EuroPED-NN: Uncertainty aware surrogate model

A. Panera Alvarez[1,2,*], A. Ho[1,3], A. Järvinen[4], S. Saarelma[5], S. Wiesen[1,6], JET Contributors[a], and the ASDEX Upgrade Team[b]

[1]DIFFER – Dutch Institute for Fundamental Energy Research, 5612 AJ Eindhoven, Netherlands
[2]Eindhoven University of Technology, 5612 AZ Eindhoven, Netherlands
[3]MIT Plasma Science and Fusion Center, Cambridge, 02139, MA, United States of America
[4]VTT Technical Research Centre of Finland, FI-02044 VTT, Finland
[5]United Kingdom Atomic Energy Authority, Culham Centre for Fusion Energy, Culham Science Centre, Abingdon, Oxon, OX14 3DB, UK
[6]Forschungszentrum Jülich GmbH, Institut für Energie- und Klimaforschung - Plasmaphysik, DE-52425 Jülich, Germany
[a]See the author list of J. Mailloux et al. 2022 Nucl. Fusion 62 042026
[b]See the author list of Zohm et al. 2024 Nucl. Fusion, U. Stroth et al. 2022 Nucl. Fusion 62 042006.
[*]E-mail: a.paneraalvarez@differ.nl

Abstract

This work successfully generates an uncertainty-aware surrogate model of the EuroPED plasma pedestal model using the Bayesian neural network with noise contrastive prior (BNN-NCP) technique. This model is trained using data from the JET-ILW pedestal database and subsequent model evaluations, conforming to EuroPED-NN. The BNN-NCP technique has been proven to be a suitable method for generating uncertainty-aware surrogate models. It matches the output results of a regular neural network while providing confidence estimates for predictions as uncertainties. Additionally, it highlights out-of-distribution (OOD) regions using surrogate model uncertainties. This provides critical insights into model robustness and reliability. EuroPED-NN has been physically validated, first, analyzing electron density $n_e(\psi_{\text{pol}} = 0.94)$ with respect to increasing plasma current, I_p, and second, validating the $\Delta - \beta_{p,ped}$ relation associated with the EuroPED model. This affirms the robustness of the underlying physics learned by the surrogate model. On top of that, the method was used to develop a EuroPED-like model fed with experimental data, i.e. an uncertainty aware experimental model, which is functional in JET database. Both models have been also tested in ~ 50 AUG shots.

1 Introduction

A tokamak fusion reactor must achieve a sufficiently high core plasma pressure, density and temperature in order initiate and sustain fusion reactions for net energy generation. The standard operational scenarios in tokamak devices as based on the high confinement mode (*H-mode*) [1]. This regime is characterized by the formation of a transport barrier in the outer region of the confined plasma, leading to steep gradients in the density and temperature profiles within that region. As the core plasma parameters are effectively an integration of the transport equation from the edge boundary condition, this steep gradient region lifts up the core parameters and gives this feature its namesake, the plasma *pedestal*.

This edge transport barrier has been linked to an increase in the radial electric field, E_r, shear [2] which drives a local bulk rotation shear that breaks up large turbulent eddies [3], suppressing the dominant transport mechanisms observed in that region in L-mode regimes [4]. It should be noted that the steep gradient region still contains turbulent transport processes, although they are driven by other mechanisms which are less affected by the flow shear [5–7]. While the precise mechanism of the E_r-well formation for the L-H transition is still under investigation [8–10], the resulting transport reduction causes the plasma gradients to increase, along with the plasma current due to the bootstrap effect [11], up until a point where MHD instabilities occur, e.g. edge-localized modes (ELMs) [12]. These ELMs manifest as a discrete and periodic collapse of the pedestal, effectively reinstating another, albeit higher, limit on the pressure gradient. The predominant explanation of ELMs goes via MHD theory, specifically the interaction between instabilities in the pressure driven (ballooning) modes and current driven (peeling) modes [13], due to these steep gradients in the edge region. Within the EPED model [14], this theoretical pedestal limit is then located where the pedestal pressure, p_{ped}, and width, Δ, lie on the intersection of the onset of ideal MHD instabilities and the onset of kinetic ballooning mode (KBM) turbulence, which the latter is

represented by a semi-empirical relation. The large type-I ELMs are well described by this model [15], tracing out stability and instability regions in the current density-pressure gradient (j-α) space. The EuroPED model [16] effectively predicts these pedestal limits along these same theoretical principles.

Unfortunately, the computational cost of both EuroPED and EPED is too high for highly iterative applications such as plasma scenario optimization [17] and motivates the development of a surrogate model which uses less computational resources while to representing the original model reasonably accurately. Various approaches in the fusion community utilize machine learning algorithms to develop surrogate models and their subsequent applications [18–26]. The fast computational throughput also opens the door to explore the behaviour of the model and the physics behind it. This work investigates the applicability of Bayesian neural networks (BNNs) to generate a pedestal surrogate model, potentially incorporating their ability to provide uncertainty information in a single forward pass as part of a prediction application. This approach aligns with previous work in using machine learning to enhance plasma pedestal modeling [21, 27].

In the Bayesian approach, the variables in the neural networks are treated as distributions, which are propagated throughout the calculation to provide statistical information about the model prediction, labelled in this study as output *uncertainties*. The training process needs to determine the distributions of each output using a variational inference approximation [28] allowing the construction of a feasible loss function. However, the main drawback of variational inference is the high degree of sampling necessary for the training. This is circumvented using another acceleration technique called the Noise Contrastive Prior (NCP) [29], which has a secondary advantage of estimating both the *epistemic* and *aleatoric* uncertainties. The epistemic uncertainties (or model uncertainties) are defined as uncertainties due to the lack of knowledge about the system, such as missing input parameters or insufficent data volume, and the aleatoric uncertainties (or data uncertainties) are defined as uncertainties resulting from random chance, such as repeat measurements or experiments providing different results [30, 31]. This study expands on the methodology and interprets the results of the trained model based on these definitions.

Section 2 describes the specific application of the BNN-NCP model in this study, including the considerations in determining the user-defined parameters impacting the aleatoric and epistemic uncertainty predictions. In Section 3, the precision and practicality of this model is discussed, focusing on its capacity to accelerate computations, gather physics insights, and adapt to a different machine. Section 4 extends the methodology developed to experimental JET-ILW data and tests it in AUG data. Finally, Section 5 provides a summary of the main conclusions and potential future work inspired by this study.

2 Development of EuroPED-NN

In this section, the implementation of the BNN-NCP model towards building a plasma pedestal surrogate model is detailed.

2.1 Model description, inputs and outputs

The EuroPED model is conceptually identical to the EPED model, which uses an ideal MHD stability description to predict the critical plasma pressure gradient, α, as a function of the pedestal width, Δ. This curve is then crossed against a semi-empirical Δ-$\beta_{p,ped}$ relation reflecting the KBM constraint [14] to provide a prediction of the pedestal properties. The assumption of a two-species plasma in equilibrium is then used to estimate the top-pedestal electron pressure, $p_{e,ped}$, and further decompose that into the top-pedestal electron temperature, $T_{e,ped}$, by requiring the top-pedestal electron density, $n_{e,ped}$, as input.

Thus, for the EPED model, the required input parameters include:

- plasma parameters:
  - plasma current, I_p, in [MA];
  - toroidal magnetic field, B_t, in [T];
  - line-integrated effective charge, Z_{eff};
  - top-pedestal electron density, $n_{e,ped}$, in [$\times 10^{19}$ m^{-3}];
  - normalized plasma pressure, β_N;
- and magnetic geometry parameters:
  - triangularity, δ;
  - elongation, κ;
  - minor radius, a, in [m];
  - major radius, R_0, in [m].

In practice, one drawback of the EPED model is the necessity of providing $n_{e,ped}$, β_N and Z_{eff} as inputs, as this precludes the model from predictive exercises as these parameters are often unknown before performing an actual experiment.

To remedy this, the EuroPED model was developed by coupling an EPED-like pedestal parameter calculation workflow to the Bohm/gyro-Bohm (BgB) turbulent transport model [32] inside a 1D transport solver. This removes the need to explicitly define β_N and $n_{e,ped}$, as the former is provided by the kinetic profiles resulting

from the self-consistent transport calculation and the latter is defined either via a regression or a neutral penetration model provided within EuroPED. However, this model additionally requires two engineering parameters as input:

- injected auxiliary heating power, P_{tot};
- separatrix electron density, $n_{e,sep}$.

This is seen as an improvement over the EPED model as β_N or $n_{e,ped}$ are difficult to estimate prior to conducting an experiment. Although often the experiments are run in feedback mode where these parameters can actually be chosen. However, from the model point-of-view EPED is still not fully predictive. While the $n_{e,sep}$ is equally difficult to know prior to an experiment, the option to estimate it via the fuelling gas flow rate, R_{gas}, and a neutral penetration model is provided within EuroPED. This connection was excluded from this study as it was not used in the construction of the particular database forming the training set.

The EuroPED model then returns as outputs:

- the top-pedestal electron density, $n_{e,ped}$;
- the top-pedestal electron temperature, $T_{e,ped}$;
- and pedestal width, Δ.

However, due to the implementation details connecting the EPED-like workflow to the transport model, the actual radial locations of the top-pedestal quantities may not coincide in practice. This has some implications on the construction of an appropriate surrogate model, which is further discussed in Section 2.4.

2.2 BNN-NCP Technique

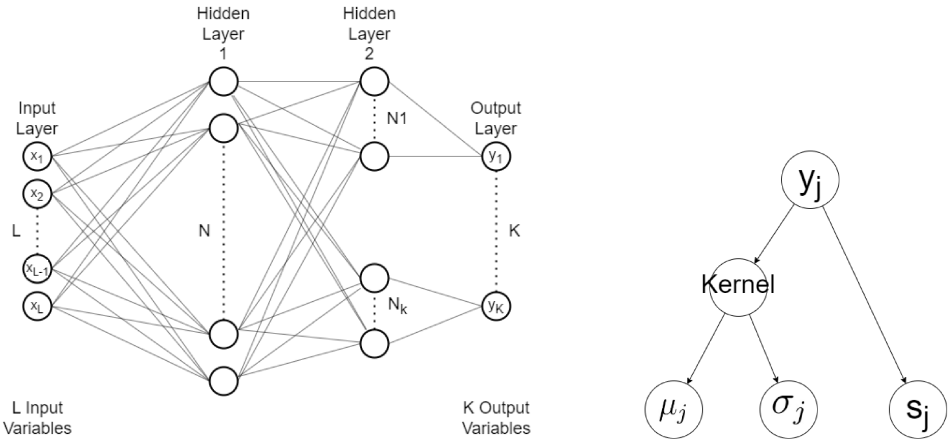
As mentioned earlier, Bayesian neural networks are a specific class of NNs trained using Bayesian inference methods, involving an estimation of a posterior distribution over network weights given observed data. One of the difficulties with training BNNs is that they can be computationally expensive, particularly for large networks with many parameters. By providing a fast and efficient approximation to the posterior distribution, the Noise Contrastive Prior technique can speed up BNN training.

Within this technique, the posterior distribution is approximated by a mixture of the prior distribution over the output and a noise distribution. The network is trained to distinguish between observed data and noise samples generated by the noise distribution during training. Essentially, it uses simulated noise to shift the hypothesis from the probability distribution of the neural network weights to a probability distribution in the data space, requiring less computational resources. In practice, this means that a random point from the joint input distribution is sampled per training epoch. This sampled point is passed forward through the current BNN configuration, and the resulting mean and epistemic uncertainty is compared to a user-defined prior output distribution via some distance metric. This significantly reduces the number of samples required per epoch compared to a typical Monte Carlo method by leveraging the large number of epochs required for training the model to effectively accumulate the necessary statistics, which is similar to the training advantages offered by mini-batching and stochastic gradient descent. Furthermore, the NCP method improves the detection of out-of-distribution (OOD) regions in data space by directly accounting for input variations within its epistemic uncertainty estimation. In summary, this method allows for an approximate calculation of the output distribution without the need to sample the weights to arrive at a reliable estimate of the uncertainty. This approach is presented in [29] for one input and one output problems, in this study a multidimensional approach to fit the needs of the plasma pedestal models is built.

2.3 Neural network architecture and loss function

As a brief overview, the BNN-NCP architecture consists of a number of fully connected feed-forward (i.e. dense) layers with a custom layer inserted just before the output layer, as shown in Figure 1a. This custom layer consists of a *variational layer* and a dense layer with a “softplus" activation function, as shown in Figure 1b, each with only a single neuron. This variational layer represents its weights and biases as joint probability distribution, also known as a *kernel*, which effectively allows them to be sampled to assemble its predictive distribution, $p(\mathbf{y}^*|\mathbf{x}^*, \mathbf{x}, \mathbf{y}, \theta)$, where the * notation is used to differentiate a generic point from one which is explicitly within the training dataset. The “softplus" activation function on the dense layer enforces its output to be positive, allowing it to represent an independent uncertainty estimate. All other neurons in the network have a “leaky ReLU" activation function, which is effectively a piecewise function of two linear components, both with non-zero gradients and joined together at the origin.

Due to the extension of this technique to multiple output variables, the final dense layer before the custom layer is separated into dedicated blocks per output variable. This serves as an attempt to ease the training process by reducing the amount of cross-talk between the network predictions of each variable. It also introduces a degree of flexibility in the last hidden layer, allowing adjustments to suit the complexity of each variable, e.g. the most difficult variables to fit can have a large number of neurons to account for their dependencies. This



(a) Main network scheme

(b) Output layer scheme

Figure 1: Architecture schema of the BNN-NCP developed in this study. In (a), L and K represents the number of input and output variables respectively; N, the number of neurons in the first layer; N_1 ... N_K represent the number of neurons in each of the K packs of neurons associated with every output neuron. In (b), y_j represents a output neuron; kernel, the probability distribution of weights and bias; μ_j and σ_j, the mean and standard deviation of the Kernel; and s_j, an independent neuron.

results in a multi-input multi-output model, transforming input points, $\mathbf{x} = \{x_i\}$, to output points, $\mathbf{y} = \{y_j\}$, which can then be trained on a labelled training dataset, denoted as $D = \{(\mathbf{x}, \mathbf{y})_k\}$.

As a result of the BNN-NCP architecture, there are two modes of evaluating it on a given input point, $\mathbf{x}$:

1. By sampling the probability distribution represented by the kernel in the variational layer, resulting in a stochastic output, $\hat{\mathbf{y}}$;

2. By computing the joint statistical moments of the kernel in the variational layer, resulting in a deterministic mean output, μ, and its standard deviation, σ.

Both of these evaluation modes return another deterministic output from the softplus neuron, $\mathbf{s}$. The predictive output value of the BNN-NCP model is taken as the mean value, i.e. $\mathbf{y}^* \equiv \mu$, from the second evaluation method due to its deterministic nature, although the first evaluation method is still described here due to its importance for the training methodology. Thus, the entire BNN-NCP model can be mathematically represented as a black box function, $\{\mu, \sigma, \mathbf{s}\}^* = f(\mathbf{x}^*)$.

Then, based on these descriptions, the output standard deviation from taking the statistical moment of the kernel, σ, was labelled as the epistemic uncertainty prediction and the output of the independent softplus neuron, $\mathbf{s}$, as the aleatoric uncertainty prediction. Both uncertainty outputs are taken to represent a value of one standard deviation when converting them from statistical moments back into probability distributions.

The cost function used in this study is similar to the one introduced in Ref. [29] except extended into the multidimensional approach. This results in expression provided in Equation (1). The terms are then summed over each output variable, represented by the subscript j.

$$\begin{aligned}\mathcal{L}(\theta) = &- \sum_j \gamma_{\text{nll},j} \cdot \ln p\left(y_j|\mathbf{x},\theta\right) \\ &+ \sum_j \gamma_{\text{epi},j} \cdot \text{KL}\left(p_{\text{prior,epi},j} \,||\, q(\tilde{\mu}_j|\tilde{\mathbf{x}},\theta)\right) \\ &+ \sum_j \gamma_{\text{alea},j} \cdot \text{KL}\left(p_{\text{prior,alea},j} \,||\, \mathcal{N}\left(0, \tilde{s}_j^2\right)\right)\end{aligned} \tag{1}$$

where γ represents a user-defined weight coefficient used to adjust the associated loss term, θ represents the distribution of weights in the BNN, and the $\sim$ notation is used to indicate quantities related to the OOD region sampling. The sampled OOD input values, $\tilde{\mathbf{x}}$, were generated from a normal distribution about each dataset point, $\mathbf{x}$, according to a user-defined input noise term, $\boldsymbol{\Sigma}_{\mathbf{x}}$. This input noise should generally be chosen as a large enough value to ensure the OOD sampling goes adequately beyond the training dataset input values, $\mathbf{x}$, but not so large that the OOD sampling would have difficulty resolving well the area just outside the training dataset.

The first term of the loss function corresponds to the negative log-likelihood (NLL) which represents a goodness-of-fit metric for the approximate predicted posterior distribution, selected to be:

$$p(y_j|\mathbf{x},\theta) \sim \mathcal{N}(\hat{y}_j, s_j) \tag{2}$$

By minimizing this term with respect to the weights and biases, θ, the maximum likelihood value is obtained for the network prediction. As the aleatoric uncertainty prediction, s_j, is used as the weighting parameter for this metric, it also encourages the aleatoric uncertainty to grow where the network cannot match the specific data

point where the combination of other data points and regularization discourage it, e.g. where data noise is present.

The second term represents a metric to anchor and regularize the predicted epistemic uncertainty with weight coefficients, $\gamma_{\text{epi},j}$, which are used to adjust its relative importance in the loss function per output variable. The last term represents the same anchoring and regularization metric for the predicted aleatoric uncertainty, with its own relative weight coefficients per output variable, $\gamma_{\text{alea},j}$.

Both uncertainty regularization terms use the Kullback-Leibler (KL) divergence metric, which effectively compares the similarity of two distributions. For the epistemic term, the approximate output distribution predicted by the network using an OOD input distribution, $q(\mu_j|\tilde{\mathbf{x}}, \theta)$, was compared against an epistemic prior distribution, $p_{\text{prior,epi},j}$, selected to be:

$$p_{\text{prior,epi},j} = \mathcal{N}(y_j, \sigma_{y_j}^2) \tag{3}$$

This represents a normal distribution where the mean is the corresponding output value in the dataset, y_j, and the standard deviation, σ_{y_j}, is a user-defined parameter encapsulating any expected epistemic uncertainty already known to be associated with the data-generating system. This term is expected to generate a significant loss for OOD samples far away from the training point, effectively encouraging the epistemic uncertainty at those OOD samples to grow in order to minimize the overall loss. In order to save computational time during training, only one OOD point is sampled per training point per epoch, which relies on the stochastic gradient descent process to generalize this feature to the entire OOD domain and is one of the essential acceleration tricks provided by the NCP methodology. For the aleatoric term, a normal distribution with zero mean created from the aleatoric uncertainty prediction of the network, $\mathcal{N}(0, s_j^2)$, was compared against an aleatoric prior distribution, selected to be:

$$p_{\text{prior,alea},j} = \mathcal{N}(0, s_{y_j}^2) \tag{4}$$

Again, this represents another normal distribution where the standard deviation, s_{y_j}, is another user-defined parameter encapsulating any expected aleatoric uncertainty already known to be associated with the dataset output values. This is effectively meant to provide a competing pressure on the aleatoric uncertainty prediction, s_j, preventing it from growing to the point where a flat prediction with infinite aleatoric uncertainties becomes the most likely solution.

Basically, this way of treating the output results into the loss function is the main difference with respect to a regular neural network. For this matter, it is particularly meaningful to define several parameters that will help to adjust for the problem to solve. These user-defined values, $\mathbf{\Sigma_x} = \{\sigma_{x_i}\}$, $\mathbf{\Sigma_y} = \{\sigma_{y_j}\}$ and $\mathbf{s_y} = \{s_{y_j}\}$, are meaningful for the model as they are hyperparameters that introduce knowledge from the system and the data region to be modelled into the neural network. While in principle, this would allow extensions to the labelled training dataset, $D' = \{(\mathbf{x}, \mathbf{y}, \mathbf{\Sigma_x}, \mathbf{\Sigma_y}, \mathbf{s_y})_k\}$, for refining the uncertainty predictions, it was decided to leave these extra hyperparameters constant across the entire original training dataset, D, for this study. This is primarily for ease of demonstration for the BNN-NCP methodology, as it simultaneously reduces the burden of collecting the information necessary to refine these priors and the complexity needed to be captured by the training process. More details about the values used for these priors is available in Section 2.5.

2.4 Zero-Dimensional Pedestal Database

During this study, the dataset is derived from a specialized collection JET-ILW plasma discharges [33] and their corresponding EuroPED predictions [16]. While the dataset includes several plasma quantities, this study focused on the input and output variables needed to predict the appropriate pedestal characteristics within a given plasma scenario accroding to EuroPED, as discussed in Section 2.1.

However, in order to improve the generality of the model via the introduction of dimensionless variables, the inverse aspect ratio, $\epsilon = a/R_0$, was used instead of the minor radius, a, and major radius, R_0. This process was also applied to the plasma current, I_p, replacing it by a safety-factor-like parameter, μ, expressed as:

$$\mu = \frac{\mu_0}{2\pi a}\frac{I_p}{B_t} \tag{5}$$

where I_p is expressed in units of A, B_t in T and a in m. While it is uncertain whether pedestal phenomena follow any specific dimensionless scaling, it was attempted to improve its extrapolation capability to other tokamaks.

However, due to the particular application of the pedestal predictions inside the 1D transport solver within EuroPED, the exact top-pedestal temperature and density outputs of the ideal MHD calculation are no longer representative of the plasma state described by the profiles. Thus, the variables provided in this dataset used as the target outputs were:

- electron density at $\psi_{\text{pol}} = \psi_0$, $n_e(\psi_0)$;
- electron temperature at $\psi_{\text{pol}} = \psi_0$, $T_e(\psi_0)$;
- and pedestal width, Δ

where ψ_{pol} is the poloidal magnetic flux, which is used as a geometry-independent radial coordinate, and $\psi_0 \equiv 0.94$

for the purposes of this study. This specific flux coordinate value was chosen due to its statistical similarity to the $\beta_{\text{pol},ped}$ computed from EuroPED.

The pedestal dataset is composed of 1317 entries. However, 257 of them were removed in the training and testing because some of the variables of interest were not available. During the training of the model, the dataset was split into 90% for training and 10% for testing.

In Figure 2, the input data (train and test) used in EuroPED-NN is shown as histograms. The distribution of all the input variables as well as the range of values can be observed.

2.5 Hyperparameters for epistemic and aleatoric uncertainty

As observed in the loss function, Equation (1), several user-defined parameters and hyperparameters related with both the epistemic and aleatoric uncertainty predictions need to be determined. These values to the training routine such that the method to the particular problem. Therefore, additional considerations were required to provide reasonable targets for these values in training the EuroPED-NN.

The width of the OOD region for input sampling, $\mathbf{\Sigma_x}$, was chosen to span a constant proportion of the standard deviation within the dataset of the corresponding input quantity. These specific proportionality values are detailed Appendix A.

Regarding the target epistemic uncertainty, $\mathbf{\Sigma_y}$ in Equation (3), a small constant value was chosen across the entire dataset. So that, when comparing, using KL divergence, in the training loop the target distribution and the output distribution with OOD input, the OOD output distribution will be forced to become wider to match the thin target distribution.

Regarding the target aleatoric uncertainty, $\mathbf{s_y}$ in Equation (4), another small constant value was chosen across the entire dataset. This represents the prior knowledge that the EuroPED model is deterministic and thus, no data error should be present within the output values of the training dataset. The specific values used in this study are available in Appendix A.

All of these selections have been the result of different attempts and the personal reasoning of the authors regarding the nature of the model. Further details of these sensitivities are provided in Section 3, but the main reason is due to the fact that the epistemic and aleatoric uncertainties has been designated their own loss terms, each with their own weight coefficients, γ_{epi} and γ_{alea}, in Equation (1).

Once these parameters informing the BNN-NCP priors have been decided, a rudimentary configuration optimization exercise yielded an architecture with a common dense layer with 20 neurons and specialized dense layers with 8, 8, and 10 neurons for $n_e(\psi_0)$, $T_e(\psi_0)$, and Δ outputs, respectively. Then, the BNN was trained using the *Keras* package [34] in *Tensorflow* [35] (version 2.15.0) with 2000 training epochs and a mini-batch size of 53, taking around 1.7 CPU-hours (10 minutes in 10 cores $\approx$ 100 minutes). The *Adam* optimizer [36] was used with a learning rate of 0.001 and a exponential decay rate of 0.1.

In order to balance the relative importance of the various loss terms in Equation (1), a rudimentary optimization was performed. This resulted in the uncertainty weight coefficients of $\gamma_{\text{epi}} = 0.1$ and $\gamma_{\text{alea}} = 0.1$ for all outputs, j, and individual NLL weight coefficients of $\gamma_{\text{nll}} = \{7, 7, 2\}$ for Δ, $T_e(\psi_0)$, and $n_e(\psi_0)$, respectively.

3 Verification and Characterization of EuroPED-NN

Using the methods and principles described previously, the surrogate model of EuroPED, referred to as EuroPED-NN within this study, was trained. This section details the various methods used to verify its performance via whether it replicates expected behaviours.

3.1 Performance

To test the performance of the model, the EuroPED-NN predictions are first compared against the data points within the training set itself. Figure 3 shows the accuracy for the output variables pedestal width, Δ, electron temperature, $T_e(\psi_0)$, and electron density, $n_e(\psi_0)$, respectively. While they are all reasonably accurate, it can be visually inferred that Δ returns a worse match with the EuroPED database. This may indicate a higher degree of complexity behind the determination of this particular output, as it was found that it already required more neurons in the last BNN layer to achieve this fit quality. Overall, it can be said that the accuracy obtained through the training of the neural network is remarkably good, in spite of the typical difficulties of training multi-output regression networks.

However, another important point of investigating the BNN architecture revolves around its potential to differentiate interpolation regions from extrapolation regions via its uncertainty predictions [29]. This study examines this capability using the epistemic uncertainty prediction, where a generally lower uncertainty is expected in the interpolation regions. Figure 4 shows the relative predicted epistemic and aleatoric uncertainty for Δ, $T_{e,ped}$, and $n_{e,ped}$ using randomly generated input data, as a function of the Euclidean distance of the input data to the centroid of the training input dataset. This distance is calculated after re-scaling these random samples

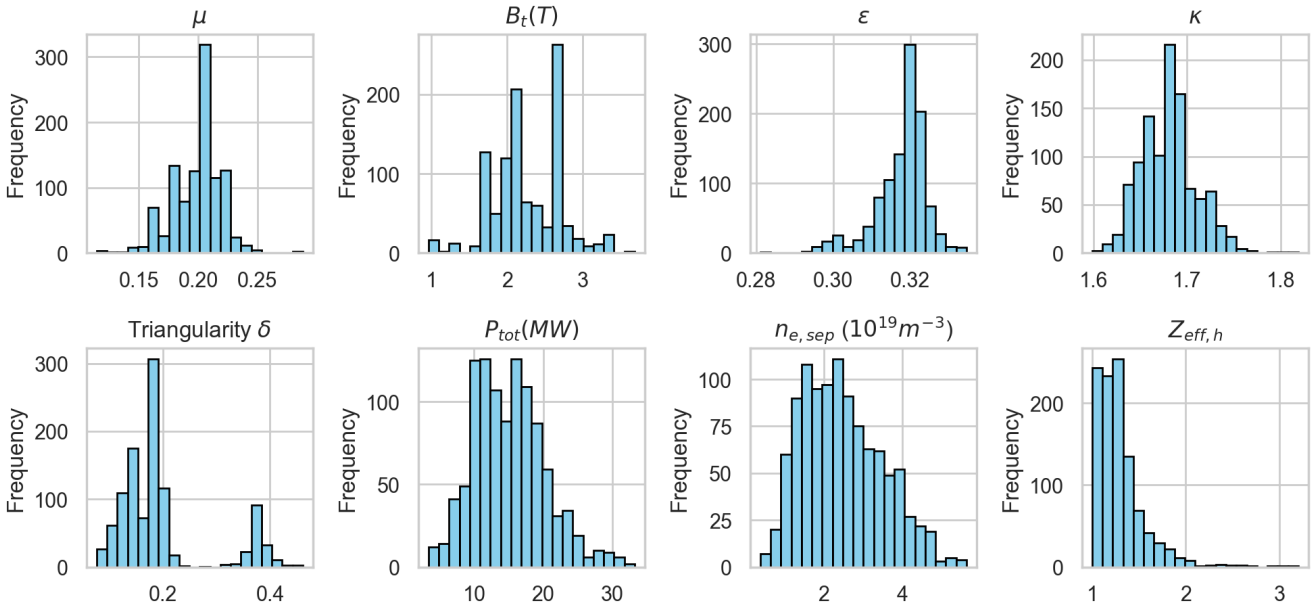


Figure 2: Input data histograms for EuroPED-NN from the dataset used in this study, obtained from [33]. μ stands for the parameter defined in equation 5, B_t for toroidal magnetic field, ϵ for inverse aspect ratio, κ for elongation parameter, δ for triangularity, P_{tot} for injected auxiliary heating power, $n_{e,sep}$ for separatrix electron density and $Z_{eff,h}$ for line-integrated effective charge.

with the same relation used to re-scale the input data distributions to have a mean of 0 and a standard deviation of 1. This is done to avoid introducing any bias on the distance metric due to the orders of magnitude differences spanned by the absolute values of the various input variables. It is also noted that some of the random input data points used may not resemble realistic operational plasma conditions, though this is not a requirement for demonstrating the characteristics of the uncertainty predictions under extrapolation. The training data centroid was chosen as the reference point as it should best represent the region which has the highest data density. Thus, as the distance increases, the epistemic error should increase as well because the neural network would have less information to train its predictive capability. 4000 input entries were randomly generated using two uniform distributions: the first one over the central 60 % of the dataset and the second over the central 98 %. This has been done intentionally to highlight both the points inside the training convex hull and those far from it.

Indeed, Figure 4b demonstrates that when the distance increases, the epistemic uncertainty also increases notably. The random points that fall inside the convex hull of the training data region are distinguished in these plots. In the epistemic case, the uncertainty is expected to be low inside the hull and growing as the points move further out of it. In the aleatoric case, the uncertainty should be higher inside the hull and decrease as the point move further out of it. Indeed, this is roughly observed in Figure 4. As the neural network catches the variance of the training input data through the aleatoric uncertainty, going far from training data means low aleatoric uncertainty as determined by the target s_j values chosen in this study. On the other hand, epistemic uncertainty reflects the precision of the neural network regarding the stochastic variance of the weights and biases with respect the training data, then out of the data region this precision should be much lower, due to the lack of information. Although the general behavior of the uncertainty is well captured in Figure 4, a 1D toy example can is provided in Section 3.2 where it is more intuitively depicted while simultaneously verifying the model via a physical scan.

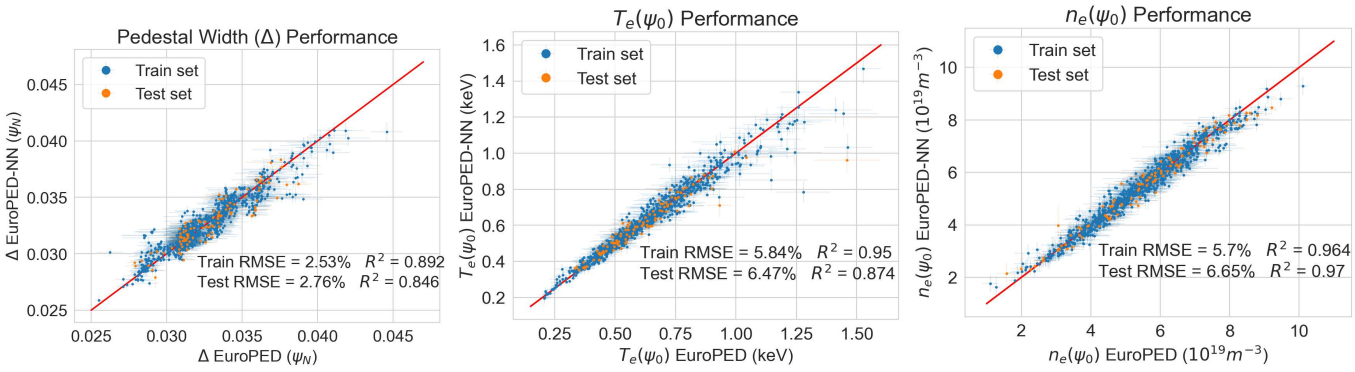
Additionally, in Appendix B a study over the input uncertainties effect on the output result of EuroPED-NN has been done. It can be observed that in most cases the aleatoric uncertainty accounts for the input uncertainty.

3.2 Independent physical analysis

In order to verify the model based on existing physical intuition about pedestal behaviours, this section compares the EuroPED-NN predictions against known empirical relations.

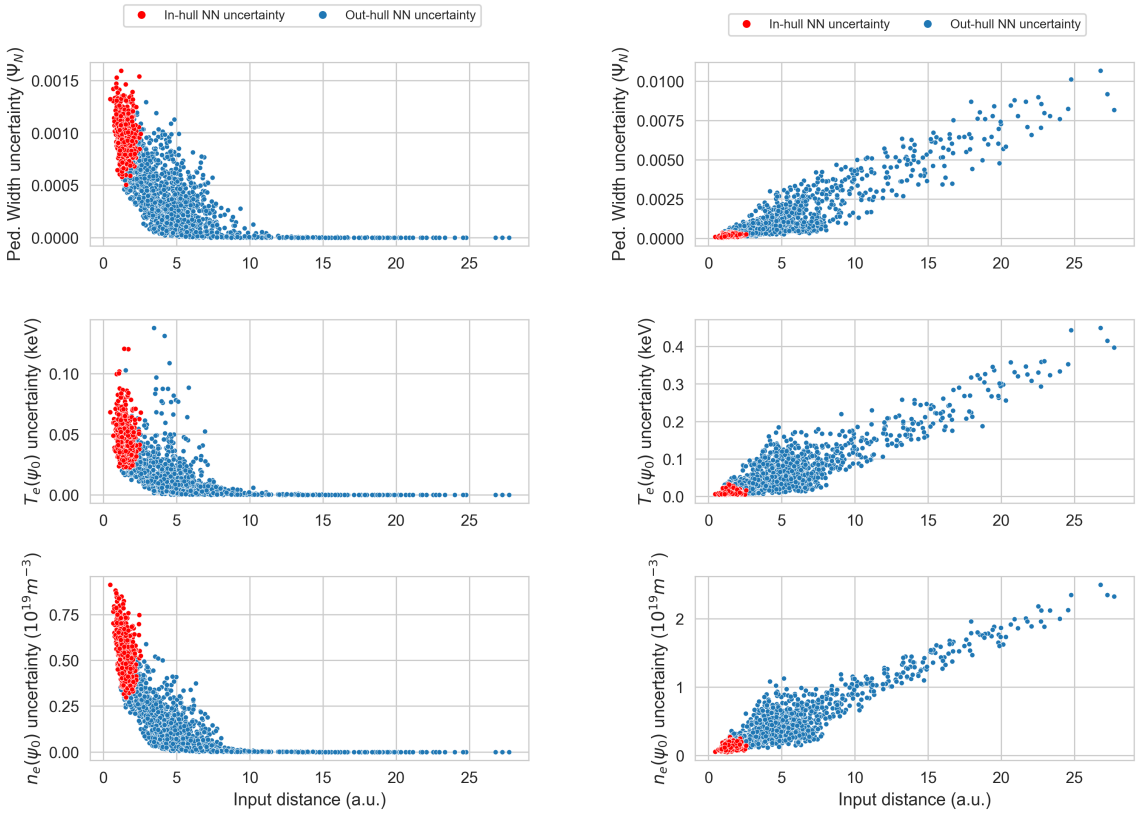
3.2.1 I_p scan

Although the precise behaviour is dependent on many variables, it is generally expected that $n_{e,ped}$ grows when increasing I_p [37]. As $n_e(\psi_0)$ is one of the outputs of the model and I_p is used in the calculation of the model input, μ, given in Equation (5), this trend should also



(a) Δ (b) $n_{e,ped}$ (c) $T_{e,ped}$

Figure 3: EuroPED-NN performance for its three output variables. Train and test data are shown in blue and orange respectively. Epistemic and aleatoric uncertainties are displayed in vertical and horizontal axis respectively.



(a) Absolute epistemic uncertainties (b) Absolute aleatoric uncertainties

Figure 4: Absolute uncertainties of the EuroPED-NN surrogate model predictions against the distance from the input position to the input centroid of the training data. Points represents the predictions for 4000 randomly generated (uniform distribution) input data entries inside the limits of the training dataset. Points inside data region convex hull are presented in red. This is shown for the variables Ped. Width (Δ), $T_e(\psi_0)$ and $n_e(\psi_0)$.

be present in EuroPED-NN. To verify this, a 1D scan is performed over I_p while keeping the other inputs at a fixed value.

Figure 5 clearly shows that $n_e(\psi_0)$ increases with I_p, especially in the training data region between the dotted lines. While it is expected that not all input combinations in this scan correspond to physically achievable plasma scenarios, this trend was found to be present across a majority of random combinations of the other input parameters, of which the 1D scan plots are not shown for brevity. It can then be inferred that it extends to the subspace of physically-relevant parameters

as well. Additionally, as mentioned in Section 3.1, it is noticeable how the region populated by the training database exhibits significantly lower epistemic uncertainty and higher aleatoric uncertainty than the regions outside of it.

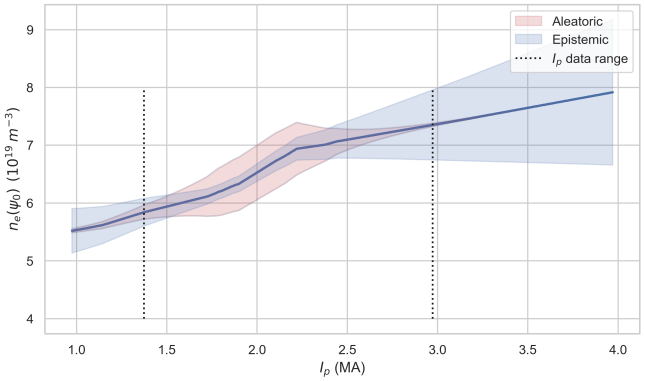


Figure 5: One dimesional scan for input variable I_p. Model prediction for $n_e(\psi_0)$ is shown for different values of I_p and fixed values for the rest of the inputs. Bands for *epistemic* and *aleatoric* uncertainties are appreciated. The data region in I_p is represented as the space between two vertical dashed lines.

3.2.2 $\Delta - \beta_{p,ped}$ relation

As mentioned in Section 2.1, the semi-empirical relation between the pedestal width, Δ, and the normalised plasma pressure, $\beta_{p,ped}$, is an important component of both EPED and EuroPED models. Thus, it is relevant to check whether the EuroPED-NN predictions replicate this relation. However, as the model outputs were reframed to provide the kinetic parameters at $\psi = \psi_0$ while retaining $\beta_p(\psi_0) \approx \beta_{p,ped}$ for purposes of consistency in integrated modelling, it is also important to illustrate that the surrogate model retains this feature by plotting Δ against $\beta_p(\psi_0)$. This is due to the fact that the outputs of EuroPED-NN are: the electron temperature prediction $T_e(\psi_0)$, electron density prediction $n_e(\psi_0)$ and the pedestal width Δ, and all of them are included somehow in the plot. Due to the fact that $T_e(\psi_0)$ and $n_e(\psi_0)$ are needed for calculating $\beta_p(\psi_0)$. Moreover, to perform this calculation, given in Equation (6), some of the model inputs are required: plasma current, effective charge, triangularity, elongation and minor radius. These values are used to calculate β_p, which ultimately is a measure of the confinement, expressed as:

$$\beta_p = \frac{\langle p \rangle}{\langle B_p \rangle^2 / 2\mu_0} \tag{6}$$

where B_p is the poloidal component of the magnetic field, and the angle brackets represent flux-surface averages made around the poloidal angle.

Then this $\Delta - \beta_p(\psi_0)$ representation shows roughly the behavior of the model, as it uses all the output information and some input information. The similarities between the plot generated from predictions of EuroPED and from the reduced model predictions will help to verify the model with a physics-based representation and to measure the performance of the surrogate model.

Figure 6a shows the behaviour of the original EuroPED original predictions for JET-ILW experiment database in the $\Delta - \beta_p(\psi_0)$ space. The original square root relation used in EuroPED and EPED models can be appreciated as: $\Delta = c\sqrt{\beta_{p,ped}}$. This relation was found in the tokamak DIII-D to be held with $c = 0.076$ [14]. However, as EuroPED-NN gives the outputs at $\psi = \psi_0$ and not at top pedestal for convenience purposes, the relation cannot be hold in the exact same way, then some scatter around the square root fit can be found for both EuroPED and EuroPED-NN predictions. This is a result from the discrepancy between $\beta_{p,ped}$ and $\beta_p(\psi_0)$.

Figure 6b show the EuroPED-NN predictions with epistemic uncertainty, across both the training and test datasets from the pedestal database. This means that the plotted uncertainties should correspond to the interpolation region, where the surrogate model is expected to provide decent predictions.

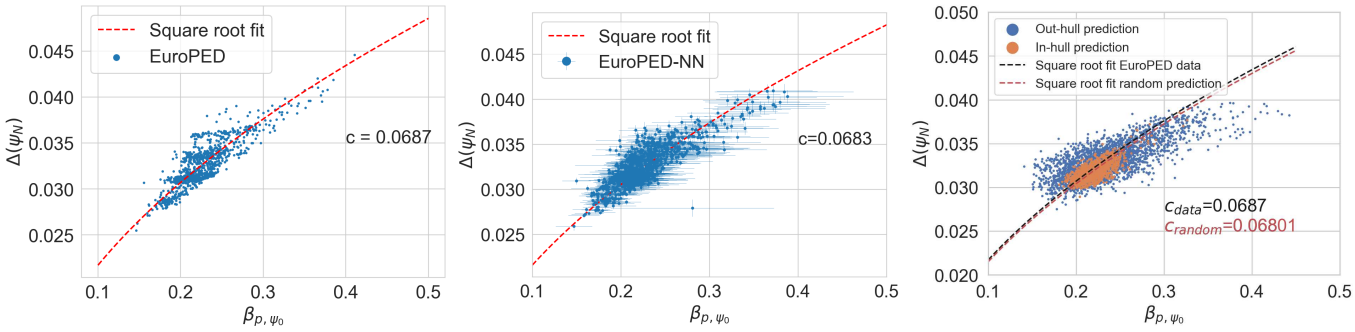
In Figure 6c, the EuroPED-NN predictions for random inputs inside the range of the dataset inputs are shown. 4000 input entries have been generated using two uniform distributions, one using the central 90% range of the dataset and the other the central 60%, this has been done to adequately populate the $\Delta - \beta_p(\psi_0)$ space. A convex hull using the training data has been created in order to check which random points fall inside of it. It can be appreciated in the plot how the points inside the convex hull are following the square root fit with higher precision than the rest. Then the $\Delta - \beta_p(\psi_0)$ relation has been caught and generalized adequately to the training space by EuroPED-NN.

Overall, it is clear that the EuroPED-NN predictions in the $\Delta - \beta_p(\psi_0)$ space is quite similar to the original one in Figure 6a, moreover the random points seem to follow the trend, specially those inside the training convex hull, so the model is capturing the behavior reasonably well.

3.3 Contextualization in data-driven pedestal models

Data-driven methods, in particular neural networks, have been applied to predict plasma pedestal quantities in the past. To put EuroPED-NN in its context a comparison will be carried out with EPED1-NN [21] and PENN [27] models, to know which are their main attributes and where EuroPED-NN stands in relation to them.

Regarding EPED1-NN, the model was developed to



(a) EuroPED dataset (b) EuroPED-NN predictions on dataset points (c) EuroPED-NN predictions on random points

Figure 6: Data from various sources represented in pedestal-width (Δ)-$\beta_p(\psi_0)$ space, with errorbars showing the epistemic uncertainty. Predictions (blue dots) are fitted to show the semi empirical relation $\Delta = c\sqrt{\beta_{p,ped}}$ (square root fit in dashed red). The pedestal width is measured in normalised magnetic flux units ψ_{pol} and $\beta_p(\psi_0)$ is a dimensionless quantity. In **(c)**, prediction points inside the training convex hull are shown in orange.

replicate EPED pedestal model [14]. The inputs of the model are the inputs of EPED: $n_{e,ped}$, Z_{eff}, β_N, I_p, B_t, a, R, κ and δ. The outputs are pedestal pressure p_{ped} and pedestal width. EPED1-NN is built using a regular feed-forward neural network, and its accuracy achieves values around $R^2 = 0.987$ and $R^2 = 0.964$, for the respective outputs, when compared with EPED predictions as described in [21].

On the other hand PENN model is intended to replicate experimental data from JET-ILW pedestal database [33]. The inputs of the model are Z_{eff}, β_N, I_p, B_t, a, κ, δ_{up}, δ_{low}, κ, NBI Power, Total Power, plasma volume and q_{95}. Where δ_{up}, δ_{low} and q_{95} stands for upper triangularity, lower triangularity and q value at $\psi = 0.95$ respectively. PENN is also build using a regular feed-forward neural network and the output variables are the pedestal density $n_{e,ped}$ and temperature $T_{e,ped}$, for which the accuracy in test data when compared with experimental data is $R^2 = 0.91$ and $R^2 = 0.93$, respectively.

Therefore, we can compare both models characteristic with EuroPED-NN, where the set of inputs (explained in the Section 2.1) differ from both EPED1-NN and PENN, then the usage cases are different, as happens between EPED and EuroPED. Depending on the available data one model can be more suitable than the other. The accuracy is reasonably good in all of them, including EuroPED-NN with $R^2_{\Delta,test} = 0.846$, $R^2_{T_e,test} = 0.874$, and $R^2_{n_e,test} = 0.970$ (see Figure 3). However, this accuracy is quoted against the respective datasets of each model, meaning that the insights that can be extracted from this comparison are limited. Regarding uncertainty analysis, only EuroPED-NN is an uncertainty-aware model inherently providing its corresponding model and data uncertainty estimates. Regarding the model applicability, it is known that neither EPED nor EuroPED includes the effect of plasma resistivity. Thus, the empirical model PENN could be a better option when predicting the properties of a more resistive plasma. However, to explore the physics and behavior of the pedestal models in a fast way, both EPED1-NN and EuroPED-NN are also applicable.

3.4 Extrapolation of EuroPED-NN to AUG data

Finally, an experimental pedestal database from ASDEX-Upgrade (AUG) tokamak [38, 39] with 50 shots, and its subsequent EuroPED predictions were available. This data is used to compare the EuroPED-NN predictions with the EuroPED evaluations. As the AUG database parameters lie outside the hull of JET parameters, the exercise constitutes an extrapolation of the model. This will test the adaptability of EuroPED-NN out of JET database values. The distribution of input values from the 50 AUG shots can be compared in Figure 7 with the JET database. The biggest difference between both of them are found in μ, ϵ, P_{tot} and triangularity δ variables.

In the Figure 8 the EuroPED-NN predictions performance in AUG data is displayed for the three output variables. The points are split between those whose input points lie inside the central 99% of JET database range, and those who doesn't. This central 99% boundary was chosen to effectively exclude outliers from artificially expanding the range limits. It can be appreciated how most of the points are outside this range, so it is therefore an extrapolating exercise. For $T_e(\psi_0)$ and $n_e(\psi_0)$ the match with the EuroPED predictions is reasonable with $R^2 = 0.645$ and $R^2 = 0.641$, respectively. However, in the case of the pedestal width, the extrapolation does not look to have succeeded. The reason for this is that the AUG predictions were run using a KBM constant of 0.09 (instead of the standard 0.076) as AUG has found that the standard prediction gives too narrow values. For this reason, the pedestal width variable cannot

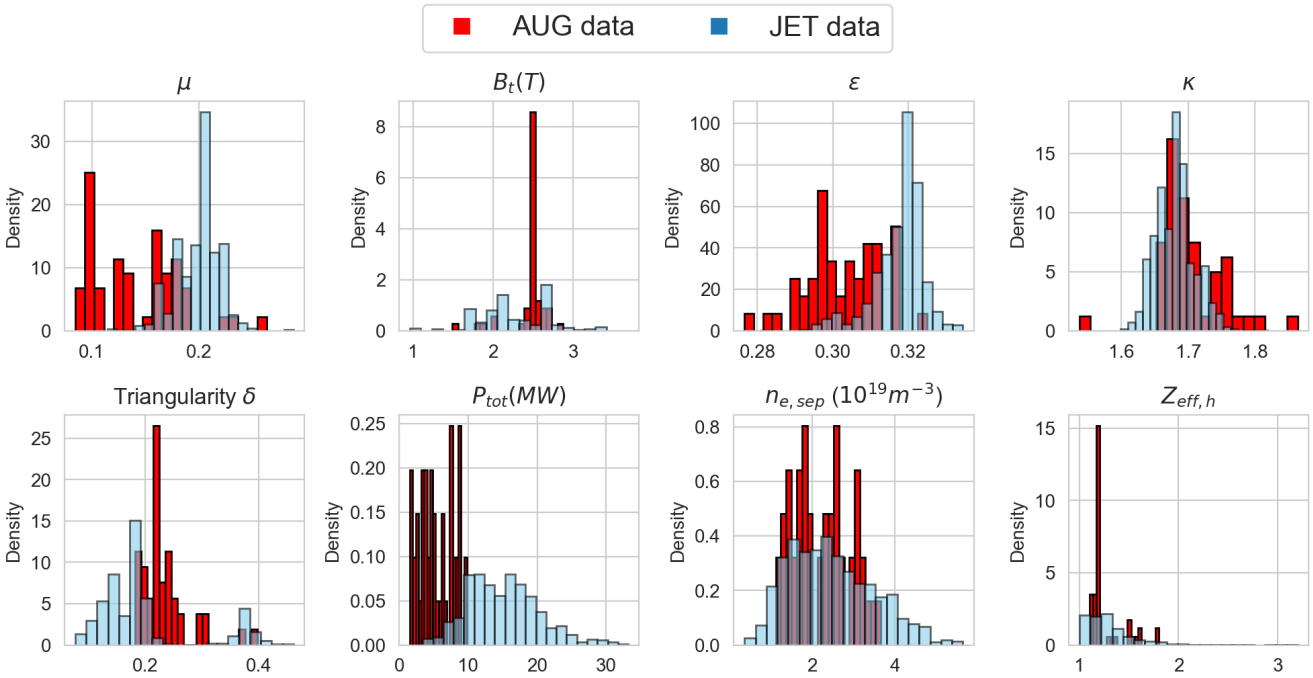


Figure 7: Input data histograms for EuroPED-NN from the JET dataset used to train and test in blue, and from AUG data in red, obtained from [33] and from [38, 39] respectively.

be extrapolated, while the rest variables perform much better.

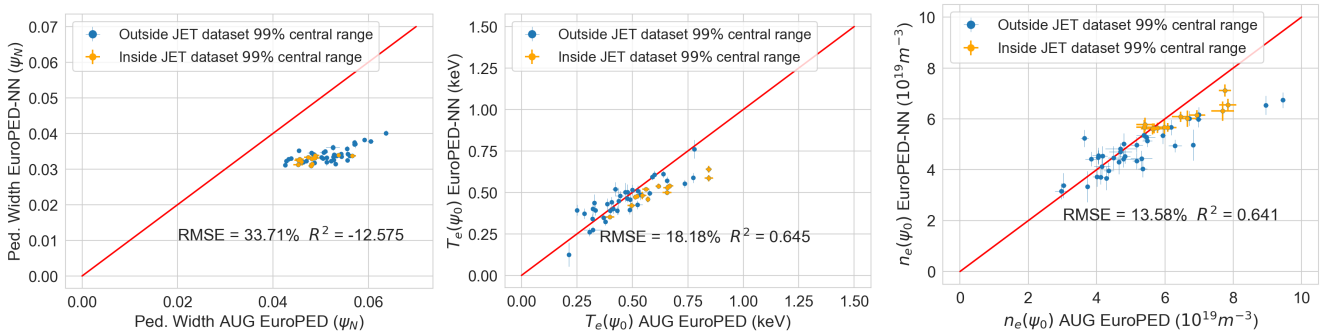
The fact that EuroPED-NN performance in AUG data is reasonable good in the pedestal height quantities, even though the model is extrapolating, might mean that EuroPED-NN is adequately catching the trend between the input data and the pedestal quantities, rather than just adapting to the training data. However, to increase its accuracy and generalization power, future work training EuroPED-NN with data from other machines is encouraged.

4 Extension of the methodology to experimental data and extrapolation to AUG experimental data.

During the training of the EuroPED-NN model, it was found that some of the EuroPED output predictions do not appear to align with the experimental data. Figure 9 shows a comparison between the experimental data and EuroPED-predicted outputs in the database used in this study, where it becomes evident that the prediction points do not follow the 1:1 line. While the objective of this study was to train a surrogate model for EuroPED, its accuracy is ultimately constrained to the accuracy of EuroPED itself. To push the topic of general pedestal surrogate modelling further, an experimental model was constructed using the same EuroPED inputs while swapping the output data to their corresponding experimental values. By employing the same BNN-NCP structure to create this upgraded surrogate model, it exhibits a much better alignment for the same variable, as illustrated in Figure 10 if we compare it with Figure 9. Then this experimental model clearly outperforms EuroPED in the JET-ILW database.

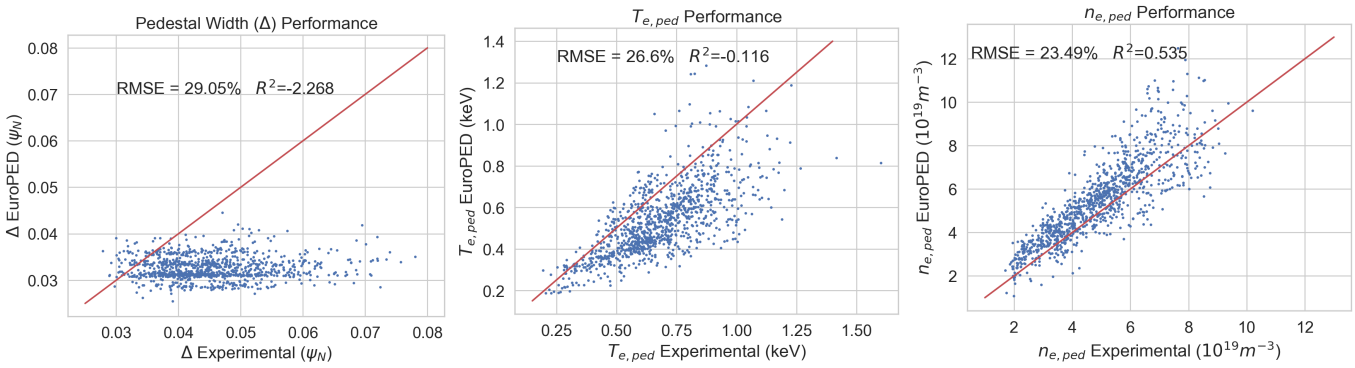
However, when comparing Figures 3 and 10, it becomes evident that the accuracy of this "experimentally-enhanced" surrogate model is significantly lower than that of the EuroPED surrogate model. This suggests that the chosen set of input variables for EuroPED may be insufficient to sufficiently characterize the pedestal physics contained in the dataset, based on its inability to obtain a similarly high model performance, as determined by its RMSE and R^2. Guided by recent pedestal physics literature, an attempt was made to include the normalized plasma collisionality, ν^*, as an input variable. However, these efforts have proven inadequate in explaining why the model performance cannot be recovered between the models. Nevertheless, this marks a promising initial study into the development of a sufficient set of parameters to fully describe pedestal physics and further research in this direction is strongly recommended.

It may seem reasonable to compare this experimental pedestal model with the previously described PENN model [27] as both of them are based on experimental data, although the set of input variables differ from each other. As mentioned, PENN model has an accuracy in test data of $R^2 = 0.91$ and $R^2 = 0.93$, for $n_{e,ped}$ and



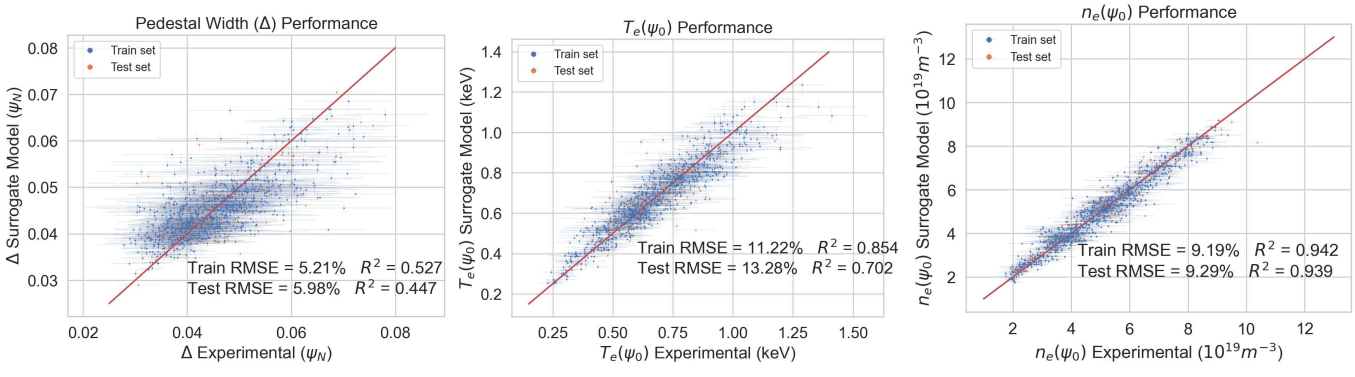
(a) Δ (b) $T_e(\psi_0)$ (c) $n_e(\psi_0)$

Figure 8: EuroPED-NN performance in AUG data for its three output variables. Predictions corresponding to input points inside and outside 99 % central JET data range are shown in orange and blue, respectively. Epistemic and aleatoric uncertainties are displayed in vertical and horizontal axis respectively.



(a) Δ (b) $T_{e,ped}$ (c) $n_{e,ped}$

Figure 9: EuroPED predictions performance compared with experimental values for its three output parameters.



(a) Δ (b) $T_{e,ped}$ (c) $n_{e,ped}$

Figure 10: Experimental surrogate model predictions for its three output parameters compared with the respective experimental data. Epistemic and aleatoric uncertainties are displayed in vertical and horizontal axis respectively.

$T_{e,ped}$ respectively, while the experimental model developed in this section is associated with test set accuracy of $R^2 = 0.94$ and $R^2 = 0.70$. Then it can be seen how the experimental model performs better for $n_{e,ped}$, while PENN performs significantly better for $T_{e,ped}$. The likely explanation lies in the input parameters of each model. The model utilized in this study benefits from $n_{e,sep}$ as input parameter, which is closely related with $n_{e,ped}$. In contrast, the PENN model boasts a greater number of inputs, enhancing its ability to provide more comprehensive information for accurate predictions. It is worth noting that the experimental model inputs have been purposefully selected to correspond with the EuroPED inputs, so an input parameter analysis has not been carried out.

Finally, the experimental pedestal database from

ASDEX-Upgrade (AUG) tokamak [38, 39] will be used to compare the experimental model predictions with the data from AUG. Again, as the AUG database parameters lie outside the hull of JET parameters, the exercise constitutes an extrapolation of the model. Also note that EuroPED model does not play any role in this exercise, as the experimental model has been trained exclusively with experimental data, not with the subsequent EuroPED evaluations, although both have the same set of inputs. In Figure 11, the accuracy of the experimental model in the AUG cases is represented for pedestal width, $T_{e,ped}$ and $n_{e,ped}$, the prediction points are split between those whose input values fall inside the central 99% of JET training database range and those that do not. It can be inferred from the accuracy metrics that the experimental NN model performs significantly worse in experimental AUG pedestal values, although in the plots it can be seen how in the case of $T_{e,ped}$ and $n_{e,ped}$ it roughly gets the trend of the 1:1 line. Interestingly, the model does not return pedestal values corresponding to the AUG data even for the AUG points which lie within the JET training parameter space. A possible explanation for this can be due to the fact that the training dataset contains ∼1000 entries, which may not be enough for the NN to learn the necessary relations to generalize the problem in the experimental case, even within its own training space. Another explanation could be the election of input parameters, which, as commented, may not be enough to recover the experimental pedestal values. A more thorough investigation into the input variables selection or the incorporation of other physical constraints is recommended to improve the results of cross-machine comparisons, especially if the databases used cannot be significantly expanded for future studies.

5 Conclusion

In this study, it has been shown that EuroPED-NN, trained using the BNN-NCP technique, is able to reproduce the results and behavior of EuroPED JET database with good accuracy, whose mean prediction provides similar performance to that of a simple feed-forward neural network, showing performances. Obtaining errors under 7 % for $T_e(\psi_0)$ and $n_e(\psi_0)$ and under 3 % for pedestal width. However the use of this BNN-NCP technique has allowed to identify extrapolation and interpolation regions via the epistemic and/or aleatoric uncertainties, as the epistemic uncertainty generally grows with distance from dataset convex hull, while the aleatoric uncertainty falls. On top of that, the use of these aleatoric and epistemic uncertainties allows the model to be uncertainty aware in the predictions, so that the prediction's confidence can be known.

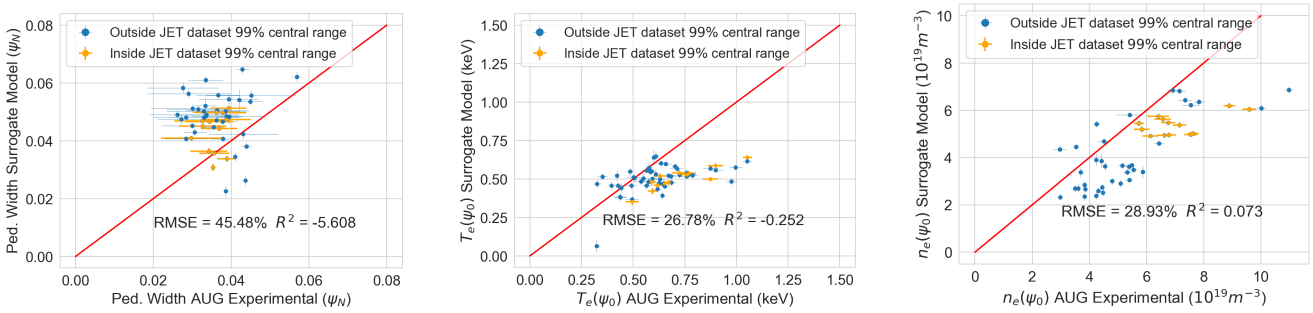
EuroPED-NN has also been tested using physical reasoning. First, a one-dimensional scan calculating $n_e(\psi_0)$ for different values of I_p has been performed proving the expected growth of I_p, while showing the behavior of the prediction's uncertainty. Second, the $\Delta - \beta_{p,ped}$ semi-empirical relation included inside EuroPED has been tested to hold in EuroPED-NN, using both JET dataset inputs and random inputs, so that the relation is correctly generalized.

Finally, after explaining the method, building the surrogate model and performing the posterior analysis, it can be said that the EuroPED-NN is completely functional in JET-ILW data. Additionally, the extrapolation of EuroPED-NN to AUG tokamak data has been tested, obtaining reasonable accuracy for $T_e(\psi_0)$ and $n_e(\psi_0)$ of around $R^2 = 0.64$. However, the pedestal width accuracy was much lower as EuroPED was run in AUG data using a different constant for the KBM constraint. In general, better performance in this exercise would typically require a bigger set of training data and/or training data from other tokamaks. Future work is encouraged in this direction.

To sum up, EuroPED-NN predictions closely match EuroPED predictions in JET input data, with errors of less than 7% for both electron temperature ($R^2 = 0.874$) and electronic density ($R^2 = 0.970$), and less than 3% for pedestal width ($R^2 = 0.846$). The model identifies the areas where the prediction uncertainty is higher and lower, which is helpful for understanding its reliability. Furthermore, the model's performance was validated by comparing its predictions with known physical principles, demonstrating its understanding of some plasma physics principles involved in EuroPED.

The surrogate model EuroPED-NN is mainly intended to be used in plasma control, scenario development and physics exploration of pedestal plasma. The advantage with respect to EuroPED model is not only the fast computation but also the uncertainty awareness. While this uncertainty is split in model uncertainty (epistemic) and data uncertainty (aleatoric) for further clarity. EuroPED-NN proves to be a valuable tool for various studies that can leverage any of this features. The model architecture and weights are located in this GitHub link `https://github.com/alexpanera/EuroPED-NN`, any feedback on the use of the model will be appreciated.

In Section 4, the same method and architecture developed for EuroPED-NN was also applied to experimental data using the same JET database, proving the adaptability of the method and bridging the gap between EuroPED predictions and experimental data, a challenge that EuroPED-NN inherently cannot address. While its performance surpassed that of EuroPED when compared with JET experimental data, it did not match the performance of EuroPED-NN compared to EuroPED. Fur-



(a) Δ (b) $T_{e,ped}$ (c) $n_{e,ped}$

Figure 11: Experimental surrogate model predictions for pedestal width, $T_e(\psi_0)$ and $n_e(\psi_0)$ compared with the respective experimental data from AUG tokamak. Points are split regarding the position of the input values with respect to the JET training database. Epistemic and aleatoric uncertainties are displayed in vertical and horizontal axis respectively

thermore, this experimental model was tested with AUG data, revealing its limited extrapolating power due to its low performance in this different tokamak. This could be due to the low number of training data points (~ 1000) and/or the election of the EuroPED input variables. Future work is encouraged towards building a pedestal experimental model that can be generalised to various machines, likely using data from different machines.

In the Appendix B, some additional analysis on EuroPED-NN uncertainties is performed, and in Appendix C the same methodology described has been implemented in the EPED model to prove again the adaptability of the method.

Furthermore, the method developed in this study (multidimensional BNN-NCP), based on a previous study detailing its methodology [29], can be extended to build uncertainty aware models even beyond plasma physics, offering opportunities in diverse fields. In essence, the multidimensional BNN-NCP model with modified architecture that has been developed in this study holds immense potential for both scientific and non-scientific applications, by enabling the construction of precise uncertainty aware neural network models across multiple dimensions, improving the speed of regular Bayesian neural networks.

Acknowledgements

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 author(s) 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. In particular, the work has been carried out within the framework of ENR-MOD.01.FZJ project.

Data availability

Raw experimental data were generated at the Joint European Torus (JET) facility. Derived data and simulation results supporting the findings of this study are available from the corresponding author upon reasonable request.

References

[1]F. Wagner, "A quarter-century of H-mode studies", Plasma Physics and Controlled Fusion **49**, B1 (2007).

[2]R. J. Groebner, K. H. Burrell, and R. P. Seraydarian, "Role of edge electric field and poloidal rotation in the L-H transition", Phys. Rev. Lett. **64**, 3015–3018 (1990).

[3]K. H. Burrell, "Effects of E×B velocity shear and magnetic shear on turbulence and transport in magnetic confinement devices", Physics of Plasmas **4**, 1499–1518 (1997).

[4]P. H. Diamond, S.-I. Itoh, K. Itoh, and T. S. Hahm, "Zonal flows in plasma – a review", Plasma Physics and Controlled Fusion **47**, R35 (2005).

[5]M. Curie, D. Hatch, M. Halfmoon, J. Chen, D. Brower, E. Hassan, M. Kotschenreuther, S. Mahajan, R. Groebner, and D.-D. team, "Gyrokinetic simulations compared with magnetic fluctuations diagnosed with a Faraday-effect radial interferometer-polarimeter in the DIII-D pedestal", Nuclear Fusion **62**, 126061 (2022).

[6]G. Ren, J. Li, L. Wei, and Z.-X. Wang, "Role of secondary long wavelength fluctuation in nonlinear saturation of the kinetic ballooning mode in tokamak plasmas", Nuclear Fusion **62**, 096034 (2022).

[7]D. Hatch, M. Kotschenreuther, S. Mahajan, M. Pueschel, C. Michoski, G. Merlo, E. Hassan, A. Field, L. Frassinetti, C. Giroud, J. Hillesheim, C. Maggi, C. P. von Thun, C. Roach, S. Saarelma, D. Jarema, F. Jenko, and J. Contributors, "Microtearing modes as the source

of magnetic fluctuations in the JET pedestal", Nuclear Fusion **61**, 036015 (2021).

[8]E. Solano, E. Delabie, G. Birkenmeier, C. Silva, J. Hillesheim, P. Vincenzi, A. Nielsen, J. Rasmussen, A. Baciero, S. Aleiferis, I. Balboa, A. Boboc, C. Bourdelle, I. Carvalho, P. Carvalho, M. Chernyshova, R. Coelho, T. Craciunescu, R. Dumont, P. Dumortier, E. la Luna, J. Flanagan, M. Fontana, J. Fontdecaba, L. Frassinetti, D. Gallart, J. Garcia, E. Giovannozzi, C. Giroud, W. Gromelski, R. Henriques, L. Horvath, P. Jacquet, I. Jepu, A. Kappatou, D. Keeling, D. King, E. Kowalska-Strzęciwilk, M. Lennholm, E. Lerche, E. Litherland-Smith, V. Kiptily, K. Kirov, A. Loarte, B. Lomanowski, C. Maggi, M. Mantsinen, A. Manzanares, M. Maslov, A. Meigs, I. Monakhov, R. Morales, D. Nina, C. Noble, V. Parail, F. Diaz, E. Pawelec, G. Pucella, D. Réfy, E. Righi-Steele, F. Rimini, T. Robinson, S. Saarelma, M. Sertoli, A. Shaw, S. Silburn, P. Sirén, Ž. Štancar, H. Sun, G. Szepesi, D. Taylor, E. Tholerus, S. Vartanian, G. Verdoolaege, B. Viola, H. Weisen, T. Wilson, and J. Contributors, "Recent progress in L-H transition studies at JET: tritium, helium, hydrogen and deuterium", Nuclear Fusion **62**, 076026 (2022).

[9]L. Vermare, P. Hennequin, C. Honoré, M. Peret, G. Dif-Pradalier, X. Garbet, J. Gunn, C. Bourdelle, F. Clairet, J. Morales, R. Dumont, M. Goniche, P. Maget, R. Varennes, and the WEST Team, "Formation of the radial electric field profile in the WEST tokamak", Nuclear Fusion **62**, 026002 (2021).

[10]C. Silva, E. Solano, J. Hillesheim, E. Delabie, S. Aleiferis, G. Birkenmeier, L. Gil, C. Giroud, E. Litherland-Smith, R. Morales, D. Nina, A. Silva, and J. Contributors, "Structure of the JET edge radial electric field in He and D plasmas", Nuclear Fusion **61**, 126006 (2021).

[11]A. G. Peeters, "The bootstrap current and its consequences", Plasma Physics and Controlled Fusion **42**, B231 (2000).

[12]A. W. Leonard, "Edge-localized-modes in tokamaks", Physics of Plasmas **21**, 090501 (2014).

[13]H. R. Wilson, P. B. Snyder, G. T. A. Huysmans, and R. L. Miller, "Numerical studies of edge localized instabilities in tokamaks", Physics of Plasmas **9**, 1277–1286 (2002).

[14]P. B. Snyder, R. J. Groebner, A. W. Leonard, T. H. Osborne, and H. R. Wilson, "Development and validation of a predictive model for the pedestal height", Physics of Plasmas **16**, 056118 (2009).

[15]H. Zohm, "The physics of edge localized modes (ELMs) and their role in power and particle exhaust", Plasma Physics and Controlled Fusion **38**, 1213 (1996).

[16]S. Saarelma, L. Frassinetti, P. Bilkova, C. D. Challis, A. Chankin, R. Fridström, L. Garzotti, L. Horvath, and C. F. Maggi, "Self-consistent pedestal prediction for JET-ILW in preparation of the DT campaign", Physics of Plasmas **26**, 072501 (2019).

[17]S. V. Mulders, F. Felici, O. Sauter, J. Citrin, A. Ho, M. Marin, and K. van de Plassche, "Rapid optimization of stationary tokamak plasmas in RAPTOR: demonstration for the ITER hybrid scenario with neural network surrogate transport model QLKNN", Nuclear Fusion **61**, 086019 (2021).

[18]J. Degrave, F. Felici, J. Buchli, M. Neunert, B. Tracey, F. Carpanese, T. Ewalds, R. Hafner, A. Abdolmaleki, D. de las Casas, C. Donner, L. Fritz, C. Galperti, A. Huber, J. Keeling, M. Tsimpoukelli, J. Kay, A. Merle, J.-M. Moret, S. Noury, F. Pesamosca, D. Pfau, O. Sauter, C. Sommariva, S. Coda, B. Duval, A. Fasoli, P. Kohli, K. Kavukcuoglu, D. Hassabis, and M. Riedmiller, "Magnetic control of tokamak plasmas through deep reinforcement learning", Nature **602**, 414–419 (2022).

[19]A. Ho, J. Citrin, C. Bourdelle, Y. Camenen, F. J. Casson, K. L. van de Plassche, and H. Weisen, "Neural network surrogate of QuaLiKiz using JET experimental data to populate training space", Physics of Plasmas **28**, 032305 (2021).

[20]K. L. van de Plassche, J. Citrin, C. Bourdelle, Y. Camenen, F. J. Casson, V. I. Dagnelie, F. Felici, A. Ho, and S. Van Mulders, "Fast modeling of turbulent transport in fusion plasmas using neural networks", Physics of Plasmas **27**, 022310 (2020).

[21]O. Meneghini, S. P. Smith, P. B. Snyder, G. M. Staebler, J. Candy, E. Belli, L. Lao, M. Kostuk, T. Luce, T. Luda, J. M. Park, and F. Poli, "Self-consistent core-pedestal transport simulations with neural network accelerated models", Nuclear Fusion **57**, 086034 (2017).

[22]S. Morosohk, A. Pajares, T. Rafiq, and E. Schuster, "Neural network model of the multi-mode anomalous transport module for accelerated transport simulations", Nuclear Fusion **61**, 106040 (2021).

[23]J. Zhu, C. Rea, K. Montes, R. Granetz, R. Sweeney, and R. Tinguely, "Hybrid deep-learning architecture for general disruption prediction across multiple tokamaks", Nuclear Fusion **61**, 026007 (2020).

[24]J. Kates-Harbeck, A. Svyatkovskiy, and W. Tang, "Predicting disruptive instabilities in controlled fusion plasmas through deep learning", Nature **568**, `10.1038/s41586-019-1116-4` (2019).

[25]O. Meneghini, G. Snoep, B. Lyons, J. McClenaghan, C. Imai, B. Grierson, S. Smith, G. Staebler, P. Snyder, J. Candy, E. Belli, L. Lao, J. Park, J. Citrin, T. Cordemiglia, A. Tema, and S. Mordijck, "Neural-network accelerated coupled core-pedestal simulations with self-consistent transport of impurities and compatible with iter imas", Nuclear Fusion **61**, 026006 (2020).

[26]E. U. Zeger, F. M. Laggner, A. Bortolon, C. Rea, O. Meneghini, S. Saarelma, B. S. Sammuli, S. P. Smith, and J. Zhao, "Prediction of diii-d pedestal structure from externally controllable parameters", IEEE Transactions on Plasma Science **49**, 3212–3227 (2021).

[27]A. Gillgren, E. Fransson, D. Yadykin, L. Frassinetti, P. Strand, and J. Contributors, "Enabling adaptive pedestals in predictive transport simulations using neural networks", Nuclear Fusion **62**, 096006 (2022).

[28]D. P. Kingma and M. Welling, "Auto-encoding variational bayes", arXiv preprint arXiv:1312.6114 (2013).

[29]D. Hafner, D. Tran, T. Lillicrap, A. Irpan, and J. Davidson, "Noise contrastive priors for functional uncertainty", in Proceedings of the 35th uncertainty in artificial intelligence conference, Vol. 115, edited by R. P. Adams and V. Gogate, Proceedings of Machine Learning Research (22–25 Jul 2020), pp. 905–914.

[30]E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods", Machine Learning **110**, 457–506 (2021).

[31]A. D. Kiureghian and O. Ditlevsen, "Aleatory or epistemic? does it matter?", Structural Safety **31**, Risk Acceptance and Risk Communication, 105–112 (2009).

[32]M. Erba, T. Aniel, V. Basiuk, A. Becoulet, and X. Litaudon, "Validation of a new mixed Bohm/gyro-Bohm model for electron and ion heat transport against the ITER, Tore Supra and START database discharges", Nuclear Fusion **38**, 1013 (1998).

[33]L. Frassinetti, S. Saarelma, G. Verdoolaege, M. Groth, J. Hillesheim, P. Bilkova, P. Bohm, M. Dunne, R. Fridström, E. Giovannozzi, F. Imbeaux, B. Labit, E. de la Luna, C. Maggi, M. Owsiak, and R. Scannell, "Pedestal structure, stability and scalings in JET-ILW: the EUROfusion JET-ILW pedestal database", Nuclear Fusion **61**, 016001 (2020).

[34]F. Chollet et al., *Keras*, (2015) `https://keras.io`.

[35]M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al., "Tensorflow: large-scale machine learning on heterogeneous distributed systems", arXiv preprint arXiv:1603.04467 (2016).

[36]D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", in 3rd international conference on learning representations, ICLR 2015, san diego, ca, usa, may 7-9, 2015, conference track proceedings, edited by Y. Bengio and Y. LeCun (2015).

[37]A. E. Hubbard, "Physics and scaling of the h-mode pedestal", Plasma Physics and Controlled Fusion **42**, A15 (2000).

[38]T. Luda, C. Angioni, M. Dunne, E. Fable, A. Kallenbach, N. Bonanomi, T. Lunt, P. Schneider, M. Siccinio, G. Tardini, the ASDEX Upgrade Team, and the EUROfusion MST1 Team, "Validation of a full-plasma integrated modeling approach on asdex upgrade", Nuclear Fusion **61**, 126048 (2021).

[39]T. Luda, C. Angioni, M. Dunne, E. Fable, A. Kallenbach, N. Bonanomi, P. Schneider, M. Siccinio, G. Tardini, T. A. U. Team, and T. E. M. Team, "Integrated modeling of asdex upgrade plasmas combining core, pedestal and scrape-off layer physics", Nuclear Fusion **60**, 036023 (2020).

A OOD Hyperparameters

This section provides some data for the OOD related hyperparameters referenced in the main body. These values were specifically chosen for this study based on the training attempts and expert reasoning of the authors. They must be reevaluated when applying the BNN-NCP technique to other applications.

In the case of the OOD sampling, as mentioned in the Section 2.5, the width of OOD region for input sampling was chosen to be 2 times the standard deviation of the dataset input variables. This range resulted wide enough to get the desired results.

Regarding the target epistemic uncertainty $\mathbf{\Sigma_y}$, the value chosen in the three output variables was $\mathbf{\Sigma_y} = 0.001$ in normalised units. The target aleatoric uncertainty was also chosen to be small, $\mathbf{s_y} = 0.0001$ also in normalised units.

B Input uncertainty study in EuroPED-NN

In the methodology developed in this work two types of uncertainties are distinguished, epistemic and aleatoric. However, it is interesting to study how the input uncertainties are translated into the output uncertainties propagating them through the EuroPED-NN model. Those input uncertainties are estimated from the measurement signals and the values appear in Table 1.

	σ_x
μ	0.02
B_t (T)	0.001
ϵ	0.01
κ	0.01
Triangularity δ	0.01
P_{tot} (MW)	0.5
$n_{e,sep}$ $(10^{19}m^{-3})$	0.7
Z_{eff}	0.05

Table 1: Values for the expected uncertainty in the input data.

The method to develop this study was similar to the procedure used in OOD sampling. The dataset input values were modified summing to each a sample of 1000 values generated by the distribution $\mathcal{N}(0, \sigma_x)$. Then this 1000 points generated with just one input entry are passed through EuroPED-NN, from this 1000 values for each of the outputs we obtain the standard deviation, which correspond with the input uncertainty translated to the output.

Figure 12 shows the results of this study comparing the sampled uncertainty with the aleatoric uncertainty for all the points in the dataset. Aleatoric uncertainty was chosen for comparison due to the closest relation with the data variation, on top of that epistemic uncertainty was consistently lower than both aleatoric and sampled uncertainty. It is noticeable how in the output variables $n_e(\psi_0)$ and $T_e(\psi_0)$ the aleatoric uncertainty accounts for all the sampled uncertainty, serving as a lower bound. In the case of Δ both aleatoric and sampled uncertainty are in the same order. It is expected that the aleatoric uncertainty captures the input uncertainties translated to the outputs, as the aleatoric uncertainty catches the variance in the data, which is highly related with the measurements uncertainty.

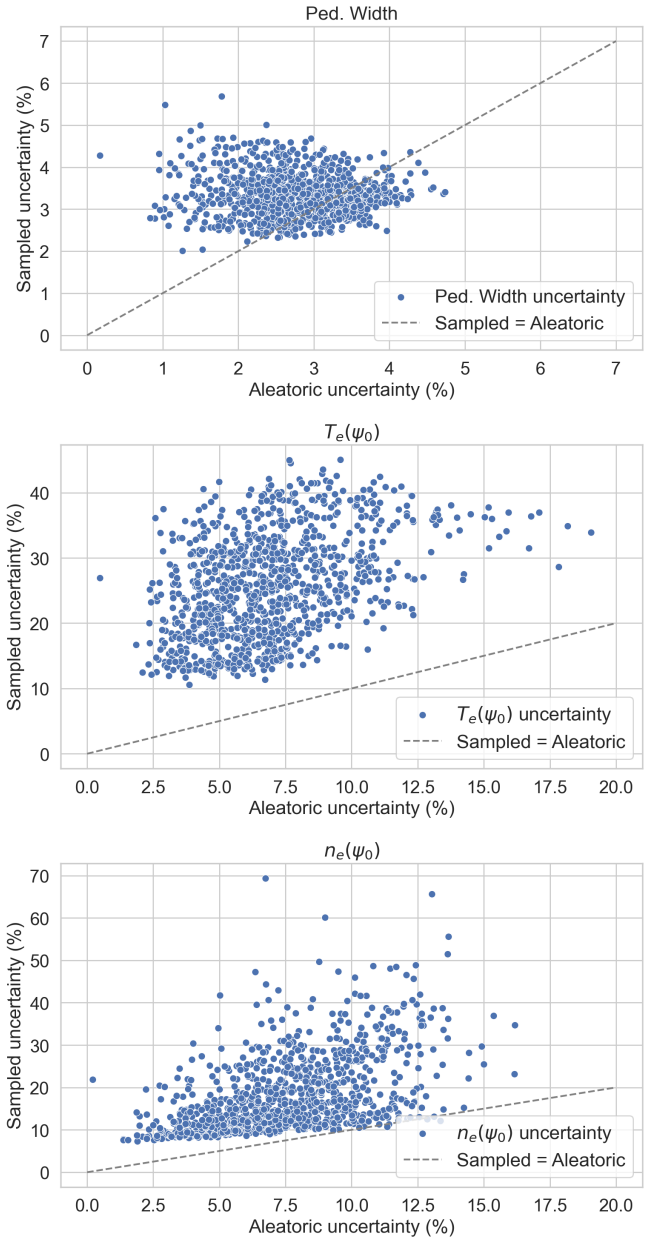


Figure 12: Comparison between sampled uncertainty coming from the input uncertainties and aleatoric uncertainty for all dataset points.

Note that the $n_{e,sep}$ uncertainty dominates the results,

due to the large σ_x values, as shown in Table 1, and due to the importance of the variable $n_{e,sep}$ in the calculation of the final output. To test impact of this choice, the sampling was repeated with lower values of σ_x for $n_{e,sep}$, obtaining significantly lower sampled uncertainties.

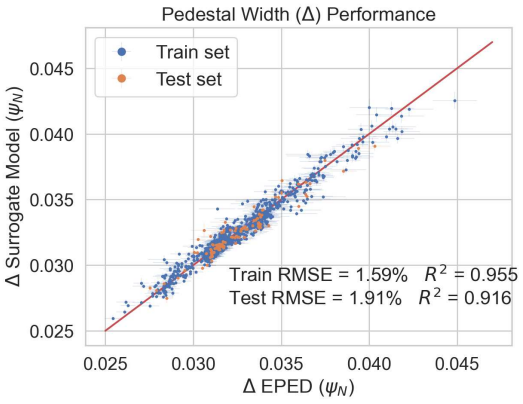
Including the effect of the input uncertainty into the methodology of the neural network would be a desired feature for the model. In this work some attempts were performed, however the result was not satisfactory in most cases. Then it is suggested for future work.

C EPED surrogate

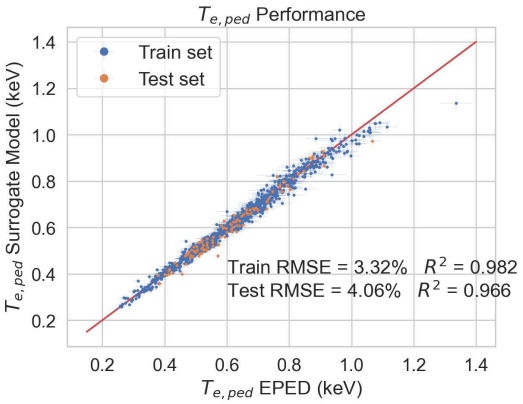
Following the same principles and methodology developed for EuroPED-NN, and also to show the adaptability of it, the surrogate model of EPED has been created. Taking the inputs and outputs of EPED listed in the Section 2.1, a BNN-NCP with the same number of neurons per layer and the same architecture has been trained to check if the BNN captures the model and the $\Delta - \beta_{p,ped}$ relation. In this case the outputs will correspond to the values of T_e at top pedestal ($T_{e,ped}$) and the pedestal width Δ. The dataset used will be the same as in EuroPED-NN, where the columns for EPED runs were already present. The model has been trained for 1000 epochs. The performance is shown in the Figures 13a and 13b. Both of them seem to have a good fit with the 1:1 line, and the performance is always lower than 5 % in both test and train datasets. RMSE is consistently lower than in EuroPED-NN, which evidence the lower complexity of EPED that allows better performance with the same neural network architecture.

As in EuroPED, the semi empirical relation $\Delta = c\sqrt{\beta_{p,ped}}$ is quite relevant in EPED, so we expect the EPED surrogate to hold the same relation. Then we will perform the physical validation as before. Figure 14a shows the EPED predictions and the square root relation in them. In Figure 14b the predictions of the EPED surrogate are also shown in the pedestal width (Δ)-$\beta_{p,ped}$ space, where the square root fit coefficient is almost the same as in the original EPED model. Finally in Figure 14c 4000 random input entries were generated in the same manner as in Section 3.2.2 and they were passed through EPED surrogate. It is appreciable that prediction points that fall inside the training convex hull follow better the trend line, although the out-hull points also follow it reasonably well. So again, the surrogate model generalises correctly the physic relation.

Additionally, it makes sense to compare this EPED surrogate model with the model EPED1-NN [21] that also replicates EPED using a neural network, as described in section 3.3. In both cases the inputs are the same, although EPED1-NN is trained in tokamak DIII-D range of values and the EPED surrogate developed in this study is trained in JET range of values. Output variables are different only because EPED1-NN uses p_{ped} and this study uses $T_{e,ped}$, however both of them are highly related. The architectures are different, EPED1-NN uses a regular feed-forward neural network while in this study the BNN-NCP technique has been applied to obtain uncertainty aware predictions. Regarding performance, EPED1-NN accuracy corresponds with $R^2 = 0.987$ and $R^2 = 0.964$ for p_{ped} and pedestal width Δ, respectively, while, as depicted in figure 13, EPED surrogate model accuracy lies around $R^2 = 0.966$ for $T_{e,ped}$ and $R^2 = 0.916$ for pedestal width Δ. Both models have reasonably high accuracy for the intended outputs, although the one developed in this study is slightly lower.



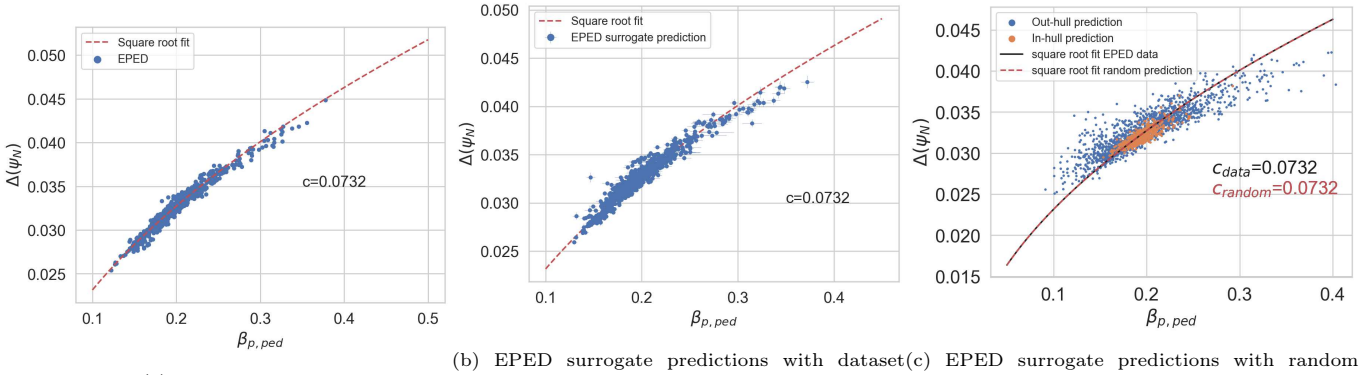
(a) Δ



(b) $T_{e,ped}$

Figure 13: EPED surrogate performance for its two output variables. Train and test data are shown in blue and orange respectively. Epistemic and aleatoric uncertainties are displayed in vertical and horizontal axis respectively.

Overall, in this section it has been shown the adaptability of the method to another plasma pedestal model, obtaining reasonably good performance and comparing it with a similar model.



(a) EPED dataset

(b) EPED surrogate predictions with dataset points

(c) EPED surrogate predictions with random points

Figure 14: Data from various sources represented in pedestal-width (Δ)-$\beta_{p,ped}$ space, with errorbars showing the epistemic uncertainty. Predictions (blue dots) are fitted to show the semi empirical relation $\Delta = c\sqrt{\beta_{p,ped}}$ (square root fit in dashed red). The pedestal width is measured in normalised magnetic flux units ψ_N and $\beta_{p,ped}$ is a dimensionless quantity. In **(c)**, prediction points inside the training convex hull are shown in orange.