Physics of Plasmas ARTICLE pubs.aip.org/aip/pop

Large-database cross-verification and validation of tokamak transport models using baselines for comparison

Cite as: Phys. Plasmas **31**, 042506 (2024); doi: 10.1063/5.0190908
Submitted: 9 December 2023 · Accepted: 15 March 2024 ·
Published Online: 11 April 2024

View Online Export Citation CrossMark

J. Abbate,[1] E. Fable,[2] B. Grierson,[3] A. Pankin,[1] G. Tardini,[2] and E. Kolemen[1,a)]

AFFILIATIONS

[1]Princeton Plasma Physics Laboratory, Princeton, New Jersey 08540, USA
[2]Max-Planck-Institut für Plasmaphysik, Garching 85748, Germany
[3]General Atomics, San Diego, California 92121, USA

[a)]**Also at:** Department Mechanical and Aerospace Engineering, Princeton University; and Andlinger Center for Energy and the Environment, Princeton University. **Author to whom correspondence should be addressed:** ekolemen@princeton.edu

ABSTRACT

State-of-the-art 1D transport solvers ASTRA and TRANSP are verified, then validated across a large database of semi-randomly selected, time-dependent DIII-D discharges. Various empirical models are provided as baselines to contextualize the validation figures of merit using statistical hypothesis tests. For predicting plasma temperature profiles, no statistically significant advantage is found for the ASTRA and TRANSP simulators over a baseline empirical (two-parameter) model. For predicting stored energy, a significant advantage is found for the simulators over a baseline empirical model based on confinement time scaling. Uncertainty in the results due to diagnostic and profile fitting uncertainties is approximated and determined to be insignificant due in part to the large quantity of discharges employed in the study. Advantages are discussed for validation methodologies like this one that employ (1) large databases and (2) baselines for comparison that are specific to the intended use-case of the model.

I. INTRODUCTION

Tokamak plasma dynamics are governed by a variety of physics, each on distinct timescales. While computational models capable of tracking individual particles through phase space, enabling a comprehensive consideration of all relevant timescales and internal physics phenomena, these simulations are currently constrained by excessive computational time requirements for executing a full discharge. Moreover, they necessitate various assumptions regarding the interaction of the plasma with external actuators and boundary conditions. In the context of realistic "whole-device modeling", the evolution of a plasma is approximated using adjustable parameters such as injected power, plasma shape, and plasma current.

The forefront of this field is represented by "integrated modeling", which employs separate approximations for the diverse timescales and physics. This approach involves assuming a diffusion-like partial differential equations (PDEs) on a coarser (transport) timescale as an ansatz solution. In this framework, the plasma state is defined by one-dimensional flux-surface-averaged profiles encompassing temperature, density, rotation, current, and geometric parameters that characterize the equilibrium. The equilibrium is either given as input or calculated based on the boundary shape (fixed boundary) or external coil currents (free boundary). The one-dimensional profiles (for any profile X over spatial coordinate x) are assumed to evolve ($\sim\frac{\partial X}{\partial t}$) with a combination of a source term, diffusion ($\sim-\nabla^2 X$), and pinch ($\sim-\nabla X$). These equations collectively account for the conservation of particles, momentum, energy, and current in the system. More details can be found in Ref. 1.

Most state-of-the-art integrated modeling efforts employ simplified calculations (e.g., MMM,[2] QuaLiKiz,[3] GLF23[4]/TGLF[5]) to predict turbulent flux given the plasma state, which can be used to back out diffusivity and/or pinch terms. Additionally, an approximation for the neoclassical transport is typically incorporated especially for the ion channel, e.g., from simple analytic estimates like the Chang–Hinton model,[6] or more complete codes like NCLASS[7] or NEO.[8]

Separately from neoclassical and microturbulent transport, global MHD instabilities can set the effective transport in many cases, e.g.,

tearing modes, sawteeth, and ELMs.[9] Tearing modes and other macroscale MHD activity are sometimes included (see, e.g., Ref. 10). Sawteeth are sometimes modeled by expulsion of current, particles, and heat from the core at a specified temporal periodicity when the safety factor drops below 1 (see Refs. 11 and 12 for its use in reactor studies). For ELMs, many believe the plasma pressure pedestal and height in H-mode plasmas is determined primarily by global MHD.[13] For H-mode plasmas, then, another constraint is added to the methodology: the pedestal height must be guessed or calculated, and the transport equations are solved inward from that boundary condition. At present, EPED[14] is a state of the art model but still requires the pedestal density as input to back-determine the temperature (e.g., using a two-point model from the scrape-off layer[15]) Furthermore, as elucidated in Ref. 16, most integrated modeling assumes an empirically determined finite diffusion coefficient in the core. This is necessary to reproduce the reality of core flux due to macroscopic current-driven instabilities like sawteeth near the magnetic axis, which are neglected in most gyrokinetic simulations.[17] In addition, for representing the influence of pinch phenomena (expressed as $\sim -\nabla X$), an empirically determined quantity is commonly employed.

Regarding the source terms, particles are primarily sourced from ionization of neutral atoms inject from various mechanisms such as gas puffing, neutral beam fueling, and gas pellet ablation. On the other hand, heat within the plasma is predominantly generated through the absorption of electromagnetic waves and the impact of neutral beams. A multitude of computational models and codes have been developed to simulate each of these specific phenomena.

Integrated modeling suites orchestrate all of these external models to determine the diffusivity, pinch, pedestal, and source terms. Time-dependent codes (e.g., TRANSP,[18] ASTRA,[19,20] and JETTO[21]) evolve the coupled PDEs forward, while steady-state codes (TGYRO[22]) use a shooting method to find a self-consistent solution. Most attempts at model-based control (e.g., RAPTOR[23] and COTSIM[24]) and reactor design studies (e.g., for SPARC[25] and ITER[26]) employ this methodology to supplement empirical scalings as the most accurate prediction of startup and flattop operation.

Understanding the accuracy and generalizability of this methodology is important given its use in scenario and reactor planning, and many papers have reported "figures of merit" outlined by the ITER working group in 1999.[26] Key among these metrics are σ and f, which measure discrepancies in profiles X across spatial points R_j, and $\Delta\bar{R}_W$, which measures discrepancies in stored energy W,

$$\varepsilon_j = X_{\text{prediction}}(R_j) - X_{\text{truth}}(R_j), \tag{1a}$$

$$f = \frac{\frac{1}{N}\sum_{j=1}^{N}\varepsilon_j}{\sqrt{\frac{1}{N}\sum_{j=1}^{N}X_{\text{truth}}(R_j)^2}}, \tag{1b}$$

$$\sigma = \frac{\sqrt{\frac{1}{N}\sum_{j=1}^{N}\varepsilon_j^2}}{\sqrt{\frac{1}{N}\sum_{j=1}^{N}X_{\text{truth}}(R_j)^2}}, \tag{1c}$$

$$\bar{\sigma} = \sqrt{\langle\sigma^2\rangle}, \tag{1d}$$

$$R_W = \frac{W_{\text{prediction}}}{W_{\text{truth}}}, \tag{1e}$$

$$\Delta R_W = R_W - 1, \tag{1f}$$

$$\Delta\bar{R}_W = \sqrt{\langle\Delta R_W^2\rangle}, \tag{1g}$$

where brackets indicate an average over all samples in a dataset (in this study considering different timesteps of the same discharge as independent samples, as discussed later).

The working group assessed a variety of models and reported the figures of merit on small-ELM H-mode, large-ELM H-mode, and L-mode cases across a variety of tokamaks. Many similar validation studies have been conducted since then, alongside discussions of the philosophy behind verification and validation using these metrics and others.[5,16,27–36]

In this study, the most up-to-date versions of the ASTRA and TRANSP codes are cross-verified for the first time. The codes are then validated against experimental data with workflows and validation metrics similar to previous studies. However, this validation adds a few important and novel features.

First, a large set of semi-randomly discharges is chosen by employing automated workflows to setup runs. This gives a sense of performance on unexpected distributions, more akin to what can be expected when predicting for a new scenario or device. It also provides more independent samples of similar discharges so that the law of large numbers guarantees lower uncertainty in the estimates of the mean for figures of merit, like $\bar{\sigma}$.

Second, multiple different transport models are evaluated compared to both one another and to experiment, in the spirit of independent validation as done periodically by the ITER working group. In particular, the state-of-the-art integrated modeling suites TRANSP and ASTRA are compared alongside simple empirical rules of thumbs for predicting plasma profile evolution (e.g., based on the ITER H89 and H98 scalings for energy confinement time).

The paper concludes with uncertainty quantification and a discussion of the benefits of validating with a large database of low-fidelity profile fits as opposed to a small number of hand-fit discharges.

II. METHODOLOGY

Within the One Modeling Framework for Integrated Tasks (OMFIT),[37] a module[38] had already been developed for generating TRANSP input files and initiating TRANSP simulations on the TRANSP computing grid. A separate module within OMFIT was established as part of this study to facilitate the generation and execution of ASTRA runs for various DIII-D discharges, marking the first instance of running ASTRA on DIII-D's computer cluster.

For the workflow discussed in this paper, the same inputs and, where feasible, identical settings as those used in the TRANSP setup have been employed. Because a study over a large database requires robust and automatic workflows for generating plasma profiles, this study utilizes relatively low-fidelity plasma profile estimates. General Atomics's standard Grad–Shafranov solver EFIT[39] using only edge magnetic constraints ("EFIT01") is used for safety factor q. For other profiles, automatically fitted profiles (ZIPFITs) that incorporate data solely from Thomson Scattering and Charge Exchange Recombination

Spectroscopy (CER). An uncertainty quantification and discussion in Sec. V argues that this choice is reasonable, and outlines how this methodology could be improved to use higher-fidelity fits in the future.

Discharges are semi-randomly selected from the database. Specifically, only DIII-D discharges from 2010 and later are used, to avoid issues with data consistency further in the past. Only discharges where all automatic ZIPFIT profile fits and other needed diagnostics for the workflow are available are used, which excludes about half of discharges. Though arguably more relevant for reactors and making up about a third of DIII-D experiments, discharges with wave heating/current drive (mostly Electron Cyclotron, but also Helicon and Fast Alfvén Wave) are also excluded to keep the number of actuators (and therefore the considerations for the validation) as low as possible. Only times in discharges during current flattop are considered. An effort is also made to exclude discharges that have non-D2 gas in the valves, and also those with externally applied 3D field perturbations.

Following a relatively simple methodology among the many validation studies referenced in the introduction, two transport channels are simulated: the core electron and ion temperatures. Options for running TRANSP and ASTRA are chosen to make them as identical as possible while respecting differences inherent to their current implementations. The models take as input the time-dependent boundary temperatures (outside of $\rho = 0.8$, where ρ is the normalized equivalent radius of toroidal magnetic flux surface throughout this paper), along with time-dependent profiles of density, rotation, q, and Z_{eff}. In the TRANSP simulation, all equilibrium surfaces are specified by an automated (EFIT01) Grad–Shafranov solution constrained only by edge magnetics (smoothed to 16 Fourier coefficients). In the ASTRA simulation, because in the current version the functionality of specifying all equilibrium surfaces is not well-validated, the SPIDER equilibrium solver[40] is employed to calculate the internal plasma magnetic flux, prescribing just the time-dependent last-closed flux surface boundary shape from EFIT01 (also smoothed to 16 Fourier coefficients). TGLF with the SAT2 saturation rule[41] is employed as the transport solver for both simulators, with the same settings for TRANSP and ASTRA. Following many of the references in the introduction, an *ad hoc* minimum value for the core heat diffusion coefficients (at 0.1 cm^2/s inside $\rho = 0.1$) is set. No sawtooth model is used, so (following Ref. 5) figures of merit are only considered outside of the $q = 1$ surface during validation. In TRANSP, the Monte Carlo code NUBEAM[42] is used to calculate beam deposition and fast ion pressure; in ASTRA the realtime-capable RABBIT code[43] is used. This decision to use different beam deposition models was largely made because the standard TRANSP workflow employs NUBEAM, while the standard ASTRA workflow employs RABBIT. The fast ion distribution and transport coefficients are updated every 10 ms in both codes. In TRANSP, a Chang-Hinton neoclassical diffusivity for electrons and ions[6] is added to the corresponding turbulent contributions. In ASTRA, the neoclassical additive contribution comes from an Angioni–Sauter model[44] for ions and a Galeev–Sagdeev model[45] for electrons.

Both codes are run with 50 radial grid points. In both ASTRA and TRANSP, the Pereverzev-Corrigan solver[46] is used to evolve the partial differential equations. The solver time step dynamically changes (based on, e.g., the time-variation in profiles). TGLF is evaluated for each of the 50 radial grid points, parallelized across 16 CPUs for ASTRA and 64 for TRANSP. With the settings used for this study, the internal solver for TRANSP is more self-consistent but time-consuming than ASTRA's. Specifically, in ASTRA the diffusion coefficients are updated every 10 ms based on the plasma state at the time of calculation (i.e., TGLF is called every 10 ms). In the TRANSP runs, diffusion coefficients are calculated at every single solver step; and each time a calculation is made up to one hundred Newton iterations are allowed in order to find implicitly self-consistent gradients and diffusion coefficients (i.e., TGLF is generally called much more often).[47] In the ASTRA runs the diffusivity profiles X_{in} are smoothed with a small parameter $\alpha = 0.001$ such that the employed diffusivity profiles X_{out} minimize the functional

$$\int_0^1 \alpha\left(\frac{\partial X_{out}}{\partial \rho}\right)^2 + (X_{in}(\rho) - X_{out}(\rho))^2 d\rho. \quad (2)$$

Meanwhile, in TRANSP there is no such explicit smoothing of the diffusivity. The authors find that when running ASTRA with unsmoothed TRANSP diffusivity profiles, ASTRA runs often crash.

For reference, the detailed execution time for an example run is shown in Table I. TRANSP runs currently must be submitted to a dedicated PPPL computer cluster (the TRANSP grid) so that the runtime comparison between TRANSP and ASTRA is necessarily done on different hardware. Additionally, in contrast to ASTRA, TRANSP's execution time is fairly variable due to the variable number of TGLF calls (also shown) and the variability in the NUBEAM runtime with, e.g., the details of the beam geometry; for reference we show a slower TRANSP execution case, a faster execution case, and the ASTRA execution time (which is about the same for the two cases). For the settings used in this study, ASTRA is about an order of magnitude faster than TRANSP with the provided settings.

TRANSP–NUBEAM and ASTRA–RABBIT are both given an excessive amount of time of 300 ms to warmup. Here warmup primarily means allowing the fast ion distribution to develop without evolving profiles; the beam slowing down time at DIII-D is order tens of milliseconds. The predictions start at a randomly selected time (within the available diagnostics and during current flattop), and the electron temperature and ion temperature are evolved predictively for 900 ms past the warmup. No fine-tuning or reruns are done: if a run fails for either or both simulators the discharge is excluded completely from the analysis. Figure 1 reports the number of discharges that did not crash for the full run by both, one, or neither of the codes. Of the 218 attempted runs, just over half converged for both. For cases where only one of the codes runs, TRANSP converges for almost three times as many cases as ASTRA, though this is likely at the cost of the added computation time for TRANSP's self-consistent diffusivity calculations. These 116 cases are those used in the remainder of this study.

TABLE I. Solver computation time (in hours) for a slow and fast TRANSP run vs ASTRA. TRANSP is variable across runs primarily due to the variable number of TGLF calls.

	TRANSP (slow)	TRANSP (fast)	ASTRA
Solver wall-clock	4.6	3.2	1.1
Solver CPU	294.4	204.8	17.6
# TGLF calls	3634	2627	90

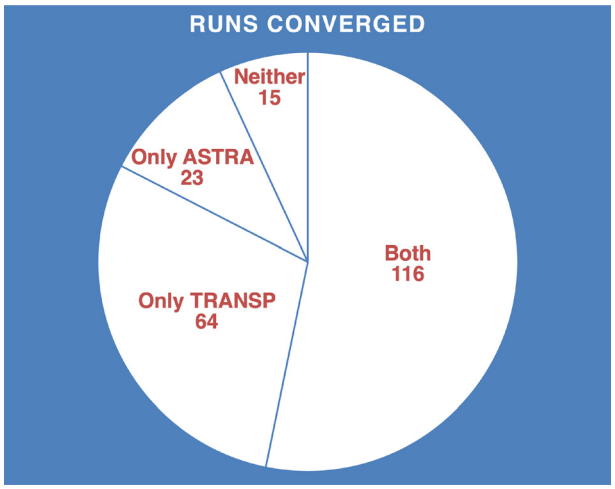

FIG. 1. Of the 218 runs automatically attempted, 53% converged for both ASTRA and TRANSP and were used for validation in this study.

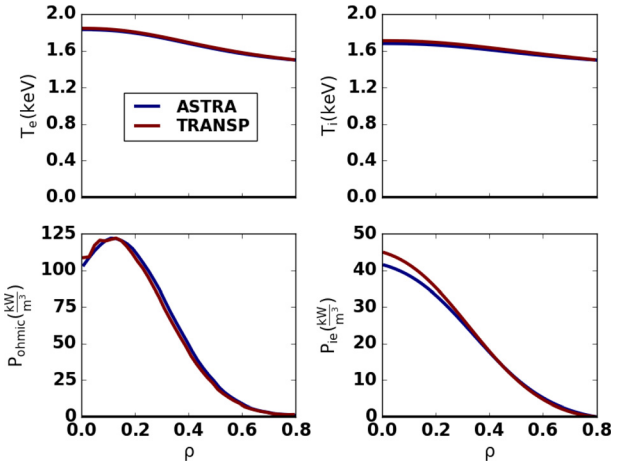

FIG. 2. Steady-state spatial verification between ASTRA and TRANSP. T_e and T_i predictions are within 2% with only minor discrepancies in internally calculated quantities like Ohmic power and ion–electron heat exchange.

III. CROSS-VERIFICATION

A fundamental step in validating the physical accuracy of computational models is to verify that these models solve their respective equations accurately, free from numerical errors or software bugs. While verification alone is not sufficient to guarantee the correctness of a code, it is an essential preliminary step. It ensures that the code is implementing the intended mathematical models and algorithms accurately and efficiently. Cross-verification of codes, which involves comparing different codes against each other (in this case ASTRA and TRANSP), is a valuable method in this process. If the different codes, developed independently and using slightly different numerical methods and assumptions, yield consistent results when applied to the same problem, it provides an indication that the codes are correctly solving the equations they are designed to model.

A. Simple steady-state verification

A simplified, steady-state, contrived plasma state is developed based loosely on DIII-D discharge 175 711, a limited plasma. The state is generated by taking the base discharge's state at 2.6s and fixing all profiles and the equilibrium. The neutral beam power is turned off, and the electron and ion heat diffusivities are set at a constant 1 cm^2/s plus a neoclassical-like diffusivity. Rotation is set to 0 for simplicity. The core electron and ion temperatures are then evolved, with all other profiles and the boundary values fixed, to the new steady state. By using this same procedure for both codes, many of the internal features of the codes can be cross-verified: PDE-solving for calculating T_e and T_i; internal smoothing of inputs; geometric parameters calculated from the equilibria and used in transport equations like differential volume ($\frac{dV}{d\rho}$); calculation of the Ohmic power (P_{Ohmic}) and electron-ion heat exchange (P_{ie}). Figure 2 shows the final agreement of electron and ion temperature to be within about 2%, with slight discrepancies in P_{Ohmic} and P_{ie}. Though not shown here, the time-dependent trajectory from initial trajectory to the steady-state solution is also within 2% at all times and for all spatial points, and agreement for other relevant quantities (such as geometric parameters like differential volume) are essentially identical.

B. Real discharge database verification

When predicting core heat transport for a real discharge, however, the discrepancy is closer to 10%–20% in many cases. Some of the factors not considered in the simple example with near-perfect agreement include time-dependent inputs; physics calculations of the heating power by external codes RABBIT, NUBEAM, and TGLF; and the more self-consistent solver in TRANSP (described earlier) vs the smoothed diffusivity in ASTRA.

A comparison of ASTRA vs TRANSP T_i-related quantities of interest for all the real-discharge database cases are shown in Fig. 3, in both a core and edge region of the plasma. Specifically, ASTRA-RABBIT systematically underpredicts the flux-surface-averaged temperature $\langle T_i \rangle$ relative to TRANSP-NUBEAM. The differential heat flux in each of the regions Q_i also appears to be underpredicted by ASTRA-RABBIT vs TRANSP-NUBEAM. The discrepancy in the heat diffusivity χ_i is more random. The ρ-averaged total ion power distribution $\langle P_{i,tot} \rangle$ is somewhat random; Neutral Beam Injection power $\langle P_{i,NBI} \rangle$ is overpredicted at high values and $\pm 20\%$ at low values; and parallel Neutral Beam Injection fast ion pressure $\langle p_{NBI,\parallel} \rangle$ is underpredicted by about 20%. The upshot is that there is a confluence of discrepancies from many sources. Though not shown here, the results are analogous for the electron channel.

The primary differences between TRANSP–NUBEAM and ASTRA–RABBIT are

1. Computation of heat fluxes and heat sources (primarily beam deposition, NUBEAM in TRANSP vs RABBIT in ASTRA)
2. Equilibrium flux surfaces (EFIT01 from experiment in TRANSP vs self-consistent SPIDER in ASTRA)
3. More self-consistent solver in TRANSP vs smoothed diffusivity in ASTRA

The geometric parameters that matter for the transport equations (primarily differential volume $\frac{dV}{d\rho}$ and flux-surface-averaged flux squared $\langle \rho^2 \rangle$) are confirmed to be correct to within a percent, so this is likely not the cause of significant error. To further dissect the sources of error, particularly differentiating between those arising from heat/

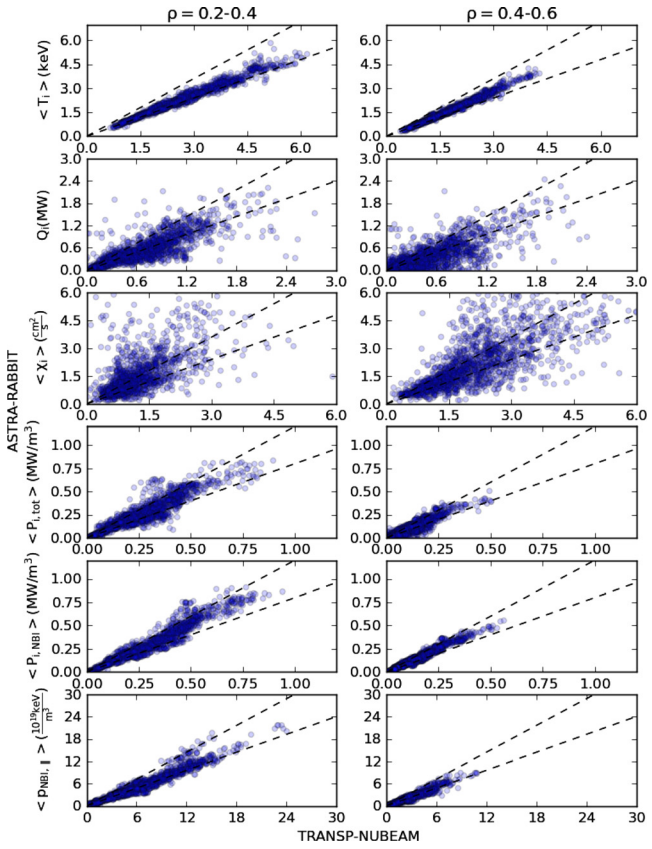

FIG. 3. Timeslice-by-time slice comparison of transport-relevant signals for the ion temperature channel over all runs in the database for ASTRA–RABBIT vs TRANSP–NUBEAM. The left and right columns correspond to core (averaged between $\rho = 0.2$–0.4 and edge (averaged between $\rho = 0.4$–0.6) values, respectively. The dashed black lines correspond to $\pm 20\%$ agreement between the codes. ASTRA–RABBIT systematically underpredicts T_e and T_i by nearly 20% relative to TRANSP–NUBEAM.

diffusivity calculations vs the solver, a comprehensive approach is undertaken.

C. Analysis of ASTRA–TRANSP discrepancy

All discharges in the database are resimulated using ASTRA, but with specific parameters such as heat source (including Ohmic, neutral beam, electron-ion exchange, and radiation loss terms), fast ion pressure, and diffusivities directly input from the values previously computed by TRANSP (though still at 10 ms fixed intervals and smoothed to avoid a much higher rate of ASTRA runs crashing). This reprocessing results in only a partial improvement in the alignment of results, as detailed in Table II using $\bar{\sigma}$ [Eq. (1d)] as a metric for discrepancy. This suggests that other factors (such as the self-consistent solver in TRANSP and smoothing of diffusivity in ASTRA) are still significant in explaining the discrepancy between the codes. Further work is needed to fully dissect the discrepancy.

The offset f [Eq. (1b)] between TRANSP and ASTRA for the two run cases is also shown in Fig. 4. This shows that both the discrepancies in heat/diffusivity and solving systematically bias ASTRA's predictions lower relative to TRANSP.

TABLE II. Deviation from TRANSP–NUBEAM for ASTRA–RABBIT (top) and ASTRA using heat deposition and diffusivity profiles from TRANSP's calculation (bottom). The difference in heat and diffusivity account for a few percent of the overall discrepancy between TRANSP and ASTRA calculations in the database, the rest potentially due to differences in the internal smoothing and solving.

	T_e	T_i	W_{MHD}
ASTRA–RABBIT	15.3%	14.5%	18.6%
ASTRA with TRANSP heat and χ	10.4%	12.4%	17.5%

These findings underscore the complexity of accurately modeling plasma transport phenomena and highlight the importance of continuous refinement and cross-verification of models to enhance their predictive reliability. Nonetheless, comparing simulators to one another says nothing about how well either of them describes the physical reality being simulated: it is unclear from these results whether ASTRA or TRANSP is "more correct", it is just shown that there is a difference between them.

IV. VALIDATION

In Sec. III, it was verified in a simple case that the ASTRA and TRANSP implementations of the basic physics and transport equations yield results that are fairly similar. It was then shown that when modeling a time-dependent discharge there are compounding errors which yield significant discrepancies between the models. In this section, the setups are validated against both experimental data and simple empirical models to study the "correctness" of the models.

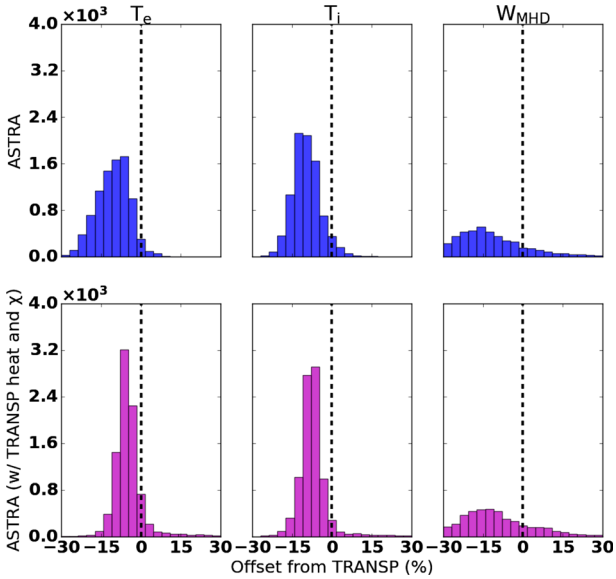

FIG. 4. Histograms showing number of timeslices in database with offset between TRANSP-NUBEAM and ASTRA–RABBIT (top) and ASTRA run with TRANSP's heat and diffusivity (bottom). This suggests the difference in calculation of heat and diffusivity profiles accounts for a few percent of the discrepancy between TRANSP–NUBEAM and ASTRA–RABBIT.

As mentioned in the introduction, the ITER profile database was constructed in 2000 to facilitate easy comparison among many different transport models as the community looked to predict-first simulations of ITER.[48] Many of the validation studies referenced in the introduction used profiles from this database, and also used the now-standard $\bar{\sigma}$ metric [Eq. (1d)] and $\Delta\bar{R}_W$ metric [Eq. (1g)] for determining accuracy of a given transport model. Departing from the ITER profile database and most other validation studies, the stored energy considered here is the total (including fast ions) rather than just thermal

$$W_{MHD} = \frac{3}{2}\int (P_{\text{thermal}} + P_{\text{fast ions}})dV, \tag{3}$$

where the experimental W_{MHD} is measured by magnetics data (in this case using the integral of the pressure calculated from the EFIT01 Grad–Shafranov solver). Including fast ions further validates the beam deposition codes' ability to track the fast ion population (accounting also for pressure). Including this metric also has the benefit that W_{MHD} is an independent measurement (only Thomson and CER data are used for the profiles, while only external magnetics are used for W_{MHD}). Note that the fast ion pressures calculated by NUBEAM and RABBIT are only slightly anisotropic, but for completeness the average between the perpendicular and parallel pressures $P_{\text{fast ions}} = \frac{1}{2}(P_{\text{fast ions},\parallel} + P_{\text{fast ions},\perp})$ is used for calculating W_{MHD}.

To calculate the figures of merit, the models are interpolated onto the space and time base of the corresponding experimentally fitted profiles (usually 10–20 ms for both T_e and T_i, with fewer T_i datapoints due to lack of charge exchange spectroscopy data when neutral beams are off; and millisecond timescale for W_{MHD}).

These figures of merit are often used as a reference for comparing models and characterizing sensitivity to input parameters, but (as discussed by, e.g., Ref. 5) there is no one good metric for understanding the accuracy of simulators. In this work, a simple empirical baseline model is therefore employed to contextualize each of the figures of merit.

A. Description of baselines for comparison

As a baseline for comparison for T_e and T_i predictions, a few simple plasma physics assumptions are employed. "Profile consistency"[49] is the concept that plasma profiles in tokamaks tend to maintain a consistent shape regardless of the details of the heat sources. More specifically, through the 1990s it was posited that in tokamaks the primary transport mechanism comes from ion temperature gradients (ITG) microinstabilities, which tend to produce profiles of the form $T = ke^{-C\rho}$ for ρ the radial position.[50] Additionally, as described in the introduction, it is generally assumed that near the magnetic axis of the tokamak plasma (inside $\rho \sim 0.2$) the profiles flatten due to macroscopic MHD instabilities like sawteeth.[11] Simple linear regression as shown in Fig. 5 is used to estimate the temperature $\hat{T}_{\text{core}}$ at $\rho = 0.2$ from the temperature T_{edge} at $\rho = 0.8$. The edge and core temperatures are then connected by the unique curve of the form $T = ke^{-C\rho}$, and the core value is floated inward from $\rho = 0.2$ to $\rho = 0$. Throughout the paper, this model will be referred to as the empirical profile-consistency model. Explicitly

$$T(\rho) = \begin{cases} \hat{T}_{\text{core}} & \text{if } \rho \le \rho_{\text{core}} \\ \hat{T}_{\text{core}} e^{-\left(\frac{\rho-\rho_{\text{core}}}{\rho_{\text{edge}}-\rho_{\text{core}}}\right) ln\left(\frac{\hat{T}_{\text{core}}}{T_{\text{edge}}}\right)} & \text{if } \rho_{\text{core}} \le \rho \le \rho_{\text{edge}}, \end{cases} \tag{4}$$

where

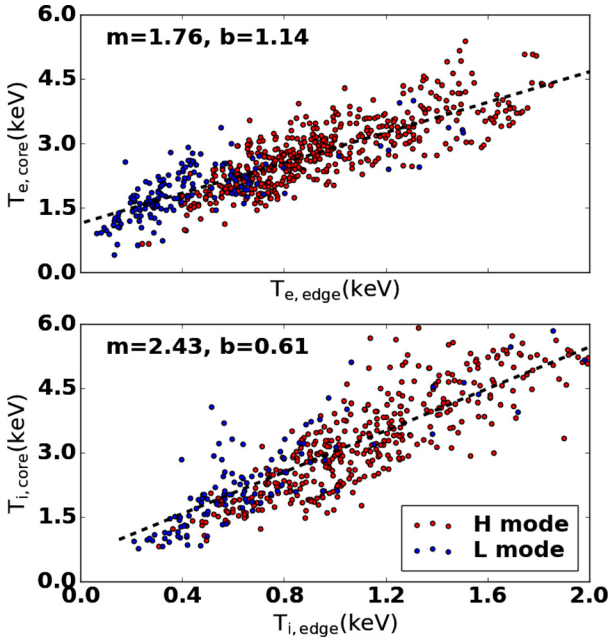

FIG. 5. Simple linear regression is used for each T_e and T_i to estimate the core ($\rho = 0.2$) value from the boundary ($\rho = 0.8$) value. This regression model is used in the profile-consistency empirical profile predictor for σ comparison.

$$\hat{T}_{\text{core}} = mT_{\text{edge}} + b, \tag{5}$$

and m and b are the slope and intercept from simple linear regression over the database of runs used for the study, as shown in Fig. 5. Figure 6 gives an illustration of this process.

The ASTRA and TRANSP modeling suites use experimental profile information from density, rotation, and safety factor profiles, in addition to the boundary temperature values. Meanwhile, this empirical baseline model uses only the boundary temperature values.

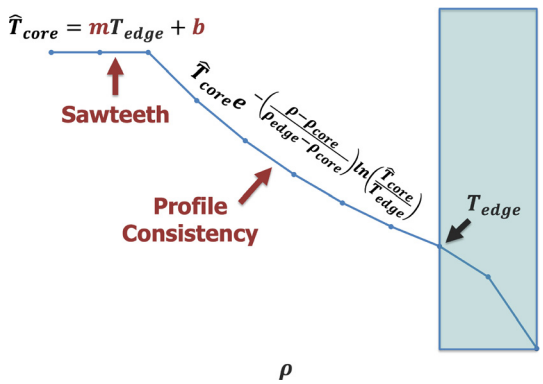

FIG. 6. Illustration of the profile-consistency empirical model used as a baseline for T_e and T_i predictions. As in the codes, the boundary T_{edge} is taken as input. The empirical model then computes $\hat{T}_{\text{core}}$ using the slope and intercept from the linear regression in Fig. 5, and this value is used between ρ of 0 and 0.2, loosely related to sawteeth flattening the core profile. The core and edge are then joined with an exponential, loosely related to profile consistency.

As a baseline for W_{mhd} predictions, the traditional ITER H_{98} (H-mode) and H_{89} (L-mode) scalings for confinement time are considered, which take the form

$$\tau_E = \alpha_0 I_p^{\alpha_1} B_t^{\alpha_2} \langle n_e \rangle^{\alpha_3} P^{\alpha_4} R_0^{\alpha_5} \kappa^{\alpha_6} a^{\alpha_7}, \quad (6)$$

for α_i log–log plot regressed fit parameters over an ITER database of discharges, I_p the plasma current, B_t the toroidal magnetic field, $\langle n_e \rangle$ the line-averaged density, P the total power absorbed, R_0 the major radius, κ the elongation, and a the minor radius.[26]

In this study's discharges, the heating is dominated by beam injected power P_{NBI} but with some usually net-positive contribution from other sources. As a crude estimate, it is therefore assumed the total power is just P_{NBI} plus a background power of $P = 0.55MW$ as a proxy for Ohmic power (ignoring radiated power), which is the average Ohmic power across the database. As a reminder, the two modeling suites use experimental profile information from density, rotation, and safety factor, along with the boundary values of temperature. Meanwhile, this empirical model takes in just the (mostly controllable) signals in Eq. (6).

Figure 7 shows the evolution over time at $\rho = 0.3$ by each model: TRANSP–NUBEAM (red), ASTRA–RABBIT (blue), the profile-consistency empirical model (green) for T_e and T_i, and the $H98$ (H-mode, dashed orange) and $H89$ (L-mode, dashed cyan) empirical models for W_{MHD}. In this discharge, the neutral beams (top) drop out at 3.5 to 3.88s, causing a marked drop then reestablishment of temperature and stored energy. Note that TRANSP–NUBEAM, ASTRA–RABBIT, and the profile consistency model all accurately predict the trend. Also note that while neutral beams are off, the charge exchange diagnostic needed for T_i profiles is unavailable (and these timesteps are excluded from the validation for T_i over the database).

For reference, the profiles corresponding to the predictions at 3.9s of the same discharge as discussed previously, just after the neutral beams and all diagnostics return, is shown in Fig. 8. Note that there is non-monotonicity in the TRANSP–NUBEAM simulation due to the rising boundary condition. The safety factor profile q is also plotted (bottom). Recall that for the full-database validation, only points outside of the $q = 1$ surface and inside the $\rho = 0.8$ simulation boundary are considered.

Note in Fig. 7 that the H-mode scaling naturally performs better in the H-mode regions (shaded orange), while the L-mode scaling performs better in the L-mode regions (shaded cyan). Across the validation database, the L-mode scaling is used as the empirical baseline for timesteps in discharges that are L-mode, and the H-mode scaling is used for H-mode. The baseline empirical model with this discrete choice will be called the "τ_E scaling" for the rest of this work. To manage this, each timepoint across the database is labeled as either H-mode or L-mode via an automated workflow implemented in OMFIT and described in Ref. 51. This H-mode detector is stated to be accurate to within about 30 ms of an L–H or H–L transition and applicable for a wide range of DIII-D scenarios. The error due to labeling is not considered in this study for its use in the empirical scaling, though future work considering this error would be of some value. Of the 116 discharges in the database, 53% are in H-mode for more than 90% of the shot, and 18% are in L-mode for more than 90% of the shot. Across all the discharges, 72% of the timeslices are H-mode.

B. Visual comparison of prediction errors

In Fig. 9 the figures of merit for L-mode vs H-mode cases are compared across the simulators and the corresponding empirical models (profile-consistency for T_e and T_i, and τ_E scaling for W_{MHD}). By visual inspection, note that there does not appear to be a significant

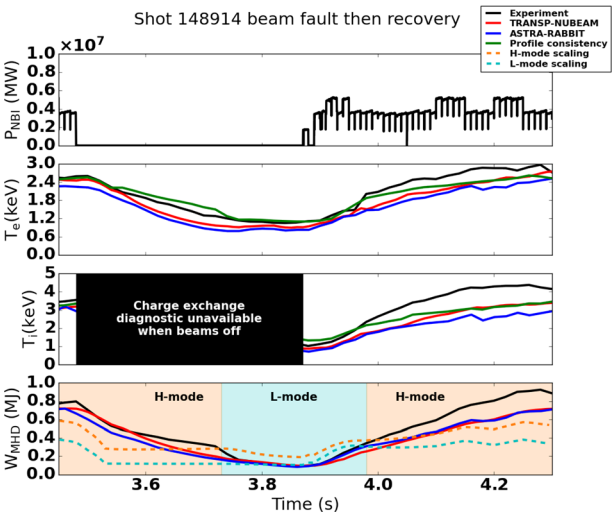

FIG. 7. Timetrace of simulators and baselines at $\rho = 0.3$ during a discharge wherein the neutral beams (top) drop out at 3.5 s through 3.88 s. Note that TRANSP–NUBEAM, ASTRA–RABBIT, and the profile consistency model give results within a few percent for this case in predicting T_e and T_i. The $H98$ (H-mode, dashed orange) and $H89$ (L-mode, dashed cyan) empirical models for W_{MHD} are also shown relative to simulator predictions, with regions of H- and L-mode shared.

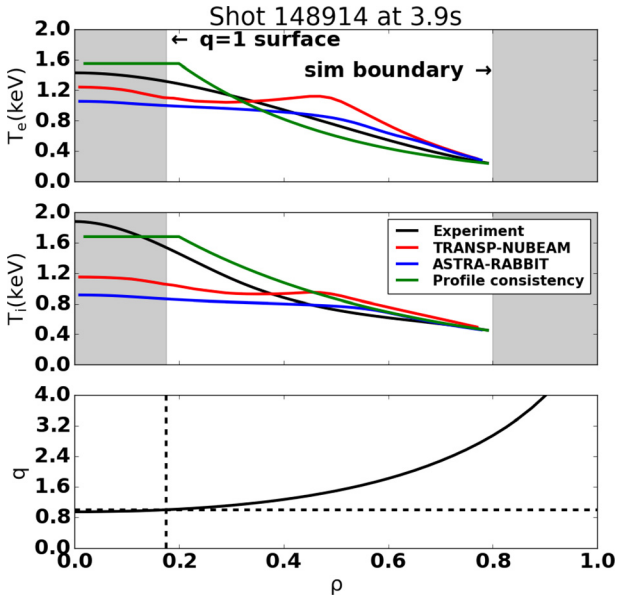

FIG. 8. Timeslice comparison of the various predictors (TRANSP–NUBEAM in red, ASTRA–RABBIT in blue, baseline profile consistency model in green) against experimental truth (black). The timetrace corresponding to $\rho = 0.3$ is shown in Fig. 7.

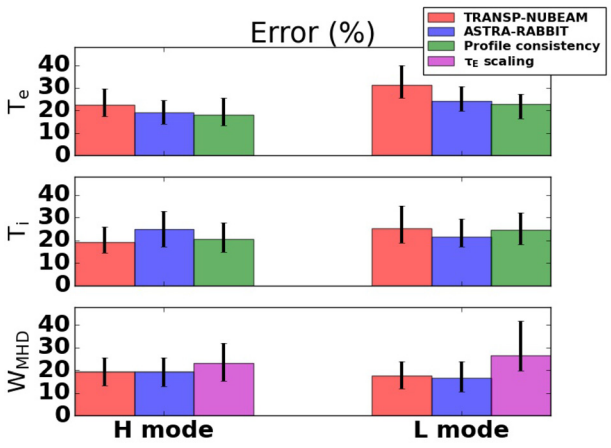

FIG. 9. Error by standard FIGS of merit ($\bar{\sigma}$ for electron and ion temperature profiles, $\Delta\bar{R}_W$ for W_{mhd}) for H and L mode timeslices in the database. The 25th to 75th percentiles of the distribution over timesteps are also shown (black lines). Performance of empirical baselines of comparison are also shown to the right of the TRANSP–NUBEAM (red) and ASTRA–RABBIT (blue) performance. For temperatures T_e and T_i, a two-parameter linear regression profile-consistency fit is employed. For stored energy W_{MHD}, a τ_E scaling is used.

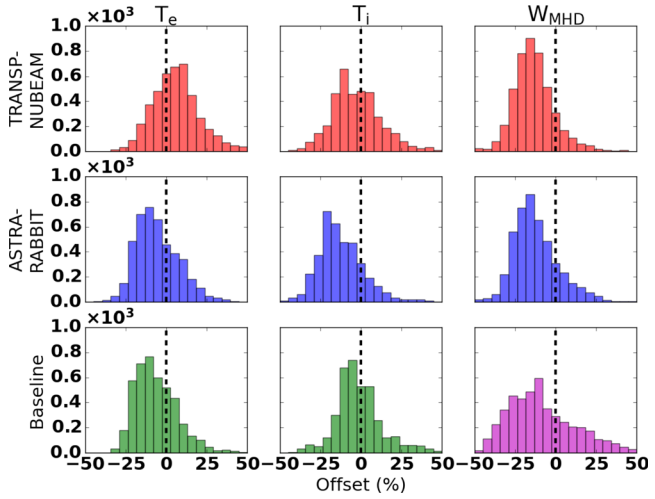

FIG. 10. Histogram of number of timeslices in database with offsets f [Eq. (1b)] from experimental values for the various simulators. Performance of an empirical model based on the ITER *H* scalings is shown for stored energy, and performance of the simple linear regression profile-consistency fit is shown for the electron and ion temperature profiles.

advantage of simulators over the profile consistency baseline for T_e and T_i predictions. On the other hand, the τ_E scaling for W_{MHD} appears to potentially perform significantly worse than the simulators.

C. Hypothesis test for advantage of simulation over baseline

For a more rigorous assessment of the performance of the simulators relative to baseline, a two-sample paired *t* test was conducted with a standard significance threshold of $p = 0.05$. Six hypotheses are made, one for each simulator (TRANSP–NUBEAM and ASTRA–RABBIT) and one for each signal (T_e, T_i, and W_{MHD}). In each case, the hypothesis is that the average simulator prediction error is lower than the baseline (i.e., the simulator performs better than baseline). For T_e and T_i, the baseline is the profile consistency model and the prediction error is considered as σ^2 (so that $\bar{\sigma}^2$ is the metric ultimately tested). For W_{MHD}, the baseline is the τ_E scaling and the prediction error is considered as ΔR_W^2 (so that $\Delta\bar{R}_W^2$ is the metric ultimately tested).

A t-test assumes that all samples are drawn independently. If each time step across discharges is considered as a sample, then this may yield significant dependence among samples since nearby points in time are obviously correlated. To be as conservative as possible, the hypothesis test instead considers each of the 116 *discharges* as the samples, and the σ^2 and ΔR_W^2 are considered as the average across timepoints within each discharge.

A t-test must also use approximately normally distributed data. The metrics here are squared values, due to the desire to compare error rather than simply offset. While offsets f and ΔR_W are fairly close to normally distributed (see Fig. 10), the squared metrics being tested are not (they tend to be highest concentrated toward 0 but are always non-negative). However, with enough samples, the central limit theorem implies a normality of the mean values which relaxes the need for an exactly normal distribution of the individual samples. To maintain as many samples as possible, therefore, the hypothesis tests are considered only across the full dataset, rather than for H- and L-mode subsets separately.

As intuited from Table III shows the *p*-values associated with the hypothesis tests. Using the significance threshold of $p = 0.05$, both simulators have a statistical advantage over the τ_E scaling baseline for W_{MHD}. However, there is not a significant advantage of either simulator over the profile consistency baseline for T_e and T_i.

D. Simulation offsets relative to experiment

A histogram is shown for the offset figure of merit (f for T_e and T_i, ΔR_W for W_{MHD}) over all relevant timeslices from all discharges in Fig. 10. It was shown in the cross-verification section that ASTRA–RABBIT systematically underpredicts all quantities relative to TRANSP–NUBEAM. Now it is shown that relative to experiment, T_e is slightly underpredicted by ASTRA–RABBIT while it is slightly overpredicted by TRANSP–NUBEAM. Meanwhile, there is very little bias in TRANSP–NUBEAM for predictions of T_i while for ASTRA–RABBIT there is a shortfall. Finally, both simulators systematically

TABLE III. *p*-Values corresponding to the two-sample paired *t* test for the hypothesis that the simulator outperforms the baseline for each signal. With a *p*-value threshold of 0.05, there is a statistically significant advantage of both TRANSP–NUBEAM and ASTRA–RABBIT over the τ_E scaling baseline for W_{MHD} predictions. However, there is no statistically significant advantage of either simulator over the profile consistency baseline for T_e and T_i predictions.

	TRANSP–NUBEAM	ASTRA–RABBIT
T_e	9.98×10^{-1}	9.57×10^{-1}
T_i	1.02×10^{-1}	7.48×10^{-1}
W_{MHD}	2.09×10^{-5}	1.87×10^{-5}

underpredict W_{MHD}. The profile-consistency model slightly underpredicts T_e while being fairly unbiased for T_i. Meanwhile, the τ_E scaling model used as a baseline for W_{MHD} not only underpredicts, but also has a much larger spread relative to ASTRA–RABBIT and TRANSP–NUBEAM.

E. Discussion

These results indicate that the error in ASTRA–RABBIT and TRANSP–NUBEAM are both of order 10%–20%, consistent with previous validation studies referenced in the introduction. They additionally show that a simple linear regression model with only two parameters can achieve no worse $\bar{\sigma}^2$ accuracy for predicting the full profiles. However, the two-fit model provides no insights into the scaling of predictions with other parameters (such as density gradient, the details of beam geometry, etc.). It is therefore of limited utility for extrapolating to new regimes or understanding trends.

Meanwhile, it was shown that the empirical W_{MHD} model based on τ_E scaling performed significantly worse than ASTRA–RABBIT and TRANSP–NUBEAM. However, the empirical model uses only controllable parameters whereas the ASTRA–RABBIT and TRANSP–NUBEAM runs require uncontrollable inputs (such as other profiles and the boundary conditions). The only implicit bleeding of information to the model from the boundary condition and other profiles is the H/L-mode label.

This work suggests maintaining a baseline is helpful to contextualizing the error in a model, since considering or comparing error metrics in isolation does little to understand the benefit of more complicated models over simple baselines. The specific baseline will be different for different applications of the transport models, and this study demonstrates just two examples. The profile consistency and τ_E scaling baselines are relevant in this case of core transport prediction. For use in designing future reactors in fundamentally different regimes, a different baseline may be chosen, such as a more holistic machine learning model (as in e.g., Ref. 52).

V. UNCERTAINTY QUANTIFICATION

The plasma profiles have experimental uncertainties and are relatively low fidelity. The following section includes a simple uncertainty quantification to argue that the details of the fits do not significantly impact the results of Sec. IV.

The plasma profile fits employed in this study contain errors both in the diagnostic measurements and in the fitting techniques used to interpolate the measurements onto a spatial ρ grid. It is important to contextualize both (1) the uncertainty in the ground truth measurements and (2) the sensitivity of the model results to the uncertainty in the model inputs.

For kinetic profiles, this study employs a standard fitting procedure from General Atomics (ZIPFIT[53]) The procedure bins measurements in a ± 20 ms time window, acquires multiple fits (splines and modified hyperbolic tangents), and returns the fit with the lowest χ^2 error. Electron temperature and density fits are based on Thomson measurements. Carbon impurity temperature, density, and rotation are measured by charge exchange spectroscopy. The main (deuterium) ions are assumed to have the same temperature and rotation, and to have the density needed to maintain quasineutrality assuming carbon is the only impurity. For safety factor q measurements, the study employs the EFIT01 equilibrium solution.

While ZIPFIT has the advantage of being readily available and robust for most discharges in the DIII-D database, the fits are relatively low-fidelity (unconstrained by other physics quantities, with unshifted pedestal, etc.) and also do not contain uncertainty measurements. In this section, fits from the Consistent Automatic Kinetic EFIT (CAKE)[51] are therefore employed to demonstrate uncertainty estimation and propagation through ASTRA in an example discharge with good CAKE fits: discharge 196114, with a focus on time step 2.6 s.

Unscented transform is employed, a technique for propagating probability distributions through arbitrary (potentially nonlinear) functions. The technique is similar to Monte Carlo, which samples many points from the input distribution to get broad coverage of the output distribution. However, unscented transform uses intelligently chosen sample points so that only $2n+1$ samples need to be taken, where n is the dimensionality of the input. This is important for the present case where each sample (simulation) requires about an hour of wall-clock time. Despite the lower computational cost, the unscented transform still provides strong mathematical guarantees for capturing variance due to higher order moments of the output distribution. For more details see Ref. 54.

For the present problem, variations between profiles are assumed to be independent. When using the most standard unscented transform heuristics in this simple case of a diagonal covariance matrix, the $2n+1$ input points to sample are simply the mean along with $\pm\sqrt{3}\sigma$ independently along each input dimension for σ the standard deviation. The mean point's output is weighted by $w_0 = 1 - \frac{n}{3}$ and the $\pm\sqrt{3}\sigma$ points' outputs are weighted by $w_i = \frac{1}{6}$ to calculate the output mean and variance as

$$\hat{\mu} = \sum_{i=0}^{2n+1} w_i y_i, \tag{7}$$

$$\hat{\sigma}^2 = \sum_{i=0}^{2n+1} w_i (y_i - \hat{\mu})^2, \tag{8}$$

for $\hat{\mu}$ the output distribution mean, $\hat{\sigma}$ the output standard deviation, and y_i the results of the simulation for each of the selected points $i = 0 \ldots 2n+1$. Note that the weights w_i sum to 1, which is maintained via a negative weight for the mean value when $n > 3$. This can yield an unphysical negative variance and reflects the cost of sampling linearly more points even as the number of points in the space grows exponentially with higher dimensionality. Per the curse of dimensionality, the number of varied parameters should therefore be kept as low as possible while still reasonably capturing sources of uncertainty.

With this in mind, the points within each profile are assumed to co-vary exactly, so that a single deviation in units of the local points' standard deviation is applied for every point. Similarly, the time-points are assumed to co-vary (whether a fully input profile or boundary conditions for the predicted T_e and T_i). This leaves six dimensions to vary: electron density, electron temperature, Z_{eff}, ion temperature, ion velocity, and safety factor q. CAKE provides uncertainty estimates for each of these profiles except q. For q, a somewhat arbitrary $\pm 10\%$ (applied independently at each point, so that the shear is also altered) is assumed based on the spread in multiple independent Grad–Shafranov solutions (from EFITs with different constraints alongside CAKE) being of this order.

The unscented transform methodology is used to estimate the mean and variance of the ASTRA-predicted T_e and T_i as well as σ at each time step for the example discharge. Figure 11 shows the results. The black curves show quantities over space at 2.6 s (left) and over time at $\rho = 0.3$ (right). The black curves correspond to the inputs to ASTRA and the ground truth against which ASTRA is compared in order to compute σ. The σ values over time are also shown in red. Throughout this particular discharge, the uncertainty in σ due to diagnostic and fitting uncertainty hovers around a few percent regardless of the magnitude of σ.

Table IV shows the average percentage uncertainty in each of the inputs in the left column at time step of 2.6 s shown in Fig. 11. The middle and right columns show the corresponding propagated percentage uncertainty through ASTRA as estimated by the unscented transform, for both T_e and T_i outputs. One can see that the input T_e and T_i have a significant effect on the corresponding output profiles. Safety factor q has a significant effect on both T_e and T_i in this case, as does density. Z_{eff} and velocity v only induce about a percent change in output. ASTRA orchestrates a highly nonlinear function (especially through TGLF's calculation of diffusivity), however, so these percentages will vary greatly based on the specific plasma scenario.

The primary figure of merit of interest reported for T_e and T_i in Sec. IV is $\bar{\sigma}$, the average over all timeslices and all shots of $\sigma = \frac{\sqrt{\frac{1}{N}\sum_{j=1}^{N}(\hat{T}_j - T_j)^2}}{\sqrt{\frac{1}{N}\sum_{j=1}^{N} T_j^2}}$. Figure 9 showed $\bar{\sigma}$ is of order 15%–25%. This sections demonstrated the diagnostics and profile fits only contributes on the order of 5% uncertainty to this estimate. On a more general note, because $\bar{\sigma}$ is an average over many shots and timeslices, by the law of large numbers this 5% uncertainty decreases approximately like $\frac{1}{\sqrt{N}}$ for N the number of independent samples. As discussed regarding hypothesis testing, one can most conservatively consider N to be the number of discharges in the database (116) in this study. To the extent that there is noise without bias, then, one can make the uncertainty due to diagnostics and fits arbitrarily small by using a sufficiently large database. With a fixed amount of human- and compute-time, then, we argue the methodology used in this paper of running many potentially inaccurate fits could provide a better estimate for simulator error than running just a few painstaking manual fits.

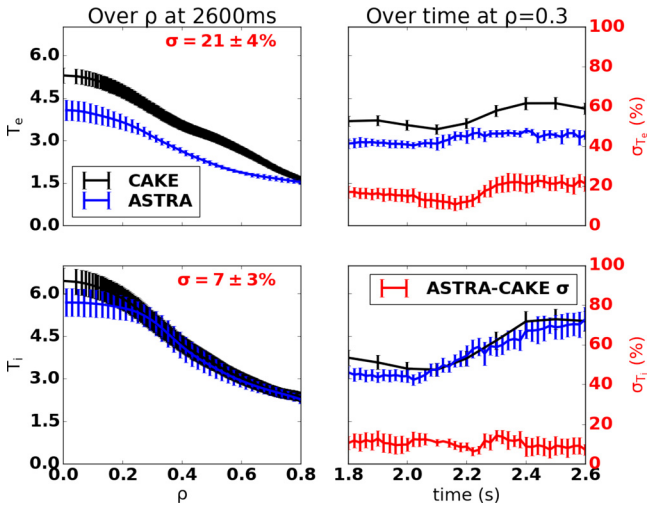

FIG. 11. Shot 196 114 using unscented transform to propagate uncertainties in electron density and safety factor q through the ASTRA simulator due to variations in the high-sensitivity inputs: electron and ion boundary temperature, safety factor q, and electron density (i.e., n = 4 variables, so 2n + 1 = 9 points in the unscented transform). The outputs are compared to the CAKE mean and uncertainty fits, demonstrating reasonable agreement for ion temperature predictions but 2σ discrepancy in electron temperature.

TABLE IV. CAKE percentage uncertainties for shot 196 114 at 2600 ms (averaged over all points within the profile up to the boundary at $\rho = 0.8$) due to variations in inputs one signal at a time (i.e., in each case with n = 1 signal, so 2n + 1 = 3 samples used in the unscented transform). The left hand side of Fig. 11 shows the corresponding time slice with point-by-point error bars.

	Input (%)	Output T_e (%)	Output T_i (%)
T_e	6.2	3.6	1.8
T_i	8.5	2.7	7.1
q	10	3.5	5.2
n_e	6.0	2.6	3.1
Z_{eff}	5.3	1.2	1.6
v	9.0	1.0	1.3

VI. CONCLUSION

For the first time, the most recent versions of TRANSP and ASTRA were cross-verified. The codes were shown to give solutions within 2% for core temperature predictions for a simple test case with identical auxiliary heat and diffusivity profiles. For time-dependent experimental cases modeling neutral beams and turbulent transport, ASTRA–RABBIT was shown to underpredict relative to TRANSP–NUBEAM by 10%–20%. Differences in calculations of the heat deposition and diffusivity are a small contribution to this error, perhaps leaving most of the difference to the solvers (fully self-consistent diffusivity calculations in TRANSP vs smoothed diffusivity in ASTRA). Figures were reported for the number of cases that failed to run, though due to the difficulty of debugging the codes the runs were not reattempted. Future work should further dissect the remaining difference in the solvers and the reasons for the run failures. In addition to more careful analysis of TRANSP and ASTRA, implementing another independent solver in a more modern computing language may also be valuable, such that it is easier to decipher and test where the fundamental differences lie. Using an automatically differentiable framework would also allow more comprehensive and natural uncertainty quantification, though implementing subroutines like TGLF and RABBIT in this framework will be a challenge.

The transport solvers, using different fast ion deposition models, were also validated against experiment. Crucially, the corresponding performance by simple empirical models were additionally included to contextualize the figures of merit. It was shown that a two-fit model based on the concept of profile consistency and core relaxation performs about as well as the more complicated one-dimensional transport solvers, though notably the two-fit model may not extrapolate well out of distribution. Total stored energy W_{MHD} is much more accurately predicted by TRANSP–NUBEAM and ASTRA–RABBIT

than by an estimate based on τ_E scaling, though notably the τ_E scaling predicts primarily from nominally controllable parameters while the TRANSP–NUBEAM and ASTRA–RABBIT runs require uncontrolled (i.e., not pre-specified) parameters as inputs. Work is forthcoming to validate transport solvers with empirical models on more equal footing: using fully-predictive transport solvers compared to empirical models (based on neural networks) which use the same inputs and outputs as the transport solvers; and testing performance both within distribution and for extrapolating out of distribution.

One primary concern in the validation portion of this work is that the error bar for the measurements used as the experimental truth are significant relative to the errors between the predictions and the truth. An uncertainty analysis was carried out for a specific discharge to demonstrate that uncertainty in the figure of merit σ due to fitting uncertainty is small for that single discharge, relative to the $\bar{\sigma}$ reported over the database. Furthermore, because of the large number of shots used in the validation, the law of large numbers implies noise in profile fits (as long as it is not systematically biased) is smoothed over so that the uncertainty due to fits over the full database is even lower. This also demonstrates a benefit, given fixed human- and compute-time, of using many potentially inaccurate fits as compared to a few painstaking manual fits. That being said, future work could repeat this methodology with higher-fidelity measurements without significant effort, e.g., using the Consistent Automatic Kinetic EFIT (CAKE) workflow[51] to achieve more reliable profiles and equilibria along with more systematic uncertainty analysis.

ACKNOWLEDGMENTS

Thanks to Greg Hammett for helpful discussions about the profile consistency baseline model, to Marina Gorelenkova for technical help with TRANSP runs, to Jeff Lestz for general feedback and discussions, and to Richard Nies and Thomas Foster for discussions on gyrokinetics. Thanks to the TRANSP and ASTRA teams for inexorable patience and scientific integrity.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Fusion Energy Sciences, using the DIII-D National Fusion Facility, a DOE Office of Science user facility, under Award(s) DE-FC02-04ER54698.

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Joseph Abbate: Data curation (lead); Formal analysis (lead); Methodology (lead). **Emiliano Fable:** Methodology (equal); Validation (equal). **Brian A. Grierson:** Validation (equal). **Alexei Y. Pankin:** Methodology (equal); Validation (equal). **Giovanni Tardini:** Methodology (equal); Validation (equal). **Egemen Kolemen:** Conceptualization (supporting).

DATA AVAILABILITY

Raw data were generated at the DIII-D large scale facility. Derived data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]G. Pereverzev and P. Yushmanov, "ASTRA automated system for TRansport analysis in a tokamak," IPP–5-98, 2002.
[2]T. Rafiq, A. H. Kritz, J. Weiland, A. Y. Pankin, and L. Luo, "Physics basis of multi-mode anomalous transport module," Phys. Plasmas **20**, 032506 (2013).
[3]C. Bourdelle, J. Citrin, B. Baiocchi, A. Casati, P. Cottier, X. Garbet, and F. Imbeaux, and JET Contributors, "Core turbulent transport in tokamak plasmas: Bridging theory and experiment with QuaLiKiz," Plasma Phys. Control. Fusion **58**, 014036 (2015).
[4]J. E. Kinsey, G. M. Staebler, and R. E. Waltz, "Simulations of internal transport barrier formation in tokamak discharges using the GLF23 transport model," Phys. Plasmas **9**, 1676–1691 (2002).
[5]J. E. Kinsey, G. M. Staebler, and R. E. Waltz, "The first transport code simulations using the trapped gyro-landau-fluid model," Phys. Plasmas **15**, 055908 (2008).
[6]C. S. Chang, "Effect of finite aspect ratio on the neoclassical ion thermal conductivity in the banana regime," Phys. Fluids **25**, 1493 (1982).
[7]W. A. Houlberg, K. C. Shaing, S. P. Hirshman, and M. C. Zarnstorff, "Bootstrap current and neoclassical transport in tokamaks of arbitrary collisionality and aspect ratio," Phys. Plasmas **4**, 3230–3242 (1997).
[8]E. A. Belli and J. Candy, "Kinetic calculation of neoclassical transport including self-consistent electron and impurity dynamics," Plasma Phys. Control. Fusion **50**, 095010 (2008).
[9]Z. Chang and J. Callen, "Global energy confinement degradation due to macroscopic phenomena in tokamaks," Nucl. Fusion **30**, 219–233 (1990).
[10]F. Turco, T. Petrie, T. Osborne, C. Petty, T. Luce, B. Grierson, T. Odstrcil, M. Van Zeeland, D. Liu, L. Casali, W. Boyes, S. Smith, H. Shen, M. Kostuk, and D. Brennan, "The physics basis to integrate an MHD stable, high-power hybrid scenario to a cool divertor for steady-state reactor operation," Nucl. Fusion **63**, 036020 (2023).
[11]B. B. Kadomtsev, "Disruptive instability in tokamaks," Sov. J. Plasma Phys. **1**, 389–391 (1975).
[12]S. E. Attenberger, W. A. Houlberg, and N. A. Uckan, "Transport analysis of ignited and current-driven ITER designs," Fusion Technol. **15**, 629–636 (2017).
[13]P. B. Snyder, H. R. Wilson, J. R. Ferron, L. L. Lao, A. W. Leonard, T. H. Osborne, A. D. Turnbull, D. Mossessian, M. Murakami, and X. Q. Xu, "Edge localized modes and the pedestal: A model based on coupled peeling–ballooning modes," Phys. Plasmas **9**, 2037–2043 (2002).
[14]P. B. Snyder, T. H. Osborne, K. H. Burrell, R. J. Groebner, A. W. Leonard, R. Nazikian, D. M. Orlov, O. Schmitz, M. R. Wade, and H. R. Wilson, "The EPED pedestal model and edge localized mode-suppressed regimes: Studies of quiescent h-mode and development of a model for edge localized mode suppression via resonant magnetic perturbations," Phys. Plasmas **19**, 056115 (2012).
[15]P. C. Stangeby, "Basic physical processes and reduced models for plasma detachment," Plasma Phys. Control. Fusion **60**, 044022 (2018).
[16]F. Felici, J. Citrin, A. Teplukhina, J. Redondo, C. Bourdelle, F. Imbeaux, O. Sauter, and JET Contributors, and the EUROfusion MST1 Team, "Real-time-capable prediction of temperature and density profiles in a tokamak using RAPTOR and a first-principle-based transport model," Nucl. Fusion **58**, 096006 (2018).

[17]G. Brochard, J. Bao, C. Liu, N. Gorelenkov, G. Choi, G. Dong, P. Liu, J. Mc. Clenaghan, J. Nicolau, F. Wang, W. Wang, X. Wei, W. Zhang, W. Heidbrink, J. Graves, Z. Lin, and H. Lütjens, "Verification and validation of linear gyrokinetic and kinetic-MHD simulations for internal kink instability in DIII-d tokamak," Nucl. Fusion **62**, 036021 (2022).
[18]R. Hawryluk, "An empirical approach to tokamak transport," in *Physics of Plasmas Close to Thermonuclear Conditions*, edited by B. Coppi *et al.* (CEC, Brussels, 1981), pp. 19–46.
[19]G. Pereverzev, P. Yushmanov, A. Dnestrovskii, A. Polevoi, K. Tarasjan, and L. Zakharov, "ASTRA - an automatic system for transport analysis in a tokamak," IPP–5/42, 2002.
[20]E. Fable, C. Angioni, A. A. Ivanov, K. Lackner, O. Maj, S. Y. Medvedev, G. Pautasso, G. V. Pereverzev, W. Treutterer, and the ASDEX Upgrade Team. "Dynamical coupling between magnetic equilibrium and transport in tokamak scenario modelling, with application to current ramps," Plasma Phys. Control. Fusion **55**, 074007 (2013).
[21]G. Cenacchi and A. Taroni, "Jetto a free boundary plasma transport code," ENEA-RT-TIB–88-5, 1988.
[22]J. Candy, C. Holland, R. E. Waltz, M. R. Fahey, and E. Belli, "Tokamak profile prediction using direct gyrokinetic and neoclassical simulation," Phys. Plasmas **16**, 060704 (2009).
[23]F. Felici and O. Sauter, "Non-linear model-based optimization of actuator trajectories for tokamak plasma profile control," Plasma Phys. Control. Fusion **54**, 025002 (2012).
[24]Y. Ou, T. Luce, E. Schuster, J. Ferron, M. Walker, C. Xu, and D. Humphreys, "Towards model-based current profile control at DIII-D," Fusion Eng. Design **82**, 1153–1160 (2007).
[25]P. Rodriguez-Fernandez, N. T. Howard, M. J. Greenwald, A. J. Creely, J. W. Hughes, J. C. Wright, C. Holland, Y. Lin, F. Sciortino, and the SPARC Team, "Predictions of core plasma performance for the SPARC tokamak," J Plasma Phys. **86**, 865860503 (2020).
[26]ITER Physics Expert Group on Confinement and Transport, "Chapter 2: Plasma confinement and transport," Nucl. Fusion **39**, 2175 (1999).
[27]P. W. Terry, M. Greenwald, J.-N. Leboeuf, G. R. McKee, D. R. Mikkelsen, W. M. Nevins, D. E. Newman, D. P. Stotler, and Task Group on Verification and Validation, U.S. Burning Plasma Organization, and U.S. Transport Task Force, "Validation in fusion research: Towards guidelines and best practices," Plasma Phys. **15**, 062503 (2008).
[28]A. Y. Pankin, A. Pletzer, S. Vadlamani, J. R. Cary, A. Hakim, S. E. Kruger, M. Miah, T. D. Rognlien, S. Shasharina, G. Bateman, A. H. Kritz, and T. Rafiq, "Simulation of anomalous transport in tokamaks using the FACETS code," Comput. Phys. Commun. **182**, 180–184 (2011).
[29]J. Kinsey, G. Staebler, J. Candy, R. Waltz, and R. Budny, "ITER predictions using the GYRO verified and experimentally validated trapped gyro-Landau fluid transport model," Nucl. Fusion **51**, 083001 (2011).
[30]C. Holland, "Validation metrics for turbulent plasma transport," Phys. Plasmas **23**, 060901 (2016).
[31]M. Greenwald, "Verification and validation for magnetic fusion," Phys. Plasmas **17**, 058101 (2010).
[32]A. Y. Pankin, S. E. Kruger, R. J. Groebner, A. Hakim, A. H. Kritz, and T. Rafiq, "Validation of transport models using additive flux minimization technique," Phys. Plasmas **20**, 102501 (2013).
[33]J. E. Kinsey, G. M. Staebler, and C. C. Petty, "Trapped gyro-Landau-fluid transport modeling of DIII-D hybrid discharges," Phys. Plasmas **17**, 122315 (2010).
[34]D. Hannum, G. Bateman, J. Kinsey, A. H. Kritz, T. Onjun, and A. Pankin, "Comparison of high-mode predictive simulations using mixed bohm/gyro-bohm and multi-mode (MMM95) transport models," Phys. Plasmas **8**, 964–974 (2001).
[35]T. Onjun, G. Bateman, A. H. Kritz, and D. Hannum, "Comparison of low confinement mode transport simulations using the mixed Bohm/gyro-Bohm and the multi-mode-95 transport model," Phys. Plasmas **8**, 975–985 (2001).
[36]J. Citrin, C. Bourdelle, F. J. Casson, C. Angioni, N. Bonanomi, Y. Camenen, X. Garbet, L. Garzotti, T. Görler, O. Gürcan, F. Koechl, F. Imbeaux, O. Linder, K. Van De Plassche, P. Strand, and G. Szepesi, "Tractable flux-driven temperature, density, and rotation profile evolution with the quasilinear gyrokinetic transport model QuaLiKiz," Plasma Phys. Control. Fusion **59**, 124005 (2017).
[37]O. Meneghini, S. P. Smith, L. L. Lao, O. Izacard, Q. Ren, J. M. Park, J. Candy, Z. Wang, C. J. Luna, and V. A. Izzo, "Integrated modeling applications for tokamak experiments with OMFIT," Nucl. Fusion **55**, 083008 (2015).
[38]B. A. Grierson, X. Yuan, M. Gorelenkova, S. Kaye, N. C. Logan, O. Meneghini, S. R. Haskey, J. Buchanan, M. Fitzgerald, S. P. Smith, L. Cui, R. V. Budny, and F. M. Poli, "Orchestrating TRANSP simulations for interpretative and predictive tokamak modeling with OMFIT," Fusion Sci. Technol. **74**, 101–115 (2018).
[39]L. L. Lao, H. E. St John, Q. Peng, J. R. Ferron, E. J. Strait, T. S. Taylor, W. H. Meyer, C. Zhang, and K. I. You, "MHD equilibrium reconstruction in the DIII-D tokamak," Fusion Sci. Technol. **48**, 968–977 (2005).
[40]S. Y. Medvedev, "New adaptive grid plasma evolution code SPIDER," in 32nd EPS Conference on Plasma Physics, Tarragona, 27 June – 1 July 2005.
[41]G. Staebler, E. A. Belli, J. Candy, J. Kinsey, H. Dudding, and B. Patel, "Verification of a quasi-linear model for gyrokinetic turbulent transport," Nucl. Fusion **61**, 116007 (2021).
[42]R. Goldston, D. McCune, H. Towner, S. Davis, R. Hawryluk, and G. Schmidt, "New techniques for calculating heat and particle source rates due to neutral beam injection in axisymmetric tokamaks," J. Comput. Phys. **43**, 61–78 (1981).
[43]M. Weiland, R. Bilato, R. Dux, B. Geiger, A. Lebschy, F. Felici, R. Fischer, D. Rittich, M. van Zeeland, and the ASDEX Upgrade Team, and the Eurofusion MST1 Team, "RABBIT: Real-time simulation of the NBI fast-ion distribution," Nucl. Fusion **58**, 082032 (2018).
[44]O. Sauter, C. Angioni, and Y. R. Lin-Liu, "Neoclassical conductivity and bootstrap current formulas for general axisymmetric equilibria and arbitrary collisionality regime," Phys. Plasmas **6**, 2834–2839 (1999).
[45]A. A. Galeev and R. Z. Sagdeev, "Neoclassical theory of diffusion," in *Advances in Plasma Physics* (Inst. of High Temperature, Moscow, 1976), Vol. 6, pp. 311–420.
[46]G. Pereverzev and G. Corrigan, "Stable numeric scheme for diffusion equation with a stiff transport," Comput. Phys. Commun. **179**, 579–585 (2008).
[47]S. C. Jardin, G. Bateman, G. W. Hammett, and L. P. Ku, "On 1d diffusion problems with a gradient-dependent diffusion coefficient," J. Comput. Phys. **227**, 8769–8775 (2008).
[48]T. I. D. M. W. Group, D. Boucher, J. Connor, W. Houlberg, M. Turner, G. Bracco, A. Chudnovskiy, J. Cordey, M. Greenwald, G. Hoang, G. Hogeweij, S. Kaye, J. Kinsey, D. Mikkelsen, J. Ongena, D. Schissel, H. Shirai, J. Stober, P. Stubberfield, R. Waltz, and J. Weiland, "The international multi-tokamak profile database," Nucl. Fusion **40**, 1955–1981 (2000).
[49]B. Coppi, "Profile consistency: Global and nonlinear transport," Phys. Lett. A **128**, 193–197 (1988).
[50]M. Kotschenreuther, W. Dorland, Q. Liu, G. Hammett, M. Beer, S. Smith, A. Bondeson, and S. Cowley, *First Principles Calculations of Tokamak Energy Transport* (IAEA, 1997).
[51]Z. Xing, D. Eldon, A. Nelson, M. Roelofs, W. Eggert, O. Izacard, A. Glasser, N. Logan, O. Meneghini, S. Smith, R. Nazikian, and E. Kolemen, "CAKE: Consistent automatic kinetic equilibrium reconstruction," Fusion Eng. Design **163**, 112163 (2021).
[52]J. Abbate, R. Conlin, and E. Kolemen, "Data-driven profile prediction for DIII-D," Nucl. Fusion **61**, 046027 (2021).
[53]S. Smith, "Recent improvements to automatic profile fitting (ZIPFIT)," (2010).
[54]S. Julier, J. Uhlmann, and H. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," IEEE Trans. Autom. Control **45**, 477–482 (2000).