



Integrated Data Analysis (IDA) for data verification

or

How to spot bad data

D.J. Stieglitz¹, R. Fischer¹

¹IPP Garching



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.

Disclaimer



Plots shown are examples of control-room tools in action and are for illustration purposes only.

Examples include off normal events for illustration purposes

Outline / Index

- My understanding of data verification / preventing misunderstandings
- Very briefly IDA: sorting the vocabulary
- Verification:
 - Basics
 - Single diagnostics
 - Multiple diagnostics
- Mapping issues
- Other topics we could discuss

“Data verification” for the next 20 minutes



Can we trust our data?

- Can we verify that our diagnostics measure what we want and where we want it?
- Can we identify issues like miscalibration, faulty/misbehaving electronics/optics?
- Can we interpret the data properly/accurately (model validation)

IDA basics



IDA basics

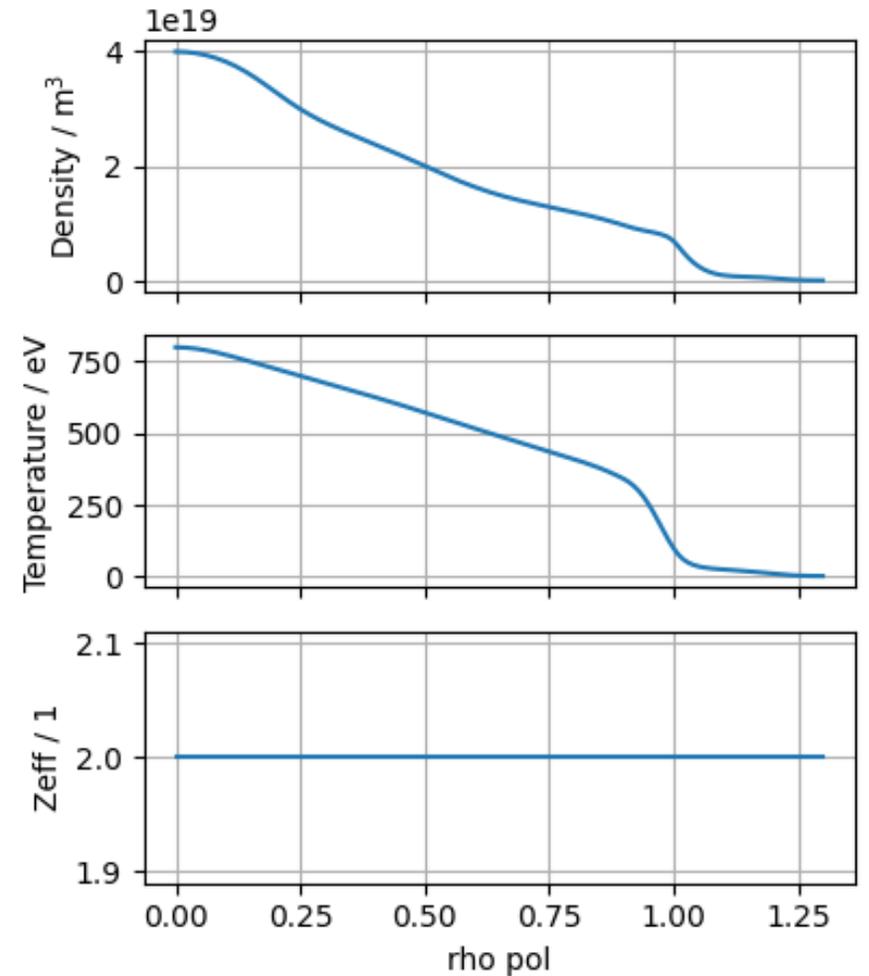


- *Keyword is Forward Modeling → Nothing is „inverted“! → handle non-linearities & outliers*

IDA basics

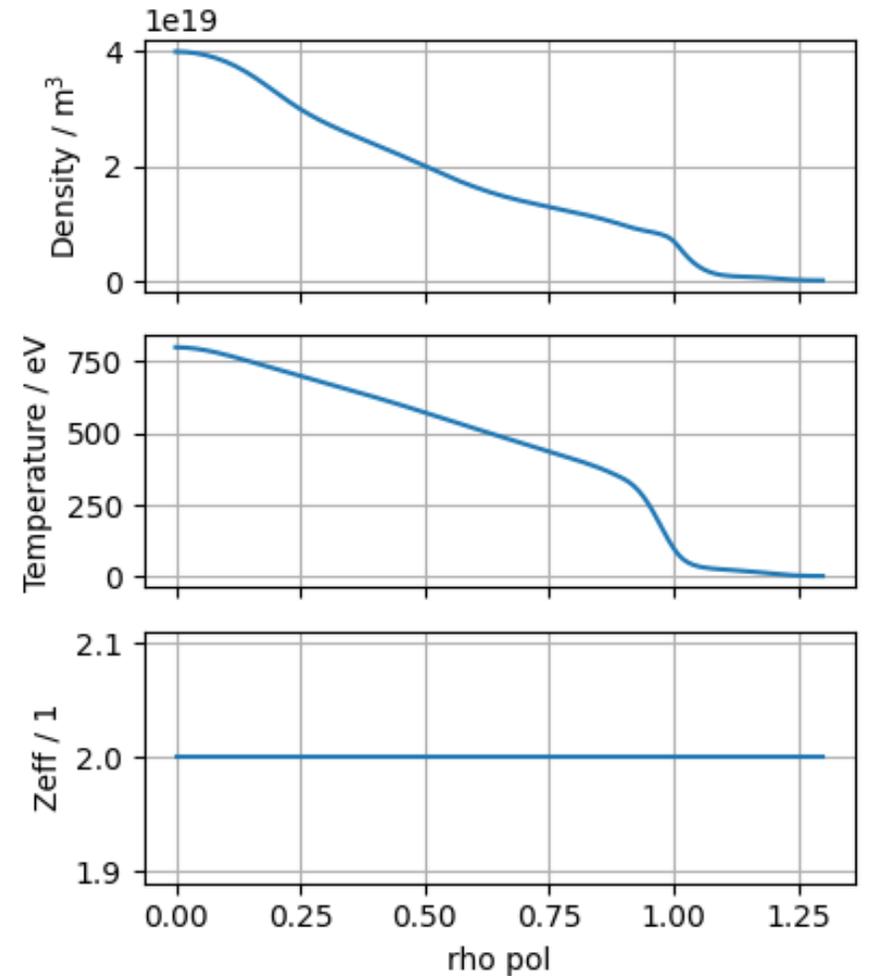


- *Keyword is Forward Modeling* → *Nothing is „inverted“!* → *handle non-linearities & outliers*
- Start with parametric model for the state of the plasma
→ continuous, differentiable profiles (where appropriate)



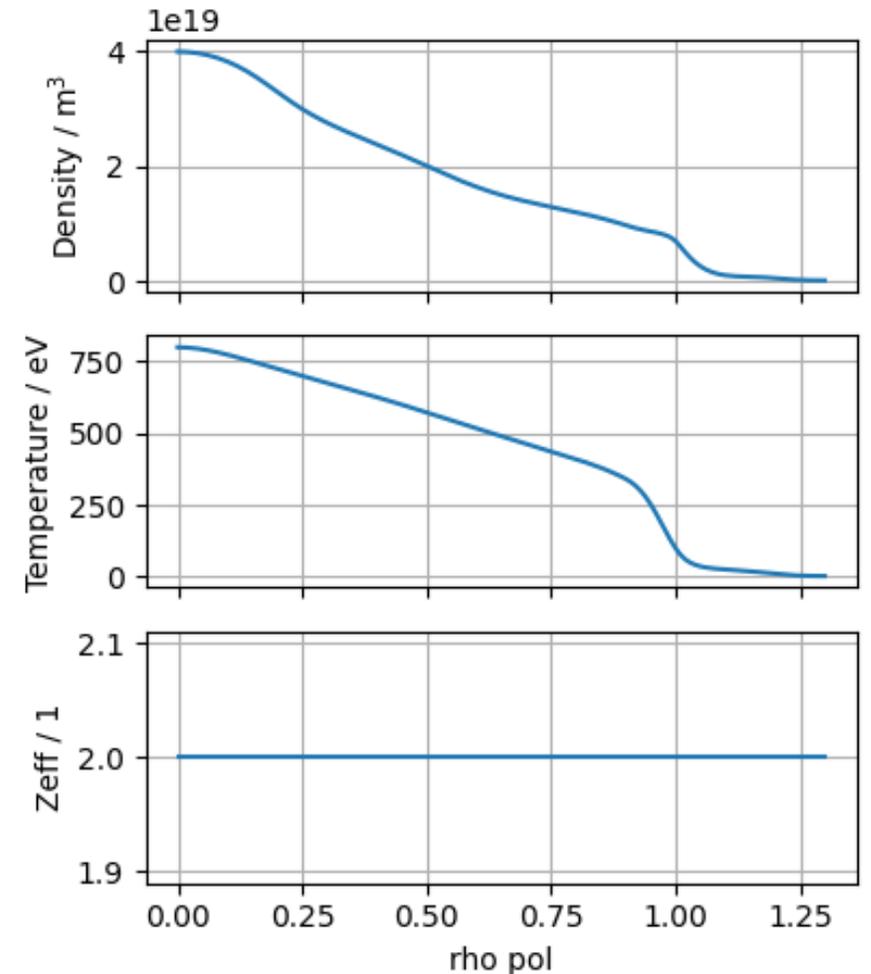
IDA basics

- *Keyword is Forward Modeling* → *Nothing is „inverted“!* → *handle non-linearities & outliers*
- Start with parametric model for the state of the plasma
→ continuous, differentiable profiles (where appropriate)
- Use forward model / synthetic diagnostic to calculate expected signal



IDA basics

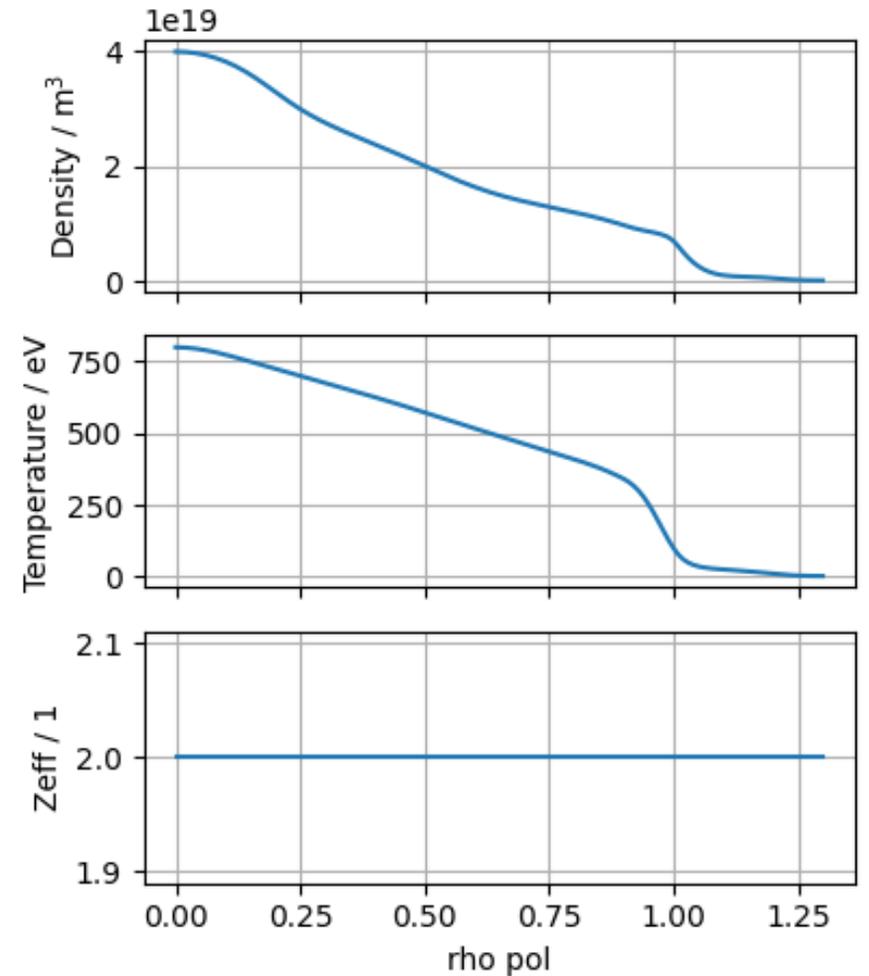
- *Keyword is Forward Modeling* → *Nothing is „inverted“!* → *handle non-linearities & outliers*
- Start with parametric model for the state of the plasma
→ continuous, differentiable profiles (where appropriate)
- Use forward model / synthetic diagnostic to calculate expected signal
- Compare predicted signal to measured signal



IDA basics



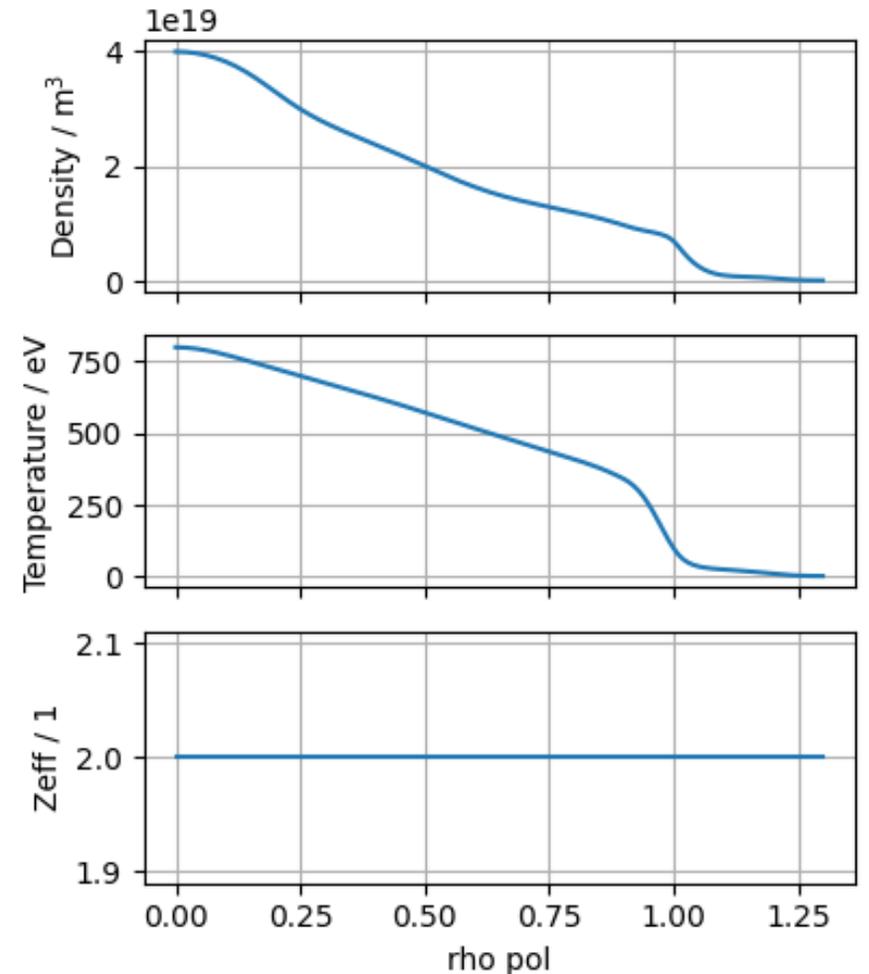
- *Keyword is Forward Modeling* → *Nothing is „inverted“!* → *handle non-linearities & outliers*
- Start with parametric model for the state of the plasma
→ continuous, differentiable profiles (where appropriate)
- Use forward model / synthetic diagnostic to calculate expected signal
- Compare predicted signal to measured signal
- Result is probability for one set of profiles, given data from *all diagnostics*



IDA basics

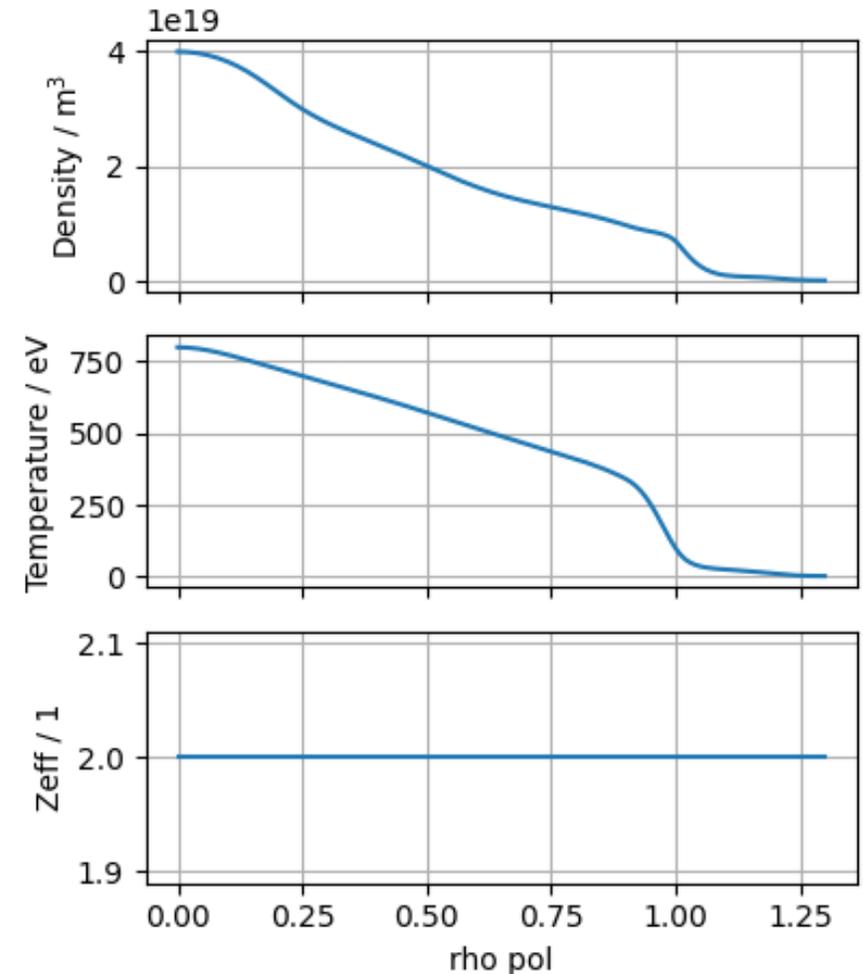


- *Keyword is Forward Modeling* → *Nothing is „inverted“!* → *handle non-linearities & outliers*
- Start with parametric model for the state of the plasma
→ continuous, differentiable profiles (where appropriate)
- Use forward model / synthetic diagnostic to calculate expected signal
- Compare predicted signal to measured signal
- Result is probability for one set of profiles, given data from *all diagnostics*
- **Consistent** description of all data, no iterations



IDA basics

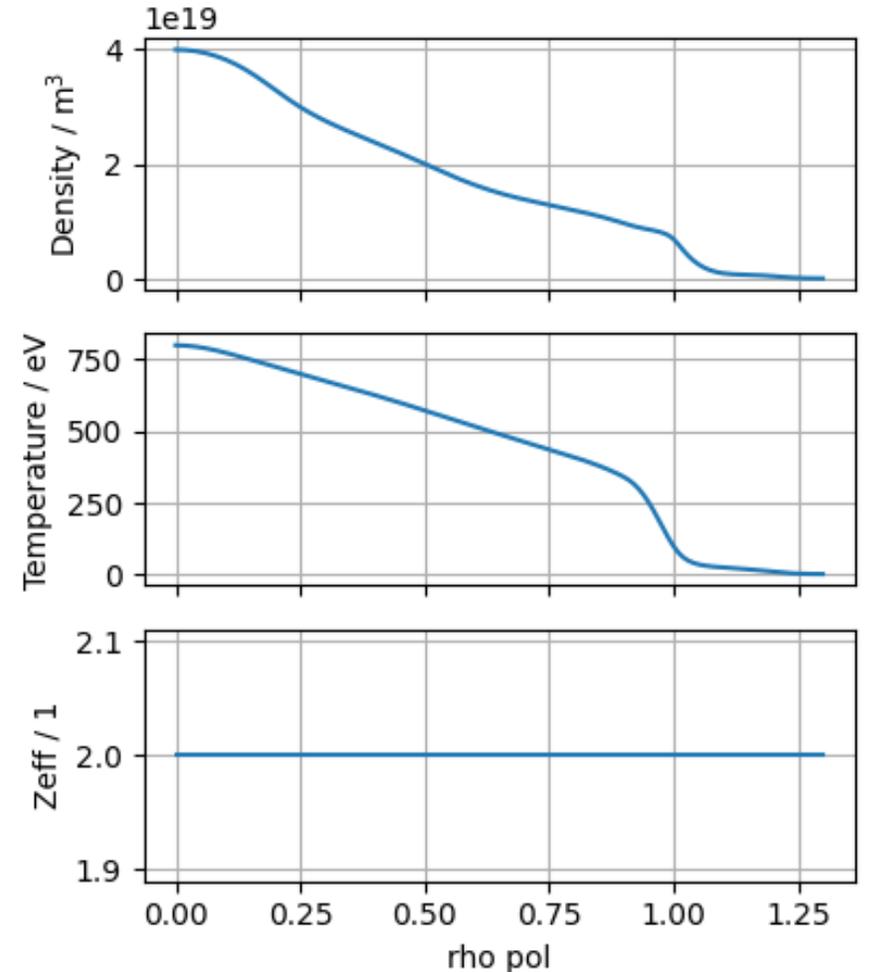
- *Keyword is Forward Modeling* → *Nothing is „inverted“!* → *handle non-linearities & outliers*
- Start with parametric model for the state of the plasma
→ continuous, differentiable profiles (where appropriate)
- Use forward model / synthetic diagnostic to calculate expected signal
- Compare predicted signal to measured signal
- Result is probability for one set of profiles, given data from *all diagnostics*
- **Consistent** description of all data, no iterations
- Reveals inconsistencies and improper calibrations



IDA basics



- *Keyword is Forward Modeling* → *Nothing is „inverted“!* → *handle non-linearities & outliers*
- Start with parametric model for the state of the plasma
→ continuous, differentiable profiles (where appropriate)
- Use forward model / synthetic diagnostic to calculate expected signal
- Compare predicted signal to measured signal
- Result is probability for one set of profiles, given data from *all diagnostics*
- **Consistent** description of all data, no iterations
- Reveals inconsistencies and improper calibrations
- Express uncertainty in system parameters
→ error bars / confidence bounds include this!



“Validated data”



- IDA does not rely on the best input data, in the sense of curated data
- If the diagnostic works and is understood, just use the data and see if the results are good

“Validated data”

- IDA does not rely on the best input data, in the sense of curated data
- If the diagnostic works and is understood, just use the data and see if the results are good
- Not every IDA profile has to be published!
 - IDA can run automatically
 - users can check agreement of diagnostics
 - diagnosticians get feedback on calibration etc.
e.g. catch drifts in calibration early, catch EM perturbations early, ...

“Validated data”

- IDA does not rely on the best input data, in the sense of curated data
- If the diagnostic works and is understood, just use the data and see if the results are good
- Not every IDA profile has to be published!
 - IDA can run automatically
 - users can check agreement of diagnostics
 - diagnosticians get feedback on calibration etc.
e.g. catch drifts in calibration early, catch EM perturbations early, ...
- Missing data do not matter, e.g. missing channels or views
 - Less data = more uncertainties

“Validated data”

- IDA does not rely on the best input data, in the sense of curated data
- If the diagnostic works and is understood, just use the data and see if the results are good
- Not every IDA profile has to be published!
 - IDA can run automatically
 - users can check agreement of diagnostics
 - diagnosticians get feedback on calibration etc.
e.g. catch drifts in calibration early, catch EM perturbations early, ...
- Missing data do not matter, e.g. missing channels or views
 - Less data = more uncertainties
- No need to restrict IDA to “validated data”, **IDA should be part of the validation!**

Long story short: look at the residues

$$(d - m) / \sigma$$



Long story short: look at the residues

$$(d - m) / \sigma$$



- How do you recognise discrepancies → not all data can be described jointly
- For single diagnostic see if all channels can be brought in agreement (with reasonable model for profiles etc)
- For several diagnostics use single „plasma state“ to describe them all
- At AUG this is checked manually (scroll through residua plots for each shot)

Long story short: look at the residues

$$(d - m)/\sigma$$



- How do you recognise discrepancies → not all data can be described jointly
- For single diagnostic see if all channels can be brought in agreement (with reasonable model for profiles etc)
- For several diagnostics use single „plasma state“ to describe them all
- At AUG this is checked manually (scroll through residua plots for each shot)

- Automatisations possible, 1st order:
$$\frac{1}{N} \sum_{i=1}^N ((d_i - m_i)/\sigma_i)^2 \approx 1$$

Long story short: look at the residues

$$(d - m)/\sigma$$



- How do you recognise discrepancies → not all data can be described jointly
- For single diagnostic see if all channels can be brought in agreement (with reasonable model for profiles etc)
- For several diagnostics use single „plasma state“ to describe them all
- At AUG this is checked manually (scroll through residua plots for each shot)
- Automatisations possible, 1st order:
$$\frac{1}{N} \sum_{i=1}^N ((d_i - m_i)/\sigma_i)^2 \approx 1$$
- This is a direct measure for the **likelihood** of the data wrt the model (profiles)

Long story short: look at the residues

$$(d - m) / \sigma$$



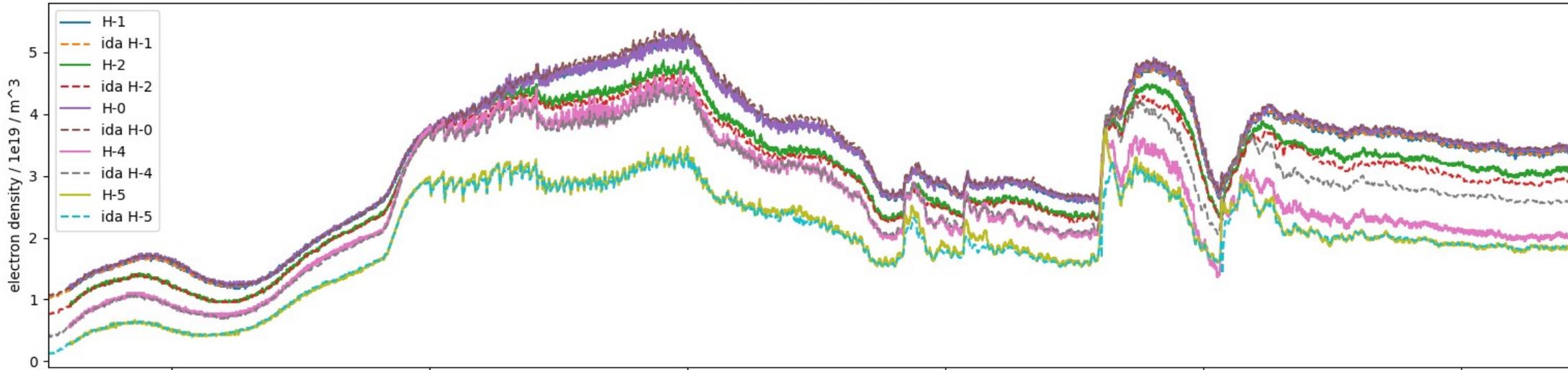
- How do you recognise discrepancies → not all data can be described jointly
- For single diagnostic see if all channels can be brought in agreement (with reasonable model for profiles etc)
- For several diagnostics use single „plasma state“ to describe them all
- At AUG this is checked manually (scroll through residua plots for each shot)
- Automatisations possible, 1st order:
$$\frac{1}{N} \sum_{i=1}^N ((d_i - m_i) / \sigma_i)^2 \approx 1$$
- This is a direct measure for the **likelihood** of the data wrt the model (profiles)
- Bayes Theorem: posterior = **likelihood** * prior / evidence

AUG Example for Interferometry (fringe jump)



38934:AUGD:DCK:1

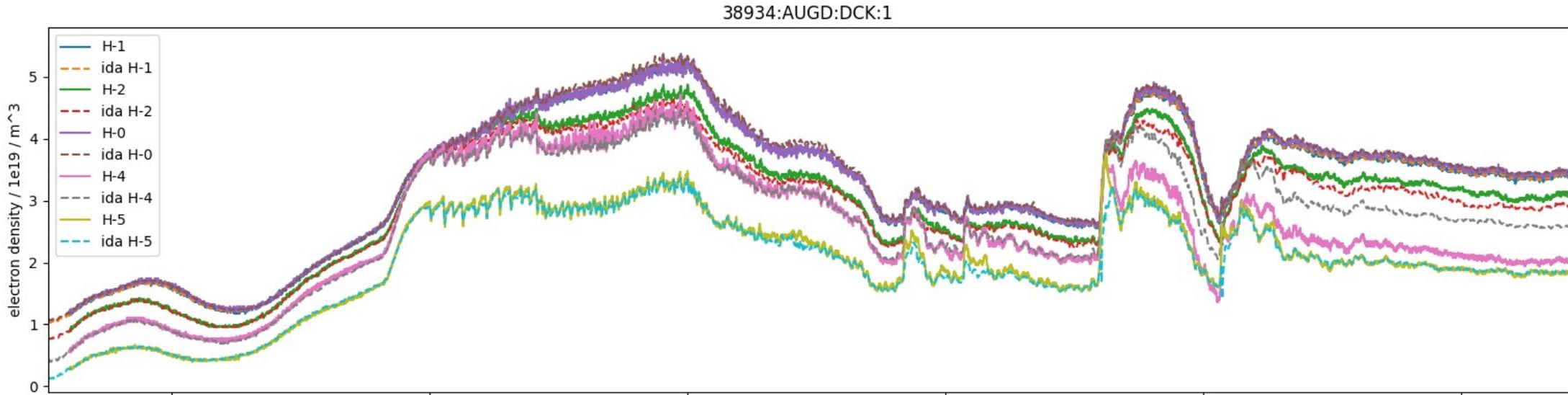
Interferometer Density



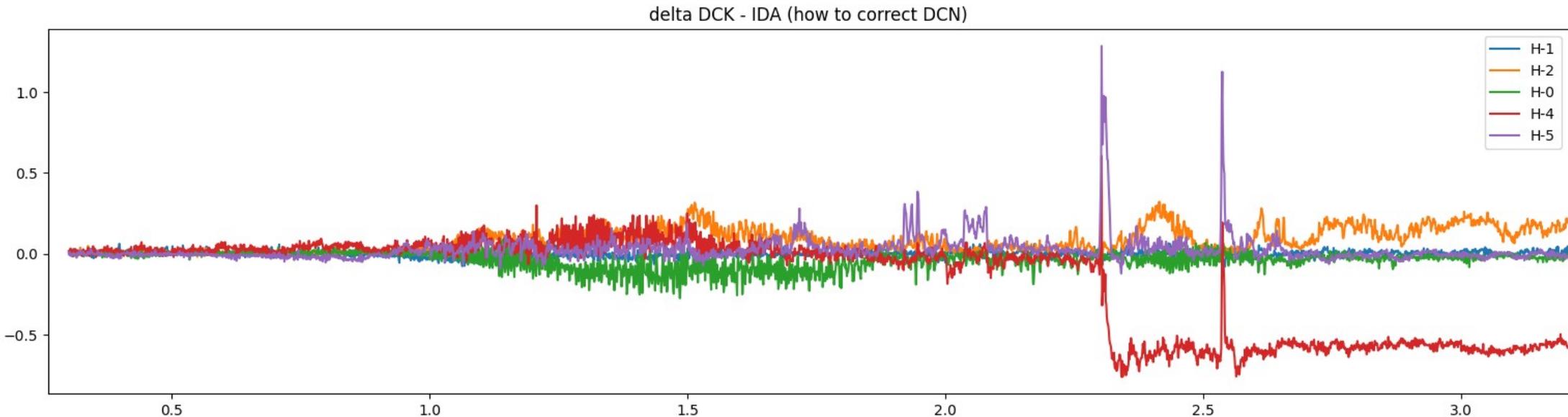
AUG Example for Interferometry (fringe jump)



Interferometer Density



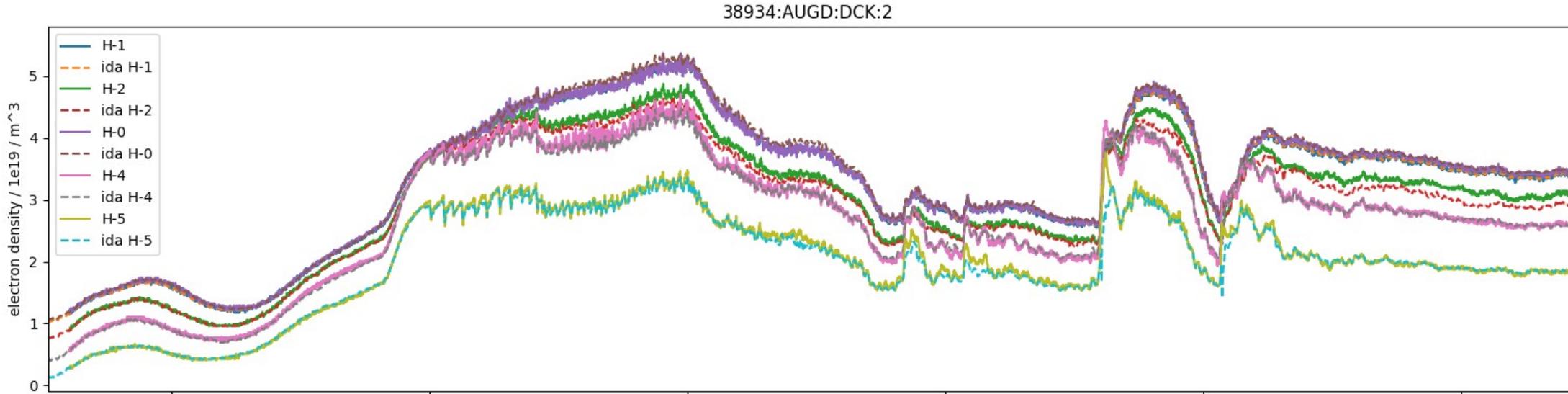
Density deviation



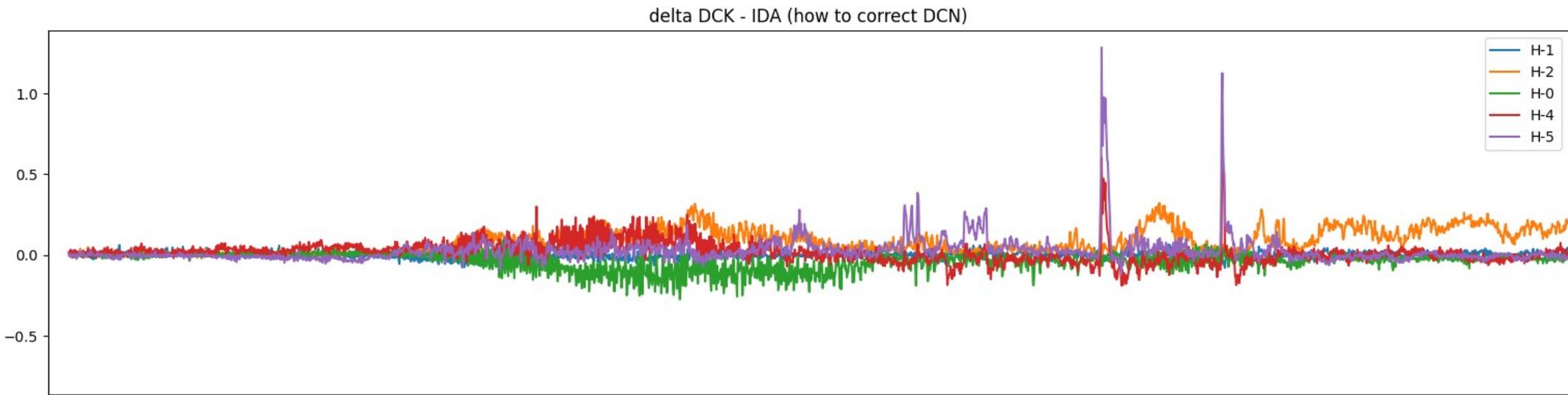
AUG Example for Interferometry (fringe jump resolved)



Interferometer Density



Density deviation



Single diagnostic validation for electron density



Single diagnostic validation for electron density



- Line integrated data like Interferometry → “Easy“

Low hanging fruit, assuming underlying function is somewhat symmetric/smooth

Single diagnostic validation for electron density

- Line integrated data like Interferometry → “Easy“

Low hanging fruit, assuming underlying function is somewhat symmetric/smooth

- Point estimates like Thomson Scattering → Hard

Analysis sensitive to

- Exact shape of profile is relevant
 - Exact calibration of all channels → Lasers? Lenses? ADC's?

Single diagnostic validation for electron density

- Line integrated data like Interferometry → “Easy“

Low hanging fruit, assuming underlying function is somewhat symmetric/smooth

- Point estimates like Thomson Scattering → Hard

Analysis sensitive to

- Exact shape of profile is relevant
 - Exact calibration of all channels → Lasers? Lenses? ADC's?
 - Expectations on the analysis include gradients etc

Single diagnostic validation for electron density

- Line integrated data like Interferometry → “Easy“

Low hanging fruit, assuming underlying function is somewhat symmetric/smooth

- Point estimates like Thomson Scattering → Hard

Analysis sensitive to

- Exact shape of profile is relevant
 - Exact calibration of all channels → Lasers? Lenses? ADC's?
 - Expectations on the analysis include gradients etc
- High time resolution?
 - very sensitivity to fast events like saw teeth etc

Multiple diagnostic validation

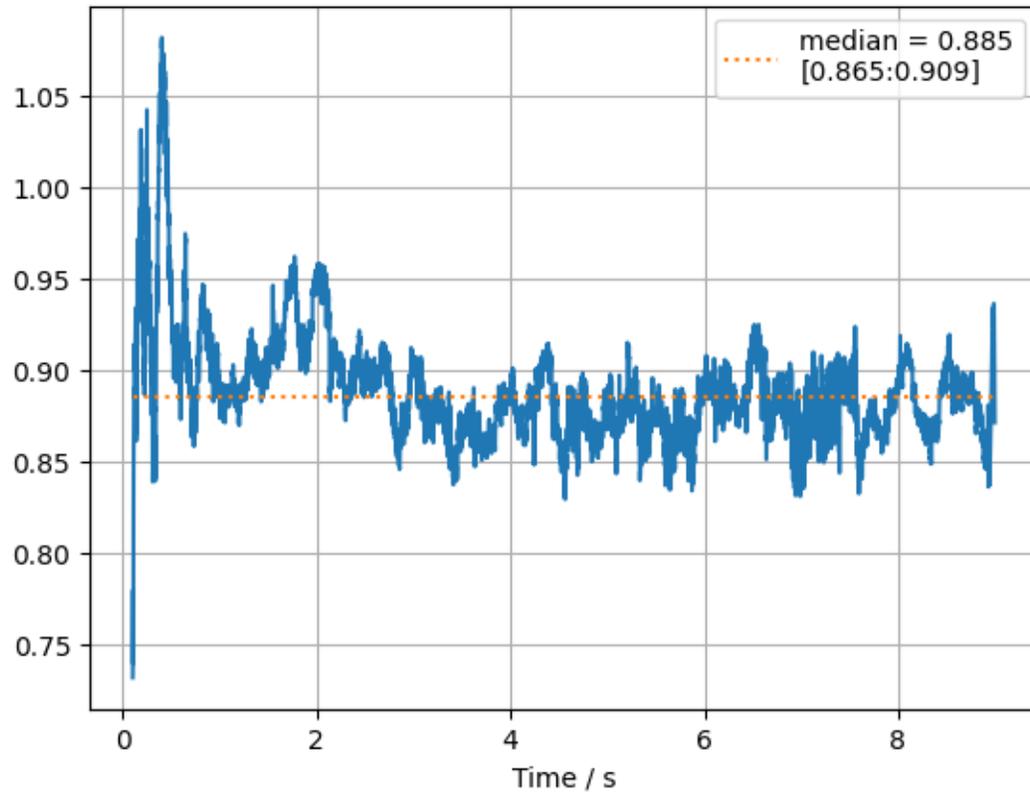
- Introduce assumptions about symmetries etc
- Allows to determine calibration factors

Thomson Scattering as example

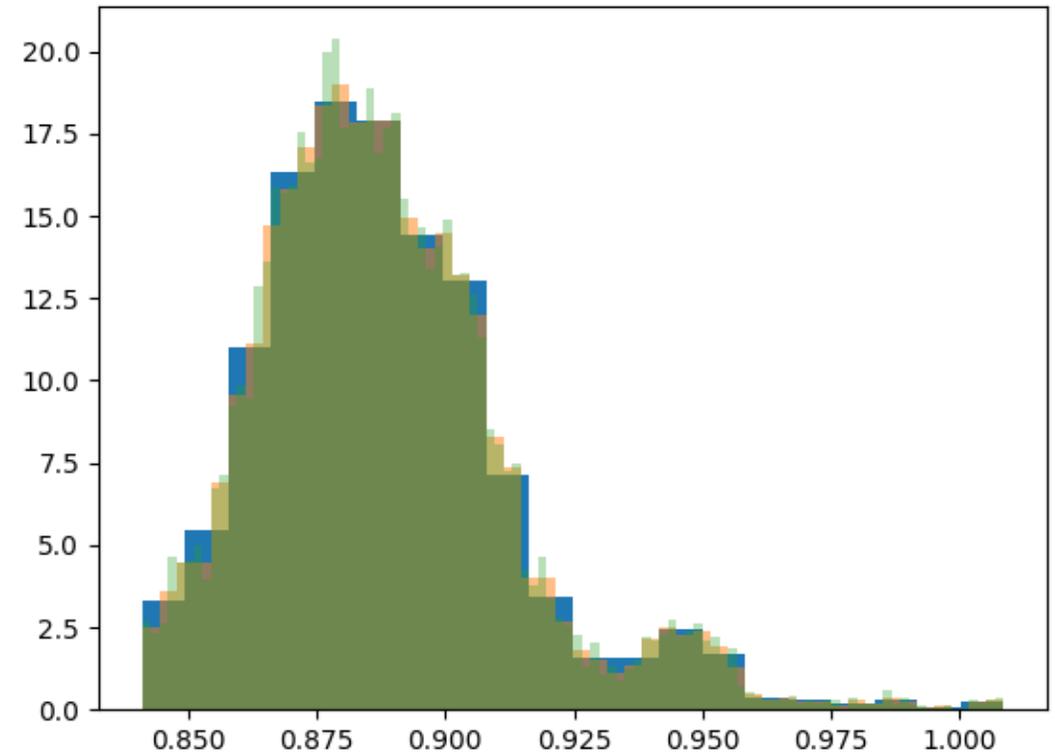
- Density depends linearly on absolute light intensity measured → hard to calibrate
- For the core system we trust the shape, but not the absolute level
- Define a ‚correction factor‘ (**tscnesc** = TS core ne scaling)
- If parameter too far off ‚expected value‘ like *1* sth is wrong → investigate

Multiple diagnostic validation

#41525 AUGD:IDA:1 sig tscnesc - Thomson Core Scaling



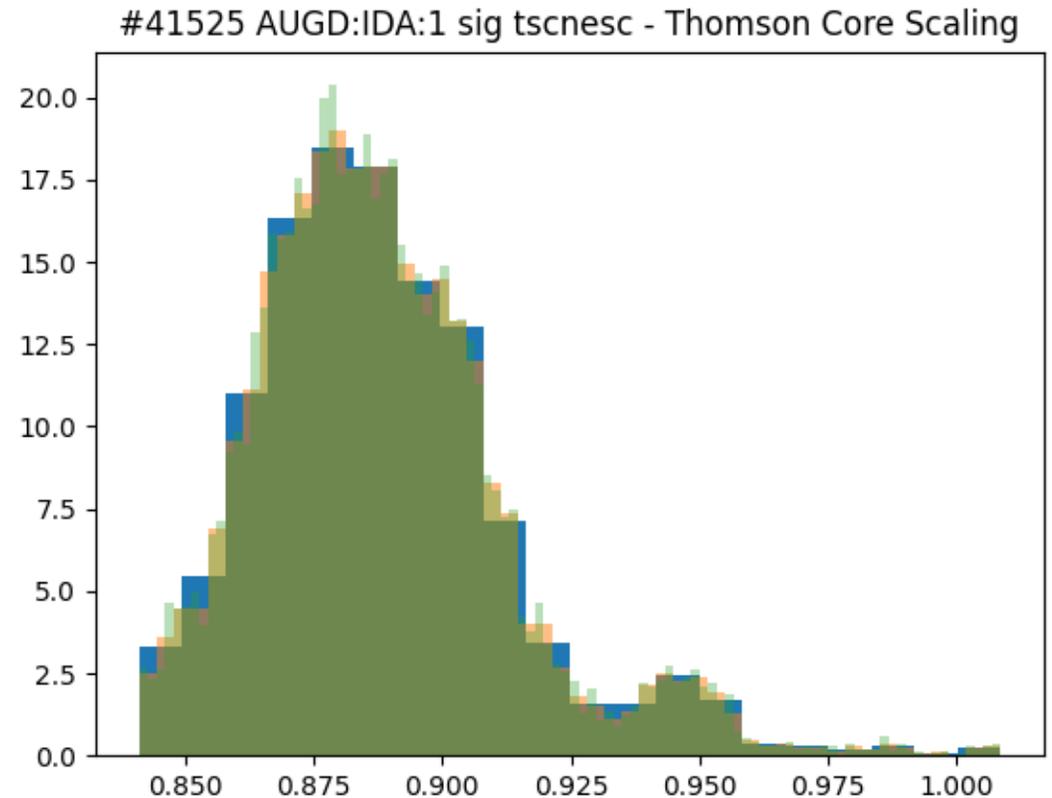
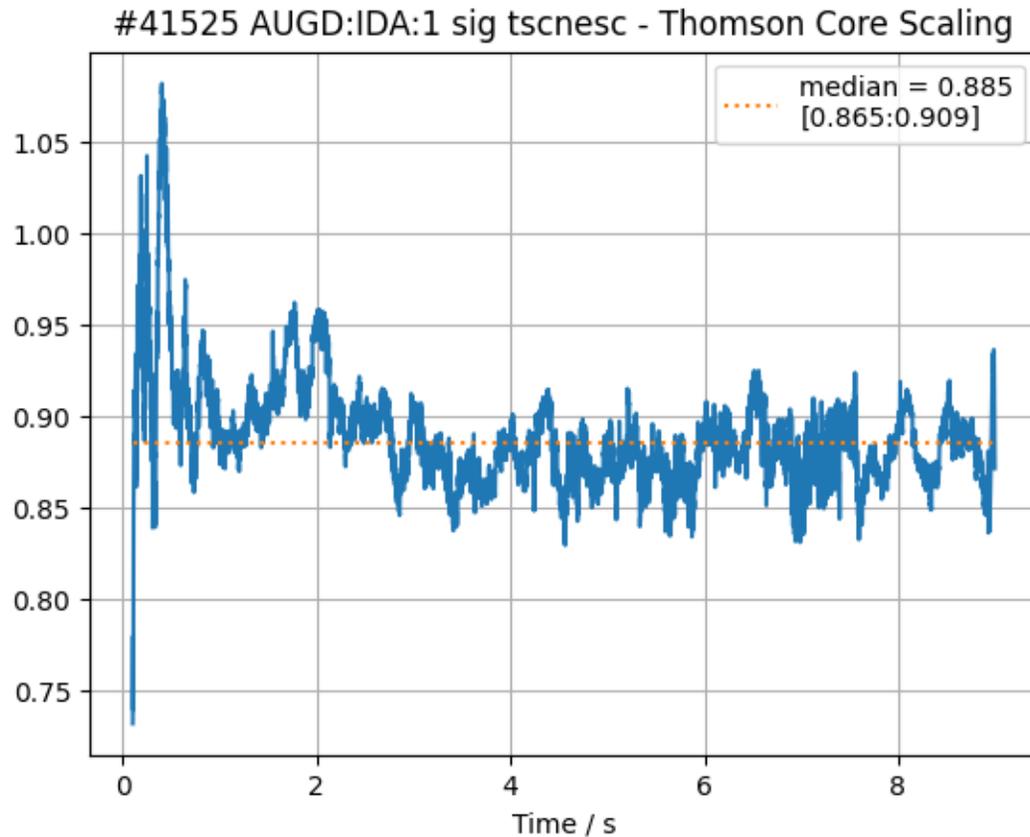
#41525 AUGD:IDA:1 sig tscnesc - Thomson Core Scaling



Multiple diagnostic validation



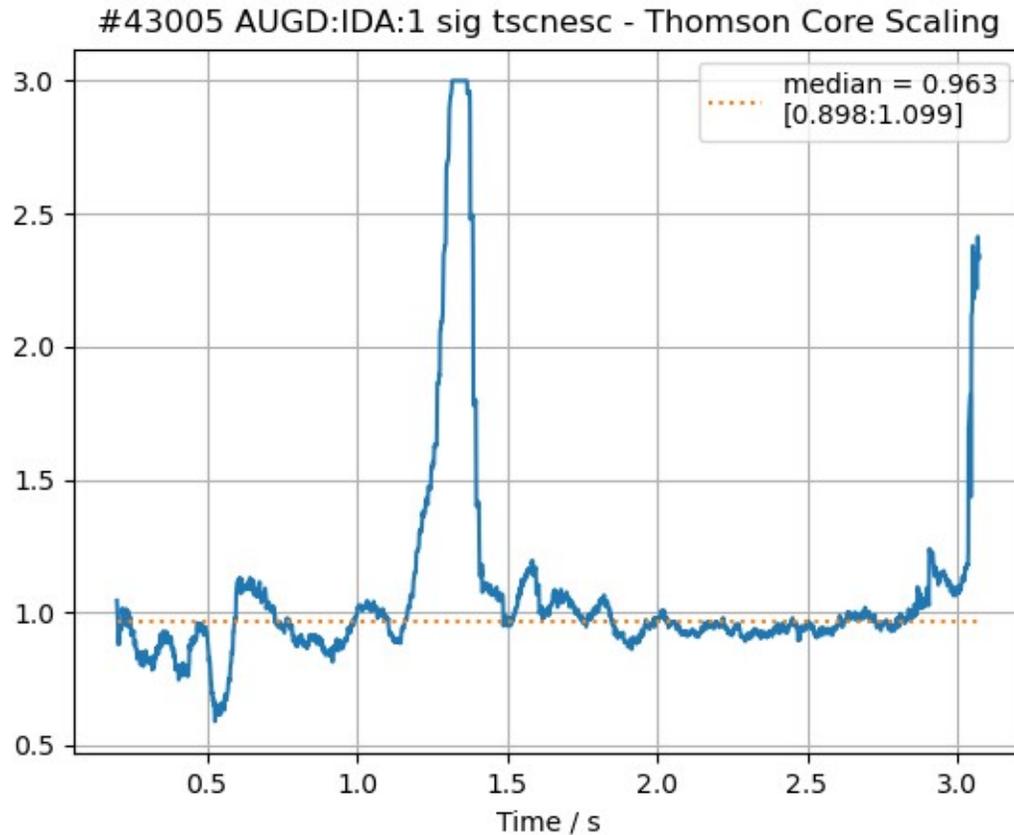
(Random) example for TS core scaling, median is 0.885 most values in [0.865 to 0.909]



Multiple diagnostic validation



Interferometer before/after plasma event?



Multiple diagnostic validation

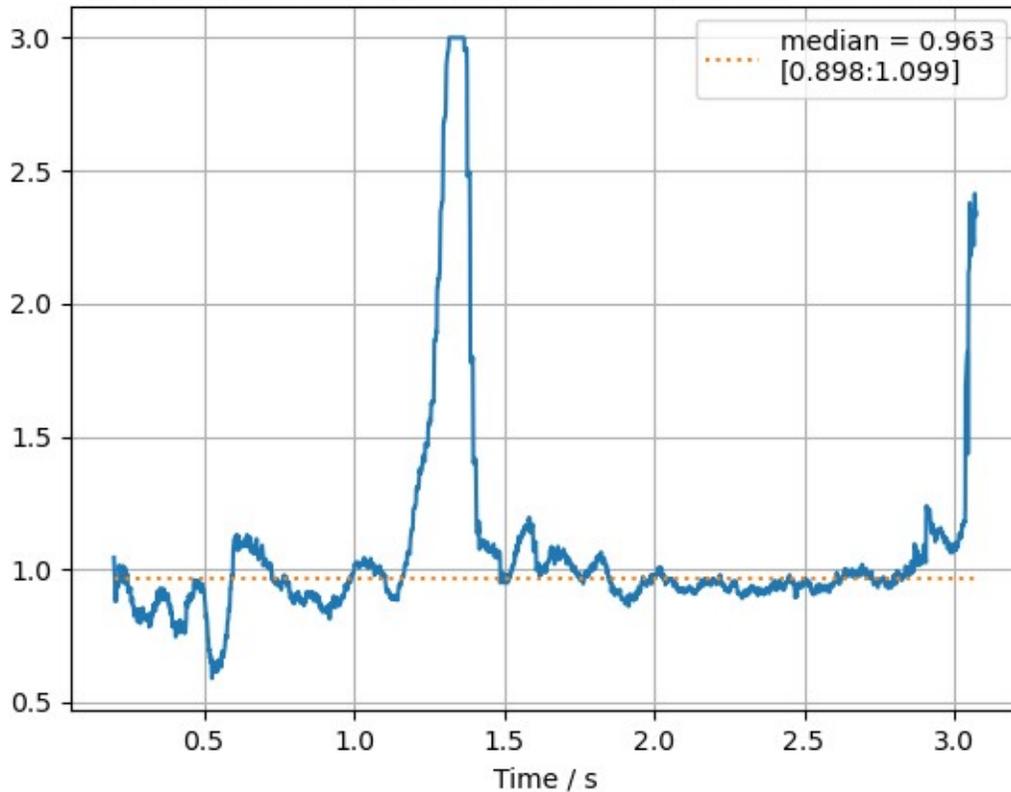


Interferometer before/after plasma event?

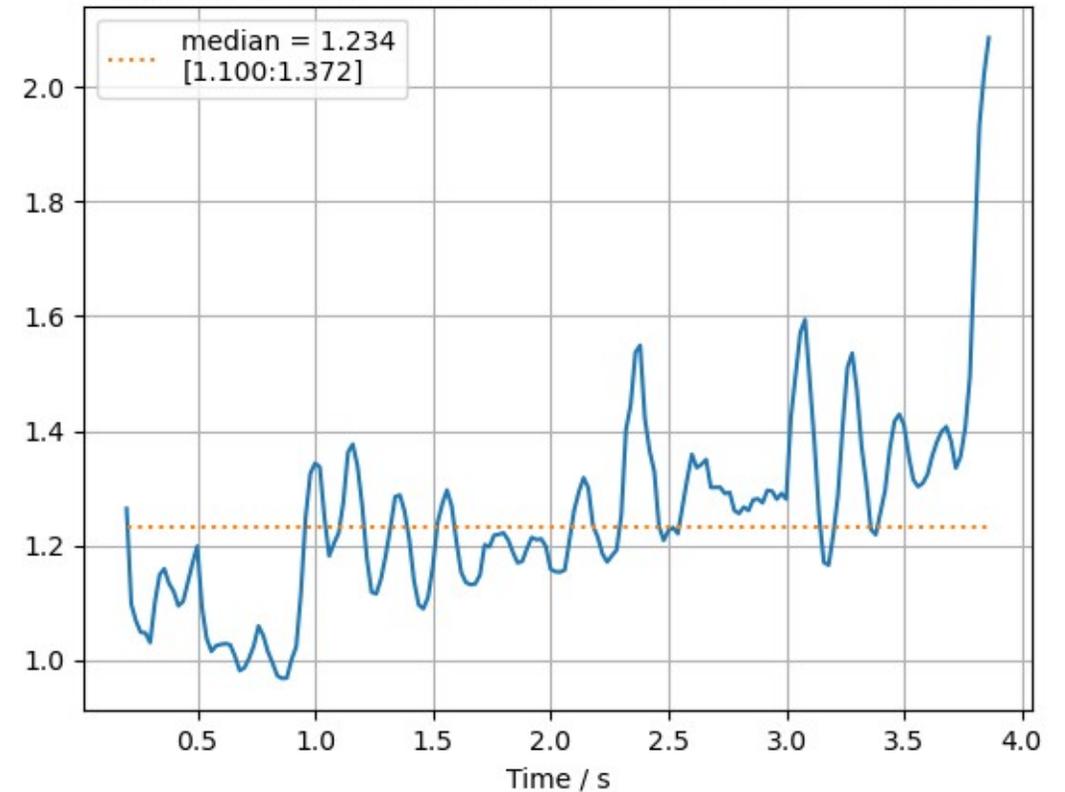
Interferometer fringe jumps missed

(what to correct for, if all channels jump?)

#43005 AUGD:IDA:1 sig tscnesc - Thomson Core Scaling



#43119 AUGI:IDA:1 sig tscnesc - Thomson Core Scaling



Issues when jointly describing different diagnostics



Possible explanations have to be investigated for each suspected disagreement

- Toroidal and poloidal asymmetries in density & temperature
 - reasonable for n_e & T_e at high densities in the edge, to some degree in core
 - W concentration has strong dependence on poloidal angle (centripetal force)
 - $n=1$ offset between equilibrium and “machine”
 - alignment vacuum vessel / diagnostics wrt field coils
- Absolute calibration of any diagnostic
 - alignment, is it fixed in space? Vibrations?
 - quality of optical components, degradation? Temperature/Field dependence?
- Wrong assumptions in models or insufficient solvers
- ...

AUG – mapping uncertainties



- “The separatrix at the low field side can only be estimated to within **5mm.**”
Rainer Fischer,

AUG – mapping uncertainties



- “The separatrix at the low field side can only be estimated to within **5mm.**”
Rainer Fischer, for as long as I know him,

AUG – mapping uncertainties



- “The separatrix at the low field side can only be estimated to within **5mm.**”
Rainer Fischer, for as long as I know him, every month or so

Other topics

- Identifying and scanning „Hyper-parameters“
Example: Plasma position at WEST → added delta z and delta R
Result: describing all 10(ish) interferometer channels & reflectometer only works with a vertical shift of few cm
- Model comparison (evaluating evidence)
We have two different ways to describe a quantity (spline with various number of knots or 1D vs 2D) → which is more appropriate?
Application: do the analysis and compare the Evidence to prevent overfitting
- Low/High-fidelity models:
Example: We have two forward models with different levels of accuracy and/or different assumptions → which to use?
Application: do the analysis for all models and compare how well diagnostics can be described jointly



Backup Slides

AUG IDA

Shotfile # 43205

AUGD other: AUG

Diagnostic

IDA

Edition 0

Time 3.00000

Time

Load Shotfile

 ne Te pe grad_ne grad_Te grad_pe

Draw

Set Scale

Reset Zoom

Time Profile

Residua

Data Fit

Contour

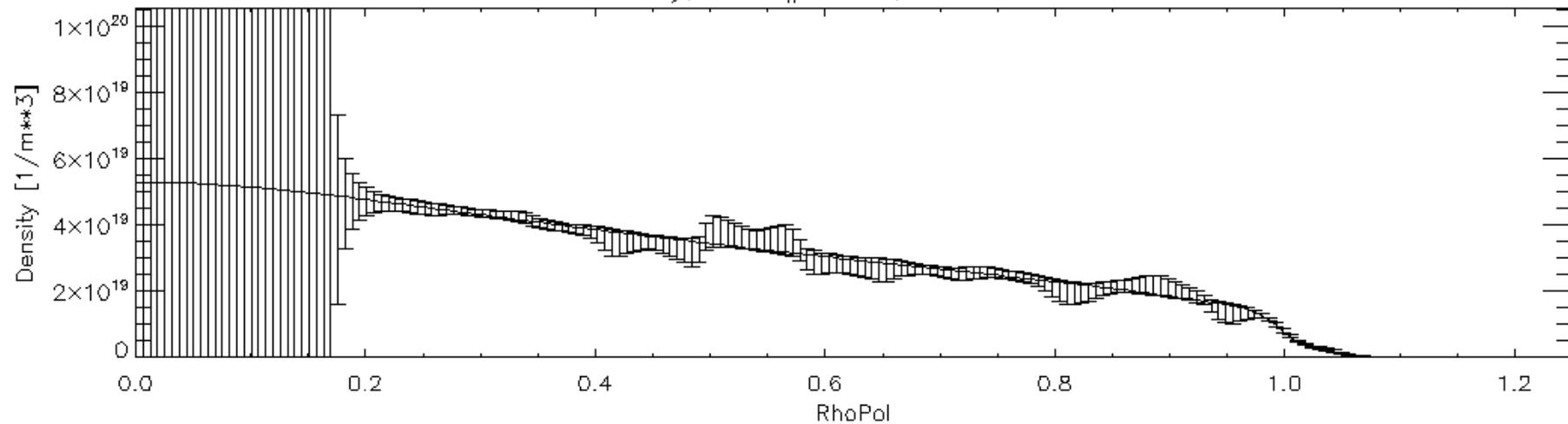
ps15m2

Print

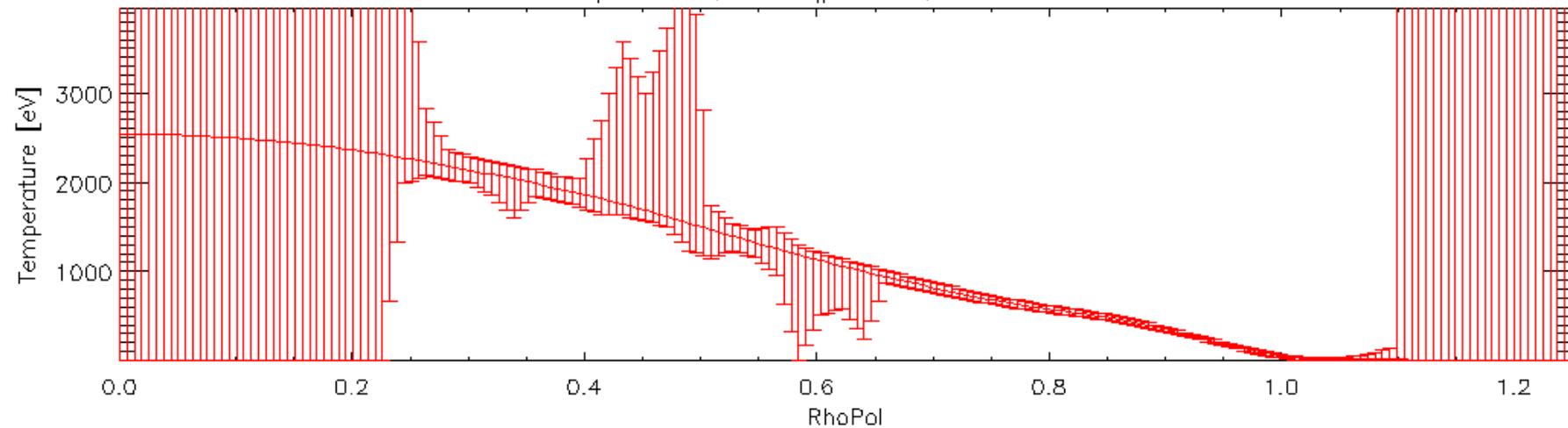
Save

Quit

Density, Shot #43205, Time: 3.00000



Temperature, Shot #43205, Time: 3.00000



AUG IDA

- ◆ LIB
- ◆ DCN
- ◆ ECE
- ◆ REF
- ◆ TS_ne
- ◆ TS_Te
- ◆ HEB

Draw

Set Scale

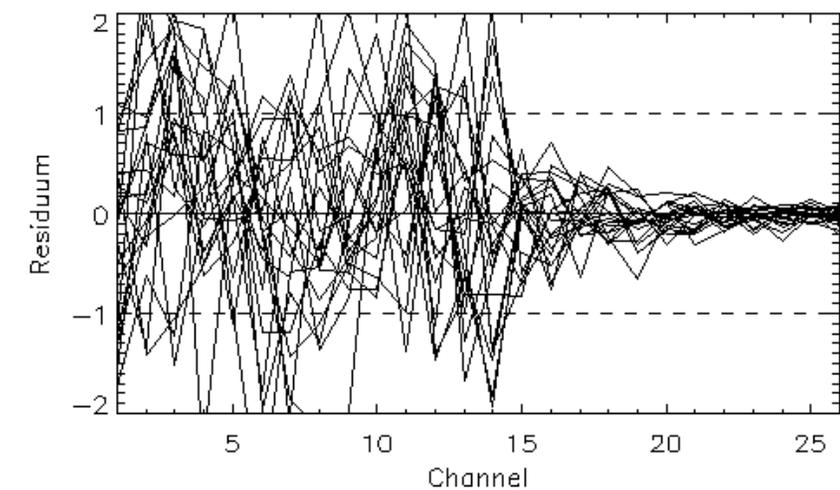
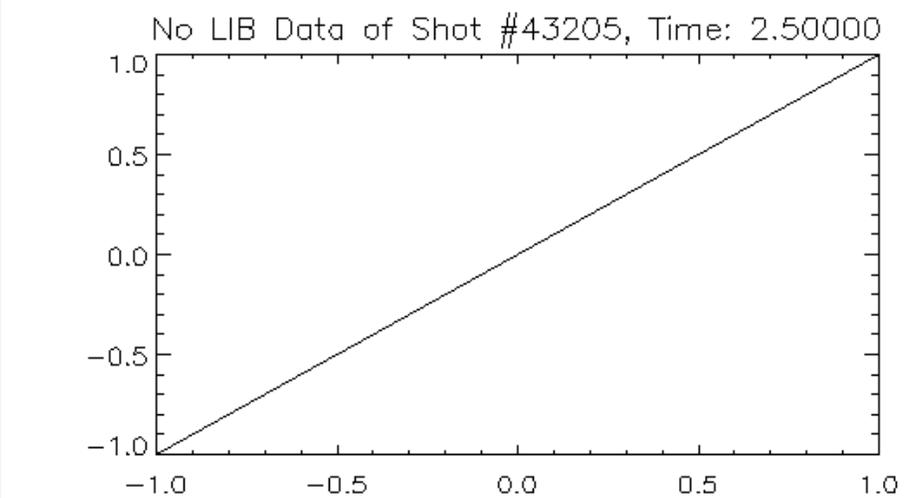
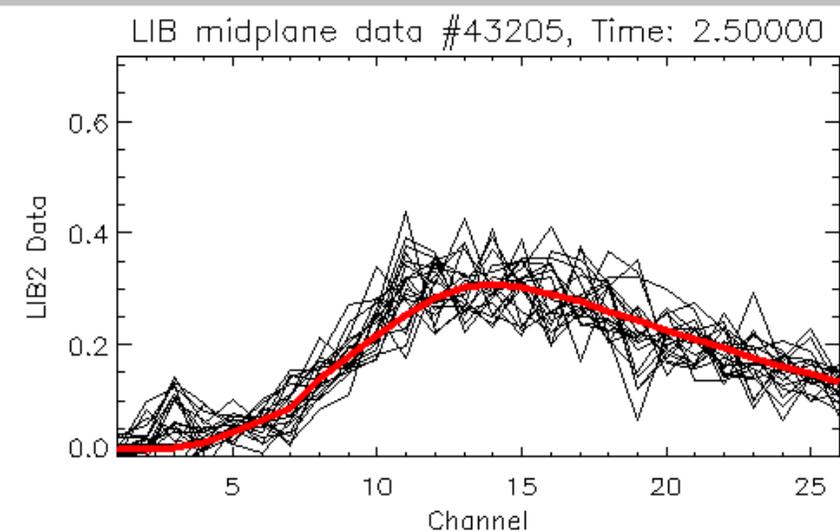
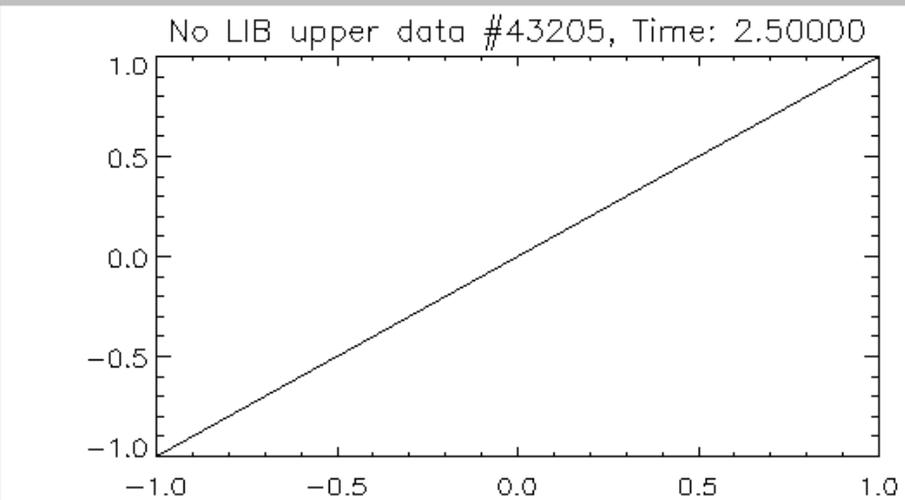
Reset Zoom

ps15m2

Print

Save

Close



AUG IDA

- ◆ LIB
- ◆ DCN
- ◆ ECE
- ◆ REF
- ◆ TS_ne
- ◆ TS_Te
- ◆ HEB

Draw

Set Scale

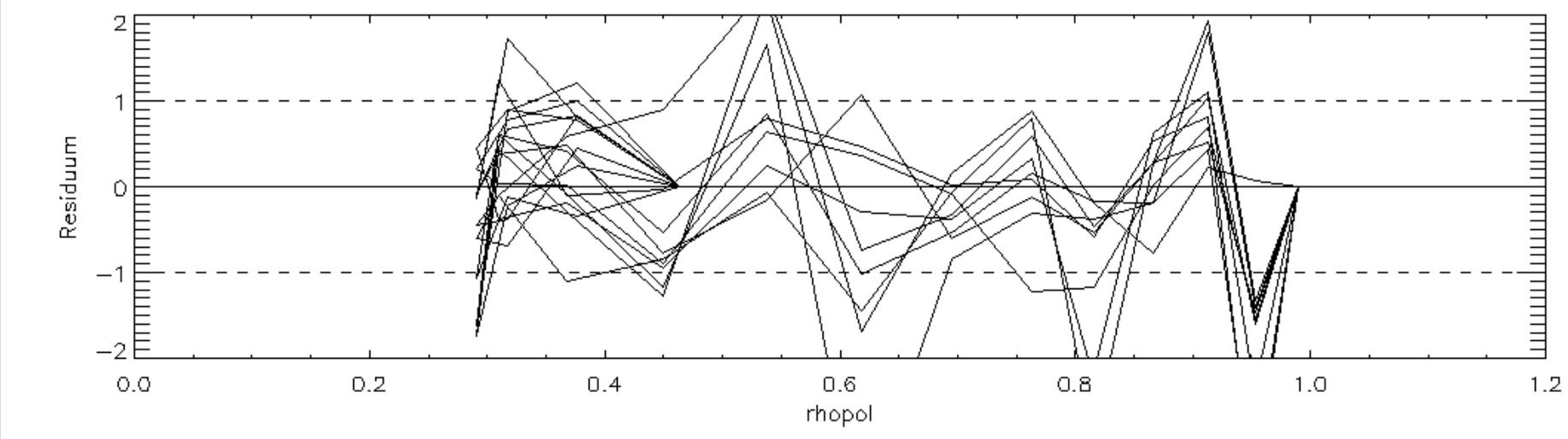
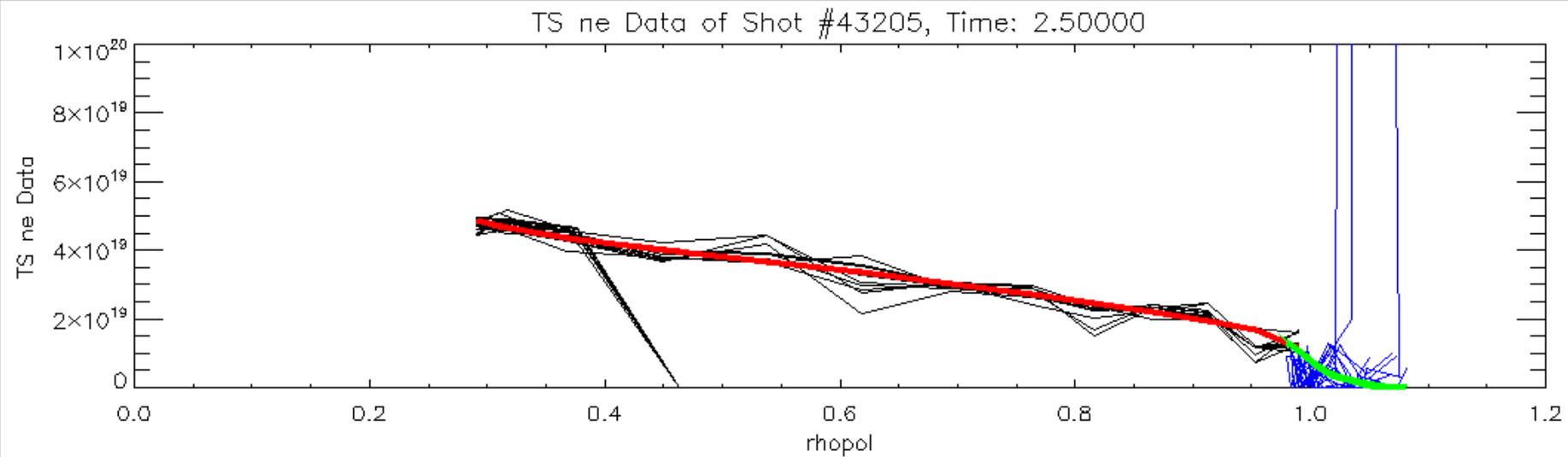
Reset Zoom

ps15m2

Print

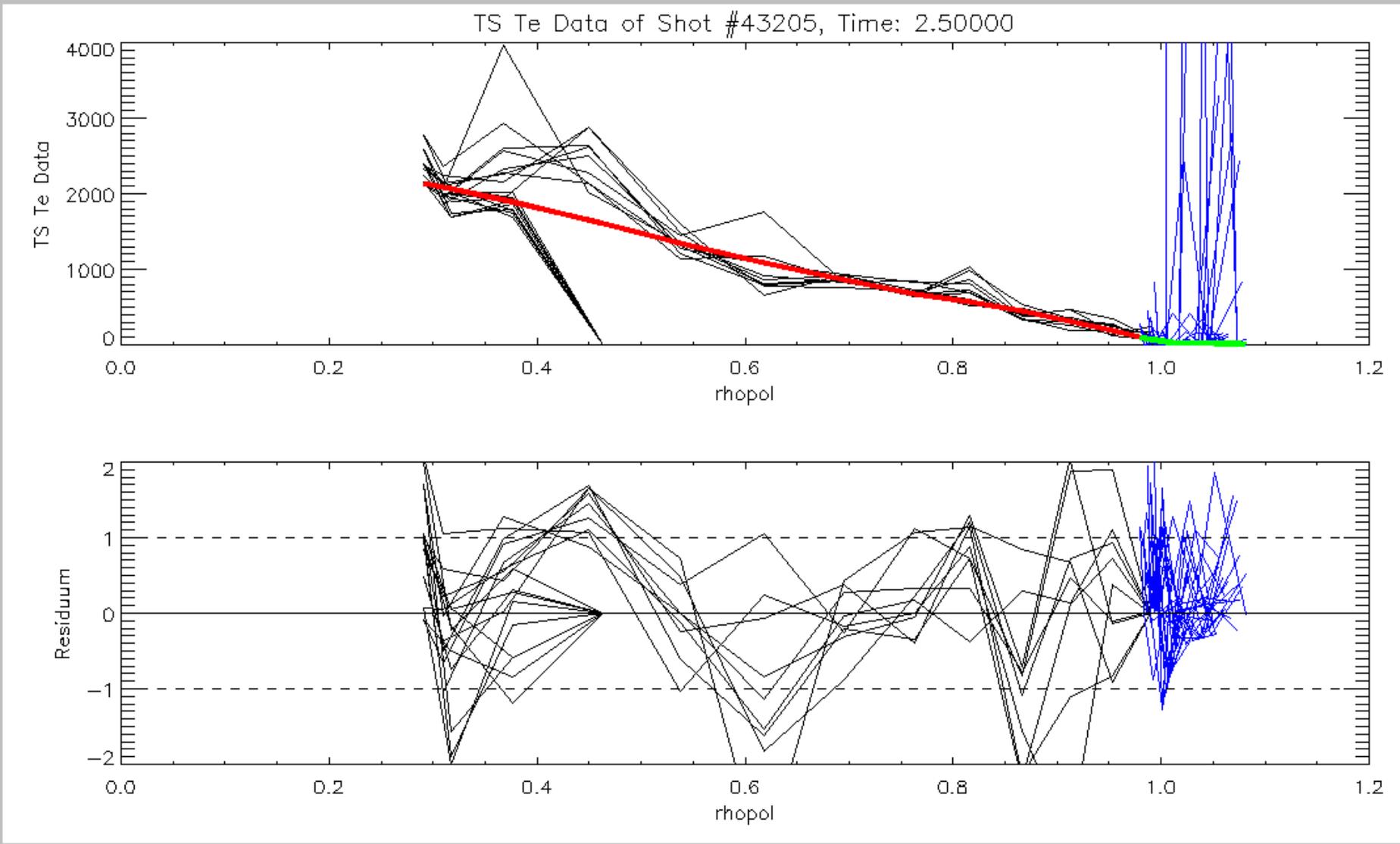
Save

Close



AUG IDA

- ◆ LIB
- ◆ DCN
- ◆ ECE
- ◆ REF
- ◆ TS_ne
- ◆ TS_Te
- ◆ HEB



- Draw
- Set Scale
- Reset Zoom

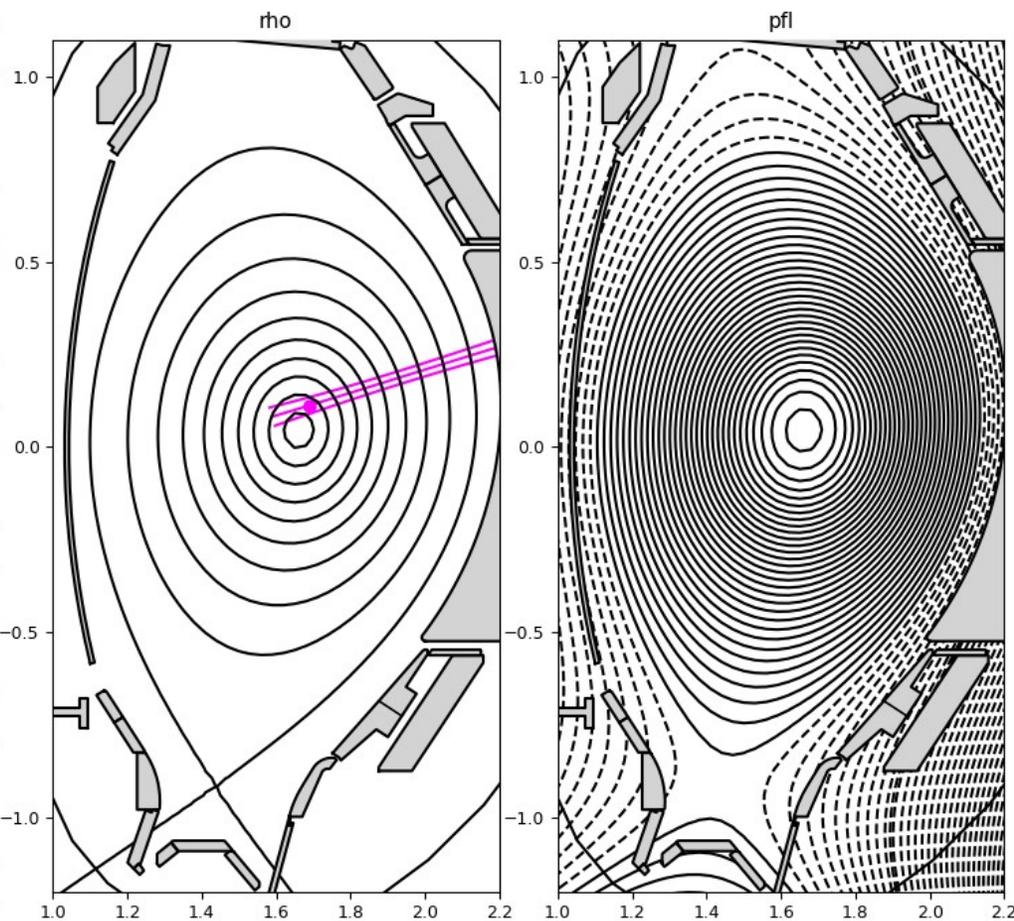
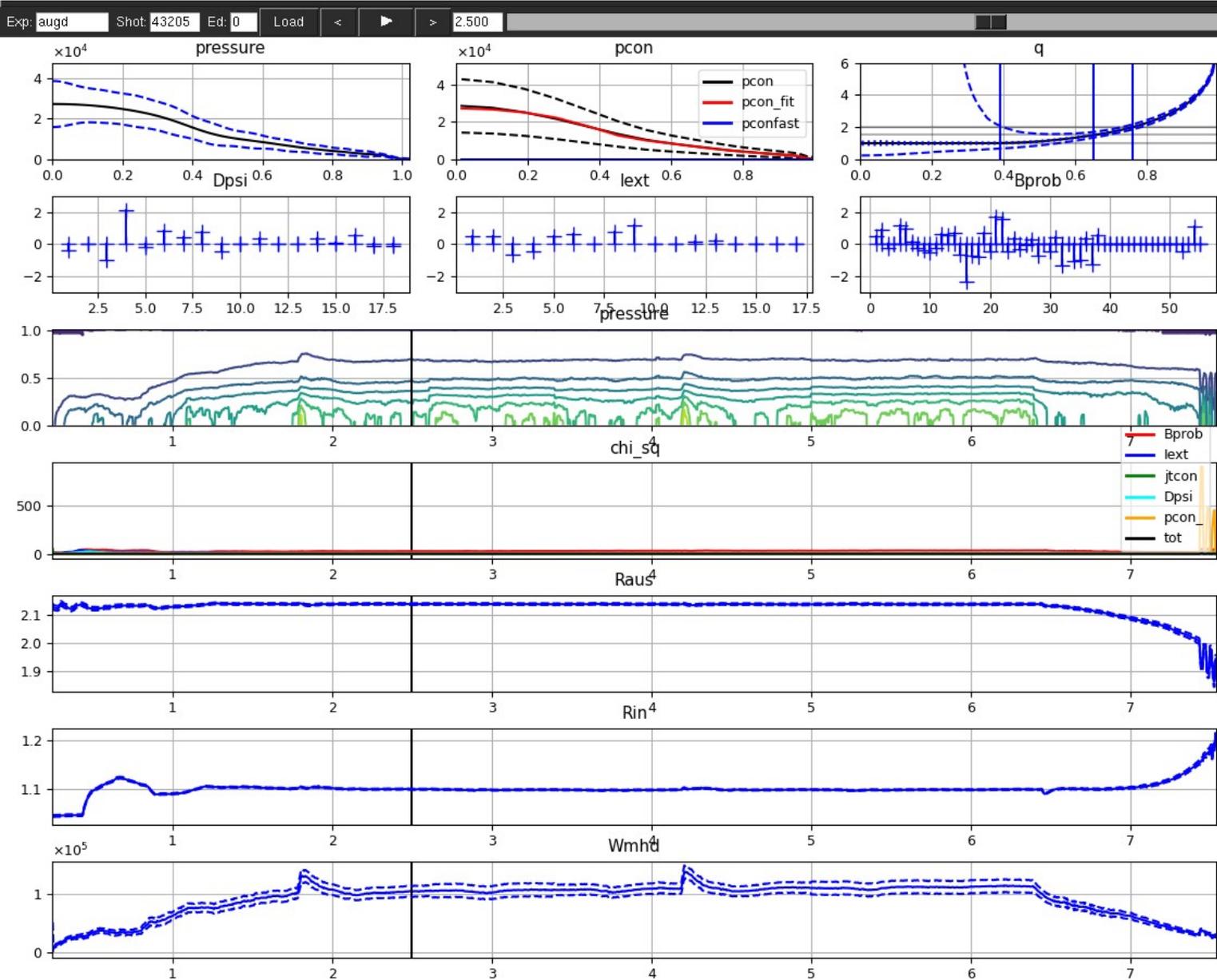
ps15m2

- Print
- Save
- Close

2.50000



AUG IDA



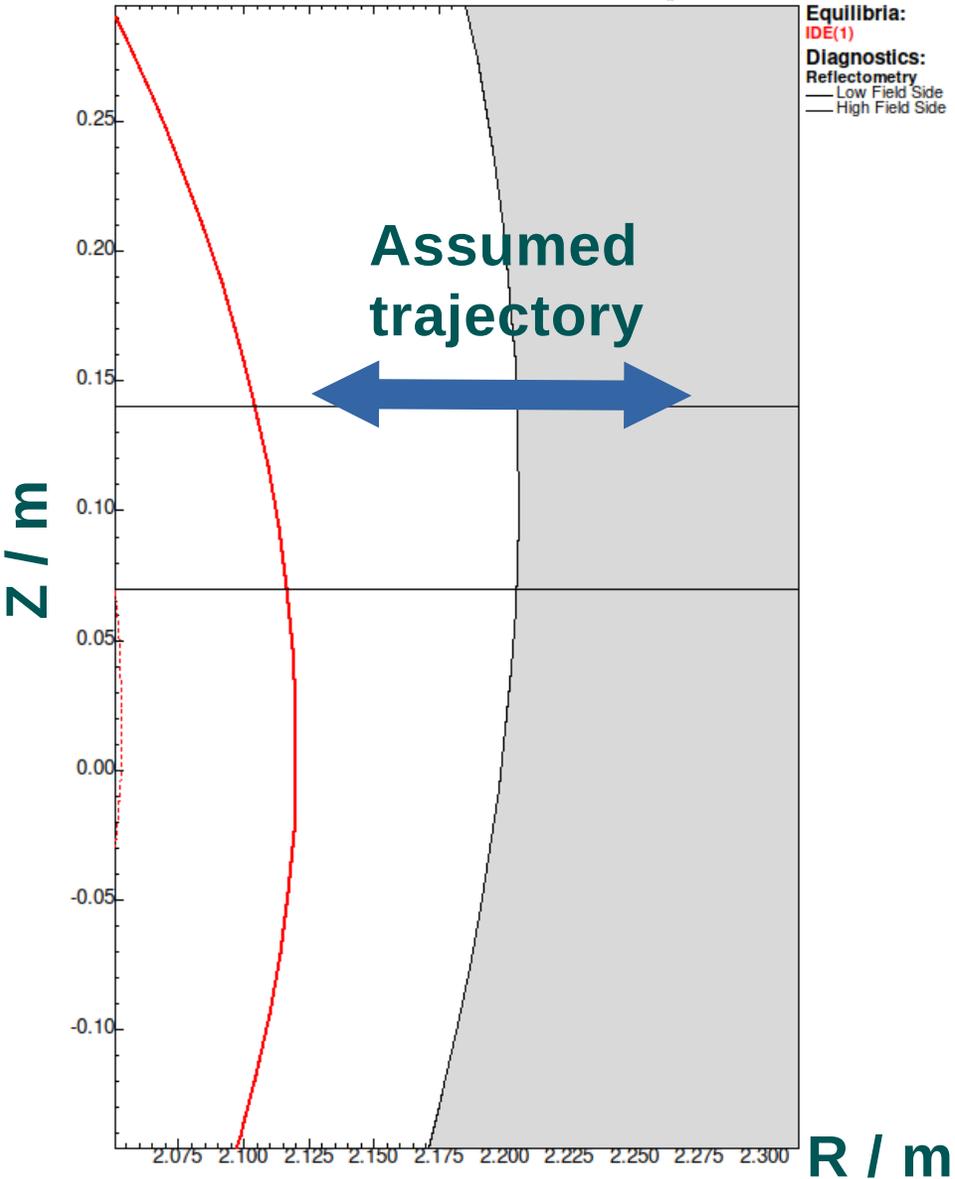
AUG – mapping reflectometry



Poloidal cross-section

#43262 2.985 s

Equilibria:
IDE(1)
Diagnostics:
Reflectometry
— Low Field Side
— High Field Side



AUG – mapping reflectometry



Poloidal cross-section

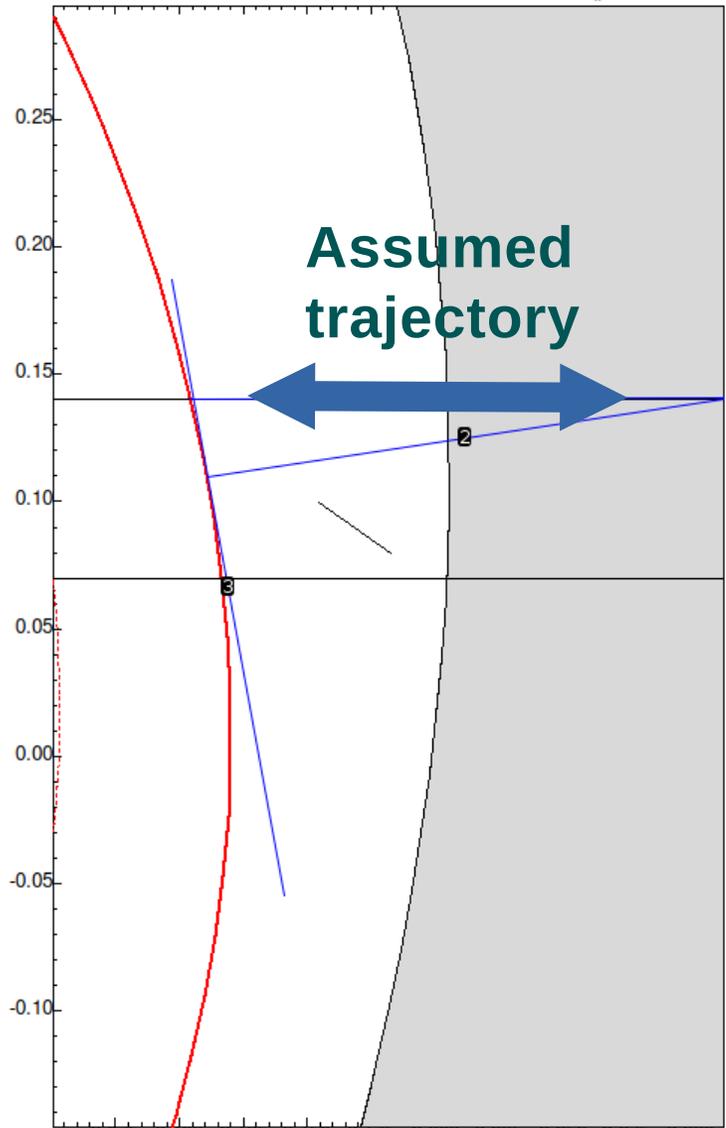
#43262 2.985 s

Equilibria:
IDE(1)
Diagnostics:
Reflectometry
— Low Field Side a
— High Field Side e

Assumed trajectory



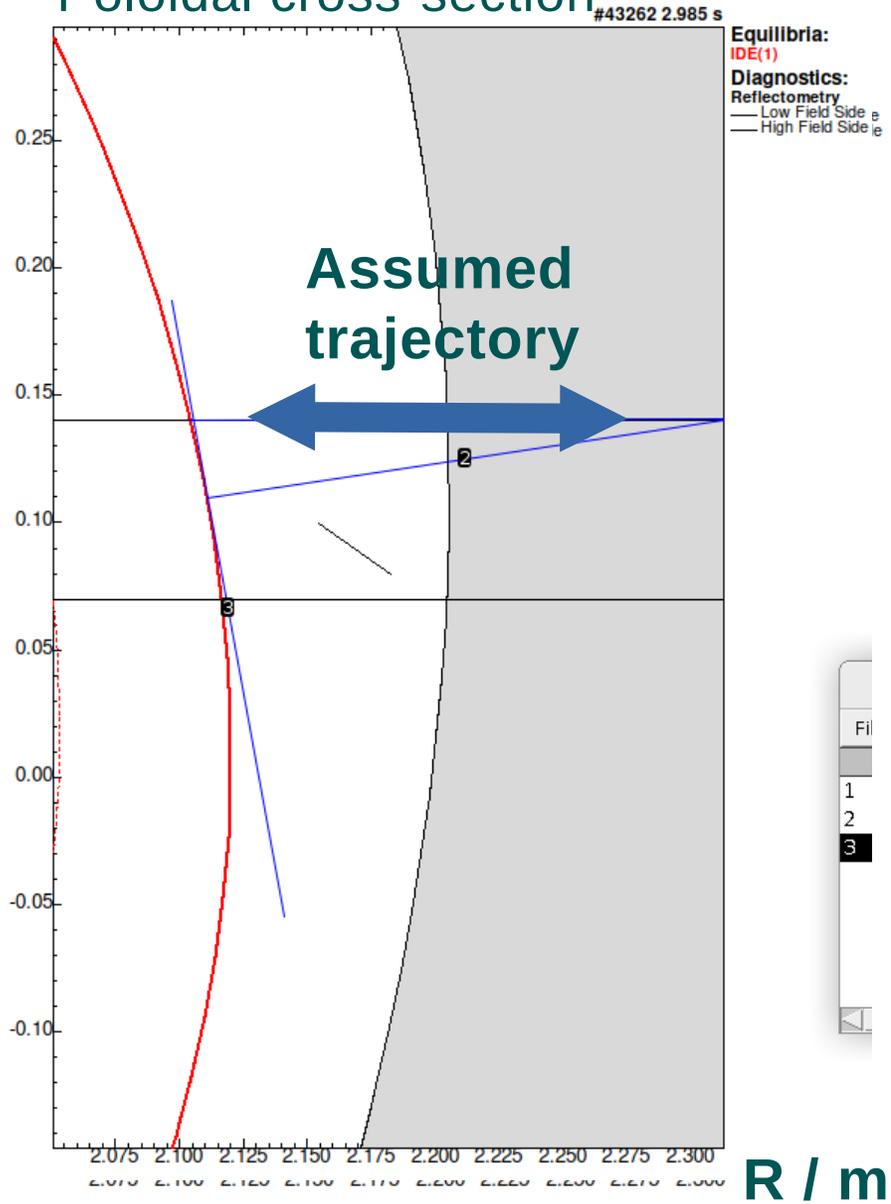
Z / m



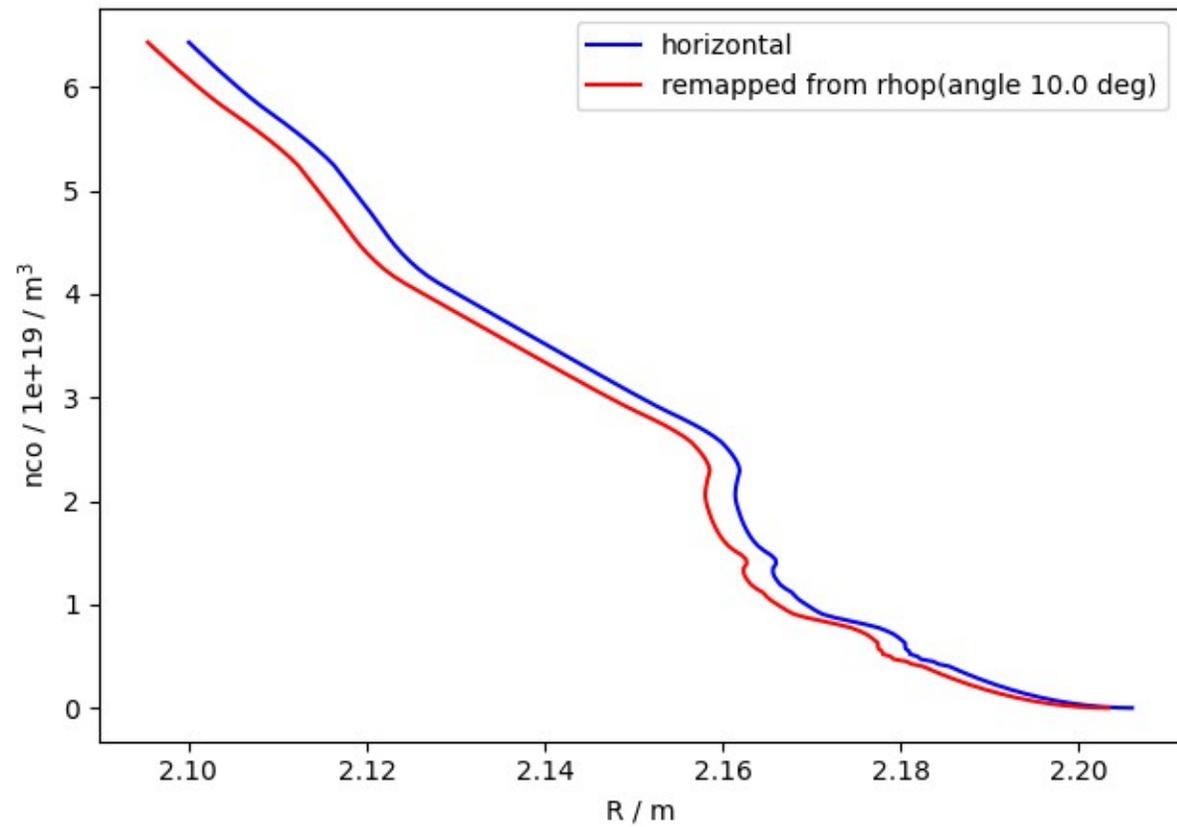
R / m

AUG – mapping reflectometry

Poloidal cross-section

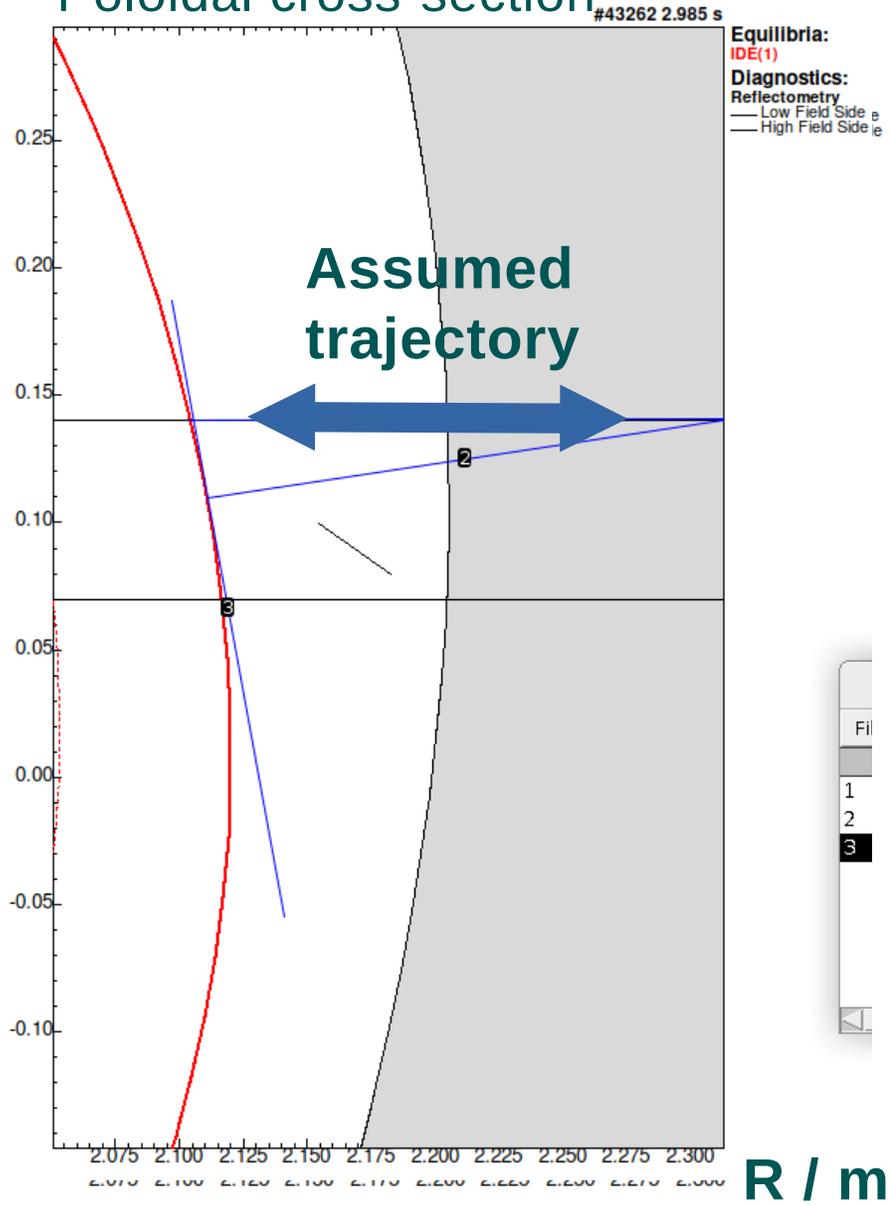


AUGD:RPS:43262 at 2.000 s

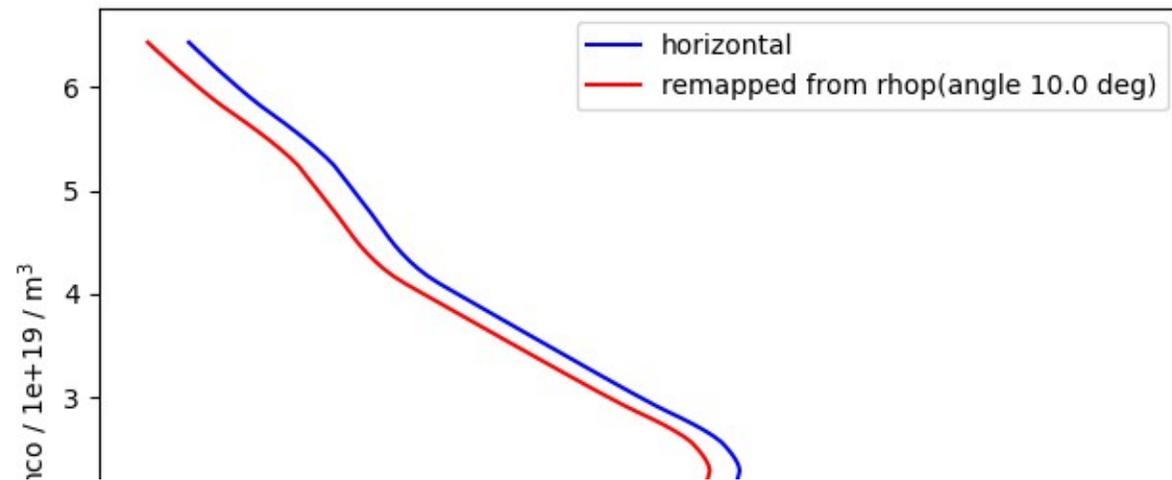


AUG – mapping reflectometry

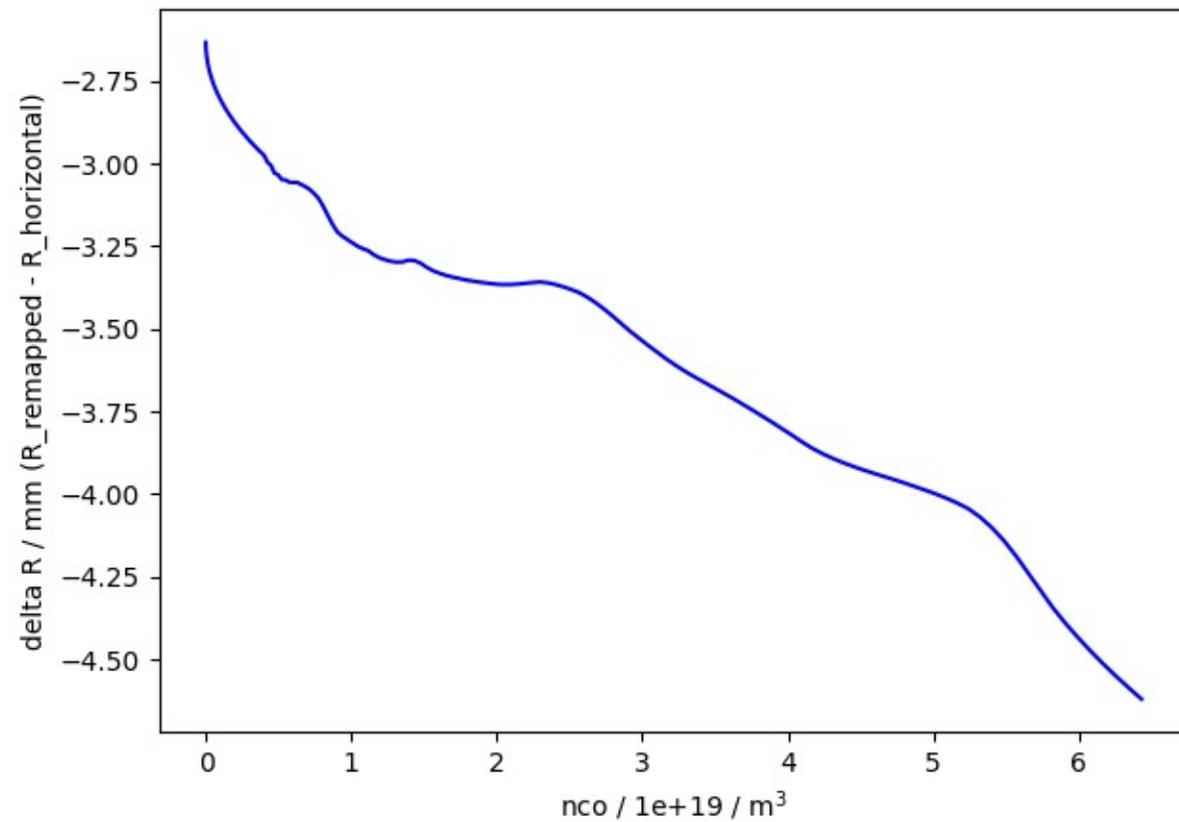
Poloidal cross-section



AUGD:RPS:43262 at 2.000 s



AUGD:RPS:43262 at 2.000 s

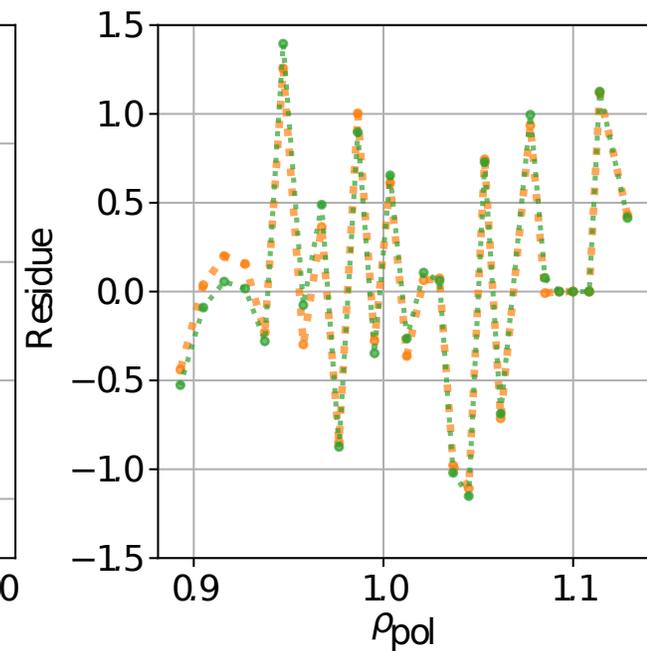
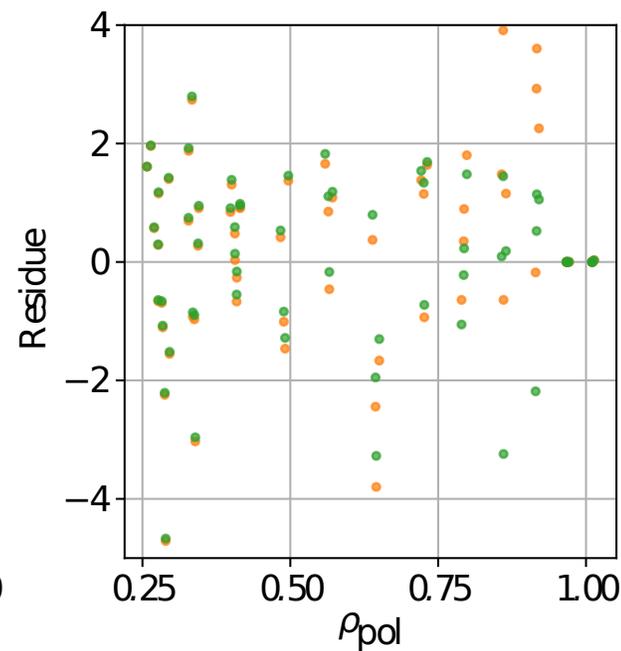
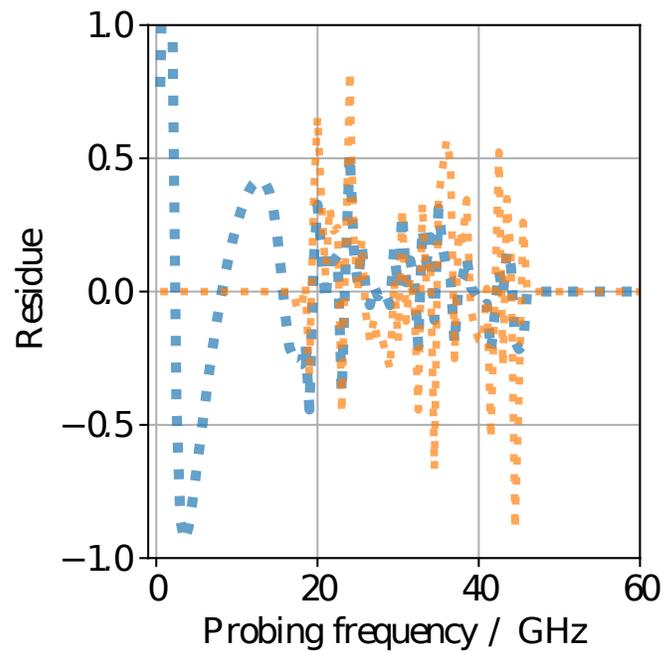
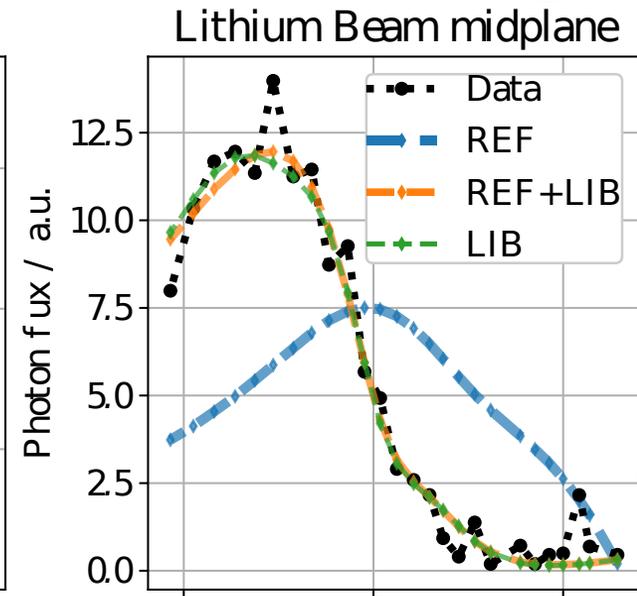
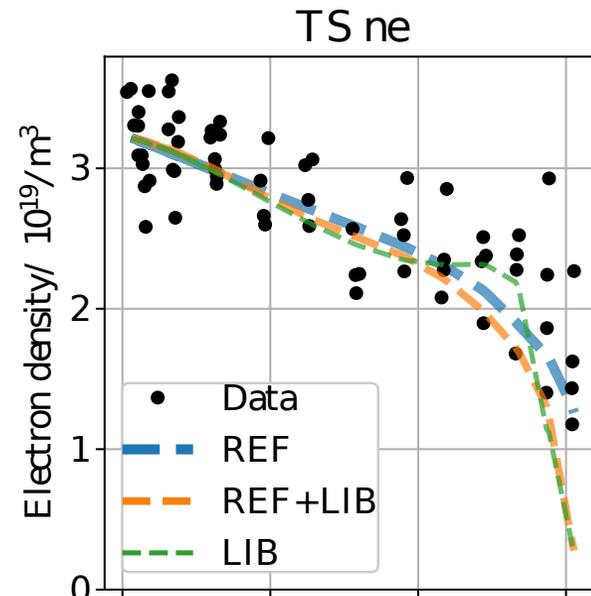
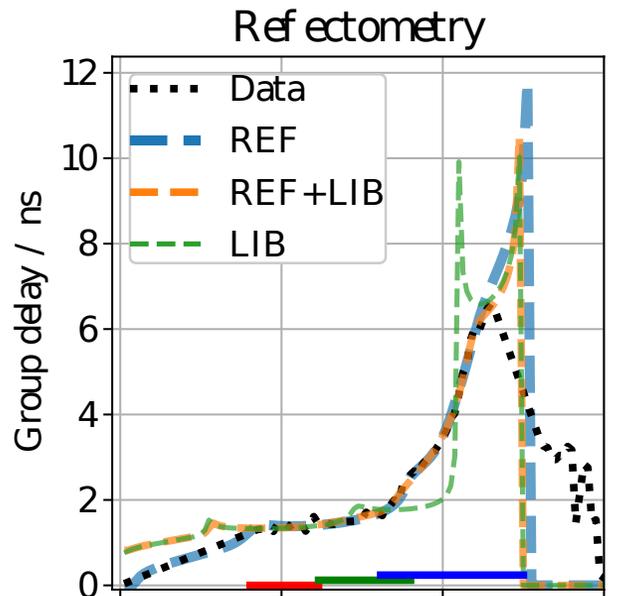
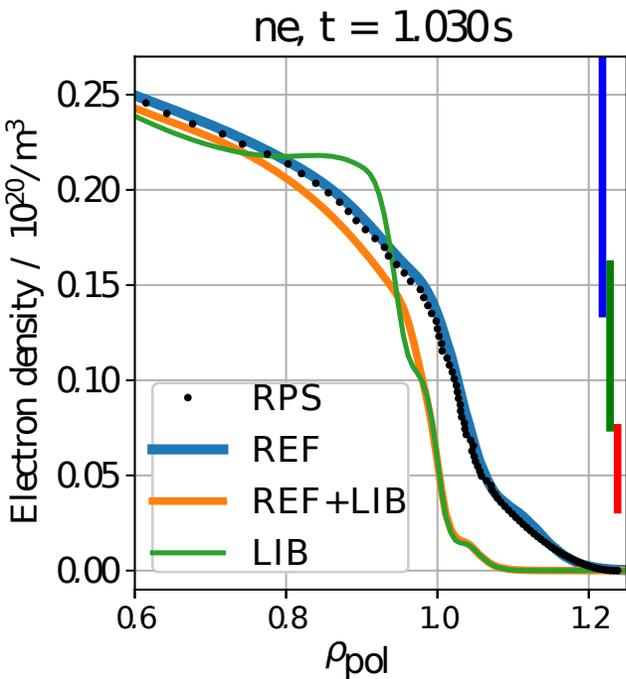


More AUG – Thomson & Lithium Beam & Reflectometry



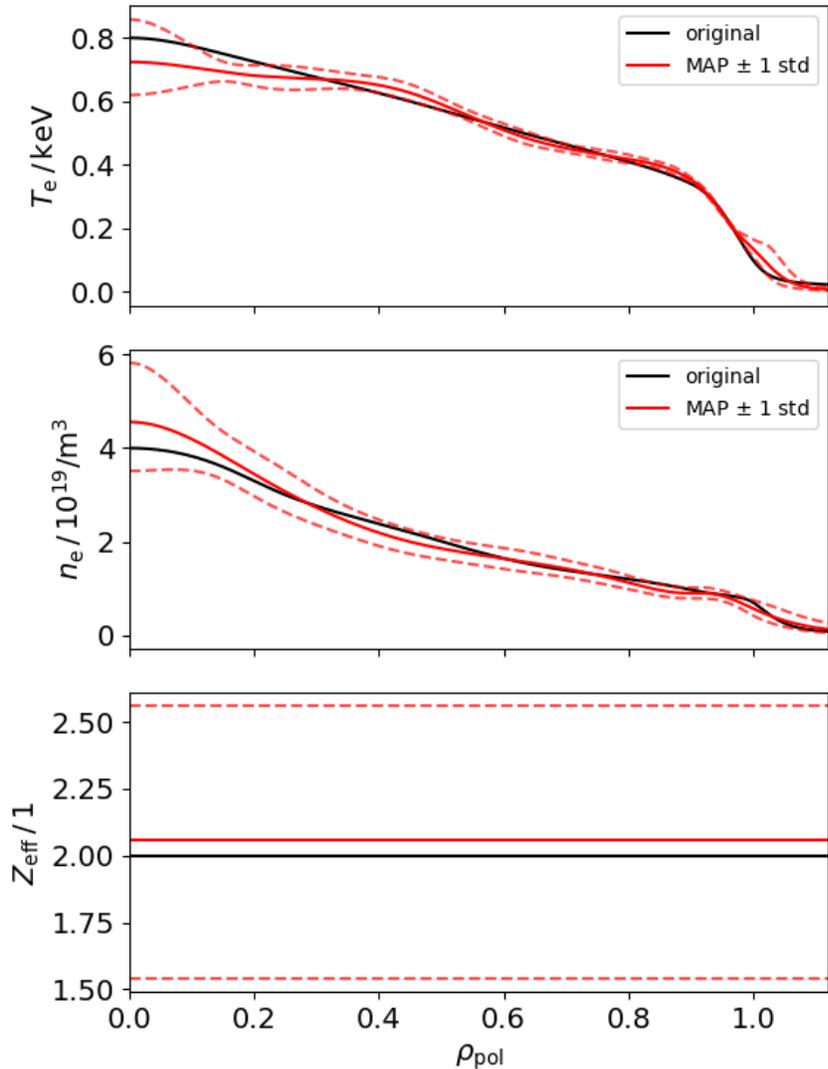
- Pedestal top at AUG is ,no-mans lands‘ → often no good measurements
- Reflectometry recently included in IDA → compare all three diagnostics at once
- Result: Lithium Beam not informative at pedestal top
- Reflectometry at odds with Thomson System
- Investigation ongoing

AUG – mapping various diagnostics



Synthetic Test-case (JT-60SA)

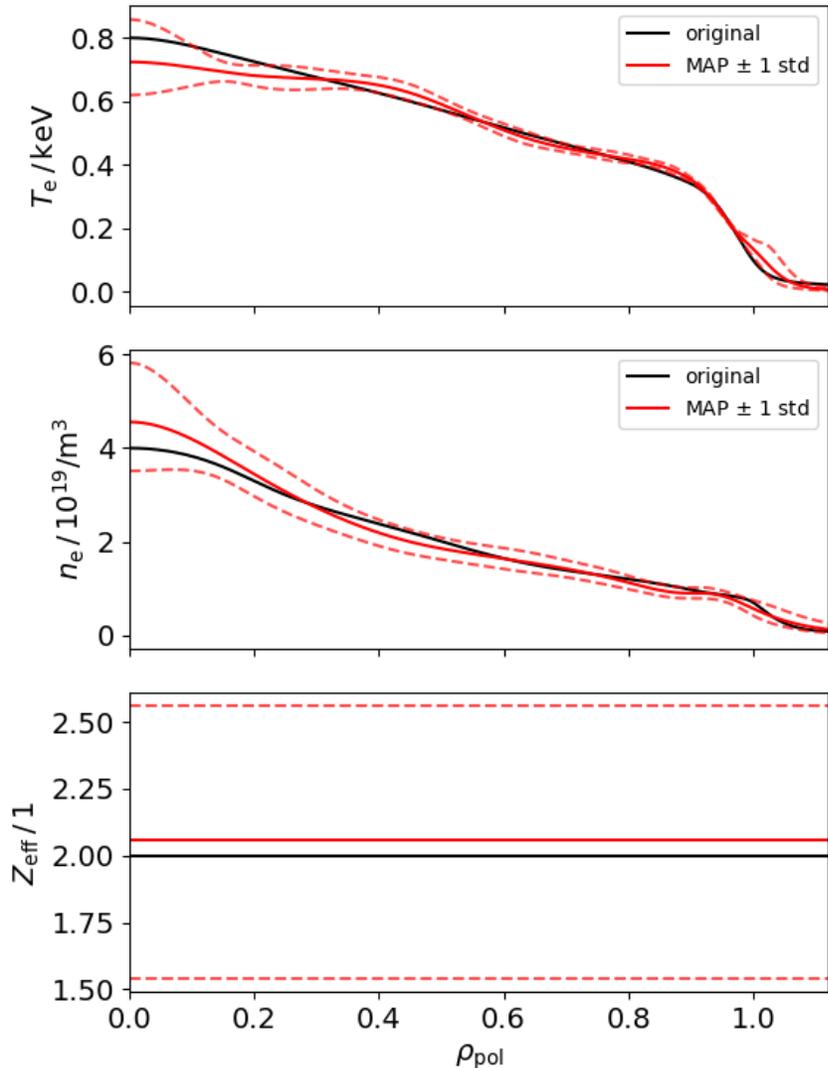
SXR only



- Reference profiles for T_e , n_e & Z_{eff} (**black lines**)
- Get synthetic diagnostic measurements
- Add noise to simulate measurement
- Infer profiles (same model) from synthetic measurements with uncertainty (**red lines**)

Synthetic Test-case (JT-60SA)

SXR only



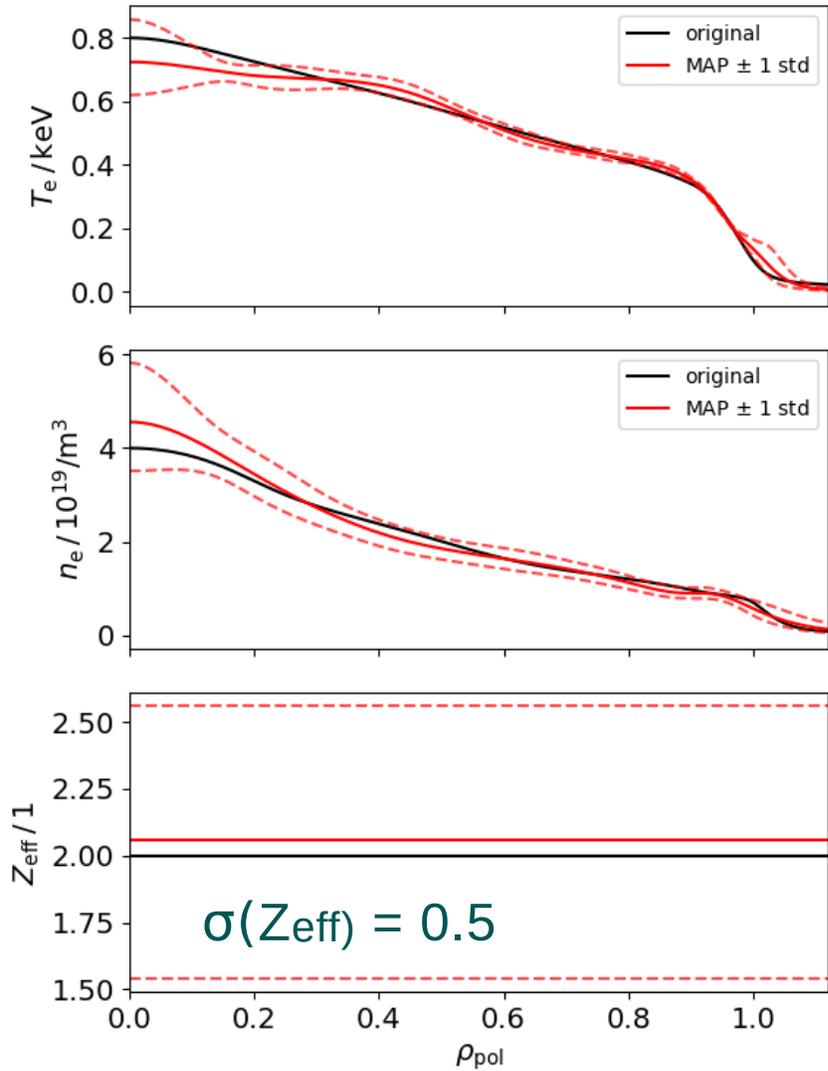
- Reference profiles for T_e , n_e & Z_{eff} (**black lines**)
- Get synthetic diagnostic measurements
- Add noise to simulate measurement
- Infer profiles (same model) from synthetic measurements with uncertainty (**red lines**)
- Profile outside separatrix not well determined!
- Prior leads to smooth decay $\rightarrow 0$ towards the edge
- SXR assumes calibration perfectly known

Synthetic Test-case (JT-60SA)



SXR only

$Z_{\text{eff}} \text{ prior} = 2 \pm 0.5$



IDATION

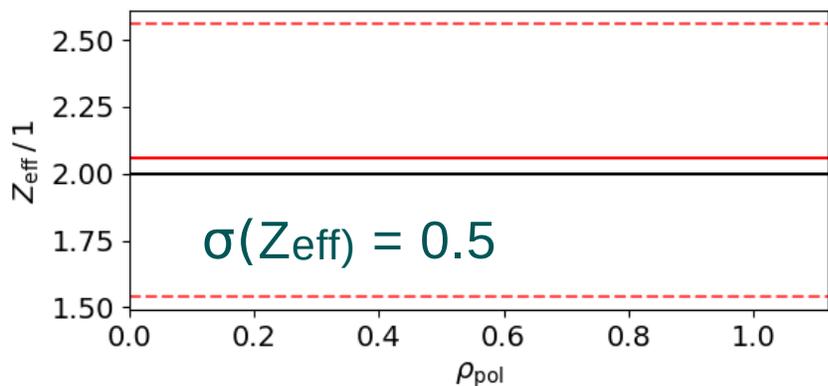
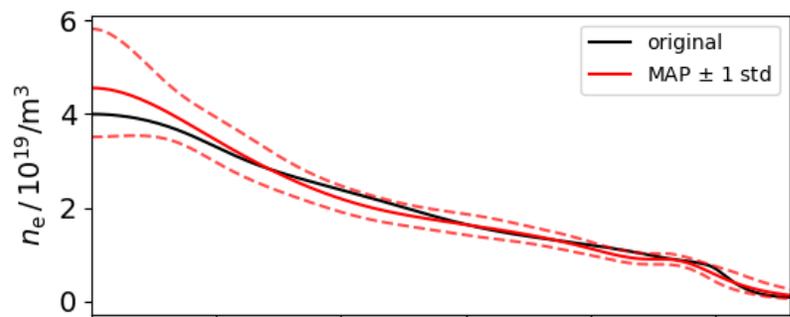
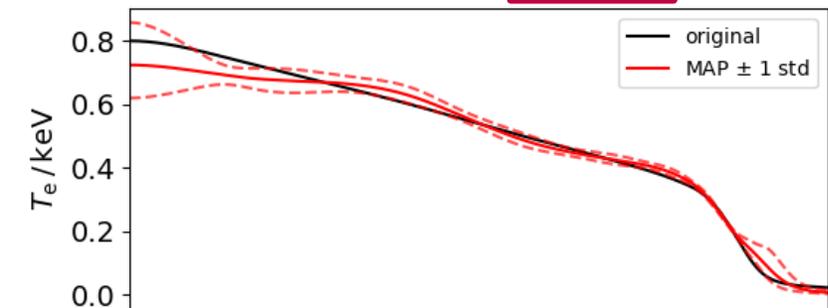
IMAS DATA MAPPING WORKSHOP, CADARACHE, 2026

Synthetic Test-case (JT-60SA)



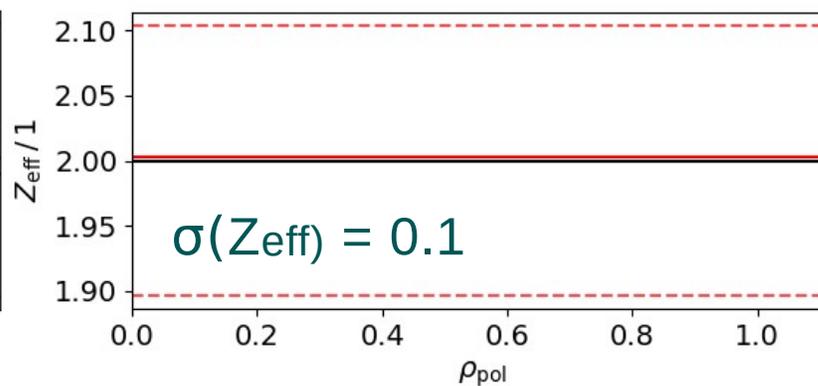
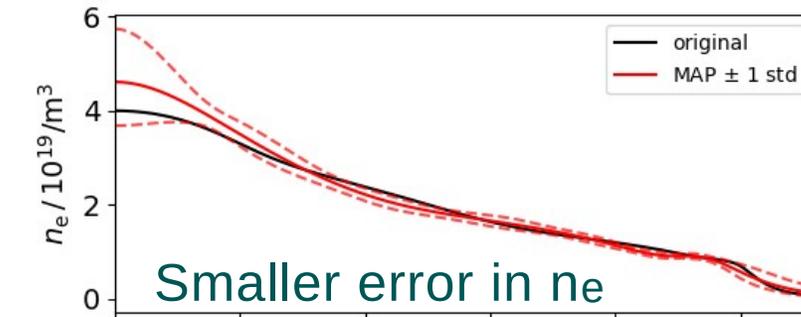
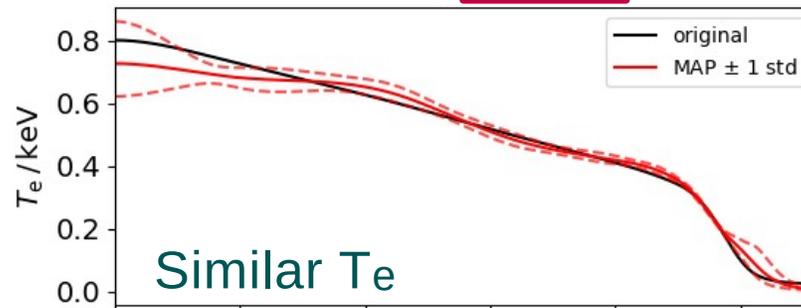
SXR only

Z_{eff} prior = 2 ± 0.5



SXR only

Z_{eff} prior = 2 ± 0.1



Synthetic Test-case (JT-60SA)



SXR only

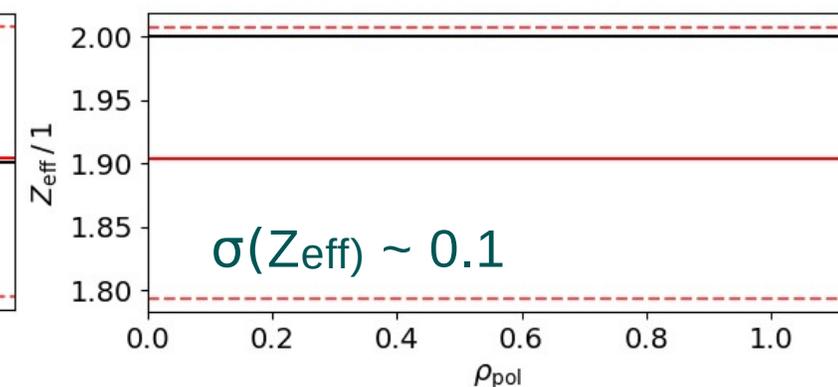
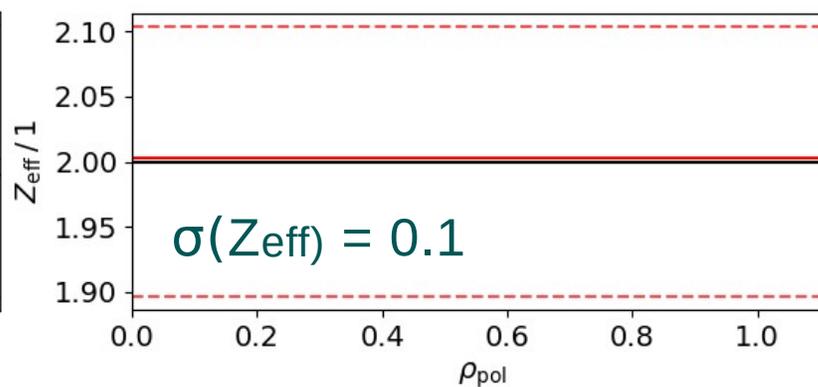
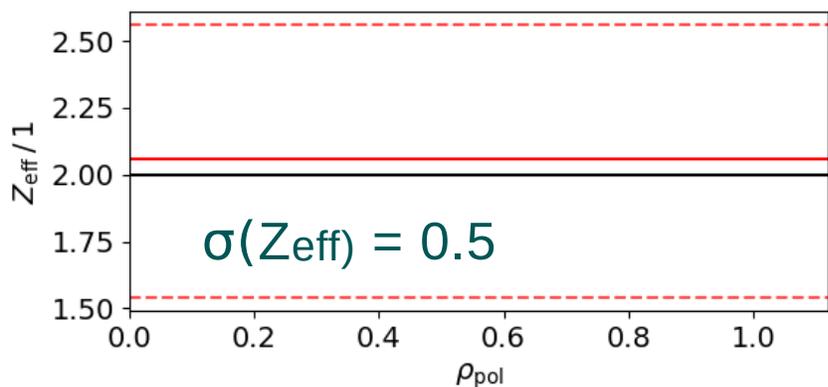
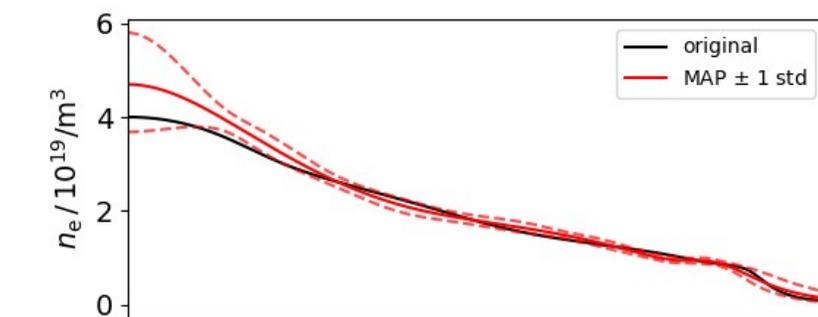
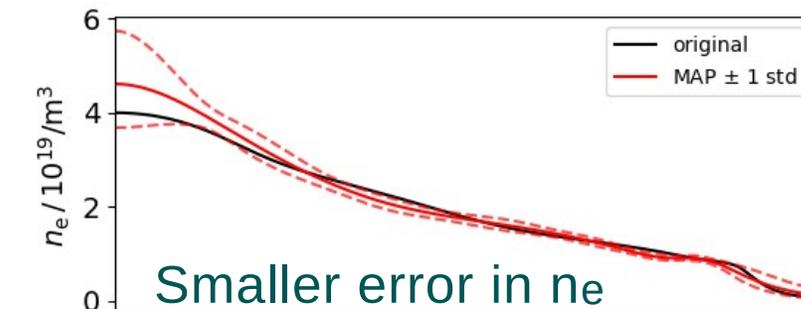
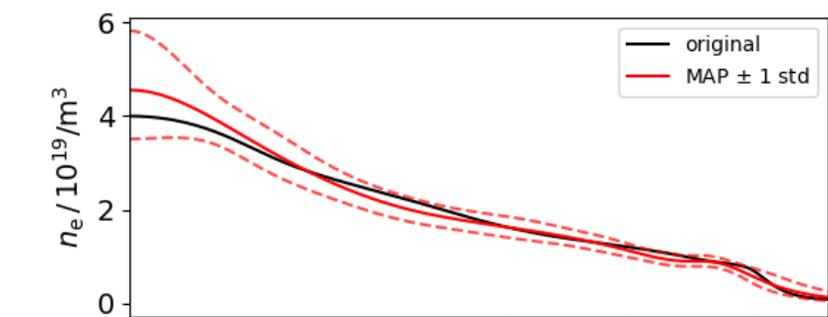
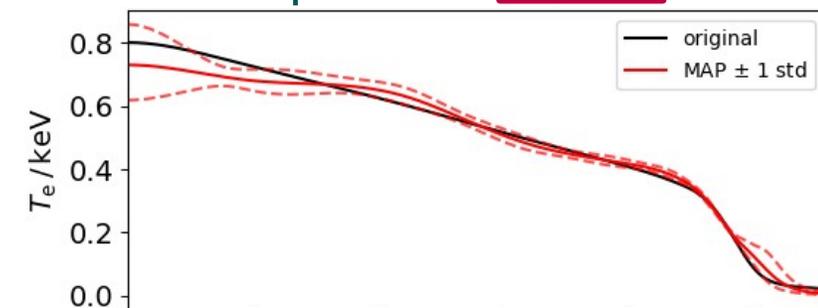
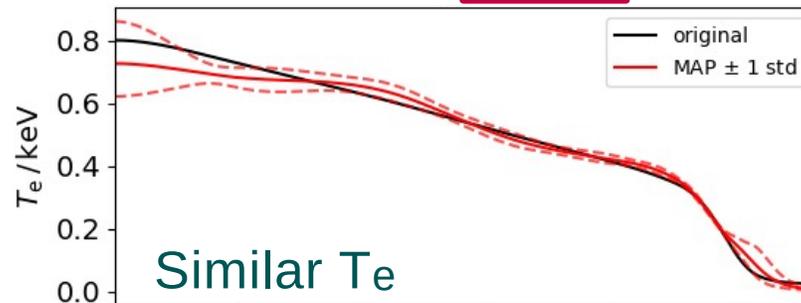
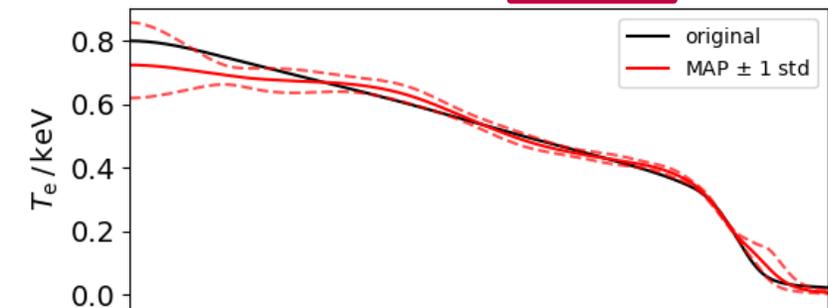
Z_{eff} prior = 2 ± 0.5

SXR only

Z_{eff} prior = 2 ± 0.1

SXR & CO₂

Z_{eff} prior = 2 ± 0.5



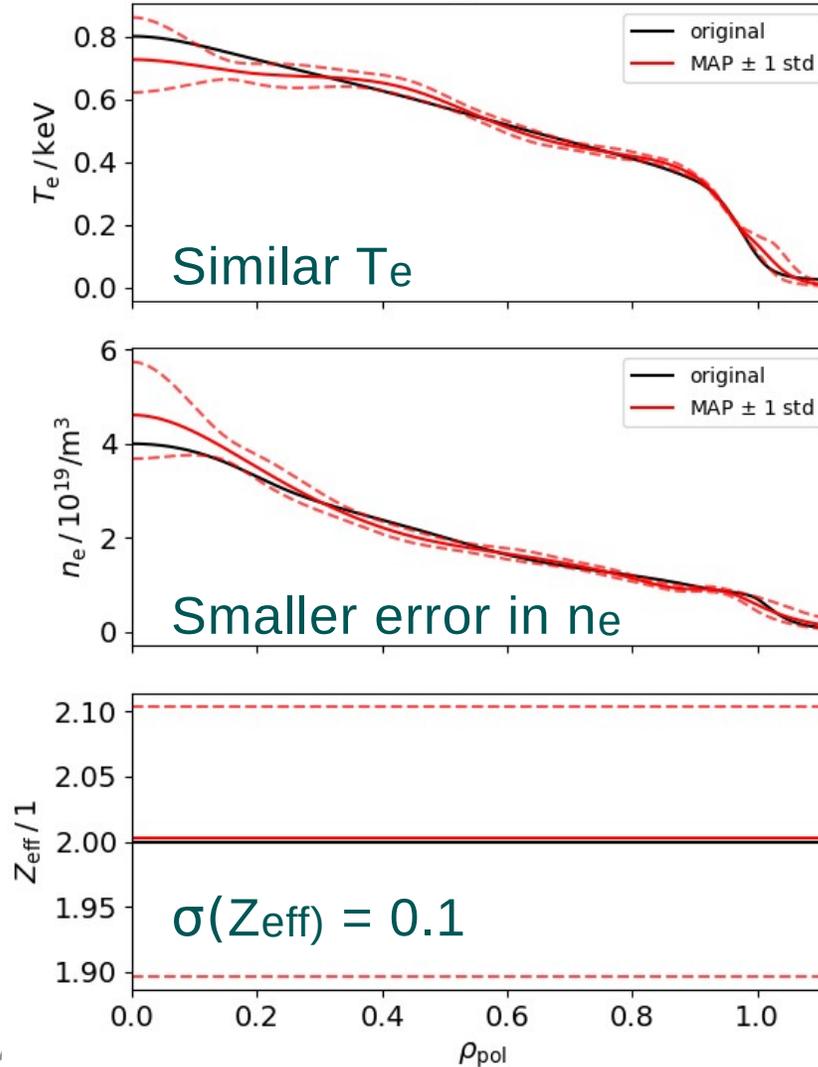
Synthetic Test-case (JT-60SA)



- Here Z_{eff} is guessed, VIS will help
- SXR Calibration mostly determined when combined with CO2
SXR $\sim n_e^2 \times Z_{\text{eff}}$
- Relative calibration only as good as Z_{eff} guess
- T_e is robust

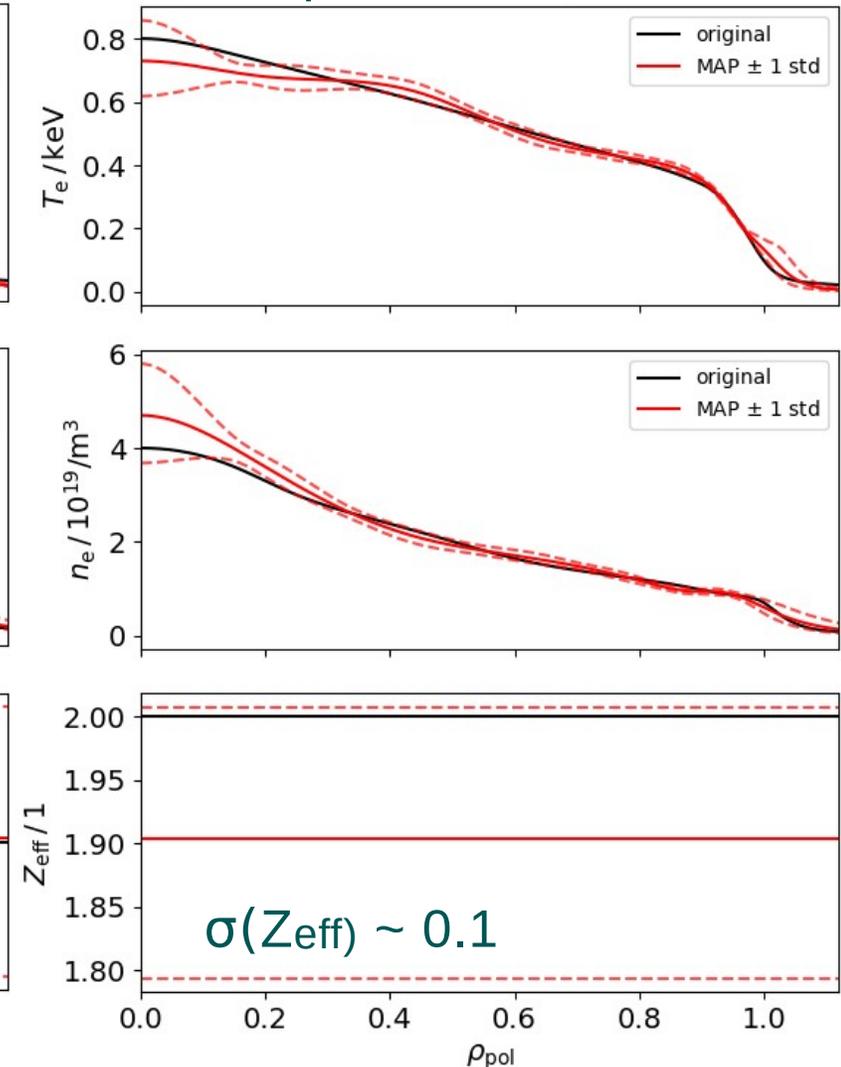
SXR only

Z_{eff} prior = 2 ± 0.1



SXR & CO2

Z_{eff} prior = 2 ± 0.5



IDS example for SXR

Nested data structures (dictionaries) with name-value pairs

https://sharepoint.iter.org/departments/POP/CM/IMDesign/Data%20Model/sphinx/4.0/generated/ids/soft_x_rays.html

- `ids_properties`
- `channel`
 - `name`
 - `description`
 - `detector`
 - `etendue`
 - ...

Access in Python as attribute, e.g.
`sxr_ids.channel[0].line_of_sight.first_point.r`

Thomson Scattering at W7X

- Signal sensitive to
 - Laser path: all mirrors in and around the machine
 - Geometry of LOS: all lenses, mirrors and polychromators
 - Sensitivity of detector → not homogeneous!

- Current Solution:

High dimensional fit via auto-encoder assuming “true” profile shape to get corrections based on the distribution of data in a single pulse.

See https://pinboard.euro-fusion.org/repository/pinboard/EFDA-JET/conference/120698_2025_lapd_fuchert.pdf
(Invited 2 talk at 1st International Symposium on Laser-Aided Plasma Diagnostics)