Divertor Power Load Predictions Based on Machine Learning

M Brenzke[1], S Wiesen[1], M Bernert[2], D Coster[2], J Jitsev[3], Y Liang[1], U von Toussaint[2], ASDEX Upgrade team[4] and EUROfusion MST1 team[5]

[1]Forschungszentrum Juelich, Institut fuer Energie- und Klimaforschung - Plasmaphysik, 52425 Juelich, Germany
[2]Max Planck Institute for Plasma Physics, 85748 Garching, Germany
[3]Juelich Supercomputing Center (JSC), Institute for Advanced Simulation (IAS), Research Center Juelich (FZJ), 52425 Juelich, Germany
[4]See the author list of H. Meyer et al. 2019 Nucl. Fusion 59 112014
[5]See the author list of B. Labit et al. 2019 Nucl. Fusion 59 086020

E-mail: m.brenzke@fz-juelich.de

Abstract. Machine learning based data-driven approaches to thermal load prediction on the divertor targets of ASDEX Upgrade are presented. After selecting time averaged data from almost six years of operation of ASDEX Upgrade and applying basic physics-motivated cuts to the data we find that we are able to train machine learning models to predict a scalar quantifying the steady state thermal loads on the outer divertor target given scalar operational parameters. With both random forest and neural network based models we manage to achieve decent agreement between the model predictions and the observed values from experiments. Furthermore, we investigate the dependencies of the models and observe that the models manage to extract trends expected from previous physics analyses.

1. Introduction

Modeling thermal loads expected at the divertor targets during a discharge is a crucial task for the operation of current and future fusion devices. The thermal loads at the divertor targets must not exceed critical values in order to limit damage caused to the divertor material (for the limits for ITER see e.g. [1]). Both, predictions of expected thermal loads as well as on-line power load control (cf. [2]) during a discharge are integral to guarantee safe operation. Analytical models for the former task do exist. However, these models only apply in the regime of attached plasma conditions, which is not the envisioned regime for the operation of future fusion devices [3]. To include the regime of detached plasma operation, more sophisticated methods are required such as the coupling of fluid and Monte Carlo simulation codes [4]. Such codes commonly have the disadvantages of very long runtime until convergence and requiring approximations

and manual selection of parameters to match simulations and experimental data.
These circumstances hinder detailed large scale studies of power exhaust in fusion devices. Moreover, existing models and simulations require parameters that are not necessarily well known before performing a discharge, such as the power crossing the separatrix, or a model for anomalous (turbulent) transport which introduce further approximations into the analysis.
To mitigate the hindrances of both analytical models and fully fledged simulation codes, we present machine learning approaches for data-driven construction of surrogate models. The presented models have the advantage of causing significantly less computational overhead than simulation codes once they are set up and allow for more complex dependencies than simplified analytical models.
Machine learning methods have been used in fusion physics in the past, mainly with regard to the prediction of disruptions in tokamaks (cf. [5], [6]) but also for other applications such as reconstructing magnetic configurations in stellerators [7] or to speed up turbulent transport simulations [8]. In this work we focus on a supervised learning approach linking vectors of scalar plasma parameters to scalar divertor conditions during regular operation to predict expected thermal loads at the divertor targets. To this end we tested different machine learning approaches to find and investigate models describing the dependencies of thermal loads at the divertor targets on more general plasma parameters. We present the results obtained with a random forest approach and an approach based on fully-connected, multi-layer neural networks. We train and test the models once with a larger set of input variables and once with a reduced set of input variables limited to direct and indirect engineering parameters. The models with the reduced set of inputs could be used to plan a discharge with regard to divertor conditions and as a lightweight method to assess operational constraints.
The following sections firstly describe the data base and how we constructed it from the experimental data including first analysis steps. In section 3.1 we explain how we set up and analysed a model based on random forests and evaluate its main dependencies. In the following section 3.2 we present our neural network based approach and the results we achieved including the model's main dependencies. Finally we draw conclusions and give an outlook on potential further work in section 4.

2. Data Base and Data Selection

For the presented analysis we utilized the large experimental data base of ASDEX Upgrade (AUG) with a full-W metal wall with the motivation of potentially extracting additional physics insights from the data instead of limiting our models to replicating the performance of any given simulation code which could also be used to train and evaluate the models. We investigated over 7000 discharges carried out between 2012 and 2018 (discharge numbers 28000 to 35489). Of all these discharges we selected the ones carried out in lower single null configuration and neglected discharges with reversed toroidal magnetic field as the data base contained only a few such discharges.

We averaged all relevant signals over 0.2 s time intervals if the plasma current was stable within these and the following 0.2 s in order to extract data from the flattop phase of a discharge and to exclude disruptions from the analysis which are identified by an abrupt reduction of the plasma current. We defined a stable plasma current by requiring that no data point within the time window had a deviation larger than 10% of the mean value from the mean within the time frame. The time averaged values are indicated either by $< \cdot >$ or by the prefix *mean*.

The value of 0.2 s was chosen since a typical time for the energy transport in AUG is about 50 ms. Thus, given a change in the core plasma, the divertor conditions should usually reach a steady state within the selected time frame. Short transient events in this process (such as ELMs) are not further investigated in this analysis. Therefore, signals such as the radiated power will still be effected by such events. The parameter to be predicted in this analysis is an AUG specific quantity called T_{div} [9]. We chose this quantity since it was proven to be a good real-time estimate of the electron temperature in the outer divertor assuming detachment of the inner divertor and because it is commonly recorded for every discharge. The quantity that is actually measured in the experiments is the thermo-electric current to the divertor (cf. [10]). Therefore, since the T_{div} measurement does not directly probe the electron temperature, the quantity might also reach negative values. Small T_{div} values around zero can be an indicator of detachment in the outer divertor characterized by a reduction of the peak heat flux into the divertor [11]. T_{div} is an ELM-filtered quantity and a good proxy for steady state thermal loads at the divertor target.

The input quantities to the models and the ranges of their values in the training data are listed in table 1. We chose these parameters because they are standardly available and comparable between different tokamaks with the exception of the strike line position. The comparability between tokamaks could be interesting to investigate scalings between different reactors. Furthermore, these parameters are most likely to have a significant influence on the divertor conditions. Some parameters show values that lie outside the range of typical operational conditions, such as a very small plasma current or density. These data points either come from the operational regime of the ramp-up or ramp-down of the plasma current that we do not filter out completely (cf. figure 1) or from discharges that probed less common operational scenarios. For example, the lowest heating power comes from a drop in heating power in a discharge (number 35135) with very low plasma current and an overall low heating power.

All values have been extracted from the measurements performed during the experiments. Note that the density measurements have not necessarily been corrected for fringe jumps. We used all the quantities in the table to set up models linking these parameters to T_{div} predictions. Furthermore, we also set up models using only the reduced number of bold-faced inputs in table 1 which are direct and indirect engineering parameters and are accessible for predictions prior to performing a discharge. Hence, this reduced parameter set resembles the actual conditions of a pulse planning phase more closely than the full set of inputs despite the fact that the values used were still

Quantity	Symbol	Value Range
plasma current	$\mathbf{I}_p$	[0.2, 1.2] MA
toroidal magnetic field	$\mathbf{B}_t$	[-3.2, -0.99] T
heating power	$\mathbf{P}_{tot}$	[0.096, 20] MW
radiated power	$\mathrm{P}_{rad,tot}$	(0, 17.9] MW
core electron density (line-integrated)	**H-1**	$[2.21{\cdot}10^{16}, 2.86{\cdot}10^{20}]\ \frac{1}{m^2}$
edge electron density (line-integrated)	H-5	$[6.98{\cdot}10^{13}, 2.86{\cdot}10^{20}]\ \frac{1}{m^2}$
neutral density in divertor	$\mathrm{n}_{div,ist}$	$[0, 2.19{\cdot}10^{22}]\ \frac{1}{m^3}$
stored energy	W_{MHD}	$[0.49, 1.25{\cdot}10^{6}]$ J
lower triangularity	δ_{untn}	[0.021, 0.54]
upper triangularity	δ_{oben}	[-0.085, 0.49]
elongation	κ	[1.05, 1.96]
strike line position	Suna2b	[0, 1.26] m
hydrogen throughput	$\mathbf{H}_{tot}$	$[0, 3.54{\cdot}10^{22}]\ \frac{el.}{s}$
deuterium throughput	$\mathbf{D}_{tot}$	$[0, 8.13{\cdot}10^{22}]\ \frac{el.}{s}$
helium throughput	$\mathbf{He}_{tot}$	$[0, 8.83{\cdot}10^{21}]\ \frac{el.}{s}$
neon throughput	$\mathbf{Ne}_{tot}$	$[0, 5.03{\cdot}10^{21}]\ \frac{el.}{s}$
nitrogen throughput	$\mathbf{N}_{tot}$	$[0, 4.41{\cdot}10^{22}]\ \frac{el.}{s}$

Table 1. Input parameters of the models; bold-faced parameters were used for models with reduced set of inputs; the value range gives the minimum and maximum value in the training data after applying the selection criteria to the data

extracted from measurements during the discharges.

To remove erroneous measurements we applied a number of filters to the extracted data. We removed all data points with obviously wrong measurements, such as negative density measurements in the plasma core or the edge region, a confined energy of ≤ 0 J or radiated power of ≤ 0 W. Moreover, we required the fraction of total radiated power over heating power to be smaller than one and the total deposited heating power to be in the range (0, 20] MW. T_{div} was limited to [-5, 30] eV, motivated by the reliability of T_{div} as an estimate of the electron temperature in the outer divertor which only holds under the assumption of a detached inner divertor since the measured current is actually determined by the temperature difference between inner and outer divertor [12]. Since large values of T_{div} can be indicative of an attached inner divertor, data points outside the given range of T_{div} were removed from this analysis. This might also limit the reliability of the models' predictions to the given range.

After applying these cuts, data from about 4500 separate discharges remained. Furthermore, all input features were standardized to zero mean and unit variance for training and testing the models using the statistics obtained from the training data.

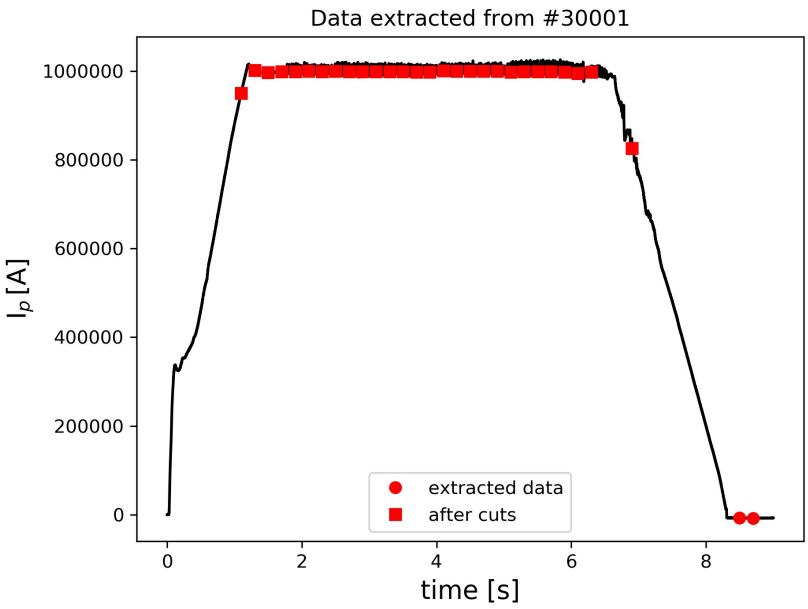


Figure 1. Extraction of data points from time traces exemplified for one discharge; red dots mark data points fulfilling the stability criterion of I_p; squares indicate data points remaining after additional cuts

Figure 1 shows an example of the data extraction process. The black line is the time trace of the plasma current measured during the discharge. The red dots are data points fulfilling the stability criterion imposed on the plasma current and red squares indicate data points remaining after applying the aforementioned additional cuts to the data. It is obvious that we manage to primarily select data points from the flattop phase, the regime this analysis focuses on, with the rather simple criteria of our data selection. However, some data points are also extracted from the ramps in the plasma current.

To set up and evaluate the models we split the data set into a training and a test set. For this we randomly selected 70% of the discharges and assigned them to the training set. All other discharges were assigned to the test set.

To analyse the data split into training and test set we investigated the distributions of the target quantity for both data sets. Figure 2 shows this distribution of extracted T_{div} values for both the training and the test data. Both distributions are peaked around 10 eV and show a similar shape which indicates a reasonable division of training and test data.
We further analysed the data by determining the pairwise Pearson correlation coefficient between all quantities from the training data set. The resulting correlation matrix is depicted in figure 3. The strongest correlations of T_{div} that we observe are a negative correlation coefficient of -0.3 with the neutral density in the divertor, rudimentarily comparable to the results obtained in [10] with the nitrogen concentration in the divertor,

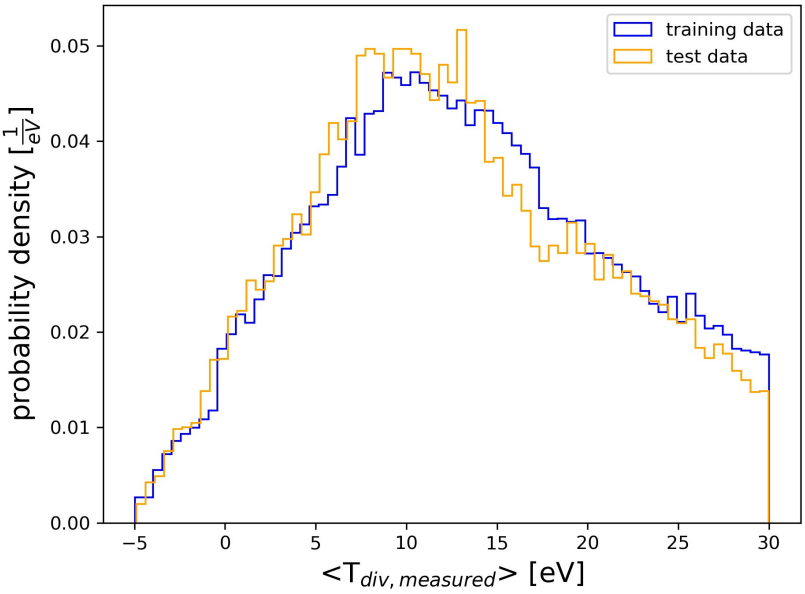


Figure 2. Distribution of T_{div} values, the target quantity of the models, averaged over 0.2 s time intervals for training and test data

and a positive correlation coefficient of 0.3 with the lower triangularity. However, these coefficients are only indicative of pairwise linear dependencies, whereas T_{div} could depend non-linearly on any of the given variables or their combinations.

Furthermore, we observe correlations indicative of expected dependencies which underline the viability of the acquired data. For example we observe strong correlations between core and edge densities and between heating power and radiated power. Above this, we also observe that the deuterium and nitrogen throughput are correlated with the neutral density in the divertor, as could be expected.

Another noteworthy point is that on one hand we observe correlations driven by physics such as the correlation between neutral density in the divertor and the core and edge density. On the other hand our data base also contains correlations imposed by operational boundary conditions of the experiment, e.g. the correlation between heating power and core and edge density which does not emerge from a direct interplay between these quantities but is set by the operational restrictions of the experiment. We do not differentiate between these types of correlations in this analysis. Missing correlations of the hydrogen, helium and neon throughputs result from the lack of variation in these signals as the data base contained little data with significant hydrogen, helium or neon throughput.

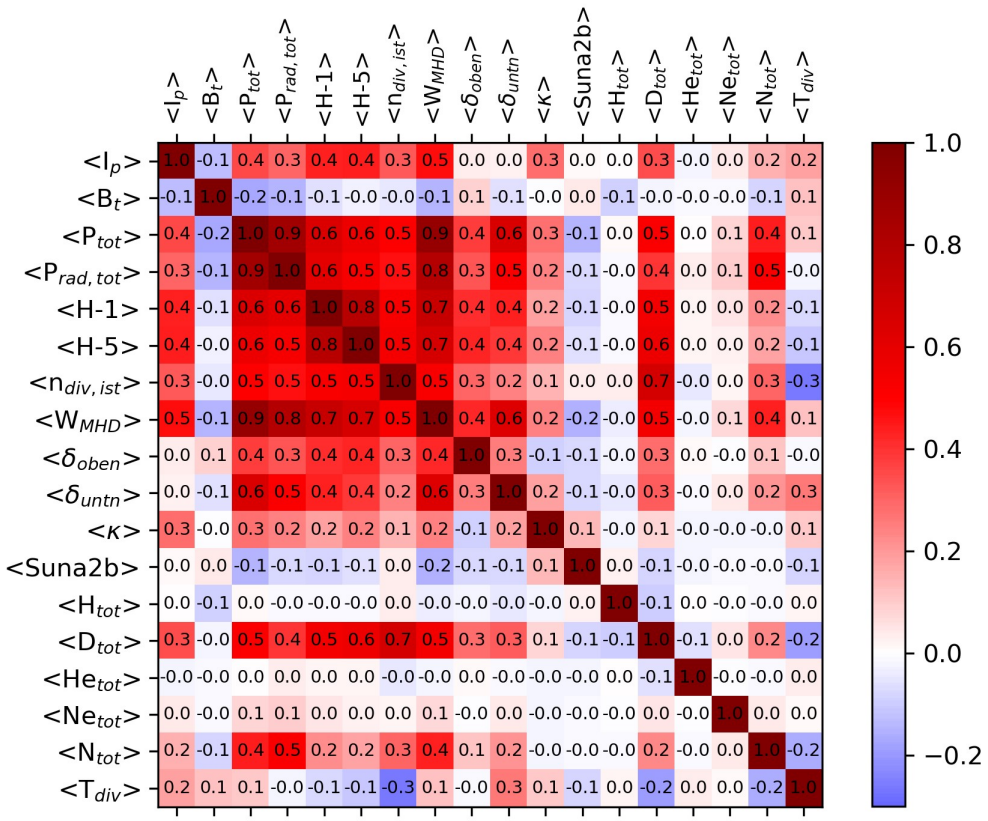


Figure 3. Matrix of Pearson correlation coefficients of the parameters investigated in this analysis; rounded to one decimal; training data set

To further test the viability of the data selection we aimed at reproducing a result from [13] where the authors found that the separatrix density depends on the divertor neutral pressure, P_0, as

$$n_{e,sep} \sim P_0^{0.31} \quad . \tag{1}$$

Therefore, we investigated the dependency of the time averaged values of the line averaged edge density on those of the neutral density in the divertor. The result is depicted in figure 4 showing the decadic logarithm of the edge density versus the decadic logarithm of the neutral density in the divertor. The dashed blue line in the figure shows the function resulting from a simple least squares fit to the data. The continuous orange line shows the result of a repeated median regression [14]. As the data shows significant outliers the least squares fit could be expected to be strongly influenced by those whereas the repeated median regression is more robust to outliers. With both approaches we manage to roughly reproduce the exponential dependency reported in [13] even if the achieved coefficients of determination here are somewhat smaller (R^2=0.59 for the least squares fit resulting in an exponent of 0.325 and R^2=0.57 for the repeated median regression resulting in an exponent of 0.277 as compared to R^2=0.7 with a resulting exponent of 0.31 in the referenced publication). Nevertheless, this shows that with the

rather rough data selection applied here we achieve results comparable to that of a more refined data selection.

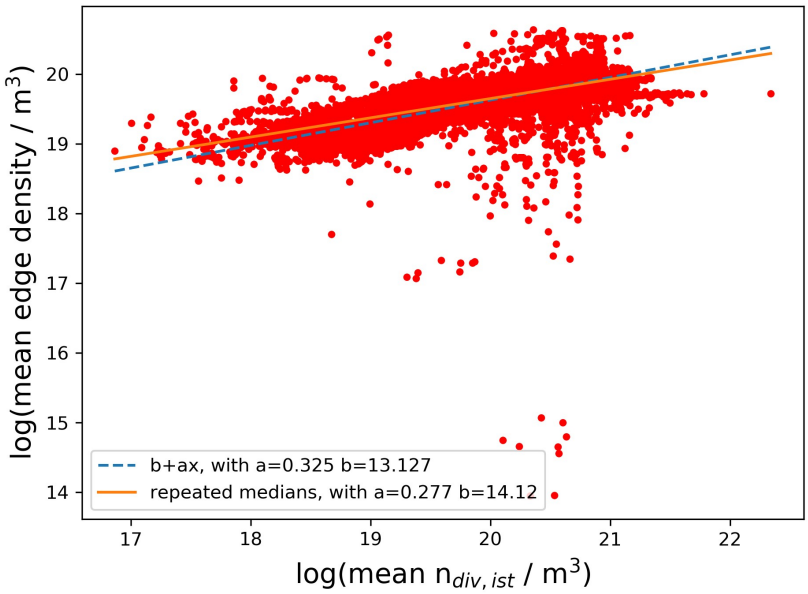


Figure 4. Dependency of the line averaged edge density on the neutral density in the divertor; dashed blue line shows the result of a least squares fit; continuous orange line shows the result of repeated median regression

3. Models and Performance

3.1. Random Forest

We tested a random forest regression model [15] to obtain predictions for T_{div} given the input variabless defined in section 2. The random forest model tested consists of 10 decision trees without limitation in depth where all decision trees could access all input quantities. The tree depth was not limited as a parameter scan of depths between 5 and 50 showed improvements in terms of predictions on the test set up to a depth of at least 20. With such a large depth the tree has a size of $\sim 10^6$ nodes which is equivalent to not limiting the tree depth in our application. This test was carried out with a model using the full set of inputs. The large number of nodes could lead to overfitting issues. We used scikit-learn [16] to set up and train the model.

The obtained performance on the test set is depicted in figure 5. The graphic shows the probability density of predictions obtained from the model versus the values extracted from the experimental data. The blue line indicates a 1:1 relation (i.e. the model predictions equal the values extracted from the experiments) which is the desired result for accurate predictions.

The distribution shows that the model predictions follow the trend of a 1:1 relation. The peak in probability density around measured values of 5 eV to 15 eV stems from

the facts that most data lies in this regime (cf. figure 2) and that the model performs best around these values. To measure and compare model performances we determined the median and the central 68^{th} percentile of the distribution of the absolute differences between model predictions and experimental values as well as the achieved adjusted R^2 coefficient on the test data. The random forest model achieved an adjusted R^2 value of 0.77 with a median absolute difference between model predictions and experimental data of $1.8^{+3.1}_{-1.3}$ eV.

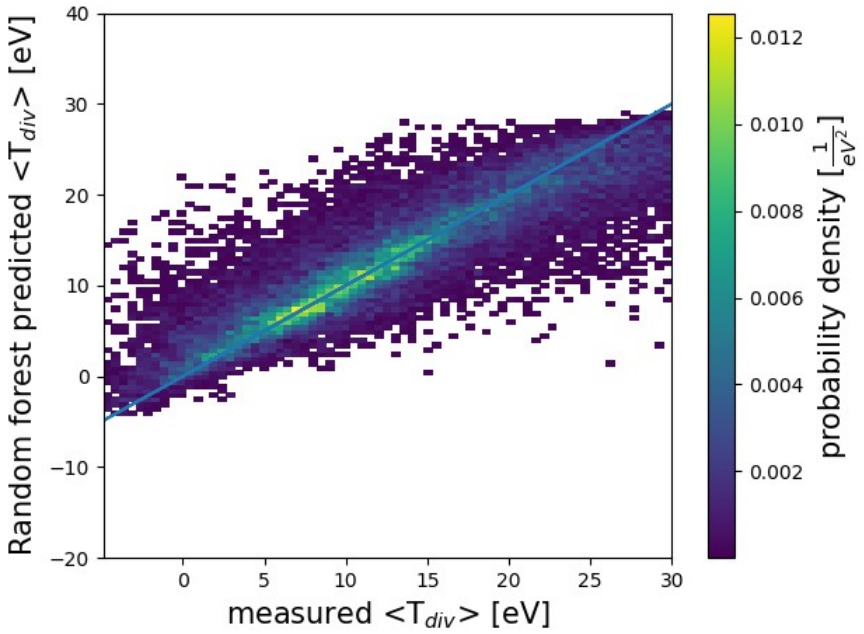


Figure 5. Probability density of predicted T_{div} values of the random forest model using all input parameters versus true values extracted from experimental data; test set; the blue line indicates a 1:1 relation

In addition we also tested a random forest model with the same criteria by using a subset of all input parameters (bold-faced in table 1). This was done since not all the input variables are available before performing a discharge. Hence, to predict a discharge, the model needs to make accurate predictions with fewer input variables. The resulting predictions on the test data are depicted in figure 6. Again, the model obviously manages to achieve a trend around a 1:1 relation between its predictions and the values from the experimental data. The adjusted R^2 value of this model is 0.71 and the median absolute difference between model predictions and experimental data amounts to $2.1^{+3.6}_{-1.5}$ eV. Thus, we manage to achieve a performance comparable to that of the model utilizing the full set of inputs which indicates that the reduced set of inputs might suffice to obtain predictions of thermal loads for a planned discharge.
However, due to the structure of random forest models and the cut applied to the range of T_{div} values in this analysis, this type of model can not predict values beyond the range to which we limited T_{div}. This limits the applicability of such models in new

experimental scenarios.
There is no evidence for a significant effect of overfitting caused by the large number of nodes in the distribution of model predictions and measured values.

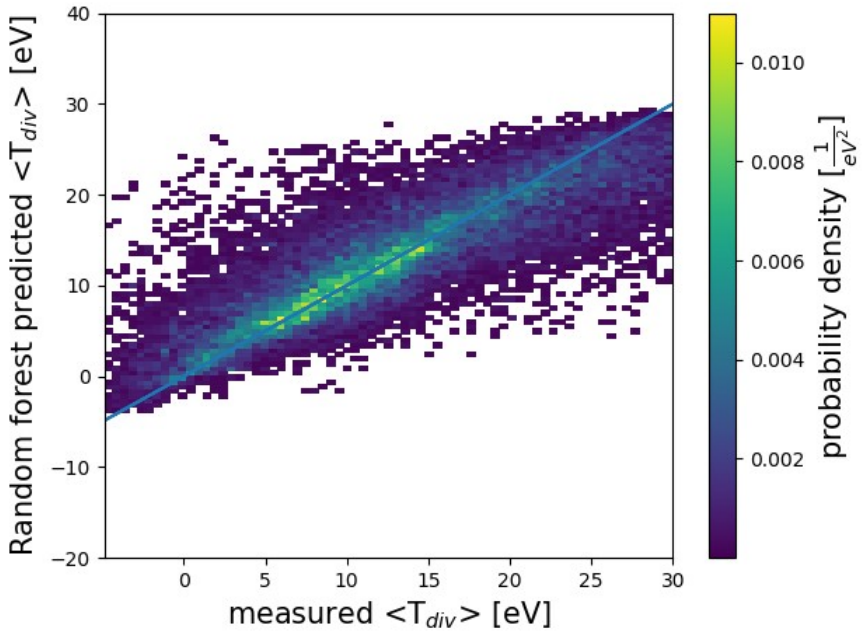


Figure 6. Probability density of predicted T_{div} values of the random forest model using a subset of input parameters versus true values extracted from experimental data; test set; the blue line indicates a 1:1 relation

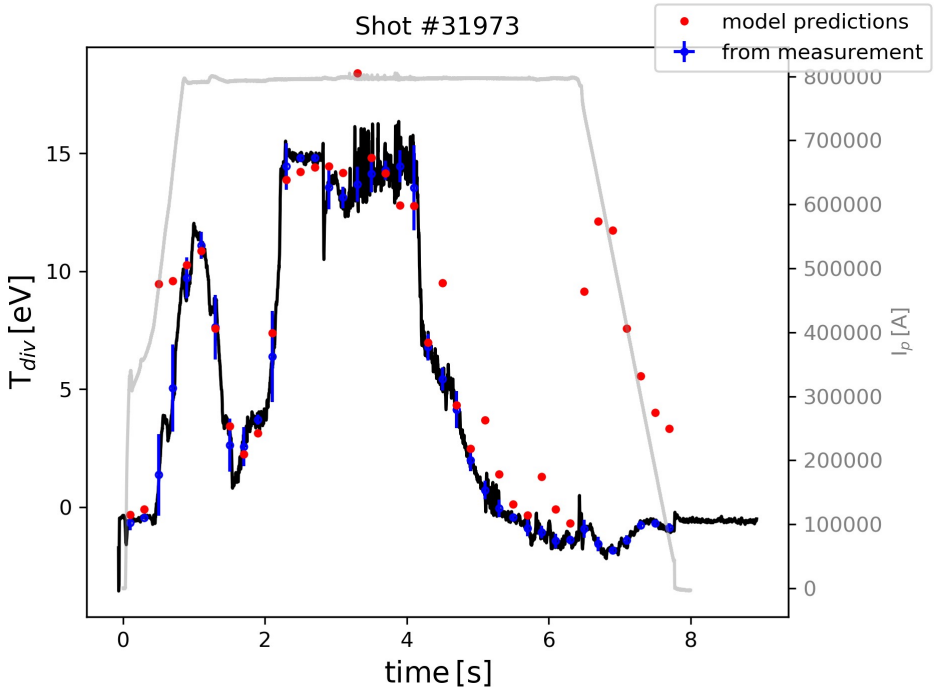


Figure 7. Random forest predictions for a complete discharge (red dots) with full set of model inputs; the black line indicates the measured T_{div} curve and the blue points are the extracted averages over 0.2 s (without applying any cuts to the data); standard deviations stem from the variation within 0.2 s; the grey curve shows the plasma current

Figure 7 shows the model predictions of the random forest model using all inputs for a complete discharge. Here, we selected a shot number from the test data and calculated the average T_{div} values for every 0.2 s interval. We also tested the model on all these data points regardless of the cuts applied on the training data to visualize the model's performance when applied to a new discharge in a worst case scenario. The red points mark the model's predictions whereas the blue dots are the averaged values of T_{div} obtained from the experimental signal (black curve). The uncertainty on the values from the experiment is given by the standard deviation of T_{div} within the 0.2 s intervals. The grey curve shows the plasma current during the discharge.

During the flattop phase, for which this model was trained, it manages to predict most data points within one standard deviation. The model also predicts the falling tendency in the flank starting at around 4 s. At this point the nitrogen seeding was significantly increased in this discharge. The discrepancy between model predictions and experimental values might be lowered by taking into account that the injected nitrogen does not immediately have an effect on the plasma temperature, considering that the presented model only uses the instantaneous values of the current gas throughput as input.

Also noteworthy are the data points in the current ramp-up and ramp-down. Where the model manages to still predict the very first data points, even though this regime

was out of the scope of the training data, it fails to predict the other data points in the ramps. Nevertheless, considering the crude data selection in this test, the model performs rather well, at least in the regime it was trained for.

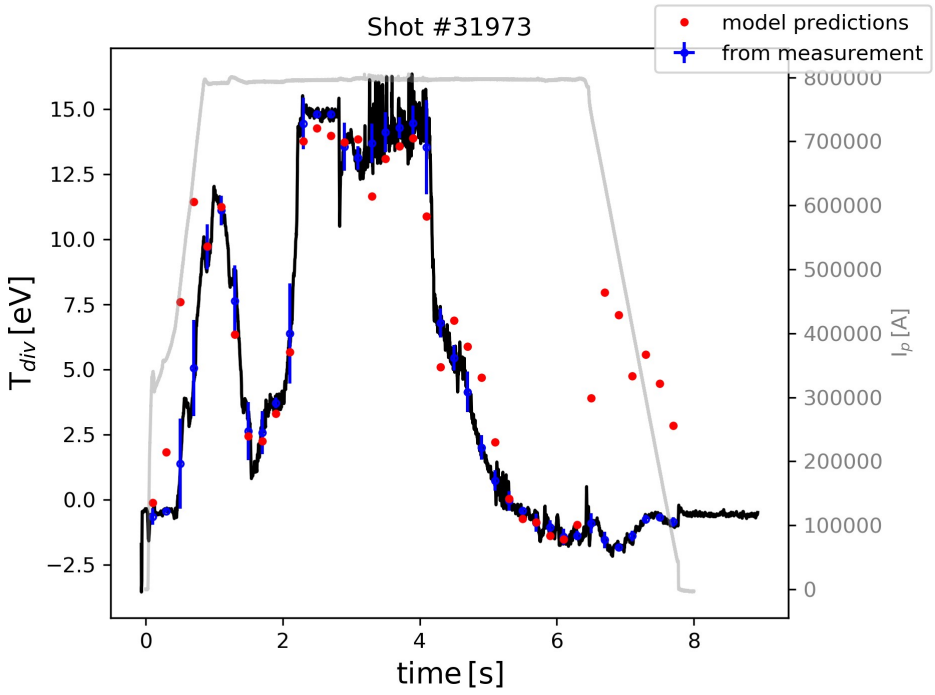


Figure 8. Random forest predictions for a complete discharge (red dots) with reduced set of model inputs; the black line indicates the measured T_{div} curve and the blue points are the extracted averages over 0.2 s (without applying any cuts to the data); standard deviations stem from the variation within 0.2 s; the grey curve shows the plasma current

Looking at predictions for the same discharge but using the model with a limited set of inputs shows similar trends. The results for this model are depicted in figure 8. The most notable change in comparison to the model with the full set of inputs is the performance in the falling flank from 4 s onward. Here, the model with fewer inputs manages to achieve more accurate predictions despite removing the radiated power as input which could be expected to have a significant influence on model predictions in impurity seeded scenarios.

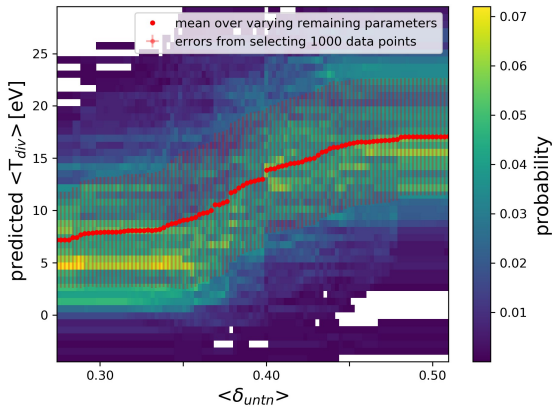
To investigate the dependency of the models on the various input quantities we varied every input quantity in a range between q_1-0.1$|q_1|$ and q_{98}+0.1$|q_{98}|$ where q_1 and q_{98} refer to the 1^{st} and 98^{th} percentile of the input's distribution in the training data‡. Since the models depend on multiple inputs we set all remaining input values by randomly selecting 1000 data points from the training set. This results in a distribution of model predictions over 1000 data points at every value in the given range of the quantity under

‡ Not to be confused with the safety factor q

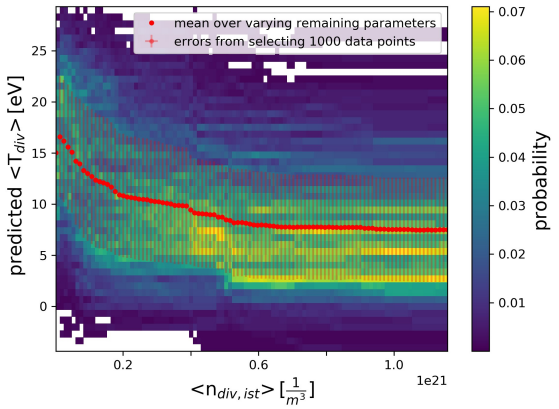
investigation (cf. [17] and [18]). The results for two of the dominant dependencies of the random forest model with full input set are depicted in figures 9(a) and 9(b) showing the partial dependency of the T_{div} predictions on the lower triangularity and the neutral density in the divertor, respectively. The red markers indicate the average value and the standard deviation of the column-wise normalized distribution obtained by inserting a given value into 1000 randomly selected data points. Beyond the depicted results we found a nearly linear dependency on the plasma current with stages indicative of the model's tree structure (figure 9(c)) whereas the predictions showed only weak or no dependency on the remaining input quantities.

As could be expected the model's predicted T_{div} decreases with increasing neutral density in the divertor as a larger density of neutral particles tends to reduce the thermal flux to the divertor targets. Moreover, we see that an increasing lower triangularity causes increasing predicted values of the model.

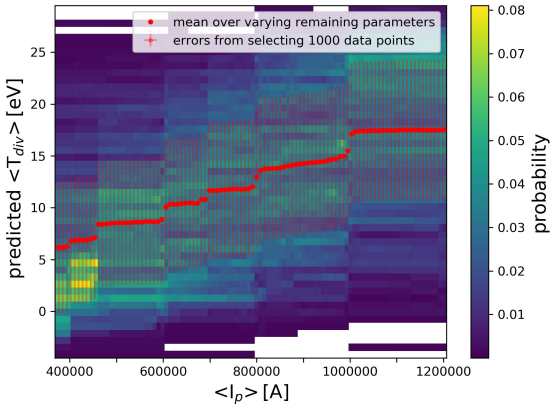
In case of the random forest model with a reduced set of input variabless the main dependency changed since we removed the neutral density in the divertor as input parameter. For this model we observe that the deuterium throughput takes over the role of the neutral density in the divertor as most important model input where an increase in deuterium throughput causes a decrease in the model's predictions as shown in figure 10(b). This, in addition with a strong Pearson correlation coefficient between neutral density in the divertor and the deuterium throughput (see figure 3), matches the observation of [19] where the authors found a direct dependency of the neutral divertor pressure on the deuterium throughput. Again we see a similar dependency of the model's predictions on the lower triangularity depicted in figure 10(a). Considering the other input variables we found again a nearly linear dependency of the predictions on the plasma current (figure 10(d)) and with the reduced set of input quantities we also observed a dependency on the core density shown in figure 10(c).



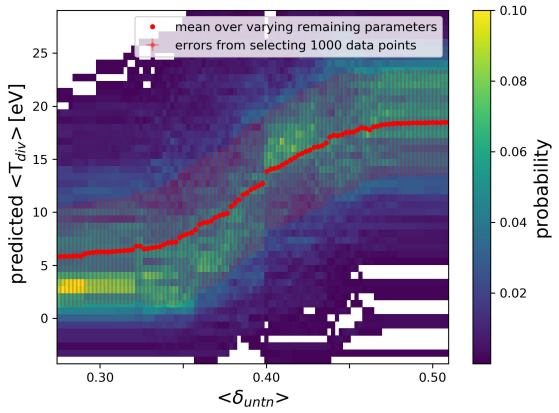
9(a) Dependency of random forest predictions using all inputs on lower triangularity, red dots indicate mean and standard deviation of the column-wise normalized distributions



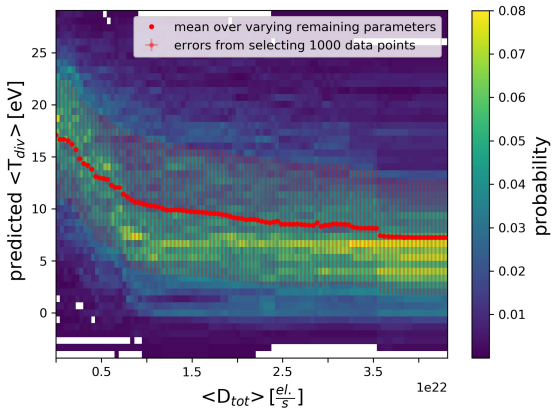
9(b) Dependency of random forest predictions using all inputs on neutral density in the divertor, red dots indicate mean and standard deviation of the column-wise normalized distributions



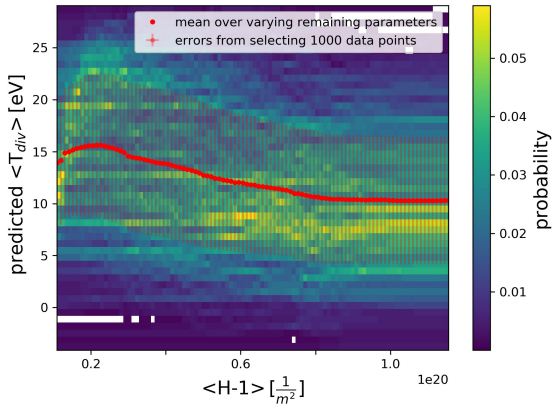
9(c) Dependency of random forest predictions using all inputs on plasma current, red dots indicate mean and standard deviation of the column-wise normalized distributions



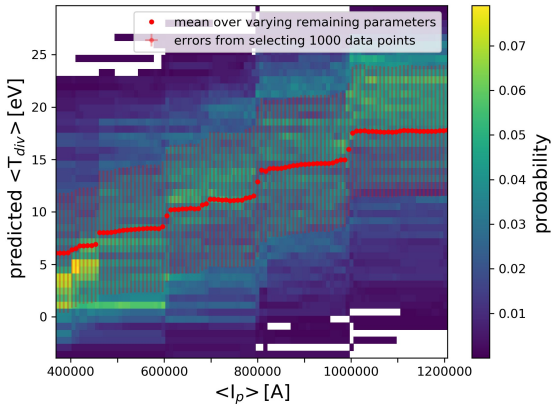
10(a) Dependency of random forest predictions using fewer inputs on lower triangularity, red dots indicate mean and standard deviation of the column-wise normalized distributions



10(b) Dependency of random forest predictions using fewer inputs on deuterium throughput, red dots indicate mean and standard deviation of the column-wise normalized distributions



10(c) Dependency of random forest predictions using fewer inputs on line-integrated core density, red dots indicate mean and standard deviation of the column-wise normalized distributions



10(d) Dependency of random forest predictions using fewer inputs on plasma current, red dots indicate mean and standard deviation of the column-wise normalized distributions

In terms of computing time, the random forest based approach is a very inexpensive approach. Setting up and training the model as well as the evaluation of its performance could all be done within a few minutes. The memory requirements for the trained model are also miniscule as it only requires about 50 MB to be stored. Thus it constitutes an inexpensive compact model for divertor power load predictions.

3.2. Neural Network

As a second approach we tested a fully connected neural network to model the dependency of T_{div} on the input quantities of section 2. The neural network used in this approach consists of three hidden layers with 100 neurons each and uses ELU activation functions [20] in the hidden layers and a linear activation in the output layer. Furthermore, we employed an L1 weight regularization term in addition to the mean squared error term in the loss function to induce weight sparsity in the model and counteract overfitting. With the full set of inputs and one output neuron the number of trainable parameters in this model amounts to about 22000. We trained the network for up to 2000 epochs on the same training data as the random forest. However, for the neural network we used 30% of the training data as validation set. We stopped the network training if the loss on the validation set had not improved by at least 0.01 over 100 epochs. The model setup as well as the training were done with Keras [21] and the Tensorflow [22] backend.
We then evaluated the model in its state of lowest validation loss.

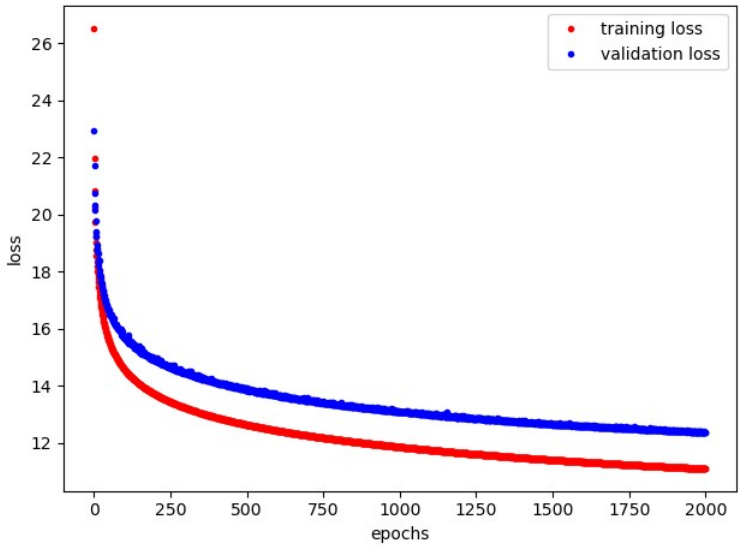


Figure 11. Evolution of neural network loss during training with the full set of inputs; loss on training data depicted in red and loss on validation data depicted in blue; losses are corrected for the regularization term

The evolution of the model's loss is depicted in figure 11. The red curve shows the loss achieved on the training data and the blue curve the one achieved on the validation

data. The contribution from the regularization term has been subtracted from both curves. The model's minimum validation loss is ∼12.4.
The model's performance on the test data set in this state is shown in figure 12. This plot again shows the probability density of the model predictions versus the values extracted from the experiment. Again we can observe that the model achieved a general tendency of a 1:1 relation. The adjusted R^2 value this model achieved is 0.75 with a median absolute difference between model predictions and experimental data of $2.3^{+3.1}_{-1.6}$ eV. Therefore, the model's performance is worse than that of the random forest model but still on a comparable level.

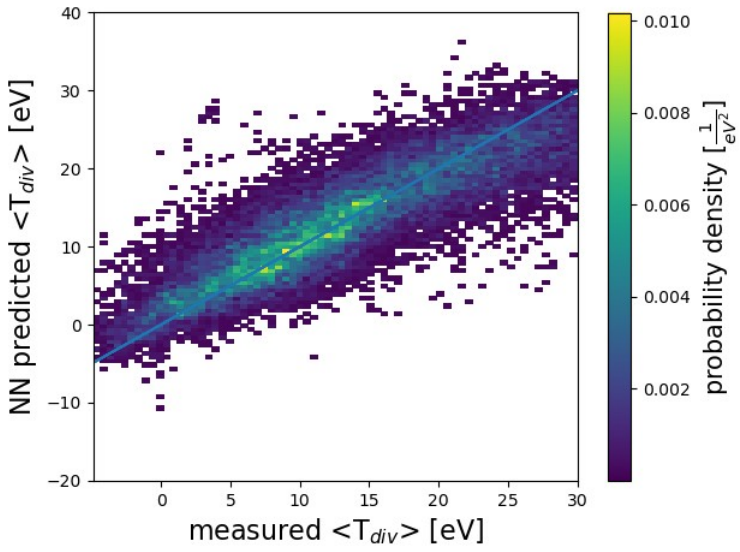


Figure 12. Probability density of predicted T_{div} values of the neural network model using all input variables versus true values extracted from experimental data; test set; the blue line indicates a 1:1 relation

We also tested this model with the reduced set of inputs. The evolution of the network's loss for this case is depicted in figure 13. The minimum validation loss achieved is ∼18.0 indicating that the model with all inputs achieved a better performance compared to this model with fewer input variables.

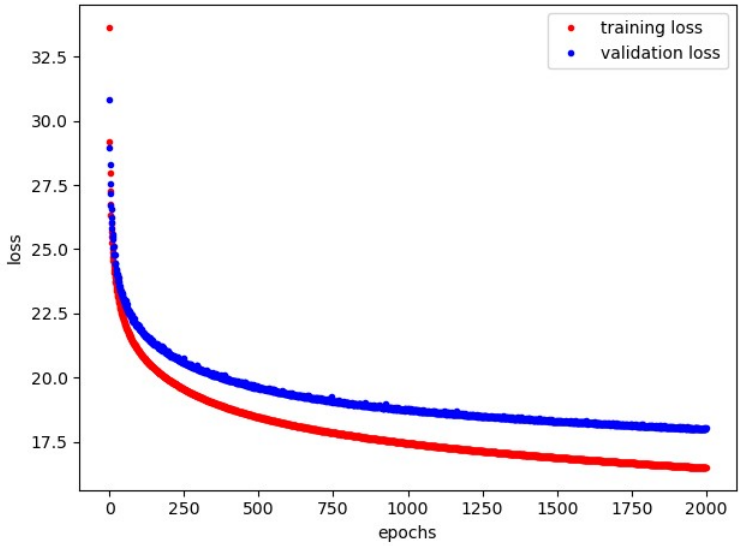


Figure 13. Evolution of neural network loss during training with the reduced set of inputs; loss on training data depicted in red and loss on validaion data depicted in blue; losses are corrected for the regularization term

This can be confirmed by the resulting predictions obtained on the test data set shown in figure 14. The distribution shows a wider spread around the indicated 1:1 relation than the model with all inputs. Moreover, this model only achieved an adjusted R^2 value of 0.65 and a median absolute difference between model predictions and experimental data of $2.7^{+3.8}_{-1.9}$ eV. This model yields the worst performance of the models under investigation even though it is still comparable with the other models.

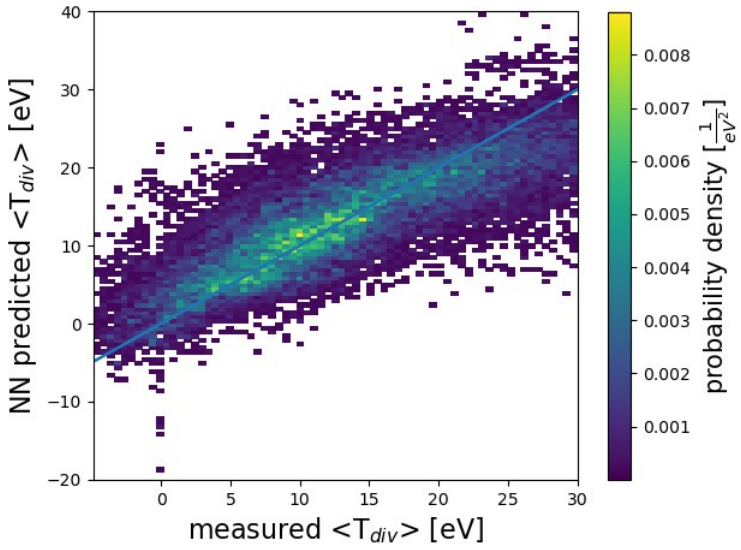
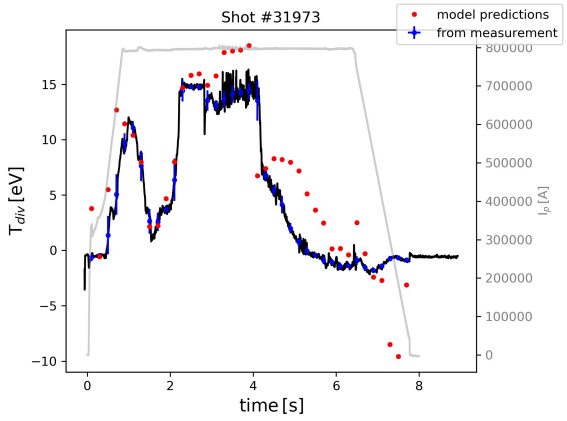
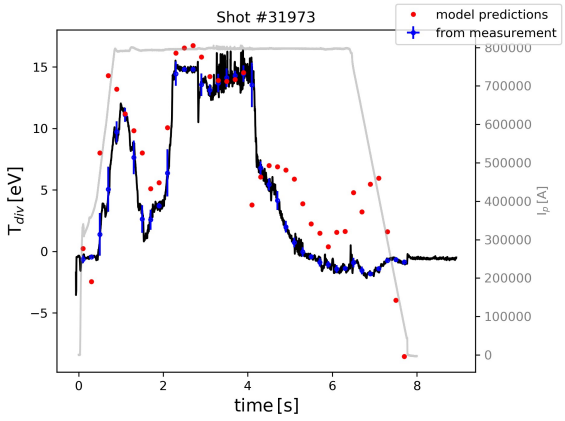


Figure 14. Probability density of predicted T_{div} values of the neural network model using fewer input variables versus true values extracted from experimental data; test set; the blue line indicates a 1:1 relation



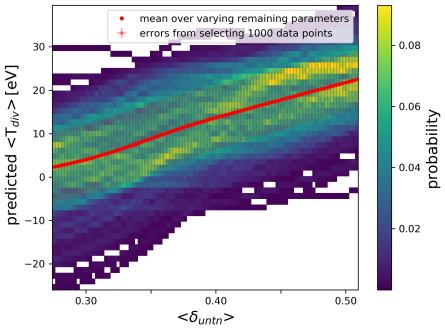
15(a) Neural network predictions for a complete discharge (red dots) with full set of model inputs; the black line indicates the measured T_{div} curve and the blue points are the extracted averages over 0.2 s (without applying any cuts to the data); standard deviations stem from the variation within 0.2 s; the grey curve shows the plasma current



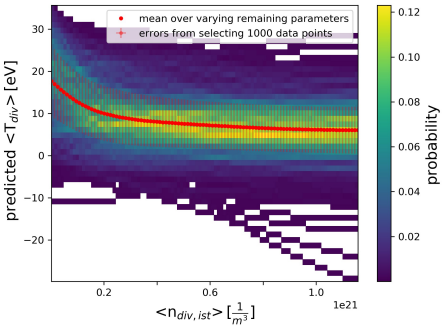
15(b) Neural network predictions for a complete discharge (red dots) with reduced set of model inputs; the black line indicates the measured T_{div} curve and the blue points are the extracted averages over 0.2 s (without applying any cuts to the data); standard deviations stem from the variation within 0.2 s; the grey curve shows the plasma current

As we did with the random forest models we also tested the neural network models on the full data of the same discharge from the test set, regardless of the data cuts mentioned in section 2. The resulting predictions for the same discharge that we used to evaluate the random forest models are depicted in figure 15(a) for the neural network model with the full set of inputs and in figure 15(b) for the neural network model with the reduced set of inputs.
It is evident that the neural network model with reduced set of inputs performs worse on this data than the respective random forest model. Except for a few data points right before the 4 s mark the remaining predictions deviate more from the experimental values than the predictions of the random forest model.
Both neural network models manage to capture the falling trend in the flank from 4 s onwards but the predictions do not match the experimental values. This might be caused by spurious dependencies that were captured by the neural networks but were omitted by the random forest models. However, the performance of the neural network model with the full set of inputs and the respective random forest model is comparable, except for the falling flank from 4 s onwards.
This tendency underlines the results we observed on the tests on the full test data set. On average, the neural networks perform slightly worse than the random forest models.

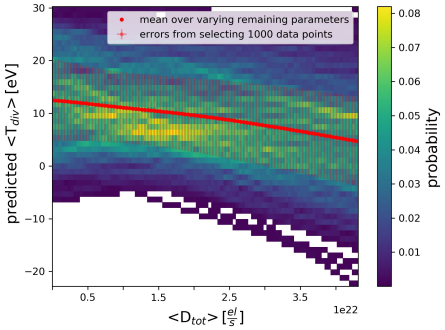
To determine the different dependencies the models captured we investigated the partial dependencies of the neural network models as described for the random forest models in section 3.1.



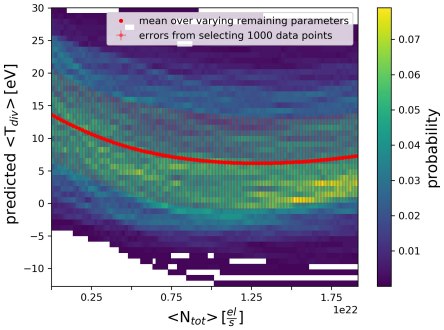
16(a) Dependency of neural network predictions using all inputs on lower triangularity, red dots indicate mean and standard deviation of the column-wise normalized distributions



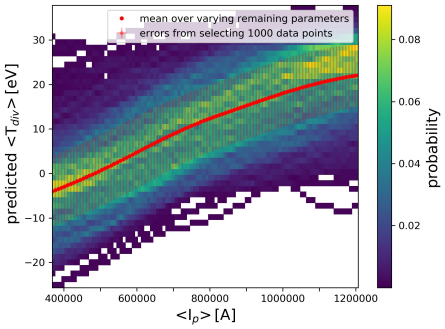
16(b) Dependency of neural network predictions using all inputs on neutral density in the divertor, red dots indicate mean and standard deviation of the column-wise normalized distributions



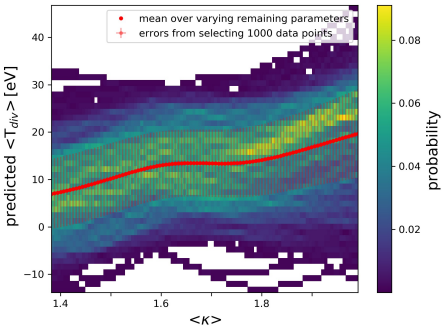
16(c) Dependency of neural network predictions using all inputs on deuterium throughput, red dots indicate mean and standard deviation of the column-wise normalized distributions



16(d) Dependency of neural network predictions using all inputs on nitrogen throughput, red dots indicate mean and standard deviation of the column-wise normalized distributions



16(e) Dependency of neural network predictions using all inputs on plasma current, red dots indicate mean and standard deviation of the column-wise normalized distributions



16(f) Dependency of neural network predictions using all inputs on plasma elongation, red dots indicate mean and standard deviation of the column-wise normalized distributions

The neural network model with the full set of inputs shows a mostly linear dependency on the plasma current up to 0.8 MA beyond this the predicted T_{div} curve slightly flattens

as can be seen in figure 16(e). The neural network with a reduced set of inputs shows a linear dependency on the plasma current with only slight deviations around 0.9 MA to 1 MA (figure 17(d)). The general tendency of increasing predictions with increasing plasma current matches the expectations obtained from physics analyses where a decreasing power fall-off length was observed for increasing poloidal magnetic field strength [23] (i.e. plasma current), thus leading to an increase in electron temperature. This confirms that the models extracted dependencies from the data that would be expected from previous physics analyses.

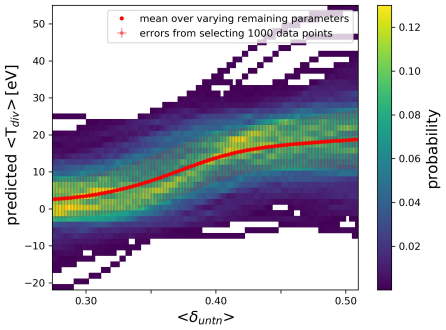
In contrast to the random forest models the neural network model with all inputs shows a nearly linear dependency of its predictions on the lower triangularity as can be seen in figure 16(a) whereas in figure 17(a) the neural network model with a reduced set of inputs shows a similar dependency on the lower triangularity as the random forest models.

Motivated by the strong correlation between the lower triangularity and the heating power observed in our data set (cf. figure 3) and by the expectation that the heating power should be more relevant to the model predictions than the triangularity, we further investigated the model's dependency on the latter. A first analysis in which we removed the lower triangularity from the set of inputs to the neural network with a reduced number of inputs showed that the model's dependency on the heating power became more pronounced and we observed a mostly linear dependency of the model's predictions on the heating power. Hence, the dependency of the model predictions on the lower triangularity could be driven by the correlation between triangularity and heating power. However, the actual underlying reason for the model's dependency on the lower triangularity and especially the shape of the dependency between T_{div} predictions and triangularity are still open questions which could be addressed in further analyses.

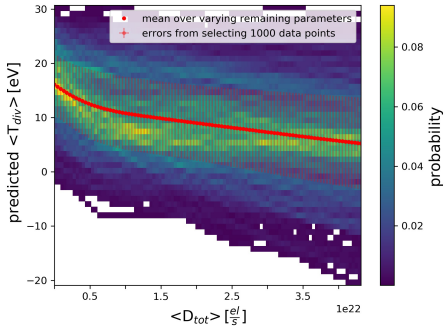
Both neural network models show a similarly shaped dependency on the plasma elongation (figures 16(f) and 17(e)). The dependency is more pronounced in the neural network with fewer input variables. In contrast to this the random forest models showed only a weak dependency on the plasma elongation. The dependency on the neutral density in the divertor that we find in the neural network model, depicted in figure 16(b), is comparable to that of the corresponding random forest model. We find a similar result when comparing the dependency of model predictions on the total deuterium throughput in the neural network, figure 17(b), and random forest models with reduced input sets. In both neural networks and random forests removing the neutral density in the divertor from the set of inputs leads to a change in the dependency on the total deuterium throughput in such a way that it mimics the dependency on the neutral density the model with all inputs shows. We already discussed this phenomenon when dealing with the random forest models.

An interesting dependency can also be observed for the total nitrogen throughput. Where the random forest models showed little dependency on the nitrogen throughput, it is evident in figures 16(d) and 17(c) that the neural network predictions do depend on the nitrogen throughput. The dependencies that we observe here indicate that an

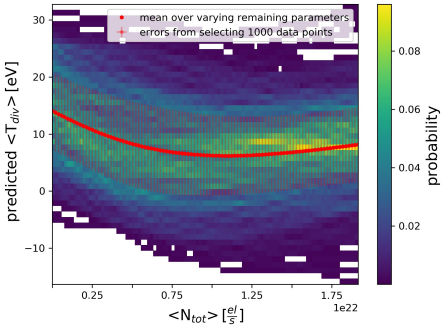
increased nitrogen throughput leads to reduced predictions for T_{div} as could be expected since nitrogen seeding tends to cool the plasma [9].



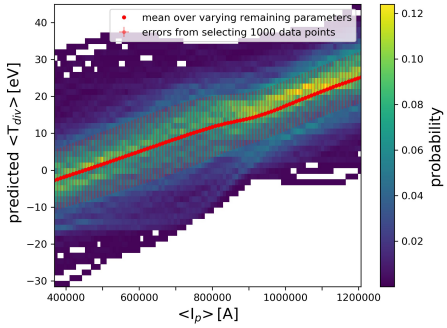
17(a) Dependency of neural network predictions using fewer inputs on lower triangularity, red dots indicate mean and standard deviation of the column-wise normalized distributions



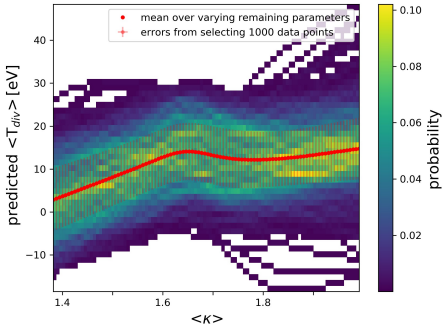
17(b) Dependency of neural network predictions using fewer inputs on deuterium throughput, red dots indicate mean and standard deviation of the column-wise normalized distributions



17(c) Dependency of neural network predictions using fewer inputs on nitrogen throughput, red dots indicate mean and standard deviation of the column-wise normalized distributions



17(d) Dependency of neural network predictions using fewer inputs on plasma current, red dots indicate mean and standard deviation of the column-wise normalized distributions



17(e) Dependency of neural network predictions using fewer inputs on plasma elongation, red dots indicate mean and standard deviation of the column-wise normalized distributions

The result we obtain for the neural network also shows an important difference between the neural network and the random forest approach for predictions beyond the range used during the training. While the random forest based models are making reasonable predictions within the limits we imposed on the range of T_{div}, they fail to make predictions outside the training range. The neural network models on the other hand can predict values outside the training range above 30 eV regardless of the cuts we applied to the data as can be seen e.g. in figure 17(a).
However, in terms of computing time requirements the neural network based approach is the more expensive one. Training the neural networks as presented required roughly two hours on a single GPU for the model with all inputs and about 1.5 hours for the model with reduced input set. On the other hand evaluating the models on the full test data only required a few minutes for each model. The fully trained models only require roughly 200 kB of storage. Hence, once the initial training is completed the neural network models also constitute a computationally cheap model for divertor power load predictions.

A first analysis of the hyperparameters of the neural networks in which we varied the number of layers, numbers of neurons in each layer, the constant of the regularizer, the batch size, activation functions in the hidden layers and also tested batch normalization [24] and dropout [25] did not indicate clear trends of performance improvements with respect to most of the hyperparameters.
We used Optuna [26] to investigate a range of hyperparameter configurations. In this approach we did not limit the number of neurons to be the same in each layer. The activation functions were chosen between the ELU function we also used for the baseline model presented before and the SELU activation function [27]. Since the latter is supposed to have a similar effect on the network as batch normalization, we excluded batch normalization if the SELU activation function was selected for a network.
If dropout was active in a network we also randomly selected the dropout rate for that network.
The most significant finding is that a small value for the constant of the regularizer as well as not using dropout seem to benefit the model's performance. Investigating the cause for this and a more detailed hyperparameter analysis could be the focus of an upcoming work.

Beyond this we also tested an ensemble based approach [28] in which we trained the presented network architecture ten times. Due to the random initial values of the network's weights each training leads to a slightly different final model. For the evaluation we then averaged the predictions of all ten networks to obtain one prediction.
For the approach with all inputs we thus achieved an adjusted R^2 on the test data of 0.77 and a median absolute difference between model predictions and values extracted from experiments of $2.16^{+2.98}_{-1.53}$ eV. The approach with a reduced set of inputs results in an adjusted R^2 of 0.67 and a median absolute difference of $2.6^{+3.7}_{-1.8}$ eV. Hence, using this

ensemble approach seems to slightly improve the model's performance but the difference thus far is not significant.

4. Conclusions and Outlook

We set up and tested two machine learning based approaches to construct a data-driven scalar surrogate model of steady state thermal loads at the divertor targets in AUG. The results we obtained show that even with our very basic data selection we manage to predict thermal loads for given operational parameters reasonably well. Moreover, the models show dependencies which were also found by dedicated physics analyses. Given that we did not differentiate between L- and H-mode discharges and detached and attached plasma conditions our models work over a wide range of parameters even including impurity seeded scenarios.

Due to our rough data selection and the limited selection of standardly available input quantities there are still missing dependencies that we have not directly included in the models, e.g. other impurity species or wall conditioning. Furthermore, the experimental data contain dependencies that are imposed by the boundary conditions of the experiments.

Both, the approaches with the full set of inputs and with a reduced set of inputs yield similar results for both types of machine learning models. Hence, a reduced set of input quantities known before performing a plasma discharge might suffice to utilize the presented models in operation. These models could already be used to plan a discharge with respect to divertor conditions.

According to our analysis of the partial dependencies of the various models all models show a strong dependency on the lower triangularity which has a similar shape for most of the models we tested. This might be driven by the correlation between heating power and triangularity but further investigation, especially into the shape of this dependency, might be required. Both neural network approaches show a dependency of their predictions on the plasma elongation whereas we did not find such a pronounced dependency on the elongation in the random forest models. Thus, considering the parameters describing the plasma shape, the lower triangularity seems to have a dominant effect. Furthermore, we observed that in both neural network based models an increase in the nitrogen throughput leads to a decrease in the predicted T_{div} values. This and the results obtained for the dependencies on the neutral density in the divertor and the deuterium throughput are in accordance with previous analyses.

The random forest based models showed a better performance than the neural network based models in our application task. However, their predictions are limited to the range of target values observed in the training data. Therefore, a neural network based approach might be more interesting for future investigation.

In the analysis of an exemplary discharge we found that the prediction accuracy of both neural network based approaches seems to suffer from changes in the impurity throughput during the discharge. To mitigate this effect convolutional or recurrent

neural networks could be employed in order to account for the temporal evolution of the impurity throughput and the corresponding plasma response.
A future analysis will also deal with model uncertainties caused by e.g. missing inputs to the model. A potential method of approaching this topic could be mixture density networks [29] which model the whole predictive distribution instead of a single point estimate. This would allow for an assessment of the reliability of the model predictions.
Another topic for further investigation could be the inclusion of more output quantities to be predicted by the models. Future fusion devices will rely on a scarce selection of available sensors available in the divertor for detachment control. The presented work could be extended to include a line-of-sight signal of a Balmer line emission in the divertor or fast integrated bolometric measurements to characterise the plasma state around the X-point [30].
In order to mitigate the obstacles of experimental observations data from simulations could also be included in this analysis similar to [7]. Work in this direction has started with results obtained from SOLPS-ITER simulations.
All of our results are based solely on data from AUG. A further point of investiagtion could be the inclusion of data from different tokamaks to include and investigate effects of varying machine size on the models. With such data included the models could then be used as a lightweight method to perform parameter scans with respect to divertor conditions and to determine operational constraints on machines of varying size and typical operational regimes.

Acknowledgements

This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 and 2019-2020 under grant agreement No 633053. The views and opinions expressed herein do not necessarily reflect those of the European Commission.
The authors gratefully acknowledge the computing time granted through JARA on the supercomputer JURECA at Forschungszentrum Jülich.

References

[1] RA Pitts et al. “Physics basis for the first ITER tungsten divertor”. In: *Nuclear Materials and Energy* (2019), p. 100696.

[2] M Bernert et al. “Power exhaust by SOL and pedestal radiation at ASDEX Upgrade and JET”. In: *Nuclear Materials and Energy* 12 (2017), pp. 111–118.

[3] M. Wischmeier. “High density operation for reactor-relevant power exhaust”. In: *Journal of Nuclear Materials* 463 (2015). PLASMA-SURFACE INTERACTIONS 21, pp. 22 –29. ISSN: 0022-3115. DOI: https://doi.org/10.1016/j.jnucmat.2014.12.078. URL: http://www.sciencedirect.com/science/article/pii/S0022311514010216.

[4] S. Wiesen et al. "Plasma edge and plasma-wall interaction modelling: Lessons learned from metallic devices". In: *Nuclear Materials and Energy* 12 (2017). Proceedings of the 22nd International Conference on Plasma Surface Interactions 2016, 22nd PSI, pp. 3 –17. ISSN: 2352-1791. DOI: https://doi.org/10.1016/j.nme.2017.03.033. URL: http://www.sciencedirect.com/science/article/pii/S2352179116300370.

[5] C Rea et al. "Disruption prediction investigations using Machine Learning tools on DIII-D and Alcator C-Mod". In: *Plasma Physics and Controlled Fusion* 60.8 (2018), p. 084004. DOI: 10.1088/1361-6587/aac7fe. URL: https://doi.org/10.1088%2F1361-6587%2Faac7fe.

[6] Julian Kates-Harbeck, Alexey Svyatkovskiy, and William Tang. "Predicting disruptive instabilities in controlled fusion plasmas through deep learning". In: *Nature* 568.7753 (2019), pp. 526–531.

[7] Daniel Böckenhoff et al. "Reconstruction of magnetic configurations in W7-X using artificial neural networks". In: *Nuclear Fusion* 58.5 (2018), p. 056009.

[8] Karel Lucas van de Plassche et al. "Fast modeling of turbulent transport in fusion plasmas using neural networks". In: *Physics of Plasmas* 27.2 (2020), p. 022310.

[9] A Kallenbach et al. "Divertor power load feedback with nitrogen seeding in ASDEX Upgrade". In: *Plasma Physics and Controlled Fusion* 52.5 (2010), p. 055002.

[10] SS Henderson et al. "An assessment of nitrogen concentrations from spectroscopic measurements in the JET and ASDEX upgrade divertor". In: *Nuclear Materials and Energy* 18 (2019), pp. 147–152.

[11] AW Leonard. "Plasma detachment in divertor tokamaks". In: *Plasma Physics and Controlled Fusion* 60.4 (2018), p. 044001.

[12] G.M. Staebler and F.L. Hinton. "Currents in the scrape-off layer of diverted tokamaks". In: *Nuclear Fusion* 29.10 (1989), pp. 1820–1824. DOI: 10.1088/0029-5515/29/10/017. URL: https://doi.org/10.1088%2F0029-5515%2F29%2F10%2F017.

[13] A Kallenbach et al. "Parameter dependences of the separatrix density in nitrogen seeded ASDEX Upgrade H-mode discharges". In: *Plasma Physics and Controlled Fusion* 60.4 (2018), p. 045006.

[14] Andrew F Siegel. "Robust regression using repeated medians". In: *Biometrika* 69.1 (1982), pp. 242–244.

[15] Leo Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[16] Fabian Pedregosa et al. "Scikit-learn: Machine learning in Python". In: *the Journal of machine Learning research* 12 (2011), pp. 2825–2830.

[17] Jerome H. Friedman. "Greedy function approximation: A gradient boosting machine". In: *Annals of statistics* 29.5 (2001), pp. 1189–1232. DOI: 10.1214/aos/1013203451.

[18] Christoph Molnar. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. https://christophm.github.io/interpretable-ml-book/. 2019.

[19] T Luda et al. "Integrated modeling of ASDEX Upgrade plasmas combining core, pedestal and scrape-off layer physics". In: *Nuclear Fusion* 60.3 (2020), p. 036023.

[20] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. Nov. 23, 2015. arXiv: 1511.07289v5 [cs.LG].

[21] François Chollet et al. *Keras*. https://keras.io. 2015.

[22] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

[23] Thomas Eich et al. "Scaling of the tokamak near the scrape-off layer H-mode power width and implications for ITER". In: *Nuclear fusion* 53.9 (2013), p. 093031.

[24] Sergey Ioffe and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift". In: *arXiv e-prints*, arXiv:1502.03167 (Feb. 2015), arXiv:1502.03167. arXiv: 1502.03167 [cs.LG]. URL: https://ui.adsabs.harvard.edu/abs/2015arXiv150203167I.

[25] Alex Krizhevsky Ilya Sutskever Ruslan Salakhutdinov Nitish Srivastava Geoffrey Hinton. "Dropout: a simple way to prevent neural networks from overfitting". In: *The journal of machine learning research* 15 (2014), pp. 1929–1958. ISSN: 1.

[26] Takuya Akiba et al. "Optuna: A Next-generation Hyperparameter Optimization Framework". In: (July 25, 2019). arXiv: 1907.10902v1 [cs.LG].

[27] Günter Klambauer et al. "Self-Normalizing Neural Networks". In: *Advances in Neural Information Processing Systems 30 (NIPS 2017)* (June 8, 2017). arXiv: 1706.02515v5 [cs.LG].

[28] Thomas G Dietterich. "Ensemble methods in machine learning". In: *International workshop on multiple classifier systems*. Springer. 2000, pp. 1–15.

[29] Christopher M Bishop. "Mixture density networks". In: *Neural Computing Research Group Report* (1994). URL: https://publications.aston.ac.uk/id/eprint/373/1/NCRG_94_004.pdf.

[30] M. Bernert et al. "X-point radiation, its control and an ELM suppressed radiating regime at the ASDEX Upgrade tokamak". In: *Nuclear Fusion* 61 (2021), p. 024001.