Max-Planck-Institut für Plasmaphysik

Scalable Bayesian inference with model based machine learning

A. Pavone

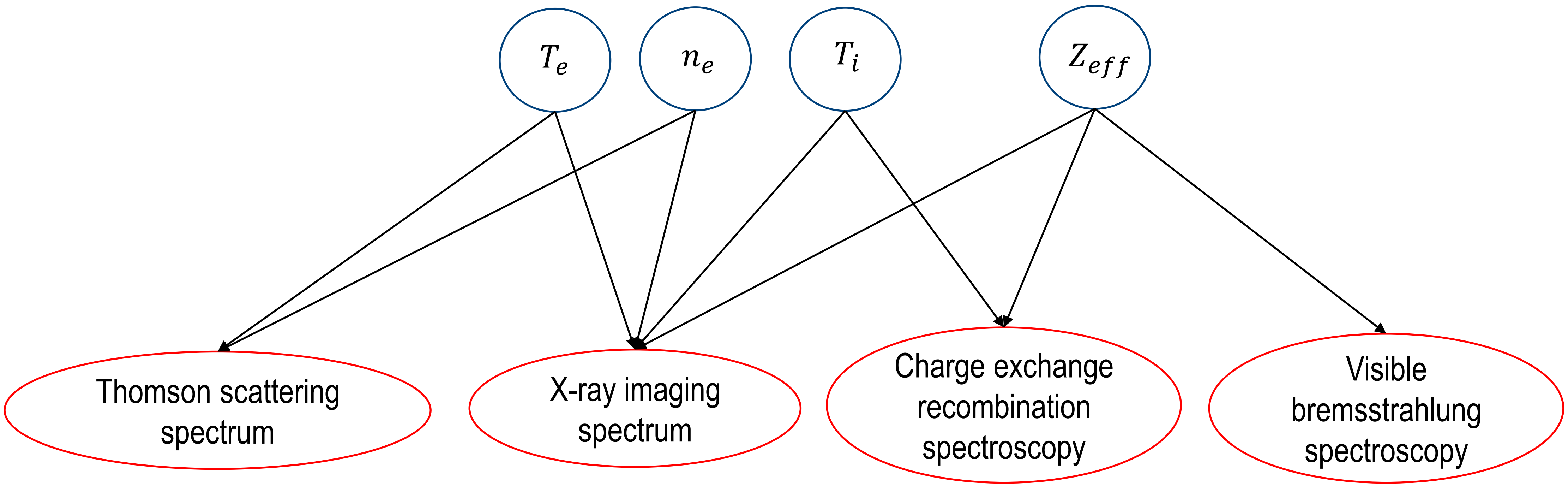
This work has been carried out within the framework of the EUROfusion Consortium and has received funding from the Euratom research and training programme 2014-2018 and 2019-2020 under grant agreement No 633053.
The views and opinions expressed herein do not necessarily reflect those of the European Commission.

Content

1. Bayesian inference in fusion experiments
2. The Minerva framework: unifying Bayesian modeling and inference
3. Machine learning based approximations of Bayesian inference
4. Proofs of concept

One model of the plasma

Few common plasma model parameters inferred from heterogenous plasma diagnostic measurements



Bayesian inference and modeling

- Model *‘m’* of a plasma process can predict observations *‘d’* (data)
- *Probability distributions p:* uncertainties in model assumptions and predictions
- Bayes rule:

$$\underbrace{p(m|d)}_{\text{posterior}} = \frac{\overbrace{p(d|m)}^{\text{likelihood}}\overbrace{p(m)}^{\text{prior}}}{p(d)} = \frac{\overbrace{p(d,m)}^{\text{joint distribution}}}{p(d)} \propto p(d,m)$$

- Joint distribution $p(d,m)$: landscape of all possible assumptions and predictions

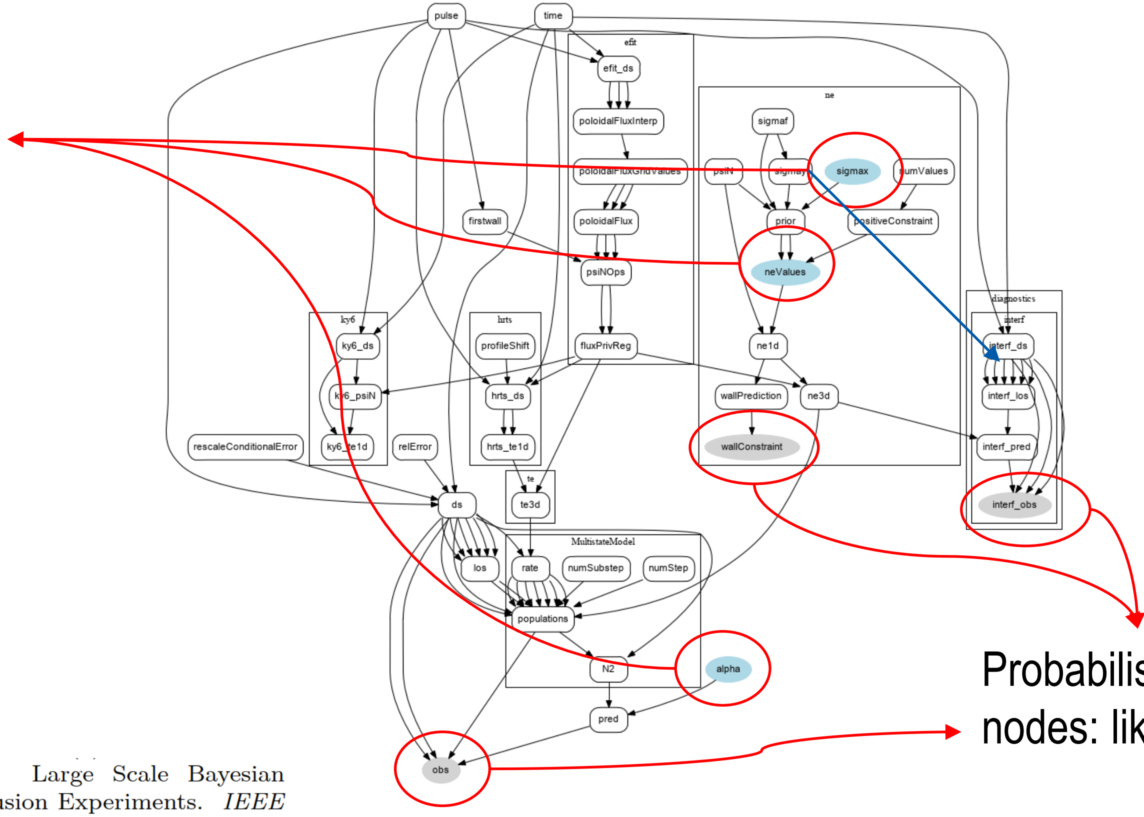
The Minerva Bayesian modeling framework

- Minerva* models are represented through graphical models

Minerva model of two different diagnostics at JET

Probabilistic nodes: priors

White nodes: deterministic nodes



Probabilistic nodes: likelihood

(*) J. Svensson and A. Werner. Large Scale Bayesian Data Analysis for Nuclear Fusion Experiments. *IEEE International Symposium on Intelligent Signal Processing*, pages 1–6, 2007.

Approximated Bayesian Inference

- Inference on posterior probability:
  - Markov Chain Monte Carlo (MCMC) sampling
  - Maximum A Posteriori (MAP) for most likely value
  - Linear methods: linear Gaussian inversion
- Non-linear methods provide robust, complete solutions
- But for complex forward model, calculations can be slow
- Tens of minutes to hours for one single data points (kHz to MHz sampling rates are common)

Machine learning to approximate Bayesian inference

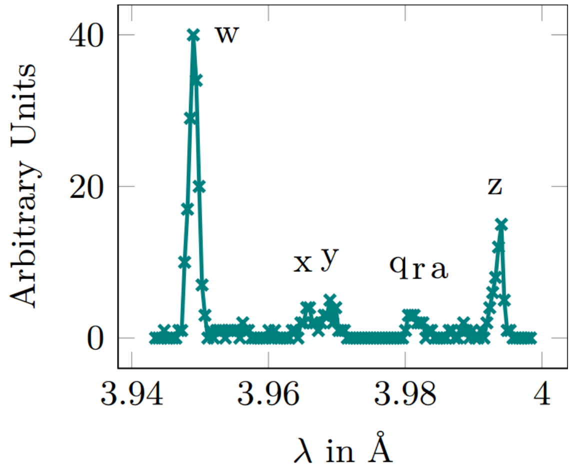
- A Bayesian model is defined by its joint distribution: $p(D, H)$
- The network is trained on samples drawn from the joint: $p(D, H) = p(D|H)p(H)$
- It can be trained to learn different mappings:
  - Forward function $f\colon H \rightarrow D$
  - Inverse function $f^{-1}\colon D \rightarrow H$
  - Mapping to the joint probability value $g\colon (D, H) \rightarrow p(D, H)$
- And then used as a fast approximation when doing inference
- Inference time with deep learning models can be reduced to ≈100 μs
- Fast estimation of plasma parameters for intershot analysis
- Initial guess for conventional Bayesian inference algorithms

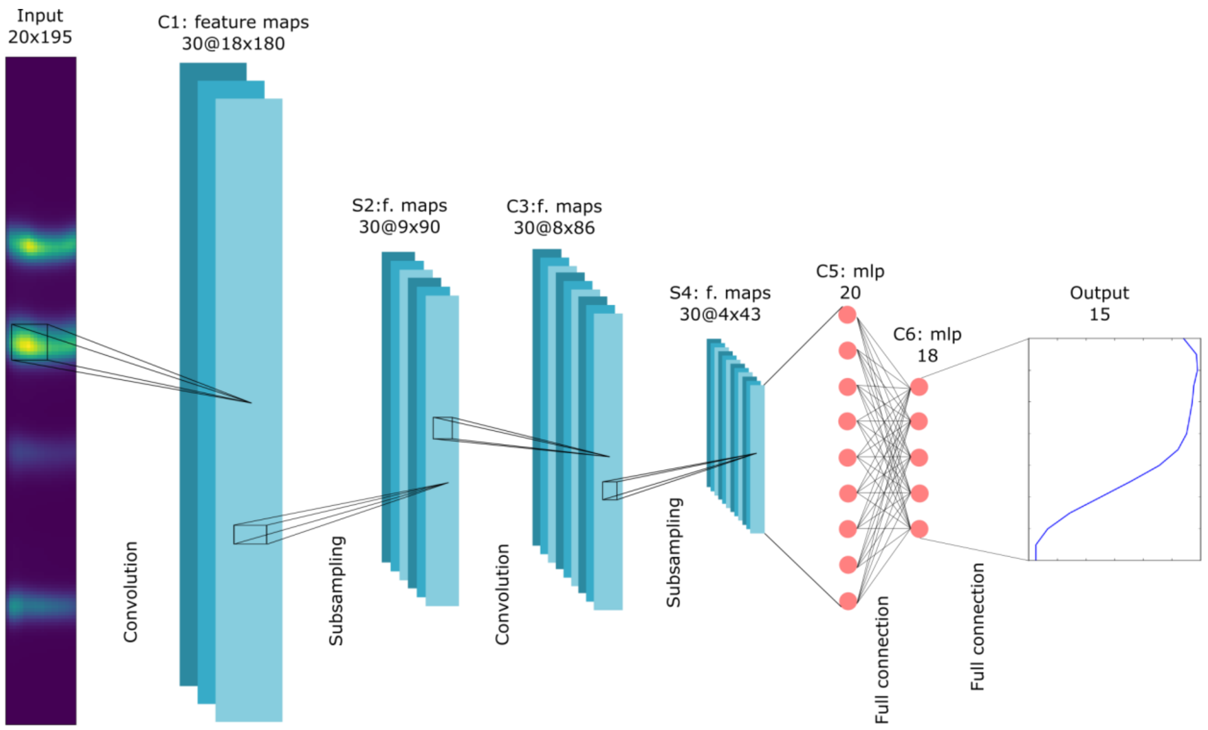
Uncertainties in deep learning: whitening the ‘black box’

- It is not luxury: more accurate and reliable models
- There is no perfect guess: uncertainties make DL more suitable for real world applications
- Bayesian neural network: $p(w|D)$ w: trainable weights, D: training data
- Laplace approximation $p(w|D) \sim N(\mu, \sigma)$
- Variational inference based approach $KL(q(w|D)|p(w|D))$ *MC dropout* (**online estimation**)
- Deep ensembles: accounting for local minima of optimization

Convolutional neural network for Te and Ti profiles at W7-X

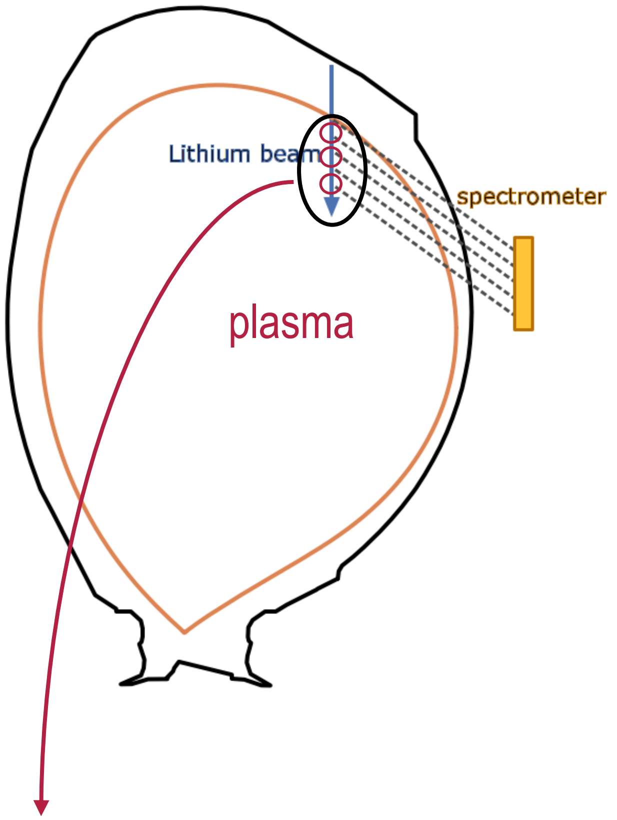
- Data: 2D X-ray imaging crystal spectrometer

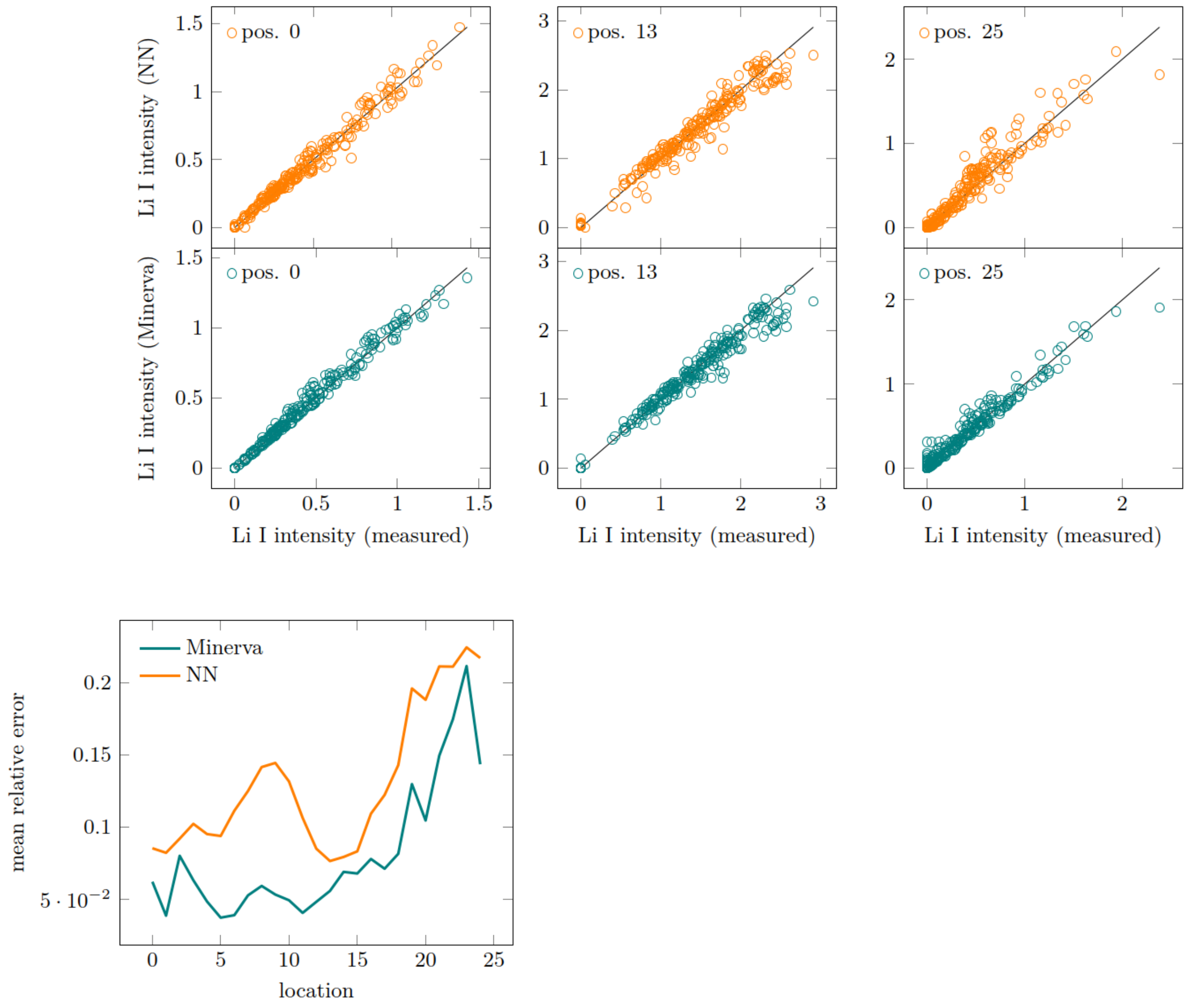




Edge electron density at the JET tokamak

JET plasma cross section





Measuring Li line intensity along 40 cm distance from the top: n_e is inferred at these edge positions

Conclusions

- Bayesian modeling of multiple plasma diagnostics
- Model based machine learning training: samples from the joint probability distribution
- Computationally sustainable and scalable inference (≈100 μs)
- DL uncertainties computable also in real time

- Real time applications possible
- NN based fast approximate inference immediately generalized for any integrated model

- Acceleration of posterior sampling (MCMC/DL-based variational inference)
- Physics constraints into ML model training