Development and validation of synthetic diagnostics and inference models

Simon Pinches[1], Jakob Svensson[2], Udo Höfel[2], Simon McIntosh[1], Paulo Abreu[1], Olivier Hoenen[1], Leonor Tauleigne Roque[1]

[1]ITER Organization, [2]Seed eScience UG

IAEA Technical Meeting on Fusion Data Processing, Validation and Analysis, 11 September 2025
IDM UID: EVJQCG

china eu india japan korea russia usa

Uncertainty propagation, verification and validation

- Aim is to extract the most information from all diagnostic measurements to maximize knowledge acquisition and accelerate ITER Research Plan
- This means having physics models for all the parameters to be inferred and all the diagnostics doing the measuring
- Want to quantify the uncertainties in the inferred quantities given all the uncertainties known about the system
- Developing inference capabilities in a generic way allows to verify and exploit the models on machines with similar diagnostics
- Verifying the correctness of the algorithms and validating them against real-world observations is essential to establishing confidence in the data processing and analysis tools

Introduction

- Preparing to interpret data from ITER plasmas requires the development and validation of models for each of the diagnostic systems installed
- Prioritising the development of synthetic diagnostics for the Start of Research Operations (SRO) has focused initial attention on models for:
  - Magnetic probes, flux loops, saddle coils and Rogowskis
    - Including white or coloured $(1/f^{\beta})$ noise on observations
  - Interferometers (TIP and DIP)
  - Polarimeter (PoPola)
    - Including full Stokes vector calculation, Faraday and Cotton-Mouton effects
  - Thomson scattering (core and edge)
  - Soft X-rays
  - XRCS (core, edge, and survey)
  - Hard X-rays
  - ECE
  - VSRS

Introduction

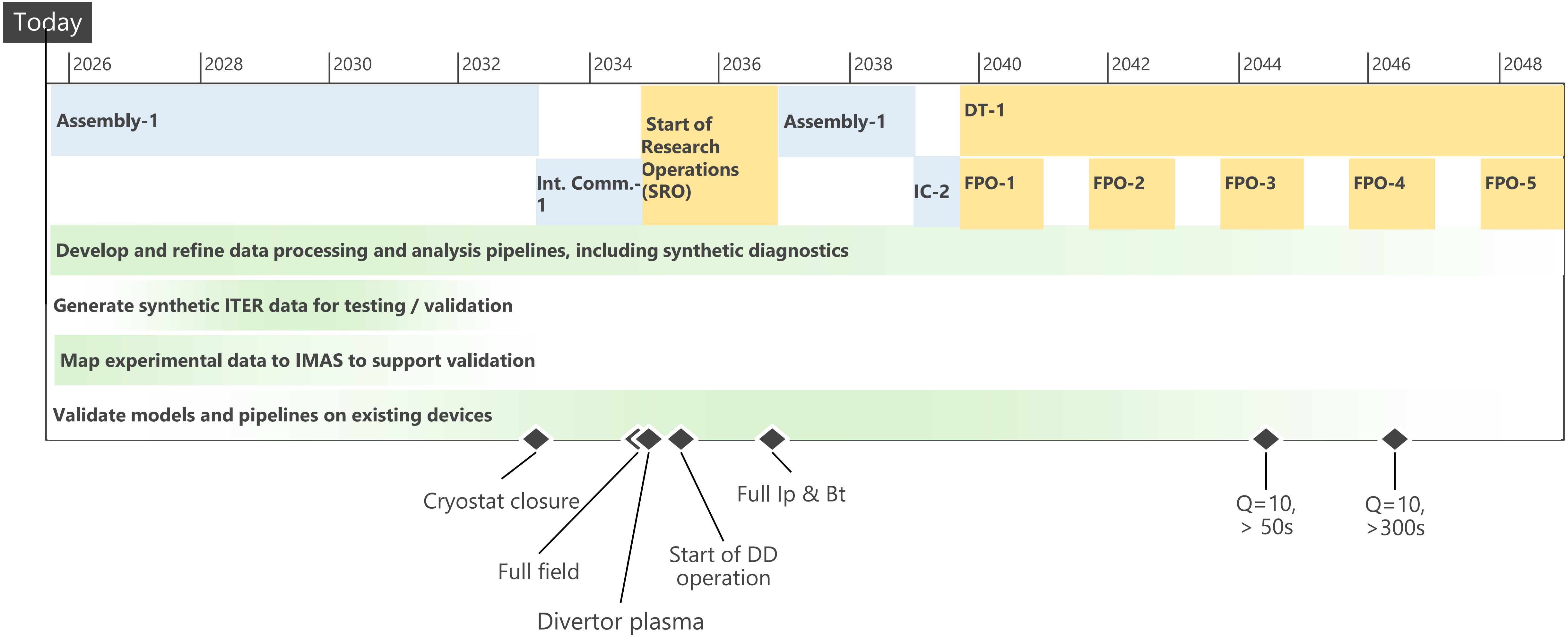
- These models have been developed using the Minerva modelling framework, generalizing the building of models
- Since all these models have been developed following the IMAS paradigm, they are applicable to any device whose configuration (Machine Description data) and data can be mapped to follow the IMAS Data Dictionary
- This allows these models to be validated by comparing their predictions with experimental measurements
- In Minerva, these synthetic diagnostic models can also be easily combined to construct inference models to infer key physics parameters such as the electron density and temperature profiles

Integrated Modelling & Analysis Suite (IMAS)

- IMAS aims to provide a comprehensive framework for integrated modelling of fusion plasmas, including support for benchmarking and validation activities
- It provides a standard way to represent fusion data
- A specific goal is to deliver tools to interpret and analyze measurement data
- Since the researchers who will analyze ITER data will come from the fusion community, it's important to establish a close collaboration to not only prepare and validate the tools that will be used, taking advantage of the wealth of existing knowledge available, but also to train the future generation who will take part in executing the ITER Research Plan

See Olivier Hoenen's talk tomorrow morning for more details on IMAS

Timeline



- IMAS software developed for ITER can be exploited and refined on today's machines

VERNAZZA
AUTOGRU

Two vacuum vessel sector modules in pit

VV sector modules 6 and 7 (KO-DA) now installed in pit

Magnetic sensors already attached to vacuum vessel

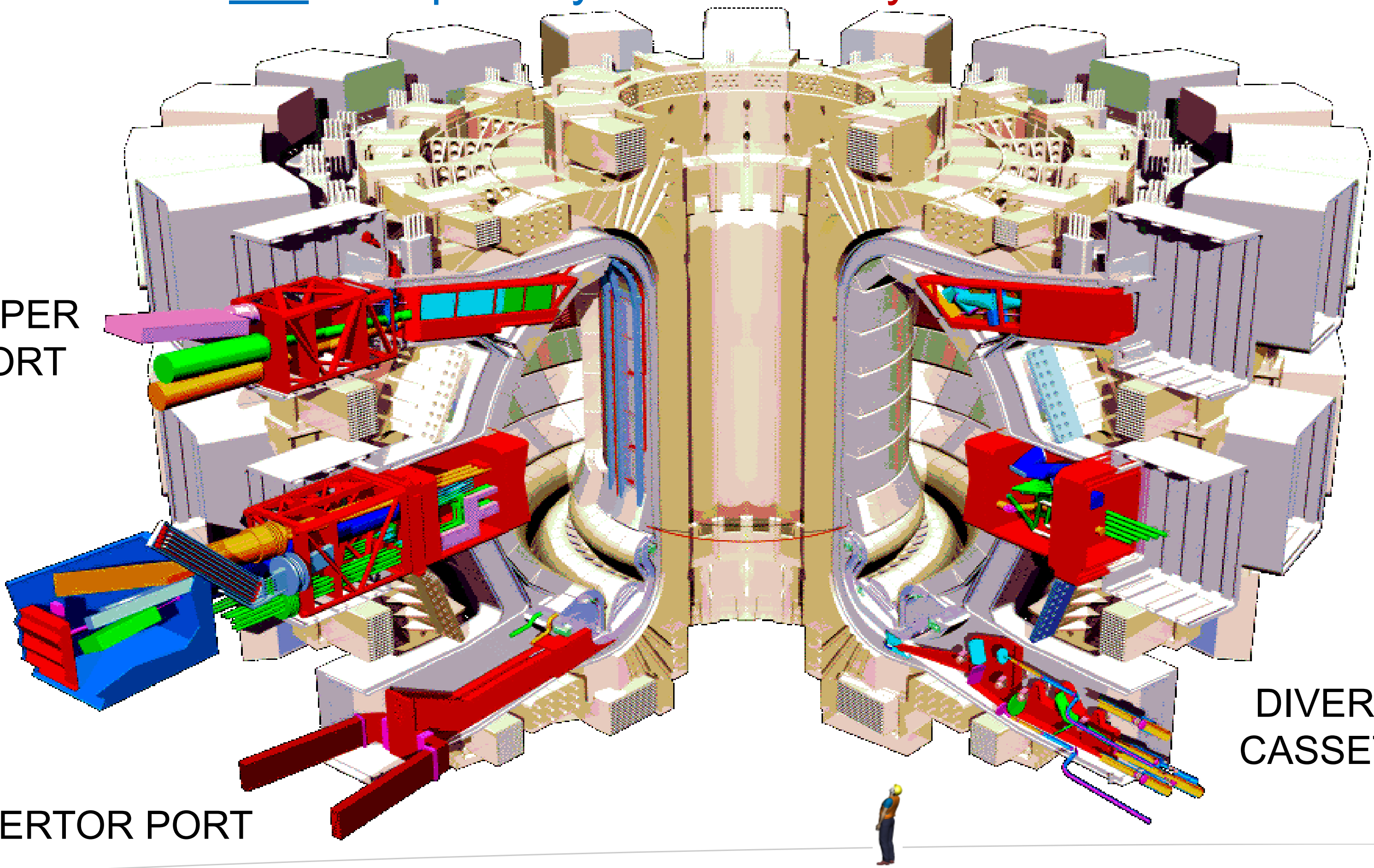
17 June 2025

Challenge in terms of data volume

About 40 major diagnostic systems (= very well diagnosed)

- For **machine protection**, **control** and **physics studies**
- Can reach 2.2 PB of raw data per day → 0.45 EB / year!

ITER will produce so-called "Big Data"



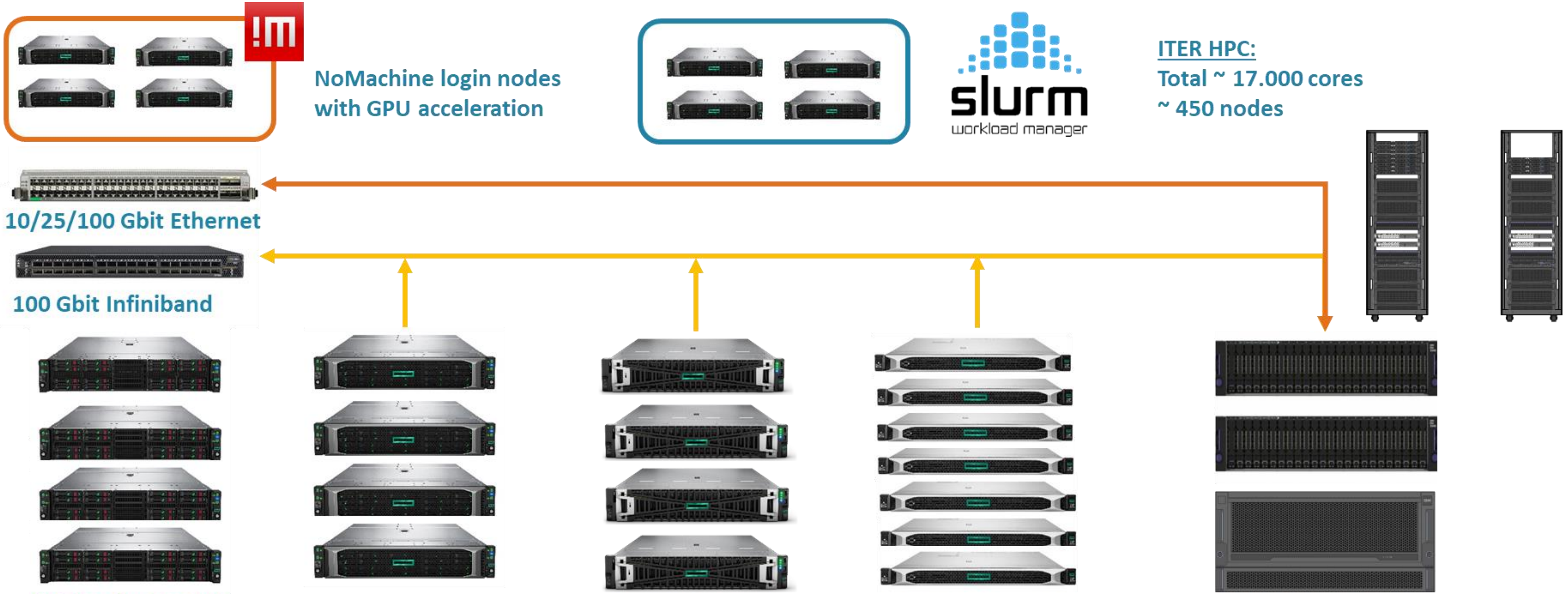
Automated Processing of Data

- Envisaged data processing chains will run concurrently (in-pulse on dedicated hardware) as soon as their input raw data dependencies (in the form of IDSs) are satisfied during a pulse
  - Whilst simple linear chains may have modest computational requirements, more complex probabilistic (Bayesian) inference chains may consume significant resources
  - Since these latter chains are envisaged to be highly parallelizable, computational capabilities should ensure delivery of processed data does not impact inter-shot time or delay next pulse

    → ITER's Scientific Data & Computing Centre (SDCC) sized for this activity
  - Close collaboration with devices which can map and serve raw data and matching Machine Description data → Allows development and validation of workflows now
  - First example/near-term target: magnetics-only equilibrium reconstruction pipeline

Recall Paulo Abreu's talk from yesterday afternoon for more information on in-pulse workflow managers

ITER's Scientific Data & Computing Centre (SDCC)



HPE Apollo Gen9 Cluster
102 Compute nodes
2 X Intel 16 core
128 GB Ram
56 Gbit Infiniband
10/25 Gbit Ethernet

HPE Apollo Gen10 Cluster
156 Compute nodes
2 X Intel 5220 18 core
192 GB Ram
56/100 Gbit Infiniband
10/25 Gbit Ethernet

HPE Apollo Gen11 Cluster
80 Compute nodes
2 X Intel 6348 28 core
512 GB Ram
100 Gbit Infiniband
10/25/100 Gbit Ethernet

HPE DL360 Gen11 Cluster
139 Compute nodes
2 X Intel 6416 18 core
512 GB Ram
100 Gbit Infiniband
10/25/100 Gbit Ethernet

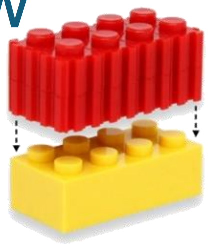
IBM ESS 3500 V7000 Storage Cluster
16 - 60 GB/s Write / 120 GB/s Read
7 PB Total Capacity (+ 4 offsite)
15+ File systems
HPC / SDCC / BACKUP
8 x 100 Gbit IB

Prototype data processing configuration

- Whilst SDCC was sized for ITER's data processing needs and data volumes, the current configuration of the SLURM job scheduling system may not be optimum for data-driven processing pipelines
  - Testing use of Kubernetes as a scalable parallel computing infrastructure for running Minerva Cloud (parallelized over time slices and models)
- Data is read and stored as Interface Data Structures (IDSs) that follow the IMAS Data Dictionary
  - See https://imas-data-dictionary.readthedocs.io for more details

Using diagnostic models

- A CLI tool on SDCC allows to run the models developed in Minera, as well as to develop new models
- Wrapping the tools in Python will enable them to be easily used within other workflows, e.g. to simulate plasma control where actuators respond to diagnostic measurements



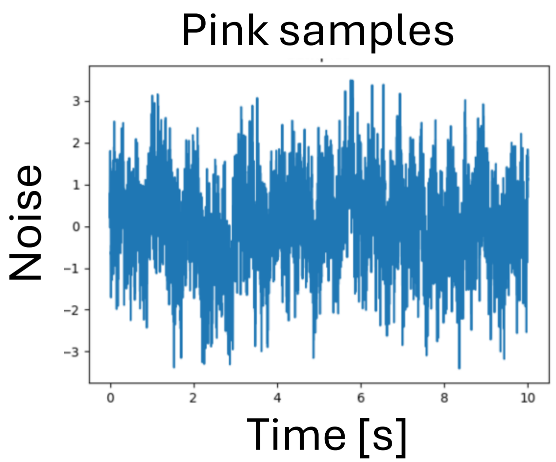
```
[pinches@98dci4-srv-1006 nclass]$ minervac list apps

------------------- Applications -------------------
```

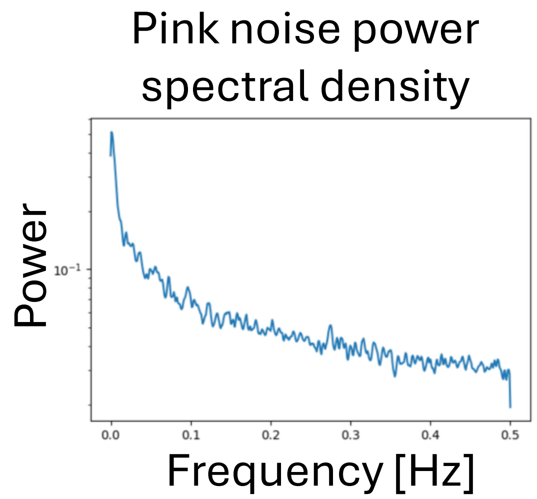
Name	Version	Description
ece-blackbody	1.0.77	ECE blackbody inference
ece-schmuck	1.0.14	ECE Schmuck model inference
ece-spece	1.0.26	ECE diagnostic prediction using SPECE
hxr-pred	1.0.18	Prediction of hard x-ray signals from the runaway electron..
interf-gpt	1.0.92	Interferometer inference using GPT
magn-pred	1.0.62	Magnetic diagnostics prediction.
polarim-pred	1.0.44	Polarimeter and interferometer prediction.
sxr-pred	1.0.24	Prediction of soft x-ray signals
thomson	1.0.22	Imas Thomson scattering app for data generation and infere..
xrcs-pred	1.0.27	Prediction of XRCS spectra
zeff	1.0.17	Imas Zeff app for data generation and inference

```
For details: minervac list apps -v [filter]
```

Digital twin for magnetics system

- Forward models for the full magnetics system allow synthetic data to be created for testing end-to-end magnetics-only reconstruction chain
  - Standard, uncorrelated, white noise contributions can be set for each diagnostic set
    - E.g. `diagnostics.magnetics.bprobes.relErr`, `diagnostics.magnetics.bprobes.absErr`
  - General coloured noise can be imposed on top of white noise
  - Coloured noise is defined through beta parameter that defines the type of coloured noise, $1/f^{\beta}$
    - $\beta = 1$ gives pink noise, $\beta = 2$ gives brown noise, etc. and is set through `coloredNoiseAmplitudeBprobes`, `coloredNoiseAmplitudeFluxloops`, etc.





```
[pinches@98dci4-srv-1006 nclass]$ minervac list apps -v magn-pred

------------------- Applications -------------------
```

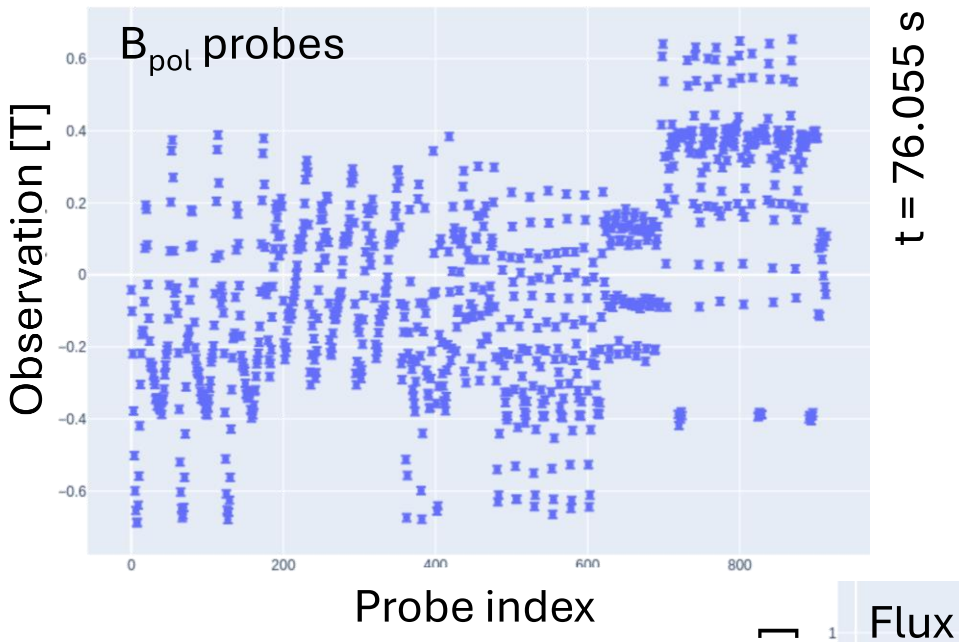
magn-pred (1.0.62) by pjds

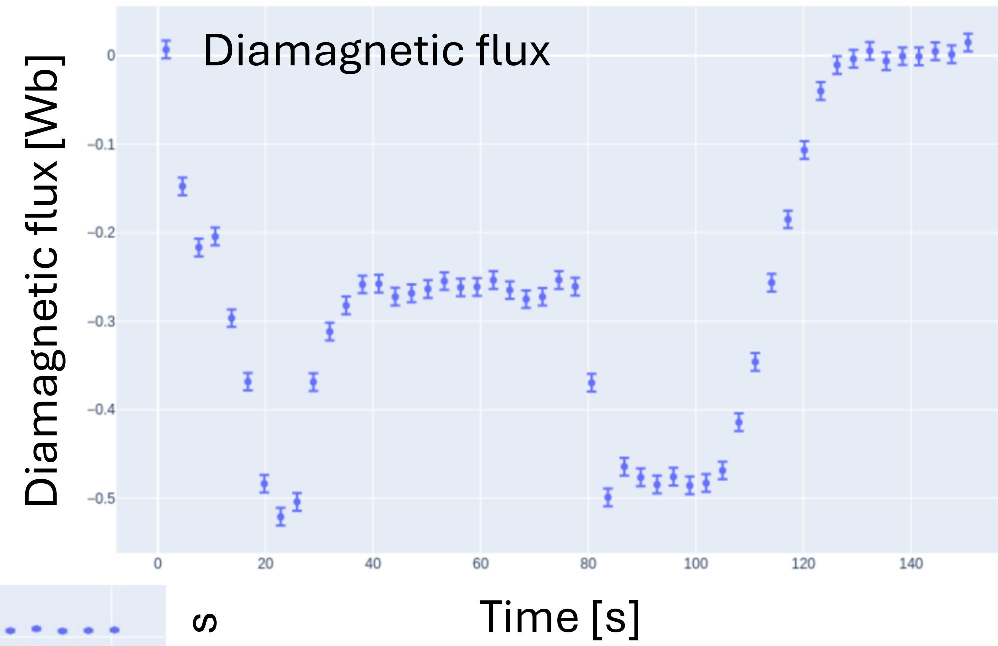
Magnetic diagnostics prediction.

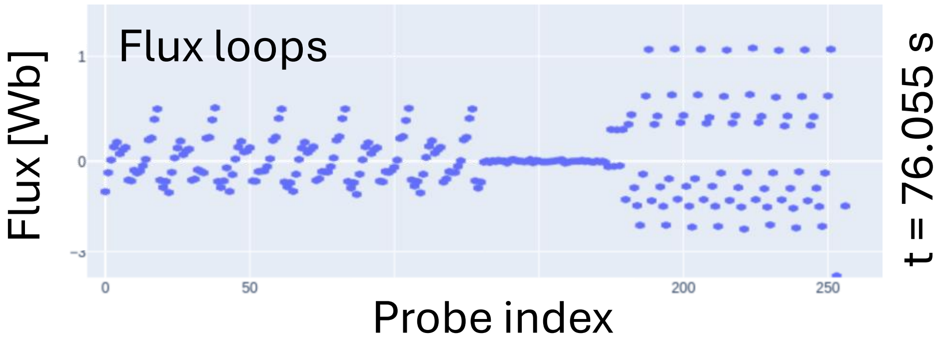
Parameter	Description	Default
mode	This app only runs in synthesize mode	synthesize
num	Number of time slices to analyze. Default (num=0) analyzes all timeslices in the time base (time vector from equilibrium datasource)	0
tstart	Start time	0
tend	End time, 0 for last time in pulse	0
equilibrium.equi_ds.uri	URI of equilibrium IDS	--
diagnostics.magnetics.mag_ds.uri	URI magnetics IDS	imas:hdf5?database=iter_md;version=3;user=public;pulse=150100;run=4
equilibrium.pfActive_ds.uri	URI of pf_active IDS	imas:hdf5?database=ITER_MD;version=3;user=public;pulse=111001;run=102
equilibrium.pfPassive_ds.uri	URI of pf_passive IDS	imas:hdf5?database=iter_md;version=3;user=public;pulse=115004;run=5
coloredNoiseAmplitudeBprobes	Colored noise (1/f^beta) amplitude bprobes	0
coloredNoiseAmplitudeFluxloops	Colored noise (1/f^beta) amplitude fluxloops	0
coloredNoiseAmplitudeRogowskis	Colored noise (1/f^beta) amplitude Rogowski coils	0
coloredNoiseAmplitudeDiamagloops	Colored noise (1/f^beta) amplitude diamagnetic loops	0
beta	Noise color, beta=1 -> pink noise, beta=2 -> rw etc	1
diagnostics.magnetics.bprobes.relErr	White noise relative error bprobes	0.02
diagnostics.magnetics.bprobes.absErr	White noise absolute error bprobes	0.001
diagnostics.magnetics.fluxloops.relErr	White noise relative error fluxloops	0.02
diagnostics.magnetics.fluxloops.absErr	White noise absolute error fluxloops	0.001
diagnostics.magnetics.rogowskis.relErr	White noise relative error rogowskis	0.02
diagnostics.magnetics.rogowskis.absErr	White noise absolute error rogowskis	0.001
diagnostics.magnetics.diamagloops.relErr	White noise relative error diamagnetic loops	0.02
diagnostics.magnetics.diamagloops.absErr	White noise absolute error diamagnetic loops	0.001

Simulated magnetics measurement data from ITER

- Synthetic magnetics for 5 MA, 1.8 T, P_{ECH} = 30 MW calculated with DINA

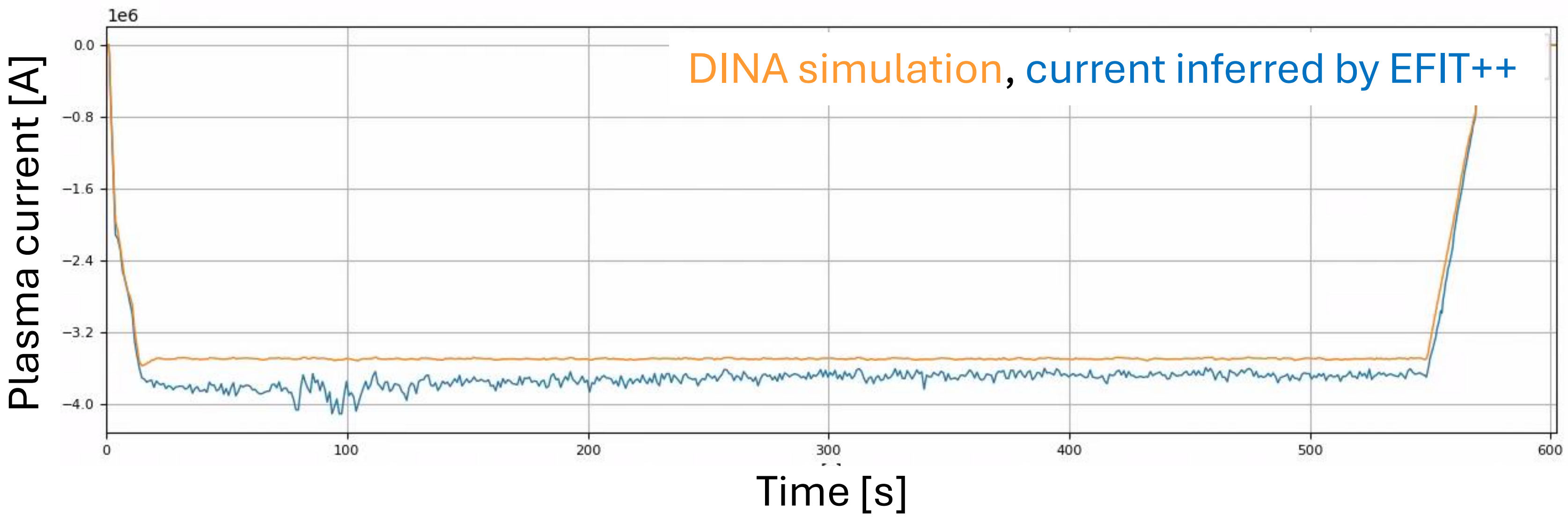


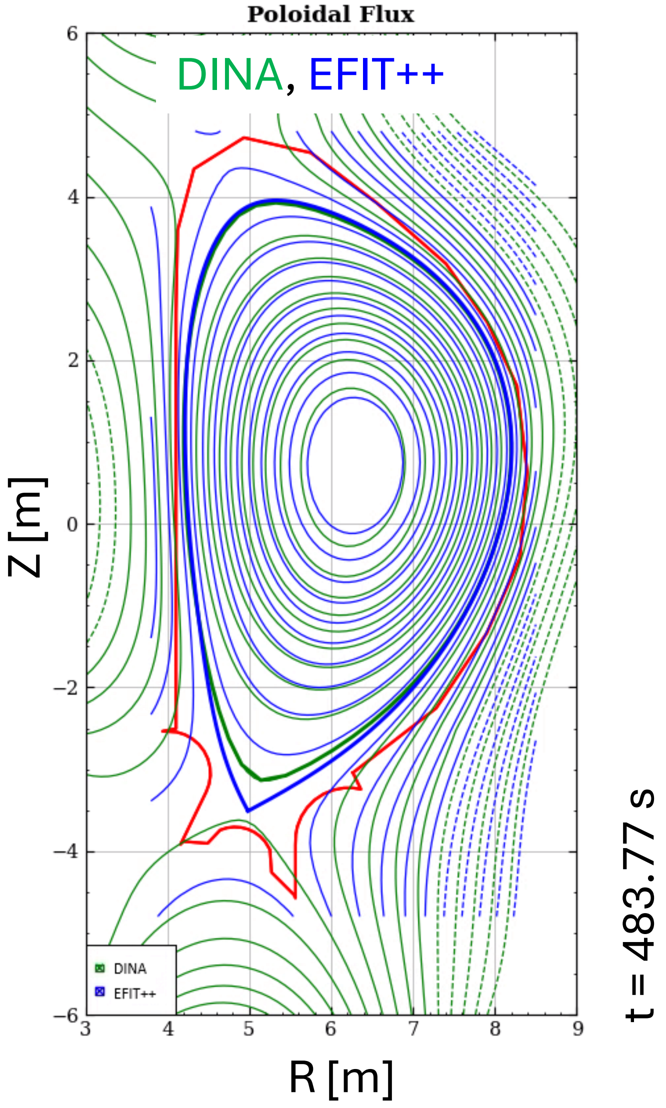




Using synthetic magnetics data to test equilibrium reconstruction

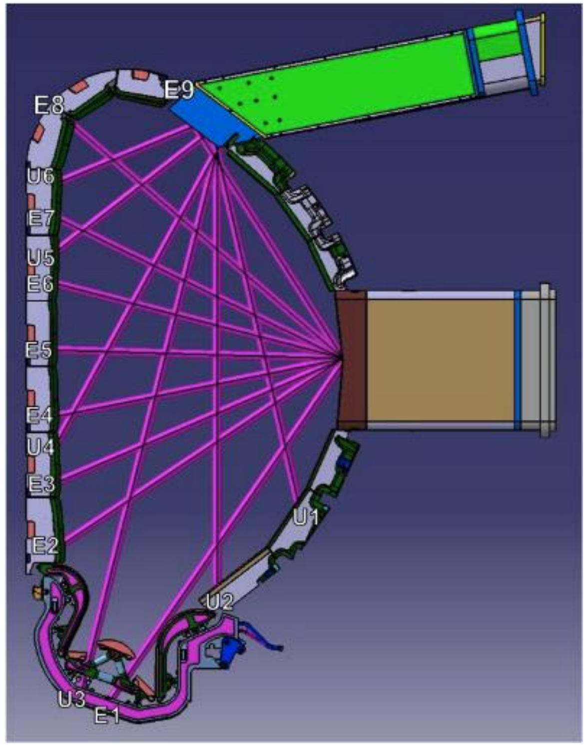
- DINA simulation of 15 MA / 5.3 T scenario
  - →Use to generate synthetic measurement data using forward model for magnetics
  - →Reconstruction equilibrium using EFIT++ and compare with original



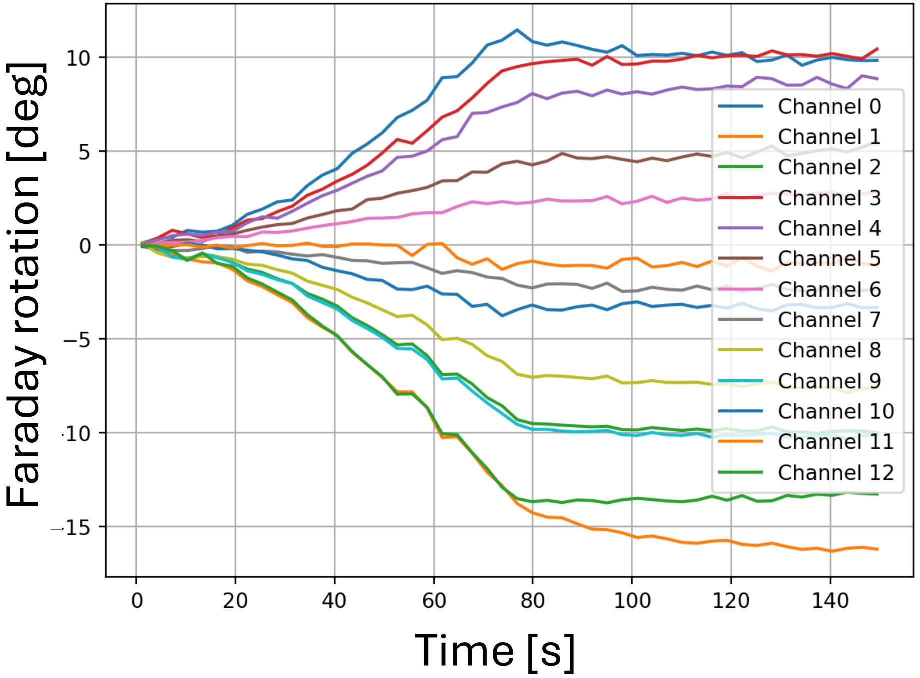


Polarimeters

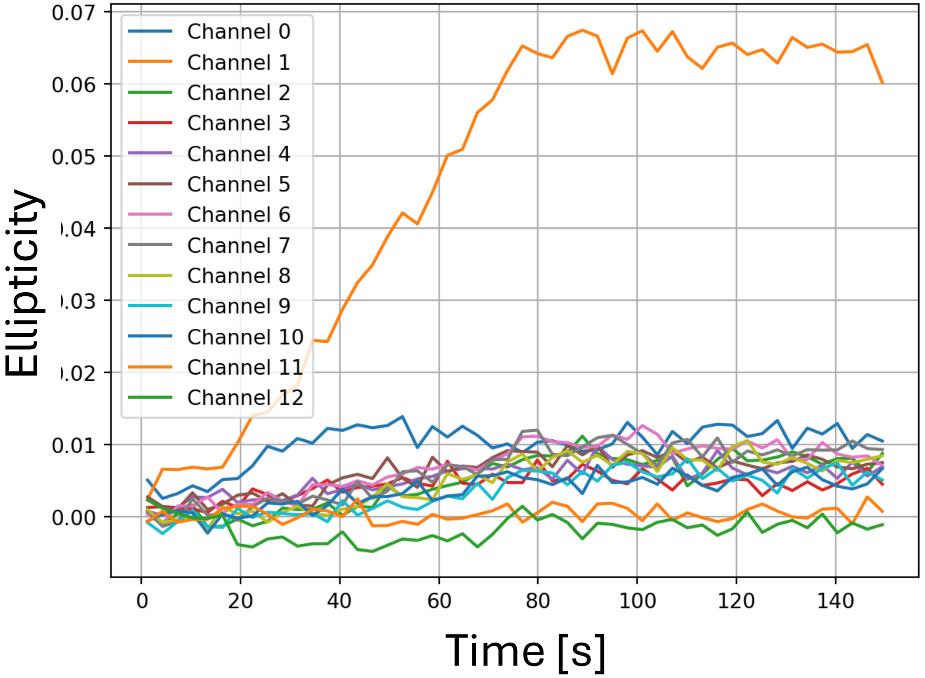
- Polarimeter predictions are now also available using `polarim-pred`:



Faraday rotation PoPola



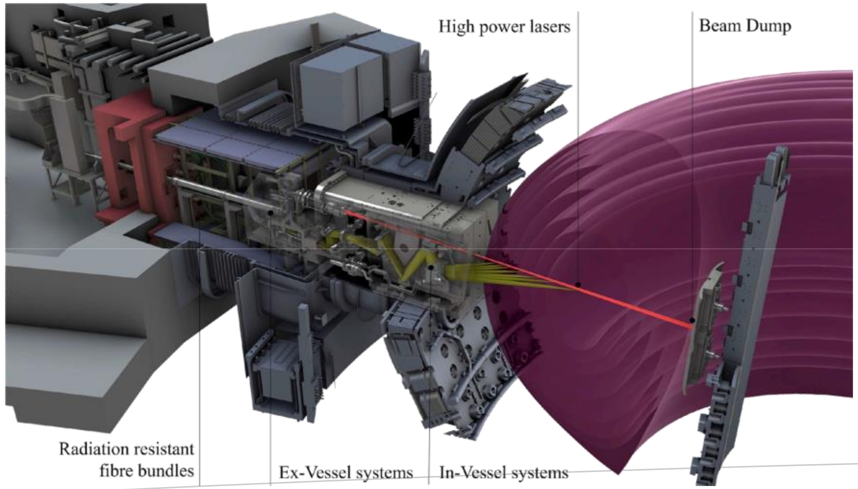
Ellipticity PoPola

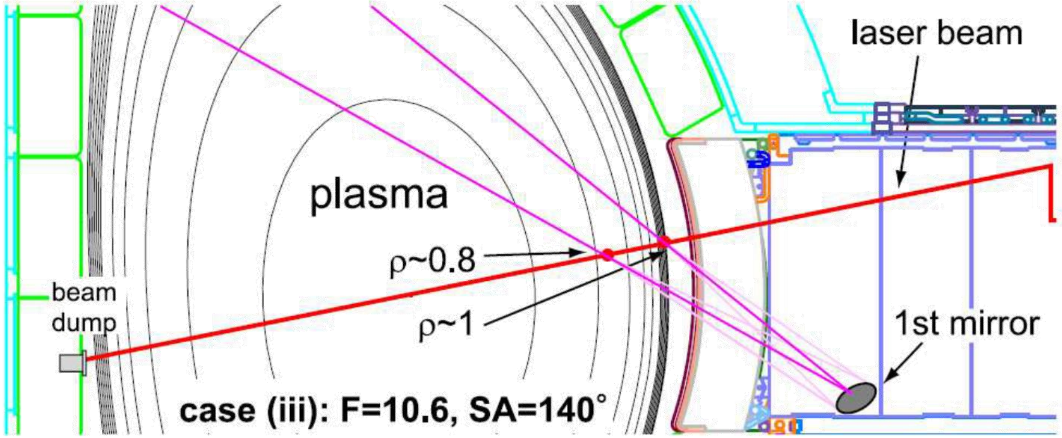


- Relative and absolute errors can be specified for Faraday rotation and ellipticity

Thomson Scattering

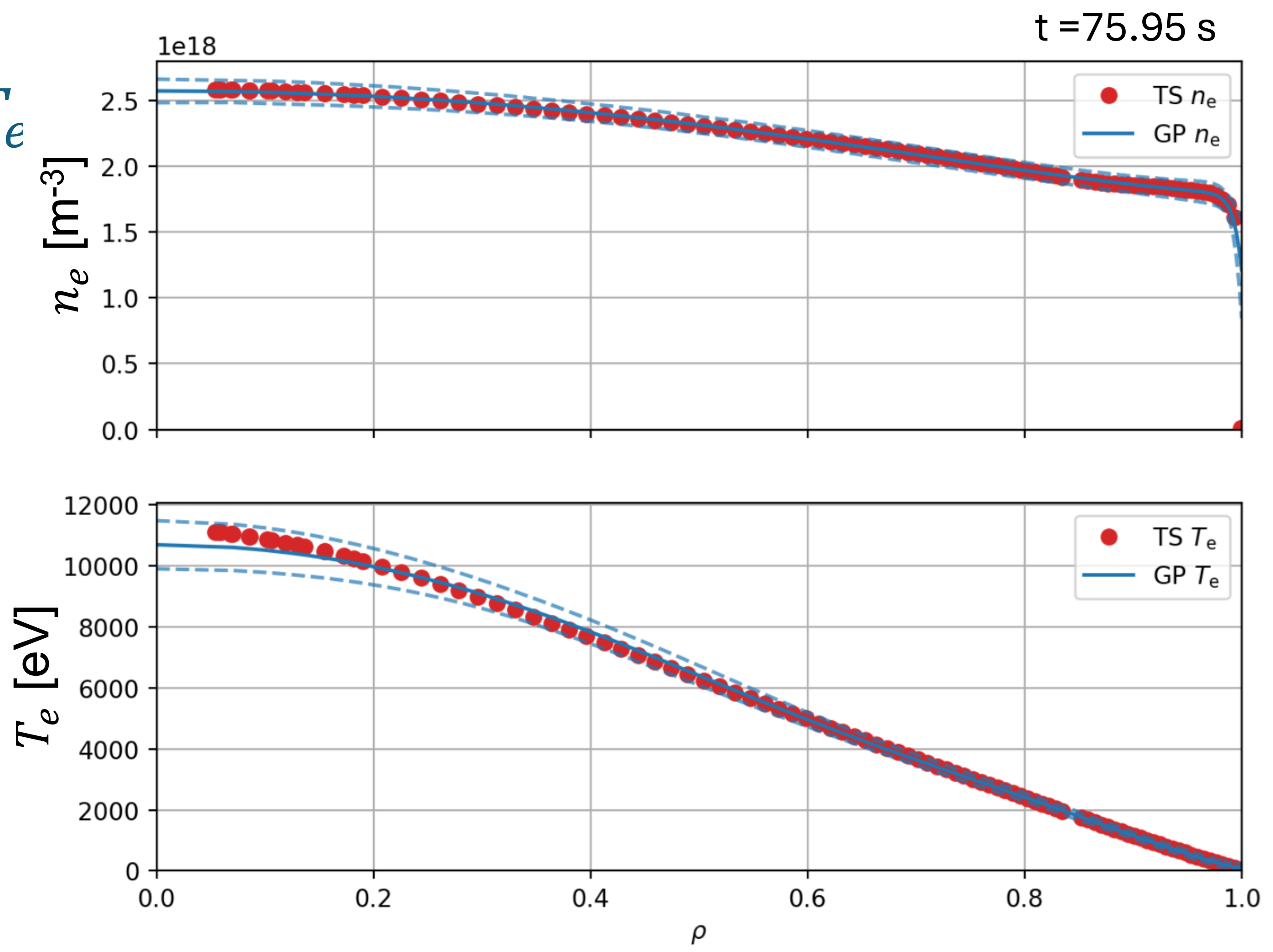
- The TS model can generate observational datasets from given n_e and T_e profiles, equilibrium, and associated Machine Description data (experimental setup of core and edge Thomson scattering systems, e.g., lines of sight, lasers, polychromators) and infer n_e and T_e profiles from spectrally integrated polychromator signals over laser pulses
- Model is applicable to both the core and edge Thomson scattering systems and for both data synthesis and inference





Inference of n_e and T_e profiles from Thomson scattering observations

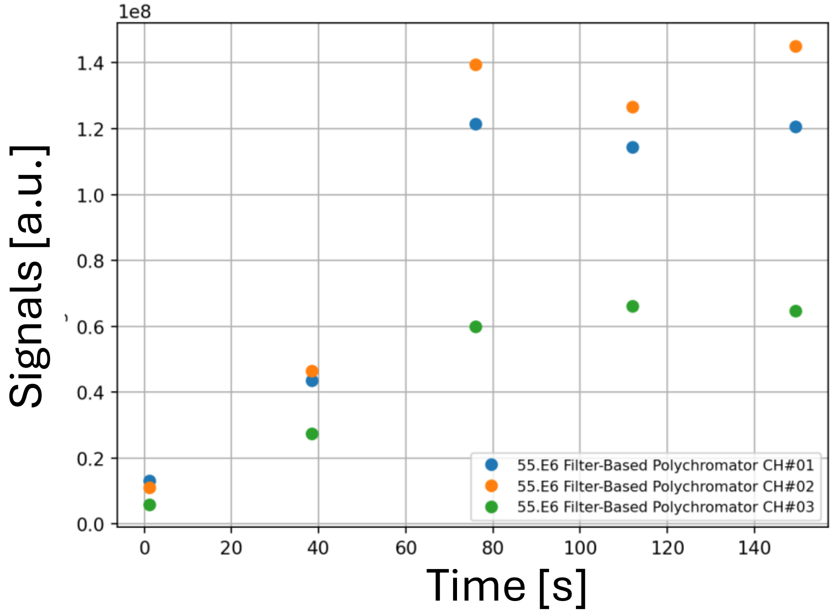
- Individual TS measurements are first used to derive local n_e and T_e values in a fully Bayesian manner, including uncertainties
- Given these local density and temperature values, Gaussian processes are employed to infer the underlying profile functions
  - Hyperparameters governing gradient (smoothness) are inferred by maximising the hyperparameter posterior, following Occam's razor



S. Kwak *et al* 2020 *Nucl. Fusion* **60** 046009

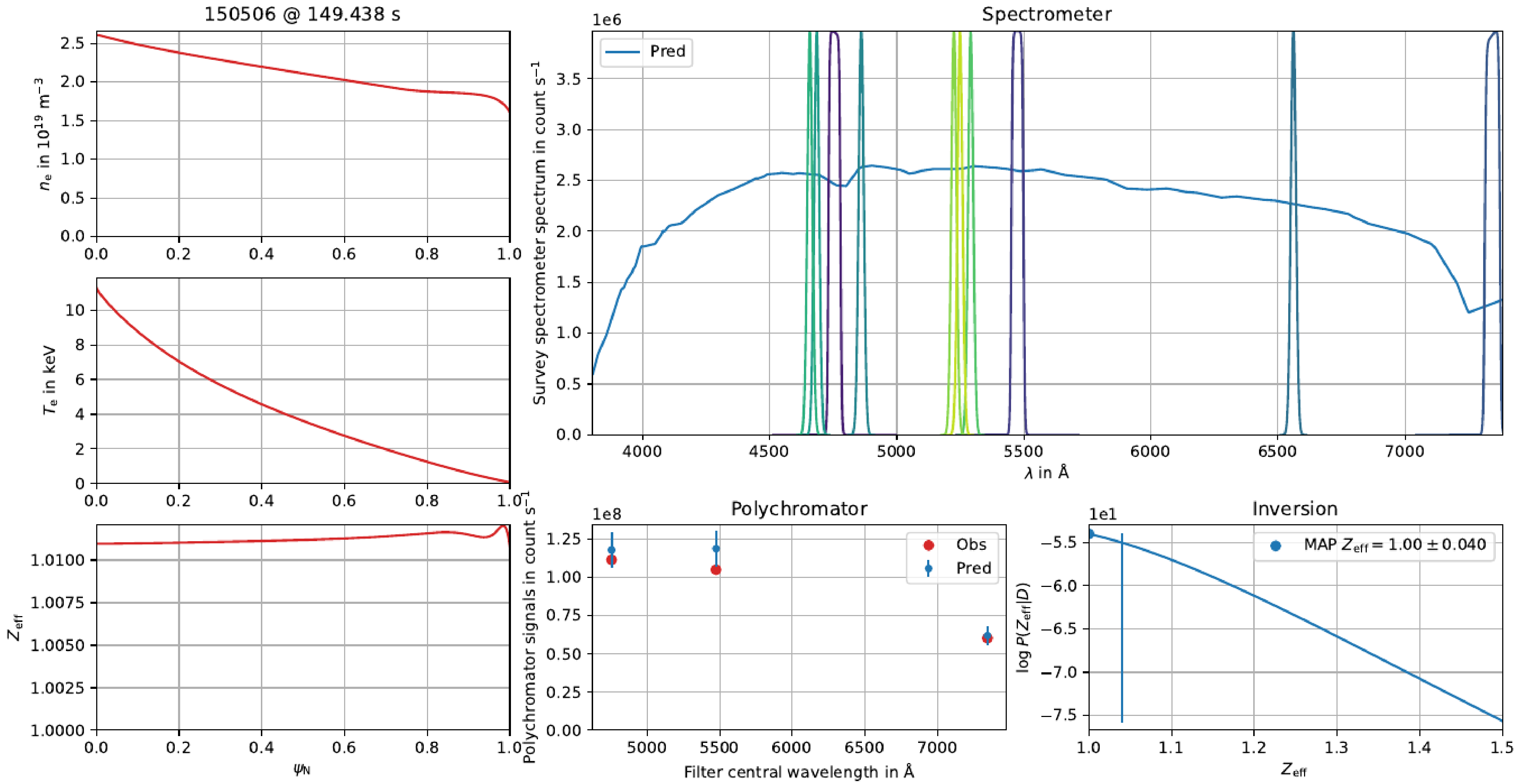
Visible Spectroscopy Reference System (VSRS)

- The VSRS model can generate observational datasets based on specified n_e and T_e profiles, Zeff, equilibrium, and the experimental setup (e.g., lines of sight, filter responses)
  - Observations for filter-based polychromator bremsstrahlung spectral channels are sampled from the specified `core_profiles` and `equilibrium` IDS URIs
- The model can also infer Zeff from spectrally integrated polychromator signals
  - The Zeff values, including uncertainties, are inferred from the VSRS datasets and written to a `core_profile` IDS



S. Kwak *et al*, *Rev. Sci. Instrum*. **92**, 043505 (2021)

Zeff inferred from synthetic VSRS observations



S. Kwak *et al*, *Rev. Sci. Instrum*. **92**, 043505 (2021)

Summary

- Diagnostic models have been developed that can be used both as forward models to generate synthetic measurement data, as well as for the inference of plasma parameters from experimental observations
- The models follow the IMAS paradigm for data and metadata (describing the diagnostic set-up) and are thus applicable to other machines with the same diagnostics and whose data has been mapped to IMAS
  - The ITER Researchers of the future are working on the machines of today
  - Mapping data to IMAS enables testing, validating and using ITER's tools
- In the near term, synthetic measurement data will be generated for all key ITER scenarios from the ITER Research Plan to support the development and testing of more extensive inference capabilities

Getting involved and contributing

- If you're a physicist or software developer, you can contribute with new physics models or help refine existing open-source components
  - https://github.com/iterorganization
- If you have data, then mapping it to the IMAS Data Model (Machine Description data and experimental / simulation data) will help to verify, test and validate components including data analysis and interpretation
- Or you can work with us on-site by applying for an internship, ITER Postdoctoral Fellowship, or to be an ITER Project Associate
  - Advertised on https://www.iter.org
- If you have more experience, then you could come to ITER as a Visiting Researcher or join the ITER Scientist Fellows Network