November 3, 2017

DE-SC0008454, Office of Fusion Energy Sciences
The University of Texas at Austin
Office of Sponsored Projects
P.O. Box 7726
Austin, TX 78712

Partnership for Edge Physics (EPSI), University of Texas Final Technical Report

DOE-UTexas-0008454

University of Texas PI: Prof. Robert Moser
University of Texas Team: Drs. V. Carey, D. Faghihi, C. Michoski

Partnership Members:

Princeton Plasma Physics Laboratory C.S. Chang (Lead PI), S. Ethier, S. Ku, D. Stotler
Oakridge National Laboratory: S. Klasky, E. DAzevedo, N. Podhorszki , R . Tchoua, P. Worley
Brown University: J. Hesthaven
California Institute of Technology: J. Cummings, Mark Stalzer
University of Colorado: S. Parker, Y. Chen
Columbia University: M. Adams
Lawrence Berkeley National Laboratory: A. Shoshani
Lehigh University: A. Kritz, T. Rafiq
Massachusetts Institute of Technology: M. Greenwald, L. Sugiyama
North Carolina State University: N. Samatova
Rensselaer Polytechnic Institute: M. Shephard
Rutgers University: M. Parashar
University of California San Diego: G. Tynan, P. Diamond

Abstract

Simulations of tokamak plasmas require a number of inputs whose values are uncertain. The effects of these input uncertainties on the reliability of model predictions is of great importance when validating predictions by comparison to experimental observations, and when using the predictions for design and operation of devices. However, high fidelity simulation of tokamak plasmas, particular those aimed at characterization of the edge plasma physics, are computationally expensive, so lower cost surrogates are required to enable practical uncertainty estimates. Two surrogate modeling techniques have been explored in the context of tokamak plasma simulations using the XGC family of plasma simulation codes. The first is a response surface surrogate, and the second is an augmented surrogate relying on scenario extrapolation. In addition, to reduce the costs of the XGC simulations, a particle resampling algorithm was developed, which allows marker particle distributions to be adjusted to maintain optimal importance sampling. This means that the total number of particles in and therefore the cost of a simulation can be reduced while maintaining the same accuracy.

Summary

This project has been focused on the development and application of uncertainty quantification techniques to simulations of tokamak plasmas with particle-in-cell (PIC) methods for the solution of gyrokinetic models using the XGC code family. The primary challenge in this application is the great cost of high-fidelity simulations, which make it impractical to perform high fidelity simulations of more than a hand full of cases. Multi-fidelity Monte-Carlo techniques and related approaches have been identified as most promising for this application. These techniques require that lower fidelity, less expensive models be identified to be used to explore the the mapping between uncertain model inputs to model outputs. To explore this possibility, two approaches have been pursued to define lower-cost surrogates for the system model, a response surface surrogate and a surrogate based on scenario extrapolation. Another way to reduce the cost of uncertainty quantification is to reduce the cost of each model evaluation. Uncertainty quantification tools were applied to develop a resampling method to increase accuracy and reduce the number of marker particles required in a PIC simulation. These were three main research thrusts pursued under this project, and they are summarized briefly below.

1. **Response Surface Surrogates:** a response surface surrogate is explored to represent the upshift (Dimits shift) in the critical gradient due to the presence of non-linear stationary states. Results demonstrate moderate variation in the critical gradient as a function of R_0 and B_0, and in a cross validation, the global surrogate agreed well with a nominal base case chosen within the parameter space. However, it is unclear how far this surrogate representation can be extrapolated, and surrogates for another quantity (the linear departure rate) were not as well behaved.

2. **Scenario Extrapolation Surrogates:** Scenario extrapolation (also dubbed "telescoping") was used to develop a hybrid scheme we call an augmented surrogate. By combining (potentially uncertain) scenario parameters with the dimensionless plasma scaling parameter ρ^*, we are able to construct reasonable response surface approximations, yielding extrapolative predictions as well as parameter space gradients at a reduced cost. Additional testing of our predictive extrapolation methodology is part of an ongoing power scaling study.

3. **Marker Particle Resampling:** The resampling algorithm developed under this project is designed to reduce the sampling noise in PIC codes while preserving important features of the distribution being represented by the particles. By resampling the distribution in bins, the general features of the distribution are preserved, and by enforcing the preservation of a specified set of moments in each bin, the dynamically relevant characteristics of the distribution remain unchanged after the resampling. In a test problem, it was found that periodic resampling allowed simulations to be performed with a factor of 10 fewer particles, while retaining the same accuracy.

Contents

1 Introduction **5**

2 Critical Gradient Surrogates **7**
2.1 Preliminary Considerations . . . 9
2.1.1 Geometry and Coordinate Systems . . . 9
2.1.2 Magnetic Field . . . 9
2.1.3 Initial Conditions . . . 11
2.1.4 Simulation Parameters . . . 12
2.1.5 Diagnostics . . . 14
2.2 Characteristics of the Stationary State . . . 15
2.3 Low-fidelity surrogates in the critical threshold of the gradient . . . 17
2.3.1 Derived features . . . 19

3 Scenario Extrapolation **22**
3.1 Scenario Extrapolation Methods . . . 23
3.2 Scenario Extrapolation Results . . . 24

4 Resampling Algorithms **25**
4.1 Conserved quantities for particle resampling . . . 28
4.2 The resampling algorithm . . . 29
4.2.1 Discretizing the phase-space. . . . 29
4.2.2 Setting the number of new particles in each bin. . . . 29
4.2.3 Determining new particle positions and velocities. . . . 30
4.2.4 Weight readjustment and conserving quantities. . . . 30
4.2.5 Comments, Remarks, and Discussion . . . 31
4.3 Implementation of the resampling into XGC gyrokinetic PIC code . . . 33
4.4 Particle Resampling Results . . . 35
4.4.1 Global particle downsampling . . . 35
4.4.2 Periodic particle resampling of a neo-classical PIC simulation . . . 40

5 Conclusions **48**

1 Introduction

Computational simulations of complex systems, such a tokamak plasmas generally require many inputs, many of which may be uncertain. These inputs could include parameters in submodels in the system, or inputs that define the conditions under which the system is operating. As part of the EPSi SciDAC partnership (Edge Physics Simulation), the team at the University of Texas pursued the development of uncertainty quantification tools and techniques for the characterization of these uncertainties and their impact on the results of simulations of edge plasmas in tokamaks.

There are two primary motivations for the quantification of these uncertainties. First, models of complex systems need to be validated by determining whether the models are consistent with experimental observations. Rigorous validation tests require careful characterization of uncertainties, since it is the uncertainties in both the experimental measurements, and in the simulation outputs that determine how well experiments and simulations can be expected to match. They also determine how strong a test any particular comparison is. In the case of tokamak simulations, a validation test generally involves using numerous experimental measurements to infer the conditions for an experimental shot (e.g. the equilibrium magnetic field), using these as inputs to a simulation and determining if simulation outputs are consistent with measurements of other quantities (e.g. temperature and density gradients in the pedestal). In this case, it is important to account for the uncertainties in the measurements that are used as inputs and the resulting uncertainty in model outputs, in addition to uncertainties in the measurements to which the model outputs are compared. This requires "forward propagation" of uncertainties through the model.

The second motivation for quantifying uncertainties in model outputs, is to characterize the reliability of model predictions when they are to be used to support decision-making about the system. An example for tokamak plasma simulations is using them to predict ITER performance. In this context, uncertainty characterization is important to ensure that simulation results are not relied on when their predictions are not sufficiently certain. As with validation, this application requires forward propagation of uncertainties from inputs to outputs.

Fundamental to forward propagation of input uncertainties is the exploration of the mapping from inputs to outputs. This generally requires the evaluation of the model for many values of the inputs. For the edge physics simulations considered here, the computational model is instantiated in the PIC (particle-in-cell) code family XGC [32, 13]. The computational cost of a single evaluation of this model can be very high, with cost increasing with the non-dimensional size of the tokamak. For example, an ITER-scale simulation may require 10's of millions of core hours. This makes many evaluations of the model impractical. The primary computational challenge is thus to develop uncertainty estimates with as few evaluations of the model as possible. Multi-fidelity Monte-Carlo techniques [43, 48] and related approaches have been identified as most promising for this application. These techniques require that lower fidelity, less expensive models be identified to be used to explore the input-output map. To explore this possibility, two approaches have been pursued to define lower-cost surrogates for the system model. In the first, a response surface surrogate is explored to represent the up-shift (Dimits shift) in the critical gradient due to the presence of non-linear stationary states (see 2). In the second, scenario extrapolation (also dubbed "telescoping"), in which sensitivities and uncertainties from a system scenario that is less expensive to simulate (e.g. a smaller tokamak) are used to inform uncertainties at the scenarios of interest. The viability of this approach was explored, as described in section 3.

Another way to reduce the cost of uncertainty quantification is to reduce the cost of each model

evaluation. In the XGC PIC code, the cost is dominated by the evolution of the particles. Thus to reduce the cost, one can reduce the number of particles needed to represent the distribution function f. The tools of uncertainty quantification can be brought to bear on this problem as well. Essentially, one wants to optimally distribute the marker particles to minimize the uncertainty introduced by the finite sample of the distribution function. As the simulation evolves, the particle distribution will generally evolve away from the optimal representation of f, so the particles will need to be "resampled" periodically, to maintain a near-optimal particle distribution. The capability to do this was developed under this project as described in section 4.

In addition to the major efforts described briefly above and in more detail in the remainder of this report, other preliminary investigations of uncertainty issues were pursued. First, the possibility of accounting for measurement uncertainties in the Grad-Shafranov inverse problem used to construct magnetohydrodynamic equilibria was explored, with the conclusion that it is feasible. This will be an important component of the validation under uncertainty. Secondly, a preliminary parameter estimation exercise with under uncertainty was conducted for contaminant transport models.

A way to model neoclassical impurity transport in plasma is by using the 1D transport approximation [58]:

$$\frac{\partial n_{_{i,Z}}}{\partial t} = Q_{_{i,Z}} - g^{-1/2}\frac{\partial}{\partial x} g^{1/2}\Gamma_{_{i,Z}}, \tag{1}$$

where $\Gamma_{_{i,Z}}$ is the transport flux, $Q_{_{i,Z}}$ the mass action of the relevant plasma chemistry, and $n_{_{i,Z}}(r)$ the radial charge density of species i in charge state Z.

In the STRAHL code [18] such an approach to match impurity transport during TESPEL injection in [59]. In this approach the transport model is specified in terms of the input fields T_i, T_e, and ϕ, which are provided from experimental measurements, given pointwise as spatial and temporal distributions. Thus in order to understand the transport of $n_{i,Z}$, it is paramount that the inputs be sampled from their respective distributions, and then propagated through the transport algorithm, leading to a distribution in $n_{i,Z}$, rather than a single solution. In this way it is possible to propagate uncertainty through neoclassical impurity transport that is consistent with evidence. Preliminary work has been conducted on this problem, which suggests strong dependencies on the uncertainties in the experimentally determined inputs. These uncertainties can then be used to optimize transport coefficients via an inverse problem, which has also been tested.

In this project, we use the gyrokinetic hybrid Eulerian-Lagrangian particle-in-cell codes XGC1 and XGCa. These codes are used to solve the gyrokinetic equations for the evolution of the particle distribution function f. The distribution f is represented by a sampling of marker particles which evolve in five-dimensional phase space including physical space and a two-dimensional representation of the velocity space. The XGC codes support three different representations of the distribution function. The full-f representation is most straight-forward as the marker particles are simply samples from the distribution f. The full-f formulation is expensive as it takes many particles to obtain an accurate sampling representation of the distribution. A cheaper alternative is the commonly used δf representation which is essentially a variance reduction approach in which only the difference between f and a reference distribution f_0 is represented by particles. However, this requires the reference distribution to be specified *a priori*, and particularly for the mean profiles to be specified, precluding the possibility for the means to evolve in response to the dynamics of the system. The so-called "total-f" formulation [36], also uses variance reduction through a reference distribution, but in this case f_0 is represented on a grid in velocity space, and is allowed to evolve

slowly to represent the macroscopic evolution of the system.

The XGC1 code [32, 13] incorporates realistic flux geometries up to the first wall, as well as a number of particle species.

2 Critical Gradient Surrogates

Present-day tokamak experiments indicate that modern reactor designs contain plasma that is particularly resilient to common global MHD instabilities. More clearly, when the instabilities in fusion plasma are not suppressed they can easily destroy confinement altogether, but modern reactors are designed to suppress these instabilities. The resulting global MHD stability, however, is in a state of delicate quasi-equilibrium with an abundance of energy available to drive more localized, and thus – as generally assumed – more benign instabilities.

These localized instabilities are prevalent in many fusion plasmas, and lead to turbulence that often reduces confinement times through convective transport. Experimentally this micro-turbulent transport can also be suppressed resulting in the high confinement mode, or H-mode – an operational regime of great importance in tokamak reactors. Of additional importance in the edge plasma is the drift-wave turbulence thought to be responsible for the anomalous heat losses observed in many tokamaks.

A dominant feature of classical fluid turbulence is the nonlinear interaction of fluctuations at different scales. These fluctuations generally reach a dynamic equilibrium in which turbulent energy is produced or injected at some (large) scale, and transferred by the nonlinear interactions to much smaller scales where it is dissipated. Recent progress in drift-wave plasma turbulence, however, suggests that turbulent saturation can also occur through interaction with damped modes at the injection scale, thus complicating the classical fluid picture [60].

In addition to setting up a cascade by which energy in the fluctuations is ultimately damped, turbulence affects transport, which modifies the large-scale flow. The result can be the inhibition of large-scale instabilities; a behavior observed in tokamak plasmas [14, 53]. In particular, the break-down of the turbulent equilibrium is characterized by a large increase in the ion heat flux, and the ion temperature gradient at which this occurs is larger than the critical gradient for linear stability [14].

The linear instability threshold discussed above is obtained by solving for the dispersion relation for infinitesimal perturbations in the reactive fluid model for quasilinear scaling of turbulent drift-wave transport, as discussed by B. Coppi and further detailed by [67]. The base flow is determined by a Braginskii closure obtained by expanding the kinetic theory in an orthogonal basis (Sonine polynomials) and deriving first-order corrections [23]. In addition to other linear scaling formulations, this model for the linear instability is used in transport codes such as TRANSP or JINTRAC for analysis of experimental results. Linear gyrokinetics is one of the more advanced approaches to quasi-linear transport models [7].

Experimentally it is generally observed that the confinement time increases with the magnetic field and plasma current, and decreases with increasing heating power [69]. Transport models describe the phenomenon in terms of a stability threshold, analogous to the onset of linear instability described above. It is important to note that, as discussed at length in [39], experimental scaling studies have traditionally been performed using dimensional scalings such as the ITER98 scalings, in contrast to, for example, the theoretically more compelling dimensionless parameters of Kadomtsev. The reasons for this are multifaceted, and include the infeasibility of performing

experimental scans in the Kadomtsev parameters. That said, it should be understood that the surrogate models discussed in this paper are designed with an eye towards experimental validation, and thus certain choices are made that align more closely with the experimental scalings than with those of the more theoretical dimensionless ones.

Here and below, we use the term "critical threshold" to indicate the boundary between the stable stationary state and the unstable state. That is, there is an approximate threshold value in the parameter $\eta_i = L_n/L_T$ (the ratio of the scaled ion gyrocentered density gradient $L_n = -\nabla \ln n_i$ to the scaled ion temperature gradient L_T as defined below) above which turbulent transport occurs.

This critical threshold is generally not the same as the critical threshold for linear instabilities. In the nonlinear regime the linearly unstable perturbations saturate creating a new stationary state in a process often referred to as "self-organized criticality." The transition to this state with enhanced transport and therefore reduced confinement times is therefore more involved than in simple linear instability theory. As a consequence, in addition to transport and confinement scaling based on linear theory, it is important to characterize this critical threshold for transition to enhanced transport in terms of experimentally accessible plasma parameters.

The principle aim of this work is to establish surrogate models that can be used to predict instability driven loss of confinement to allow determination of admissible machine configurations for current and future reactors. Sensitivity studies and parameters sweeps using high-fidelity computational models capable of representing the phenomenon (e.g. gyrokinetic models) can be prohibitively expensive. This makes it challenging to identify problems that are both pertinent to fusion and amenable to generating low-fidelity surrogates. As a result, many of the sensitivity studies that have been performed were done using reduced models, such as gyrofluid and Hasegawa-Wakatani models [30, 31, 55], or even surrogates in the linear case in the form of derived algebraic relations [27].

Gyrokinetic studies on the other hand have been constrained by their computational expense, but have nonetheless investigated dependence on the plasma β (i.e. the ratio of plasma pressure to magnetic pressure) parameter [51], and sensitivities to corrections for collisionality, nonadiabatic electron dynamics, and impurity effects [41, 63]. Most transport models such as JINTRAC [46], TGLF [56], the one used in European Transport Solver (ETS) [12], and others incorporate some data obtained from gyrokinetic simulations into their models.

Here we identify a number of problems exhibiting the upshift in the critical threshold of the ion temperature gradient (ITG) transition point due to ITG turbulence. We first provide the general context of the study and a characterization of the solution strategy. Next we provide a set of diagnostic quantities, along with an explanation of the stationary state of interest. By judiciously selecting surrogate training scenarios, we are then able to determine low-cost surrogates for the temperature gradient (R/L_T), as a function of the toroidal magnetic field (B_0) and the major radius (R_0) of the tokamak. Finally, after arriving at surrogate models for the onset of the stationary regime, we closely examine the possibility of deriving more informative features from the surrogates, including rates of departure and attraction in

In this effort, we use XGC1's total-f formulation, which is able to evolve to a self-organized dynamic equilibrium characteristic of turbulence, and is therefore appropriate for the study pursued here, i.e. the inhibition of large-scale instabilities due to turbulence. Here, we restrict attention to electrostatic ITG simulations in low-β (β here is the plasma beta, or the plasma pressure over the magnetic pressure). We target the low-β limit here because our focus is on circular geometries, which means that the Shafranov shift is negligible.

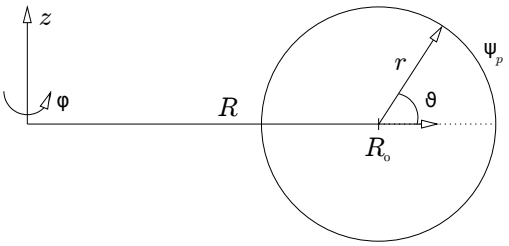


Figure 1: Coordinate system for the torus. Here ψ_p is an isocontour of the poloidal flux surface.

The simulations described below are global full-radius and self-determined – that is, background profiles are co-evolved, including the inhomogeneous terms in the magnetic field. Here we use linearized responses of Maxwell-Boltzmann distributions to the electric field to model the electrons, e.g. $n_e = n_0 \left(1 + \frac{e(\phi - \langle\phi\rangle)}{T_e}\right)$, and the ion distribution f_0 is initialized as a centered Maxwell-Boltzmann distribution. Magnetic equilibria for the simulations described here were generated using gengrid2 (as described below), and the meshes are subsequently produced using the Simmetrix library [72]. The simulated plasma is collisionless, and the particles satisfy the standard "bounce condition" at the boundaries. The Poisson solver uses the standard XGC multiscale Dirichlet conditions on the potential at the outer boundary [32].

2.1 Preliminary Considerations

To characterize the dependence of the ion temperature gradient on B_0 and R_0, a series of plasma simulations in idealized tokamak configurations have been performed. These simulation scenarios are defined in the following subsections.

2.1.1 Geometry and Coordinate Systems

A simple toroidal geometry with circular cross-section is considered, with major radius R_0 and minor radius $a < R_0$. The aspect ratio is defined as $\epsilon = a/R_0$. There are two convenient coordinate systems for the interior of the torus. The first are cylindrical coordinates (R,φ,z), with axis coincident with the axis of the torus, as shown in figure 1. The second coordinate system is a scaled toroidal system ($\hat{r}$,φ,ϑ), with $\hat{r} = r/R_0$. The relationship between cylindrical and toroidal coordinates is given by:

$$R = R_0(1 + \hat{r}\cos\vartheta), \quad z = R_0\hat{r}\sin\vartheta. \tag{2}$$

Vectors defining the coordinate directions in r, φ and ϑ will be denoted $\vec{\nabla}r$, $\vec{\nabla}\varphi$ and $\vec{\nabla}\vartheta$, respectively, with magnitudes 1, $1/R$ and $1/r$.

In the remainder of this paper, quantities that are averaged over the toroidal volume, and over surfaces of constant r will be computed. For an arbitrary scalar quantity A, they will be denoted $\{A\}$ and $\langle A\rangle$, respectively

2.1.2 Magnetic Field

The magnetic field for each simulation case must be defined so that as R_0 and B_0 vary, the field is otherwise consistent from case to case. Consider first the toroidal component of the magnetic field

B_T. It is defined to vary as $1/R$, and be independent of z and φ. Consistency from case to case requires that the toroidal component of the magnetic field at $R = R_0$ scale with B_0, and as such B_T is defined by $B_T = B_0 R_0/R$.

The poloidal component B_p is defined in terms of the poloidal magnetic flux function ψ, which depends only on r, through the relation $B_p = \psi'/R$, where ψ' is the derivative of ψ with respect to r. As B_0 and therefore B_T are varied, B_p should also vary to maintain the same magnetic configuration. The relevant measure of the magnetic configuration is the safety factor $q(r, \vartheta)$:

$$q = \frac{\vec{B} \cdot \nabla\varphi}{\vec{B} \cdot \nabla\vartheta} = \frac{r B_T}{R B_p} \tag{3}$$

While q can be kept the same with varying B_0, it cannot remain exactly the same while R_0 varies. Instead we insist that the ϑ-averaged q (i.e. $\langle q \rangle$) remain the same, where

$$\langle q \rangle = \frac{1}{2\pi} \oint \frac{\vec{B} \cdot \nabla\varphi}{\vec{B} \cdot \nabla\vartheta} \, d\vartheta = \frac{B_0 R_0 r}{2\pi\psi'} \oint \frac{1}{R} \, d\vartheta. \tag{4}$$

Making use of (2), rewriting to give ψ' as a function of $\langle q \rangle$ and evaluating the integral yields

$$\psi' = \left(\frac{B_0 R_0 \hat{r}}{\langle q \rangle \sqrt{1 - \hat{r}^2}} \right). \tag{5}$$

The average safety factor $\langle q \rangle$ is then specified to be quadratic in r:

$$\langle q \rangle = \beta + \alpha r^2/a^2, \tag{6}$$

where α and β are free parameters used to define $\langle q \rangle$ at $r = a$ and $r = 0$ respectively. To obtain $\psi(r)$, (5) is integrated while setting $\psi(0) = 0$ to yield

$$\psi = B_0 R_0^2 (c_1 - c_2 \tanh^{-1}(c_3 \sqrt{1 - \hat{r}^2})), \tag{7}$$

where

$$c_1 = c_2 \tanh^{-1}(c_3) \tag{8}$$

$$c_2 = \frac{R_0}{\sqrt{\alpha a (\beta R_0 + \alpha a)}} \tag{9}$$

$$c_3 = \sqrt{\frac{\alpha a}{\beta R_0 + \alpha a}}. \tag{10}$$

With these definitions the magnetic field used in the simulations is completely defined by

$$\vec{B} = B_0 R_0 \nabla\varphi + r\psi'(r) \nabla\vartheta. \tag{11}$$

Here $\vec{B}$ is specified with five parameters: R_0, B_0, a, α and β, of which a, α and β remain fixed as R_0 and B_0 are varied.

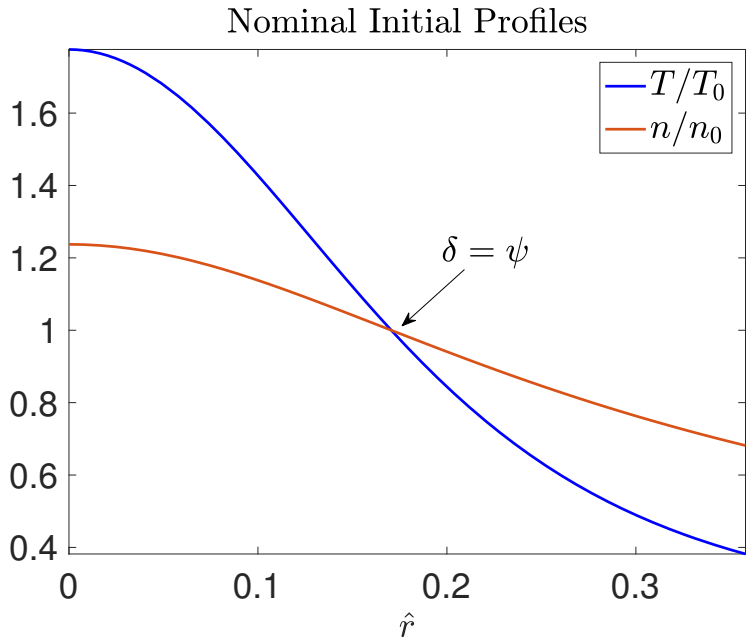


Figure 2: Initial profiles of T and n for the nominal case.

2.1.3 Initial Conditions

An initial condition must be specified for each simulation, in particular, profiles of the initial ion temperature T and ion gyrocentered density n need to be specified as a function of $\hat{r}$ in each case. The profiles are taken as

$$T = T_0 \exp\left\{\gamma_1 \tanh((\delta - \psi_{\rm n})\gamma_2)\right\}, \tag{12}$$

$$n = n_0 \exp\left\{\sigma_1 \tanh((\delta - \psi_{\rm n})\sigma_2)\right\}. \tag{13}$$

where $\psi_{\rm n}$ is the normalized ψ. As shown in Fig. 2, these profiles take on values T_0 and n_0 where $\psi_{\rm n}(\hat{r}) = \delta$. The profiles transition from a high value at $\hat{r} = 0$ to a lower value at the boundary $\hat{r} = a$. Of particular interest are the derivatives with respect to $\hat{r}$ of these profiles (Fig. 3), since it is the derivatives that define the stability parameter $\eta_i = L_n/L_T$. The scale lengths L_T and L_n are defined in terms of the logarithmic gradient of T and n by $L_T^{-1} = |\nabla T(\psi)|/T$ and $L_n^{-1} = |\nabla n(\psi)|/n$.

These profiles were chosen to introduce specified gradients (L_T and L_n) in n and T over most of the domain of r, while ensuring that the gradients were smaller near boundaries. The intent is to introduce instability and turbulence over much of the domain. In simulations performed here, δ is kept constant while B_0 and R_0 are varied. The remaining variables γ and σ were varied to keep R_0/L_n constant. Initial guesses for R_0/L_T were taken from the linear theory for η_i, though it was found that keeping R_0/L_T constant did not always result in an instability leading to turbulence, and so the value of the initial R_0/L_T, and therefore γ_1 were adjusted to ensure the onset of turbulence. Finally, rather than holding the reference temperature T_0 constant, the ratio ρ^* of the ion gyroradius ρ to the minor radius a is held constant. Because the gyroradius scales like $\sqrt{T_0}/B_0$ and a is being kept constant, this requires that $T_0 \propto B_0^2$.

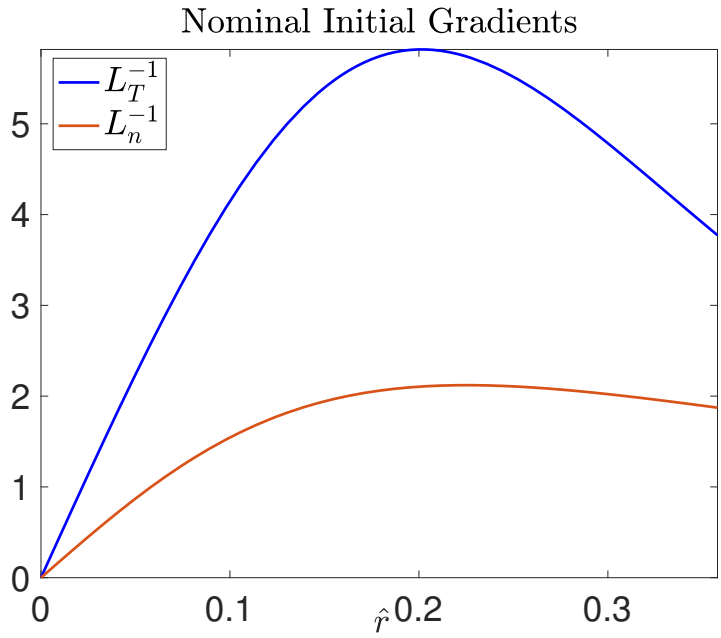


Figure 3: Initial profiles of the gradients for the nominal case.

As a comment, it turned out that extrapolating the threshold (or sufficient) γ_1 for the initial instability that induced the onset of turbulence was unsuccessful, though a number of hypothesis were tested. For example scaling R_0/L_T linearly with respect to B did not lead to the onset of turbulent states. This choice could be partially motivated by historical scale invariance approaches that suggest linear (or Bohm) scaling of χ_i in ρ^* (and therefore in $\sim 1/B$) (similarly gyro-Bohm and Goldston scalings were not observed). Ultimately the choice of γ_1 for scaling the initial R_0/L_T showed no notable scaling.

2.1.4 Simulation Parameters

To set the parameters defined in the previous subsections, a common cyclone test case was used. The values of the parameters that do not vary from case to case in this study are provided in Table 1. As discussed in Section 1, simulations scenarios are set by varying the major radius R_0 and magnetic field strength B_0. A plausible range of these parameters for a cyclone case was selected so that the aspect ratio $\epsilon \in [0.17, 0.85]$ which with a as given in Table 1 places $R_0 \in [0.716\text{m}, 3.58\text{m}]$ and $B_0 \in [1\text{T}, 5\text{T}]$.

The objective is to develop response surfaces over these intervals in R_0 and B_0 for several scalar-valued quantities (quantities of interest or QoIs). Sparse grid Smolyak interpolants are used here to limit the number of model evaluations required to obtain interpolants of a given order [54]. In particular, a univariate nested scheme of Kronrod-Patterson sequences is used to determine the evaluation points for the response surface [49]. This has two advantages, first these sequences are designed to limit the growth of the Lebesgue constant of the interpolation with increasing polynomial order g. Second, because the sequences are nested, points used for interpolation at a specified order can be reused (i.e. recycled) at higher order.

This is illustrated in figure 4. As is clear from the diagram, the second order accurate response

Table 1: Simulation Parameters Kept Constant in all Cases

Parameters	Values
α (6)	2.184
β (6)	0.8540
a (minor radius)	0.6086 m
δ (12-13)	0.3536
$R_0/L_n(r_\delta)$	2.2
T_0/B_0^2	550.76 K/T^2

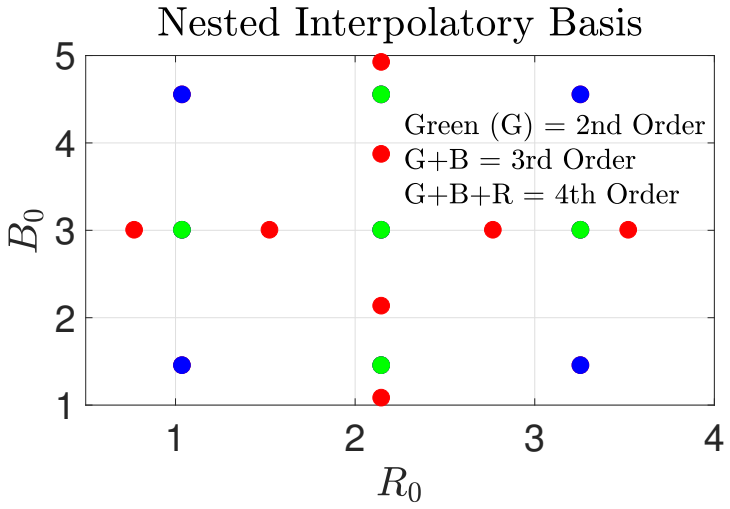


Figure 4: Univariate nested Kronrod-Patterson points. In green the $g = 2$ case, green and blue (G+B) the $g = 3$ points, and in green, blue, and red (G+B+R) the $g = 4$ points.

surface $g = 2$ requires 5 evaluation points, while the cubic response surface recycles those 5 points and requires only 4 additional evaluations. In contrast, for example, a tensor product Gaussian quadrature rule, which would require the evaluation of 13 additional points at $g = 3$. Here polynomial response surfaces of order up to $g = 3$ are considered, so 9 evaluations of the model using the gyrokinetics code XGC1 are required, labeled case 1-9 in table 2. In addition, the cyclone base case (i.e. the "nominal case," and case 0) is evaluated to be used as a test of the accuracy of the interpolant (case 0). In table 2, the values of R_0 and B_0 are listed along with several other parameters that are determined as in Section 2.1.2.

Finally, to define the numerical simulation in XGC1, several numerical parameters and a grid are required. First the number of marker particles is kept the same in all simulations at 307.2 M. The XGC1 grid is generated for the circular toroidal cross section, and 32 cross sections are equally spaced in φ. The grid in the cross sections are different in each simulation case because the grid is generated using gengrid2/simmetrix in the magnetic coordinate defined by ψ, rather than r. However, in each case, the grid is tuned to preserve mesh-conditioning. Finally, the simulation time step is set to scale with R_0, with $dt/R_0 = 3.1508 \times 10^{-3}$ s/m.

Table 2: Values of Parameters that Vary from Case to Case

Case	R_0 [m]	B_0 [T]	R_0/L_T	$\psi(a)$ [Wb]	$\hat{r}_\delta$
0	1.7	1.9056	1.19457	0.21063	0.17056
1	1.0388	3.0	2.56117	0.34913	0.28603
2	2.1480	1.4508	1.15528	0.15870	0.13435
3	2.1480	3.0	4.15400	0.32816	0.13435
4	2.1480	4.5492	10.57824	0.49763	0.13435
5	3.2572	3.0	5.25519	0.32510	0.08821
6	1.0388	1.4508	1.02740	0.34913	0.28603
7	1.0388	4.5492	13.32519	0.52942	0.28603
8	3.2572	1.4508	1.03984	0.15721	0.08821
9	3.2572	4.5492	12.75090	0.49297	0.08821

Note that only R_0 and B_0 are independently varied. The remaining quantities are derived as in Sections 2.1.2 and 2.1.3

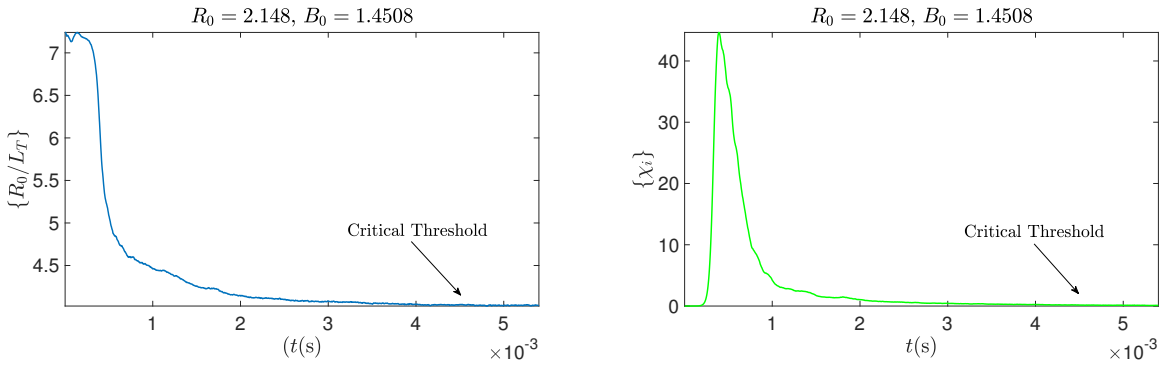


Figure 5: Basic analytics at the first support point. The critical point marks the transition to the stationary state, given in volume-averaged characteristic quantities.

2.1.5 Diagnostics

To evaluate and analyze the Dimits shift we will use a characteristic measure of the ion thermal conductivity $\{\chi_i\}$ and ion temperature gradient $\{R_0/L_T\}$, defined by:

$$\begin{aligned} \{R_0/L_T\} &= R_0|\{\nabla T\}/\{T\}|^{-1}, \\ \{\chi_i\} &= \{q_D\}/\{n\nabla T\}. \end{aligned} \tag{14}$$

To obtain these we need the surface-averaged dissipative heat flux: $q_D = \langle Q_e \rangle - q_c$. The local $\vec{E} \times \vec{B}$ energy flux is defined as

$$Q_e = \int |\vec{v} - \vec{u}|^2 \vec{v}_{\vec{E}\times\vec{B}} \cdot \vec{n} f(\vec{x}, \vec{v}) d^3v,$$

with $\vec{v}$ the velocity, $\vec{n} = \frac{\nabla\psi}{|\nabla\psi|}$ the surface normal, $\vec{u}$ the mean velocity (or the first moment in $\vec{v}$), and $\vec{v}_{E\times B} = \frac{\vec{E}\times\vec{B}}{|B|^2}$. The value is then surface-averaged in the usual way.

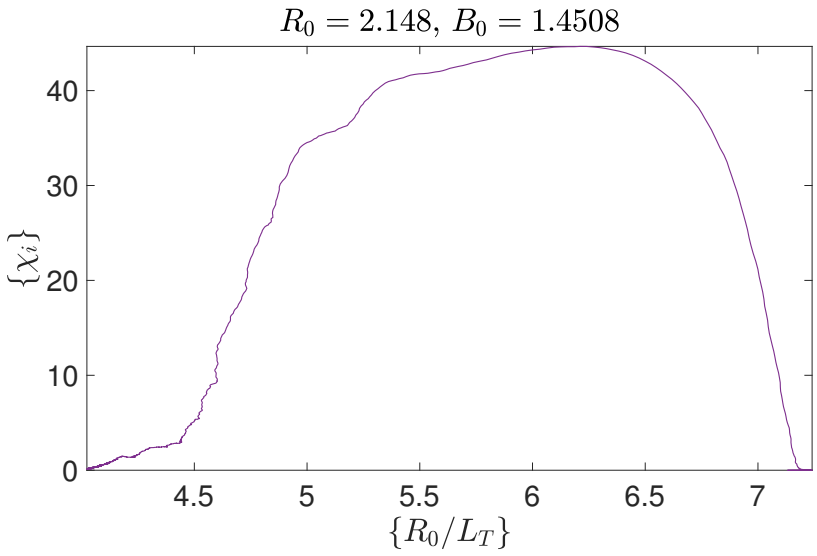


Figure 6: A phase space diagram of the quantities from Figure 5: $\{\chi_i\}$ as a function of $\{R_0/L_T\}$.

Similarly, the convective flux is defined by

$$q_c = \gamma \langle p \rangle \langle \vec{v}_{\vec{E}\times\vec{B}} \cdot \vec{n} \rangle = \gamma \langle T \rangle \langle \vec{\Gamma}_{\vec{E}\times\vec{B}} \cdot n \rangle,$$

with $\gamma = 3/2$ [57], and the $\vec{E} \times \vec{B}$ particle flux defined by $\vec{\Gamma}_{\vec{E}\times\vec{B}} = \int \vec{v}_{\vec{E}\times\vec{B}} f(\vec{x}, \vec{v}) d\vec{v}$.

2.2 Characteristics of the Stationary State

The experimentally identified "marginal stability" equilibrium is reached after an initial transient relaxes to a statistically stationary state. It is important to clarify the language here. The stationary state of the system is demonstrably stable in the sense of dynamical systems in that the system evolves toward it; that is, it is an attractor (see Section 2.3.1 for more discussion of the stationary state). It is referred to as a "marginally stability" not to indicate that, for example, the basin of attraction is particularly shallow, or to that anything particular about the attractor is known a shallow, or that anything particular about the attractor is known a priori. Rather "marginal stability" is a term used in the experimental literature to describe the observation that characteristics of the stationary state are delicate; that is, they depend sensitively on the values of system parameters.

The profiles of surface-averaged temperature $\langle T/T_0 \rangle$ and gyrocentered density $\langle n/n_0 \rangle$ change relatively little as the system evolves from the initial condition to a stationary state, as seen in Figs. 7. Shown are two examples, the nominal case (case 0 in Table 2) and case 4, which, among the cases used to inform the second order surrogate, produces the largest value of $\{R_0/L_T\}$ (see Table 3). Note that there is very little change in either the temperature or density profile in case 0, while in case 4 there is a much larger change near the center line ($r = 0$) of the torus.

Studies of the critical threshold for enhanced transport commonly examine the volume-averaged temperature gradient $\{R_0/L_T\}$ and effective thermal conductivity $\{\chi_i\}$ as functions of time. These quantities are defined in Section 2.1.5 and have been found to be good diagnostics characterizing this stationary state (marginal stability equilibrium). The upshift, or sometimes "Dimits shift," of the onset of turbulent transport relative to the onset of linear instability is often depicted in

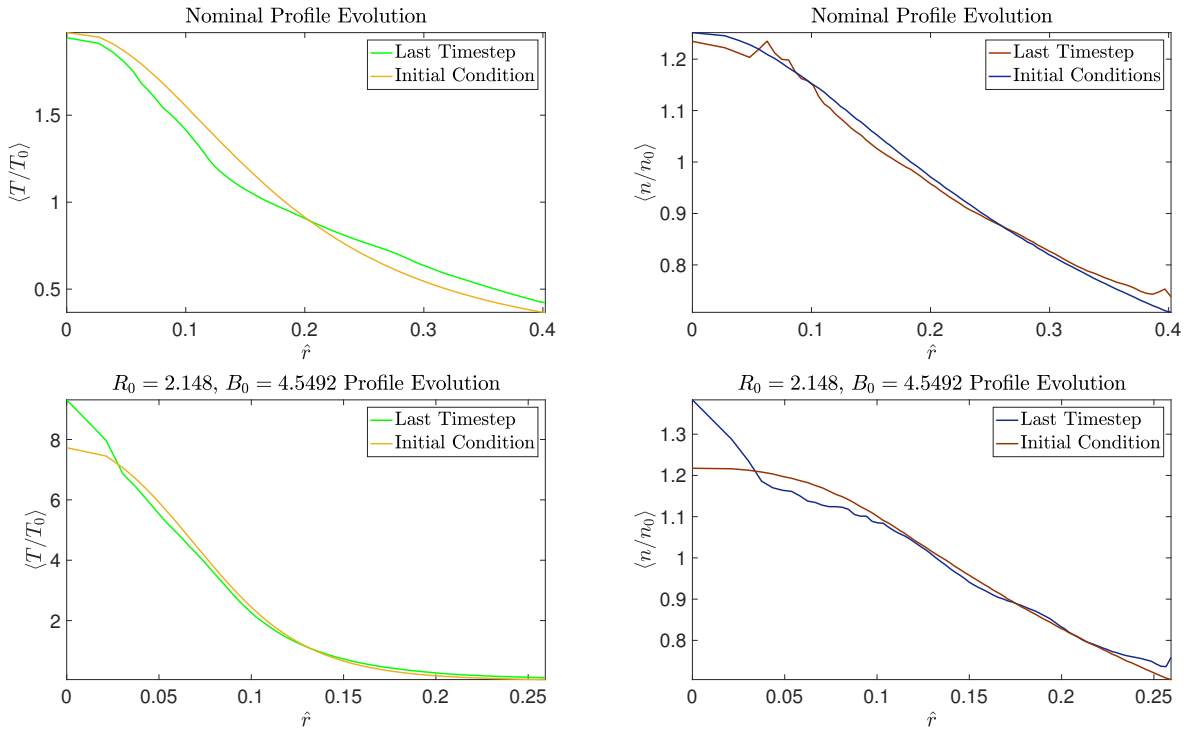


Figure 7: The numerical temperature and gyrocentered density profile evolution.

a phase space representation as in Fig. 6. In this picture the upshift corresponds to viewing the critical threshold as function of $\{\chi_i\}(\{R_0/L_T\})$. For example, in Fig. 6 the critical threshold occurs around $\{R_0/L_T\} \approx 4$, which is upshifted from the linear value, which we can infer as < 4. Alternatively the upshift is merely a translation along the vertical axis in $\{r_0/L_T\}$ of Fig. 5

Here Fig. 5 shows the time evolution, where the onset of the stationary state is approximated at a threshold value of $d\{R_0/L_T\}/dt = \epsilon_s$ for ϵ_s small. We characterize this approximate value as the "critical threshold." The dependence of the onset of the stationary state (i.e. of the critical threshold) on the magnetic field $\vec{B}$ can be approximated by understanding the scaling dependence of $\{\chi_i\}$ on $\vec{B}$. As mentioned above, there are several different well-established scaling regimes that characterize this dependence. The scale invariance models suggest that $\chi_i \propto (\rho^*)^{\alpha_\rho}$, so that there is the Bohm $\alpha_\rho = 0$, the gyro-Bohm $\alpha_\rho = 1$, and the Goldston $\alpha_\rho = -1$ scalings. Due to the fact that our parameter scan sweeps over multiple R_0 and B_0 regimes, it is perhaps not surprising that none of these scalings are strictly observed in our results. Even more so, these scalings are really indicative of local scaling in $\hat{r}$, and thus the global quantities are not necessarily anticipated to be a good indicator of the systems state. This can be emphasized by looking at experimental shots, such as those in DIII-D [65], where ρ^* scalings differ (transitioning locally between Bohm and gyro-Bohm) merely as a function of $\hat{r}$.

In this work we observe strong sensitivities to both the onset of the initial transient, as well as relaxation to the appropriate stationary state, on the magnetic field B_0. This dependence is encoded in our choice of the γ_1 parameter, and we find that this parameter choice demonstrates strong sensitivities. In contrast, the sensitivity of γ_1 as a function of the major radius R_0 is observed to have a much weaker effect on transport. It should however be noted that since in our tests we hold the $\langle q \rangle$ profile constant in each simulation, from the scaling law perspective, the observed

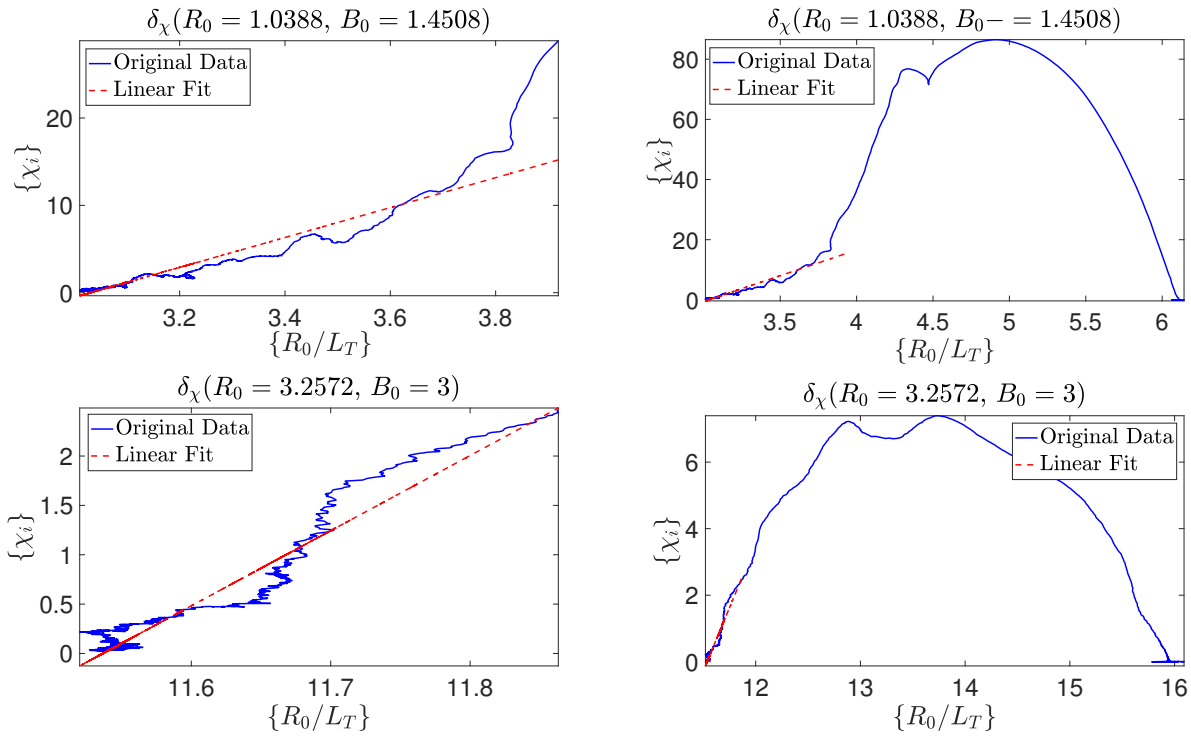


Figure 8: Departure rates as approximated from from the modal coefficients of the linear fit near the critical point in the gradient.

sensitivities are really a result of increasing both B_0 and the plasma current I simultaneously.

2.3 Low-fidelity surrogates in the critical threshold of the gradient

The response surface interpolated from the calibration data is plotted in Figure 9. Here we use second order accurate support points to recreate an interpolatory response surface in $\{R_0/L_T\}$. For the second order accurate curve six support points are taken Fig. 3, with five being the second order Kronrod-Patterson points, and the sixth being an asymptotic point easily inferred from the physics. The second order fitted surface $f = f(R_0, B_0) = f(\hat{x}, \hat{y})$ in this case takes the form:

$$f = c_0 + c_{10}\hat{x} + c_{01}\hat{y} + c_{20}\hat{x}^2 + c_{11}\hat{x}\hat{y} + c_{02}\hat{y}^2,$$

where the fitting procedure feature scales the values of $\hat{x} = (x - x_0)/\sigma_x$ and $\hat{y} = (y - y_0)/\sigma_y$, by the mean and standard deviations, respectively, such that $x_0 = 1.790$, $\sigma_x = 1.123$, $y_0 = 2.5$, and $\sigma_y = 1.568$.. The second order accurate surface is reported in table 4.

The response surface 9 shows a global polynomial fit of the calibration data. The nominal point is then plotted in red, and as is clear from visual inspection, the nominal point has a moderately small residual, with 5.64% relative error. Here this is an indicator that the model is fairly predictive to second order for the co-validation point.

To test the convergence behavior in g we run the third order accurate Kronrod-Patterson points. This requires the evaluation of four additional data points in $\{R_0/L_T\}$. To further resolve this surface we also choose additional points that obey the asymptotics of the solution. These additional

Table 3: Second order data

R_0	B_0	$\{R_0/L_T\}$
1.0388	3	5.4214
2.1480	1.4508	4.0325
2.1480	3	9.1700
2.1480	4.5492	18.73279
3.2572	3	11.5422
0	0	0

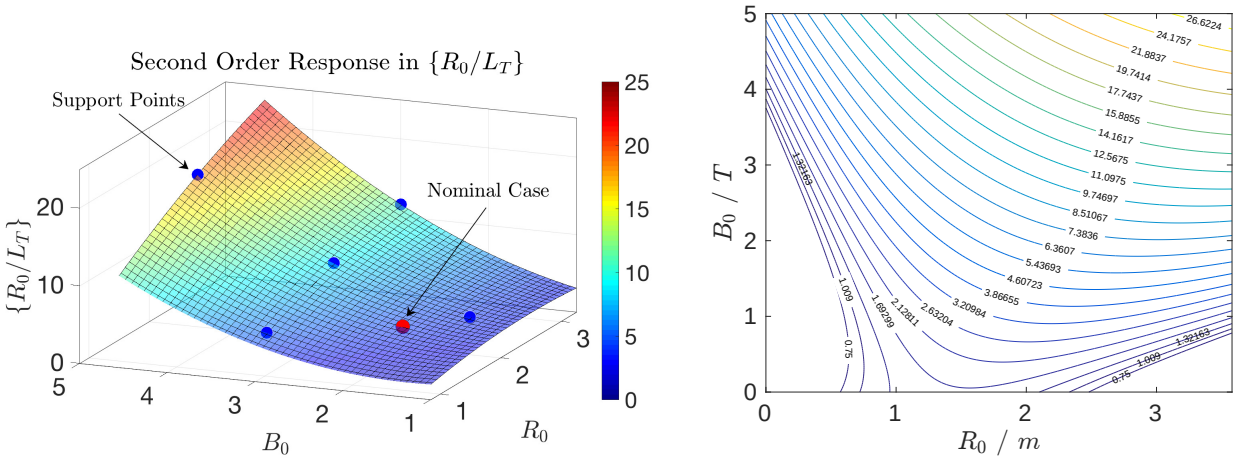


Figure 9: On the left is the second order response surface with relative error 5.64%. On the right is the contour plot

Table 4: Fitted Second Order coefficients

c_0	c_{10}	c_{01}	c_{20}	c_{11}	c_{02}
6.115	3.089	5.536	-0.7054	1.441	2.268

Table 5: Third order data

R_0	B_0	$\{R_0/L_T\}$
1.0388	1.4508	3.0274
3.2572	1.4508	4.5729
3.2572	4.5492	23.0767
1.0388	4.5492	19.4155
3.2572	3	11.5422
0	i	0
i	0	0

Table 6: Fitted Fourth Order coefficients

c_0	c_{10}	c_{01}	c_{20}	c_{11}	c_{02}
4.637	0.4007	5.329	-2.664	-0.1466	4.423
c_{30}	c_{21}	c_{12}	c_{03}	c_{40}	c_{31}
1.68	-2.418	0.7643	1.4515	0.8394	2.377
c_{22}	c_{13}	c_{04}			
-1.196	0.1799	-1.022			

points are listed in Table 5, where $i = 1, \ldots, 5$. Because of the addition of these asymptotic points, it is now possible to fit a fourth order surface to the $g = 3$ Kronrod-Patterson points. The fourth order surface $f = f(R_0, B_0) = f(\hat{x}, \hat{y})$ that is fitted is:

$$\begin{aligned} f = c_0 + c_{10}\hat{x} + c_{01}\hat{y} + c_{20}\hat{x}^2 + c_{11}\hat{x}\hat{y} + c_{02}\hat{y}^2 \\ + c_{30}\hat{x}^3 + c_{21}\hat{x}^2\hat{y} + c_{12}\hat{x}\hat{y}^2 + c_{03}\hat{y}^3 + c_{40}\hat{x}^4 \\ + c_{31}\hat{x}^3\hat{y} + c_{22}\hat{x}^2\hat{y}^2 + c_{13}\hat{x}\hat{y}^3 + c_{04}\hat{y}^4, \end{aligned}$$

where the fitted coefficients are reported in table 6. Here $x_0 = 1.504$, $\sigma_x = 1.295$, $y_0 = 2.1$, and $\sigma_y = 1.809$.

As shown in Figure 10, the third order fit shows almost the exact same relative error as the second order fit, here being 5.79%. This again suggests convergence behavior, particularly as we have used 10 asymptotic points to fit a fourth order surface to the third order data. It is worth noting here that the response surface $\{R_0/L_T\}$ is nonlinear both as a function of B_0 as well as R_0, showing substantial rippling as B_0 increases. This may indicate that as the magnetic field strength increases, the surrogate model becomes more nonlinear, and therefore possibly less predictive.

2.3.1 Derived features

The states of marginal stability of interest to experimentalists are characterized mathematically by the stationary states observed in our simulations with an approximate onset at and beyond the critical threshold. These stationary states are attractors in the classical sense of dynamical systems.

Ideally one would like to be able to infer from the simulation runs further model behavior of the turbulent regime that is amenable to representation by global surrogates. An appealing candidate here is what we refer to as the "attractor rate," which is in principal associated to the onset of

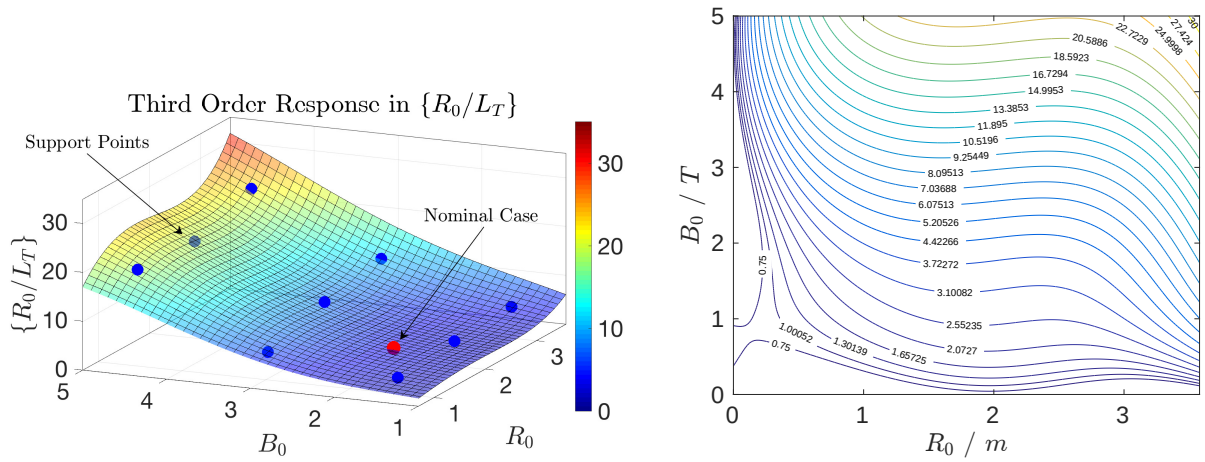


Figure 10: Third order response surface with relative error 5.79% on left; on right the corresponding contour plot.

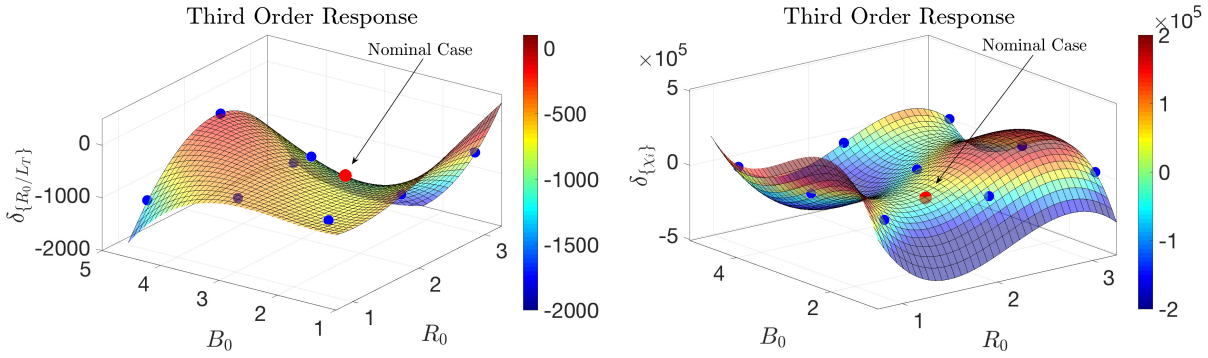


Figure 11: Left: third order response surface in $\delta_{\{R_0/L_T\}}$, with relative error 39.47%; right: third order response surface in $\delta_{\{\chi_i\}}$, with relative error 231.94%

the stationary state. Broadly, the attractor rate here will be defined as the rate of approach to the stationary state.

There are a number of ways of defining an attractor rate relative to the data from the simulation runs in Section 2.3. For example, we can define the attractor rate in $\{R_0/L_T\}$ as a function of time, such that the attractor rate $\delta_{\{R_0/L_T\}}$ indicates the rate at which the gradient approaches the stationary state. Since simple inspection of the behavior of the gradient (e.g. Fig. 5) indicates possible exponential decay, we simply plot a log-linear plot and regress a line of best fit. Clearly from Fig. 12 this gives a relatively good estimate of the decay rate approaching the critical threshold. Similarly we can define the attractor rate of $\{\chi_i\}$ using the same approach, and denote it by $\delta_{\{\chi_i\}}$.

Notice here also that the confinement times of energy and density in the tokamak – and in other magnetic confinement schemes – scales roughly as $\tau_E \sim L_a/\{\chi_i\}$, where $L_a = Ca^2$ is the characteristic scale length of the device given C a constant. In practice, all experimental discharges are assumed to be effectively close to a stationary state (i.e. marginal stability) due to the very strong transport created by deviating from the nonlinear stability condition. As a consequence, it would be very advantageous to be able to estimate χ_i as a measure of the deviation away from marginal stability [3] .

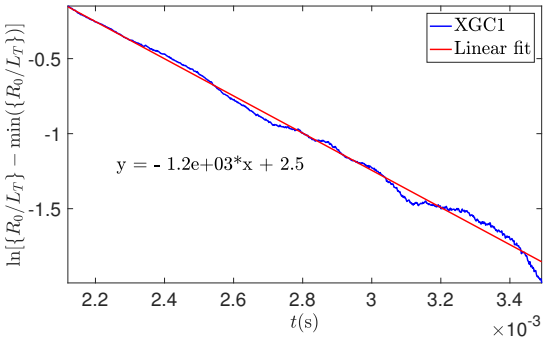


Figure 12: Log-linear plot of $\{R_0/L_T\}$ where the attractor rate is given by the slope of the best fit line, $\delta_{\{R_0/L_T\}} = -1.2\text{e+}03$.

Another way to achieve this estimate in χ_i is to use the phase space relation of $\{R/L_T\}$ to $\{\chi_i\}$ in Figure 6. In other words, we can infer the linear perturbation $\{\chi_i(R_0/L_T)\} = \delta_0\{R/L_T\}$. Notice that both definitions above have essentially relied on the fact that visually it appears that $\{\chi_i\} = e^{-cx}$, and likewise $\{R_0/L_T\} = e^{-bx}$ (e.g. from the simulation runs in Section 2.3). But since $\{\chi_i\}(\{R_0/L_T\}) = \delta_0\{R_0/L_T\}$, we get that $d\{\chi_i\}/d\{R_0/L_T\} = \delta_0$. From above we have that $d\{\chi_i\}/dt = -ce^{-cx}$ and $d\{R_0/L_T\}/dt = -be^{-bx}$, so with standard smoothness assumptions we can infer that $d\{\chi_i\}/d\{R_0/L_T\} = -\delta_0 e^{(c-b)x}$, which shows that all δ_0 is really measuring is the relationship between c and b. Since $\ln(-ce^{-cx}) = \ln(-c) - cx$ implies that $\delta_{\{\chi_i\}} = -c$, and likewise that $\delta_{\{\chi_i\}} = -b$, this really does imply that up to the $(c-b)$ in the exponent δ_0 is primarily a measure of the ratio between the two.

To establish these variational relations, we construct surrogates for each of these quantities using similar methods to those in Section 2.3. For the gradient $\delta_{\{R_0/L_T\}}$, the third order surrogate for the attractor rate is shown in Fig. 11. Notice that the substantial nonlinearity shown here leads to a very poor predictor of the surface, giving only $\sim 40\%$ relative error for the nominal point. Similarly for $\delta_{\{\chi_i\}}$ the nonlinearity is even more severe. As a consequence of this extreme nonlinearity in these two surrogates, it is not surprising that the ratio of these two, which is essentially captured in the surrogate for δ_0, is in fact globally unstable leading to a non-informative surrogate (not shown).

Clearly increasing the order g of the surrogate may be effective in capturing some of these nonlinearities, and in that context asking the question: "how high might g need to go before one can capture this behavior?," becomes interesting. However, in the present context the nonlinearity is not very surprising, and in fact, the surrogate itself, even if fit to high order and given to within sharp relative error, might very well be less informative that would be preferred. The reason for this is multifaceted.

In the case of turbulence, we are presented with a stationary state that represents a stable attractor in a dynamical system. This attractor can be viewed as a stable level set (also the "section of attraction") that nearby solutions evolve to. The first difficulty arising in this picture, is that the attractor rate can only be interpreted as meaningful in a statistical sense, because the local high dimensional manifold that the attractor is embedded within looks different from every direction. As a consequence, the attractor, as a meaningful quantity, can only make sense when averaging over all nearby routes to the attractor; that is all trajectories through the basin of attraction. In our case, however, we cannot afford to compute ensemble averages (e.g. $> 10,000$ runs), and are ultimately

trying to infer a statistical quantity from a single sample, knowing that the turbulent state positions us randomly somewhere near the basin of attraction.

Moreover, it is plausible that because the relaxation occurs mostly during the initial transient, another type of hysteresis may be responsible for dampening oscillations close to the stable point. As transport is largest on the outboard side of the torus, a parallel pressure gradient develops there in the presence of strong turbulence, and the parallel pressure gradient might respond by a geodesic acoustic oscillation (i.e. from the GAM) known to modulate transport. Such a dynamical mechanism would not necessarily influence the stationary state or even the critical threshold, but could still contaminate the attraction rate. We conjecture that it might be possible to start the simulations closer to the stability point to avoid the contaminating relaxation dynamics, but such an analysis is beyond the current scope of this paper.

Finally, one of the base assumptions we made in correlating the attractor rate to the confinement time τ_E was that the attractor rate is fundamentally correlated to the departure rate. The departure rate would be the rate that local perturbations depart from the stationary attractor. But this is not rigorously correct. The reason is, the attractor rate is the rate an unperturbed state statistically and passively moves through the basin of attraction into the stationary regime. The departure rate, in contrast, is the rate at which a state moves away from the stable attractor, which can happen due to external perturbations. It is clear then that it can more likely be the type and form of the perturbation, not the stability regime the perturbation is acting upon, that almost entirely determines this rate. In fact, this is the case so much so that it is probably more accurate to call the departure rate the "rate of external perturbation."

Because of these reasons, it is important to take great care in trying to associate local variation in the upshift of the critical gradient to infer global, or even local, properties of the turbulent regime. The general experience appears to be that while present transport models are often successful in predicting devices that they have been developed to model, they are less successful across a wide range of experiments/machine configurations. If the goal is to obtain a model with predictive capability across a large range of machine parameters, even for those not yet completed, we feel that additional effort may need to be expended exploring the vicinity (and hence the statistics) of the stable attractor.

3 Scenario Extrapolation

Prediction under uncertainty is often an expensive and complicated process[44, 45]. The trade off between the computational cost of complex models and the loss of predictive accuracy associated with simpler models may be addressed by emerging multifidelity UQ approaches[43, 48].

In magnetic confinement fusion, there is a clearly defined set of prediction scenarios, corresponding to predicting ITER or DEMO performance. Even a "high-fidelity", extreme-scale model such as XGC, still has a parameter space of large enough dimension to make a brute-force, sampling-based predictive process impossible. The traditional "solution" to this problem is to employ a combination of dimension reduction and surrogate modeling. To achieve dimension reduction, we will need gradients of the prediction scenario P with respect to scenario paramters θ_i. Even if modest dimension reduction is achievable, the resulting cost of even "low-fidelity" kinetic ITER simulations is sufficient to seriously limit the number of training points for the surrogate model. The resulting predictive error will be sufficiently large to reject the surrogate approach. We outline our workaround, an approach we will broadly label as predictive extrapolation.

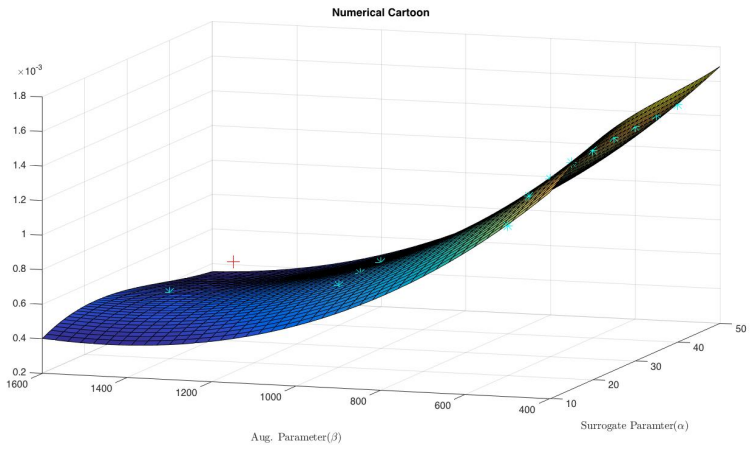


Figure 13: Numerical cartoon depicting a simple example of an augmented surrogate

A surrogate model is already a form of predictive extrapolation-an inexpensive model is fit to training data, and other samples of the prediction scenario are extrapolated from the inexpensive surrogate. In our approach, this is taken a step further-we extrapolate the prediction scenario from other nearby prediction scenarios. One way of performing this process is the construction of an augmented, or "telescoping" surrogate.

3.1 Scenario Extrapolation Methods

There are two fundamental assumptions in the construction of an augmented surrogate of a n-dimensional predictive scenario $P(\theta_1, \theta_2, \ldots, \theta_n)$.

1. There exists small number m of (usually deterministic) parameters $d_1, d_2, \ldots, d_m$ that characterize nearby n-dimensional predictive scenarios $\tilde{P}_i(\theta_1, \theta_2, \ldots, \theta_n)$.

2. The cost of a sample from the nearby scenario, $C_{\tilde{P}_i}$, is much less than the cost C_P of a sample from the desired prediction scenario.

The *augmented surrogate* is a surrogate model $S(\theta, d)$ constructed on training data $\{(\theta, d), \hat{P}(\theta, d)\}$ in the $m + n$-dimensional parameter space. We briefly illustrate this approach via the following numerical cartoon.

Given a two point boundary value problem

$$-(f(x, \alpha)u')' + \beta u' = g(x, \alpha), \quad u(0) = u(1) = 0. \tag{15}$$

where α is an uncertain parameter (θ) and β is a known parameter (d), with a prediction scenario $P(\theta)$ equal to the average of u in an subinterval of [0,1]. The (naive) simulation cost depends on both α and β, with simulation cost increasing as β is increased. In Figure 13, we take eight samples(blue *) of $P(\beta = 600)$, three samples of $P(\beta = 1000)$, and one sample at $P(\beta = 1400)$. A polynomial interpolant is constructed from the training data ($S(\alpha, \beta)$ and is compared against the true value of $P(\alpha, \beta)$(the red +).

We may employ any of the sparse grid designs discussed in Section 2. But the very nature of the augmented surrogate suggests two strategies: adaptive sparse grids, or unstructured surrogates based on radial basis function approaches.

An adaptive sparse grid method

The classical sparse grid method is dimension agnostic[8]. All interactions of the same order are treated equally. Often, a small subset of variables and interactions contributes significantly to the variability of the function $P(\theta, d)$. If the variability in the $theta$-dimensions is greater than the variability in the scenario parameters($\{d_i\}$ then the overall cost of constructing of constructing the larger dimensional surrogate due to the cheaper computational cost of $\tilde{P}_i$.

We modify the greedy algorithm for constructing h-adaptive generalized sparse grid (h-GSG) in [26]. The hierarchical surpluses for the current sparse grid are modified with a cost weight $W_C(d_1, d_2, \ldots, d_m)$ that approximates the relative cost of simulating the added sparse grid point, and a m-dimensional distance metric that penalizes sparse grid samples that are too far away from the prediction scenario $P(\theta)$. This encourages parameter exploration in θ_i-dimensions at inexpensive simulation levels, while rewarding coarse grid points that reduce the local surrogate error near the prediction scenario $P(\theta)$.

Radial basis function surrogates

Radial basis function surrogates [52] offer attractive features for the augmented surrogate approach. The built-in support for unstructured interpolation of sparse data allows the direct integration of simulations (or experiments) from relevant nearby prediction scenarios $\tilde{P}_i$.

In prediction scenarios relevant to ITER[16], there are a well described set of deterministic, dimensionless plasma scaling parameters (ρ^*, β, ν^*) used in scaling studies for confinement and transport. There are additional dimensionless parameters to describe reactor geometry that could provide additional "scaling" (d) parameters. With the setup, the augmented surrogate $S(\theta, d)$ can be trained using existing simulations and experimental data (for the same response surface) from JET, DIII-D, C-Mod, etc. In the sense, we are generating additional surrogate training data "for free" at the cost of a few deterministic parameters.

3.2 Scenario Extrapolation Results

In many ITER prediction scenarios where an electrostatic gyrokinetic simulation is valid, the dominant dimensionless scaling parameter is $\rho^* = \rho_i/a$, the ratio of the ion gyroradius to the tokamak minor radius. As a preliminary test of extrapolative prediction using an augmented surrogate, we augment our prediction scenario with one scaling parameter, $a/\rho_i = 1/rho^*$. As the gyrokinetic PIC grid must resolve the ion gyroradius with a moderate particle density per simulation cell, the cost $C_{\tilde{P(\theta)}}$ scales at least quadratically in ρ_*^{-1}. An illustration of the proposed scaling from CYCLONE-like base case(1x) simulations to ITER relevant ρ^* is given in Figure 14.

Description of Prototype Surrogate

We have one simulation parameter for the applied external heating(P_0), with a prediction scenario of estimating the averaged ($.3 < \psi_n < .7$) steady-state normalized ion temperature gradient (R/L_T). This is more an example of the use of a surrogate for parameter space exploration, optimization, or extrapolation as the R/L_T variation to measurement inaccuracies in P_0 is negligible.

We scale ρ^* by scaling the background magnetic field, rather than the tokamak minor radius a. The "base case", corresponding to a normalized a/ρ_i value of 1, is described in Table 7.

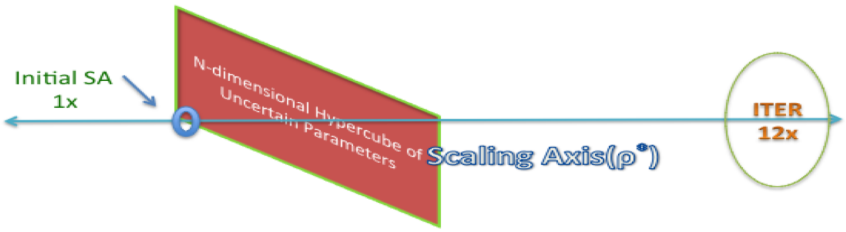


Figure 14: Illustration of prototype augmented surrogate for electrostatic extrapolation

Ion temperature at magnetic axis T_i	14keV
Magnetic field B_0	1.91 T
Simulation range	$\psi_N \in [0, .9]$
Tokamak major radius	1.75 M
adiabatic electrons	
simulation δt	$3.85 \times 10^{-8} s$

Table 7: Cyclone-like "Base Case" simulation parameters

We use four training samples at the base case, two training samples at $1.5B_0$, and one training sample at $2B_0$. The resulting surrogate response surface is plotted, and $S_{R/L_T}(P_0 = 200MW, 2B_0)$ is compared against $R/L_T(P_0 = 200MW, 2B_0)$ in Figure 15. The prediction value for R/L_T is reasonable, with excellent agreement between $\partial S/\partial P_0$ and $\partial(R/L_T)/\partial P_0$. The excellent parameter gradient motivates another common use of surrogates: estimating parameter derivatives for dimension reduction via gradient-based methods such as active subspaces[11].

4 Resampling Algorithms

Particle-in-cell (PIC) methods are well-established and widely used numerical methods for the solution and evolution of differential equations [6, 24, 22]. In such methods a set of marker particles is used to represent a distribution in some phase space, and its evolution in time. The distribution being represented could be that of physical particles (e.g. atomic scale particles or particulates in a multiphase flow), or a probability distribution. In either case, in PIC methods, the evolution of the distribution and therefore of the marker particles is governed by interaction with field variables that are represented on a grid (thus the name).

A common requirement of PIC simulations, is to have enough marker particles in each cell for good statistical accuracy. Computational efficiency also requires a uniform distribution of marker particles over the phase space across the simulation cells. A well formulated resampling procedure, which is just a way by which the number of marker particles can be decreased or increased (down- or up-sampling), or redistributed, can be helpful in achieving these requirements. In fact, as it turns out, simply resampling periodically can address a surprisingly large number of common deficiencies in raw PIC implementations.

Even when a PIC simulation begins with an ideal number and distribution of marker particles, the evolution of particles has a tendency to degrade the quality of the sampled distribution over time. This can first occur because particles evolve away from the desired importance-sampled target distribution, and thus the relative distribution of particles can accumulate into regions of less

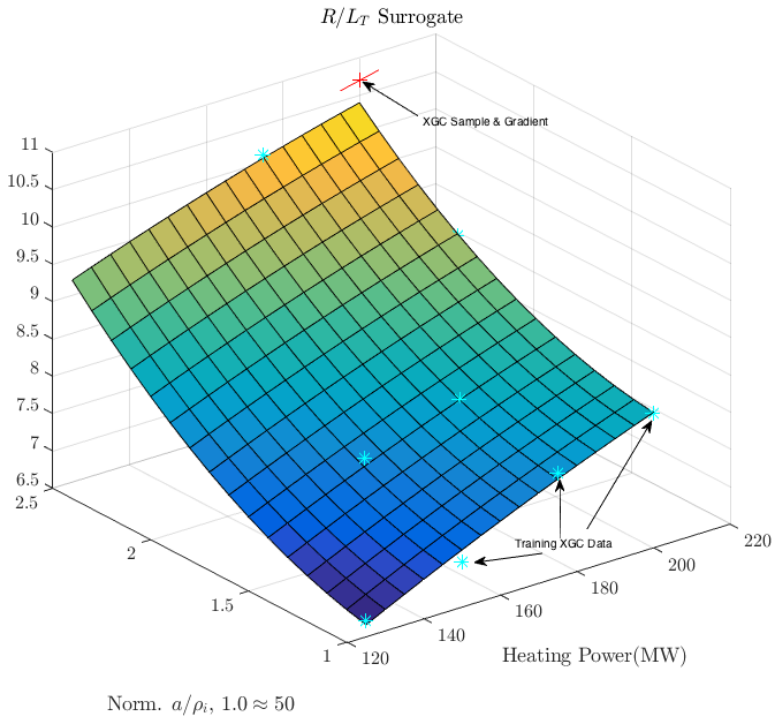


Figure 15: R/L_T surrogate as a versus input power P_0

importance, along with particle weights becoming distributed throughout the domain in inefficient ways. For example, in gyrokinetic PIC tokamak simulations, low weight particles are frequently loaded at the foot of the plasma pedestal, which overlaps with the scrape-off layer (SOL). The plasma pedestal region describes the transition zone in the plasma density, from high density in the core to low density in the SOL. Time-evolving the simulation, the low-weight particles, though initially loaded in correct proportions, mix with higher weight particles in the core, effectively spreading the marker particle weights and lowering the simulation efficiency in the critical pedestal and scrape-off layer regions. This has the effect of degrading the numerical representation of the distribution, and leads to a gross loss of resolution in the critical region of interest. This so-called 'weight spread' can however be substantially reduced by periodic resampling.

A related issue also arises with the application of the collision operator in a PIC simulation. In this context, for example, high energy atomic particles — often arising from non-Maxwellian processes — may be represented by a small number of marker particles in the region of interest. This clearly results in noisy approximations of the collision operator at high energy. By sampling the high energy part of the distribution with many lower-weight particles, however, and by maintaining this local state by periodically resampling the distribution, this problem can be alleviated.

A third setting in which resampling in PIC simulations arises, is when using the control variate technique as is done in δf simulators. In these methods, the particle weights evolve leading to sample degeneration wherein the weights become concentrated on fewer and fewer particles as the simulation proceeds; resulting in a poor representation of the distribution. This phenomena is similar to the sample degeneracy observed in particle filtering methods [19, 28, 29]. Again, periodic resampling to rebalance the particle weights can address this problem.

For these reasons, preserving the salient features of the distribution in a PIC simulation is more than just beneficial, it is in many ways essential in order to recover solutions that faithfully reproduce the continuum behavior. It is however also the case that just preserving the statistical properties of the distribution is not always enough to ensure correct solution behavior. For example, in resampling a particle distribution in a plasma PIC simulation, it can be equally important to preserve the low-order moments of the particle distribution function such as energy, mass, and momentum, etc. In this context these moments represent physical conservation laws, that further serve as sources for recovering electrostatic fields and transport phenomena. In addition, it is also crucial to conserve these corresponding grid quantities (such as current and charge density) after particle resampling. Since these too are integral quantities, we refer to these as 'grid moments.'

A large number of methods for particle resampling that preserve various moments or distributional features of the solution have been explored. Many of these methods are focused on splitting and merging original particles and are capable of preserving derived features (e..g moments) to some degree of accuracy. For example, Lapenta [38, 37] proposed a scheme in which the number of particles is increased by splitting a single particle into various particles displaced over the system space. The number of particles is decreased by coalescence of two particles close to each other in phase space, i.e. "particle coalescence"). This algorithm can be easily extended to PIC simulations with two-dimensional and three-dimensional Cartesian grids and enables preservation of the overall charge, momentum, and energy. However, this algorithm is incapable of conserving the velocity distribution function both locally and globally, and the scheme is not directly extendable to 2D and 3D unstructured grids. Teunissen and Ebert [61] improved Lapenta's particle merging algorithm using the k–d tree method to search for the nearest neighbor. Following a similar procedure, Vranic et al. [64] divided the the momentum space into smaller cells for sorting particles that resulted in better local preservation of the energy, momentum, and charge.

A different two-dimensional method of coalescing particles in PIC codes is presented by Assous [2] that conserves the particle and cell charge and current densities as well as the particle energy. However, the method is limited to two-dimensional triangular cells only and its extension to other cell geometries as well as to three-dimensional analyses is not straight forward. Moreover, in this method only specific numbers of particles per cell after coalescence are possible depending on the integration points employed in the solution. Welch et al. [68] provided an extension of Assous method [2] to coalescing particles on 2D and 3D cells. This method is limited to orthogonal grids and similar to the method of Assous, in that coalescence might not be possible in some cases.

Luu et al. [40] presented a particle merging algorithm in which the phase space of a simulation is partitioned into smaller subsets. The algorithm merges particles that are close to each other and provides direct control over errors introduced by a merging event. Examining the performance of this algorithm indicates that momentum is conserved perfectly while there are discrepancies in energy conservation.

Pfeiffer et al. [50] proposed two algorithms for particle splitting and merging that use a 3D unstructured hexahedral mesh and are expandable to any cell geometry. The first method is computationally feasible and enables preserving particle charges, currents, and energies exactly, while the grid quantities of charge and current are only preserved approximately. The second method makes fewer assumptions in the velocity distribution function resulting in better conservation of the grid quantities (i.e. current and charge density), but requires higher numerical effort compared with the first method.

In this work, we develop a general algorithm for resampling particles that conserves any desired

features of the distribution, called Moment Preserving Constrained Resampling (MPCR). It differs from the techniques described above by using the fact that the marker particles are a sampling representation of an underlying distribution, and that this distribution can therefore be resampled when convenient. To accomplish this, MPCR uses a binning strategy based on any convenient discretization of phase space, making it suitable for general geometries and mesh/grid representations. MPCR uses constrained optimization techniques to produce a new set of particle positions and velocities, and to readjust the particle weights while accurately preserving necessary particle and cell information. In particular, MPCR: (1) enables preservation of any number of derived quantities (e.g. energy, mass, momentum, current density, charge density, etc.) up to roundoff error, (2) can be applied to any discretization geometry and is easily integrated into any PIC code, (3) can use a codes native discretization for particle binning, and (4) is computationally efficient.

The resampling algorithm developed here can be applied to any problem in which PIC methods are used to represent the evolution of distributions. But, in this project, the target is obviously PIC representations of collisional plasmas, particularly for the simulation of edge physics in tokamak plasmas. The target in this case is the gyrokinetic representation used in the XGC tokamak simulation codes [32, 13]. Details of the MPCR algorithm and its implementation in XGC is described in sections 4.1–4.3. A number of example problems demonstrating the capabilities of MPCR are presented in section 4.4.

4.1 Conserved quantities for particle resampling

While resampling the particles in a PIC code it is important to preserve key quantities of the particle distribution function f as well as specified grid quantities in order to maintain the physical properties of the plasma system. The general resampling strategy proposed here enables preserving derived particle and grid quantities with a high degree of accuracy. In this section we summarize the significant features and quantities that are preserved after resampling the particles.

The moments of f coincide with physically relevant quantities, and preserving them during resampling can reduce resampling error. For example, one might be concerned about preserving the following quantities,

$$n(\mathbf{x},t) = \int f(\mathbf{x},\mathbf{v},t)d\mathbf{v}, \tag{16}$$

$$\mathbf{m}(\mathbf{x},t) = m\int \mathbf{v}f(\mathbf{x},\mathbf{v},t)d\mathbf{v}, \tag{17}$$

$$K(\mathbf{x},t) = \frac{1}{2}m\int |\mathbf{v}|^2 f(\mathbf{x},\mathbf{v},t)d\mathbf{v}, \tag{18}$$

$$T_{kl}(\mathbf{x},t) = m\int v_k v_l f(\mathbf{x},\mathbf{v},t)d\mathbf{v}, \tag{19}$$

where n is number density, $\mathbf{m}$ momentum density, K kinetic energy density, and T_{kl} the stress density.

These in no way exhaust the number of important, and/or physically relevant moments however. Depending on the specific application, any number of different quantities might also be of interest. For example, due to gauge transforms that lead to different coordinate frames in gyrokinetic tokamak plasma simulations, it is often the canonical angular momentum that is of particular

interest, rather than $\mathbf{m}$. Similarity, it might also be beneficial to preserve the moments taken with respect to configuration space $\mathbf{x}$

$$\bar{\mathbf{x}}(\mathbf{v},t) = \int \mathbf{x} f(\mathbf{x},\mathbf{v},t) d\mathbf{x}, \tag{20}$$

$$\mathcal{R}_{kl}(\mathbf{v},t) = \int x_k x_l f(\mathbf{x},\mathbf{v},t) d\mathbf{x}. \tag{21}$$

In PIC algorithms, field solvers are frequently and necessary coupled to particle tracking algorithms. As an example, in gyrokinetic PIC solvers, the charge density ρ and the current density $\mathbf{J}$, defined by

$$\rho(\mathbf{x},t) = \int q f(\mathbf{x},\mathbf{v},t) d\mathbf{v}, \tag{22}$$

$$\mathbf{J}(\mathbf{x},t) = \int q\mathbf{v} f(\mathbf{x},\mathbf{v},t) d\mathbf{v}, \tag{23}$$

are used as the forcing functions that drive the solution of the electrostatic potential ϕ. Thus ρ and $\mathbf{J}$ have to be defined on the interpolation points of the solver, resulting in the so-called grid variables. In this context, it is essential that the operator that maps the particle configuration onto the grid for use in the global solver be consistent before and after particle resampling, so that underlying electromagnetic fields are unaffected by the resampling strategy.

4.2 The resampling algorithm

The MPCR method generates a collection of M^p macro-particles relative to the collection of N^p original particles representing $f(\mathbf{x},\mathbf{v},t)$. The new particle set can be either smaller $M^p < N^p$ or larger $M^p > N^p$ than the original set, but the focus of the algorithm becomes the preservation of the required moments and any additional salient features, which will be viewed as constraints in a constrained optimization problem.

4.2.1 Discretizing the phase-space.

In a general and spatially three-dimensional kinetic PIC solver, each particle carries with it seven attributes: (1) its weight w, (2)-(4) the three spatial coordinates $\mathbf{x}$, and (5)-(7) the three velocity coordinates $\mathbf{v}$. The first step of the particle resampling process is to divide the six-dimensional phase space $\boldsymbol{\zeta} = (\mathbf{x},\mathbf{v})$ into bins and sort the particles into those bins. There are then a total of N^p particles sorted into N^b bins, with bin i containing N_i^p particles. Depending on the application, it may be convenient to use the spatial discretization of the grid solver to demarcate the bins. However, MPCR is designed to handle any cell geometry.

4.2.2 Setting the number of new particles in each bin.

The target number of particles in each bin M_i^p must be determined. In general, a particle number density function $g(\boldsymbol{\zeta})$ can be defined, $\int_{\boldsymbol{\zeta}} g\, d\boldsymbol{\zeta} = M^p$ to allow importance sampling, so that in bin i:

$$M_i^p = \int_{\boldsymbol{\zeta}^i} g\, d\boldsymbol{\zeta}, \tag{24}$$

where $\boldsymbol{\zeta}^i$ is the phase space domain of bin i, i.e. range of position and velocity coordinates in each bin.

4.2.3 Determining new particle positions and velocities.

The algorithm is designed to address both *down-sampling* and *up-sampling*. The objective of down-sampling is to reduce the number of particles in each bin, while up-sampling aims to increase particle numbers. Some applications requires the global number of particles to remain approximately unchanged , i.e. $M^p \approx N^p$. In these cases the goal of resampling is to adjust the distribution in a way that reduces error and/or mitigates noise in the particle simulation.Clearly, to accomplish this resampling when particle number is held relatively constant, requires up-sampling in some bins and down-sampling in others.

The bin-wise particle down-sampling can be performed by drawing M_i^p samples from the f distribution restricted to $\boldsymbol{\zeta}^i$. Assuming that $M_i^p < N_i^p$, this can be easily accomplished by drawing samples of the original particles with probability proportional to their weights, using weighted sampling without replacement [70]. This way the original particles with higher weights are sampled more often. However, In the up-sampling case, $M_i^p > N_i^p$, one can draw M_i^p samples from uniform distribution restricted to $\boldsymbol{\zeta}^i$, i.e. uniformly sample the range of position and velocity coordinates in each bin.

4.2.4 Weight readjustment and conserving quantities.

All that remains is to set the new weights $\tilde{w}_j^i$ of each new particle. In doing so, the particle and grid quantities discussed previously will be preserved in each bin. The requirements of the particle flow are then that for each bin i:

$$\sum_{j=1}^{M_i^p} \tilde{w}_j^i = \sum_{j=1}^{N_i^p} w_j^i, \tag{25}$$

$$\sum_{j=1}^{M_i^p} \tilde{\boldsymbol{\zeta}}_j^i \tilde{w}_j^i = \sum_{j=1}^{N_i^p} \boldsymbol{\zeta}_j^i w_j^i, \tag{26}$$

$$\sum_{j=1}^{M_i^p} (\tilde{x}_k \tilde{x}_m)_j^i \tilde{w}_j^i = \sum_{j=1}^{N_i^p} (x_k x_m)_j^i w_j^i, \tag{27}$$

$$\sum_{j=1}^{M_i^p} (\tilde{v}_k \tilde{v}_m)_j^i \tilde{w}_j^i = \sum_{j=1}^{N_i^p} (v_k v_m)_j^i w_j^i, \tag{28}$$

where k,$m \in \{1, 2\}$, variables with an over-tilde $\tilde{\cdot}$ represent resampled particle quantities, and the particles have been renumbered by bin (superscript i) to be sequentially numbered within each bin (by subscript j) for both the original and resampled particles. It can be easily inferred from these relations that (25) ensures particle number density conservation, (26) indicates preservation of momentum and first order configurational moments, and (27) and (28) denote conservation of energy and other second order and off-diagonal moments of the particle distribution function.

Moreover, by conserving grid quantities (charge and current densities) we will require that,

$$\sum_{\substack{k \\ \mathbf{x}_k \in V_{\mathbf{x}}^i}} \tilde{\Lambda}_k \tilde{w}_k^i = \sum_{\substack{k \\ \mathbf{x}_h \in V_{\mathbf{x}}^i}} \Lambda_h w_h^i, \tag{29}$$

$$\sum_{\substack{k \\ \mathbf{x}_k \in V_{\mathbf{x}}^i}} \tilde{\mathbf{v}}_k^i \tilde{\Lambda}_k \tilde{w}_k^i = \sum_{\substack{k \\ \mathbf{x}_h \in V_{\mathbf{x}}^i}} \mathbf{v}_k^i \Lambda_h w_h^i, \tag{30}$$

where $V_{\mathbf{x}}^i$ denotes the spatial volume of bin i and $\Lambda(\mathbf{x})$ represents a general shape function that accounts for the distance information between the interpolation point and particle position. A special case of this shape function for a specific PIC code with triangular mesh will be discussed in section 3.3.

Since the resampling of the particles was weighted by the w_j^i of the original particles, weights of the new particles $\tilde{w}_j^i$ would ideally be uniform within each bin. However, enforcing this would over-constrain the system. Instead, the weights are made as close to uniform over the bin as possible while satisfying the constraints (25-30). In particular, the new weights $\tilde{w}_j^i$ are set as

$$\{\tilde{w}_j^i\}_{(j \in 1,2,\ldots M_i^p)} = \underset{\tilde{w}_j^i \in \mathcal{C}^i,\ \tilde{w}_j^i > 0}{\arg\min} \sum_{j=1}^{M_i^p} (\overline{\tilde{w}}^i - \tilde{w}_j^i)^2, \tag{31}$$

where

$$\overline{\tilde{w}}^i = \frac{1}{M_i^p} \sum_{j=1}^{N_i^p} w_j^i, \tag{32}$$

and $\mathcal{C}^i$ is the set of $\tilde{w}^i$ satisfying the constraints (25-30). Notice that with a quadratic objective function (31) and linear equality constraints in $\tilde{w}_j^i$ (25-30), this becomes a straight-forward quadratic programming optimization problem.

4.2.5 Comments, Remarks, and Discussion

In order to preserve the physics of the system, it is quite important to reproduce the original velocity and position distribution functions within each bin in order to preserve particle distribution functions globally. The methods described above for setting new particle positions and velocities in each bin, i.e. weighted sampling of original particles for down-sampling and uniform sampling of phase space within each bin for up-sampling, is a highly computationally efficient approach. However, extreme spatial gradients in the position and velocity distribution functions cannot be reproduced identically with these methods. For example, if the number of new particles is much less than the number of original particles $M_i^p \ll N_i^p$ (i.e. in the down-sampling case) or the bin sizes are not sufficiently small compared to the global range of $\boldsymbol{\zeta}$ (i.e. in the up-sampling case), then one part of a particle distribution in a bin can have a specific flow and/or distribution that substantially differs from another part of bin – primarily due to unbounded variation at the bin-wise level. In such scenarios, it is possible to discretize the bins into sub-bins and perform resampling algorithm for each sub-bin.

An alternative technique is to use a kernel density estimation (KDE) of the original particle positions and velocities in the reference bin. KDE is a method to estimate a probability distribution

function of a finite data sample. After constructing the KDE of the particles in the phase space within each bin, various classes of importance sampling and Markov chain Monte Carlo (MCMC) algorithms can be employed to set new particles positions and velocities while replicating the original velocity and position distribution functions. However, the computational cost required to perform KDE and MCMC sampling in six-dimensional phase space is high. Considering the usual grid resolutions of PIC codes, such approaches are often over-kill and unnecessary. More often than not, it is more efficient to build sub-binning capabilities into the resampling routine conducted adaptively on bins with spatial gradients above a critical tolerance in both the position and/or velocity distribution functions.

It is also important to note that in some particle evolution models, the dynamics of the system require the weights to evolve. This can lead to sample degeneration in which weights become concentrated on fewer and fewer particles as the simulation proceeds, resulting in a poor distributional representation. In this regard, using the cost function (31) in the optimization step of the resampling strategy results in keeping the global markers weight relatively uniform. Similar to particle filtering methods, maintaining relative uniformity in particle weights improves the representation of the distribution over time, reducing the need for frequent resampling and improving particle diversity, see e.g. [4].

MPCR has three notable features compared to recently developed particle splitting/merging algorithms within the PIC simulation context:

1. The stochastic approach for calculating new particle positions and velocities allows application to spatial decompositions with irregular geometries, including, for example, unstructured meshes. This enables straight forward integration of the resampling method to any available PIC code with various coordinate systems and different implementations. Moreover, since the optimization problems for each bin are independent, the computations at the bin level can easily be solved in parallel (i.e. embarrassingly parallel) or built into the parallel solver of the PIC code.

2. As it will be shown in the numerical results section, the constrained optimization technique is able to conserve any number of particle and grid quantities to *near machine precision* regardless of the method employed in the PIC code for particle evolution and deposition. Some of the previously developed algorithms in the literature [37, 62, 42] are successful in preserving some of the particles quantities but fail in conserving grid values or the distribution function. The particle merging algorithm developed by Luu et al. [40] shows accuracy of preserving phase space evolution and total energy error to only 10^{-3} and 10^{-5} respectively. In addition, the statistical particle split and merge methods recently proposed by Pfeiffer et al. [50], show exact conservation of grid variables (to machine precision, or as reported 10^{-19}) for the cell mean value current and charge density deposition method, i.e. every point of the field solver gains equal current and charge density independent of the positions of the particles. However, the same approach results in much less accuracy in preserving grid quantities to only 10^{-2}) when taking into account a more complex deposition method, i.e. in which the dependency on the distance between the interpolation points and the particle positions is included in deposition.

3. The reliance on constrained optimization leads to an efficient particle resampling algorithm, but introduces a subtle parallel computing inefficiency. The complication that arises relates

to sorting the particles into bins. In a special case the target number of new particles in a bin becomes too small to allow the constraints to be imposed, and the resulting system is over-constrained. In this case, neighboring bins in the velocity space directions can be easily merged to form larger bins with more particles, until it exceeds a specified minimum particles to match the number of constraints. However, solving the optimization problem with bin merging in parallel results in a less computationally efficient algorithm. Even so, it turns out that this is not really reason for concern. Since the frequency of resampling is generally much lower than that of, e.g. pure timestepping, the additional numerical effort devoted to bin-merging are negligible compared to the computational work of the full PIC simulation. It is further worth noting that for a particular problem, global particle distribution and/or computational load balance criteria can be used to adaptively determine the period of resampling during PIC simulation.

4.3 Implementation of the resampling into XGC gyrokinetic PIC code

MPCR can be applied to resample any distribution function, but as an example application we apply this method to a gyrokinetic plasma PIC simulation. In this section, we describe the XGC PIC gyrokinetic codes for solving fusion reactor simulations [10, 9] along with the specific implementation required for integrating MPCR.

The time evolution of plasma systems is described by the six dimensional Maxwell-Boltzmann system in phase space [69]. In strongly magnetized plasmas, such as fusion plasmas, these equations can be averaged over the gyrophase (viz. the cyclotron frequency) leading to a gyrocenter tracking equation for a charged ring over the relatively slow motion of particle gyrocenters, thus treating rapid particle orbits about magnetic field lines perturbatively. Transforming to the gyrocenter coordinate then results in a phase space reduction from six to five dimensions, and the resulting system of equations is referred to as the gyrokinetic equations. While a significant simplification is achieved through the reduction of the full six-dimensional equations to five-dimensional gyrokinetic equations [22], simulating this five-dimensional system over full-scale fusion reactor geometries is still a formidable task that requires carefully formulated numerical approximation within PIC simulation.

The Lagrangian evolution equations for the marker particle positions, velocities and weights depend on exactly how f is sampled. In a straight-forward approach (called full-f), the distribution is simply sampled. However, this requires the largest number of particles to accurately represent f, and is therefore computationally most expensive. Other algorithms use a control-variate (δf) approach in which the difference between f and an ideal or simple distribution f_0 is sampled. There are a number of such δf approaches that differ in the details; see for example [34]. In the full-f representation, the particle weights do not evolve, but in the δf approaches, they do. The details of a resampling algorithm for a full-f and δf sampling scheme may differ, because of the difference in representation. In the algorithm described here, we consider a full-f sampling representation of f.

XGC provides a large collection of PIC solution strategies, allowing for full-f, delta-f, and total-f (or equivalently, hybrid-Lagrangian [34]) simulations, where in each case the XGC PIC algorithms support electrostatic turbulence over a plasma volume parameterized in a toroidal reactor geometry, following the magnetic axis across the magnetic separatrix and scrape-off layer (SOL), to just outside the sheath interfacing material boundary. The XGC PIC model uses a cylindrical

coordinate system, in which the components of the particle position vector are $\mathbf{x} = (r, z, \phi)$, and the velocity vector decomposes into into parallel and perpendicular components of the magnetic field $\mathbf{v} = (v_\parallel, v_\perp)$. The code propagates marker particles using Lagrangian motion and the corresponding electrostatic field is solved on a finite element mesh. Due to the need for field-line following parameterizations along the magnetic separatrix, scrape-off layer region, etc., XGC uses a 2D unstructured triangular mesh [1].

One of the challenges in the simulation of a tokamak plasma is the multiscale nature of the problem, both in space and time. For example, a simulation that resolves both plasma turbulence and axisymmetric neoclassical dynamics, tracking slowly evolving large-scale phenomena is extremely expensive due to the shorter time-scale small-scale phenomena that can drive them. One approach that attempts to mitigate this computational expense, takes a spatially 3D particle distribution function distorted from Maxwellian and down-samples it to a spatially 2D axisymmetric particle distribution to evaluate the neoclassical transport arising from the turbulence distortion. In this low-fidelity representation, the lower dimensionality and reduced number of particles required to represent f significantly reduces the computational cost of the simulation.

Integrating MPCR into the XGC PIC code begins by identifying the spatial discretization used by XGC (i.e. unstructured triangular meshes). In XGC the element vertices on each plane are ordered on a space filling curve starting on the magnetic axis. During parallel processing computation, the curve is divided into number of patches and each computing node manages a finite numbers of vertices depending on how many particles are assigned to each vertex. Particles are then assigned to vertices by projecting their position along the magnetic field to the midplane of the toroidal section. On the midplane, a nearest neighbor weighting (i.e. the vertex with the largest linear finite element interpolation weight) determines to which vertex a particle belongs. This results in 3D spatial discretization of the domain, or *Voronoi cell*, that is the volume in the vicinity of a field line that projects to the same vertex. Each Voronoi cell associated with a site (vertex) k contains a set of particles, such that the distance from them to k is not greater than the distance to any other vertices [15]. Each Voronoi cell in XGC is assigned to a particular MPI process, and these cells are characterized as spatial bins along $\mathbf{x} = (r, z, \phi)$ in order to avoid additional communications between processors. In contrast, in velocity space $\mathbf{v} = (v_\parallel, v_\perp)$ uniform binning is used over each Voronoi cell.

The particle preservation can be directly implemented in each bin for (25)-(28). The constraining grid quantities (29)-(30), on the other hand, utilize the spatial discretization, where the patch of vertex i is the smallest 2D triangular mesh (Figure 16) in the ϕ-midplane containing $\mathrm{Vor}(i)|_{\phi_J}$. The corresponding constraints then become:

$$\sum_{p_k \in \mathrm{Vor}(i)} \beta_k \lambda_j w_k = \sum_{p_k \in \mathrm{Vor}(i)} \tilde{\beta}_k \tilde{\lambda_j} \tilde{w_k} \quad \forall v_j \in \text{Patch}(i), \tag{33}$$

and

$$\sum_{p_k \in \mathrm{Vor}(i)} (1-\beta_k) \lambda_j w_k = \sum_{p_k \in \mathrm{Vor}(i)} 1 - \tilde{\beta}_k \tilde{\lambda_j} \tilde{w_k} \quad \forall v_j' \in \text{Patch}(i'), \tag{34}$$

as well as,

$$\sum_{p_k \in \mathrm{Vor}(i)} \beta_k \lambda_j \mathbf{v}_k w_k = \sum_{p_k \in \mathrm{Vor}(i)} \tilde{\beta}_k \tilde{\lambda_j} \tilde{\mathbf{v}}_k \tilde{w_k} \quad \forall v_j \in \text{Patch}(i), \tag{35}$$

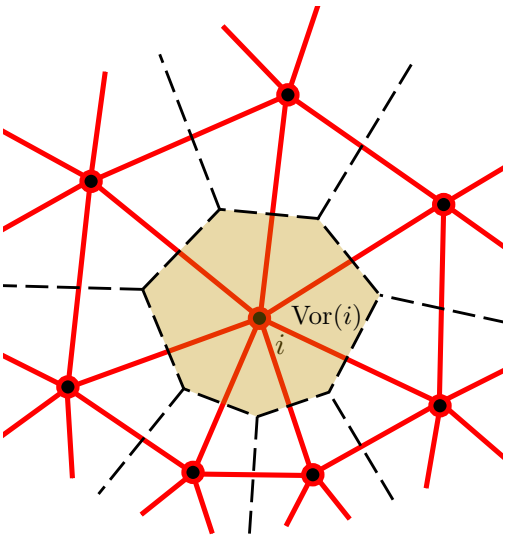


Figure 16: Illustration of a 2D unstructured mesh and Voronoi cell associated with the vertex (node) i.

and

$$\sum_{p_k \in \mathrm{Vor}(i)} (1-\beta_k)\lambda_j \mathbf{v}_k w_k = \sum_{p_k \in \mathrm{Vor}(i)} (1-\tilde{\beta}_k)\tilde{\mathbf{v}}_k \tilde{\lambda}_j \tilde{w}_k \quad \forall v'_j \in \mathrm{Patch}(i'), \tag{36}$$

where β_k is the coordinate with respect to the field line between the ϕ-midplanes, and λ_j is the barycentric coordinate with respect to vertex j in the triangle in Patch(i), and $\tilde{(\cdot)}$ represents the resampled particle quantities.

4.4 Particle Resampling Results

In this section we demonstrate and analyze numerical examples from coupling MPCR through the XGC full-f Fortran code via constrained quadratic programming (31), where the algorithm of Goldfarb & Idnani [17] is implemented using the *quadprog* Fortran library [5].

In the first set of test problems, we demonstrate the quality of MPCR in conserving features of the distribution function using global down-sampling of particles from an XGC plasma turbulence solution. Finally we show the benefits of periodic particle resampling using MPCR on a neo-classical plasma fusion simulation.

4.4.1 Global particle downsampling

In current XGC implementations, the particle down-sampling process for multi-fidelity simulations is performed by randomly choosing particles over the entire domain of f, without preserving any features, or moments, of the system. Subsequently, one might anticipate that MPCR would provide a significant improvement in the accuracy of the kinetics representation. Using this as motivation, we provide two examples where the global number of particles are reduced/increased and the ability of the resampling algorithm to preserve features of the distribution is assessed.

For the first example, consider the one-dimensional probability distribution shown in Figure 17. It might, for example, be considered to be the distorted marginal (reduced dimensional) distribution

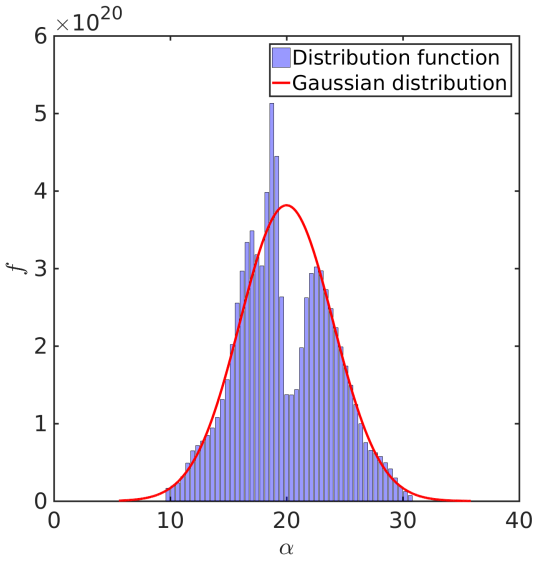


Figure 17: Initial particle distribution function with local features. In this case, total number of particles N^p is 50000. The solid red curve is a Gaussian.

of one of the velocity components, in which case Maxwell-Boltzmann statistics would yield a Gaussian distribution, and the given distribution is clearly far from Gaussian. This distribution is represented with $N^p = 50000$ samples, and down-sampled to $M^p = 1000$ and $M^p = 250$: down-sampling factors (DSF) of 50 and 200, respectively.

Down-sampling was performed using MPCR in one-dimension, given $M^b = 50$ bins in which the sample density, and first and second moments are preserved. For comparison, global random down-sampling was also used. The results are shown in Figures 18 and 19. By construction the histogram formed on the 50 down-sampling bins is identical to that of the original sample (compare Figures 18a and c with Figure 17), but the randomly down-sampled histograms show large variations from the original. Indeed, the sampling noise obscures the structure of the distribution; so much so that in the DSF=200 case, the underlying structure of the distribution is not visible at all. By construction, the low- order moments of the distribution are preserved in the proposed approach, as shown in Figure 19, whereas relative errors in the moments with random down-sampling range from 2% to 10%.

MPCR up-sampling of the distribution function from Figure 17 was also performed and compared with the simple particle reproduction method widely employed in PIC codes. Such particle reproduction consists of duplicating original particles several times until the up-sampling factor is satisfied. Both methods are capable of preserving detailed features of the distribution function, and as it is shown in Figure 20, the relative error associated with moments in both approaches are within range of machine epsilon accumulation. However, simple particle reproduction methods can result in unrealistic/unphysical PIC simulations, since duplicate particles with the exact same phase space representation have the exact same physical responses in the PIC dynamics. One common way to avoid this unrealistic particle duplication is to tweak the new particles with small (e.g. phase space) variations. However, again, such artificial particle manipulation introduces large error and results in much higher relative errors than those shown in Figure 20.

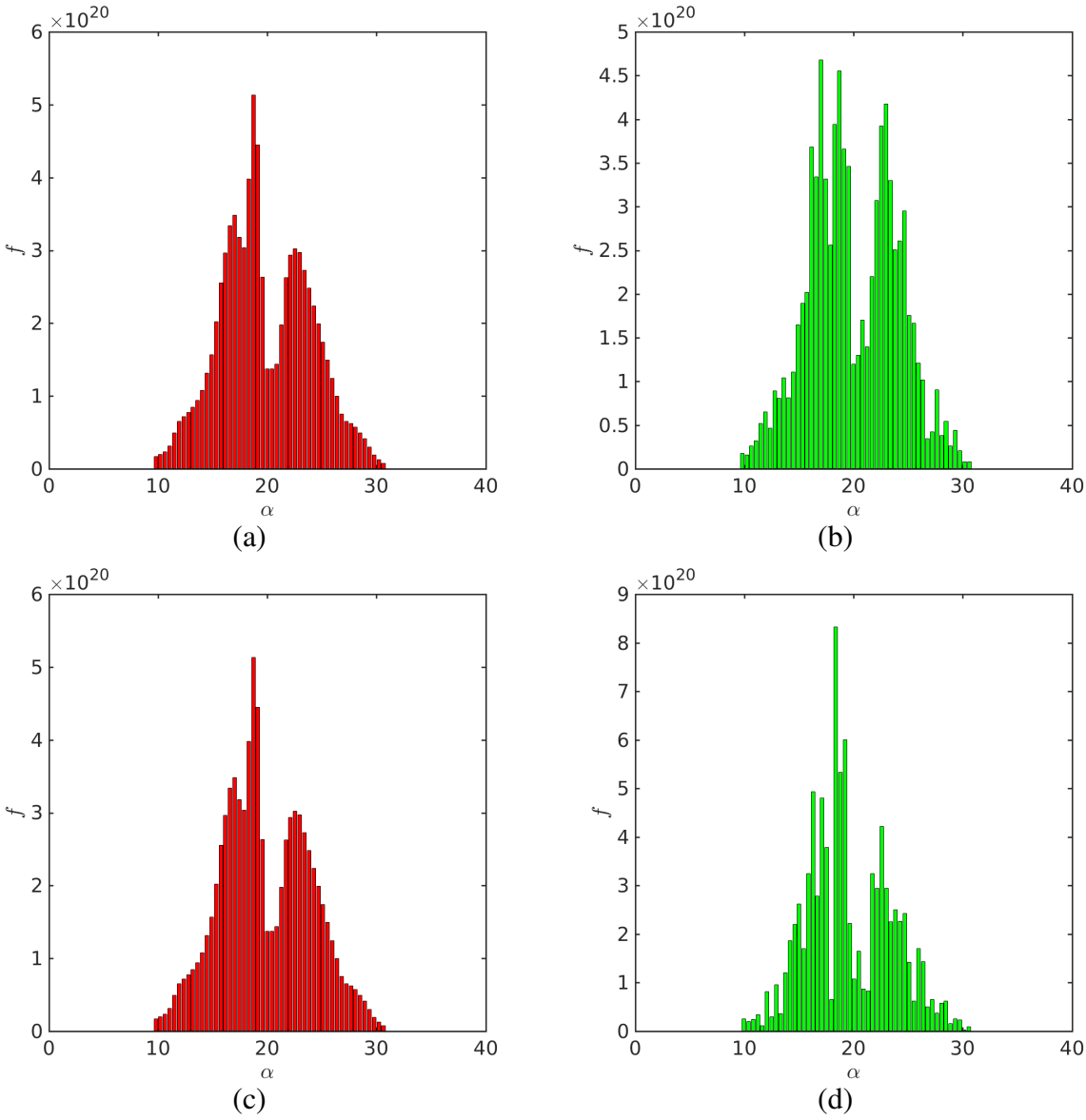


Figure 18: Local features preservation after particle down-sampling of the distribution in Figure 17 (number of bins = 50): (a) feature-based resampling method (DSF = 50); (b) random down-sampling (DSF = 50); (c) feature-based resampling method (DSF= 200); (d) random down-sampling (DSF = 200).

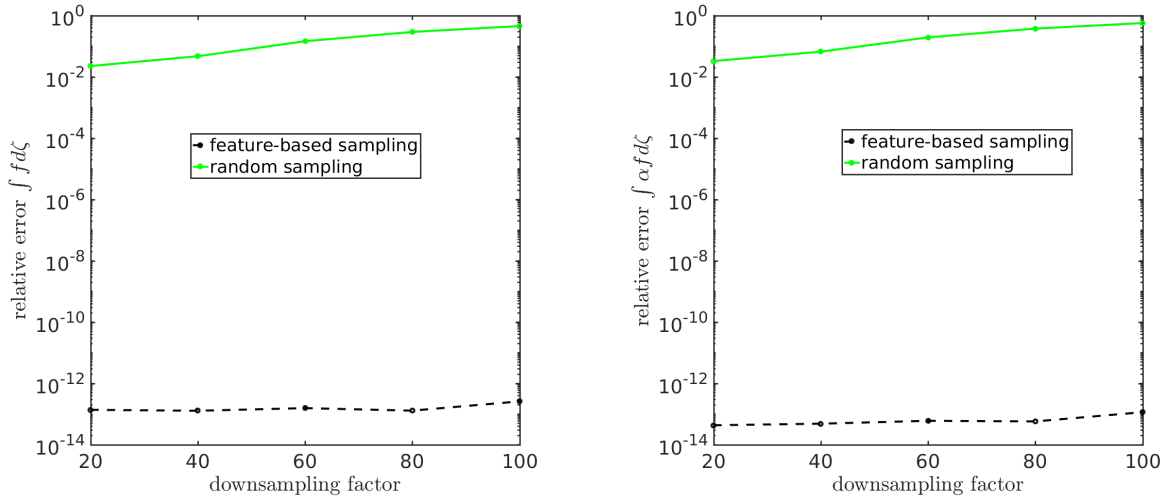


Figure 19: Relative error in global 0th and 1st moments due to feature-based/random down-sampling for different down-sampling factors (error in random down-sampling is the mean of 10^4 realizations).

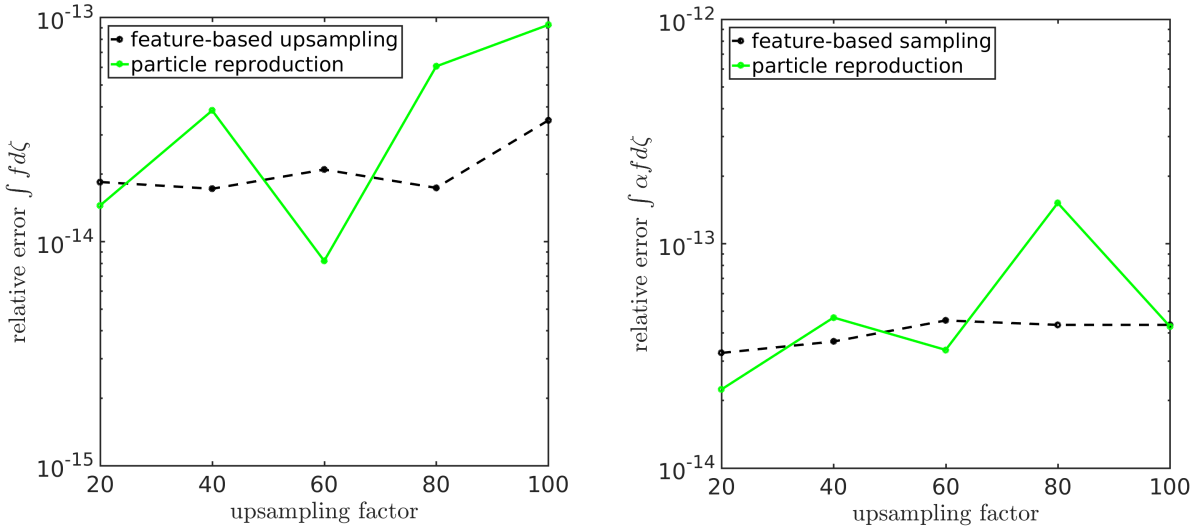


Figure 20: Relative error in global 0th and 1st moments due to feature-based up-sampling and particle reproduction for different up-sampling factors.

The second example application considered in the section examines the resampling algorithm on a particle distribution taken from a tokamak plasma simulation performed with the gyrokinetic PIC code XGC1. In a spatially three-dimensional gyrokinetic PIC code, each particle carries with it six attributes, its weight w, spatial coordinates $\mathbf{x} = (r; z; \phi)$, and velocity coordinate $\mathbf{v} = (v_\parallel; v_\perp)$. In both up-sampling and down-sampling considered in this example, the distribution f is projected on a spatially two-dimensional configuration space, which assumes the solution is statistically homogeneous in the ϕ direction. This illustrates the particle resampling required to perform spatially two dimensional neoclasical simulations that utilizes a spatially three-dimensional simulation of the turbulence as input. The data set consists of $N^p = 10^6$ particles, initially weighted to effect importance sampling. The particles have since been mixed through their evolution, so their weights no longer accomplish the desired importance sampling.

The first step of the particle resampling process is to divide the four-dimensional phase space $\boldsymbol{\zeta}$ into bins and sort the particles into those bins. In the gyrocentered XGC1 code, the particle velocity space is parameterized in terms of the normalized parallel velocity $\rho = v_\parallel/B$ and magnetic moment $\mu = v_\perp^2/2B$, where B is the magnitude of the magnetic field. For this case then, the four-dimensional down-sampled phase space is $\boldsymbol{\zeta} = (r, z, \rho, \mu)$. The $\boldsymbol{\zeta}$ domain is divided into bins, with 50 bins in each of the four phase-space directions.

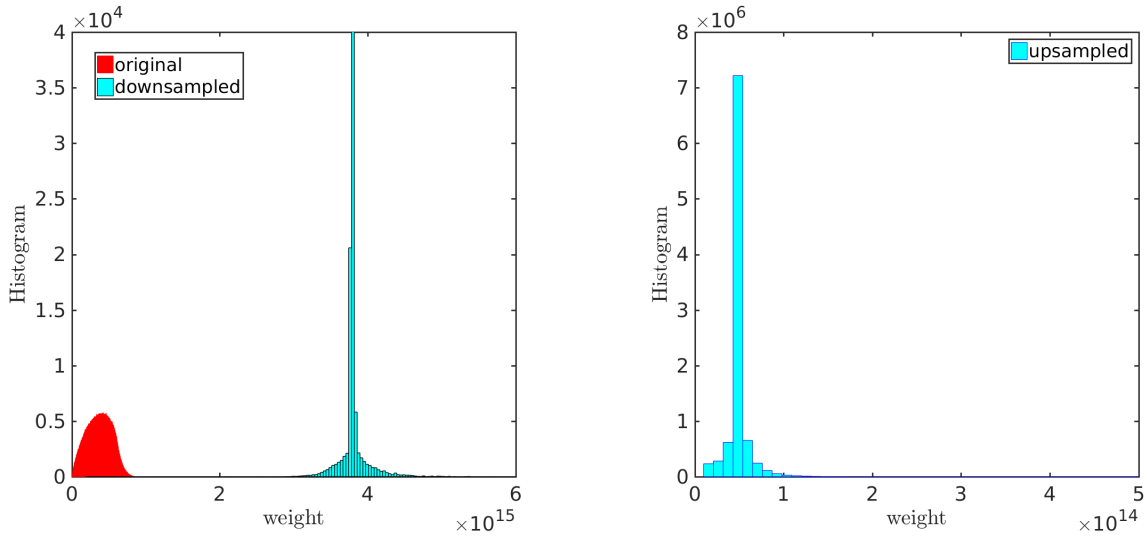
To demonstrate the adjustment of weights as part of the resampling process the importance weighting is eliminated by making $g \propto f$, where g is the marker particle distribution function in (24). The target number of particles M_i^p in each bin is then simply proportional to the sum of the weights of the original particles in the bin. Specifically,

$$M_i^p = M^p \left(\frac{\sum_{j=1}^{N_i^p} w_j^i}{\sum_{j=1}^{N^p} w_j} \right). \tag{37}$$

One complication that arises in sorting the particles into bins is that in some cases the target number of down-sampled particles M_i^p in a bin becomes too small to allow the constraints to be imposed. In this case, neighboring bins in the velocity space directions (ρ, μ) are merged to form larger bins with more particles, until M_i^p exceeds a specified minimum, which was set here to $M_i^p \geq 25$.

The distribution of weights is shown in Figure 21 for the original, down-, and up-sampled particles. Note that the original distribution is broad, while the resampled distributions are significantly more peaked; a result of the improved uniformity in the weights after resampling. Also note that the down-sampled particle weights are several times larger in magnitude than the original weights, while the up-sampled weights are about an order of magnitiude smaller — a behavior that is the result of the respective resampling factors. It is important to further observe here that the marginal distributions of the original and down-sampled particle and velocity variables, shown in Figure 22, are, as expected, effectively indistinguishable.

The accuracy of the moment preservation in the resampling approach is tested by down-sampling while testing combinations of moments preservation. That is, while preserving 1) only the zeroth moment, 2) the zeroth and first moments, 3) and the zeroth, first and second moments as described in Section 4.1. This was done for a range of down-sampling factors, and the results are shown in Figures 23 and 24, along with those for a random down-sampling. The moment error for the random down-sampling are averaged over 10^4 sampling realizations. In all cases shown in Figures 23 and 24 the relative error of random down-sampling increases as a function



(a) (b)

Figure 21: Histograms of particle weights from the original weights and the proposed resampling algorithm. In this case $N^p = 10^6$ and the number of bins in the ρ, μ, r and z directions is 50: (a) original particles and downsampled particles with down-sampling factor (DSF) = 20, (b) upsampled particles with up-sampling factor (USF) = 10.

of the down-sampling factor. In all cases, when the moments are constrained, the relative errors in the down-sampled moments are of order 10^{-10}. As expected, when only some moments are constrained, the errors in those that are not are significantly larger than in the random case, etc., showing that the MPCR algorithm is effectively preserving moments.

Similarly we analyze the upsampling case over a range of bin numbers. The results are shown in Figure 25. Here the figure indicates that, when the moments are constrained the relative errors in the up-sampled moments remain below 10^{-10}. When only the 0th moment is constrained, the errors in the first moment decreases with increasing resolution in the up-sampling algorithm, i.e. using more, and thus smaller, bins. This dependency to bin size is the result of uniform sampling of phase space within each bins and can be greatly improved by the strategies discusses in Section 4.2.

4.4.2 Periodic particle resampling of a neo-classical PIC simulation

XGCa [20, 35, 33] is a global gyrokinetic particle-in-cell (PIC) code with axisymmetric electrostatic potential solver specialized in the simulation of neoclassical transport physics in the edge plasma of toroidal magnetic confinement devices. These simulations are capable of evolving the full five-dimensional gyrocenter distribution function from the magnetic axis to the inner wall of the device by using either conventional full-f (constant particle weight) [33] or semi-Lagrangian total-f (variable particle weights) [35] methods. A logical sheath boundary condition [47] is used to prevent net currents to the material wall. Plasma lost to the material wall is replenished self-consistently by a neutral particle recycling module according to the local loss rates. Collisional physics are evaluated either by a fully nonlinear, Eulerian Fokker-Planck-Landau collision operator [71, 21] (total-f) or a linearized Monte-Carlo collision operator [66] (full-f).

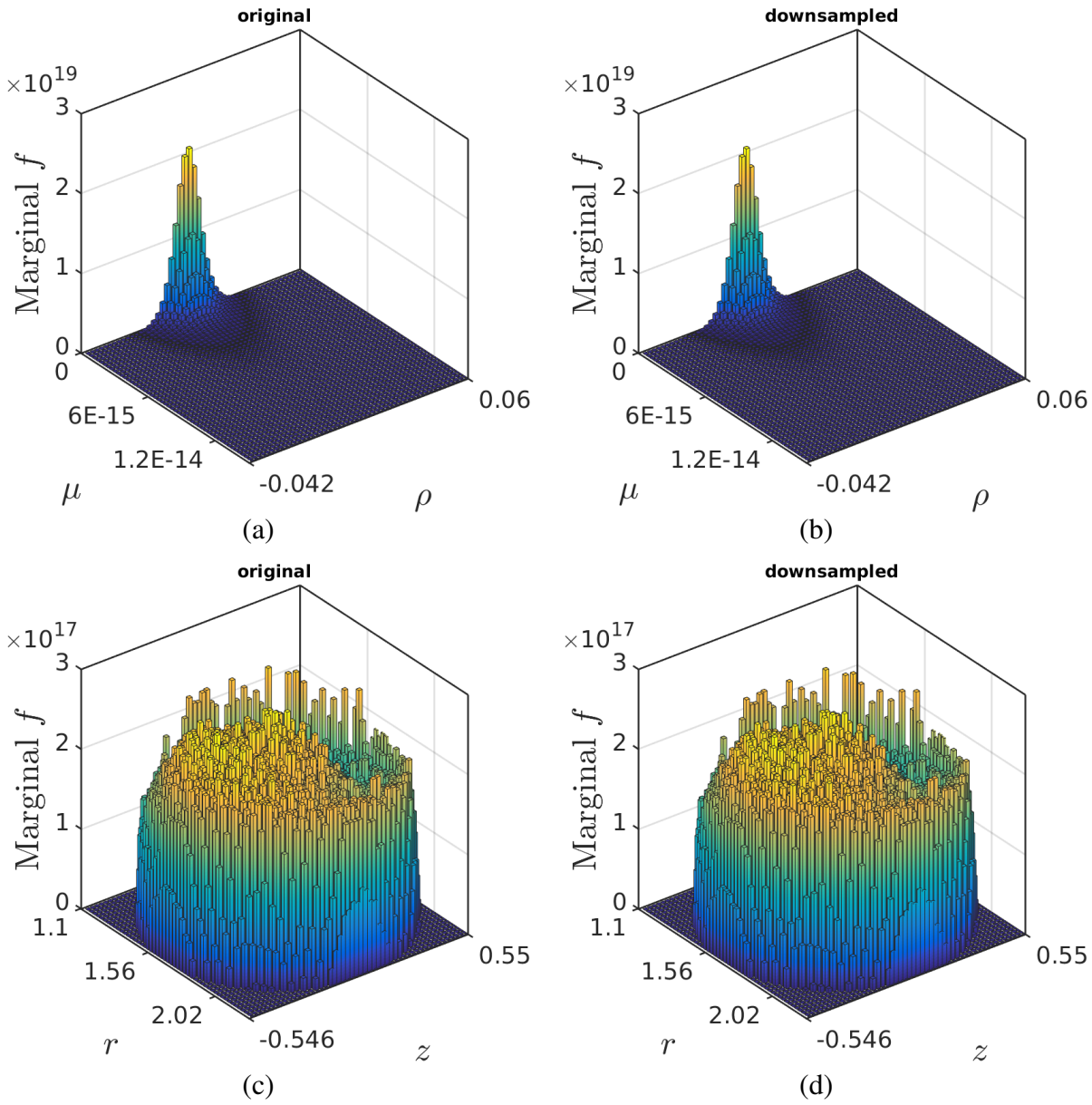


Figure 22: Marginal 2D bar plots of distribution functions before and after particle down-sampling (a and b) along ρ and μ; (c and d) along r and z. In this case, $N^p = 10^6$, DSF = 20, and the number of bins in the ρ, μ, r and z directions is 50.

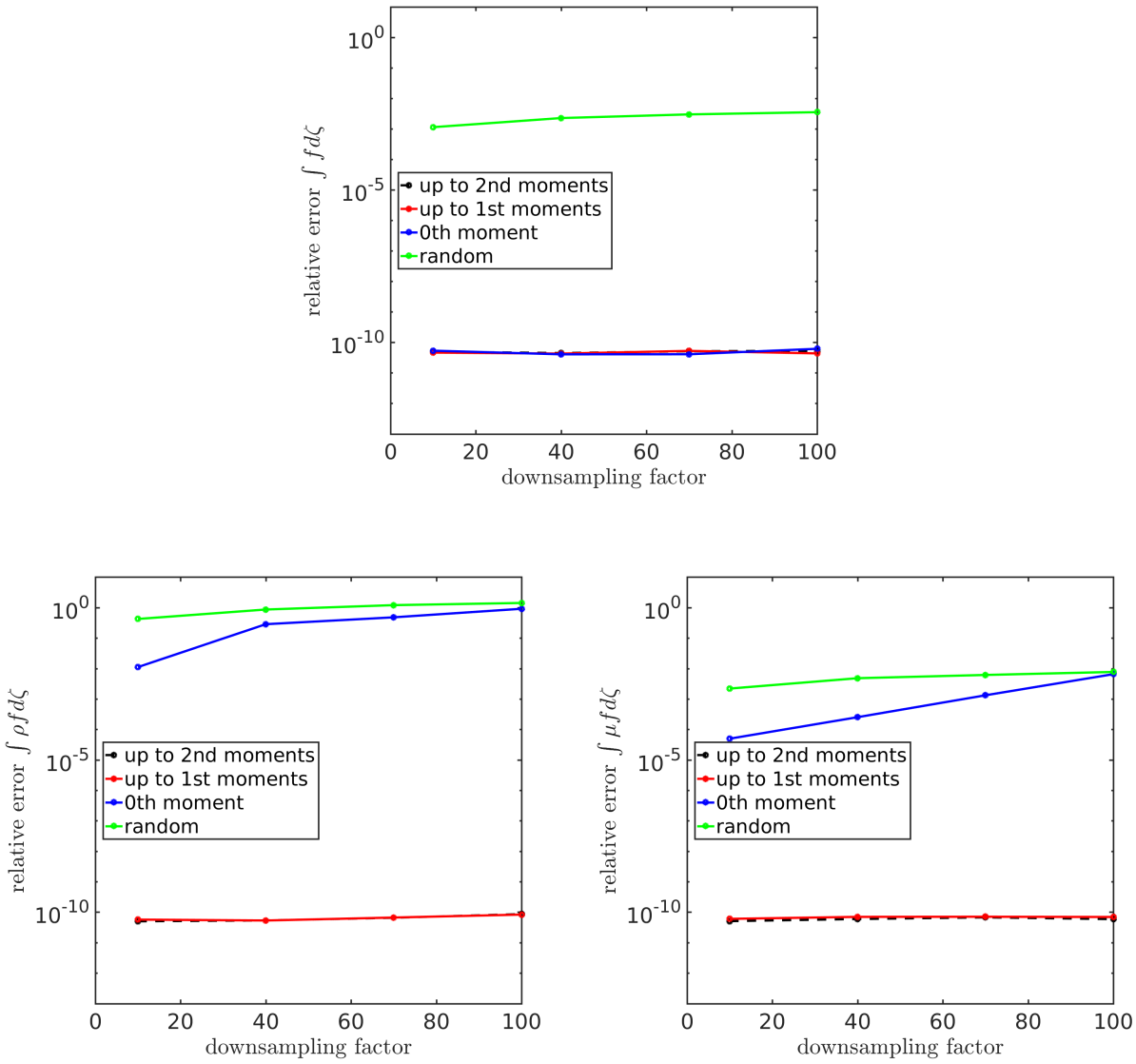


Figure 23: Relative error in global 0th and first moments due to feature-based/random down-sampling for different down-sampling factors. In this case, $N^p = 10^6$ and the number of bins in the ρ, μ, r and z directions is 10.

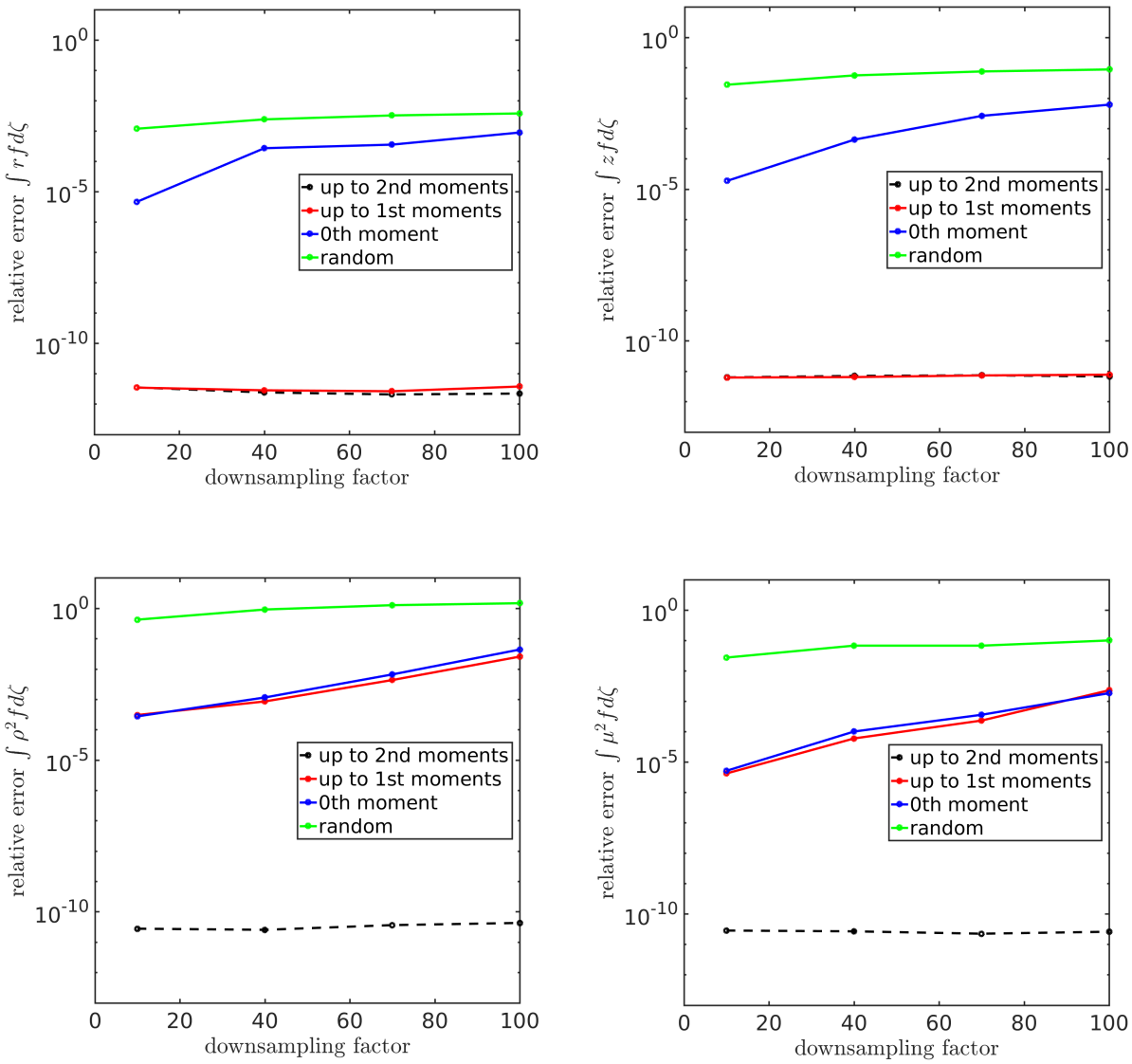


Figure 24: Relative error in global first and second moments due to feature-based/random down-sampling for different down-sampling factors. In this case, $N^p = 10^6$ and the number of bins in the ρ, μ, r and z directions is 10.

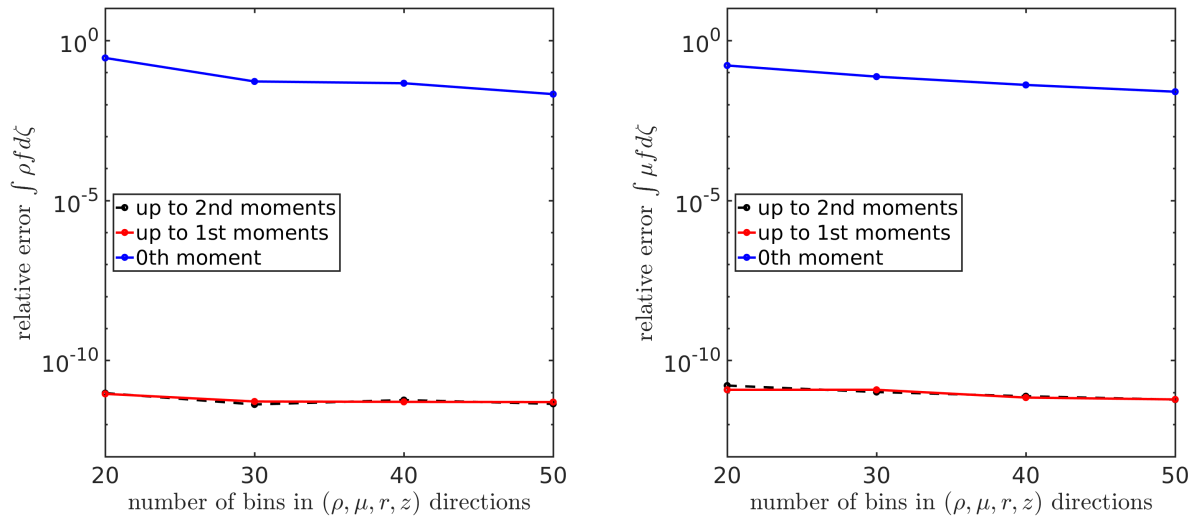


Figure 25: Relative error in global 0th and first moments due to particle up-sampling for different number of bins. In this case, $N^p = 10^6$, up-sampling factor=10, and the number of bins are equal in the ρ, μ, r and z directions.

As a proof-of-principle test of the MPCR method, we investigate the most basic problem a full-f gyrokinetic code has to solve — the formation of the background radial electric field that is needed to maintain quasi-neutrality, the divergence-free equilibrium plasma flows, and to conserve toroidal angular momentum. We use the conventional full-f method with a single ion species (deuterium) and the adiabatic electron model.

The magnetic equilibrium field is that of a generic, up-down symmetric, low-aspect ratio tokamak with a circular boundary surface and Shafranov shift. Since no material wall is included in the simulation, particles that leave the simulation region at the outer boundary are re-inserted into the simulation on the same flux-surface but at the poloidal angle, at which they would have re-entered the simulation domain had they been allowed to continue their orbit. Marker particles are loaded with a uniform distribution in configuration and velocity space and perpendicular and parallel velocities of up to approximately $3.5v_{th}$, where v_{th} is the thermal velocity. The marker particle weights w are determined such that the initial plasma distribution function is a local Maxwellian with density $n_i(\psi)$ and temperature $T_i(\psi)$ that depend only on the flux-label (generalized minor radial coordinate) ψ, i.e. each particle's initial weight is a function of its initial flux-label and velocity. The electrostatic potential is zero initially. A local Maxwellian is, however, not the neoclassical equilibrium distribution function in the presence of a pressure gradient due to the magnetic inhomogeneity drift, which causes charged particles in tokamaks to move on so-called banana-orbits with finite width $\Delta\psi_b$. Starting with a local Maxwellian distribution and vanishing radial electric field, the orbit motion of the ions together with the background pressure gradient leads to the growth of an up-down anti-symmetric pressure perturbation and a net toroidal flow, which violates the divergence-free condition of the equilibrium flow and the conservation of toroidal angular momentum. To make the equilibrium flow divergence free and cancel the net toroidal flow from the magnetic inhomogeneity drift, the plasma reacts by generating a radial electric field (guiding

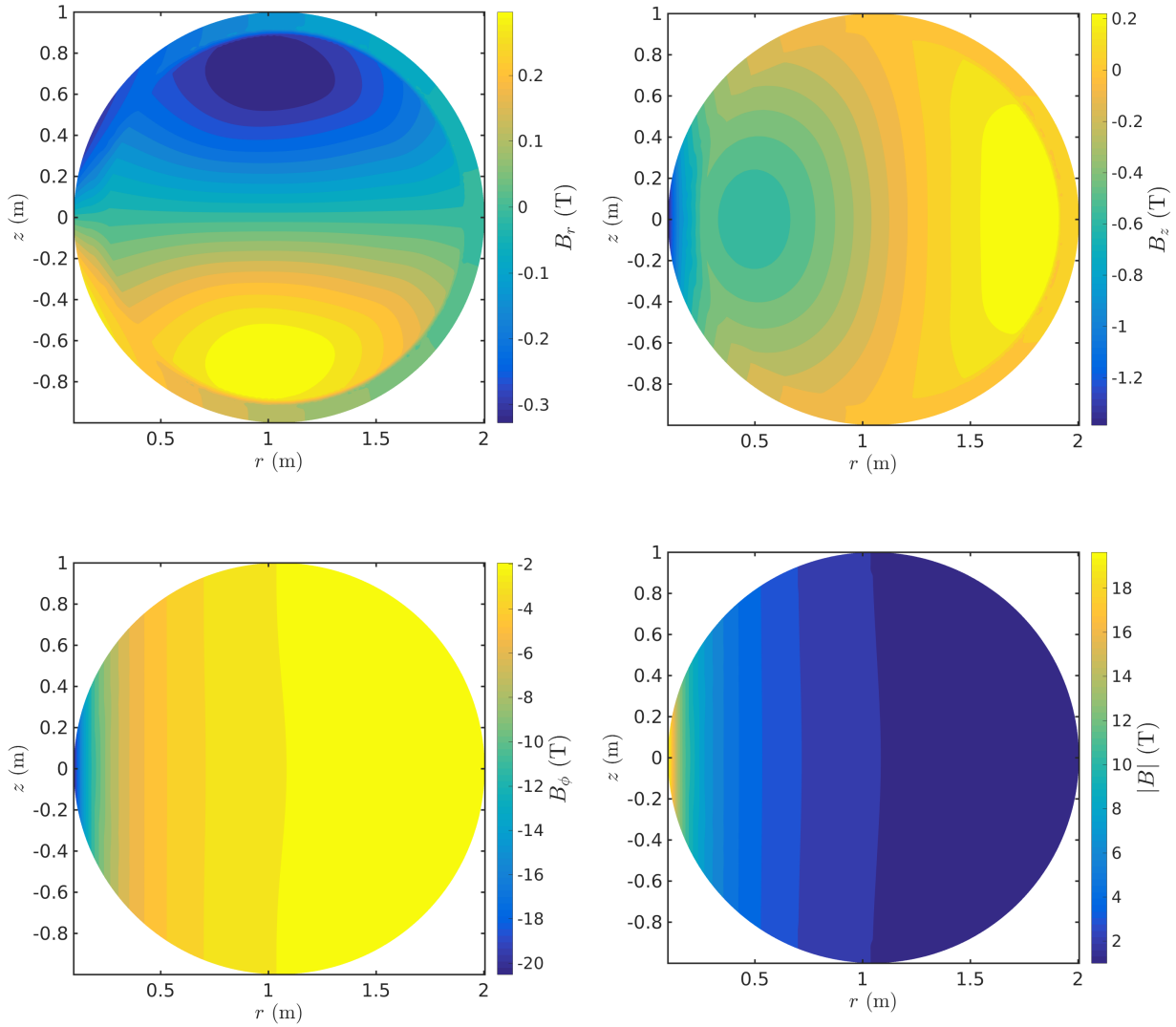


Figure 26: Spatial distribution of magnetic field components. The computed values at vertices are interpolated in these plots.

center polarization) with its corresponding $\mathbf{E} \times \mathbf{B}$ flow.

The challenging problem in the calculation of the radial electric field with a full-f PIC code, is that the orbit motion of the markers leads to mixing of particles with disparate weights. This is a considerable source of sampling noise, especially in regions with small pressure gradient and low amplitude of the electrostatic potential. The purpose of periodic particle-resampling is to reduce the sampling noise by homogenizing the particle weights in appropriately defined bins and by optimizing the phase space coverage of the particle population. We demonstrate this capability by comparing the time evolution of the radial electric field and the radial electric field profile in quasi-steady state between simulations with a total of 5 million, 50 million, and 500 million marker particles without particle re-sampling, and a simulation with 50 million particles with periodic resampling.

In all cases, the PIC simulations are run over 6000 time steps of 0.002 of the toroidal transit time. The spatio-temporal evolution of the radial electric field of the simulations without periodic particle resampling are shown in Figure 27. In all three cases, we observe increased noise in the self-consistent radial electric fields in the edge region, $0.8 < \psi < 1$. This is due to the low marker particle density in these regions.

Figures 27 and 28 show the evolution of the marker density and particle weight distribution at initial loading and after the 6000 time steps. These figures indicate that mixing low-weight particles with higher weight particles during the simulation leads to spreading of the marker particle weights and lowering the simulation efficiency in the critical edge regions. Such phenomena degrades the quality of the Monte Carlo sampling over time since particles evolve away from the desired importance-sampled distribution and results in large simulation error.

To investigate the effect of resampling in reducing simulation error while saving compute time, the 50 million particle case is subjected to periodic particle resampling. Of course, to fully optimize the efficiency of the algorithm, the resampling period must ultimately be determined adaptively using the global particle distribution along with the frequency of the dynamic load balancing. However, to illustrate the feasibility of the resampling technique in the current numerical experiment, three stages of resampling are examined, i.e. after 1500, 3000, and 4500 time steps. In the resampling algorithm, the 2D unstructured mesh in XGC is used for the spatial bins along $\mathbf{x} = (r, z)$ while velocity space $v = (v_\parallel, v_\perp)$ is uniformly discretized. At each resampling stage, the total number of resampled particles are enforced to be the same as the number of the original particles i.e. $M^p = N^p$, while the target number of resampled particles in each bin M_i^p is determined from the initially loaded particles at the beginning of the simulation.

In Figures 30 and 31 we show the evolution of the neo-classical radial electric field, marker density, and weights for the 50 million particle case with three stages of particle resampling. The noise reduction effect of particle resampling from MPCR, immediately after each stage, is due to attenuation in spreading of the particle weight and restoration along regions with low amplitude of the electrostatic potential. These favorable effects gradually disperse as the simulation continues, owing to loss in phase space resolution and simulation efficiency in the critical regions. This indicates that these features could be easily recovered with moderate increase in the frequency of resampling.

Figure 32 compares the radial electric field in the 50 million particle case with and without resampling. The corresponding high-fidelity simulation here is run using 500 million particles. Remarkably, MPCR does extremely well capturing the averaged radial electric field profile over 4000-6000 time steps as well as the time evolution of E_r in the edge and core regions. It is notable

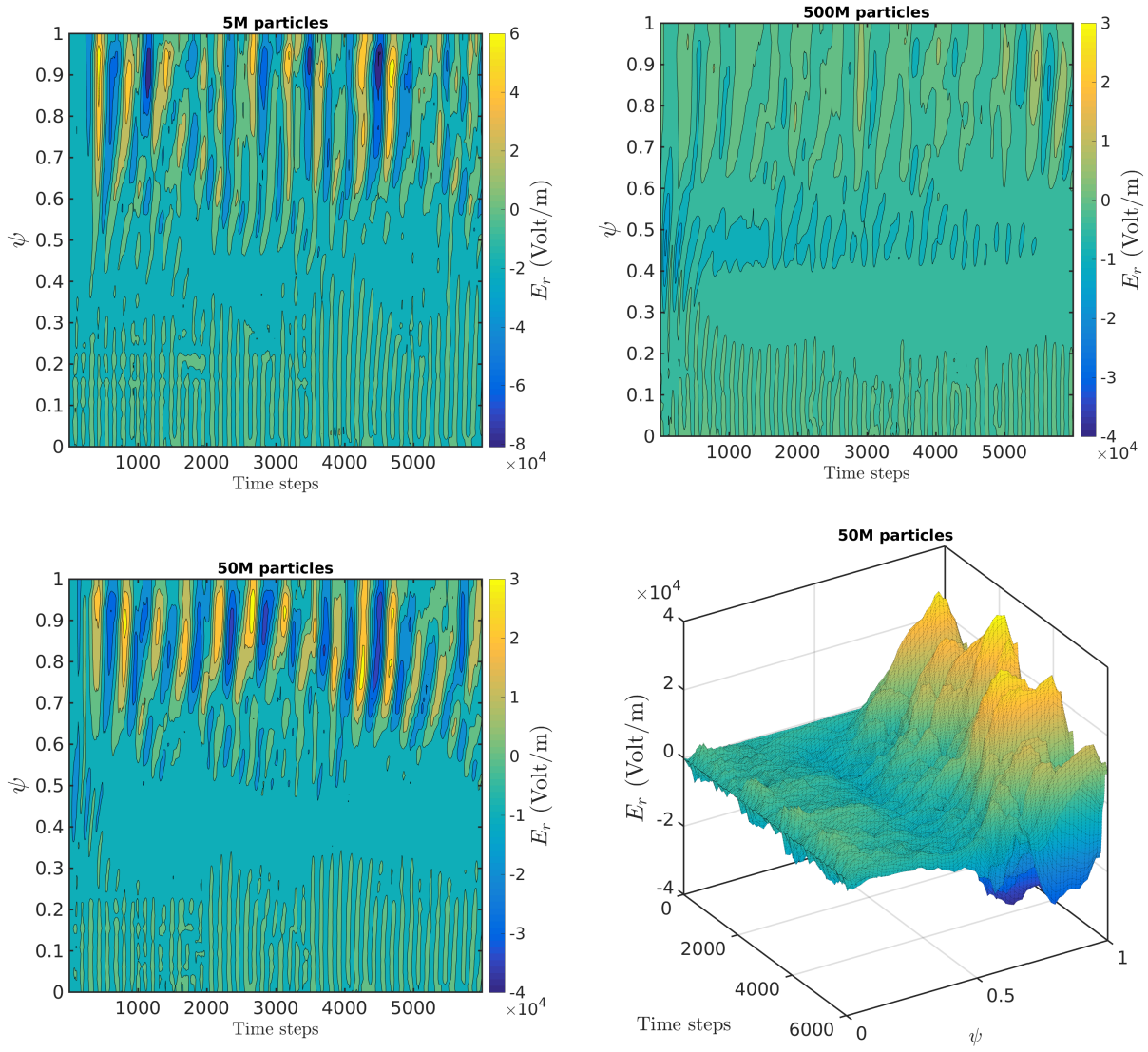


Figure 27: Spatio-temporal evolution of E_r without resampling for (a) 5 million particles, (b) 500 million particles, and (c,d) 50 million particles.

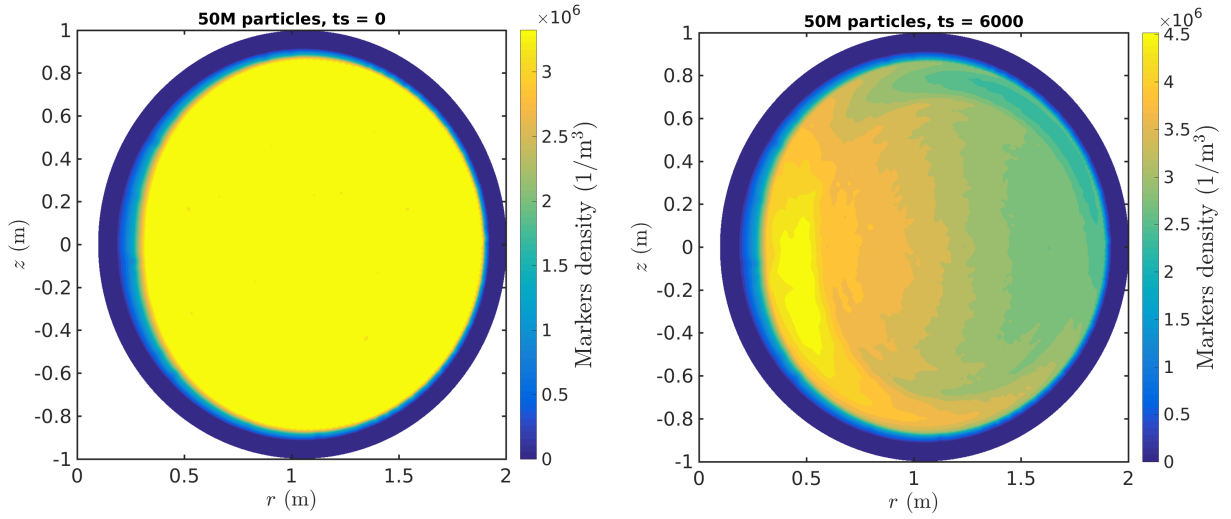


Figure 28: Spatial distribution of particle density for the simulation with 50 million particles and without resampling: (a) initial loaded particles, (b) after 6000 time steps. The computed values at vertices are interpolated in these plots.

that even with the fairly low resampling frequency of 1500 timesteps, the resampling significantly reduces the simulation error and improves the accuracy of the results obtained from the reduced order model.

5 Conclusions

The application of uncertainty quantification to high fidelity simulations of tokamak plasmas is particularly challenging because of the great computational costs of a single simulation. Also challenging is the fact that there are many uncertain inputs to the simulations, that have poorly characterized uncertainties. This latter, however, just increases the importance of UQ in the validation of the simulation models, which is the most immediate application of UQ techniques in tokamak simulations. A multi-fidelity sampling approach has been identified as the most promising to minimizing the number of high fidelity simulations that need to be run as part of such a UQ study. In this approach lower fidelity and lower cost surrogates are needed, and in this work two such have been investigated: a polynomial response surface representation (section 2) and a scenario extrapolation approach (section 3). In addition, to reduce costs of an XGC forward simulation, UQ techniques were used to optimize marker particle distributions in a particle resampling algorithm (section 4). This algorithm is currently being integrated into the XGC code base.

To explore the utility of response surface surrogates, we developed a prototype application to collisionless plasmas, and generated global surrogates from a minimal set of calibration data. These preliminary results demonstrate moderate variation in the critical gradient as a function of R_0 and B_0, and in a cross validation, the global surrogate agreed well with a nominal base case chosen within the parameter space. We see improvement of the response surface accuracy as a function of g, the order of the polynomial response surface, indicating both a well-behaved surrogate and the potential for increased accuracy with higher g. It remains unclear however, if this surrogate can be extrapolated to outside the limits of the parameter scan. In contrast to the well-

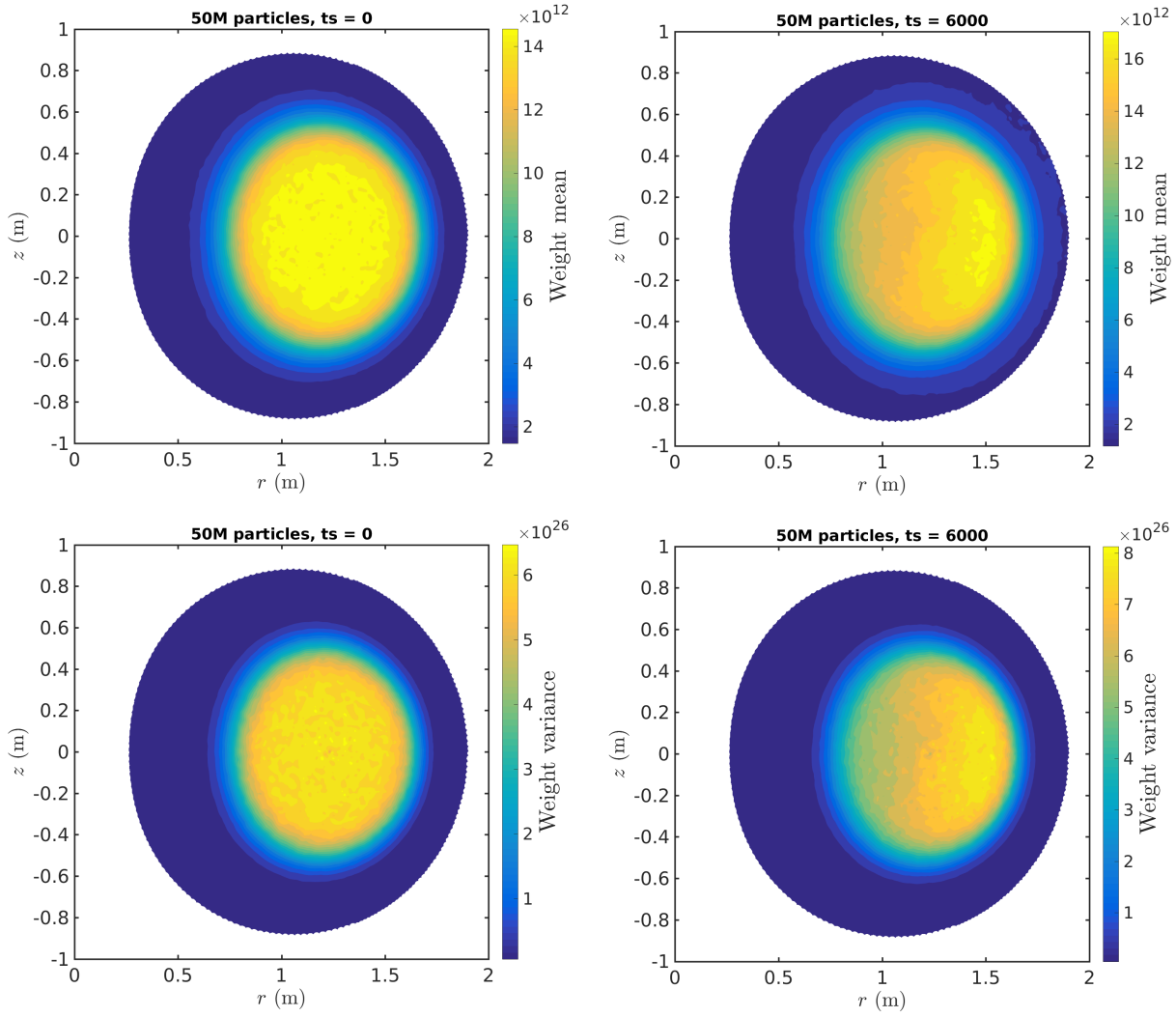


Figure 29: Spatial distribution of mean and variance of particle weights for the simulation with 50 million particles and without resampling: (a,c) initial loaded particles, (b,d) after 6000 time steps. The computed values at vertices are interpolated in these plots.

behaved response surface in the ion temperature gradient, the linear departure rates were not as well behaved. Additional calibration data would be required to hope to arrive at a usable surrogate. A paper reporting on the findings reported here on global surrogates for collisionless plasmas is in the final stages of preparation.

To address the massive computational requirements for predictive simulation and uncertainty quantification of fusion plasmas, we have developed a hybrid method for constructing surrogate models for predictive extrapolation that we have named an augmented surrogate. By combining (potentially uncertain) scenario parameters with the dimensionless plasma scaling parameter ρ^*, we are able to construct reasonable response surface approximations, yielding extrapolative predictions as well as parameter space gradients at a reduced cost. Additional testing of our predictive extrapolation methodology is part of an ongoing power scaling study based on a set of numerical experiments outlined in [25].

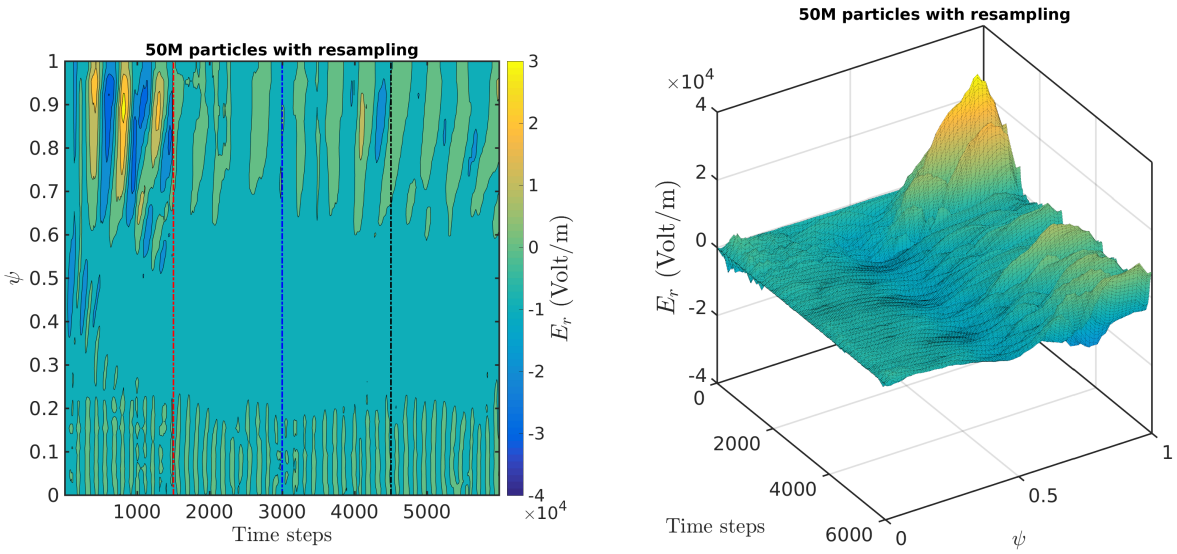


Figure 30: Spatio-temporal evolution of E_r for the 50 million particle case with three stages of resampling: after 1500, 3000, and 4500 time steps.

The MPCR algorithm presented here (section 4) is designed to reduce the sampling noise in PIC codes while preserving important features of the distribution being represented by the particles. The algorithm itself functions by first partitioning the distribution phase space into smaller subdomains, wherein binning can be performed. The general method uses sampling techniques for producing new particle positions and velocities, and employs constrained optimization techniques to adjust particle weights to accurately preserve essential characteristics of the distribution.

In addition, MPCR enables the preservation of any number of derived particle and grid quantities (e.g. energy, mass, momentum, current density, charge density), with the only limitation being the bin volume, and hence number of particles per bin. By performing this particle resampling in bins and constraining the low-order moments computed in each bin, the algorithm is able to preserves essential features of the particle distribution that are important to the physical consistency and dynamics of the system. The binning also enables the introduction and/or adjustment of importance sampling.

The quality and utility of the MPCR method was demonstrated through several examples, including application to a neoclassical plasma simulation. The tests demonstrate the ability of MPCR to significantly reduce computational cost and increase solution accuracy, while the resampling does not introduce any change in the Lagrangian and Eulerian PIC solution through local conservation of particle and grid quantities. A paper reporting on the MPCR algorithm and its utility in plasma PIC simulations is in the final stages of preparation.

Acknowledgments

The work reported here was supported by the Department of Energy under grant DE-SC0008454. We also acknowledge the Texas Advanced Computing Center (TACC, http://www.tacc.utexas.edu) at The University of Texas at Austin and the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Depart-

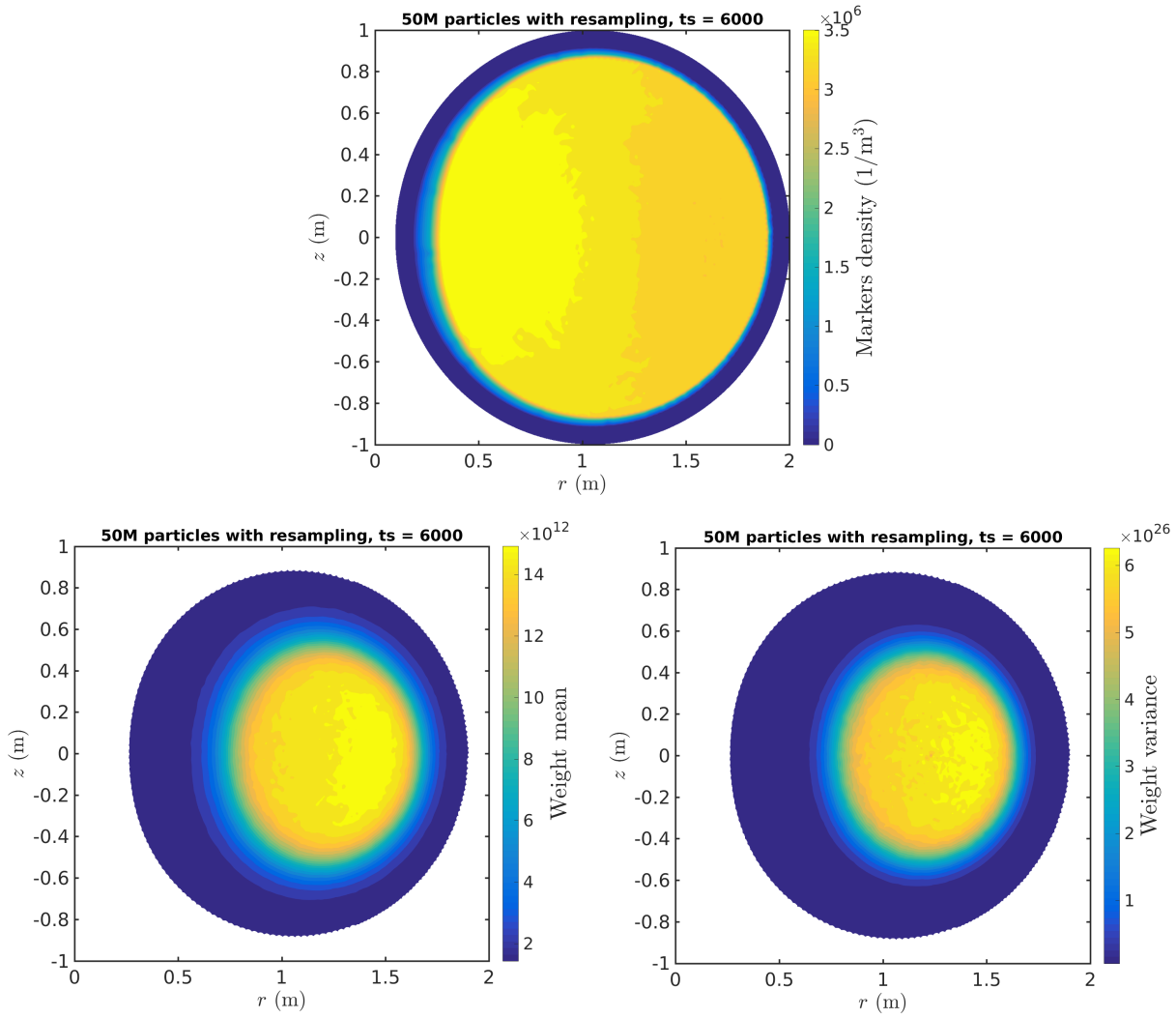


Figure 31: Spatial distribution of particle density and weights for the simulation with 50 million particles in three stages of resampling after 6000 time steps: (a) particle density, (b) mean of marker weights, (c) variance of marker weights. The computed values at vertices are interpolated in these plots.

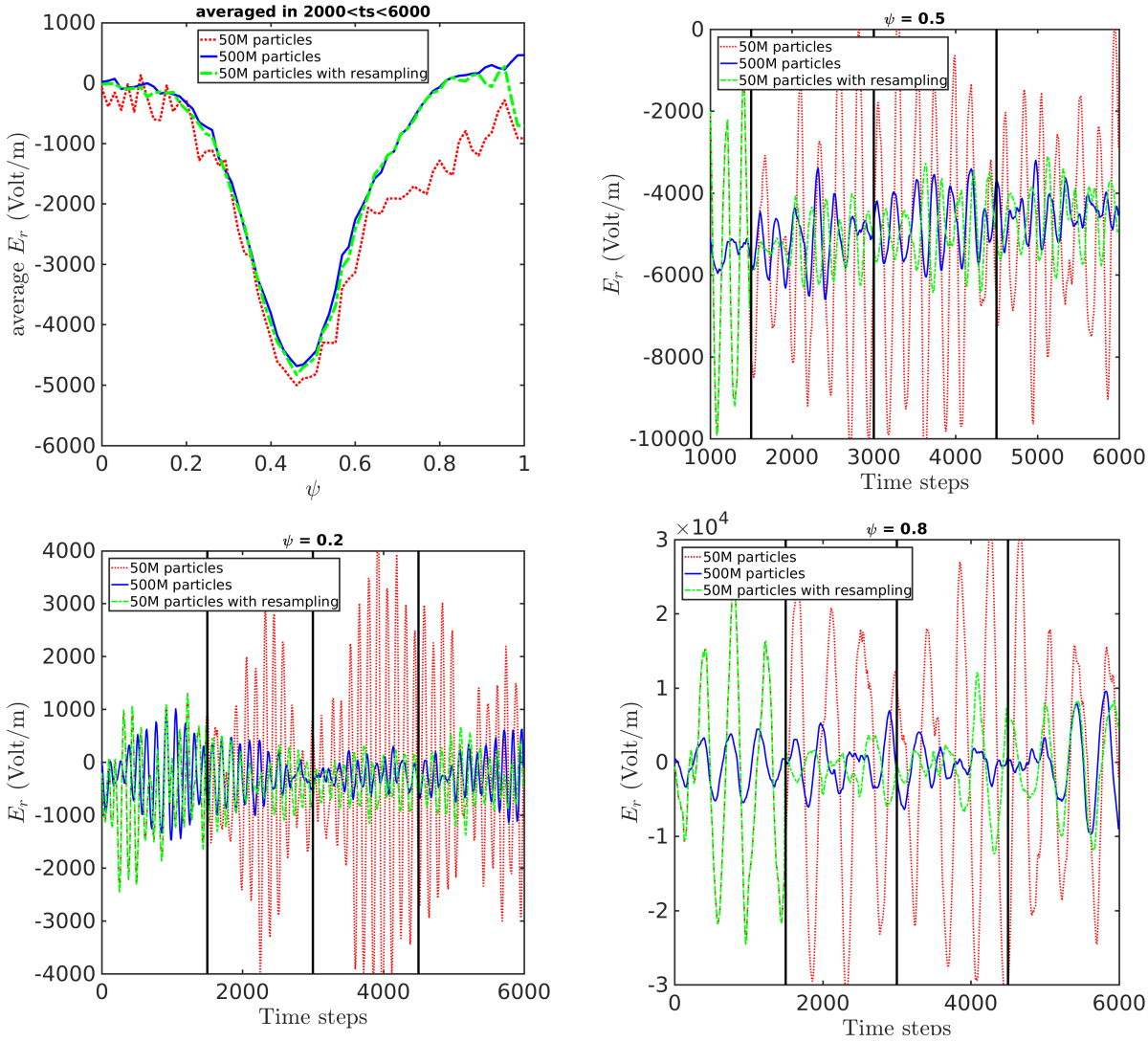


Figure 32: (a) E_r profile at 6000 time steps. Time evolution of E_r at (b) $\psi = 0.5$, (b) $\psi = 0.2$, (b) $\psi = 0.8$

ment of Energy under Contract No. DE-AC02-05CH11231 for providing HPC resources. Finally, the University of Texas team gratefully acknowledges the fruitful collaboration and many contributions of the entire EPSi team, and particularly C. S. Chang (PPPL), M. Greenwald (MIT), R. Hager (PPPL), and S. Janhunen (PPPL).

References

[1] Mark F Adams, Seung-Hoe Ku, Patrick Worley, Ed D'Azevedo, Julian C Cummings, and CS Chang. Scaling to 150k cores: recent algorithm and performance engineering developments enabling xgc1 to run at scale. In *Journal of Physics: Conference Series*, volume 180, page 012036. IOP Publishing, 2009.

[2] F. Assous, T. Pougeard Dulimbert, and J. Segre. A new method for coalescing particles in PIC codes. *Journal of Computational Physics*, 187(2):550 – 571, 2003. ISSN 0021-9991. doi: http://dx.doi.org/10.1016/S0021-9991(03)00124-4. URL `http://www.sciencedirect.com/science/article/pii/S0021999103001244`.

[3] Y. Baranov, K. Kirov, M. Goniche, J. Mailloux, M.L. Mayoral, J. Ongena, and F. Nave. Lh power modulation experiment on jet. *AIP Conference Proceedings*, 1187(1):343–346, 2009. doi: 10.1063/1.3273763. URL `http://aip.scitation.org/doi/abs/10.1063/1.3273763`.

[4] Kristopher R Beevers and Wesley H Huang. Fixed-lag sampling strategies for particle filtering slam. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 2433–2438. IEEE, 2007.

[5] A.T. Berwin. quadprog, a FORTRAN library for solving quadratic programming problems. `https://github.com/badgerodon/quadprog`, 2013.

[6] Charles K Birdsall and A Bruce Langdon. *Plasma physics via computer simulation*. CRC Press, 2004.

[7] C. Bourdelle, J. Citrin, B. Baiocchi, A. Casati, P. Cottier, X. Garbet, F. Imbeaux, and JET Contributors. Core turbulent transport in tokamak plasmas: bridging theory and experiment with qualikiz. *Plasma Physics and Controlled Fusion*, 58:014036, 2016. doi: 10.1088/0741-3335/58/1/014036.

[8] Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta Numerica*, 13:147269, 2004. doi: 10.1017/S0962492904000182.

[9] C. S. Chang, S. Ku, P. H. Diamond, Z. Lin, S. Parker, T. S. Hahm, and N. Samatova. Compressed ion temperature gradient turbulence in diverted tokamak edge. *Physics of Plasmas*, 16(5):056108, 2009. doi: 10.1063/1.3099329. URL `http://dx.doi.org/10.1063/1.3099329`.

[10] C.S. Chang and S. Ku. Spontaneous rotation sources in a quiescent tokamak edge plasma. *Physics of Plasmas*, 15(6):062510, 2008.

[11] P. Constantine. *Active Subspaces*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2015. doi: 10.1137/1.9781611973860. URL `http://epubs.siam.org/doi/abs/10.1137/1.9781611973860`.

[12] David P. Coster, Vincent Basiuk, Grigori Pereverzev, Denis Kalupin, Roman Zagorski, Roman Stankiewicz, Philippe Huynh, Frederic Imbeaux, and Task Force Integrated Tokamak Mode. The European Transport Solver. *IEEE TRANSACTIONS ON PLASMA SCIENCE*, 38 (9, 1):2085–2092, SEP 2010. ISSN 0093-3813. doi: {10.1109/TPS.2010.2056707}.

[13] G. Dif-Pradalier, P. H. Diamond, V. Grandgirard, Y. Sarazin, J. Abiteboul, X. Garbet, Ph. Ghendrih, A. Strugarek, S. Ku, and C. S. Chang. On the validity of the local diffusive paradigm in turbulent plasma transport. *PHYSICAL REVIEW E*, 82(2, 2), AUG 3 2010. ISSN 2470-0045. doi: {10.1103/PhysRevE.82.025401}.

[14] A. M. Dimits, G. Bateman, M. A. Beer, B. I. Cohen, W. Dorland, G. W. Hammett, C. Kim, J. E. Kinsey, M. Kotschenreuther, A. H. Kritz, L. L. Lao, J. Mandrekas, W. M. Nevins, S. E. Parker, A. J. Redd, D. E. Shumaker, R. Sydora, and J. Weiland. Comparisons and physics basis of tokamak transport models and turbulence simulations. *Physics of Plasmas*, 7(3):969–983, 2000. doi: 10.1063/1.873896. URL `http://dx.doi.org/10.1063/1.873896`.

[15] Rex A. Dwyer. Higher-dimensional voronoi diagrams in linear expected time. *Discrete & Computational Geometry*, 6(3):343–367, 1991. ISSN 1432-0444. doi: 10.1007/BF02574694. URL `http://dx.doi.org/10.1007/BF02574694`.

[16] ITER Physics Basis Editors, ITER Physics Expert Group Chairs, Co-Chairs, ITER Joint Central Team, and Physics Integration Unit. Chapter1: Overview and summary. *Nuclear Fusion*, 39(12):2137, 1999. URL `http://stacks.iop.org/0029-5515/39/i=12/a=301`.

[17] Donald Goldfarb and Ashok Idnani. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical programming*, 27(1):1–33, 1983.

[18] P. R. Goncharov, Y. Igitkhanov, S. Sudo, R. Dux, H. Funaba, D. Kalinina, M. Yokoyama, S. Murakami, A. Wakasa, N. Tamura, K. Kawahata, H. Yamada, and O. Motojima. Stellarator impurity strahl code development in nifs. *Plasma and Fusion Research*, 2:S1132–S1132, 2007. doi: 10.1585/pfr.2.S1132.

[19] Neil J Gordon, David J Salmond, and Adrian FM Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. In *IEE Proceedings F-Radar and Signal Processing*, volume 140, pages 107–113. IET, 1993.

[20] R. Hager and C. S. Chang. Gyrokinetic neoclassical study of the bootstrap current in the tokamak edge pedestal with fully non-linear coulomb collisions. *Phys. Plasmas*, 23(4):042503, 2016. doi: http://dx.doi.org/10.1063/1.4945615.

[21] R. Hager, ES Yoon, S Ku, EF D'Azevedo, PH Worley, and CS Chang. A fully non-linear multi-species Fokker-Planck-Landau collision operator for simulation of fusion plasma. *J. Comput. Phys.*, 315:644–660, 2016.

[22] TS Hahm. Nonlinear gyrokinetic equations for tokamak microturbulence. *Physics of Fluids (1958-1988)*, 31(9):2670–2673, 1988.

[23] P. Helander and D.J. Sigmar. *Collisional Transport in Magnetized Plasmas*. Cambridge Monographs on Plasma Physics. Cambridge University Press, 2005. ISBN 9780521020985. URL https://books.google.ca/books?id=nm-V01E0H2MC.

[24] Roger W Hockney and James W Eastwood. *Computer simulation using particles*. CRC Press, 1988.

[25] Yasuhiro Idomura and Motoki Nakata. Plasma size and power scaling of ion temperature gradient driven turbulence. *Physics of Plasmas*, 21(2):020706, 2014. doi: http://dx.doi.org/10.1063/1.4867379. URL http://scitation.aip.org/content/aip/journal/pop/21/2/10.1063/1.4867379.

[26] John D. Jakeman and Stephen G. Roberts. Local and dimension adaptive sparse grid interpolation and quadrature. *CoRR*, abs/1110.0010, 2011.

[27] F. Jenko, W. Dorland, and G. W. Hammett. Critical gradient formula for toroidal electron temperature gradient modes. *Physics of Plasmas*, 8(9):4096–4104, 2001. doi: 10.1063/1.1391261. URL http://dx.doi.org/10.1063/1.1391261.

[28] Genshiro Kitagawa. A Monte Carlo filtering and smoothing method for non-gaussian nonlinear state space models. In *Proceedings of the 2nd US-Japan joint seminar on statistical time series analysis*, pages 110–131, 1993.

[29] Genshiro Kitagawa. Monte Carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of computational and graphical statistics*, 5(1):1–25, 1996.

[30] R. A. Kolesnikov and J. A. Krommes. Transition to collisionless ion-temperature-gradient-driven plasma turbulence: A dynamical systems approach. *Phys. Rev. Lett.*, 94:235002, Jun 2005. doi: 10.1103/PhysRevLett.94.235002. URL http://link.aps.org/doi/10.1103/PhysRevLett.94.235002.

[31] R. A. Kolesnikov and J. A. Krommes. Bifurcation theory of the transition to collisionless ion-temperature-gradient-driven plasma turbulence. *Physics of Plasmas*, 12(12):122302, 2005. doi: 10.1063/1.2116887. URL http://dx.doi.org/10.1063/1.2116887.

[32] S. Ku, C.S. Chang, and P.H. Diamond. Full-f gyrokinetic particle simulation of centrally heated global itg turbulence from magnetic axis to edge pedestal top in a realistic tokamak geometry. *Nuclear Fusion*, 49(11):115021, 2009. URL http://stacks.iop.org/0029-5515/49/i=11/a=115021.

[33] S. Ku, C.S. Chang, and P.H. Diamond. Full-f gyrokinetic particle simulation of centrally heated global ITG turbulence from magnetic axis to edge pedestal top in a realistic tokamak geometry. *Nucl. Fusion*, 49(11):115021, 2009.

[34] S. Ku, R. Hager, C.S. Chang, J.M. Kwon, and S.E. Parker. A new hybrid-lagrangian numerical scheme for gyrokinetic simulation of tokamak edge plasma. *Journal of Computational Physics*, 315:467 – 475, 2016. ISSN 0021-9991. doi: http://dx.doi.org/10.1016/j.jcp.2016.03.062. URL http://www.sciencedirect.com/science/article/pii/S0021999116300274.

[35] S Ku, R Hager, CS Chang, JM Kwon, and SE Parker. A new hybrid-lagrangian numerical scheme for gyrokinetic simulation of tokamak edge plasma. *J. Comput. Phys.*, 315:467–475, 2016.

[36] S. Ku, R. Hager, C. S. Chang, J. M. Kwon, and S. E. Parker. A new hybrid-Lagrangian numerical scheme for gyrokinetic simulation of tokamak edge plasma. *Journal of Computational Physics*, 315:467–475, JUN 15 2016. ISSN 0021-9991. doi: {10.1016/j.jcp.2016.03.062}.

[37] Giovanni Lapenta. Particle rezoning for multidimensional kinetic particle-in-cell simulations. *Journal of Computational Physics*, 181(1):317 – 337, 2002. ISSN 0021-9991. doi: http://dx.doi.org/10.1006/jcph.2002.7126. URL http://www.sciencedirect.com/science/article/pii/S0021999102971263.

[38] Giovanni Lapenta and Jeremiah U. Brackbill. Dynamic and selective control of the number of particles in kinetic plasma simulations. *Journal of Computational Physics*, 115(1):213 – 227, 1994. ISSN 0021-9991. doi: http://dx.doi.org/10.1006/jcph.1994.1188. URL http://www.sciencedirect.com/science/article/pii/S0021999184711880.

[39] T C Luce, C C Petty, and J G Cordey. Application of dimensionless parameter scaling techniques to the design and interpretation of magnetic fusion experiments. *Plasma Physics and Controlled Fusion*, 50(4):043001, 2008. URL http://stacks.iop.org/0741-3335/50/i=4/a=043001.

[40] Phuc T. Luu, T. Tckmantel, and A. Pukhov. Voronoi particle merging algorithm for PIC codes. *Computer Physics Communications*, 202:165 – 174, 2016. ISSN 0010-4655. doi: http://dx.doi.org/10.1016/j.cpc.2016.01.009. URL http://www.sciencedirect.com/science/article/pii/S001046551600028X.

[41] D. R. Mikkelsen and W. Dorland. Dimits shift in realistic gyrokinetic plasma-turbulence simulations. *Phys. Rev. Lett.*, 101:135003, Sep 2008. doi: 10.1103/PhysRevLett.101.135003. URL http://link.aps.org/doi/10.1103/PhysRevLett.101.135003.

[42] EN Nerush, I Yu Kostyukov, AM Fedotov, NB Narozhny, NV Elkina, and H Ruhl. Laser field absorption in self-generated electron-positron pair plasma. *Physical review letters*, 106 (3):035001, 2011.

[43] L.W.T. Ng and M.S. Eldred. Multifidelity uncertainty quantification using nonintrusive polynomial chaos and stochastic collocation. In *Proceedings of the 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference (14th AIAA Non-Deterministic Approaches Conference)*, 2012.

[44] J. T. Oden, R. Moser, and O. Ghattas. Computer predictions with quantified uncertainty, Part I. *SIAM News*, 43(9), November 2010.

[45] J. T. Oden, R. Moser, and O. Ghattas. Computer predictions with quantified uncertainty, Part II. *SIAM News*, 43(10), December 2010.

[46] V. Parail, R. Albanese, R. Ambrosino, J. F. Artaud, K. Besseghir, M. Cavinato, G. Corrigan, J. Garcia, L. Garzotti, Y. Gribov, F. Imbeaux, F. Koechl, C. V. Labate, J. Lister, X. Litaudon, A. Loarte, P. Maget, M. Mattei, D. McDonald, E. Nardon, G. Saibene, R. Sartori, and J. Urban. Self-consistent simulation of plasma scenarios for ITER using a combination of 1.5D transport codes and free-boundary equilibrium codes. *NUCLEAR FUSION*, 53(11), NOV 2013. ISSN 0029-5515. doi: {10.1088/0029-5515/53/11/113002}.

[47] SE Parker, RJ Procassini, CK Birdsall, and BI Cohen. A suitable boundary condition for bounded plasma simulation without sheath resolution. *Journal of Computational Physics*, 104(1):41–49, 1993.

[48] B. Peherstorfer, K. Willcox, and M. Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *Technical Report, Aerospace Computational Design Laboratory*, (TR-16-1), 2016.

[49] Knut Petras. Smolyak cubature of given polynomial degree with few nodes for increasing dimension. *Numerische Mathematik*, 93(4):729–753, 2003. ISSN 0945-3245. doi: 10.1007/s002110200401. URL http://dx.doi.org/10.1007/s002110200401.

[50] M. Pfeiffer, A. Mirza, C.-D. Munz, and S. Fasoulas. Two statistical particle split and merge methods for particle-in-cell codes. *Computer Physics Communications*, 191:9 – 24, 2015. ISSN 0010-4655. doi: http://dx.doi.org/10.1016/j.cpc.2015.01.010. URL http://www.sciencedirect.com/science/article/pii/S0010465515000302.

[51] M. J. Pueschel and F. Jenko. Transport properties of finite- microturbulence. *Physics of Plasmas*, 17(6):062307, 2010. doi: 10.1063/1.3435280. URL http://dx.doi.org/10.1063/1.3435280.

[52] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[53] B. N. Rogers, W. Dorland, and M. Kotschenreuther. Generation and stability of zonal flows in ion-temperature-gradient mode turbulence. *Phys. Rev. Lett.*, 85:5336–5339, Dec 2000. doi: 10.1103/PhysRevLett.85.5336. URL http://link.aps.org/doi/10.1103/PhysRevLett.85.5336.

[54] S. Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. *Soviet Mathematics, Doklady*, 4:240–243, 1963.

[55] D. A. St-Onge. On nonlocal energy transfer via zonal flow in the dimits shift. *arxiv preprint*, pages 1–28, 2017. URL https://arxiv.org/abs/1704.05406.

[56] G. M. Staebler, J. E. Kinsey, and R. E. Waltz. A theory-based transport model with comprehensive physics. *PHYSICS OF PLASMAS*, 14(5), MAY 2007. ISSN 1070-664X. doi: {10.1063/1.2436852}. 48th Annual Meeting of the Division of Plasma Physics of the APS, Philadelphia, PA, JAN 30-NOV 03, 2006.

[57] T E Stringer. Different forms of the plasma energy conservation equation. *Plasma Physics and Controlled Fusion*, 33(14):1715, 1991. URL http://stacks.iop.org/0741-3335/33/i=14/a=001.

[58] Shigeru Sudo. A review of impurity transport characteristics in the lhd. *Plasma Physics and Controlled Fusion*, 58(4):043001, 2016. URL http://stacks.iop.org/0741-3335/58/i=4/a=043001.

[59] Shigeru Sudo, Naoki Tamura, Hisamichi Funaba, Sadatsugu Muto, Chihiro Suzuki, and Izumi Murakami. Impurity transport study with tespel injection and simulation. *Plasma and Fusion Research*, 8:2402059–2402059, 2013. doi: 10.1585/pfr.8.2402059.

[60] P. W. Terry, K. D. Makwana, M. J. Pueschel, D. R. Hatch, F. Jenko, and F. Merz. Mode-space energy distribution in instability-driven plasma turbulence. *Physics of Plasmas*, 21 (12):122303, 2014. doi: 10.1063/1.4903207. URL http://dx.doi.org/10.1063/1.4903207.

[61] Jannis Teunissen and Ute Ebert. Controlling the weights of simulation particles: adaptive particle management using k-d trees. *Journal of Computational Physics*, 259:318 – 330, 2014. ISSN 0021-9991. doi: http://dx.doi.org/10.1016/j.jcp.2013.12.005. URL http://www.sciencedirect.com/science/article/pii/S0021999113008048.

[62] AN Timokhin. Time-dependent pair cascades in magnetospheres of neutron stars–i. dynamics of the polar cap cascade with no particle supply from the neutron star surface. *Monthly Notices of the Royal Astronomical Society*, 408(4):2092–2114, 2010.

[63] T. Vernay, S. Brunner, L. Villard, B. F. McMillan, S. Jolliet, T. M. Tran, and A. Bottino. Synergy between ion temperature gradient turbulence and neoclassical processes in global gyrokinetic particle-in-cell simulations. *Physics of Plasmas*, 19(4):042301, 2012. doi: 10.1063/1.3699189. URL http://dx.doi.org/10.1063/1.3699189.

[64] M. Vranic, T. Grismayer, J.L. Martins, R.A. Fonseca, and L.O. Silva. Particle merging algorithm for PIC codes. *Computer Physics Communications*, 191:65 – 73, 2015. ISSN 0010-4655. doi: http://dx.doi.org/10.1016/j.cpc.2015.01.020. URL http://www.sciencedirect.com/science/article/pii/S0010465515000405.

[65] M. R. Wade, T. C. Luce, and C. C. Petty. Gyroradius scaling of helium transport. *Phys. Rev. Lett.*, 79:419–422, Jul 1997. doi: 10.1103/PhysRevLett.79.419. URL https://link.aps.org/doi/10.1103/PhysRevLett.79.419.

[66] W X Wang, N Nakajima, M Okamoto, and S Murakami. A new delta-f method for neoclassical transport studies. *Plasma Phys. Contr. F.*, 41(9):1091, 1999.

[67] Jan Weiland. *Collective modes in inhomogeneous plasma : kinetic and advanced fluid theory*. Plasma physics series. Institute of Physics, Bristol (UK), Philadelphia, 2000. ISBN 0-7503-0589-4. URL http://opac.inria.fr/record=b1132029.

[68] D.R. Welch, T.C. Genoni, R.E. Clark, and D.V. Rose. Adaptive particle management in a particle-in-cell code. *Journal of Computational Physics*, 227(1):143 – 155, 2007. ISSN 0021-9991. doi: http://dx.doi.org/10.1016/j.jcp.2007.07.015. URL http://www.sciencedirect.com/science/article/pii/S0021999107003245.

[69] J. Wesson and D.J. Campbell. *Tokamaks*. International Series of Monographs on Physics. OUP Oxford, 2011. ISBN 9780199592234. URL https://books.google.ca/books?id=BH9vx-iDI74C.

[70] Chak-Kuen Wong and Malcolm C. Easton. An efficient method for weighted sampling without replacement. *SIAM Journal on Computing*, 9(1):111–113, 1980.

[71] E. S. Yoon and C. S. Chang. A Fokker-Planck-Landau collision equation solver on two-dimensional velocity grid and its application to particle-in-cell simulation. *Phys. Plasmas*, 21(3):032503, 2014. doi: http://dx.doi.org/10.1063/1.4867359.

[72] Fan Zhang, Robert Hager, Seung-Hoe Ku, Choong-Seock Chang, Stephen C. Jardin, Nathaniel M. Ferraro, E. Seegyoung Seol, Eisung Yoon, and Mark S. Shephard. Mesh generation for confined fusion plasma simulation. *Engineering with Computers*, 32(2):285–293, 2016. ISSN 1435-5663. doi: 10.1007/s00366-015-0417-y. URL http://dx.doi.org/10.1007/s00366-015-0417-y.