Leveraging neural networks for real-time beam emission inference

M. Karacsonyi[1,2], O. Asztalos[1,2], A. Jalalvand[3], G.I. Pokol[1,2], B. Molnar[2], Cs. Nyitrai[1], H. Bugyi[1], J. Illerhaus[4], R. Fischer[4], E. Wolfrum[4], A.H. Nielsen[5] and the ASDEX Upgrade Team[4,6]

[1] *Budapest University of Technology and Economics, Budapest, Hungary*

[2] *HUN-REN Centre for Energy Research, Institute for Atomic Energy Research, Budapest, Hungary*

[3] *Princeton University, Princeton, NJ, USA*

[4] *Max Planck Institute for Plasma Physics, Garching, Germany*

[5] *Technical University of Denmark, Copenhagen, Denmark*

[6] *See Zohm et al (https://doi.org/10.1088/1741-4326/ad249d) for the ASDEX Upgrade Team.*

Beam emission spectroscopy [1] (BES) is an active plasma diagnostic employed for plasma density measurements. In multiple BES-associated applications such as density profile reconstruction and synthetic diagnostics computationally expensive emission inference calculations are utilized to determine the expected emission for a given density profile. The resource intensiveness of such calculations limits the applicability of BES for real-time measurements, while also restricting the speed of synthetic diagnostics.

In this work, we present a possible solution to this problem in the form of a neural network that can predict the expected emission profile on a sub-millisecond timescale.

Building on our previous work [2], we trained an extreme learning machine [3] network and show that not only does it work well on various synthetic datasets generated to simulate different BES beam flavours, but also maintains its high performance on experimental data obtained from the Li-BES system on ASDEX Upgrade [4].

We also show that even in the most extreme edge localized mode burst like cases, the emission predictions provided by the network remain smooth with no additional roughness other than the inherent roughness caused by the numerical resolution of the profiles.

A separate model was developed to provide prediction uncertainty estimates for the emission profiles predicted by the neural network.

Finally, we examined the possibility of integrating the neural network into the IDA framework [5].

Extreme Learning Machines

We worked with extreme learning machines [3],[6] which are dense neural networks (similar to perceptron networks). The only trainable parameters of these networks are the output weights going from the last layer to the output vector (W^{out} in Figure 1). The input weights as well as the weights between the hidden layers

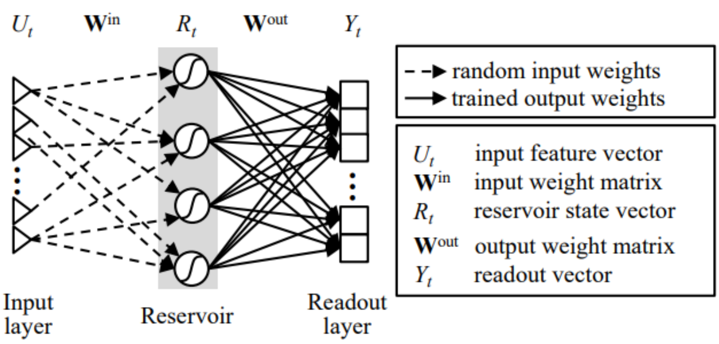


Figure 1 Architecture of a single layer Extreme Learning Machine

(if there is multiple) are randomly initialized and fixed during training.

We used batch intrinsic plasticity [7] to scale the W^{in} input weights of the first hidden layer. Our best network has 2 hidden layers with 200 neurons in each and a dropout layer which was used to provide prediction uncertainty estimates using Monte Carlo dropout [8].

Performance on experimental ASDEX dataset

We worked with an experimental dataset from ASDEX Upgrade, which contained shots with different operational scenarios and types of ELMs. In this section we present the network's performance on shot #42425.

The data was split into training, validation and test sets using a 60/20/20 split. The splitting was done without shuffling, to separate profiles in the different sets temporally.

The reconstructed density profile was used as the neural network's input, and the emission profiles modelled by the numerical BES model in the Integrated Data Analysis tool (IDA) as its output.

Figure 2 shows a prediction example. The plot includes:

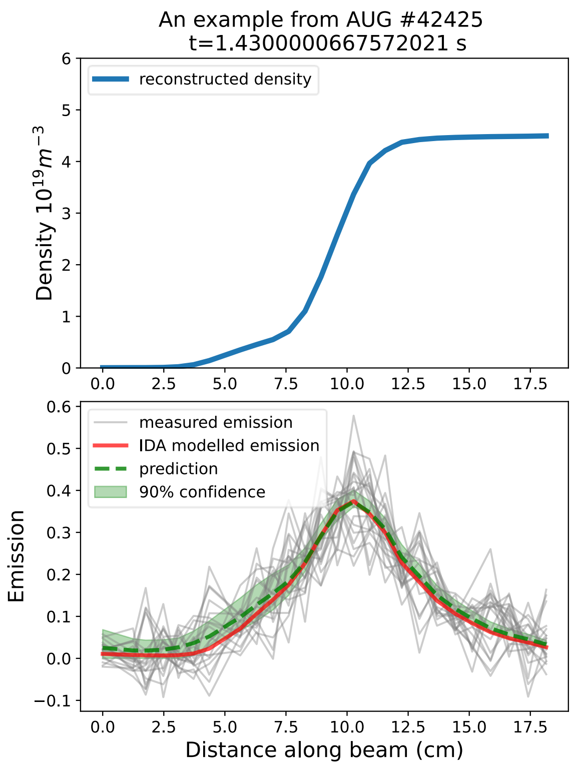


Figure 2 An example from AUG #42425 which shows the emission calculated by IDA together with the corresponding network prediction.

- Reconstructed density profile(blue)
- Measured emission profiles(gray)
- Emission profile modelled by IDA (red)
- Emission predicted by the neural network (green dashed line)
- 90% confidence interval of the neural network's prediction. (green interval)

We can see that the network prediction shows good agreement with the emission profile produced by IDA and the differences between the two are inside the prediction's confidence interval.

The relative cumulative error, defined in Equation 1, was used to assess model performance.

$$Relative\ cumulative\ error = \frac{1}{N}\sum_{i=0}^{N}\left(\frac{\sum_{j=0}^{M}\left|y_{i,j}-p_{i,j}\right|}{\sum_{j=0}^{M}y_{i,j}}\right)_i \quad (1)$$

Where $\boldsymbol{N}$ is the number of profiles, $\boldsymbol{M}$ is the number of points in a profile, $\boldsymbol{y_{i,j}}$ is the jth point in the ith IDA emission profile, and $\boldsymbol{p_{i,j}}$ is the jth point in the ith network prediction.

In Figure 3 we present the prediction uncertainty histogram for the test set of shot #42425.

We can see that model uncertainty is around or under 10% in a majority of cases, although we do have some outliers.

The average uncertainty is 10.76% which is around the uncertainty of atomic data used in BES calculations, therefore satisfactory.

Based on preliminary inference time tests in the standardized GPU-accelerated AUG Deep Learning Inference pipeline (using the methodology described in [9]) we can also say that an **inference time of 33μs per profile is possible** with this network. This can enable real-time BES measurements.

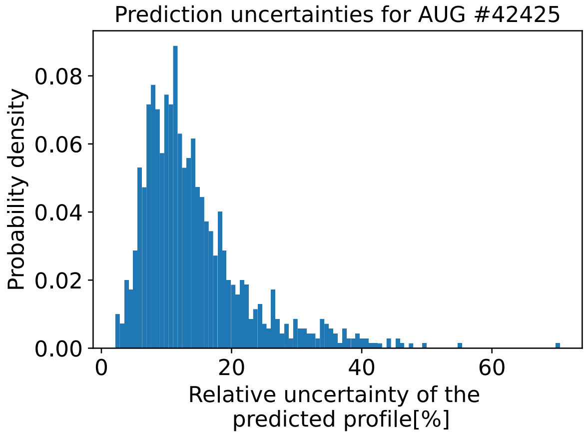


Figure 3 Prediction uncertainty histogram for AUG #42425. We can see that model uncertainty is around or under 10% in a majority of cases.

Smoothness of model predictions

To assess how smooth the predictions given by the network are we defined ”roughness”, which we calculated as follows:

1. We took the profiles one by one.
2. Removed one point from the profile (done for every point except the first and last).
3. Fitted a spline to the remaining points.
4. Calculated the absolute difference between the spline and the removed point.
5. After doing this for all points, we calculate the mean of the differences, this describes the ”roughness” of that profile.
6. We take the maximum of these profile roughness values in the dataset and use it to describe the whole dataset.

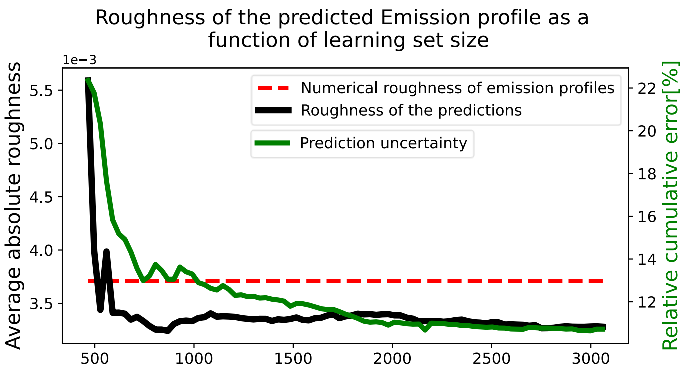


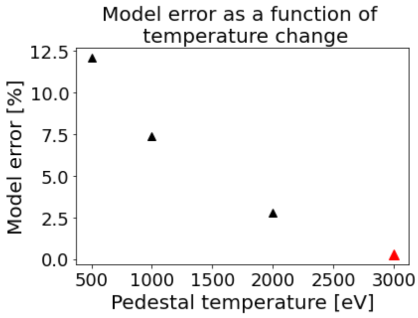
Figure 4 Roughness and prediction uncertainty as a function of learning set size. We can see that both values converge if the learning set is sufficiently large.

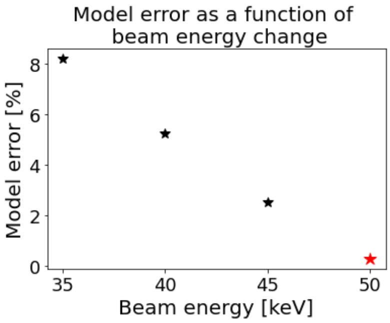
We calculated the roughness values both for the network’s predictions and for the original emission profiles modelled by IDA while changing the size of the learning set (the number of profiles used for training the network). The results, together with the network’s prediction uncertainty are presented in Figure 4.
We can see that if the learning set is big enough the network has a slight smoothening effect, the predictions are smoother than the original IDA emission profiles.
The model error and the roughness both converge roughly at a learning set size of 2000.

Sensitivity to plasma parameters

We generated multiple synthetic datasets with tanh-like density profiles using RENATE-OD [10] and used them to test the neural network’s sensitivity to plasma parameters.

The results are presented in Figures 5-7. The datasets corresponding to the red dots were used to train our network, then we used this network to make predictions for the datasets with different parameters.





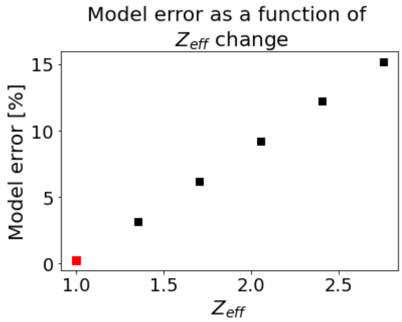


Figure 5 Sensitivity to temperature **Figure 6** Sensitivity to beam energy **Figure 7** Sensitivity to $\boldsymbol{Z_{eff}}$

We can see that as we move away from the original training scope model error increases slowly and deterministically. This leads us to the conclusion that it is enough to have a few different models for different plasma scenarios and switch between them based on the expected parameter values for a given discharge.

Conclusion

In this work we created an Extreme Learning Machine network that can predict the expected emission profile in case of experimental AUG BES data with an uncertainty of roughly 10% which is similar to the uncertainty of atomic data used for BES calculations.

The network can also provide standard deviation and confidence intervals to estimate its own uncertainty.

We have shown that network predictions remain smooth in all cases. Therefore, the network meets the three main requirements for possible IDA integration.

Preliminary inference time tests in the standardized GPU-accelerated Deep Learning Inference pipeline, as integrated into ASDEX Upgrade's discharge control system [11] have shown that the network can provide expected emission predictions under $33\mu s$. This shows that the network not only has uses in speeding up synthetic diagnostics but is light enough to provide real-time BES measurements.

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. G I Pokol and O. Asztalos acknowledge the support of the National Research, Development and InnovationOffice (NKFIH) Grant FK132134.

References

[1] D.M. Thomas *et al*. Fusion Sci. Technol., 53 487-527 (2008)
[2] M. Karacsonyi *et al*. 49th EPS, P1.029 (2023)
[3] G.-B. Huang *et al*. Neuro-computing, 70 489-501 (2006)
[4] M. Willensdorfer *et al*. Plasma Phys. Control. Fusion, 56 025008 (2014)
[5] R. Fischer *et al*. arXiv: 2411.09270 (2024)
[6] P. Steiner et al, Eng. Appl. Artif. Intell. 113., 104964 (2022)
[7] K. Neumann et al. Neuro-computing, 102 23-30 (2013)
[8] Y. Gal and Z. Ghahramani arXiv: 2007.01720 (2016)
[9] J. Illerhaus *et al*. https://doi.org/10.1016/j.fusengdes.2025.115274 [submitted for publication]
[10] O. Asztalos et al, https://github.com/gergopokol/renate-od
[11] W. Treutterer *et al.* Fusion Eng. Des., 89.3 146-154 (2014)