OPEN ACCESS
IOP Publishing | International Atomic Energy Agency
Nuclear Fusion
Nucl. Fusion **64** (2024) 096039 (18pp)
https://doi.org/10.1088/1741-4326/ad645d

Real-time capable modeling of ICRF heating on NSTX and WEST via machine learning approaches

Á. Sánchez-Villar[1,*], Z. Bai[2], N. Bertelli[1], E.W. Bethel[3], J. Hillairet[4], T. Perciano[2], S. Shiraiwa[1], G.M. Wallace[5] and J.C. Wright[5]

[1] Princeton Plasma Physics Laboratory, Princeton, NJ 08540, United States of America
[2] Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States of America
[3] San Francisco State University, San Francisco, CA 94132, United States of America
[4] CEA, IRFM, F-13108 Saint-Paul-Lez-Durance, France
[5] MIT Plasma Science and Fusion Center, Cambridge, MA 02139, United States of America

E-mail: asvillar@pppl.gov

Received 28 November 2023, revised 13 June 2024
Accepted for publication 17 July 2024
Published 12 August 2024

Abstract

A real-time capable core Ion Cyclotron Range of Frequencies (ICRF) heating model on NSTX and WEST is developed. The model is based on two nonlinear regression algorithms, the random forest ensemble of decision trees and the multilayer perceptron neural network. The algorithms are trained on TORIC ICRF spectrum solver simulations of the expected flat-top operation scenarios in NSTX and WEST assuming Maxwellian plasmas. The surrogate models are shown to successfully capture the multi-species core ICRF power absorption predicted by the original model for the high harmonic fast wave and the ion cyclotron minority heating schemes while reducing the computational time by six orders of magnitude. Although these models can be expanded, the achieved regression scoring, computational efficiency and increased model robustness suggest these strategies can be implemented into integrated modeling frameworks for real-time control applications.

Keywords: machine learning, surrogate modeling, data-driven, neural networks, ICRF, plasma heating, tokamak

(Some figures may appear in colour only in the online journal)

1. Introduction

Radio-frequency (RF) wave-heating systems are one of the major auxiliary heating systems in magnetic confinement fusion. Amongst the multiple RF-heating systems available is the heating via fast magnetosonic wave (FW) excitation and damping in the Ion Cyclotron Range of Frequencies (ICRF) [1]. ICRF heating plays a crucial role determining tokamak operational performance and stability and, as a consequence, ICRF actuators are part of the design of multiple tokamaks as JET [2, 3], EAST [4, 5], WEST [6, 7], AUG [8, 9], NSTX/NSTX-U [10] and will also be part of SPARC [11] and ITER [1].

The evolution of tokamak plasma properties is governed by physical phenomena that can be modeled through a set of non-linear coupled differential equations. As direct first principle numerical simulations are ambitious, the current paradigm

* Author to whom any correspondence should be addressed.

is the development of integrated modeling frameworks as for example TRANSP [12], CORSICA [13], ASTRA [14], CRONOS [15], TOPICS [16], JINTRAC [17], and METIS [18]. These codes employ a modular structure including different models of magnetic equilibria, transport and diffusion coefficients, heating and current drive via RF or neutral beam injection, with flexibility to select a trade-off between physics-fidelity and computational time. These models have been successfully applied to analyze and interpret experimental measurements, for specific scenario optimization and, some of them, for inter-shot predictive modeling.

The aforementioned models are still computationally expensive for real-time control and optimization applications. For that purpose, with the advent of machine learning (ML), the development of fast physics-based surrogate models (e.g. neural networks, random forests, etc) has been identified as a Priority Research Opportunity for model reduction and extraction, being these models capable to bridge the existing gap in computational time scales [19]. Some control-oriented integrated modeling tools that have adopted this methodology are RAPTOR [20] and COTSIM [21, 22]. RAPTOR has integrated QLKNN-4Dkin [23], a multilayer perceptron (MLP) [24] based surrogate model predicting turbulent transport fluxes obtained by the QuaLiKiz code in a 4D input parametric space. COTSIM integrates a neural network surrogate of the multimode anomalous transport model MMM [25]. More recently, a JINTRAC-QLKNN framework has been implemented with an extended version of the original QLKNN surrogate [26]. Other relevant surrogate development efforts include fusion applications as disruption prediction and avoidance [27], self-consistent core pedestal transport estimation without [28] and with impurities [29], equilibrium reconstruction [30], and neutral beam injection [31].

Regarding RF heating and current drive surrogate models, recently it was shown that surrogate model predictions of the lower hybrid current drive based on GENRAY-CQL3D [32, 33] are able to accurately reproduce the current drive profiles with $\mathcal{O}$(ms) average inference times [34, 35], using MLP regressors, random forest regressors (RFR) [36, 37] and Gaussian processes regressors (GPR) [38]. While all ML-algorithms achieved similar regression accuracy after training, GPRs required higher inference times (t_{I}) and significantly greater training times (t_{tr}).

In this paper we present a natural continuation of that work, where we applied a similar methodology (section 2) to develop a set of surrogate models of the core ICRF power absorption based on the most adequate performing non-linear multivariate regressions models identified in [34], the MLP and RFR. Predicting ICRF heating physics features increased complexity as both electron and multiple ion species are contributing to absorption through multi-scale plasma-wave interactions, including electron Landau damping (ELD) of the ion Bernstein wave (IBW) and FW and cyclotron damping both at the fundamental and second harmonic in IC minority heating or high harmonic cyclotron absorption. Specifically we show that the presented methodology can be applied to consistently develop surrogate models that capture the multi-species core ICRF power absorption predicted by the ICRF spectrum solver TORIC [39, 40], currently implemented in TRANSP [12, 41]. Rather than aiming to capture the entire predictive range of TORIC, which would be impractical, we here aim to obtain optimal regression accuracy and computational efficiency for specific ICRF heating scenarios. In section 3, we demonstrate the development of ML-based surrogate models that successfully capture the physics of two fusion-relevant ICRF heating schemes: the high-harmonic fast heating (HHFW) [42] at NSTX and the ion cyclotron (IC) minority heating [43, 44] at WEST. Further details on surrogate implementation and validation is shown in appendix. Section 4 describes the nature of the outliers found in the generated databases for both NSTX and WEST, showing the impact of removing the outliers driven by numerical instabilities from surrogate training. Furthermore, we demonstrate the application of surrogate models to overcome possible challenges observed in the original model when providing ICRF heating estimates in outlier parametric regimes. In section 5, we conclude and provide insight on future directives and applications of this work.

2. ICRF heating model and datasets generation

TORIC is a full-wave ICRF semispectral solver that calculates the wave electric field in arbitrary axisymmetric toroidal plasmas [39]. The model is based on the finite Larmor radius approximation and encompasses the physics of fast compressional and torsional Alfvén, as well as of IBW which are excited through linear mode conversion in the vicinity of ion-ion hybrid resonances. The code considers absorption via ion damping at the fundamental and first harmonic cyclotron regions, and also ELD and transit time damping. Additionally, the code is capable of dealing with FW propagation and absorption via high-harmonic cyclotron damping through the quasi-local approximation [40]. This TORIC version is part of the integrated physics modeling suite TRANSP/PTRANSP [12, 41], which has been employed for multiple integrated modeling analyses, including ITER [45].

2.1. Feature generation and data collection

Two databases of 12 240 and 10 000 TORIC simulations are generated for the HHFW heating in NSTX and the IC minority heating in WEST, respectively, both in flat-top tokamak operation. The code is run assuming Maxwellian plasmas. The radial profile of the plasma properties as the density (n) and temperature (T) are assumed to follow

$$T_e(\rho) = T_{e1} + (T_{e0} - T_{e1})\left(1 - \rho^{\alpha}\right)^{\beta}, \tag{1}$$

where subscripts 0 and 1 represent core and edge locations, $\rho \in [0, 1]$ is the squared root of the normalized poloidal flux function, the subscripts e and i stand for the electron and ion species, respectively, and α and β are the profile exponents. Profile exponents are assumed to be equal for both density and temperature profiles.

Table 1 details the ranges of key input parameters selected based on the expected values identified in experimental campaigns for flat-top operation of NSTX [46] and WEST [47]. To

Table 1. Summary of input parameters for the Maxwellian TORIC databases used in this work for both NSTX and WEST scenarios. The parameter spans (within squared brackets) define lower and upper limits for the selection of each parameter via Latin hypercube sampling. N_φ is the toroidal mode number, n_{e0} and n_{e1} are core and edge electron densities, T_{e0} and T_{i0} are the electron and ion core temperatures, α is the profile inner exponent (1), and X_H is the minority concentration.

Input variable	N_φ [–]	n_{e0} [$10^{19}\,\mathrm{m}^{-3}$]	n_{e1} [$10^{19}\,\mathrm{m}^{-3}$]	T_{e0} [keV]	T_{i0} [keV]	α [–]	X_H [%]
NSTX	[5 – 21]	[5 – 20]	0.5	[1 – 5]	[1 – 5]	[2 – 10]	0
WEST	[0 – 60]	[4 – 8]	[0.1 – 0.3]	[1 – 1.5]	[1 – 1.5]	[2 – 10]	[1 – 10]

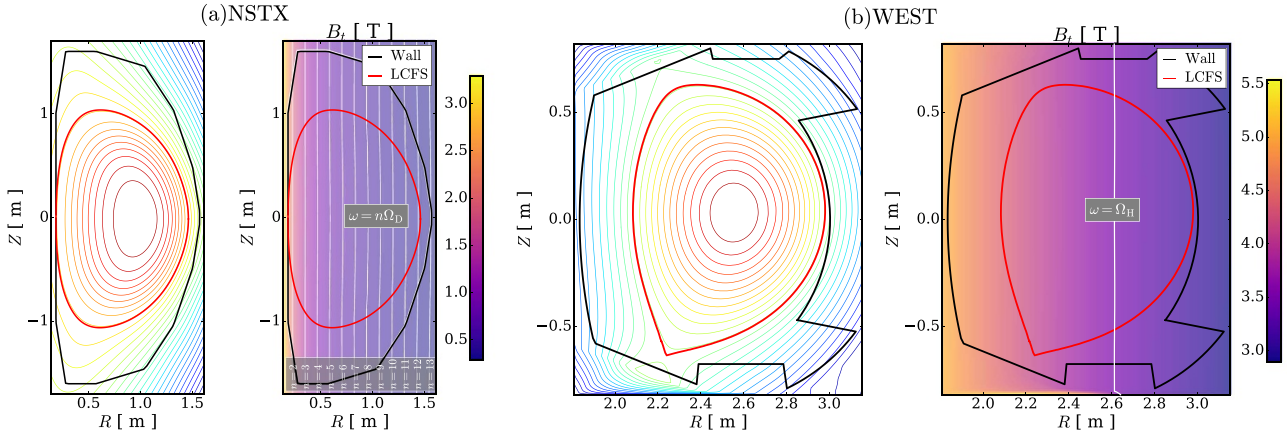

Figure 1. Equilibrium configurations used for (*a*) NSTX and (*b*) WEST databases, corresponding to shots `138 506` and `56 898`, respectively. On the left, the magnetic equilibrium is shown with highlighted LCFS (red) and plasma facing surface wall (black). On the right, the toroidal magnetic field amplitude B_t (colorbar) is shown, as well as the relevant resonance layers (white) in (*a*) NSTX (i.e. deuterium harmonic resonances, $n > 1$) and (*b*) WEST (i.e. fundamental resonance for hydrogen).

better fit the experimental profiles at the edge while minimizing the number of free parameters, we set $\beta = 3$ and $T_{e1} = T_{i1}$, being $T_{e1} = 300$ (200) eV for NSTX (WEST). For NSTX the key input parameters identified are the core electron density (n_{e0}) and temperature (T_{e0}), the toroidal mode number (N_φ), and the inner profile exponent (α). We assume that the electron density at the edge (n_{e1}) is fixed (see table 1) and $T_e = T_i$. For WEST we relax these two assumptions to test the sensitivity of the surrogates to these parameters: both ions (thermalized) and electrons featuring different temperatures and allowing the electron density to vary slightly in the edge. Additionally, we include hydrogen minority concentration X_H as it is a parameter highly impacting IC minority heating. It is worth noting that this choice is limiting the applicability of the surrogate model to capture the impact of impurities and/or the physics of different heating schemes such as the three-ion heating scheme [48]. However, the hydrogen concentration represents the key parameter determining the physics of ICRF heating in WEST experiments. As a result, NSTX (WEST) inputs are specified by sets of 4(7) inputs, these being $\{N_\varphi, n_{e0}, T_{e0}, \alpha\}$ ($\{N_\varphi, n_{e0}, n_{e1}, T_{e0}, T_{i0}, \alpha, X_\mathrm{H}\}$).

The toroidal mode number (N_φ) span is determined according to the spectrum of the antenna excitation for both NSTX [49] and WEST [6]. The plasma mixture of each device is composed by both deuterium (D) and electrons (e), while WEST includes also a minority species of hydrogen (H). The equilibrium configurations assumed for all NSTX and WEST simulations are shown in figure 1, corresponding to shots `138 506` [46] and `56 898` [50], respectively. TORIC simulation domain is the core region, delimited by the last closed surface (LCFS) (red curves in figure 1), and for sake of clarity we show the device wall (black). For each scenario we show both the equilibrium flux surfaces (left) and the toroidal magnetic field intensity (right). The resonance locations given by condition $\omega = n\Omega_s$ (ignoring the Doppler shift) are shown for deuterium at NSTX (see figure 1(*a*)) and hydrogen in WEST (see figure 1(*b*)), where $\omega = 2\pi f$ is the excitation frequency ($f = 30$ MHz in NSTX and $f = 56$ MHz in WEST), n is the harmonic number, and Ω_s is the cyclotron frequency of species s. While in WEST only the fundamental hydrogen resonance ($n = 1$) is present (and the second harmonic for deuterium, $n = 2$), in NSTX a set of deuterium harmonic resonances $n = [2, 12], n \in \mathbb{N}$ can be observed inside the LCFS.

The TORIC outputs investigated here are the radial power absorption density profiles. In NSTX we show models for electron heating (P_e) and deuterium heating (P_D), and in WEST, we focus on electrons and hydrogen (P_H), which is the ion species with dominant absorption. As it is impractical to perform mesh tuning for each case, here we generate the database using a fixed mesh resolution. More specifically, the radial resolution for NSTX is 500 nodes, higher than that of WEST (i.e. 320) to better characterize the multi-harmonic absorption

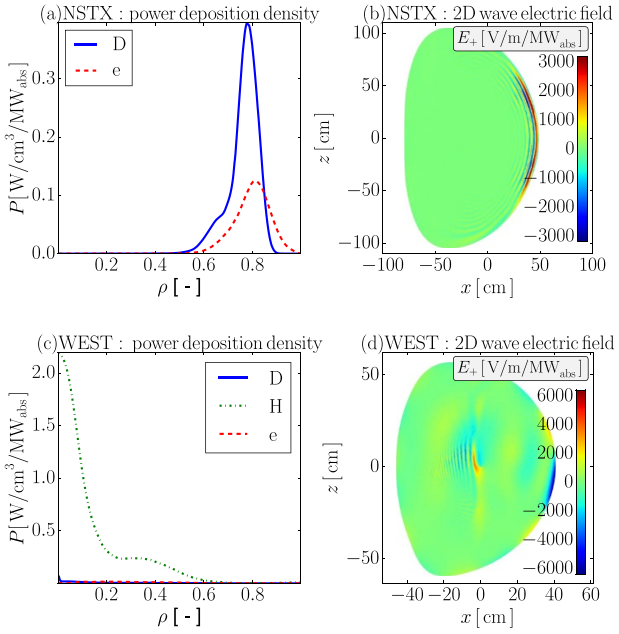

Figure 2. Examples of TORIC solutions for the typical flat-top scenarios of NSTX (*a*) and (*b*) and WEST (*c*) and (*d*). One dimensional power density profiles are shown for deuterium (blue solid), hydrogen minority (green dash–dotted) and electrons (red dashed). Figures (*b*) and (*d*) show the associated left-handed polarized 2D wave electric fields (E_+).

present at HHFW in NSTX. In the poloidal direction, the mesh resolution is chosen to be equal to 128 for both cases, keeping a sufficient number of poloidal modes in the solution. Some additional remarks on TORIC resolution parameters are discussed in section 4.

Figure 2 shows an example of TORIC solutions for the power deposition profiles (*a*) and (*c*) and the wave electric fields (*b*) and (*d*) in the typical plasma scenarios of both NSTX and WEST. In NSTX, the power deposition is characterized by heating peaked off-axis due to strong single pass absorption. In particular, deuterium absorption occurs at about $n = 8 - 11$ harmonic resonances (see figure 1(*a*)) and electron absorption via Landau damping. In the case of the WEST database, the power absorption is more intricate as three species (H, D, and *e*) simultaneously compete. The expected scenario, and the most typical, is a hydrogen minority dominated absorption through cyclotron damping at the fundamental resonance, which is located close to the magnetic axis in the WEST tokamak (see figure 1(*b*)). The resulting dominant hydrogen power absorption profile is peaked at or close to $\rho \sim 0$ (see figure 2(*c*)). Figure 2(*d*) shows the corresponding wave propagation and the FW-IBW mode conversion.

On the technical side, data collection is carried out by the development of a framework for codes featuring input parametric sets and outputs in both ASCII or NetCDF formats. Automated routines deliver and launch the batch of cases to the super-computing workstation. TORIC simulations are computed at Cori, a multi-core scientific supercomputer managed by the National Energy Research Computing Center (NERSC). In producing a suitable database of TORIC output representative regimes of interest, a Latin hypercube sampling of TORIC inputs enables pseudo-random sampling across the input multidimensional parametric space [51]. This method stratifies each marginal distribution function corresponding to each input parameter to reduce the sampling statistical bias over high-dimensional parameter spaces, that would be induced for instance by selecting a gridded sampling method.

2.2. Data processing

A crucial feature in the implementation of a surrogate model is the form of the dataset used for training. Adequate data processing can provide significant boost to surrogate scoring. While RFRs do not require data normalization or standardization, neutral networks are sensitive to both the features and targets scales. Therefore, inputs and outputs have been independently standardized using training data, so that, for instance, standardized input is computed as $z = (x - \mu)/\sigma$, being x the input value, and μ and σ its respective mean and standard deviation in the training dataset. Standardized data is used for both surrogate training and testing purposes. Inverse transformation is required to compare the obtained standardized predictions to original data. Standardization processes are observed to be more important for the WEST scenario as the data features higher variance in the profile shapes and scale. In the case of P_e, the maximum and mean profile values are 0.236 and 0.005 W m^{-3} per MW of absorbed power (MW$_{\mathrm{abs}}$), respectively, indicating the presence of high variance in profiles scale as the mean is roughly 2% of the maximum local value.

The surrogate models developed for WEST scenario suggested required the use of extra data processing to achieve satisfactory surrogate performance. Application of principal component analysis (PCA) [52] to WEST dataset after standardization allowed a dimensionality reduction of the output data, from 287 nodes in the radial profiles to only 10(27) coefficients to reconstruct the electron (hydrogen) power absorption profiles. PCA was selected opposed to the use of a coarser simulation grid to preserve finer features within the power deposition profiles that would not be resolved with a coarser mesh. We set the PCA decompositions to preserve a 99% of the explained variance of the training dataset output. As it is well known, dimensionality reduction can lead to simplification of the network structure [53] and improvement of surrogate scoring, specially to the MLP regressor results. As a result, not only we achieved a reduced computational time, but also an improved surrogate regression accuracy.

In the achievement of the surrogate models described in section 3, an exploratory analysis of each database is carried out. While in WEST scenario we employed the entire dataset for training and testing the surrogate models, in the case of NSTX scenario we removed a portion of the HHFW dataset featuring numerical instabilities. A detailed analysis on the nature of the cases removed, their impact in surrogate training, and a proposed solution for ML-based predictions in the outlier regime are discussed in section 4.

Table 2. TORIC-ML surrogate scoring metrics for each species, dataset, and ML-method. The coefficient of determination and the mean squared error (MSE) are shown for both training ($R^2_{\rm tr}$, $\rm MSE_{tr}$) and testing (R^2, MSE) data. Both surrogate training time ($t_{\rm tr}$) and average profile inference time ($t_{\rm I}$) are shown for one single thread at a CPU Perlmutter computing node.

Target	Dataset	Method	$R^2_{\rm tr}$	$\rm MSE_{tr}$	$t_{\rm tr}$ [s]	R^2	MSE	$t_{\rm I}$ [μs]
$P_{\rm e}$	NSTX	RFR	0.99	3.3×10^{-6}	29	0.97	1.8×10^{-5}	49
$P_{\rm e}$	NSTX	MLP	0.97	2.1×10^{-5}	23	0.97	2.1×10^{-5}	2
$P_{\rm D}$	NSTX	RFR	0.99	1.5×10^{-5}	29	0.90	1.1×10^{-4}	54
$P_{\rm D}$	NSTX	MLP	0.96	4.7×10^{-5}	48	0.94	5.2×10^{-5}	4
$P_{\rm e}$	WEST	RFR	0.96	2.6×10^{-6}	4	0.72	1.7×10^{-5}	28
$P_{\rm e}$	WEST	MLP	0.92	3.5×10^{-6}	1067	0.81	9.8×10^{-6}	15
$P_{\rm H}$	WEST	RFR	0.94	2.9×10^{-3}	5	0.63	2.0×10^{-2}	27
$P_{\rm H}$	WEST	MLP	0.84	5.9×10^{-3}	1325	0.70	1.1×10^{-2}	2

3. Regression results

The datasets generated are employed to develop a set of surrogate models to provide accelerated ICRF heating predictions in NSTX and WEST under the regimes investigated. In this section the performance of these surrogates, summarized in table 2, is described in terms of computational efficiency and regression accuracy. Further attention is devoted to the latter, where we analyze the regression results for the surrogate models developed for each fusion device investigated here.

Two metrics are used to assess surrogate computational performance: the training time ($t_{\rm tr}$) and, more importantly, the average inference time ($t_{\rm I}$), which represents the mean time taken by the surrogate to provide a heating prediction accross the test dataset. The determining factor driving the training time $t_{\rm tr}$ is the ML-regressor algorithm used (i.e. RFR/MLP). While affected by the network architecture (see table A1), the MLP training times reported in table 2 are mainly determined by the number of epochs needed for convergence, which are higher in WEST scenario. While in some applications, as it is the case of WEST, the RFR-based models can require negligible training times (4–5 s) compared to the MLP regressors (1067–1325 s), the overall trend is that the MLP-based models outperform the RFR in terms of inference speed by up to more than one order of magnitude (i.e. 2–15 against 27–54 μs). This result implies that both regressors could be used for different applications. For instance, while the RFR shows a simplified and faster (re)training process, the MLP regressor appears to beat the RFR models in terms of inference speed. Overall, the inference speed boost achieved by the surrogates developed ($\mathcal{O}(\mu{\rm s})$) represents roughly six orders of magnitude acceleration with respect to TORIC serial mode solutions obtained at Cori, which require $\mathcal{O}(\rm min)$ inference times. However, two considerations have to be taken into account: (*i*) the inference time used for TORIC solutions is to achieve a complete solution, including 2D wavefields, power deposition maps and other variables, in addition to the 1D power integrated profiles, (*ii*) the profiles are obtained self-consistently while the surrogates developed obtain each profile independently. Further details on the computational efficiency, together with the methodology used to implement the present surrogate models can be found in appendix.

The quantification of the regression accuracy is characterized here using two scoring metrics: the MSE, and the coefficient of determination (R^2), defined as

$$\rm MSE = \frac{1}{N_y}\sum_{j=1}^{N_y}\frac{1}{N}\sum_{i=1}^{N}\left(y_{ij}-\hat{y}_{ij}\right)^2, \tag{2}$$

$$R^2 = \frac{1}{N_y}\sum_{j=1}^{N_y}\left(1-\frac{\sum_{i=1}^{N}\left(y_{ij}-\hat{y}_{ij}\right)^2}{\sum_{i=1}^{N}\left(y_{ij}-\bar{y}_j\right)^2}\right), \tag{3}$$

where y_{ij} and $\hat{y}_{ij}$ represent the ground truth (TORIC result) and predicted value of output variable j for case i, $\bar{y}_j$ is the mean (amongst all cases) of the ground truth values for output variable j, N_y is the number of output variables, and N the number of cases. While MSE provides an estimate of the overall error of the model, R^2 features a higher sensitivity to the capability of the model to capture the variance in the dataset.

The regression accuracy results of the metrics in (2) and (3) are summarized in table 2. The metrics are shown for both the data used for training ($\rm MSE_{tr}$, $\rm R^2_{tr}$) and testing ($\rm MSE$, $\rm R^2$) of the surrogates developed for NSTX and WEST, respectively. As expected, both MSE and R^2 are observed to agree for each surrogate. Overall, the values of the metrics achieved show that the surrogates successfully capture the profiles with excellent (good) accuracy for NSTX (WEST) database. In the following, we analyze in detail the ML-regression results achieved for the ICRF heating profiles of each fusion device investigated.

3.1. HHFW heating predictions on NSTX

As observed in table 2, an excellent regression accuracy is featured by the ML-surrogates implemented for NSTX, featuring test coefficients of determination $R^2 \in [0.90\text{–}0.97]$ and mean squared error $\rm MSE \in [1.8\text{–}5.2\times10^{-5}]$ W/cm^3/MW$_{\rm abs}$. These metrics can be better interpreted by analyzing the logarithmic scale histograms of the pairwise comparison between TORIC estimations (abscissa) and TORIC-ML predictions (ordinate) for all points in all power deposition profiles within the training database (see figure 3). In this scenario, the diagonal character of the histograms in the training column (figures 3(*a*), (*c*), (*e*) and (*g*)) shows a satisfactory surrogate training

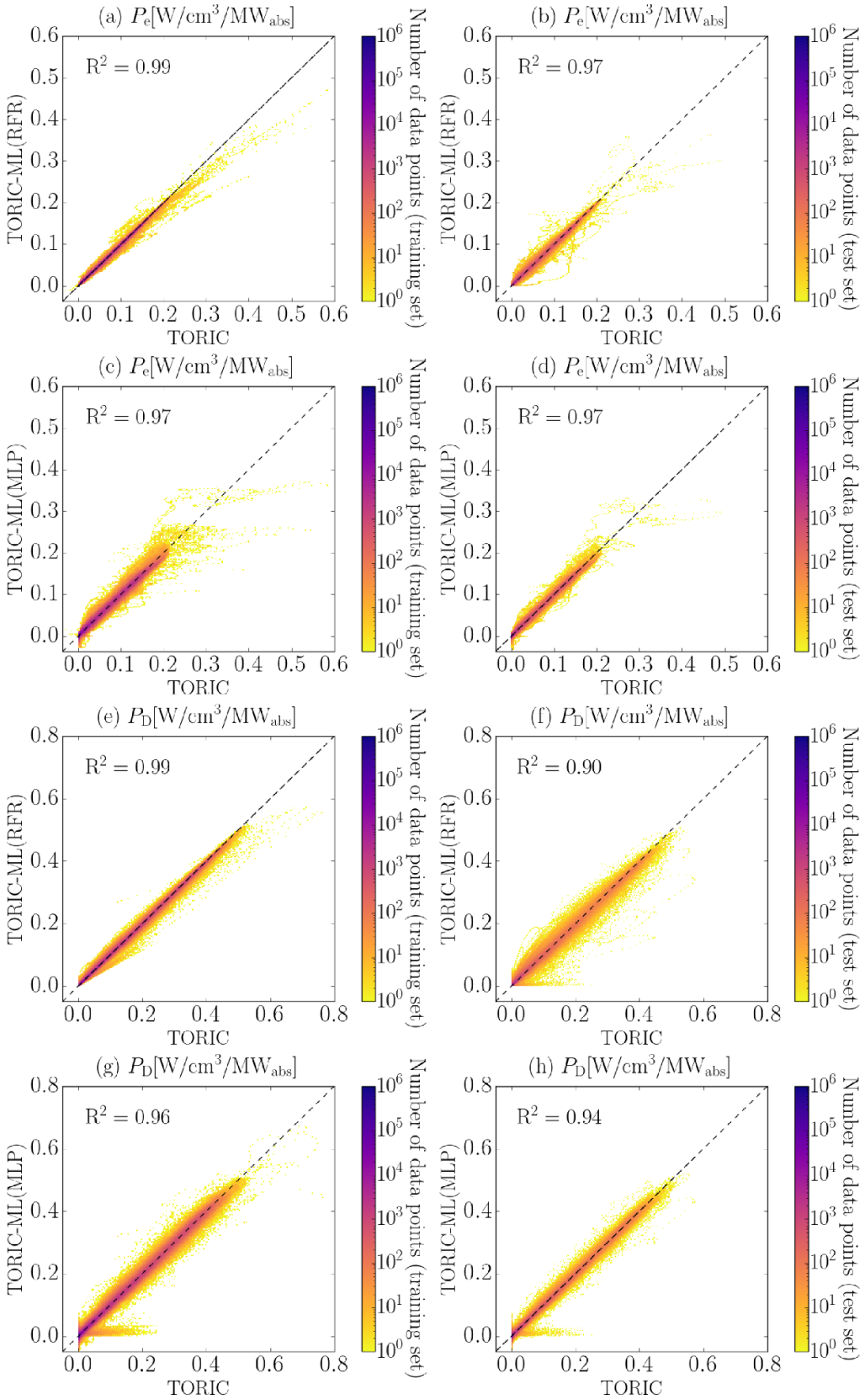

Figure 3. Log-scale histograms of regressor predictions for NSTX 1D power absorption data for training (left column) and test (right column) data. Electron power absorption results are shown for RFR (*a*) and (*b*) and MLP (*c*) and (*d*). Deuterium power absorption results are shown for RFR (*e*) and (*f*) and MLP (*g*) and (*h*).

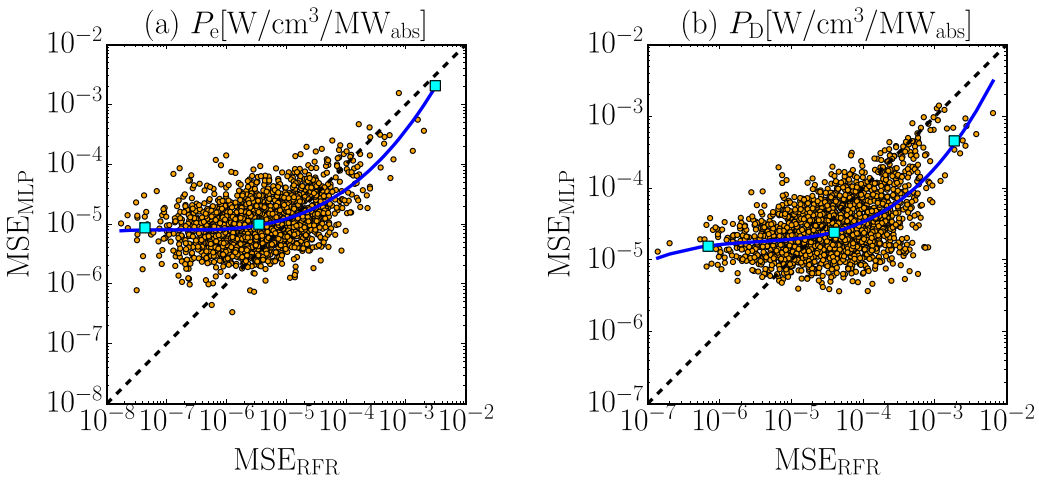

Figure 4. Scatter plot of the pairwise comparison of RFR and MLP mean squared errors for all NSTX test cases (orange circle) for electron (*a*) and deuterium (*b*) power absorption profile predictions obtained by TORIC-ML, highlighting representative cases (cyan square) of good, average and poor fitting. The location where both surrogates regression accuracy is equal (dashed) is shown as well as the cubic spline fit of the scatter cloud (solid).

process. Regarding the capability to predict unseen cases (i.e. test), the surrogate predictions are observed to preserve the regression accuracy, following TORIC estimations (see figures 3(*b*), (*d*), (*f*) and (*h*)). The surrogate models capture slightly more accurately electron power deposition profiles ($R^2 = 0.97$) than deuterium power deposition profiles ($R^2 \in$ [0.90–0.94]), especially in the case of the RFR, as can be observed in figure 3(*f*). This result may be motivated by the increased complexity of the multi-pass absorption characteristic of weak damping regime, where profiles are more unpredictable and MLP outperforms RFR performance. TORIC predictions rarely provide a negative local value of the 1D power absorption as can be seen in the abscissa of all subfigures in figure 3. While the RFR cannot intrinsically provide negative predictions if there are none in the training dataset (see figures 3(*a*), (*b*), (*e*) and (*f*)), the MLP predictions are not restricted to be of the same sign as the training set. For this reason, and the high sensitivity to capture negligible heating values, the MLP regressors can sometimes predict negative values for positive but negligible ground truth values (see figures 3(*c*), (*d*), (*g*) and (*h*)).

A further detailed comparison of both surrogates in terms of their capability to predict all cases in the test dataset is shown via a scatter plot of the pairwise comparison of RFR and MLP MSE (see figure 4). As can be seen the correlation between both regressor performances, estimated by a cubic spline fitting the scatter cloud (blue solid line), deviates from equal values (black dashed diagonal) for either electron or deuterium surrogates. This means that both surrogates do not perform equally in all regions of parametric space. In square markers we highlight examples of representative cases of 'good', 'average', and 'poor' predictions based on the MSE obtained by the surrogates (i.e. close to the cubic spline). Figure 5 shows a comparison of the highlighted cases in figure 4, representing examples good (*a*) and (*d*), average (*b*) and (*e*) and worst (*c*) and (*f*) cases for fitting in the electron (top row) and deuterium (bottom row) radial power absorption profiles in NSTX test dataset. TORIC-ML RFR (red dashed) and MLP (blue dash–dotted) profile predictions are compared to TORIC solutions (black solid). The mean squared error for each prediction is shown on figure titles and parameters in the figure caption. In the 'good' and 'average' part of the predictions RFR outperforms the MLP, although MLP still features excellent regression accuracy (see figures 5(*a*)–(*d*)) for (good, average) electron and deuterium predictions, respectively). These scenarios match the typical scenario mentioned in section 2.1 of FW single pass absorption. In the cases exhibiting highest MSE (see figures 5(*c*) and (*f*) for electron and deuterium, respectively) the MLP regressor is capable of providing significantly better results for these rather infrequent scenarios, while the RFR features larger errors in this part of the cloud. We highlight that even in the worst scenarios, which are rather infrequent in the database (see figure 4), both surrogates are able to provide a good prediction accuracy. Additionally, note that the highest errors in the RFR poor estimates are higher for smaller radius which represent a lower impact in terms of volume integrated total absorption. Overall, RFR performs better in typical scenarios and MLP outperforms in predicting rarer scenarios, for instance, featuring multi pass absorption (see figure 5(*f*)).

3.2. IC minority heating predictions on WEST

Table 2 shows that the surrogate models obtained for the WEST scenario feature a decreased scoring (i.e. $R^2 \in$ [0.63–0.81]) compared to those for the NSTX scenario (i.e. $R^2 \in$ [0.90–0.97]). This result could be expected from the increased variance observed in WEST dataset arising from (*i*) the higher number of input parameters and their relative importance in the model predictions (i.e. 7 for WEST vs. 4 for NSTX), and (*ii*) the increased complexity of the absorption physics both by the addition of a minority species fundamental for absorption and a more complex coupling with a further diverse type of wavefields including the IBW.

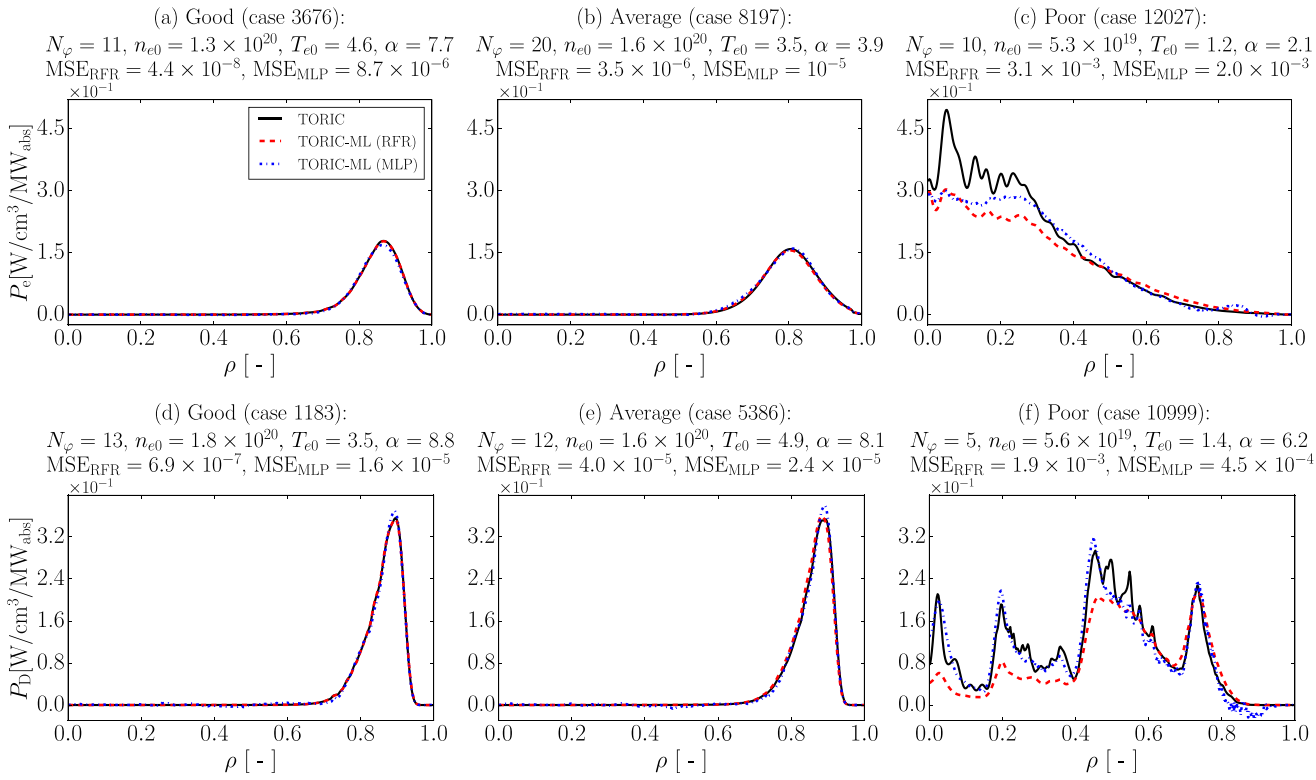

Figure 5. Good (*a*) and (*d*) average (*b*) and (*e*) and worst (*c*) and (*f*) cases for fitting in the electron (top row) and deuterium (bottom row) radial (ρ) power absorption profiles in NSTX test dataset. TORIC-ML RFR (dashed) and MLP (dash–dotted) profile predictions are compared to TORIC solutions (solid). The mean squared error for each prediction is shown on figure titles, as well as key input parameters (n_{e0}[m^{-3}] and T_{e0}[keV]) for each case.

Comparison of both training and test mean squared error obtained shows that $\mathrm{MSE_{tr}}$ is roughly one order of magnitude smaller than MSE. This denotes a higher level of overfitting in WEST surrogates compared to those of NSTX. The overfitting is observed to be specially pronounced in the RFR models. Further evidence is found in the difference between training and test coefficient of determination (i.e. $R^2_{\mathrm{tr}} \in [0.84\text{–}0.96]$ vs. $R^2 \in [0.63\text{–}0.81]$). A possible explanation for this result is an insufficient amount of information in the database, suggesting an improvement could be achieved by expanding the dataset size.

Electron and hydrogen power deposition predictions for all WEST test cases are summarized in figure 6. As can be noticed the typical values of electron power deposition are in a much smaller span (i.e. $P_e \in [0\text{–}0.25]$ W/cm^3/MW$_{\mathrm{abs}}$) compared to that of hydrogen (i.e. $P_{\mathrm{H}} \in [0\text{–}5]$ W/cm^3/MW$_{\mathrm{abs}}$), denoting again the dominant character of hydrogen absorption in IC minority heating schemes (as explained in section 2.1). Certain level overfitting can be observed comparing the training (left column) and test (right column) logscale histograms for WEST predictions shown in figure 6, where training predictions follow much closer the ideal regression compared to test predictions. An overall trend to underpredict larger power density values is observed in RFR predictions, even in the training dataset (see figures 6(*a*) and (*e*)). On the contrary, figures 6(*c*) and (*g*) show that this effect is not featured by MLP-based models, where the training regression accuracy is found to be more successful, and the predictions follow better the ideal (dashed line) regression. When it comes to testing, the overfitting is clearly observed for larger values, which coherently to the overall metrics described in table 2, is more pronounced in the case of RFR based models (6(*b*) and (*f*)) compared to MLP-based models (figures 6(*d*) and (*h*)). Regarding predictions of small values, the RFR predictions tend to outperform the MLP models, which not only sometimes estimate negative values (as also occurred in the NSTX scenario) but overall present difficulties predicting negligible absorption values. These trends can also be observed in a pairwise comparison of both RFR and MLP MSE errors (see figure 7) where the cubic spline fit of the scatter cloud is shown as a solid blue line. This cubic spline shows a similar behavior, so that for low values the RFR regressor outperforms the MLP regressor while the MLP regressor is more accurate for both electron and hydrogen large power density predictions. Also note that the MSE cloud appears to be further scattered likely featured by a higher variance in the data for the hydrogen models (see figure 7(*b*)) compared to electron models (see figure 7(*a*)).

Figure 8 shows the cases highlighted in figure 7(cyan square) for both electron and hydrogen predictions using RFR (red dashed) and MLP (blue dash–dotted) compared to the TORIC ground truth (black). Depending on the MSE value, the (electron, hydrogen) ICRF heating predictions of unseen cases are classified into 'good' MSE (see figures 8(*a*) and (*d*)), 'average' MSE (see figures 8(*b*) and (*e*)), and high MSE or 'poor'

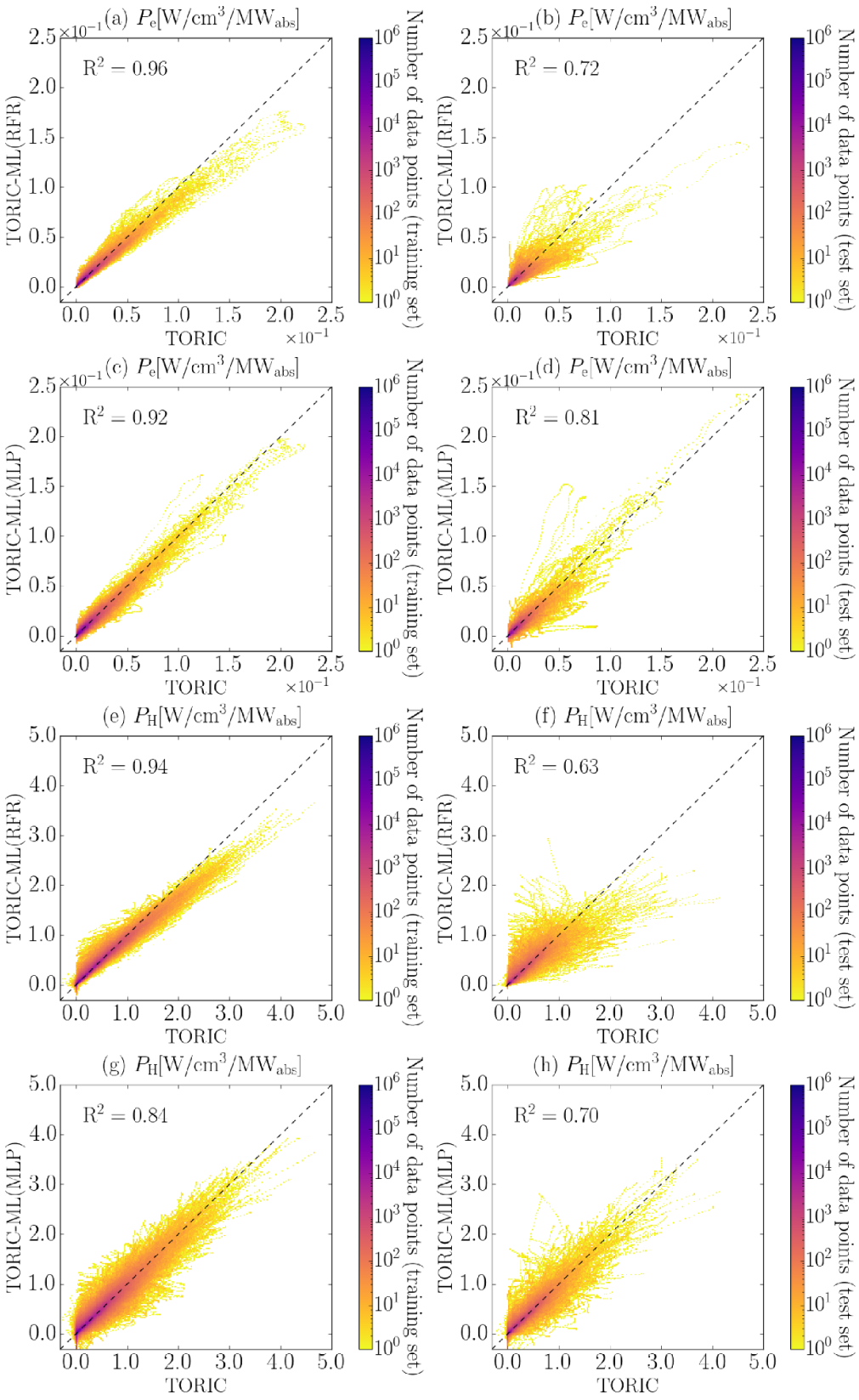

Figure 6. Log-scale histograms of regressor predictions for WEST 1D power absorption data for training (left column) and test (right column) data. Electron power absorption results are shown for RFR (*a*), (*b*) and MLP (*c*), (*d*). Hydrogen power absorption results are also shown for RFR (*e*), (*f*) and MLP (*g*), (*h*).

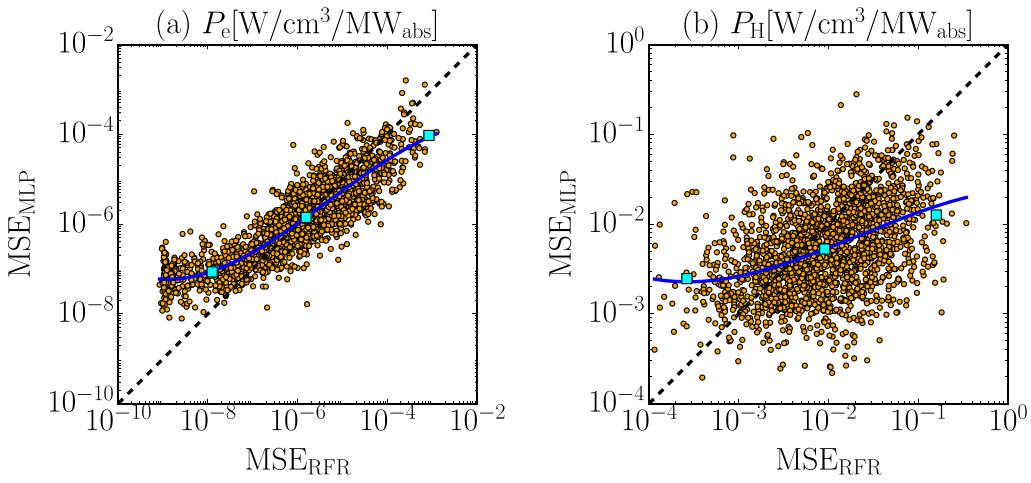

Figure 7. Scatter plot of the pairwise comparison of RFR and MLP mean squared errors for all WEST test cases (orange circle) for electron (*a*) and hydrogen (*b*) power absorption profile predictions obtained by TORIC-ML, highlighting representative cases (cyan square) of good, average and worst fitting. The location where both surrogates regression accuracy is equal (black dashed) is shown as well as the cubic spline fit of the scatter cloud (blue solid).

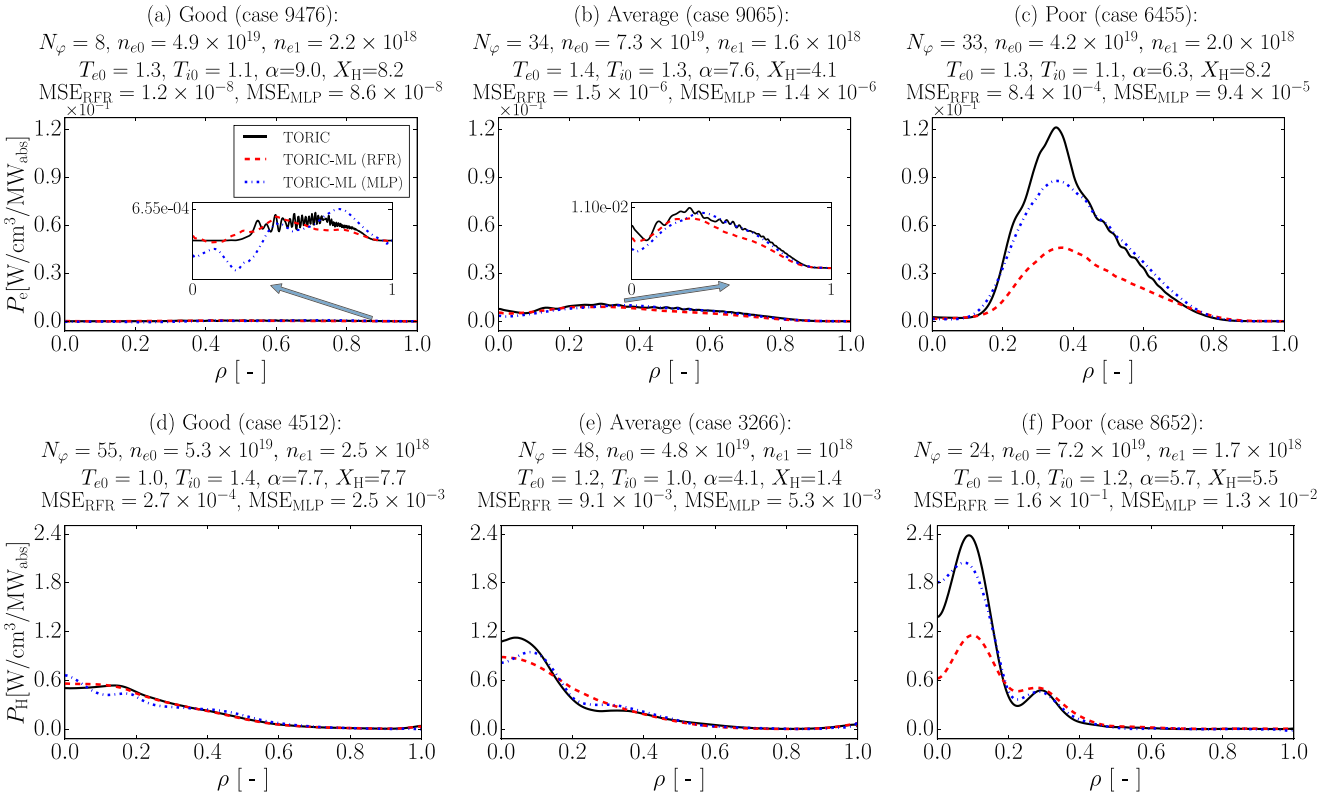

Figure 8. Good (*a*) and (*d*), average (*b*) and (*e*) and worst (*c*) and (*f*) predictions of RFR (dashed), MLP (dash–dotted), of the TORIC (solid) the electron (*a*)–(*c*) and deuterium (*d*)–(*f*) radial (ρ) power absorption profiles in WEST test dataset. Zoomed in profiles are shown in insert axis when their maxima are lower than 20% of the maximum of all predictions shown for each species.The mean squared error for each prediction is shown on figure titles, as well as key input parameters ($n_{e0|e1}$[m^{-3}], $T_{e0|i0}$[keV], X_H[%]) for each case.

(see figures 8(*c*) and (*f*)) categories. Insert axes show zoomed in profiles when their maxima are below 20% of the maximum for predictions of that species shown in figure 8. In the case of electron power absorption predictions, it is observed often that 'good' cases tend to be characterized by negligible absorption (see case 9476 in figure 8(*a*)). In these cases, RFR predictions are far more accurate than the MLP counterpart (see zoom region in figure 8(*a*)). Nevertheless, for those cases, MLP MSE remains still negligible and irrelevant to characteristic absorption profile magnitudes. Note that this is not the case of 'good' hydrogen absorption (see example case 4512 in figure 8(*d*)), as it is the species dominating absorption in the ample majority of cases, and MLP is still able to capture these profiles. However, the typical MSE of MLP (i.e. 2.5×10^{-3}W/cm^3/MW$_{abs}$) is still

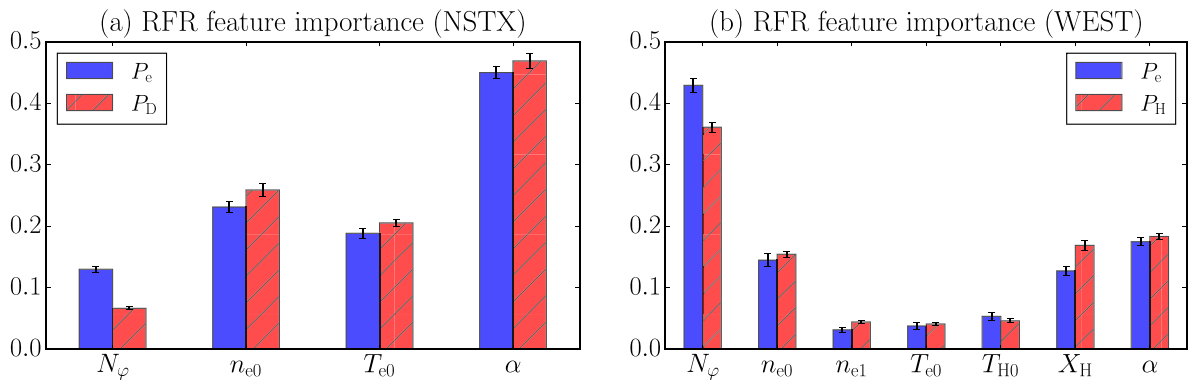

Figure 9. Histograms of the RFR feature importance for electron (plain bar) and dominant ion power absorption (dashed bar) for (*a*) NSTX and (*b*) WEST scenarios. The error bars quantify the standard deviation in the feature performance of the 100 estimators used in the ensemble RFR models.

higher than that of RFR (i.e. 2.7×10^{-4}W/cm^3/MW$_{abs}$), consistent with the location of the cubic spline fit in figures 7(*a*) and (*d*) in 'good' predictions. In both electron and hydrogen 'average' predictions both regressors show the capability to capture the profiles trends (see example cases 9065 and 3266 in figures 7(*b*) and (*e*)). Regarding 'poor' regression cases, it is observed that while the RFR can follow the ground truth shape, it is generally unable to estimate the peak absorption value (see example cases 6455 and 8652 in figures 8(*c*) and (*f*)). This effect is also observed on MLP predictions but to a lower amount, with roughly one order of magnitude smaller MSE than that of RFR. In summary, the different surrogates trained for WEST can be combined to obtain good predictions for both negligible (i.e. using RFR) and significant (i.e. using MLP) absorption cases.

3.3. *Feature importance of NSTX and WEST*

The RFR algorithm allows us to identify the relative importance of the features used in our models to predict the targets. The feature importance is computed as the normalized averaged accuracy gain when using a feature in the decision trees of the RFR algorithm to estimate target variables. This accuracy gain is measured in terms of MSE difference between value before and after separating according to each decision. The average is taken over all the trees and normalized with respect to the total importance to provide a relative value (i.e. from [0–1]). RFR surrogates feature importance is only barely affected by the species both in NSTX (see figure 9(*a*)) and WEST (see figure 9(*b*)) surrogates. In the NSTX scenario, the RFR models show that the most important feature is the profile inner exponent α (see figure 9(*a*)), followed by the core electron density, temperature and the least important being the toroidal mode number. This result is consistent with the expected behavior of the model given the majority of scenarios in the database are coherent with the typical one described in section 2.1, namely, strong FW single pass absorption. The location of the absorption is then more influenced by the gradients in the plasma property profiles, followed by the core electron temperature and density. Thus, the parameter influencing less the predictions in the filtered NSTX dataset (for more details see section 4) is the toroidal mode number. The WEST scenario, however, appears to be highly determined by the toroidal mode number (see figure 9(*b*)) and the minority concentration. These parameters are known to play a major role in determining the distance between the fundamental resonance and the ion–ion hybrid resonance, which amongst other effects, highly influences FW to IBW mode conversion efficiency [44, 48, 54, 55]. The toroidal mode number is found to be most dominant parameter influencing the profiles as could be expected from its wide parameter span. Two other relevant parameters are the profile inner exponent and next the electron core density, similarly to the case of NSTX. Finally, the rest of the parameters (core ion and electron temperature, and edge density) are less relevant.

4. Discussion on outliers

We perform an exploratory analysis of each TORIC database to identify and classify cases that do not follow the targeted behavior, generally termed as outliers, allowing us to choose if they are to be maintained or not in the surrogate training. While the majority of scenarios in the datasets for both HHFW and IC minority heating schemes follow the typical phenomenology described above in section 2.1 (see figure 2), some outlier scenarios are found in each database.

In the NSTX database, the outliers observed feature non-physical behavior. In these outliers a significant spike appears in the deuterium/electron absorption profiles, as can be observed in the example shown in figure 10(*a*). This power deposition corresponds to the electric wavefield shown in figure 10(*b*). The wavefield solutions present numerical instabilities [39, 56]. These numerical instabilities appear within the plasma volume in the vicinity of the IC resonances and denote a spurious wavefront with a wavelength correlated with the mesh-element size. Moreover, these scenarios correspond to non-converged cases in terms of energy balance. In order to identify these outliers in each profile, a numerical metric has been designed and implemented. The metric allows to identify outliers, ordering them by the relative importance of the spike appearing in the power deposition profile. The metric

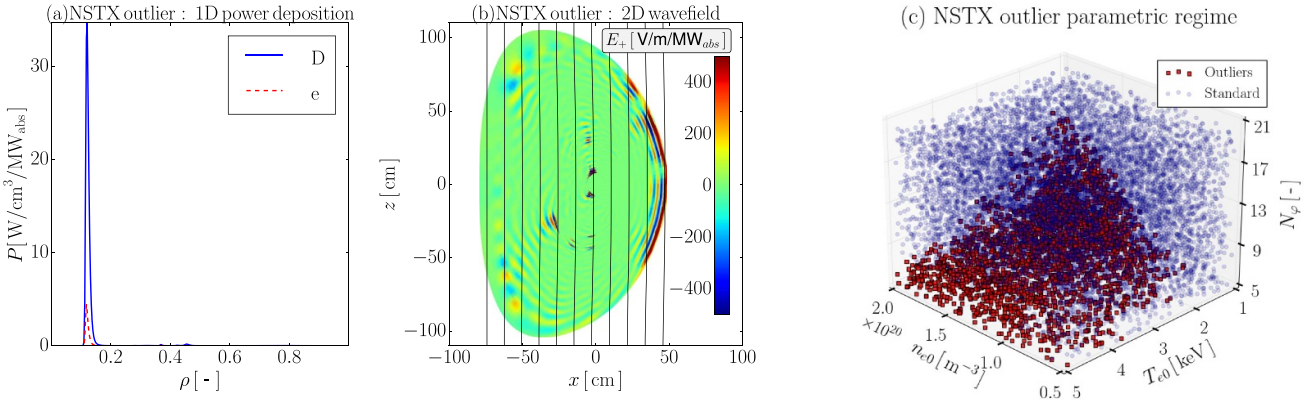

Figure 10. Outliers in NSTX database: (*a*) 1D power deposition of deuterium (solid) and electrons (dashed) and (*b*) 2D left-handed polarized wave electric field. In (*b*) solid vertical lines represent the harmonic ion cyclotron resonances within the plasma volume. Colormaps show the wavefield amplitude saturated to ± 500 V/m/MW$_{\mathrm{abs}}$. A scatter plot (*c*) of the location of outliers (red square) and standard cases (blue circle) is shown within the TORIC input parametric space investigated for the HHFW heating in NSTX.

is defined by the difference of the Gaussian-filtering smoothed signal of each power profile to its original value normalized by the profile maximum value. This metric enabled fast identification of 22.5% (i.e. 2752) outliers in the NSTX database that were removed from the database. As a result, 9488 cases comprise the filtered dataset employed for training the surrogates shown in section 3.1. Figure 10(*c*) shows a scatter plot of the location of outliers (square-red) and standard cases (circle-blue) within the TORIC input parametric space investigated for HHFW heating in NSTX (core electron density and temperature, and toroidal mode number). One can clearly observe that the parametric regime where these outliers appear is a delimited region defined by (*i*) low toroidal mode number, (*ii*) high electron density and low electron temperature. Amongst these parameters the toroidal mode number appears to be the key parameter driving these instabilities. In fact, for a sufficiently high toroidal mode number this instability does not appear in the solution. Even though the numerical parameters used to generate the database are fixed for simplicity, we find that these numerical instabilities can be mitigated or even, sometimes, eliminated, by reducing the poloidal resolution and thus the maximum number of poloidal modes in the spectral solution.

Regarding WEST database, our analysis did not show any non-physical cases featuring numerical instabilities unlike the NSTX database. For this reason, the entire IC minority heating database is employed for surrogate training. However, some challenging cases are still present. An example of these outliers is shown in figure 11 indicating the presence of under-resolved and underdamped IBW modes at the high field side. The parametric regime in which these outliers appear is correlated exclusively to a low toroidal mode number.

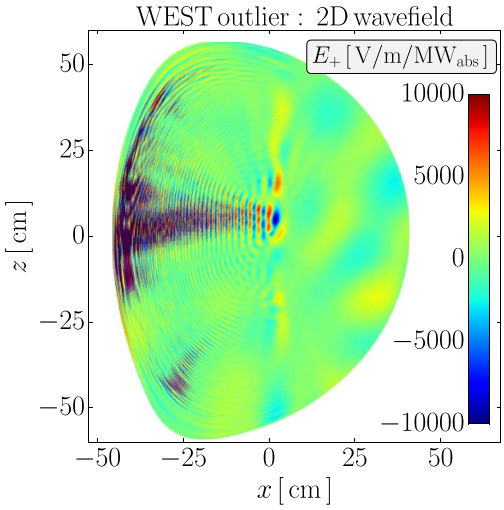

Figure 11. Example 2D left-hand polarized wave electric field solution representative of outlier cases in the WEST database. Wavefield amplitude (colormap) is saturated to $\pm 10\,000$ V/m/MW$_{\mathrm{abs}}$.

4.1. Impact of NSTX outliers on surrogate performance

In order to assess the impact of NSTX outliers identification and removal from the database on the achieved regression accuracy, a set of surrogate models is trained with the original NSTX database including outliers (i.e. the totality of 12 240 cases) in comparison with the filtered database (9488 cases) used in section 3.1. Figure 12 shows the impact on surrogates' regression accuracy (MSE) of training RFR (*a*) and (*b*) and MLP (*c*) and (*d*) surrogates for electron (*a*) and (*c*) and deuterium (*b*) and (*d*) HHFW heating profiles with NSTX original (plain bar) or filtered (dashed bar) dataset (i.e. with/without outliers). Histograms represent the proportion of test cases featuring an MSE within an specific accuracy range (abscissa). One can observe that the error distributions are similar for both electron figures 12(*a*), (*c*) and deuterium figures 12(*b*) and (*d*) power predictions, for both RFR and MLP methods. In all cases, the larger complexity of the deuterium power deposition increases the errors in the predictions compared to that of electrons. While the RFR features a higher variance in the predictions' MSE (see figures 12(*a*) and (*b*)), MLP MSE coalesce around higher average values (see figures 12(*c*) and (*d*)). The

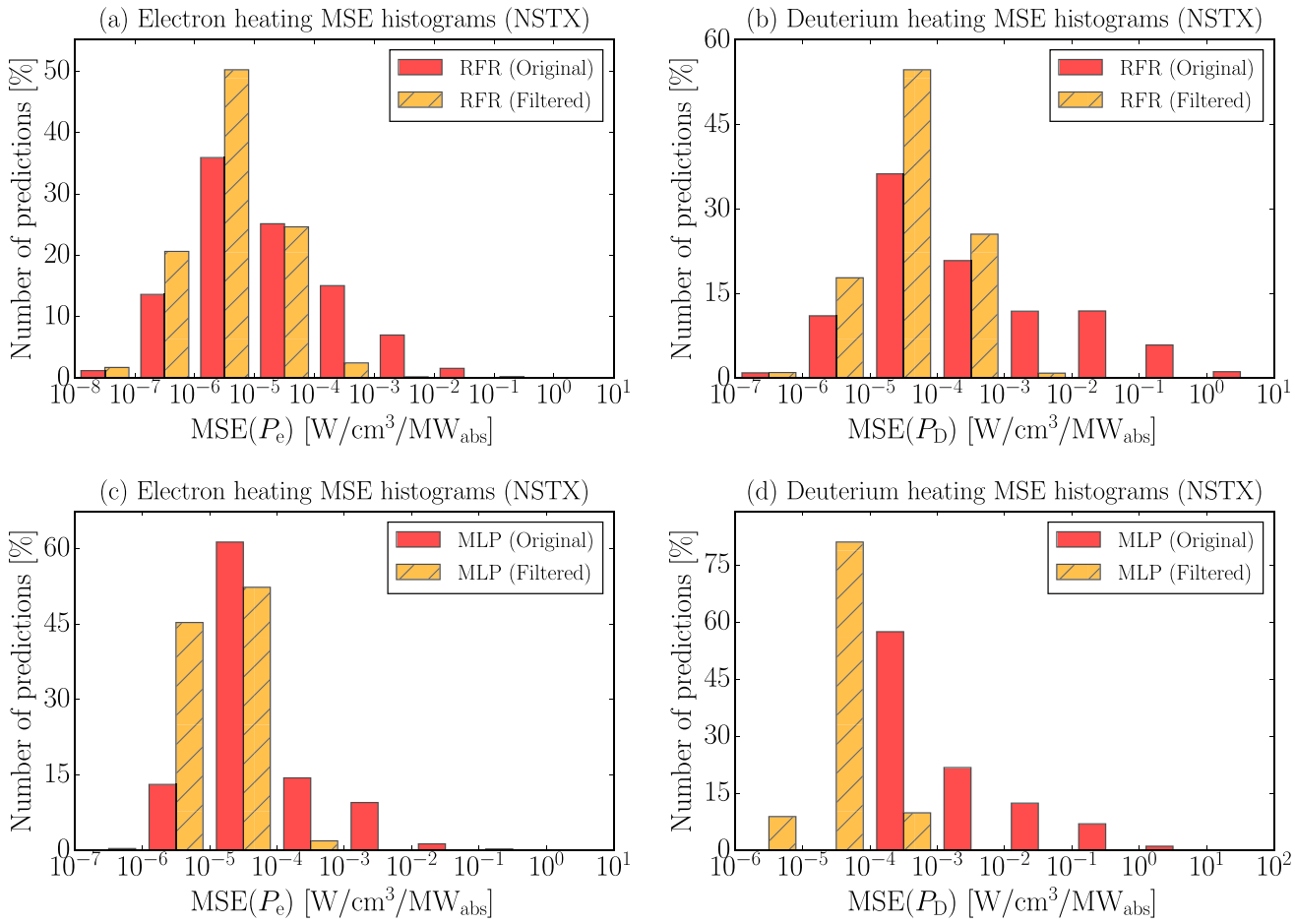

Figure 12. Impact on surrogates' regression accuracy (MSE) of training RFR (*a*), (*b*) and MLP (*c*), (*d*) surrogates for 1D electron (*a*), (*c*) and deuterium (*b*), (*d*) HHFW heating profiles with NSTX original (plain bar) or filtered (dashed bar) dataset (i.e. with/without outliers). Histograms represent the proportion of test cases featuring an MSE within an specific accuracy range (abscissa) for each of the surrogates developed.

impact of outlier filtering on RFR predictions appears to eliminate higher errors of the distribution while maintaining the overall shape. On the contrary, MLP predictions denote higher sensitivity to outliers in the training data, as can be observed (see figures 12(*c*) and (*d*)) in the noticeable left-shift experienced by the MSE distribution when training with a filtered dataset (dashed bar) compared to the original dataset (plain bar).

Although the surrogates trained with the original dataset are capable of capturing the features of the major part of the signal while exhibiting large errors in some predictions, specially for deuterium power absorption, their regression accuracy significantly drops with respect to training with a filtered dataset. The regression accuracy drop is not only correlated to the unpredictability of power profiles in nonphysical outlier scenarios, but also to the failed attempt to learn from these outliers and predicting them in standard scenarios. Overall, the set of models trained with outliers in the NSTX database feature poor to good scoring in the predictions ($R^2 \sim 0.51$–0.75) depending on the ML-method and species, and further justifies the removal of these outliers from the database used for surrogate training.

4.2. ML-based predictions in TORIC-HHFW non-physical outlier scenarios

Outlier identification and elimination from the HHFW database results in TORIC-ML being limited to a parametric regime where outliers are not present. This parametric space is a subspace of the total parametric space of interest. In order to address this issue, we train surrogates on the outlier-free part of the dataset to test their capability to predict profiles in non-physical outlier scenarios. Here we classify non-physical scenarios into three categories (minor, major and critical) regarding the outlier identification metric employed, which quantifies the nonconformance degree of the outlier portion of the profile with respect to the physical counterpart. A minor nonconformance outlier is characterized by a negligible type of outlier part of the signal compared to the total signal. A major nonconformance outlier features an outlier part comparable in magnitude to the signal scale. Critical nonconformance outliers occur when the outlier part of the signal considerably exceeds the physical one. Figure 13 shows examples of minor (*a*) and (*d*), major (*b*) and (*e*), and critical (*c*) and (*f*) nonconformance outlier (electron, deuterium) solutions obtained with

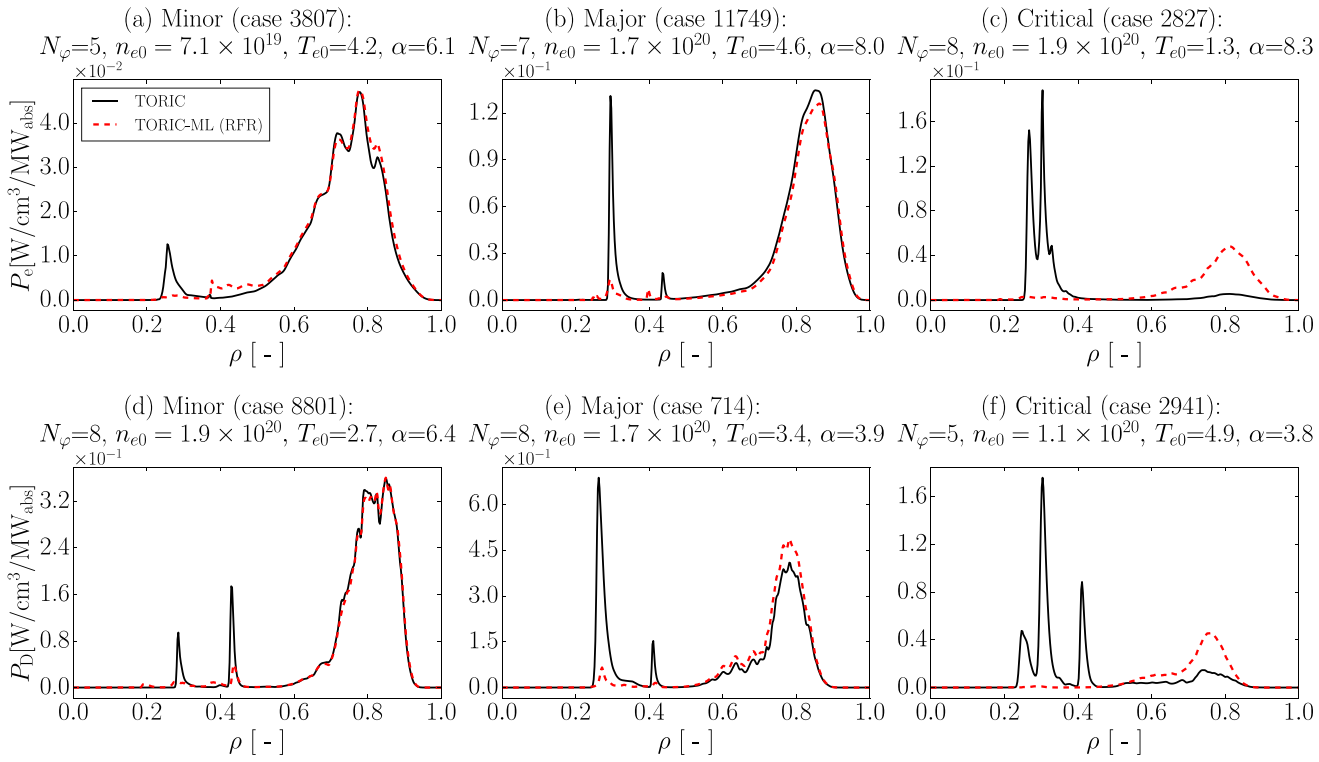

Figure 13. TORIC-ML RFR predictions (red dashed) of electron (top) and deuterium (bottom) power deposition in minor (*a*), (*d*), major (*b*), (*e*) and critical (*c*), (*f*) outlier scenarios compared to TORIC solutions (black solid). Key input parameters (n_{e0} [m^{-3}] and T_{e0}[keV]) are shown in the title for each case.

TORIC. While TORIC-ML predictions using MLP are unsatisfactory in these outlier scenarios, TORIC-ML RFR predictions featured interesting characteristics. Firstly, the RFR predictions eliminate the outlier part of the signal for all scenarios, independently of the amount of sharp maxima in the signal. Secondly, the surrogate predictions preserve the main physical aspects of HHFW heating absorption predicted by TORIC, in both minor and major nonconformance outlier solutions. Thirdly, the predictions feature a correction that correlates to the nonconformance degree of the TORIC prediction. As a result, in critical nonconforming outliers (see figures 13(*c*) and (*f*)) TORIC-ML predicts an outlier-free power profile with an increased power deposition at locations of physical deposition predicted by TORIC. Note that further verification of the physical conformity of these predictions remains to be carried out with other models capable of predicting HHFW physics in the outlier regime found using TORIC. Overall, these features not only extend the parametric regime capability of TORIC including minor and major outliers but also represent an increase in model robustness.

5. Concluding remarks and prospective directions

This paper describes a first proof of principle of acceleration of high-fidelity predictions of ICRF power absorption in both HHFW and IC minority heating schemes via ML approaches. The achieved prediction times, in the order of $\mathcal{O}(2\text{–}50)\,\mu$s, represent a significant increase in computational efficiency with respect to TORIC, the original model, featuring average inference times per case of $\mathcal{O}(1\text{–}5)$ min. Excellent agreement is found between TORIC solutions and TORIC-ML predictions of NSTX HHFW power deposition profiles in unseen cases, with coefficients of determination of $R^2 = 0.97(0.94)$ for electron (deuterium) species. Good surrogate accuracy is achieved for both electron and hydrogen heating profiles in WEST (i.e. $R^2 = 0.81(0.70)$ for electron (hydrogen) species). In both NSTX and WEST scenarios, the RFR achieves more accurate predictions in typical single pass absorption scenarios. MLP regressors outperform RFR predictions in less frequent heating scenarios in the database, as it is the case of multi-pass absorption. The RFR are shown to both simplify training and reduce training times. However, MLP regressors feature shorter inference times. The RFR training process also provides insight on the relative feature importance in the recognition of the different profiles in each database. While in NSTX the majority of single-pass absorption scenarios in the database results in plasma properties dominating the type of profile observed, in WEST it is the toroidal mode number and the minority concentration, which strongly influence the mode conversion efficiency. Furthermore, an exploratory analysis of the databases shows that these contain some numerically challenging scenarios termed as outliers, when using the TORIC version in TRANSP, in both HHFW and IC minority heating schemes. An investigation on the nature of these outliers shows that while the outlier scenarios in WEST have a physical explanation, the outliers found in NSTX are non-physical and arise from a numerical instability [39, 56]. The

instabilities are shown to appear in well-defined parametric regions and to be correlated with IC resonances. The impact of filtering these outliers from the training database is shown to provide a significant boost in the achieved surrogates' regression accuracy. A ML-based solution to predict HHFW heating profiles in outlier parametric regimes is presented, extending the operative parametric range of the ICRF model, and providing a reasonable profile where the original model fails. These profiles are shown to preserve the main physical mechanisms of HHFW absorption while eliminating the outlier portion of the predictions. Overall, the significant increase in computational efficiency, the high regression scoring achieved and the enhanced robustness in specific HHFW scenarios, make this work a critical step to achieve a complete real-time capable ICRF heating model which could be incorporated in plasma control systems or integrated modeling frameworks. Future extensions of these surrogates expect a significant broadening in input parametric space dimensionality to include more complex heating scenarios such as the three-ion heating scheme [48], non-Maxwellian effects [57], the impact of impurities, generalized plasma properties and equilibria and core/edge coupling [58].

Data Availability Statement

The data that support the findings of this study are openly available in Princeton Data Commons at https://doi.org/10.34770/g2gt-4694.

Acknowledgments

The authors Á. Sánchez-Villar, N. Bertelli, and S. Shiraiwa thank R. Bilato and M. Brambilla for the fruitful discussion on the TORIC code and the outlier scenarios found. This work was supported by the U.S. Department of Energy under Contract Number DE-AC02-09CH11466. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC Award FES m3716 for 2023. The United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

Appendix Surrogate implementation

A.1. ML algorithms

Providing predictions of the power profiles via surrogate modeling is a process that involves the solution of a regression problem, which is a core application of supervised ML algorithms. Supervised learning is based on labeled data on which an algorithm is trained to capture the main characteristics of the transfer function between inputs and outputs. The methods used in this work involve two ML-algorithms: the RFR and the MLP.

RFR is a ML-algorithm that employs an ensemble of decision trees that select a value depending on a sequence of conditions. The algorithm adjusts the conditions of each tree by maximizing the information gain per condition. Ensemble methods provide improved prediction accuracy as they make use of prediction averages of independently trained surrogates to obtain a novel prediction. These methods can reduce the over-fitting arising from training to specific features of a part of the dataset.

MLP is an artificial neural network composed of multiple layers of perceptrons, the most elementary type of networks often used for linearly separable problems. The MLP has shown to be successful for implementation of non-linear regressions of multi-dimensional data. MLP connects multiple perceptrons in a network with an input layer, a series of hidden layers and an output layer. In the networks considered in this work the structure is fully-connected, meaning each layer neuron is connected to all neurons in both previous and next layers. A linear combination of the previous layer predictions is fed to this neuron's activation function that provides an output, transferred to the next layer neurons. A back-propagation algorithm optimizes the weights and biases in the network to minimize the prediction error.

A.2. Surrogate implementation

The ML workflow has been carried out on Perlmutter scientific supercomputer at NERSC facilities. The implementation has been performed using Jupyter (https://docs.nersc.gov/services/jupyter/) on a Perlmutter exclusive CPU node, which consists of 2xAMD EPYC^{TM}7713 (Milan) 2.45 GHz CPUs featuring 64 cores, 2 threads per core, and 512 GB of DDR4 memory total. This study evaluates the performance of surrogate models that we implemented for the two species dominating absorption, which are electrons and deuterium (hydrogen) for the NSTX (WEST) scenario. The developed surrogate models make use of Python ML libraries. The RFR and MLP models use the `RandomForestRegressor` and the `MLPRegressor` of `scikit-learn` library [59] within the `ensemble` and `neural_network` modules, respectively. Each model's hyperparameters were tuned with a different methodologies depending on the regressor. Sensitivity analyses to RFR hyperparameters, which are mainly the number of trees in the forest (i.e. `n_estimators`), the maximum depth of the tree (i.e. `max_depth`) and the maximum number of features to consider when splitting the node (i.e. `max_features`), showed minor variation of the error for `n_estimators`>100, while featuring significant increase in both training and average inference times, apart from the increase in memory size of the surrogate. While `max_depth` barely affected surrogate performance, the default `max_features` = 2 was found to be optimal for surrogate performance.

The number of relevant hyperparameters is higher for MLP than for RFR, and are also more intricate to tune. The main hyperparameters of the MLP regressor are: (*i*) the `activation` function (e.g. 'identity/logistic/tanh/relu'),

Table A1. Set of hyperparameters selected for each `MLPRegressor`-based surrogate model using both grid searches and methodical scanning.

Target	Scenario	Activation	Hidden layer sizes	Solver	Batch size	L^2-term strength
P_e	NSTX	`ReLU`	(15,35,15)	`Adam`	40	0.01
P_D	NSTX	`ReLU`	(45,105,45)	`Adam`	40	0.001
P_e	WEST	`ReLU`	(30,70,30)	`Adam`	30	0.0001
P_H	WEST	`ReLU`	(50,50,50,50)	`Adam`	30	0.0001

Table A2. Cross validation of the multiple models using MSE. MSE mean (μ) and standard deviation (σ) of the 5-fold for each model are compared to the corresponding MSE of the final model developed (Test).

			RFR	MLP
P_e (NSTX)	5-fold	μ(MSE)	2.4×10^{-5}	2.4×10^{-5}
		σ(MSE)	1.8×10^{-6}	1.7×10^{-6}
	Test	μ(MSE)	1.8×10^{-5}	2.1×10^{-5}
P_D (NSTX)	5-fold	μ(MSE)	1.2×10^{-4}	5.2×10^{-5}
		σ(MSE)	6.2×10^{-6}	2.1×10^{-6}
	Test	μ(MSE)	1.1×10^{-4}	5.2×10^{-5}
P_e (WEST)	5-fold	μ(MSE)	1.6×10^{-5}	9.4×10^{-6}
		σ(MSE)	1.9×10^{-6}	8.6×10^{-7}
	Test	μ(MSE)	1.7×10^{-5}	9.8×10^{-6}
P_H (WEST)	5-fold	μ(MSE)	2.1×10^{-2}	1.2×10^{-2}
		σ(MSE)	1.7×10^{-4}	8.0×10^{-4}
	Test	σ(MSE)	2.0×10^{-2}	1.1×10^{-2}

(*ii*) network structure given by `hidden_layer_sizes`, (*iii*) the `solver` for used optimization, (*iv*) the size of the mini-batches used for the stochastic optimizer (`batch_size`), (*v*) the `learning_rate` defining the step-size in the updating process of the weights, (*vi*) and the weight of the L^2-norm regularization term `alpha`, and (*vii*) the iteration number controlled by the number of iterations `n_iter_nochange` without a change in the order of a tolerance `tol`, with a maximum number of iterations `max_iter`. Parameter `alpha` allows to prevent model overfitting by constraining the weights size. The sets of MLP hyperparameters selected after tuning are shown in table A1 and resulted in different combinations for each surrogate. A methodical scanning of the hyperparameters was carried out starting with low node number and coming with a layer proportion optimal combination together with other hyperparameters as the `batch_size` and `alpha`. For this, a series of grid searches using `GridSearchCV` were found useful to tune hyperparameters. While the optimal `activation` function and `solver` were consistently found to be `ReLU` and `Adam` [60], respectively, other hyperparameters had a minor effect on surrogate scoring, as it was the case for both `learning rate` method and initial value. The optimal `batch_size` was found to be 40(30) for both surrogates obtained for NSTX(WEST). Optimal `alpha` was found to be larger for NSTX cases (i.e. 0.01 and 0.001) compared to those of WEST (i.e. 0.0001 for both). Optimal network structures (i.e. `hidden_layer_sizes`) were found to have three/four layers. The optimal number of nodes per layer was found to be symmetric and peaked at the middle hidden layer for most surrogates, except that of hydrogen absorption in WEST, found to be 50 for all layers. Additionally, the tolerance `tol` was tuned according to the sensitivity of the convergence to this parameter for each network. In case the convergence of the loss function is not achieved, parameters `n_iter_no_change` and `max_iter` were increased to extend the number of epochs of the optimization process.

A.3. Cross-validation

To determine the model sensitivity to the randomization used when splitting the dataset into test and training sets, a cross-validation of each surrogate model is carried out via k-fold validation. We split the dataset randomly into $k = 5$ folds or subsets. For each fold, we train a surrogate model on the remaining part of the dataset and test the fold computing the MSE. After all models are trained and tested, we compute both the MSE mean (μ) and standard deviation (σ) amongst the different models. Table A2 shows the results of the 5-fold cross-validation for each target profile in each scenario. It can be noticed that both RFR and MLP regressors are not very sensitive to the dataset split, consistently featuring $\sigma(\text{MSE}) \ll \mu(\text{MSE})$. Additionally, the cross-validated averages (5-fold) can be compared to the MSE obtained from the final models developed (Test) using randomize splitting. In all models, the

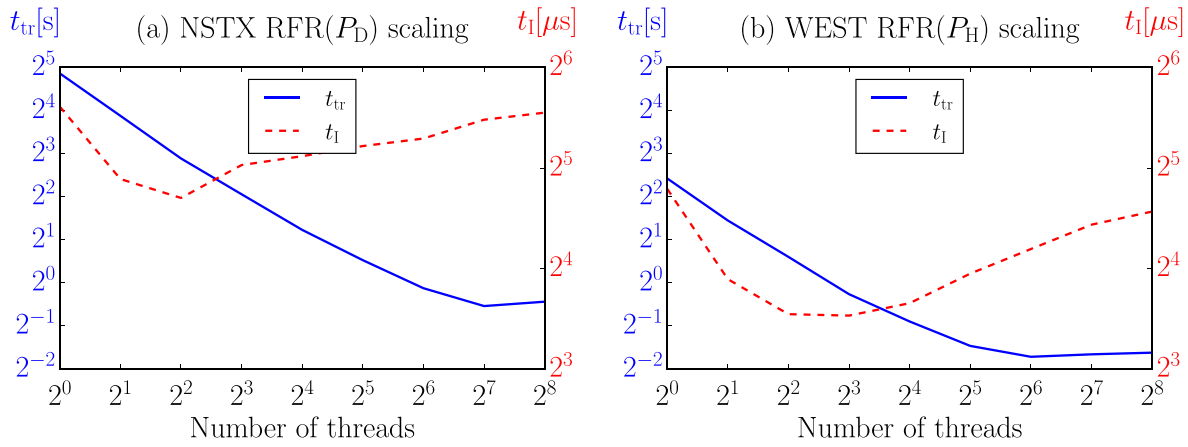

Figure A1. Scaling of computational efficiency figures of merit of the RFR-based surrogates for the ICRF heating of (*a*) deuterium in NSTX and (*b*) hydrogen in WEST. Scales are shown in base-2 logarithmic plot for both training time t_{tr} (solid), and inference time t_I (dashed). Independent tick labels are shown for each figure of merit, being training time on the left, and average inference time on the right.

μ(MSE) values of the 5-fold and the final model are similar, denoting the negligible bias included in the training due to randomized data splitting for the final models shown in section 3.

A.4. Multi-threading impact on surrogate computational efficiency

Here we analyze the impact of multi-threading and parallelism in the computational efficiency of the ML-models developed. While the `MLPRegressor` of `sklearn` does not take advantage of any type of parallelism, `RandomForestRegressor` uses different levels of parallelism including higher-level (i.e. pipelining) and lower-level via OPENMP configuration and multithreading capabilities of `NumPy` and `SciPy` libraries. Here we focus on the higher-level parallelization of `RandomForestRegressor`, which is controlled by parameter `n_jobs`, which specifies the number of threads used. Figure A1 shows a sensitivity analysis to the number of threads. It can be noticed that while training time decreases almost linearly with the number of cores used, a saturation minimum occurs at 128(64) for NSTX(WEST) scenario. The training time appears to be more restrictive for the NSTX RFR model for deuterium (see figure A1(*a*)) than for the WEST RFR model for hydrogen (see figure A1(*b*)) by more than one order of magnitude. Regarding the average inference time, saturation occurs rapidly and above 4–8 threads for both scenarios, inference not only does not decrease by using more cores, but in fact increases, probably due to ineffective use of the parallelization strategy. In the rest of this work computational efficiency is compared for single thread (i.e. serial).

ORCID iDs

Á. Sánchez-Villar https://orcid.org/0000-0003-3727-9319
Z. Bai https://orcid.org/0000-0002-3092-0903
N. Bertelli https://orcid.org/0000-0002-9326-7585
E.W. Bethel https://orcid.org/0000-0003-0790-7716
J. Hillairet https://orcid.org/0000-0002-1073-6383
T. Perciano https://orcid.org/0000-0002-2388-1803
S. Shiraiwa https://orcid.org/0000-0001-5249-0441
G.M. Wallace https://orcid.org/0000-0001-6303-3684
J.C. Wright https://orcid.org/0000-0001-6524-6149

References

[1] Wilson J. and Bonoli P. 2015 *Phys. Plasmas* **22** 021801
[2] Start D. *et al* 1999 *Nucl. Fusion* **39** 321
[3] Rimini F. *et al* 1999 *Nucl. Fusion* **39** 1591
[4] Yang H. *et al* 2021 *Nucl. Fusion* **61** 035001
[5] Zhang X. *et al* 2022 *Nucl. Fusion* **62** 086038
[6] Hillairet J. *et al* 2021 *Nucl. Fusion* **61** 096030
[7] Colas L. *et al* 2021 *Nucl. Fusion* **62** 016014
[8] Bobkov V. *et al* 2020 *AIP Conf. Proc.* **2254** 040005
[9] Ochoukov R. *et al* 2020 *AIP Conf. Proc.* **2254** 030005
[10] Menard J. *et al* 2017 *Nucl. Fusion* **57** 102006
[11] Lin Y., Wright J. and Wukitch S. 2020 *J. Plasma Phys.* **86** 865860506
[12] Hawryluk R. *et al* 1981 *Physics of Plasmas Close to Thermonuclear Conditions* ed B. Coppid *et al* p 19
[13] Crotinger J.A. *et al* 1997 *LLNL Report* UCRL-ID-126284
[14] Pereverzev G.V. and Yushmanov P. N. 2002 ASTRA. Automated System for Transport Analysis in a Tokamak (IPP 5/98) (Max-Planck-Institut für Plasmaphysik)
[15] Artaud J.F. *et al* 2010 *Nucl. Fusion* **50** 043001
[16] Hayashi N. *et al* 2010 *Phys. Plasmas* **17** 056112
[17] Romanelli M. *et al* 2014 *Plasma Fusion Res.* **9** 34030231
[18] Artaud J.F. *et al* 2018 *Nucl. Fusion* **58** 105001
[19] Humphreys D. *et al* 2020 *J. Fusion Energy* **39** 123
[20] Felici F. *et al* 2018 *Nucl. Fusion* **58** 096006
[21] Pajares A. and Schuster E. 2021 *Fusion Eng. Des.* **170** 112526
[22] Morosohk S. *et al* 2021 *Nucl. Fusion* **61** 106040
[23] Citrin J. *et al* 2015 *Nucl. Fusion* **55** 092001
[24] Hinton G.E. 1990 *Connectionist Learning Procedures in Machine Learning* (Elsevier) pp 555–610
[25] Rafiq T. *et al* 2013 *Phys. Plasmas* **20** 032506
[26] van de Plassche K.L. *et al* 2020 *Phys. Plasmas* **27** 022310
[27] Rath K. *et al* 2022 *J. Plasma Phys.* **88** 895880502
[28] Meneghini O. *et al* 2017 *Nucl. Fusion* **57** 086034
[29] Meneghini O. *et al* 2020 *Nucl. Fusion* **61** 026006
[30] Bonotto M., Abate D. and Pigatto L. 2024 *Fusion Eng. Des.* **200** 114193
[31] Boyer M., Kaye S. and Erickson K. 2019 *Nucl. Fusion* **59** 056008

[32] Smirnov A., Harvey R. and Kupfer K. 1994 *Bull. Am. Phys. Soc.* **39** 1626
[33] Harvey R. and McCoy M. 1992 *Proc. IAEA TCM on Advances Simulation and Modeling of Thermonuclear Plasmas* (*Montreal, Canada*, *1992*) pp 489–526
[34] Wallace G. *et al* 2022 *J. Plasma Phys.* **88** 895880401
[35] Wallace G.M. *et al* 2023 *AIP Conf. Proc.* **2984** 090008
[36] Breiman L. 2001 *Mach. Learn.* **45** 5
[37] Geurts P., Ernst D. and Wehenkel L. 2006 *Mach. Learn.* **63** 3
[38] Liu H. *et al* 2020 *IEEE Trans. Neural Netw. Learn. Syst.* **31** 4405
[39] Brambilla M. 1999 *Plasma Phys. Control. Fusion* **41** 1
[40] Brambilla M. 2002 *Plasma Phys. Control. Fusion* **44** 2423
[41] Brambilla M. and Bilato R. 2006 *Nucl. Fusion* **46** S387
[42] Ono M. 1995 *Phys. Plasmas* **2** 4075
[43] Stix T.H. 1975 *Nucl. Fusion* **15** 737
[44] Perkins F.W. 1977 *Nucl. Fusion* **17** 1197
[45] Budny R. 2009 *Nucl. Fusion* **49** 085008
[46] Taylor G. *et al* 2012 *Phys. Plasmas* **19** 042501
[47] Bucalossi J. *et al* 2022 *Nucl. Fusion* **62** 042007
[48] Kazakov Y.O., Fülöp T. and Van Eester D. 2013 *Nucl. Fusion* **53** 053014
[49] Bertelli N., Shiraiwa S. and Ono M. 2022 *Nucl. Fusion* **62** 126046
[50] Maquet V. *et al* 2023 arXiv:2301.10754
[51] Stein M. 1987 *Technometrics* **29** 143
[52] Wold S., Esbensen K. and Geladi P. 1987 *Chemometr. Intell. Lab. Syst.* **2** 37
[53] Bishop C.M. and Nasrabadi N.M. 2006 *Neural Networks for Pattern Recognition* vol 4 (Springer)
[54] Brambilla M. and Ottaviani M. 1985 *Plasma Phys. Control. Fusion* **27** 1
[55] Stix T.H. 1992 *Waves in Plasmas* (American Institute of Physics)
[56] Brambilla M. and Bilato R. 2024 Private communication
[57] Bertelli N. *et al* 2017 *Nucl. Fusion* **57** 056035
[58] Shiraiwa S., Wright J., Lee J. and Bonoli P. 2017 *Nucl. Fusion* **57** 086048
[59] Pedregosa F. *et al* 2011 *J. Mach. Learn. Res.* **12** 2825
[60] Kingma D.P. and Ba J. 2014 arXiv:1412.6980