradient	ω_{T_i}	$\mathcal{U}(7.5000, 12.5000)$
θ_3	electron log temperature gradient	ω_{T_e}	$\mathcal{U}(7.5000, 12.5000)$

Table 2: The three uncertain input parameters for the three-dimensional Cyclone Base Case and their probability distributions. By $\mathcal{U}(a, b)$ we denote a uniform distribution on $[a, b]$.

For this scenario, it is known [15] that around bi-normal wave-number $k_y \rho_s = 0.6$ there is a mode transition from an ion-temperature gradient (ITG) to a trapped electron (TEM)/electron temperature gradient (ETG) hybrid mode. Therefore, we consider two wave numbers in our experiments: one for which the micro-instability is driven by ITG, i.e., $k_y \rho_s = 0.3$ and another one for which we have TEM/ETG micro-instability, i.e., $k_y \rho_s = 0.8$.

Data-driven low-fidelity models. For both values of $k_y \rho_s$ we construct two low-fidelity models, $f^{(1)}$ and $f^{(2)}$. One data-driven low-fidelity model is based on the sensitivity-driven dimension-adaptive sparse grid interpolation approximation summarized in Section 3.2. To construct it at $k_y \rho_s = 0.3$, we employ the tolerance $\boldsymbol{\tau} = (10^{-1}, 10^{-1}, 10^{-1}, 10^{-1})$. At $k_y \rho_s = 0.8$, we set $\boldsymbol{\tau} = (10^{-6}, 10^{-6}, 10^{-6}, 10^{-3})$. We choose smaller tolerances at $k_y \rho_s = 0.8$ because we know from previous experiments that, compared to $k_y \rho_s = 0.3$, more effort is needed to obtain sufficiently accurate low-fidelity models. Note, however, that the sparse grid model is hierarchical, therefore one can start constructing low-fidelity models with some initial tolerances and re-use those model evaluations if the tolerances need to be decreased.

In both cases, the maximum grid level is set to $L_{\max} = 20$. Only four high-fidelity evaluations with GENE were needed to create the sparse grid low-fidelity model at $k_y \rho_s = 0.3$ and 34 GENE evaluations were needed for $k_y \rho_s = 0.8$. The other low-fidelity model is based

²<https://www.lrz.de/services/compute/linux-cluster/>

on the deep-network model summarized in Section 3.3. To train the network, we make use of existing GENE evaluations available from previous experiments. At $k_y\rho_s = 0.3$, we had available 10^5 evaluations. For the second wave-number, we trained on 2.5×10^4 data samples.

We show the correlation coefficients, single-core runtimes and variances of the high- and low-fidelity models in Table 3. For clarity, we use the following notation for the two low-fidelity models: the superscripts denote the fidelity of the model and to distinguish between the sparse grid and the machine learning models, we use the subscripts SG and ML, respectively. At $k_y\rho_s = 0.3$, the first low-fidelity model in the hierarchy is the deep-network low-fidelity model, i.e., $f_{\text{ML}}^{(1)}$, and the second low-fidelity model is the sparse grid approximation, i.e., $f_{\text{SG}}^{(2)}$. At $k_y\rho_s = 0.8$, the order of the low-fidelity models is reversed due to the values of their Pearson correlation coefficients: $f_{\text{SG}}^{(1)}$ is the first low-fidelity model and $f_{\text{ML}}^{(2)}$ is the second low-fidelity model. The three quantities were estimated using 1,000

$f^{(j)}$	ρ_j	w_j [sec]	σ_j^2	$f^{(j)}$	ρ_j	w_j [sec]	σ_j^2
$f^{(0)}$	1.0000	260.1697	0.0170	$f^{(0)}$	1.0000	240.5123	0.0754
$f_{\text{ML}}^{(1)}$	0.9998	0.0019	0.0165	$f_{\text{SG}}^{(1)}$	0.9819	0.0166	0.0747
$f_{\text{SG}}^{(2)}$	0.9989	0.0008	0.0144	$f_{\text{ML}}^{(2)}$	0.9708	0.0017	0.0703

(a) $k_y\rho_s = 0.3$

(b) $k_y\rho_s = 0.8$

Table 3: Correlation and costs of high-fidelity and low-fidelity models for modified Cyclone Base Case.

evaluations of the high- and low-fidelity models. The evaluation costs of the low-fidelity models are 4 – 5 orders of magnitude smaller compared to the costs of the high-fidelity model. At $k_y\rho_s = 0.3$, both low-fidelity models have correlation coefficients close to 1.0, whereas at $k_y\rho_s = 0.8$, the low-fidelity models are less accurate and thus poorer correlated to the high-fidelity model.

Estimating the expectation. We employ MFMC to estimate the expectation and variance of the growth rate at both $k_y\rho_s = 0.3$ and $k_y\rho_s = 0.8$ for budgets $p \in \{5 \times 10^2, 10^3, 5 \times 10^3, 10^4, 5 \times 10^4\}$ seconds. For comparison purposes, we also compute MC approximations for the same budgets p .

We first compare the MC and MFMC expectation estimators in terms of their MSE that is estimated from N replicates as:

$$e_{\text{MSE}}(\hat{E}_n^{(\cdot)}) = \frac{1}{N} \sum_{n=1}^N (\hat{\mu}_{\text{ref}} - \hat{E}_n^{(\cdot)})^2, \quad (4.1)$$

where $\hat{\mu}_{\text{ref}}$ serves as the reference mean estimator and $\hat{E}_n^{(\cdot)}$ is either an MFMC or an MC estimator. The reference $\hat{\mu}_{\text{ref}}$ is obtained with MFMC using a large budget $p_{\text{ref}} = 10^5$ seconds. At $k_y\rho_s = 0.3$, the reference is $\hat{\mu}_{\text{ref}} = 0.6811$ and at $k_y\rho_s = 0.8$, we obtain $\hat{\mu}_{\text{ref}} = 0.5560$.

For a broader overview, we consider two MFMC estimators: one in which we have the sparse grid low-fidelity model only and another in which we consider both the sparse grid

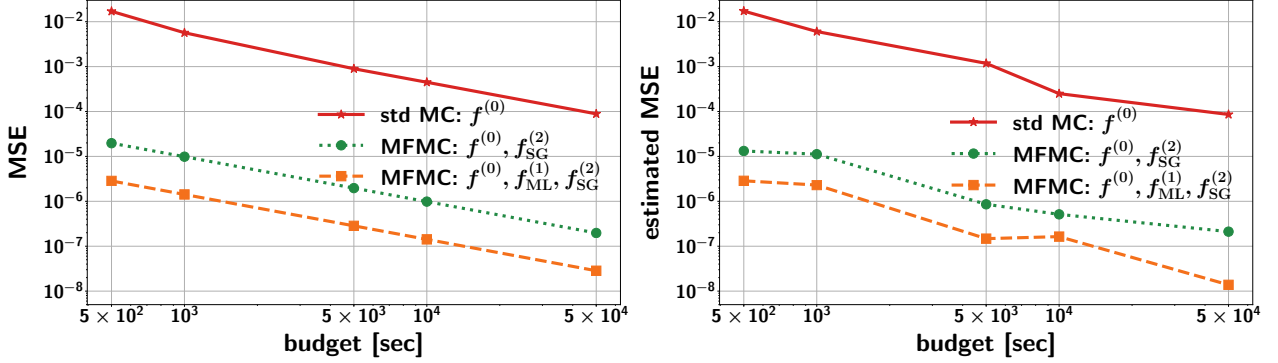


Figure 2: Analytical MSE (left) and estimated MSE (right) of standard MC and MFMC for the Cyclone Base Case with 3 uncertain inputs and $k_y \rho_s = 0.3$.

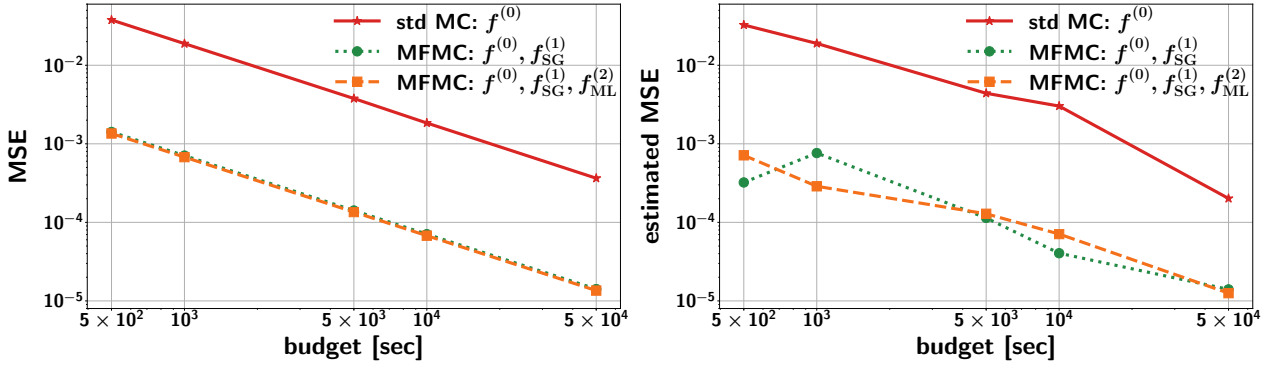


Figure 3: Analytical MSE (left) and estimated MSE (right) of standard MC and MFMC for the Cyclone Base Case with 3 uncertain inputs and $k_y \rho_s = 0.8$.

and the machine-learning-based low-fidelity models. We first compute the analytic MSE obtained via (2.4) and (2.13) based on the analysis in [39]. The analytic MSE is obtained from the measured values in Table 3 and gives an indication what variance reduction and speedup we expect to see in the numerical simulations. Notice that the analytic MSEs are computed purely from the values given in Table 3 and require no additional numerical simulations. Additional analyses of the MFMC approach to estimate how much variance reduction one can expect before conducting numerical experiments are presented in [38]. The analytic MSEs are shown in the left panel of Figure 2 and 3. The estimated MSE, computed via (4.1) using $N = 10$ replicates, are visualized in the right panel in each figure. MFMC is around three orders of magnitude more accurate in terms of the estimated MSE for $k_y \rho_s = 0.3$ and about one order of magnitude more accurate for $k_y \rho_s = 0.8$, compared to the MC estimator. The results in Figure 2 for $k_y \rho_s = 0.3$ indicate that adding the machine learning model to MFMC decreases the MSE further. This is explained by the high correlation and the low evaluation costs of the two low-fidelity models as shown in Table 3. In contrast, for $k_y \rho_s = 0.8$, the results in Figure 3 indicate that adding the machine learning model leads to little improvement, which is explained by the poorer correlation coefficient as shown in Table 3.

We also compute the MSE of the MFMC estimator in which we explicitly account for the construction cost of low-fidelity models. We compute the estimator which depends on the high-fidelity and the sparse grid low-fidelity model and focus on the estimator with budget $p = 5 \times 10^4$. At $k_y \rho_s = 0.3$, four high-fidelity evaluations, i.e., $4 \times w_0 = 4 \times 260.1697 \approx 1,040$ seconds were required to construct the sparse grid low-fidelity model $f_{\text{SG}}^{(2)}$. We subtract the construction budget from the total budget of 5×10^4 seconds and use the remaining budget $p' = 5 \times 10^4 - 1,040 = 48,960$ seconds for MFMC sampling, i.e., around 98% of the original budget. Analogously, at $k_y \rho_s = 0.8$ the budget used to construct the low-fidelity model $f_{\text{SG}}^{(1)}$ was $34 \times w_0 \approx 5,772$ seconds. The remaining budget for MFMC sampling is therefore $p' = 5 \times 10^4 - 5,772 = 44,228$ seconds, i.e., roughly 88% of the original p . We visualize the corresponding results in Figure 4 (left, $k_y \rho_s = 0.3$; right, $k_y \rho_s = 0.8$). We compare the MSE of standard MC sampling, standard MFMC sampling (without offline cost) and MFMC sampling in which we explicitly consider the construction cost. As expected, subtracting the sparse grid construction cost of only four high-fidelity evaluations at $k_y \rho_s = 0.3$ has a negligible effect on the MSE of the MFMC estimator. At $k_y \rho_s = 0.8$, subtracting the construction cost of the low-fidelity model has an insignificant effect as well: the MFMC estimator remains more than one order of magnitude more accurate than the standard MC estimator despite the fact that the sampling budget was decreased by roughly 12%.

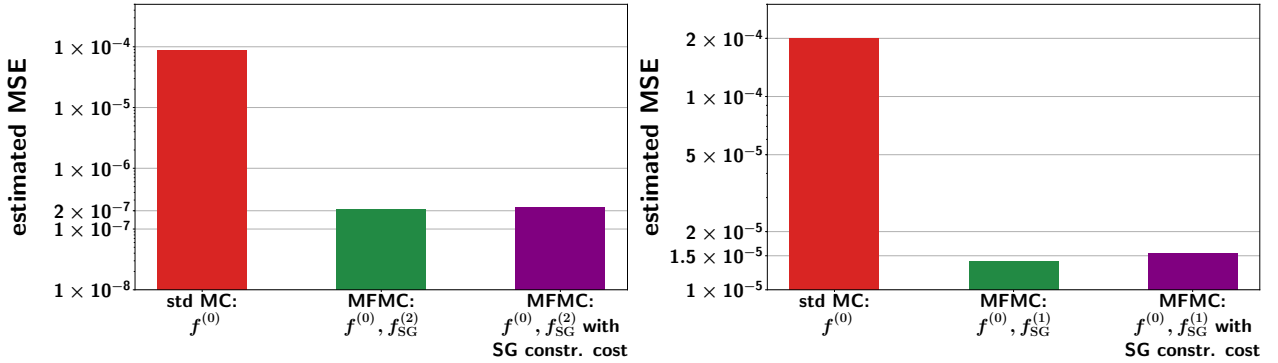


Figure 4: MSE of mean estimators for $p = 5 \times 10^4$ for standard MC and MFMC using the low-fidelity model $f_{\text{SG}}^{(1)}$ with and without considering the construction cost of the sparse grid model for the Cyclone Base Case with three uncertain inputs and $k_y \rho_s = 0.3$ (left) and $k_y \rho_s = 0.8$ (right).

Speedup. To put the computational savings provided by our MFMC approach into perspective, let us take a closer look at how the obtained variance reduction translates into computational speedups. High-fidelity simulations are performed on 32 cores, whereas low-fidelity simulations are performed on a single core only because they are so much cheaper. At both $k_y \rho_s = 0.3$ and $k_y \rho_s = 0.8$, the MFMC estimator with the smallest MSE was obtained for budget $p = 5 \times 10^4$ seconds, hence this will serve as reference in the following. Notice that the budget of $p = 5 \times 10^4$ seconds on a single core translates into $5 \times 10^4 / 60 / 32 \approx 26$ minutes of high- and low-fidelity simulations on 32 cores. Computing a standard MC estimator with the same MSE as the MFMC estimator requires a budget of about 104,583,190

seconds on a single core, which corresponds to $104,583,190/3,600/24/32 \approx 38$ days of high-fidelity simulations on 32 cores. Finally, the MFMC estimator with only the sparse grid low-fidelity model needs a budget of about 249,916 seconds, i.e., about 130 minutes of high- and low-fidelity simulations in 32 cores. To summarize, at $k_y \rho_s = 0.3$, the variance reduction obtained by our MFMC approach leads to a runtime reduction from 38 days to 26 minutes on 32 cores if both low-fidelity models are used. At $k_y \rho_s = 0.8$, a budget of about 1,339,482 seconds on a single core is needed to obtain a standard MC estimator with the same MSE as the MFMC estimator with budget $p = 5 \times 10^4$ on a single core, which means that a standard MC estimator would require about 12 hours of high-fidelity simulations on 32 cores. In contrast, MFMC requires only 26 minutes of high- and low-fidelity simulations on 32 cores. Hence, we see that even though the low-fidelity models at $k_y \rho_s = 0.8$ were not as accurate as at $k_y \rho_s = 0.3$, using them in the context of MFMC lead to a runtime reduction from 12 hours to 26 minutes on 32 cores.

Observations. The reported results demonstrate two aspects of MFMC: first, even models with limited accuracy can be useful for MFMC and can lead to orders of magnitude improvements in terms of MSE compared to standard MC estimators. Especially in the context of plasma physics, where one high-fidelity simulation can require significant computational resources, this MFMC feature can make the estimation of the expectation (or of other quantities of interest, such as variance or Sobol' indices for sensitivity analysis) computationally feasible. Second, adding more than one low-fidelity model to MFMC does not necessarily lead to a further reduction of the MSE but depends on the properties of the low-fidelity model. Figure 5 shows the distribution of the number of samples among the models. At most 0.0497% of the total number of samples are with the high-fidelity model in MFMC, whereas standard MC sampling employs the high-fidelity model exclusively. This provides a significant computational advantage compared to single-fidelity sampling MC schemes.

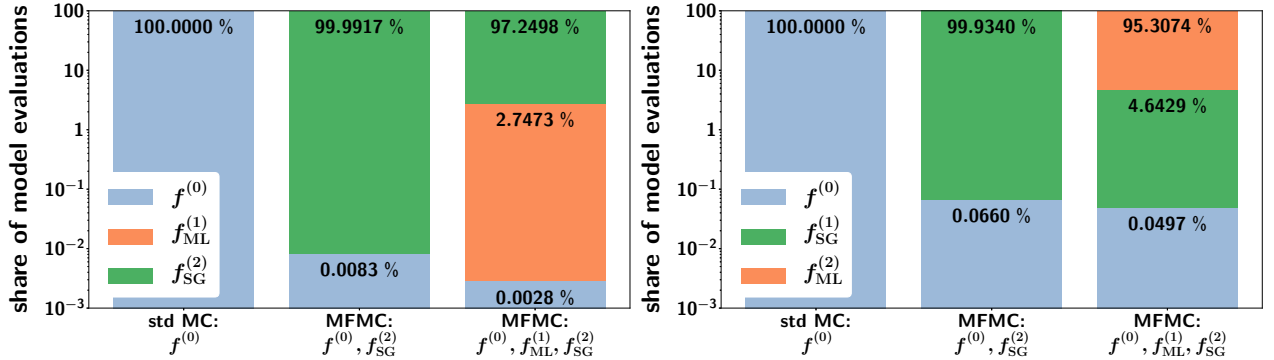


Figure 5: The distribution of evaluations over the models for standard MC and MFMC with two and three models for $k_y \rho_s = 0.3$ (left) and $k_y \rho_s = 0.8$ (right). On the logarithmic y-axis, the percentage of evaluations of the different models is given.

Estimating the variance. Table 4 shows the variance estimates obtained in our experiments. We used the same low-fidelity models as for the expectation estimation. To distinguish

between the MFMC estimators using one and two low-fidelity models, we use the superscripts $MFMC, 1$ and $MFMC, 2$, respectively. For $k_y\rho = 0.3$, no estimate is obtained with budget $p = 500$ because at least two samples are needed to estimate the variance and one evaluation of the high-fidelity model takes already $260.1698 > p/2$ seconds; cf. Table 3. Using the same setup as for estimating the reference means, we obtain the reference variance $\hat{\sigma}_{ref}^2 = 0.016765$ at $k_y\rho_s = 0.3$ and $\hat{\sigma}_{ref}^2 = 0.0757$ at $k_y\rho_s = 0.8$. We observe that again MFMC yields more accurate results than the standard MC estimator: at $k_y\rho_s = 0.3$, both MFMC estimators have an extra digit of accuracy, whereas at $k_y\rho_s = 0.8$, the MC and MFMC estimators have two digits of accuracy for most budgets, but the MFMC estimator with one low-fidelity model yields a variance similar to the reference.

p	$\widehat{\text{Var}}^{MC}$	$\widehat{\text{Var}}^{MFMC,1}$	$\widehat{\text{Var}}^{MFMC,2}$	p	$\widehat{\text{Var}}^{MC}$	$\widehat{\text{Var}}^{MFMC,1}$	$\widehat{\text{Var}}^{MFMC,2}$
500	-	-	-	500	0.0749	0.0753	0.0724
1,000	0.015314	0.017189	0.016677	1,000	0.0912	0.0723	0.0805
5,000	0.015497	0.016826	0.016725	5,000	0.0695	0.0751	0.0801
10,000	0.017299	0.016873	0.016762	10,000	0.0702	0.0738	0.0727
50,000	0.017132	0.016714	0.016769	50,000	0.0759	0.0757	0.0750

(a) $k_y\rho_s = 0.3$, reference $\hat{\sigma}_{ref}^2 = 0.016765$

(b) $k_y\rho_s = 0.8$, reference $\hat{\sigma}_{ref}^2 = 0.0757$

Table 4: The estimated variance of the output of interest at $k_y\rho_s = 0.3$ (left) and $k_y\rho_s = 0.8$ (right) obtained with MC and MFMC estimators in the Cyclone Base Case with 3 uncertain inputs. The superscripts $MFMC, 1$ and $MFMC, 2$ refer to estimators obtained from MFMC using one and two low-fidelity models, respectively.

4.1.2. Cyclone Base Case with eight uncertain inputs

For a broader overview of this scenario, we extend the setup from three to eight uncertain parameters. Besides the logarithmic temperature and density gradients, we consider five additional uncertain parameters: the ion-electron temperature ratio, T_i/T_e ; β , the ratio of kinetic to magnetic pressure inside the plasma; the normalized collision frequency, ν_c ; the safety factor, q , which describes the relationship of the number of toroidal turns of a magnetic field line to number of poloidal turns; and its derivative, the magnetic shear, \hat{s} . We model these eight inputs as independent uniform variables with bounds of 25% around their nominal values, as shown in Table 5.

We focus on one perpendicular wave number, i.e., $k_y\rho_s = 0.3$. In MFMC, we consider only the sensitivity-driven sparse grid low-fidelity model summarized in Section 3.2. We do not employ the deep-network-based low-fidelity model because we do not have available training data from previous experiments. Running numerical simulations to generate sufficient training data is beyond our means for this work.

Data-driven low-fidelity model. To find the low-fidelity sparse grid model $f^{(1)}$, we prescribe tolerances $\tau = 10^{-6} \cdot \mathbf{1}_9$ and a maximum level of refinement $L_{\max} = 20$. With this setup, we need only 52 GENE evaluations to construct the low-fidelity model. Table 6 reports

	parameter	symbol	probability distribution
θ_1	ion/electron log density gradient	ω_n	$\mathcal{U}(1.665, 2.775)$
θ_2	ion log temperature gradient	ω_{T_i}	$\mathcal{U}(7.500, 12.500)$
θ_3	electron log temperature gradient	ω_{T_e}	$\mathcal{U}(7.500, 12.500)$
θ_4	plasma beta	β	$\mathcal{U}(0.598 \times 10^{-3}, 0.731 \times 10^{-3})$
θ_5	collision frequency	ν_c	$\mathcal{U}(0.238 \times 10^{-2}, 0.322 \times 10^{-2})$
θ_6	temperature ratio	T_i/T_e	$\mathcal{U}(0.950, 1.050)$
θ_7	magnetic shear	\hat{s}	$\mathcal{U}(0.716, 0.875)$
θ_8	safety factor	q	$\mathcal{U}(1.330, 1470)$

Table 5: The eight uncertain input parameters for the eight-dimensional Cyclone Base Case and their probability distributions.

the correlation coefficients, single-core runtimes and variances of the high- and low-fidelity models, obtained using 1,000 MC evaluations from both models. We see that the low-fidelity

$f^{(j)}$	ρ_j	w_j [sec]	σ_j^2
$f^{(0)}$	1.0000	257.0304	0.0168
$f^{(1)}$	0.9999	0.1006	0.0169

Table 6: Overview of the high-fidelity and low-fidelity models for $k_y \rho_s = 0.3$ used in the MFMC estimation in the eight-dimensional Cyclone Base Case.

model is highly correlated with a correlation coefficient of 0.9999, while achieving four orders of magnitude cost reduction.

Estimating the expectation. We consider budgets $p \in \{5 \times 10^3, 10^4, 5 \times 10^4\}$ seconds. Compared to the case with three inputs, we increased the smallest budget from 500 to 5,000 seconds to allow for at least one high-fidelity evaluation in the MFMC estimator. We compare the standard MC and MFMC expectation estimators with respect to the reference $\hat{\mu}_{\text{ref}} = 0.6794$, computed using MFMC with a budget $p_{\text{ref}} = 10^5$ seconds.

In Figure 6, we compare MC and MFMC in terms of both the analytic (left) and the estimated MSE (4.1) computed using $N = 10$ replicates (right). The MFMC estimate is about three orders of magnitude more accurate than the standard MC estimate for the same budget. This speedup is achieved because an accurate low-fidelity model is used that is much cheaper to evaluate than the high-fidelity model.

Speedup. This translates into the following runtime reduction: to obtain an MSE similar to the smallest MSE, i.e., the MSE of our MFMC estimator with budget 5×10^4 seconds on a single core, the standard MC estimator requires a budget of about 43,447,972 seconds on a single core. Using 32 cores per high-fidelity simulation, this budget requires about 16 days of simulations. In contrast, our MFMC approach requires only about 26 minutes of high- and low-fidelity simulations on 32 cores.

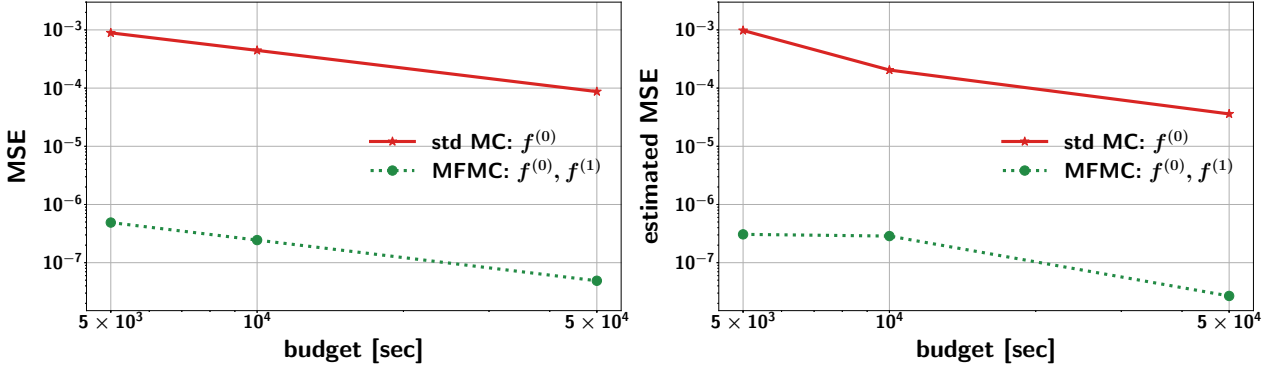


Figure 6: Analytical MSE of the mean estimator (left) and estimated MSE of standard MC and MFMC results (right) for the 8-dimensional Cyclone Base Case for $k_y \rho_s = 0.3$.

Estimating the variance. The reference variance, obtained using the same setup as for the expectation, is $\hat{\sigma}_{\text{ref}}^2 = 0.01716$. We show our results in Table 7. As for the mean estimator, the MFMC variance estimator is more accurate than the standard MC estimator for the same budget for all investigated cases.

p	\widehat{Var}^{MC}	$\widehat{Var}^{MFMC,1}$
5,000	0.02074	0.01719
10,000	0.01771	0.01716
50,000	0.01754	0.01715

Table 7: The estimated variance of the output of interest for the Cyclone Base Case with eight inputs. The superscript $MFMC, 1$ refers to estimators obtained from MFMC using the sparse grid interpolation low-fidelity model. The reference variance is $\hat{\sigma}_{\text{ref}}^2 = 0.01716$.

4.2. Turbulence suppression by energetic particles

This scenario is inspired by [11, 12], which studied the effect of energetic particles such as Neutral Beam Injection (NBI) fast deuterium and Ion Cyclotron Resonance Heating (ICRH) Helium-3 in suppressing ITG turbulence in the Joint European Torus (JET) tokamak. It was shown numerically in [11, 12] that the turbulence suppression observed in tokamak experiments can be explained via the combination of two distinct effects: (i) a quasi-linear wave-particle resonance interaction between fast ions and the bulk ITG-driven micro-instabilities and (ii) a nonlinear wave-wave interaction between marginally stable energetic particle-driven modes and ITG turbulence. Simulations supported by theoretical results have shown that the former effect takes place when the drift-frequency of the supra-thermal particles gets close to the frequency of the underlying ITG, thus allowing a free energy redistribution. Experimental evidence showing signatures of improved plasma confinement due to this quasi-linear effect have been observed in tokamaks [9, 11, 12], and predicted also for optimized stellarator devices [8]. The second stabilization effect involves a nonlinear wave-wave coupling between ITG turbulence and marginally stable electromagnetic modes driven by

fast ions, which depletes the ITG turbulent drive. In strong electromagnetic regimes, these nonlinearly excited fast ion modes act as a catalyst of energy into axisymmetric perturbations called zonal flows, which further reduce turbulent transport [10, 42].

To emphasize the effect of fast ions in stabilizing plasma turbulence, we show in Figure 7 a poloidal cut of the density fluctuations from two nonlinear simulations. The left panel shows the case without fast ions and the right panel the case when fast ions are injected into the plasma. Using fast ions leads to less elongated streamers and therefore less radial transport, as a result of the combined effect of reduced turbulent drive and increased zonal flow activity, meaning more turbulence stabilization than in the case without fast ions. Besides the two types of energetic particles, NBI fast deuterium and ICRH ^3He , and the

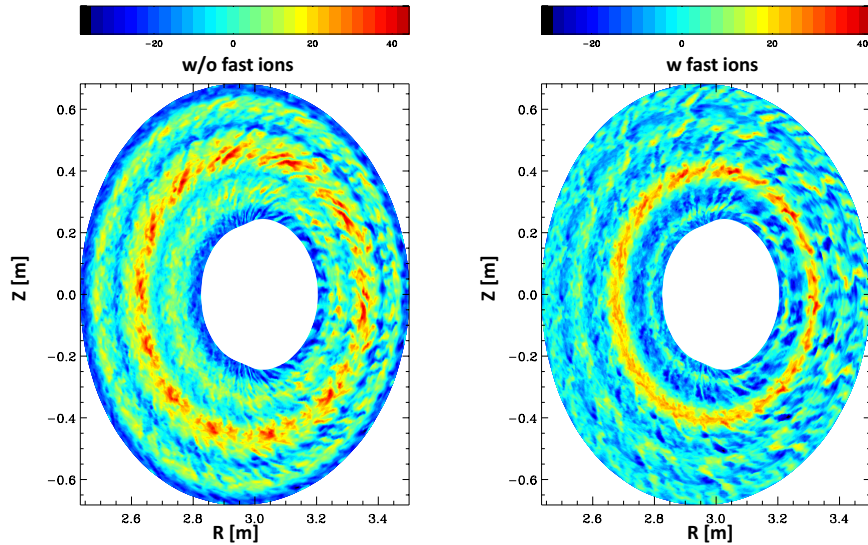


Figure 7: Poloidal cut (at a selected time step) of the density fluctuations in two nonlinear simulations without (left) and with (right) fast ions. When using fast ions, the streamers are less elongated and do not travel outward as in the case without fast ions. Therefore, there is less radial transport and hence more turbulence stabilisation in the simulation with fast ions.

other two usual particle species in a plasma, deuterium ions and electrons, we also consider Carbon impurities to mimic JET-like C-Wall plasma conditions. This scenario therefore has a total of five particle species.

The study of turbulence suppression by energetic particles is therefore a problem of high practical relevance in the fusion community, since both experiments and simulations suggest that supra-thermal particles generated via external heating schemes might lead to a significant improvement of the fusion output and overall plasma performances. And since a potentially large number of parameters characterizing the properties of the magnetic geometry, the bulk species, and the energetic particles are typically affected by e.g., noise in the experimental measurements, it is paramount to perform numerical simulations using an uncertainty propagation framework.

Setup. To discretize the 5D gyrokinetic state space, we use 21 Fourier modes in the radial (x) direction and 24 points along the field line in the z direction. In velocity space, we employ 32 equidistant symmetric parallel velocity grid points and 16 Gauss-Laguerre distributed magnetic moment points. This gives a total of 258,048 degrees of freedom. The magnetic geometry is characterized by the analytical Miller equilibrium [31].

The GENE simulations were performed using 240 cores on eight Intel Xeon E5-2697 nodes of the CoolMUC-2 Linux Cluster. The low-fidelity models were simulated on an Intel Core i5-8250U CPU running at 1.60 GHz. As in the first scenario, all calculations were performed in double precision arithmetic.

The parameters of interest in this problem are the ones associated to the underlying particle species. Therefore, in our simulations, we consider 12 uncertain parameters characterizing the main ions, electrons, fast deuterium and ^3He , and two additional parameters associated to the magnetic geometry, i.e., the safety factor, q , and the magnetic shear, \hat{s} . This gives a total of 14 uncertain parameters. We model the uncertain parameters as uniform random variables with 25% bounds around their nominal values as shown in Table 8. Throughout our simulations, the impurity parameters are kept fixed to their nominal values. Quasi-neutrality is ensured by prescribing the values for the density of the deuterium ions and the logarithmic density gradient of electrons in terms of the values of the density and logarithmic density gradients of the other four species, respectively. We perform our simulations at a fixed wave number, $k_y \rho_s = 0.5$, which is known to be the most unstable ITG eigenmode in flux-tube simulations [11].

	parameter	symbol	probability distribution
θ_1	safety factor	q	$\mathcal{U}(1.3230, 2.1705)$
θ_2	magnetic shear	\hat{s}	$\mathcal{U}(0.3920, 0.6533)$
θ_3	ion log temperature gradient	ω_{T_i}	$\mathcal{U}(3.4230, 5.7050)$
θ_4	ion log density gradient	ω_{n_i}	$\mathcal{U}(0.0047, 0.0078)$
θ_5	ion temperature	T_i	$\mathcal{U}(0.7500, 1.2500)$
θ_6	fast deuterium log temperature gradient	ω_{T_D}	$\mathcal{U}(0.7742, 1.2903)$
θ_7	fast deuterium log density gradient	ω_{n_D}	$\mathcal{U}(3.5413, 5.9022)$
θ_8	fast deuterium density	n_D	$\mathcal{U}(0.0450, 0.0750)$
θ_9	fast deuterium temperature	T_D	$\mathcal{U}(7.3500, 12.2500)$
θ_{10}	Helium-3 log temperature gradient	$\omega_{T_{^3\text{He}}}$	$\mathcal{U}(5.5543, 9.2573)$
θ_{11}	Helium-3 log density gradient	$\omega_{n_{^3\text{He}}}$	$\mathcal{U}(0.3770, 0.6283)$
θ_{12}	Helium-3 density	$n_{^3\text{He}}$	$\mathcal{U}(0.0525, 0.0875)$
θ_{13}	Helium-3 temperature	$T_{^3\text{He}}$	$\mathcal{U}(9.0000, 15.0000)$
θ_{14}	electron log temperature gradient	ω_{T_e}	$\mathcal{U}(1.6695, 2.7825)$

Table 8: Input parameters and their distributions used in the scenario of studying turbulence suppression with 14 uncertain inputs. By $\mathcal{U}(a, b)$, we denote a uniform distribution on $[a, b]$.

Data-driven low-fidelity models with full and reduced stochastic dimension. The low-fidelity models are constructed using sensitivity-driven dimension-adaptive sparse grid interpolation.

Initially, we compute a sensitivity-driven low-fidelity model using all 14 parameters of Table 8; this will be the first low-fidelity model, $f^{(1)}$, in the multi-fidelity hierarchy. We prescribe tolerances $\boldsymbol{\tau} = 10^{-6} \cdot \mathbf{1}_{15}$ and maximum level $L_{\max} = 20$. With this setup, we need 96 GENE evaluations to construct the 14-dimensional low-fidelity model.

We exploit the sensitivity information provided by $f^{(1)}$ and construct reduced-dimension low-fidelity models as well. To this end, we compute the associated total Sobol' indices, which are visualized in Figure 8. We observe that at most six parameters (θ_1 , θ_3 , θ_5 , θ_{10} , θ_{11} and θ_{14}) are important in this scenario; the total Sobol' indices of the remaining eight parameters fall below 0,001. In addition, the total indices corresponding to θ_2 , θ_4 , θ_6 and θ_7 are in $O(10^{-4})$ or smaller. Therefore, the values of the total Sobol' indices suggest decreasing the stochastic dimension from 14 to nine or even down to six. For a more comprehensive overview, we consider two reduced-dimension low-fidelity models: $f^{(2)}$, which includes the nine most important parameters and $f^{(3)}$, which includes the six most important stochastic parameters. We note that the values of the total Sobol' indices in Figure 8 are consistent with what is expected from a physics perspective: since we have ITG-driven micro-turbulence, the main ions' logarithmic temperature gradient is expected to be the most important parameter. Additionally, for the considered parameter ranges in Table 8, the Helium-3 logarithmic temperature gradient and its density are expected to be important as well [11, 12, 19].

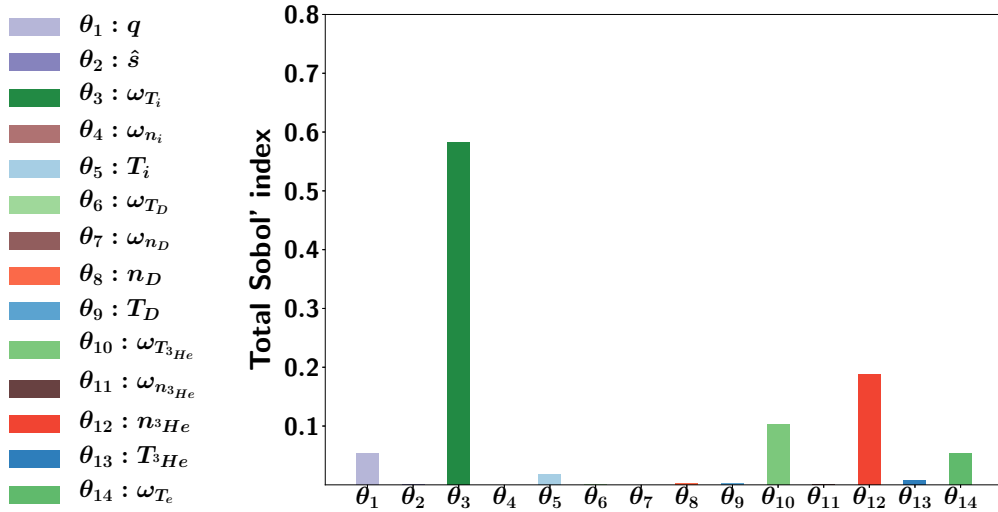


Figure 8: Total Sobol' indices of the 14 stochastic inputs from Table 8 computed by sensitivity-driven dimension-adaptive sparse grid interpolation.

Tables 9 and 10 report the single-core runtimes, correlation coefficients and variances of the high-fidelity model $f^{(0)}$ and the three low-fidelity models $f^{(1)}$, $f^{(2)}$ and $f^{(3)}$ with their respective inputs. Since the evaluation of the correlation coefficient and of the variance of the high-fidelity model entails high-fidelity evaluations, it is important to keep the associated cost as low as possible. To this end, we evaluate the aforementioned quantities using only 50 MC samples generated independently in an offline step. Additionally, we also consider the case in which we generate 50 MC samples to estimate this quantities and re-use these 50 samples in the MFMC estimators. In other words, in the latter case we bias the MFMC

estimators, the trade-off being that we do not require additional high-fidelity evaluations to estimate the correlation coefficient and the high-fidelity variance. To distinguish between the two strategies, we will use in the following the label “re-use” for the latter case. We see that all low-fidelity models have a high correlation with the high-fidelity model, showing that even the reduced-dimension models will be useful for MFMC. This is in contrast with traditional single-fidelity approaches in which the full set of uncertain parameters is used. At the same time, the runtime of the reduced-dimension models is much lower than that of the 14-dimensional models. The six-dimensional model $f^{(3)}$ is 186 times faster to evaluate than the 14-dimensional low-fidelity model $f^{(1)}$ and about 1.8×10^5 times faster than the high-fidelity model $f^{(0)}$.

$f^{(j)}$	d_j	θ_j	w_j [sec]
$f^{(0)}$	14D	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{12}, \theta_{13}, \theta_{14}\}$	11574.8697
$f^{(1)}$	14D	$\{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7, \theta_8, \theta_9, \theta_{10}, \theta_{12}, \theta_{13}, \theta_{14}\}$	11.6724
$f^{(2)}$	9D	$\{\theta_1, \theta_3, \theta_5, \theta_8, \theta_9, \theta_{10}, \theta_{12}, \theta_{13}, \theta_{14}\}$	0.3838
$f^{(3)}$	6D	$\{\theta_1, \theta_3, \theta_5, \theta_{10}, \theta_{12}, \theta_{14}\}$	0.0627

Table 9: Inputs of low-fidelity models and their runtimes for the turbulence suppression scenario.

$f^{(j)}$	ρ_i	ρ_i re-use	σ_i^2	σ_i^2 re-use
$f^{(0)}$	1.0000	1.0000	0.004379	0.002931
$f^{(1)}$	0.9978	0.9967	0.004450	0.002897
$f^{(2)}$	0.9977	0.9963	0.004445	0.002885
$f^{(3)}$	0.9929	0.9875	0.004415	0.002899

Table 10: Correlation coefficients and variances of the low-fidelity models for the turbulence suppression scenario estimated from 50 independent samples and 50 samples which are also used in MFMC.

Estimating the expectation. We compare standard MC and MFMC for the expectation estimation for budgets $p \in \{5 \times 10^4, 10^5, 5 \times 10^5, 10^6\}$ seconds, ensuring that at least one high-fidelity model evaluation is within the budget. We consider two MFMC estimators: one estimator that uses only the 14-dimensional low-fidelity model and another estimator that uses all low-fidelity models. In addition, for each estimator we consider the case in which the correlation coefficient and high-fidelity model’s variance were computed from separate high-fidelity evaluations and also the case in which the samples used to estimate the two quantities are re-used in MFMC. The reference expectation and variance were computed using MFMC with a budget $p_{\text{ref}} = 10^7$ seconds. We obtain $\hat{\mu}_{\text{ref}} = 0.1972$.

We depict the analytic (left) and the estimated MSE obtained using (4.1) with $N = 20$ replicates in Figure 9. We observe that the MFMC estimator that uses the high-fidelity model and the 14-dimensional low-fidelity model is about two orders of magnitude more efficient than standard MC. Adding the computationally cheap reduced-dimension models

leads to an improvement of about a factor 2 compared to the MFMC with only the full-dimension low-fidelity model. We see therefore that adding the reduced-dimension low-fidelity models into our data-driven multi-fidelity approach leads to more accurate results. In addition, we see that reusing the samples used to estimate the correlation coefficient and high-fidelity model's variance has an insignificant effect on the MSE of the MFMC estimators.

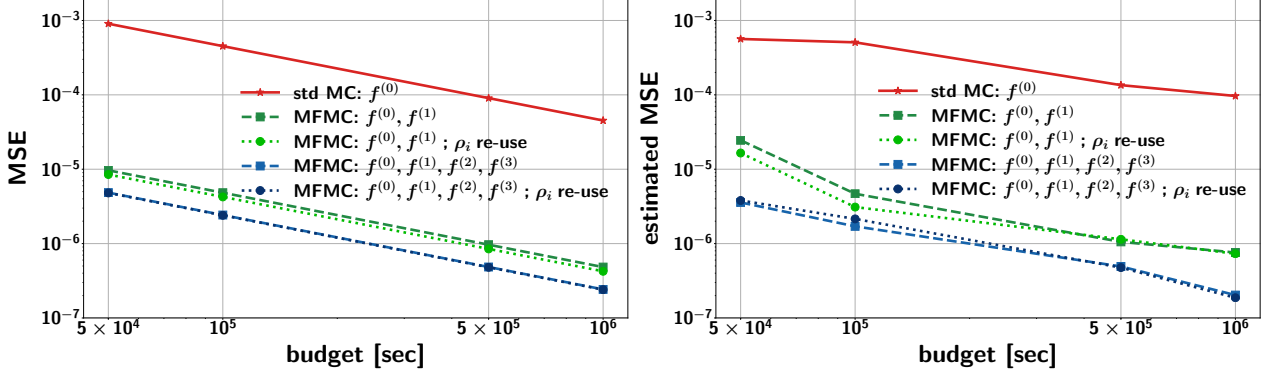


Figure 9: Analytical MSE (left) and estimated MSE (right) of standard MC and MFMC for the turbulence suppression scenario.

Speedup. The obtained variance reduction translates in the following speedup. To obtain the smallest MSE, i.e., the MSE corresponding to the MFMC estimator with all three low-fidelity models with budget $p = 10^6$ seconds, we need around one hour of high- and low-fidelity simulations on 240 cores. The MFMC estimator with only the full-dimension low-fidelity model requires about two hours of high- and low-fidelity simulations on 240 cores to yield the same MSE. Finally, for the same task, the standard MC estimator requires a budget of approximately 165,630,360 seconds on a single core. To perform the corresponding high-fidelity simulations on 240 cores requires a total of about eight days for the standard MC estimator. Thus, our MFMC approach reduces the total runtime on 240 cores from eight days to one hour when using the estimator that considers both full- and reduced-dimension low-fidelity models.

Estimating the variance. In Table 11, we show the variance estimates obtained with MC and MFMC. The reference variance is $\hat{\sigma}_{\text{ref}}^2 = 0.00397$, computed using the same procedure as for the expectation reference. The MFMC estimates are more accurate than the standard MC estimates and adding the dimension-reduced models improves the accuracy of the estimates further. We observe that between the two MFMC estimators, the most accurate is the MFMC estimator that considers all low-fidelity models. We therefore see that adding reduced-dimension low-fidelity models improves MFMC variance estimates in this example. Finally, as we observed for the expected value, we see that reusing the samples used to estimate the correlation coefficient and high-fidelity model's variance has an insignificant effect on the MFMC variance estimators.

p	$\widehat{\text{Var}}^{MC}$	$\widehat{\text{Var}}^{MFMC,1}$	$\widehat{\text{Var}}_{\rho_i \text{ re-use}}^{MFMC,1}$	$\widehat{\text{Var}}^{MFMC,3}$	$\widehat{\text{Var}}_{\rho_i \text{ re-use}}^{MFMC,3}$
5×10^4	0.00274	0.00405	0.00420	0.00405	0.00406
1×10^5	0.00372	0.00401	0.00404	0.00400	0.00400
5×10^5	0.00392	0.00404	0.00403	0.00399	0.00399
1×10^6	0.00407	0.00400	0.00401	0.00398	0.00398

Table 11: The estimated variance of the output of interest in the turbulence suppression scenario. The superscripts $MFMC, 1$ and $MFMC, 3$ refer to estimators obtained from MFMC using one ($f^{(1)}$) and three ($f^{(1)}, f^{(2)}, f^{(3)}$) low-fidelity models, respectively. The reference variance is $\hat{\sigma}_{\text{ref}}^2 = 0.00397$.

5. Conclusions

In the present paper, we demonstrated the feasibility of uncertainty propagation in large-scale gyrokinetic simulations by means of multi-fidelity Monte Carlo sampling together with structure-exploiting data-driven low-fidelity models. The high-fidelity model was given by the GENE code, while the low-fidelity models were constructed using the sensitivity-driven dimension-adaptive sparse grid interpolation procedure which exploits that in many real-world problems, the intrinsic dimension is smaller than the ambient dimension and that the stochastic inputs are anisotropically coupled. In addition, we also considered a deep-network-based low-fidelity model in a scenario where we had available a large database of numerical experiments to train the model. In both of our numerical experiments, the MFMC estimators with data-driven low-fidelity models achieved several orders of magnitude speedups compared to standard MC estimators. Especially for the second numerical example, which studied the role of supra-thermal particles in suppressing instabilities, the speedups achieved with MFMC with reduced-dimension low-fidelity models allowed to perform a much more comprehensive uncertainty propagation study than a standard, single-fidelity Monte Carlo estimator. Thus, the present work opens the door towards studies involving a large number of gyrokinetic simulations under real-world conditions, including uncertainty quantification and optimization.

Acknowledgments

I.-G.F., A.D.S. and F.J. were supported by the Exascale Computing Project (No. 17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration. B.P. acknowledges support by the National Science Foundation under grants CMMI-1761068 and IIS-1901091. We gratefully acknowledge the compute and data resources provided by the Leibniz Supercomputing Centre (www.lrz.de).

References

- [1] Alsup, T., Peherstorfer, B., 2020. Context-aware surrogate modeling for balancing approximation and sampling costs in multi-fidelity importance sampling and Bayesian inverse problems. arXiv:2010.11708.
- [2] Berrut, J.-P., Trefethen, L. N., 2004. Barycentric lagrange interpolation. SIAM Review 46 (3), 501–517.

- [3] Bishop, C. M., 2006. Pattern Recognition and Machine Learning. Springer-Verlag, Berlin, Heidelberg.
- [4] Brizard, A., Hahm, T., 2007. Foundations of nonlinear gyrokinetic theory. *Reviews of Modern Physics* 79.
- [5] Bungartz, H.-J., Griebel, M., 2004. Sparse grids. *Acta Numerica* 13, 147–269.
- [6] Constantine, P. G., Eldred, M. S., Phipps, E. T., 2012. Sparse pseudospectral approximation method. *Computer Methods in Applied Mechanics and Engineering* 229–232, 1–12.
- [7] Dannert, T., Jenko, F., 2005. Gyrokinetic simulation of collisionless trapped-electron mode turbulence. *Physics of Plasmas* 12 (7), 072309.
- [8] Di Siena, A., Bañón Navarro, A., Jenko, F., Sep 2020. Turbulence suppression by energetic particle effects in modern optimized stellarators. *Physical Review Letters* 125, 105002.
- [9] Di Siena, A., Bilato, R., Görler, T., Navarro, A. B. n., Poli, E., Bobkov, V., Jarema, D., Fable, E., Angioni, C., Kazakov, Y. O., Ochoukov, R., Schneider, P., Weiland, M., Jenko, F., the ASDEX Upgrade Team, Jul 2021. New high-confinement regime with fast ions in the core of fusion plasmas. *Physical Review Letters* 127, 025002.
URL <https://link.aps.org/doi/10.1103/PhysRevLett.127.025002>
- [10] Di Siena, A., Görler, T., Poli, E., Navarro, A. B. n., Biancalani, A., Bilato, R., Bonanomi, N., Novikau, I., Vannini, F., Jenko, F., et al., 2021. Nonlinear electromagnetic interplay between fast ions and ion-temperature-gradient plasma turbulence. *Journal of Plasma Physics* 87 (2), 555870201.
- [11] Di Siena, A., Görler, T., Doerk, H., Poli, E., Bilato, R., mar 2018. Fast-ion stabilization of tokamak plasma turbulence. *Nuclear Fusion* 58 (5), 054002.
- [12] Di Siena, A., Görler, T., Poli, E., Bilato, R., Doerk, H., Zocco, A., 2019. Resonant interaction of energetic ions with bulk-ion plasma micro-turbulence. *Physics of Plasmas* 26 (5), 052504.
URL <https://doi.org/10.1063/1.5087203>
- [13] Dimarco, G., Pareschi, L., 2019. Multi-scale control variate methods for uncertainty quantification in kinetic equations. *Journal of Computational Physics* 388, 63 – 89.
- [14] Dimarco, G., Pareschi, L., 2020. Multiscale variance reduction methods based on multiple control variates for kinetic equations with uncertainties. *Multiscale Modeling & Simulation* 18 (1), 351–382.
- [15] Dimits, A. M., Bateman, G., Beer, M. A., Cohen, B. I., Dorland, W., Hammett, G. W., Kim, C., Kinsey, J. E., Kotschenreuther, M., Kritiz, A. H., Lao, L. L., Mandrekas, J., Nevins, W. M., Parker, S. E., Redd, A. J., Shumaker, D. E., Sydora, R., Weiland, J., 2000. Comparisons and physics basis of tokamak transport models and turbulence simulations. *Physics of Plasmas* 7 (3), 969–983.
- [16] Farcaş, I.-G., Görler, T., Bungartz, H.-J., Jenko, F., Neckel, T., 2020. Sensitivity-driven adaptive sparse stochastic approximations in plasma microinstability analysis. *Journal of Computational Physics* 410, 109394.
- [17] Farcaş, I.-G., 2020. Context-aware model hierarchies for higher-dimensional uncertainty quantification. Dissertation, Technische Universität München, München.
- [18] Farcaş, I.-G., Sârbu, P. C., Bungartz, H.-J., Neckel, T., Uekermann, B., 2018. Multilevel adaptive stochastic collocation with dimensionality reduction. In: Garcke, J., Pflüger, D., Webster, C. G., Zhang, G. (Eds.), *Sparse Grids and Applications - Miami 2016*. Springer International Publishing, Cham, pp. 43–68.
- [19] Farcaş, I.-G., Siena, A. D., Jenko, F., apr 2021. Turbulence suppression by energetic particles: a sensitivity-driven dimension-adaptive sparse grid framework for discharge optimization. *Nuclear Fusion* 61 (5), 056004.
- [20] Freethy, S. J., Görler, T., Creely, A. J., Conway, G. D., Denk, S. S., Happel, T., Koenen, C., Hennequin, P., White, A. E., 2018. Validation of gyrokinetic simulations with measurements of electron temperature fluctuations and density-temperature phase angles on asdex upgrade. *Physics of Plasmas* 25 (5), 055903.
- [21] Gerstner, T., Griebel, M., 2003. Dimension-adaptive tensor-product quadrature. *Computing* 71 (1), 65–87.
- [22] Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, <http://www.deeplearningbook.org>.
- [23] Görler, T., Lapillonne, X., Brunner, S., Dannert, T., Jenko, F., Merz, F., Told, D., 2011. The global

- version of the gyrokinetic turbulence code gene. *Journal of Computational Physics* 230 (18), 7053 – 7071.
- [24] Hegland, M., 2003. Adaptive sparse grids. In: Burrage, K., Sidje, R. B. (Eds.), *Proc. of 10th Computational Techniques and Applications Conference CTAC-2001*. Vol. 44. pp. C335–C353.
 - [25] Hu, J., Jin, S., 2017. *Uncertainty Quantification for Kinetic Equations*. Springer International Publishing, Cham, pp. 193–229.
 - [26] Jantsch, P., Webster, C. G., Zhang, G., 06 2018. On the Lebesgue constant of weighted Leja points for Lagrange interpolation on unbounded domains. *IMA Journal on Numerical Analysis* 39 (2), 1039–1057.
 - [27] Jenko, F., Dorland, W., Kotschenreuther, M., Rogers, B. N., 2000. Electron temperature gradient driven turbulence. *Physics of Plasmas* 7 (5), 1904–1910.
 - [28] Konrad, J., 2019. Multifidelity monte carlo sampling in plasma microturbulence analysis. Bachelor’s thesis, Technical University of Munich.
 - [29] Lakhili, J., Hoenen, O., Luk, O. O., Coster, D. P., 2020. Uncertainty quantification for multiscale fusion plasma simulations with vecma toolkit. In: Krzhizhanovskaya, V. V., Závodszky, G., Lees, M. H., Dongarra, J. J., Sloot, P. M. A., Brissos, S., Teixeira, J. (Eds.), *Computational Science – ICCS 2020*. Springer International Publishing, Cham, pp. 719–730.
 - [30] Marzouk, Y., Moselhy, T., Parno, M., Spantini, A., 2016. *Sampling via Measure Transport: An Introduction*. Springer International Publishing, Cham, pp. 1–41.
 - [31] Miller, R. L., Chu, M. S., Greene, J. M., Lin-Liu, Y. R., Waltz, R. E., 1998. Noncircular, finite aspect ratio, local equilibrium model. *Physics of Plasmas* 5 (4), 973–978.
 - [32] Narayan, A., Jakeman, J. D., 2014. Adaptive Leja sparse grid constructions for stochastic collocation and high-dimensional approximation. *SIAM Journal on Scientific Computing* 36 (6), A2952–A2983.
 - [33] Nelson, B. L., 1987. On control variate estimators. *Computers & Operations Research* 14 (3), 219 – 225.
 - [34] Ng, L., Willcox, K., 2014. Multifidelity approaches for optimization under uncertainty. *International Journal for Numerical Methods in Engineering* 100 (10), 746–772.
 - [35] Nobile, F., Tempone, R., Webster, C. G., 2008. A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis* 46 (5), 2309–2345.
 - [36] Oates, C. J., Girolami, M., Chopin, N., 2017. Control functionals for Monte Carlo integration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 79 (3), 695–718.
 - [37] Peherstorfer, B., 2019. Multifidelity Monte Carlo estimation with adaptive low-fidelity models. *SIAM/ASA Journal on Uncertainty Quantification* 7, 579–603.
 - [38] Peherstorfer, B., Gunzburger, M., Willcox, K., 2018. Convergence analysis of multifidelity Monte Carlo estimation. *Numerische Mathematik* 139 (3), 683–707.
 - [39] Peherstorfer, B., Willcox, K., Gunzburger, M., 2016. Optimal model management for multifidelity Monte Carlo estimation. *SIAM Journal on Scientific Computing* 38 (5), A3163–A3194.
 - [40] Peherstorfer, B., Willcox, K., Gunzburger, M., 2018. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review* 60 (3), 550–591.
 - [41] Qian, E., Peherstorfer, B., O’Malley, D., Vesselinov, V., Willcox, K., 2018. Multifidelity Monte Carlo estimation of variance and sensitivity indices. *SIAM/ASA Journal on Uncertainty Quantification* 6 (2), 683–706.
 - [42] Siena, A. D., Görler, T., Poli, E., Navarro, A. B., Biancalani, A., Jenko, F., sep 2019. Electromagnetic turbulence suppression by energetic particle driven modes. *Nuclear Fusion* 59 (12), 124001. URL <https://doi.org/10.1088/1741-4326/ab4088>
 - [43] Sobol’, I. M., 2001. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation* 55 (1), 271–280.
 - [44] Sudret, B., 2008. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety* 93 (7), 964–979.
 - [45] Vaezi, P., Holland, C., 2018. An improved approach to uncertainty quantification for plasma turbulence validation studies. *Fusion Science and Technology* 74 (1-2), 77–88.
 - [46] Xiu, D., 2009. Fast numerical methods for stochastic computations : A review. *Communication in*

Computational Physics 5, 242–272.

- [47] Yang, M., Zhang, G., del Castillo-Negrete, D., Stoyanov, M., Beidler, M., 2020. A sparse-grid probabilistic scheme for approximation of the runaway probability of electrons in fusion tokamak simulation. arXiv:2001.05800.