Interpretability guided transfer learning approaches for tritium pedestal predictions

Downloaded from: https://research.chalmers.se, 2026-09-05 05:06 UTC

Citation for the original published paper (version of record):
Gillgren, A., Yadykin, D., Strand, P. (2026). Interpretability guided transfer learning approaches for tritium pedestal predictions. Plasma Physics and Controlled Fusion, 68(6).
http://dx.doi.org/10.1088/1361-6587/ae6dfa

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Plasma Physics and
Controlled Fusion

PURPOSE-LED
PUBLISHING™

PAPER • **OPEN ACCESS**

Interpretability guided transfer learning approaches for tritium pedestal predictions

To cite this article: A Gillgren *et al* 2026 *Plasma Phys. Control. Fusion* **68** 065007

View the article online for updates and enhancements.

You may also like

- Introducing Machine-Learning in Spectroscopy for Plasma Diagnostics and Predictions
M Koubiti and M Kerebel
- Physics-informed deep learning model for line-integral diagnostics across fusion devices
Cong Wang, Weizhe Yang, Haiping Wang et al.
- Optimisation of physics-informed neural network architecture and training for tokamak equilibrium reconstruction
Novella Rutigliano, Andrea Murari, Pasquale Gaudio et al.

This content was downloaded from IP address 129.16.68.206 on 11/06/2026 at 12:24

Plasma Physics and Controlled Fusion

Check for updates

OPEN ACCESS

RECEIVED
16 February 2026

REVISED
27 April 2026

ACCEPTED FOR PUBLICATION
14 May 2026

PUBLISHED
29 May 2026

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

PAPER

Interpretability guided transfer learning approaches for tritium pedestal predictions

A Gillgren*, D Yadykin, P Strand and JET Contributors[1]

Chalmers University of Technology, Gothenburg, Sweden

[1] See Maggi *et al* (https://doi.org/10.1088/1741-4326/ad3e16) for JET Contributors.

* Author to whom any correspondence should be addressed.

E-mail: andreas.gillgren@chalmers.se

Keywords: fusion, pedestal, isotope, tritium, machine learning, interpretability, transfer learning

Abstract

We explore transfer learning approaches to extend data-driven pedestal models trained on deuterium (D) plasmas to tritium (T) and DT mixtures. Specifically, we use models pre-trained on JET D pulses, and JET T/DT data for the transfer learning. We use model interpretability to guide our choice of transfer learning strategy. Analysis of model behavior post transfer learning reveals that sparsity and multicollinearity in the T/DT data lead to severe overfitting when fine-tuning the weights of the pre-trained neural network-based models. Therefore, we instead use a more robust and simple output calibration approach to facilitate the impact of isotope composition. This yields models with R^2 between 0.66-0.87, performing significantly better than uncalibrated models though not matching the original D-only performance. The scaling coefficients obtained qualitatively agree with previous research, namely that the pedestal density scales positively with increased isotope mass, while pedestal temperature exhibits a weak negative scaling with isotope mass. This work highlights the importance of understanding model behavior in transfer learning to ensure reasonable functional mappings, which is particularly relevant for fusion research where sparse, multicollinear data are encountered.

1. Introduction

A fuel mixture of deuterium (D) and tritium (T) is a leading candidate for future fusion reactors like ITER due to its high reaction rate under reactor-relevant conditions. Therefore, an important area of research focuses on how plasma properties in magnetic confinement devices change as the isotope composition shifts from being D-dominated, the most studied case in existing machines like JET, to DT mixtures or T-dominant [1–4].

One approach to studying such isotope scalings is through simulations, which typically involve solving first-principles-based equations using numerical methods. However, there are certain aspects of the plasma in fusion devices that remain challenging to accurately simulate from first-principles, such as the pedestal region in tokamak devices. The pedestal is characterized by steep temperature and density gradients that form near the plasma edge when the input power exceeds a certain threshold [5, 6]. While the exact physics behind this phenomena is still a subject of active research, the consensus points to a self-organizing mechanism involving sheared flows and turbulence suppression. Specifically, near the edge of the plasma, a strong gradient in the radial electric field develops, which creates a velocity shear that breaks up turbulent transport structures. This leads to a more steep pressure gradient that strengthens the radial electrical field and the transport suppression in a feedback loop. Therefore, the pedestal keeps building until the pressure gradient exceeds a threshold governed by instabilities that trigger edge localized modes (ELMs), or other transport mechanisms that restrict further build-up of the pedestal. Because heat and particle transport is suppressed when there is a developed pedestal, this operational mode is referred to as a high-confinement mode (H-mode), as opposed to the non-pedestal low-confinement mode (L-mode).

© 2026 The Author(s). Published by IOP Publishing Ltd

The temperature and density at the top of the steep region, the pedestal top values, strongly influence the temperature and density of the plasma core. This is because core transport is stiff, meaning that the maximum allowable kinetic gradient lengths are relatively fixed. Hence, to raise the core pressure, the pressure at the edge (the pedestal pressure) needs to be increased. Because the pedestal impacts the core, limited predictive capability for the pedestal inherently limits the ability to reliably simulate overall plasma behavior in H-mode, including cases where the impact of isotope composition is of interest.

As an alternative to first-principles-based pedestal models [7–11], previous work has demonstrated the feasibility of predicting pedestal top values using data-driven models trained on experimental data [12–15]. For example, in [12], an interpretable neural network-based method called *NeuralBranch* was used to predict the pedestal top at JET while transparently revealing how the pedestal depends on key tokamak parameters. However, this model was only trained on D plasmas due to the significantly smaller amount of data available for other isotope compositions in the JET pedestal database [16]. There is therefore no guarantee that the same level of accuracy will be achieved when directly applying the model presented in [12], or those in [13–15] to T/DT plasmas, especially since prior research reports a shift in pedestal characteristics with changing isotope composition, for instance at JET and AUG [1, 2, 17–20]. There are ongoing efforts aimed at capturing the isotope impact on the pedestal in first-principles models. However, previous work show, for instance, that the established KBM constraint is insufficient in describing the isotope impact on the pedestal width [2]. Moreover, [2] demonstrates examples of the necessity to include resistive physics to reproduce experimentally observed changes in the pedestal when changing isotope composition, where the inclusion of resistive physics entails high computational demand [21]. Hence, data-driven approaches may serve as a useful complement, both because they provide fast predictions once trained, but also to potentially highlight qualitative discrepancies between models trained on experimental data and first-principles based models.

1.1. Scope of work

In this work, we explore transfer learning approaches for scaling data-driven pedestal models trained on experimental D plasmas to improve their predictive performance on T and DT mixture plasmas. Specifically, we use the pre-trained NeuralBranch models from [12], which were trained on D data from the JET pedestal database. For scaling to T/DT predictions, we use data entries from the JET DTE2 experimental campaign [3] included in a newer version of the database. For completeness, we also investigate isotope scaling using hydrogen (H) plasma data entries.

This work has two main goals. First, we demonstrate how model interpretability can be used to analyze the behavior of a machine learning model post transfer learning, which enables us to make informed choices regarding appropriate scaling approaches. Second, the more specific aim is to extend the applicability of the existing data-driven pedestal prediction tool first introduced in [12] by incorporating the impact of isotope composition.

2. Dataset

All of the data used in this work comes from the JET pedestal database [16], which was created prior to this work. The database contains pedestal parameters derived using curve-fitting techniques on temperature and density measurements accumulated during steady-state time intervals. The specific parameters we predict in this work are the pedestal top values of the electron density and temperature just before ELMs are triggered. These pedestal top values are illustrated in figure 1.

More details on the creation of the database, including the handling of experimental measurements and the curve-fitting techniques used to obtain radial profiles can be found in [16].

2.1. Data filtering

As in [12], we here exclude data entries that are associated with impurity seeding or ELM mitigating techniques (RMPs, kicks, and pellets), due to the impact that these techniques have on the pedestal. Moreover, all of the included data entries are from experiments that were run with the beryllium/tungsten wall at JET (the ITER-like wall). Another feature is that the dataset consists primarily of Type-I ELMs, and we have excluded pulses that show characteristics of other ELM types (as defined in [6]). That said, we cannot rule out that a small fraction may represent other ELM types. This fraction is likely too small to have noticeable impact on the trained models, but may cause occasional significant prediction error. For instance, a model might overpredict on a small ELM sample since it is biased towards predicting large type-I ELM pedestals.

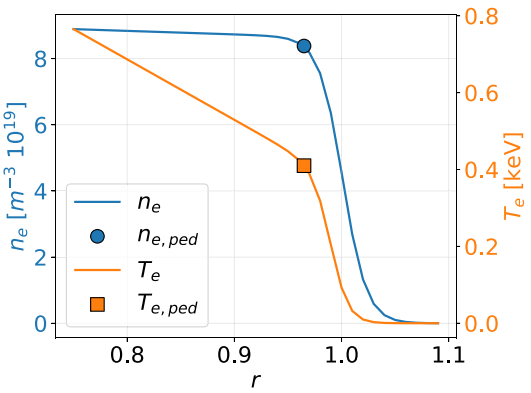


Figure 1. Sketches of the electron density n_e and temperature T_e profiles in the plasma edge region. Here, the radial coordinate r is normalized, with $r = 1$ representing the last closed flux surface (LCFS). The pedestal top values $n_{e,ped}$ and $T_{e,ped}$, indicated at the top of the steep regions, are the two values we are predicting with the models in this work.

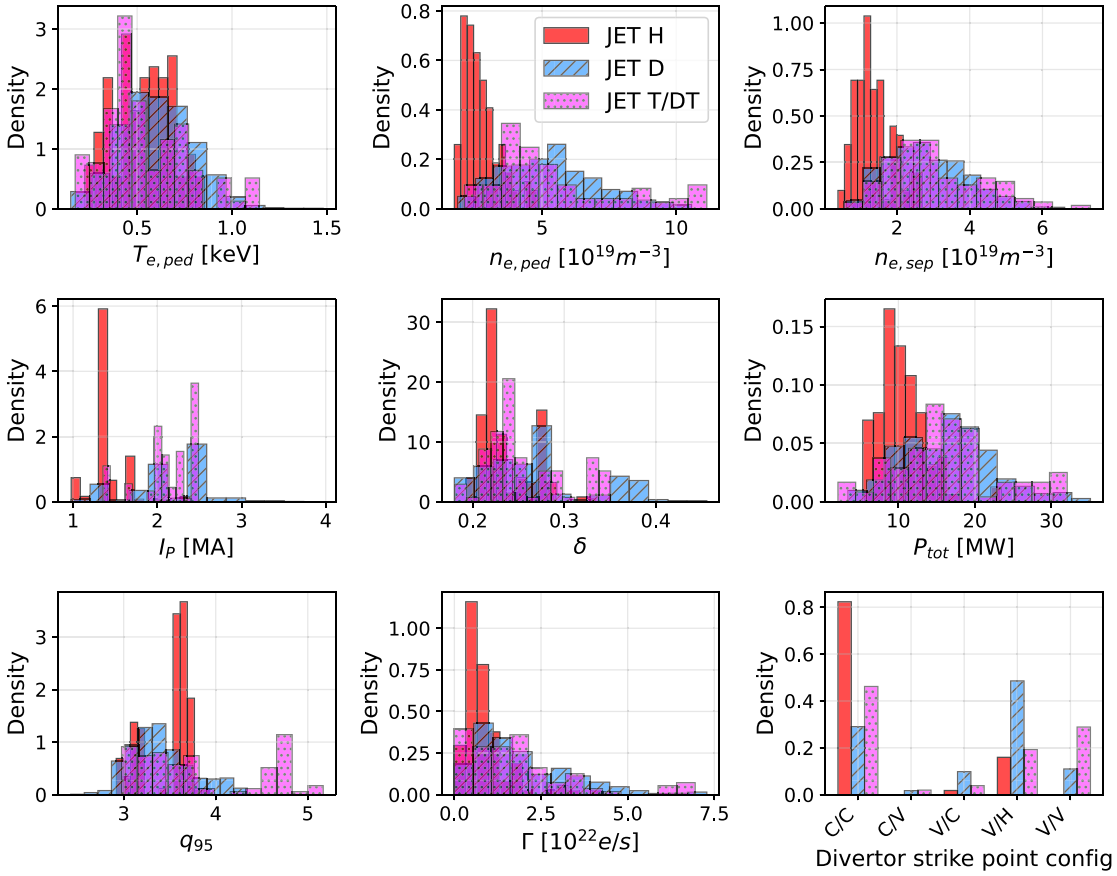


Figure 2. Distribution histograms of key parameters in the different D, H, T/DT subsets of the data. Due to the difference in data size, the distributions are normalized such that the area under each individual curve is 1. In the divertor strike point configuration subplot, acronym 'V' refers to a vertical target, 'C' to a corner target, and 'H' to a horizontal target. For instance, 'V/C' indicates a configuration where the inner strike point is on the vertical target and the outer strike point is on the corner target.

2.2. Dataset distributions

In total, 1043 deuterium (D) entries were used in the pre-training of the NeuralBranch models in [12]. In figure 2, the distribution of this data with respect to key parameters is shown along with the distributions of the T/DT data (105 entries), and the H data (112 entries). Here, we show a joint distribution of the T and DT data to improve readability, given that additional observations we conducted indicate qualitatively similar distributions for these two subsets (with the exception that some of the pure T pulses are associated with higher gas fueling compared to the DT data). The key parameters in figure 2 include the pedestal top values $T_{e,ped}$ and $n_{e,ped}$ since these are the prediction outputs in this work. Key parameters also include the inputs of the pre-trained models: separatrix density $n_{e,sep}$, plasma current

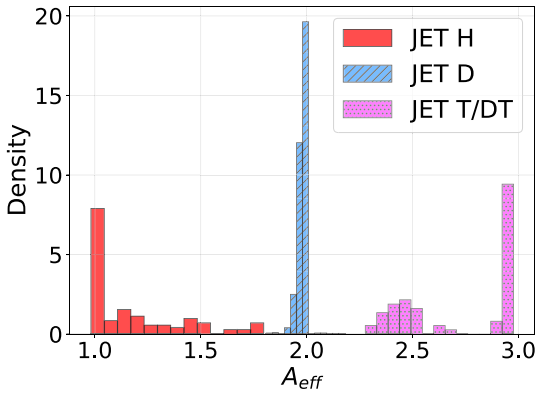


Figure 3. Normalized distributions of the effective mass A_{eff} in the three data subsets (after excluding high q_{95} entries from the T/DT subset).

I_{P}, triangularity δ, total input power P_{tot} (NBI + ICRH + Ohmic—Shine through), and gas fuel rate of the main ion Γ. Here, the triangularity is the average of the upper and lower triangularity of the plasma. Figure 2 also includes the distribution of the safety factor q_{95} and the divertor strike point configuration. The reason for displaying the distribution of q_{95} is due to a significant portion of the T/DT data showing higher q_{95} values compared to the D pre-training data. Therefore, to isolate the effect of changing the isotope composition from D to T, we choose to not include the T/DT entries where $q_{95} > 4.35$ when testing the different scaling strategies (4.35 being the maximum q_{95} value in the D data). This reduces the dataset size of the T/DT data from 105 to 71.

We see in figure 2 that the T/DT data shows similar distributions compared to the D data. However, the same cannot be said for the H data, which generally shows much more narrow distributions compared to the D data. This reflects the inherently narrower operational window for H-mode access in hydrogen plasmas, which constrains the achievable parameter space. However, this inhibits our ability to perform isotope scaling using the H data, as fine-tuning on a narrow parameter space can cause the model to misattribute changes to isotope impact rather than to the specific characteristics and biases of that narrow domain. Indeed, as will be seen in section 5, using the H data for scaling can result in misleading isotope dependencies that contradict previous research. Hence, in the remainder of the paper we will focus on scaling using the T/DT data, and explore how the T/DT data can be used to potentially extrapolate to H data.

Figure 3 shows the effective mass A_{eff} distributions for the three data subsets (after excluding high q_{95} entries from the T/DT subset). Essentially, A_{eff} describes the ion content (for instance, $A_{\text{eff}} = 2$ for pure D plasmas, $A_{\text{eff}} = 2.5$ for a 50/50 DT mix, and $A_{\text{eff}} = 3$ for pure T plasmas). These distributions show that while D/H and D/T mixes represent a non-negligible fraction, the highest data density occurs at the pure H, D, and T samples.

3. Pre-trained pedestal models

The models we use as the base for scaling in this work are NeuralBranch models [12], which means that they are neural network-based, but arranged in sub-networks (branches) to enable interpretability (to avoid the black-box problem). Before we describe in more detail how NeuralBranch models work, we first introduce the four specific NeuralBranch models from [12] that we consider in this work:

(i) Model 1: Predicts $n_{\text{e,ped}}$ from I_{P}, δ, $n_{\text{e,sep}}$

(ii) Model 2: Predicts $n_{\text{e,ped}}$ from I_{P}, δ, Γ, P_{tot}

(iii) Model 3: Predicts $T_{\text{e,ped}}$ from I_{P}, P_{tot}, $n_{\text{e,sep}}$

(iv) Model 4: Predicts $T_{\text{e,ped}}$ from I_{P}, δ, Γ, P_{tot}

Two models for each output are considered to provide flexibility depending on whether $n_{\text{e,sep}}$, which acts as a better proxy for the neutral pressure compared to the gas fueling Γ [16], and therefore improves the accuracy of models predicting the pedestal as shown in [10–12], is available or not as an input (it might

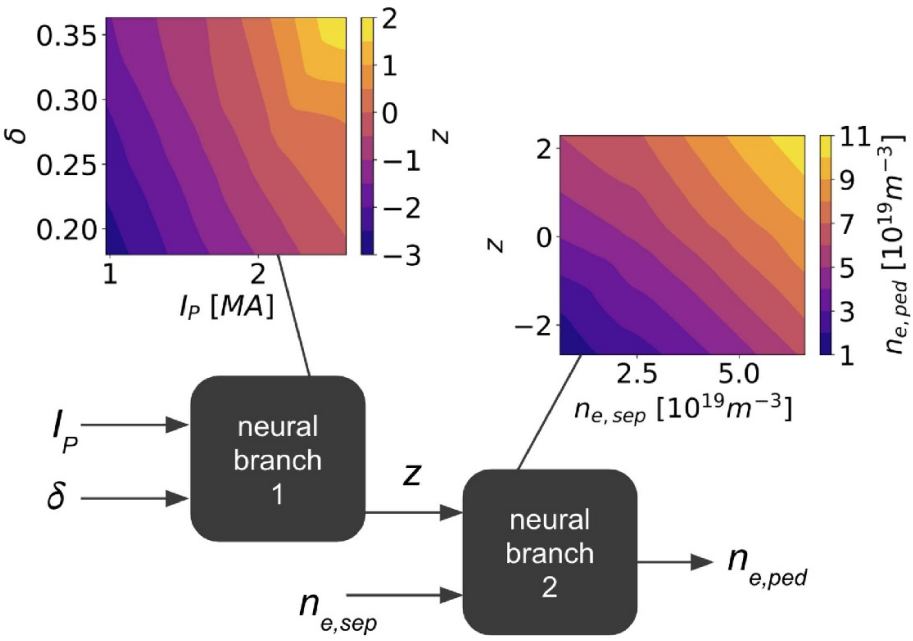


Figure 4. The NeuralBranch architecture and associated interpretable visualizations for Model 1 (predicting $n_{e,ped}$ from I_P, δ and $n_{e,sep}$, first introduced in [12]). Each neural branch (illustrated with boxes) contains fully connected neural network layers. The visualizations show the mappings that were learned when the model was trained on deuterium data in [12].

not always be available in future use cases as it is not a strict engineering parameter). Additionally, as the main goal in [12] was to analyze the input-output relationships, it was relevant to examine how the parameter dependencies changed when $n_{e,sep}$ was excluded, forcing the models to rely on inputs like the fuel rate Γ instead (which is of interest in this work as well).

As described in [12], the inputs ($n_{e,sep}$ and the others) were selected based on a feature importance method. In summary, the input candidates from [12] that were found to not enhance accuracy in any of the models include: toroidal field B, minor radius a, elongation κ, plasma volume, safety factor q_{95}, effective ion charge Z_{eff}. This does not necessarily mean that these parameters universally are unimportant for the pedestal. Rather, it means that generally across the available JET D dataset, these parameters do not contribute with unique information useful for predicting the pedestal that is not already embedded in the other inputs.

3.1. NeuralBranch method

Neural networks and other high capacity machine learning models are known for being black boxes, meaning that the qualitative mapping from input to output that have emerged during training is difficult to interpret. The NeuralBranch framework resolves this by arranging neural network layers in separate sub-networks that are called neural branches, where each neural branch only has two inputs and one output, as illustrated for Model 1 in figure 4. Since a neural branch only has two inputs, its output can be plotted versus its two inputs, which provides a qualitative approach for interpreting how the output of a neural branch depends on its inputs. For instance, in figure 4, we first interpret how the intermediate/latent variable z depends on I_P and δ, and then we interpret how the model output $n_{e,ped}$ depends on z and $n_{e,sep}$, which together reveals the full mapping of the model. In this particular case, $n_{e,ped}$ increases when both $n_{e,sep}$ and z increases, and z increases when I_P and δ increases. Hence, the model displayed in figure 4 shows, without going into nuanced details, a positive relationship between $n_{e,ped}$ and all of the three inputs.

Note that the latent variable z does not need to be known beforehand, as the neural branches are trained jointly as a single model, with z being shaped during the process. Other aspects of the NeuralBranch framework, such as the method used to determine the branch architecture and the strategy for appropriately assigning inputs to each neural branch, are detailed in [12]. Full architectural details and analyses of Models 2–4 are also available in [12]. Below, we provide a brief summary:

- **Model 1: $[I_P, \delta, n_{e,sep}] \rightarrow n_{e,ped}$**
  I_P, δ, and $n_{e,sep}$ are all positively correlated with the output $n_{e,ped}$. There is a slight positive interaction between I_P and δ, meaning that higher I_P amplifies the impact that δ has on $n_{e,ped}$, and that higher δ amplifies the impact that I_P has on $n_{e,ped}$.

- **Model 2: $[I_P, \delta, \Gamma, P_{tot}] \rightarrow n_{e,ped}$**
  I_P, δ, and Γ are positively correlated with the output $n_{e,ped}$, and P_{tot} is negatively correlated with $n_{e,ped}$. However, the negative impact of P_{tot} on $n_{e,ped}$ saturates at high P_{tot} and the impact of Γ on $n_{e,ped}$ is slightly stronger at lower Γ. The same positive interaction between I_P and δ that is present in Model 1 is also present here.

- **Model 3: $[I_P, P_{tot}, n_{e,sep}] \rightarrow T_{e,ped}$**
  I_P and P_{tot} are positively correlated with the output $T_{e,ped}$, and $n_{e,sep}$ is negatively correlated with $T_{e,ped}$. However, there is also an attenuating interaction: when I_P is high, the impact of P_{tot} and $n_{e,sep}$ on $T_{e,ped}$ weakens. Conversely, when P_{tot} is high and $n_{e,sep}$ is low, the impact of I_P on $T_{e,ped}$ weakens.

- **Model 4: $[I_P, P_{tot}, \Gamma, \delta] \rightarrow T_{e,ped}$**
  I_P and P_{tot} are positively correlated with the output $T_{e,ped}$, and Γ and δ are negatively correlated with $T_{e,ped}$. The same attenuating interaction that is present in Model 3 is also present here, with the exception that Γ effectively replaces $n_{e,sep}$.

4. Isotope scaling

In this section, we discuss three alternatives for isotope scaling from D to T/DT predictions: 1) classical transfer learning in the sense of fine-tuning the weights of the pre-trained neural network-based models, 2) simple calibration applied on top of the predictions of the pre-trained models, and 3) simple calibration on top of low-capacity power scalings. We first discuss alternative 1, and in particular why fine-tuning of the weights is not suitable for our case.

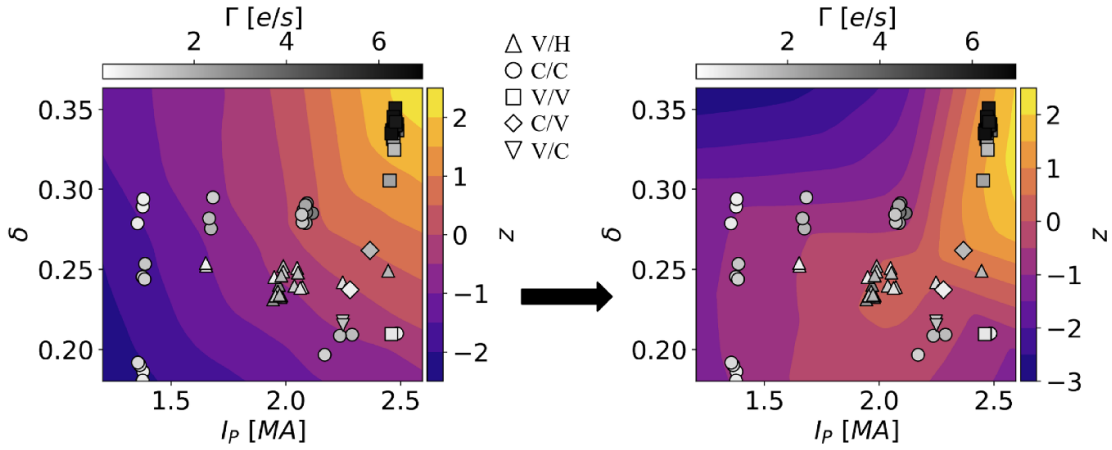
4.1. Alternative 1: fine-tuning of weights

Fine-tuning of weights for transfer learning can in many scenarios be a powerful technique, such as for fine-tuning foundational language or vision models that are pre-trained through, for instance, self-supervised learning [22, 23]. Although modern vision and language models operate at an entirely different scale than the models in this paper, starting from a favorable initial condition is a property we may attempt to exploit here as well. Specifically, the strategy is to make very small adjustments to an already reasonable mapping grounded in the diverse and more populated D data. Ideally, if only minor adjustments are required to achieve high accuracy, the model remains in close proximity to the pre-trained D model. This proximity may act as a form of regularization to maintain a more stable mapping in the regions between sparse T/DT data points than would be possible when training from scratch.

However, multicollinearities between the inputs in the T/DT data makes the fine-tuning approach unsuitable for our case, which we exemplify using Model 2, which predicts $n_{e,ped}$ from I_P, δ, P_{tot} and Γ. In figure 5, the mapping of the neural branch that has the inputs I_P and δ in Model 2 is shown, along with scatter points of the T/DT data. Before the fine-tuning, the neural branch shows a consistent pattern coherent with previous results obtained using D data [16]. Specifically, the branch output z, which for Model 2 is proportional to the final output $n_{e,ped}$, increases with increased I_P and δ, where there also is an amplifying interaction between I_P and δ. This coherent and consistent pattern is likely a result of the diverseness in the larger deuterium dataset used to initially train the model. However, after we have fine-tuned the model sufficiently to increase the accuracy on the T/DT data, we observe undesirable behavior due to the data characteristics:

- There is high class separability in the divertor strike point configuration in I_P/δ space, where the V/H and V/V points are strongly clustered. This leads to, for instance, undesirable artifacts like the peninsula-like pattern (in color) enclosing the V/H points (triangles pointing upwards), as these are associated with higher $n_{e,ped}$.

- Other parameters also exhibit problematic multicollinearity. For example, as shown in figure 5, all entries with the highest fueling Γ are concentrated at the highest δ and I_P values. Consequently, when the model is fine-tuned, it becomes ambiguous whether high output values of $n_{e,ped}$ should be attributed to high I_P/δ or Γ. This ambiguity can lead the model to over- or underestimate the true influence of these parameters.



a) Before fine-tuning b) After fine-tuning

Figure 5. Visualization of the mapping of the neural branch that takes I_P and δ as inputs in Model 2 (the model predicting $n_{e,ped}$ from I_P, δ, Γ, and P_{tot}). The mapping of the neural branch is shown before (a) and after (b) fine-tuning of the weights of the model using the T/DT data (scatter points). The marker style (divertor strike point configuration) and marker color (Γ) are used to illustrate problematic multicollinearities in the T/DT data. z (shown in color) is the output of this neural branch, which for Model 2 is approximately proportional to the final model output $n_{e,ped}$.

- The amount of fine-tuning required to achieve high accuracy has caused the to model behave unreasonably outside the T/DT data distribution, in regions that were previously covered by the D data. For instance, as shown in figure 5, there is a region at high δ and low I_P that is not populated by the T/DT data. After fine-tuning, the output z, and hence $n_{e,ped}$, scales negatively with increased δ in this region, which is inconsistent with previous research and just an undesirable artifact from the fine-tuning process. It is not uncommon for machine learning models to be unreliable outside their training distribution, but it is a drawback that the usable domain shrinks due to very small weight adjustments not being sufficient.

Is it noteworthy that all these three issues can be accompanied with a misleading high prediction accuracy on a validation set (which we observe during testing). For instance, the peninsula (color) that forms due to the V/H data (see figure 5) will be well suited for predicting on other held-out V/H data that occupies the peninsula. In other words, a good validation set prediction accuracy alone is not sufficient since we also value the qualitative behavior of the model.

The extent to which the issues listed above arise during fine-tuning depends on aspects such as the learning rate, number of training iterations, train/validation data split, freezing certain neural network layers, and regularization techniques, such as enforcing monotonic dependencies with respect to the inputs. Through experimentation with these aspects, we did not identify a solution that sufficiently addresses these qualitative issues while achieving a prediction accuracy surpassing the more simple and robust method that will be presented next.

Another aspect worth emphasizing is that we treat all of the T/DT entries as if they have the same A_{eff} value in this section. This choice was made to simplify the demonstration of how the multicollinearity discourages the fine-tuning approach, which persists whether or not the model includes a branch accounting for specific A_{eff} values in the DT range.

In summary, the interpretability aspect of the NeuralBranch framework reveals limitations in the fine-tuning approach for this application, that would have remained hidden when relying solely on black-box model performance metrics. Moreover, although we have demonstrated the limitations of the fine-tuning approach using only one pre-trained model, the fundamental limitations imposed by the T/DT data characteristics are inherent to all four cases.

4.2. Alternative 2: simple output calibration

A more simple approach to perform isotope scaling is to freeze the entire pre-trained model, and to perform an adjustment on top of the prediction, which effectively serves as an output calibration. For instance, using Model 1 as an example, one of the simplest possible approaches would be to add an offset c to the pre-trained model

$$n_{e,ped}^{T/DT} = n_{e,ped}^{D}\left(I_P, \delta, n_{e,sep}\right) + c, \tag{1}$$

Table 1. R^2 values of the different calibration methods applied to the different pre-trained models.

	R^2: Model 1 $n_{e,ped}(n_{e,sep})$	R^2: Model 2 $n_{e,ped}(\Gamma)$	R^2: Model 3 $T_{e,ped}(n_{e,sep})$	R^2: Model 4 $T_{e,ped}(\Gamma)$
$y^T = y^D$ (no calibration)	0.80	0.58	0.62	−0.38
$y^T = y^D + c$	0.84	0.69	0.65	−0.15
$y^T = y^D \cdot c$	0.86	0.72	0.66	0
$y^T = y^D \cdot c_0 A_{eff}^{c_1}$	0.87	0.72	0.66	0.10
$y^T = y^D \cdot \left(\frac{A_{eff}}{2}\right)^c$	0.87	0.71	0.66	−0.14
Original R^2 on D data	0.88	0.84	0.82	0.66

or to scale the output of the pre-trained model through multiplication

$$n_{e,ped}^{T/DT} = c \cdot n_{e,ped}^{D}\left(I_P, \delta, n_{e,sep}\right), \tag{2}$$

where c in both cases is optimized to minimize, in our case, the mean squared error on the T/DT data. Here, the pre-trained model $n_{e,ped}^{D}$ remains entirely unmodified.

Given the characteristics of the T/DT data, this simpler output calibration approach presents certain advantages over the fine-tuning method:

- Since only one or a few parameters are optimized, it reduces the risk of overfitting to specific characteristics of the T/DT data, such as the separation between divertor strike point configuration classes.
- Optimizing few parameters also makes it feasible to fit and evaluate the model on the entire T/DT dataset, which eliminates the need for the train/validation split required in the fine-tuning approach. This is particularly beneficial for the relatively small T/DT set, as it prevents results from being affected by the variability introduced through train/validation splits.

We do however also acknowledge that preserving the relationship between the outputs ($n_{e,ped}$, $T_{e,ped}$) and the inputs (I_P, $n_{e,sep}$, δ, Γ, P_{tot}) learned during training on the D data may have drawbacks. In particular, such a model would be unable to capture genuine changes in parameter interactions that arise when the plasma becomes more T dominated. For instance, in [1] and [18], it is observed that isotope composition impacts how much the gas fueling affects the pedestal, which is a parameter interaction a simple output calibration cannot capture in detail. Moreover, pedestal stability analysis performed in [2] suggests that the isotope impact gets stronger with increased resistivity. Hence, the purpose of a simple output calibration is to approximate the average impact of isotope composition rather than capturing all of the more intricate interactions.

The simple output calibration can be done in several ways, as exemplified by equation (1) and equation (2). The choice ultimately depends on the desired properties and the accuracy of the model. One of the properties we may introduce is sensitivity to the effective mass A_{eff}. For instance, taking inspiration from power scalings commonly used in the field of fusion research [16, 24–26], an alternative is

$$n_{e,ped}^{T/DT} = n_{e,ped}^{D} \cdot c_0 A_{eff}^{c_1}, \tag{3}$$

where we have omitted the notation that $n_{e,ped}^{D}$ depends on the other inputs to improve readability.

Another property that can be introduced is the enforcement $n_{e,ped}^{T/DT} \to n_{e,ped}^{D}$ as $A_{eff} \to 2$. This can be achieved by modifying equation (3)

$$n_{e,ped}^{T/DT} = n_{e,ped}^{D} \cdot \left(\frac{A_{eff}}{2}\right)^{c}. \tag{4}$$

In table 1, we display the accuracy of the different simple calibration techniques applied to the different pre-trained models. We use the coefficient of determination R^2 to evaluate our models, where $R^2 = 1$ represents a perfect model, $R^2 = 0$ represents a model with the same predictive capacity as a model that simply predicts the mean of the dataset, and where $R^2 < 0$ represents a model that is worse than a model that predicts the mean of the dataset. The main takeaways are:

- There is always an improvement compared to not calibrating the output.
- The models that have $n_{e,sep}$ as an input performs significantly better.

- Enforcing the properties that the calibration should be A_{eff}-dependent, and $y^{\text{T}} \to y^{\text{D}}$ as $A_{\text{eff}} \to 2$ (equation (4)), appears to not degrade performance (except for Model 4, but no alternative works well for Model 4).
- All models except Model 1 shows a significantly lower accuracy compared to when the models were originally evaluated on the D data.

It is not surprising that the models that have $n_{\text{e,sep}}$ as an input performs better when going from D to T plasmas, since $n_{\text{e,sep}}$ itself has previously been observed to depend on the isotope composition [2]. In other words, to an extent the change in isotope composition is embedded in $n_{\text{e,sep}}$, which leads to the models performing fairly well as long as they already are capable of mapping $n_{\text{e,sep}}$ (together with the other inputs) to $n_{\text{e,ped}}$ and $T_{\text{e,ped}}$.

Another unsurprising result is that Model 4 has the lowest overall accuracy, as it was also the least accurate pre-trained model when originally evaluated on the D data [12]. It appears that $T_{\text{e,ped}}$ is harder to predict accurately from machine parameters when $n_{\text{e,sep}}$ is excluded, which is only amplified when considering T plasmas, so much so that simply predicting the mean $T_{\text{e,ped}}$ in the T/DT dataset would yield an equally good result. We revisited the fine-tuning alternative to see if there was any way to improve the accuracy of Model 4, but we found no solution that surpassed $R^2 = 0.15$ while not showing signs of severe overfitting.

The fact that all models except Model 1 shows a lower performance on the T data compared to D data, even with the calibration, suggest that the parameter interactions learned from the D data likely change to some extent when the plasma becomes more tritium dominated. However, due to the restrictions imposed by the T/DT data, we are unable to make confident conclusions about the details of these hypothetical changes.

Out of the different calibration methods, the alternative that best aligns with our objective is the one based on equation (4). Specifically, this approach allows us to enforce $y^{\text{T}} \to y^{\text{D}}$ as $A_{\text{eff}} \to 2$ without sacrificing accuracy significantly. Additionally, this approach allows for a A_{eff}-dependency in the calibration, which is a property we desire in order to capture variations in the DT-mix range. With the fitted coefficients explicitly given, the output calibrated models based on equation (4) are:

$$n_{\text{e,ped}}^{\text{T/DT}} = n_{\text{e,ped}}^{\text{D}}\left(n_{\text{e,sep}}, I_{\text{P}}, \delta\right) \cdot \left(\frac{A_{\text{eff}}}{2}\right)^{0.30\pm0.16} \tag{5}$$

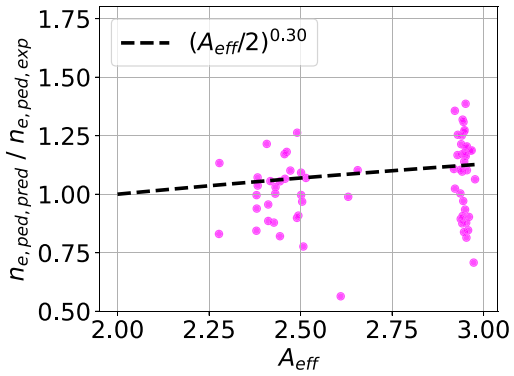
$$n_{\text{e,ped}}^{\text{T/DT}} = n_{\text{e,ped}}^{\text{D}}\left(I_{\text{P}}, \delta, P_{\text{tot}}, \Gamma\right) \cdot \left(\frac{A_{\text{eff}}}{2}\right)^{0.41\pm0.16} \tag{6}$$

$$T_{\text{e,ped}}^{\text{T/DT}} = T_{\text{e,ped}}^{\text{D}}\left(n_{\text{e,sep}}, I_{\text{P}}, P_{\text{tot}}\right) \cdot \left(\frac{A_{\text{eff}}}{2}\right)^{-0.18\pm0.15}. \tag{7}$$

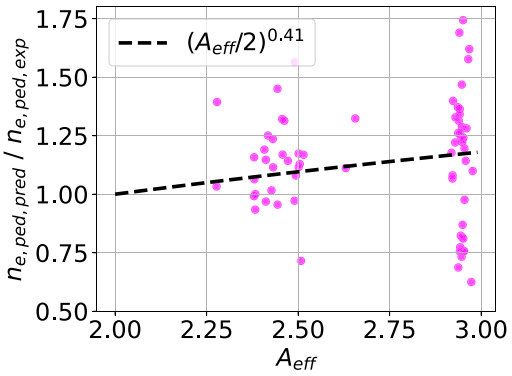
Here (and in the rest of the paper), the uncertainty in the coefficients are obtained by measuring how sensitive the accuracy (R^2) is to the coefficients. Specifically, we sweep the coefficients and use a tolerance of $\Delta R^2 = 0.02$ to estimate the uncertainty band (in other words, the numbers do not represent actual uncertainty but rather estimates based on sensitivity). With a less strict tolerance, larger uncertainty band estimates would have been obtained.

Note that we are not presenting the result for Model 4 due to its low associated accuracy, which makes it unreliable for analysis purposes and usage. (see table 1). However, the other model predicting $T_{\text{e,ped}}$ (Model 3) shows a modest negative correlation with effective mass A_{eff}, consistent with previous research suggesting a slight negative [2] or even negligible [20] dependency between $T_{\text{e,ped}}$ and A_{eff}. Furthermore, both models that predict $n_{\text{e,ped}}$ indicate a more distinct positive scaling with increased A_{eff}, which is also consistent with previous research [1, 2, 18, 20, 27]. The $n_{\text{e,ped}}$ model that includes $n_{\text{e,sep}}$ as an input (equation (5)) shows a weaker dependence compared to its counterpart, which is expected since, as previously discussed, the impact of the isotope is likely already partially embedded in $n_{\text{e,sep}}$.

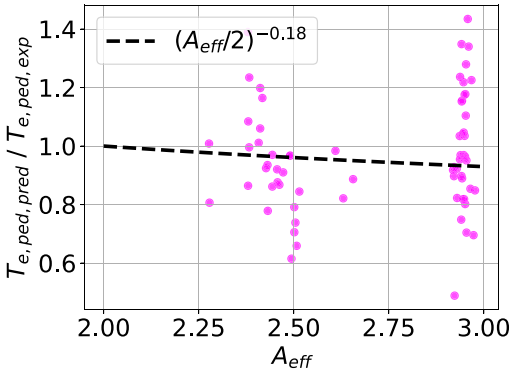
Figure 6 shows the predictions of the NeuralBranch models (Model 1-3) on the T/DT subset in more detail. Specifically, the ratio of the predictions (before calibration) to the actual experimental values are shown, along with the multiplicative calibration factors in (5)–(7). As illustrated, the spread of the ratio (which is closely related to relative error) is noticeably larger than the systematic shift captured by the calibrations. In other words, the impact of A_{eff} found by the calibrations is relatively small in comparison to other compounding factors that cause predictions to deviate from the experimental values. This



(a) Model 1.



(b) Model 2.



(c) Model 3.

Figure 6. The points represent the ratio of the predictions made by the NeuralBranch models (before calibration) to the actual experimental pedestal values in the T/DT data. The dashed lines represent the calibration scalings (multiplicative factors in (5)–(7)) employed to remove the systematic bias (the average deviation of this ratio from unity) as the effective mass A_{eff} increases.

result justifies the use of simple and robust approaches for the A_{eff}-calibration, since more complex alternatives like neural networks would not only offer marginal accuracy gains, but would also introduce significant instability. Specifically, the calibration would become highly sensitive to the stochasticity of validation splits, increasing the risk of the model overfitting to coincidental correlations or noise in sparse regions of the A_{eff} parameter space. A simpler approach ensures the calibration remains consistent and less prone to such variances.

4.3. Alternative 3: output calibration on power scalings

The hypothesis and indication of changed parameter interactions when transitioning from D to T, as discussed in the previous section, raises another question. If the more complicated parameter dependencies and interactions learned from the D data are less accurate or relevant for the T data, can we still justify using a more complex base model trained on the D data like NeuralBranch over a simpler model such as a power scaling? To answer this question, we try a third approach, which is to use power scalings trained on D data that were presented together with the NeuralBranch models in [12], together with the same output calibration that we found in the previous section, namely the calibration based on equation (4). This gives us the models:

$$n_{\text{e,ped}}^{\text{T/DT}} = 3.71 n_{\text{e,sep}}^{0.46} I_{\text{P}}^{0.58} \delta^{0.42} \cdot \left(\frac{A_{\text{eff}}}{2}\right)^{0.37\pm0.16} \tag{8}$$

$$n_{\text{e,ped}}^{\text{T/DT}} = 10.59 I_{\text{P}}^{1.15} \delta^{0.68} P_{\text{tot}}^{-0.26} \Gamma^{0.11} \cdot \left(\frac{A_{\text{eff}}}{2}\right)^{0.46\pm0.16} \tag{9}$$

$$T_{\text{e,ped}}^{\text{T/DT}} = 0.18 n_{\text{e,sep}}^{-0.60} P_{\text{tot}}^{0.48} I_{\text{P}}^{0.54} \cdot \left(\frac{A_{\text{eff}}}{2}\right)^{-0.30\pm0.15} \tag{10}$$

where these achieve the R^2-values: 0.87, 0.66, 0.49 respectively, whereas Model 4 still achieve an accuracy too low ($R^2 < 0$) to be presented here. In these power scalings, only the exponents of the effective mass A_{eff} are estimated by minimizing the mean squared error on the T/DT data, while the remaining exponents are taken directly from [12] (power scalings trained on D data). We avoid fitting all exponents directly to the T/DT data because, although the limited dataset size alone might still allow for such a fit, the strong multicollinearities among the inputs would lead to severe variance inflation in the coefficient estimates. This is confirmed by analyzing the variance inflation factor (VIF) value for the inputs, which is an estimate of how predictable an input is from the others. Specifically,

$$\text{VIF}_i = \frac{1}{1 - R_i^2} \tag{11}$$

where R_i^2 is the coefficient of determination of predicting an input i from the other inputs using a linear regression model. High multicollinearity leads to high R^2 values, which in turn lead to high VIF values. By fitting directly on the T/DT data, we obtain VIF values exceeding 10 for four of the five inputs (up to 57 for I_{P}), indicating severe multicollinearity. This ill-conditioning of the problem would make the exponents statistically unstable and unreliable, and therefore not suitable for usage once the model is fitted.

By comparing the R^2 values of the power scalings with output calibration (equations (8)–(10)) with the ones from the NeuralBranch models with output calibration, see table 1, we see that the only power scaling that matches the accuracy of the NeuralBranch model is the first case (equation (8)). This is likely because Model 1, the NeuralBranch model for $n_{\text{e,ped}}(n_{\text{e,sep}}, I_{\text{P}}, \delta)$, is the one that shows the least complicated interactions in [12], which makes it the most similar to a power scaling. However, the two remaining power scalings (equations (9) and (10)) do not achieve the same accuracy as the corresponding NeuralBranch models. This suggests that, although the more complicated parameter interactions and dependencies learned for D plasmas are likely altered by the increased tritium content, they remain at least partially relevant.

5. Fitting on hydrogen data results in misleading A_{eff} dependencies

We demonstrate here how utilizing the H data for calibration can yield non-physical isotope dependencies, likely stemming from the restricted parameter space (see figure 2) and inherent systematic biases in the H-mode discharges.

Applying the output calibration method to Model 3, but now for the H data instead of the T/DT data, yields

$$T_{\text{e,ped}}^{H} = T_{\text{e,ped}}^{\text{D}}\left(n_{\text{e,sep}}, I_{\text{P}}, P_{\text{tot}}\right) \cdot \left(\frac{A_{\text{eff}}}{2}\right)^{0.16 \pm 0.07}, \tag{12}$$

which shows a positive dependency with respect to A_{eff}, contradicting the negative dependency obtained when calibrating on the T/DT data (equation (7)). There is no established physical mechanism suggesting that pedestal temperatures should peak specifically for deuterium, and previous studies [2, 20] do not support that $T_{\text{e,ped}}$ increases with increased A_{eff} when the other parameters are held constant.

Crucially, it must be acknowledged that the A_{eff} exponents in both the H (12) and T/DT (7) calibrations are weak (partly due to $n_{\text{e,sep}}$ incorporating the effect of A_{eff} to an extent). Given the limited size of both the H and T/DT datasets, we cannot exclude the possibility that the observed variations in these expressions are a consequence of statistical instability (small datasets can produce exponents that are highly sensitive to the specific samples included).

However, other power scalings in the literature [16] fit on the JET H-mode hydrogen data also show a similar positive dependency between $T_{\text{e,ped}}$ and A_{eff} that seem to contradict other research [2, 20]. Hence, we argue that these results (ours and the ones in [16]) possibly reflect the specific constraints of the H-mode domain in hydrogen rather than a pure isotope effect. The systematic overpredictions of $T_{\text{e,ped}}$ on the H data could be due to, for instance, the majority of H samples being associated with a C/C divertor strike point configuration, bias towards low $n_{\text{e,sep}}$ and low I_{P}, or narrow ranges in other parameters that speculatively may have a weaker or stronger impact on H pedestals compared to their impact on D pedestals (that simple calibrations cannot fully capture).

Because these confounding factors cannot be statistically decoupled in the hydrogen subset due to the lack of data diversity, we conclude that we cannot justify using the H data for reliable model calibration (which also applies for $n_{\text{e,ped}}$). Furthermore, this limitation is intrinsic to the available data distribution

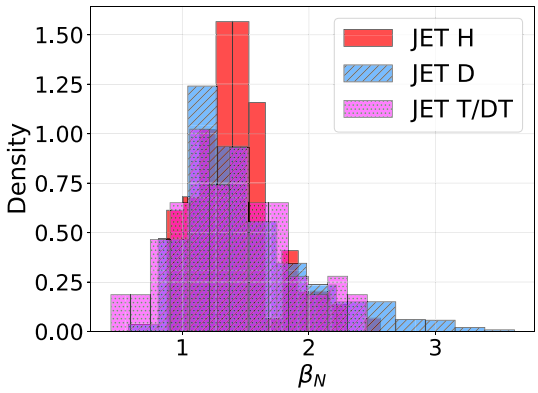


Figure 7. Distributions of β_N in the D, T/DT, and H subsets used in this paper.

and cannot be bypassed through, for instance, changing the combination of inputs, as the background biases will still be present.

We can however investigate how well the models calibrated on the T/DT data extrapolate to the H data. The model that predicts $T_{e,ped}$ with $n_{e,sep}$ as an input (equation (7)), yields $R^2 < 0$ on the H data. Even the corresponding calibration based on the biased H data (12) yields a low score ($R^2 = 0.43$). The models predicting $n_{e,ped}$ (equations (5) and (6)) yield $R^2 = 0.49$ (with $n_{e,sep}$) and $R^2 = 0.51$ (with Γ and P_{tot} instead of $n_{e,sep}$) on the H data. In summary, we conclude that the models calibrated on the T/DT data do not extrapolate particularly well to the H data. This is likely both due to how the true relationships change when going from H to T, beyond what a simple calibration can capture, but also possibly due to the evaluation being done on a narrowly distributed dataset with intrinsic biases.

6. Replacing P_{tot} with β_N as an input

Pedestal prediction models that rely solely on machine parameters known prior to experiments offer practical advantages for operational planning. However, for integrated modeling applications, incorporating plasma parameters as inputs can facilitate coupling between the pedestal and the plasma core. We therefore present alternative models that replace the total heating power P_{tot} with the global plasma β_N as an input parameter. Among the input parameters, P_{tot} is the logical choice for replacement since β_N reflects the plasma pressure achieved in response to heating. Moreover, β_N is known to impact pedestal stability via the Shafranov shift [28], and it serves as a core-edge coupling variable in established integrated modeling frameworks. As an example, the first-principles pedestal model EPED [7–9], used in the OMFIT integrated modeling framework [29], takes β_N as an input rather than heating power. This enables iterative core-pedestal calculations where the core transport codes provide β_N to the pedestal model, which in turn returns boundary conditions back to the core until convergence to self-consistent solutions is reached. The reason for excluding P_{tot}, as opposed to keeping both P_{tot} and β_N as inputs, is that we find that they provide redundant information. Specifically, when P_{tot} is included, β_N does not improve prediction accuracy, which means that the models may learn to ignore β_N, which in turn would defeat the purpose of its inclusion.

Before creating models that include β_N as an input, we first check the distribution of β_N in the dataset. As can be seen in figure 7, the distributions of β_N overlap between the D and T/DT subsets (and also the H data). Hence, we can be more confident that when we apply transfer learning from D to T/DT plasmas, we are not entangling the impact of isotope mass with the effect of changing β_N. Moreover, while there is some multicollinearity between β_N and the other inputs, we observe no linear correlation scores surpassing the highest ones (in magnitude) between the other inputs reported in [12] (the most significant being -0.59 between β_N and $n_{e,sep}$).

When pre-training NeuralBranch models and power scalings on D data, we find that β_N does not contribute to accuracy when predicting $n_{e,ped}$ (even when P_{tot} is removed). Hence, there is no point in training $n_{e,ped}$ models that include β_N, because such models will have no incentive to learn meaningful relationships for β_N. However, when training models to predict $T_{e,ped}$, we find that β_N is able to compensate for the removal of P_{tot}. Specifically, in the case where $n_{e,sep}$ is included, the same accuracy is achieved when β_N replaces P_{tot} ($R^2 = 0.82$). In the case where $n_{e,sep}$ is not included as an input,

replacing P_{tot} with β_N yields a slight improvement ($R^2 : 0.66 \rightarrow 0.70$). Moreover, when β_N replaces P_{tot}, we observe approximately the same accuracy when using simple power scalings and more complicated machine learning models. This suggests that the more complicated input parameter interactions found in [12] that involve P_{tot} are no longer present when β_N is included instead (in [12], more complicated models were needed to capture these interactions). Hence, now when β_N is included, there is no motivation to use a more complicated model compared to the simple power scaling approach. Note that so far, these results only concern pre-training on the D data.

To obtain models that incorporate isotope scaling, we first fit power scaling expressions for all inputs except the effective mass A_{eff} on the D data. We then fit the coefficient for A_{eff} on the T/DT data. This yields the results

$$T_{\text{e,ped}}^{\text{T/DT}} = (0.29 \pm 0.07)\, n_{\text{e,sep}}^{-0.39\pm0.13} \beta_N^{0.63\pm0.18} \cdot I_{\text{p}}^{1.13\pm0.21} \cdot \left(\frac{A_{\text{eff}}}{2}\right)^{-0.37\pm0.14} \tag{13}$$

$$T_{\text{e,ped}}^{\text{T/DT}} = (0.17 \pm 0.05)\, \beta_N^{0.81\pm0.16} I_{\text{p}}^{0.98\pm0.23} \cdot \delta^{-0.20\pm0.10} \Gamma^{-0.07\pm0.04} \cdot \left(\frac{A_{\text{eff}}}{2}\right)^{-0.54\pm0.15} \tag{14}$$

where (13) achieves $R^2 = 0.80$ on the D data (0.69 on T/DT data), and where (14) achieves $R^2 = 0.69$ on the D data (0.51 on T/DT data). In other words, when β_N replaces P_{tot}, we obtain models that show some predictive capability of $T_{\text{e,ped}}$ for T/DT plasmas, even when $n_{\text{e,sep}}$ is excluded as an input (when using P_{tot}, we never surpass $R^2 = 0.15$ when $n_{\text{e,sep}}$ is excluded). For completeness, we additionally find that both (13) and (14) extrapolate poorly to the H data ($R^2 = 0.26$ when $n_{\text{e,sep}}$ is included, and $R^2 = 0.05$ when $n_{\text{e,sep}}$ is excluded).

By analyzing the coefficients in (13) and (14), we observe that $T_{\text{e,ped}}$ scales positively with β_N. Moreover, we observe that $T_{\text{e,ped}}$ scales negatively with increased effective mass A_{eff}, similarly to when P_{tot} is included. The other inputs show similar dependencies compared to the models where P_{tot} is included.

7. Conclusions

The scope of this work was to explore transfer learning approaches for scaling data-driven pedestal models originally trained on experimental JET D data to JET T/DT data. Beyond the specific goal of extending the applicability of an already existing pedestal model [12], the goal was also to demonstrate how interpretability can be leveraged to reveal the qualitative change in model behavior post fine-tuning.

The interpretability aspect of the models revealed that multicollinearity clusters in the relatively small T/DT dataset caused overfitting when fine-tuning the weights of neural network-based models (which would not have been obvious by just analyzing validation set accuracy alone). Therefore, the fine-tuning of weights approach was deemed inappropriate for our problem, which also likely applies to other problems beyond that of the pedestal where JET T/DT data is used to adjust models.

As an alternative approach, we used a more robust and simple output calibration technique inspired by power scalings to incorporate the impact of isotope scaling. This resulted in moderate to good accuracy on the T/DT data (R^2 in the range 0.66–0.87), with the exception of the model predicting $T_{\text{e,ped}}$ without $n_{\text{e,sep}}$ as an input ($R^2 < 0$). The isotope scaling coefficients obtained indicate positive correlation between pedestal density $n_{\text{e,ped}}$ and isotope mass A_{eff}, and weak negative correlation between pedestal temperature $T_{\text{e,ped}}$ and A_{eff}, which is consistent with previous research [1, 2, 18, 20, 27]. Moreover, our conclusion is that the simple output calibration technique captures a decent approximation of the average isotope impact on the pedestal, but not the full picture, which is supported by the more intricate parameter interaction changes with changed isotope composition reported in [1, 2, 18], and the fact that the pre-trained models are more accurate on the D data.

In addition to scaling with respect to T/DT data, we demonstrated why scaling with respect to the H data is problematic. Specifically, the hydrogen H-mode domain is more narrowly accessible at JET compared the D and T/DT cases, which leads to a narrow data domain that contains biases. Hence, when scaling with respect to the H data, we cannot disentangle the effects of the characteristics of the narrow H-mode domain and the isotope composition. As shown, this can lead to isotope dependencies that contradict previous research.

Finally, we created alternative models for $T_{\text{e,ped}}$ where the input power P_{tot} was replaced with β_N as an input. The purpose of this replacement was to facilitate core-edge coupling for when the models are

implemented in simulation frameworks in the future. Qualitatively similar results were obtained compared to when P_{tot} is used as an input. A benefit of using β_N as an input is that the $T_{e,ped}$ model that does not include $n_{e,sep}$ as an input gains moderate predictive power ($R^2 < 0 \rightarrow R^2 = 0.51$).

Aside from implementation of models in simulation frameworks, future work includes analysis of how the pedestal empirically depends on the isotope composition on large scale datasets (not just a few pulses) at other machines. Future work also includes qualitative comparisons between predictions from theory-based frameworks and models trained on experimental data, with the purpose of highlighting discrepancies.

In summary, this work demonstrates the importance of interpretable models for understanding model behavior in transfer learning tasks, which in our case guided us to use a more simple scaling strategy. Given that sparse data with multicollinearities are encountered in magnetic confinement fusion research, this approach may be important for other tasks within the field beyond pedestal studies.

Acknowledgments

The authors would like to thank Lorenzo Frassinetti for valuable discussions and for providing the pedestal dataset. The authors would also like to thank Boel Brandström for providing useful feedback on the manuscript.

This work has been carried out within the framework of the EUROfusion Consortium, funded by the European Union via the Euratom Research and Training Programme (Grant Agreement No 101052200- EUROfusion). Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them. Moreover, the work has been funded by the Swedish Research Council under the diary No. 2020-05465.

Data availability statement

The data cannot be made publicly available upon publication because they are owned by a third party and the terms of use prevent public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Author contributions

A Gillgren 0000-0002-3810-2913
Conceptualization (lead), Data curation (lead), Formal analysis (lead), Investigation (lead), Methodology (lead), Software (lead), Visualization (lead), Writing – original draft (lead)

D Yadykin
Supervision (equal), Writing – original draft (supporting)

P Strand 0000-0002-8899-2598
Funding acquisition (lead), Project administration (lead), Resources (lead), Supervision (equal)

References

[1] Schneider P A *et al* (JET Contributors) 2023 Isotope physics of heat and particle transport with tritium in JET-ILW type-I ELMy H-mode plasmas *Nucl. Fusion* **63** 112010
[2] Frassinetti L *et al* (JET Contributors) 2023 Effect of the isotope mass on pedestal structure, transport and stability in D, D/T and T plasmas at similar ßn and gas rate in JET-ILW type I ELMy H-modes *Nucl. Fusion* **63** 112009
[3] Maggi C F 2024 Overview of T and D–T results in JET with ITER-like wall *Nucl. Fusion* **64** 112012
[4] Weisen H *et al* 2020 Isotope dependence of energy, momentum and particle confinement in tokamaks *J. Plasma Phys.* **86** 905860501
[5] Wagner F *et al* 1982 Regime of improved confinement and high beta in neutral-beam-heated divertor discharges of the ASDEX tokamak *Phys. Rev. Lett.* **49** 1408–12
[6] Fenstermacher M E *et al* 2025 Progress in pedestal and edge physics: chapter 3 of the special issue: on the path to tokamak burning plasma operation *Nucl. Fusion* **65** 053001
[7] Snyder P B, Groebner R J, Leonard A W, Osborne T H and Wilson H R 2009 Development and validation of a predictive model for the pedestal heighta) *Phys. Plasmas* **16** 056118

[8] Snyder P B, Groebner R J, Hughes J W, Osborne T H, Beurskens M, Leonard A W, Wilson H R and Xu X Q 2011 A first-principles predictive model of the pedestal height and width: development, testing and iter optimization with the eped model *Nucl. Fusion* **51** 103016
[9] Saarelma S *et al* (JET Contributors) 2019 Self-consistent pedestal prediction for JET-ILW in preparation of the DT campaign *Phys. Plasmas* **26** 072501
[10] Saarelma S *et al* (JET Contributors) 2023 Testing a prediction model for the H-mode density pedestal against JET-ILW pedestals *Nucl. Fusion* **63** 052002
[11] Saarelma S *et al* (STEP team, JET Contributors and the Eurofusion Tokamak Exploitation Team) 2024 Density pedestal prediction model for tokamak plasmas *Nucl. Fusion* **64** 076025
[12] Gillgren A, Ludvig-Osipov A, Yadykin D and Strand P (JET contributors) 2025 Investigating pedestal dependencies at JET using an interpretable neural network architecture *Nucl. Fusion* **65** 056033
[13] Gillgren A, Fransson E, Yadykin D, Frassinetti L, Strand P and Contributors J E T 2022 Enabling adaptive pedestals in predictive transport simulations using neural networks *Nucl. Fusion* **62** 096006
[14] Kit A, Järvinen A E, Frassinetti L and Wiesen S (JET Contributors) 2023 Supervised learning approaches to modeling pedestal density *Plasma Phys. Control. Fusion* **65** 045003
[15] Järvinen A E, Kit A, Poels Y R J, Wiesen S, Menkovski V, Frassinetti L and Dunne M (ASDEX Upgrade Team and JET Contributors) 2024 Representation learning algorithms for inferring machine independent latent features in pedestals in JET and aug *Phys. Plasmas* **31** 032508
[16] Frassinetti L *et al* (JET contributors) 2020 Pedestal structure, stability and scalings in JET-ILW: the eurofusion JET-ILW pedestal database *Nucl. Fusion* **61** 016001
[17] Laggner F M *et al* (EUROfusion MST1 Team and ASDEX Upgrade Team) 2017 Pedestal structure and inter-elm evolution for different main ion species in ASDEX upgrade *Phys. Plasmas* **24** 056105
[18] Schneider P A, Hennequin P, Bonanomi N, Dunne M, Conway G D and Plank U (the ASDEX Upgrade Team and the EUROfusion MST1 Team) 2021 Overview of the isotope effects in the ASDEX upgrade tokamak *Plasma Phys. Control. Fusion* **63** 064006
[19] Schneider P A *et al* (the ASDEX Upgrade Team, the EUROfusion MST1 Team and JET Contributors) 2021 The dependence of confinement on the isotope mass in the core and the edge of aug and JET-ILW H-mode plasmas *Nucl. Fusion* **62** 026014
[20] Maggi C F *et al* (JET Contributors) 2017 Isotope effects on l-h threshold and confinement in tokamak plasmas *Plasma Phys. Control. Fusion* **60** 014045
[21] Nyström H, Frassinetti L, Saarelma S, Huijsmans G T A, von Thun C P, Maggi C F and Hillesheim J C (JET contributors) 2022 Effect of resistivity on the pedestal MHD stability in JET *Nucl. Fusion* **62** 126045
[22] Brown T B et al 2020 Language models are few-shot learners (arXiv:2005.14165)
[23] Kaiming H, Zhang X, Ren S and Sun J 2015 Deep residual learning for image recognition (arXiv:1512.03385)
[24] Cordey J G (for the ITPA H-Mode Database Working Group and the ITPA Pedestal Database Working Group) 2003 A two-term model of the confinement in Elmy H-modes using the global confinement and pedestal databases *Nucl. Fusion* **43** 670
[25] (ITER Physics Expert Group on Confinement and Transport, ITER Physics Expert Group on Confinement Modelling, Database and ITER Physics Basis ed) 1999 Chapter 2: plasma confinement and transport *Nucl. Fusion* **39** 2175
[26] Martin Y R and Takizuka T (the ITPA CDBM H-mode Threshold Database Working Group) 2008 Power requirement for accessing the H-mode in ITER *J. Phys.: Conf. Ser.* **123** 012033
[27] Horvath L *et al* 2021 Isotope dependence of the type I ELMY H-mode pedestal in JET-ILW hydrogen and deuterium plasmas *Nucl. Fusion* **61** 046015
[28] Snyder P B *et al* 2004 Elms and constraints on the h-mode pedestal: peeling-ballooning stability calculation and comparison with experiment *Nucl. Fusion* **44** 320–8
[29] Meneghini O *et al* (The AToM Team) 2015 Integrated modeling applications for tokamak experiments with OMFIT *Nucl. Fusion* **55** 083008