Bayesian Inference for the LHD Experiment Data

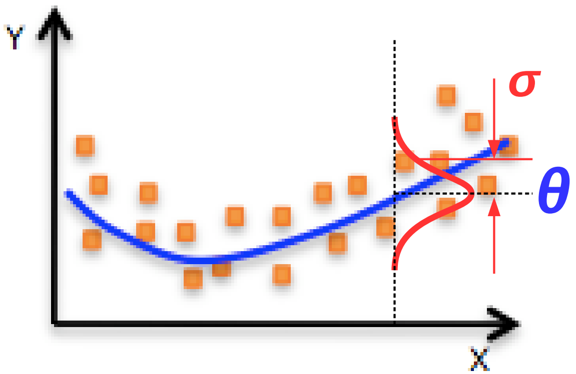
Keisuke Fujii

Bayes rule

Posterior
Full knowledge of θ
Incl. mean and
standard deviation

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$$

Bayes rule



Likelihood
How data (or noise) behaves.
e.g.
Gaussian with mean θ

Posterior
Full knowledge of θ
Incl. mean and standard deviation

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$$

Bayes rule

Posterior
Full knowledge of θ
Incl. mean and
standard deviation

Likelihood
How data (or noise)
behaves.
e.g.
Gaussian with mean θ

Prior
Our assumption on data.

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$$

Opt. Eng. Lab.

Bayes rule

Posterior
Full knowledge of θ
Incl. mean and
standard deviation

Likelihood
How data (or noise) behaves.
e.g.
Gaussian with mean θ

Prior
Our assumption on data.

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$$

Probabilistic modeling

- Quantify what we assume.

Advantage

- Uncertainty quantification
- Assumption selection (model selection)

Opt. Eng. Lab.

Outline

- Brief introduction
- Evaluation of fractional abundance data for W
  - Avoiding over and under fitting

    -model selection-
- Evaluation of systematic noise of

  LHD Thomson scattering system.
- Summary

Outline

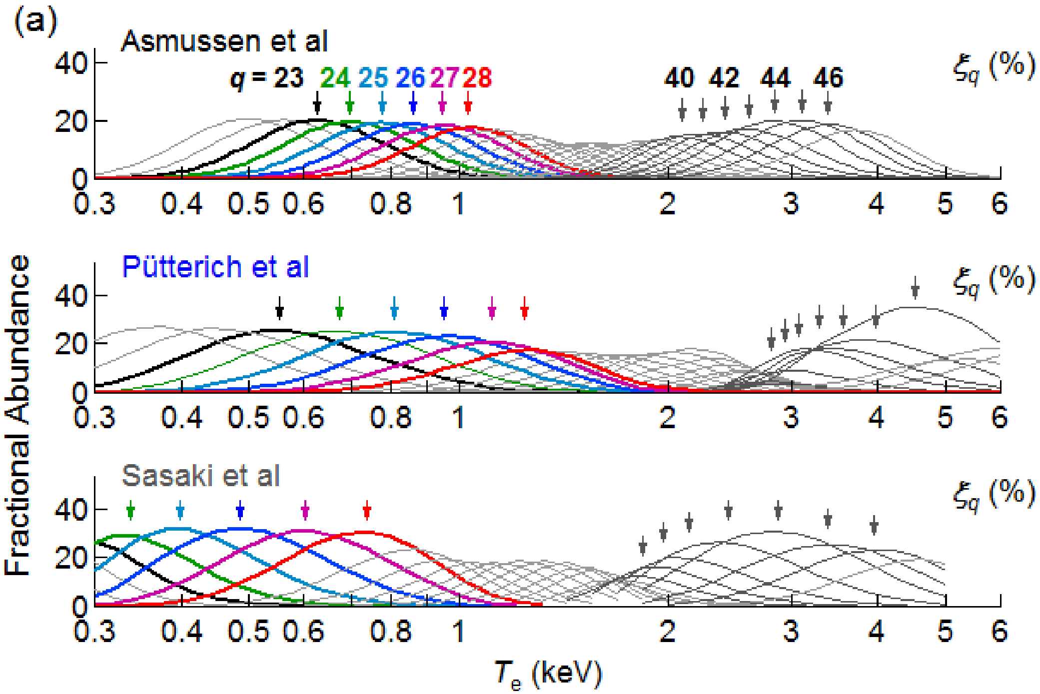
- Brief introduction
- Evaluation of fractional abundance data for W
  - Avoiding over and under fitting
    -model selection-
- Evaluation of systematic noise of
  LHD Thomson scattering system.
- Summary

Opt. Eng. Lab.

Background:
Fractional abundance of W

Measured intensity — Emission coef. — Fractional abundance

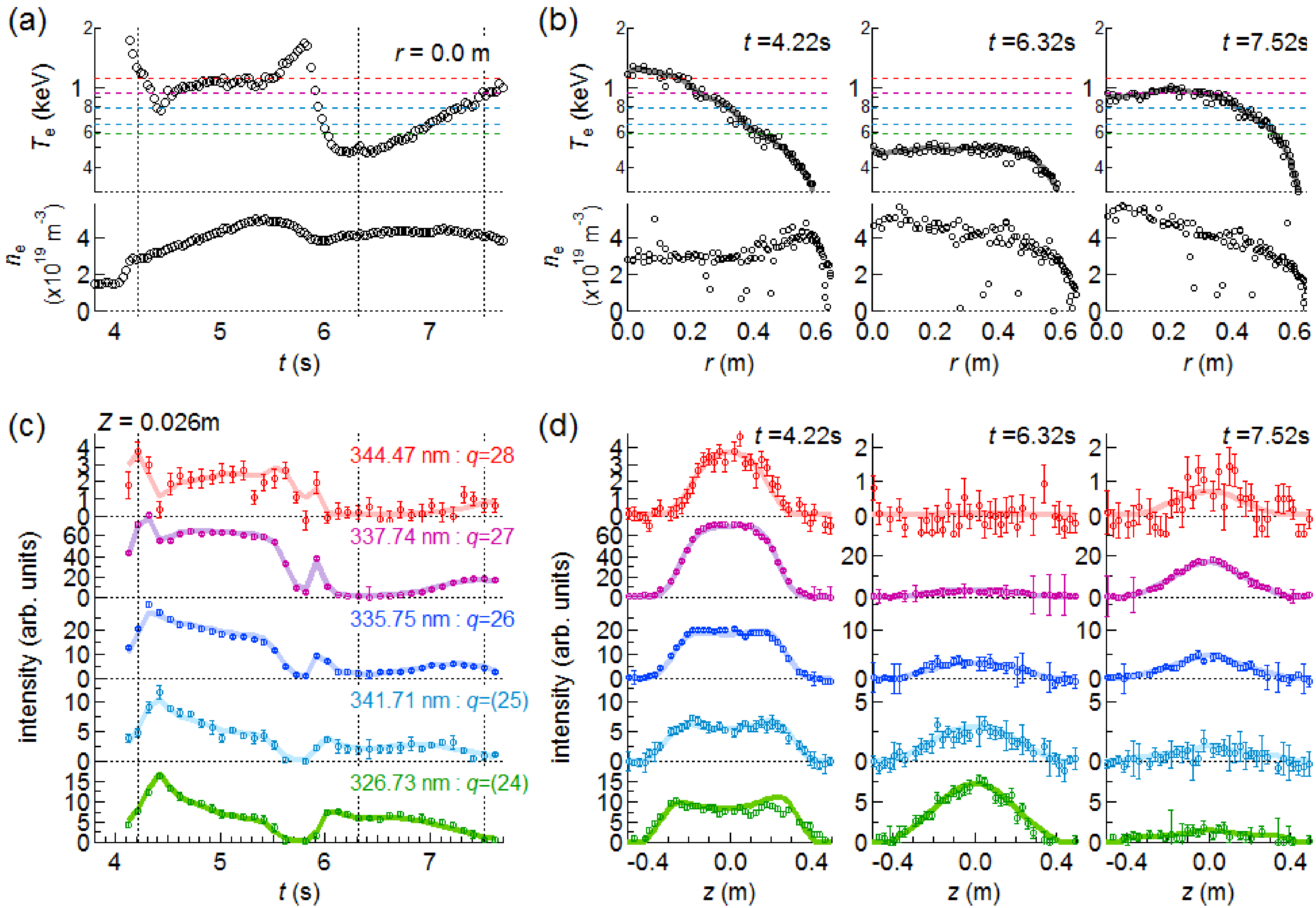
$$d^{\mathrm{model}} = \int_{\mathrm{LOS}} \epsilon_i(n_\mathrm{e}, T_\mathrm{e}) n_\mathrm{e} \xi_q(T_\mathrm{e}) n_\mathrm{W} \mathrm{d}x$$



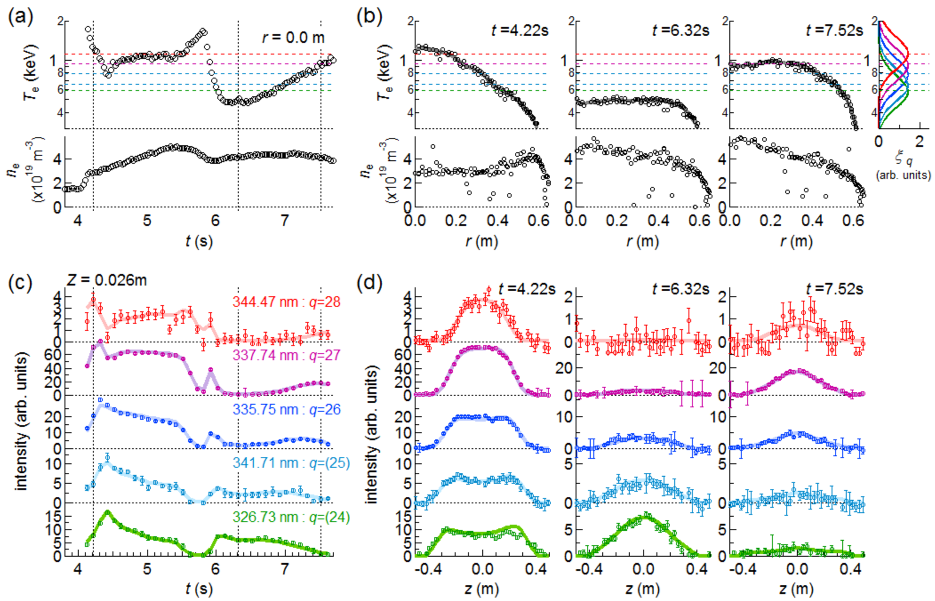
Fractional abundance of W is an important data essential to the tungsten transport diagnostics.

Significant disagreement has been reported among the results by different groups, in particular $q < 30$.

Measurement



Objective:
Inference of ξ_q from the experimental data



unknown

$$d^{\text{model}} = \int_{\text{LOS}} \epsilon_i(n_e, T_e)\, n_e\, \xi_q(T_e)\, n_W \, dx$$

known — known

Parameters:		Assumptions
$\varepsilon_i n_e$	Emission rate per 1 ground state ion.	Independent of n_e and T_e
ξ_q	Fractional abundance	Smooth function of T_e
n_W	Total tungsten density distribution	Smooth function of r and t

Opt. Eng. Lab.

How much we should assume

Parameters:		Assumptions
$\varepsilon_i n_e$	Emission rate per 1 ground state ion.	Independent of n_e and T_e
ξ_q	Fractional abundance	Smooth function of T_e
n_W	Total tungsten density distribution	Smooth function of r and t

How smooth profile we should assume?

Too strong assumption.

Too weak assumption.

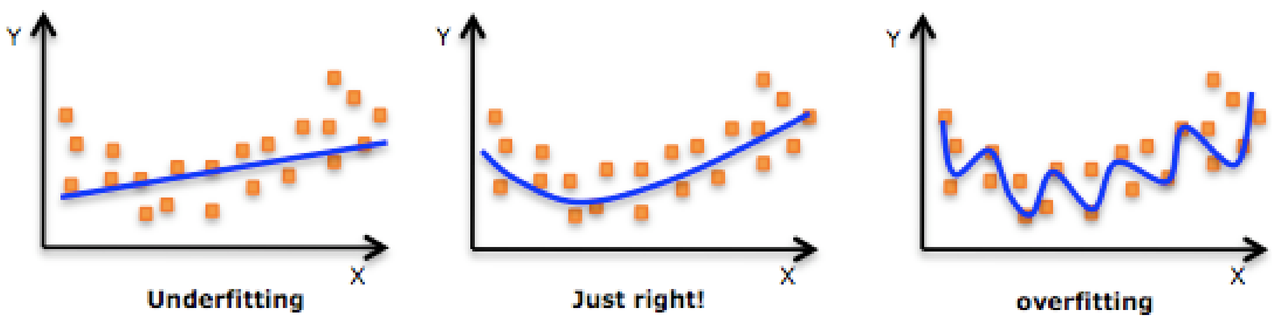


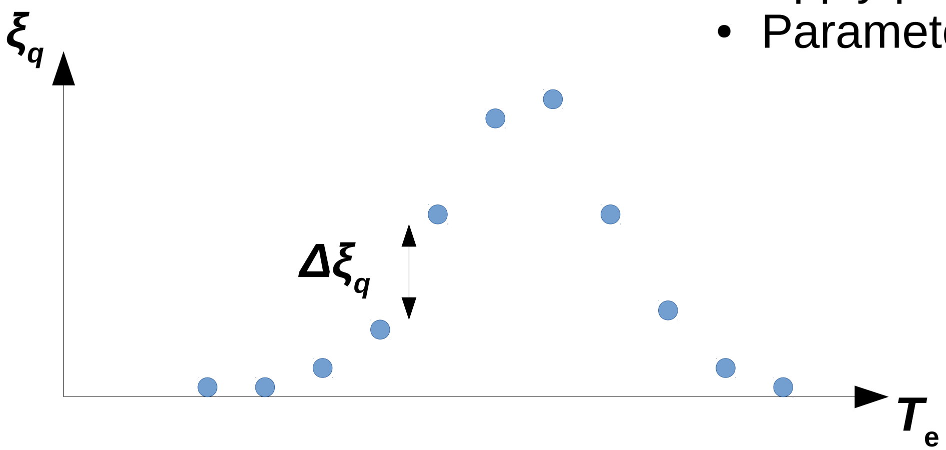
Image from http://pingax.com/regularization-implementation-r/

Introduce hyperparameter

Parameters:		Assumptions
ξ_q	Fractional abundance	Smooth function of T_e
n_W	Total tungsten density distribution	Smooth function of r and t

It is necessary to quantify the smoothness.

- Discretize the profile into finite number points
- Apply prior distribution for the difference $\Delta\xi$
- Parameterize the prior by hyperparameter



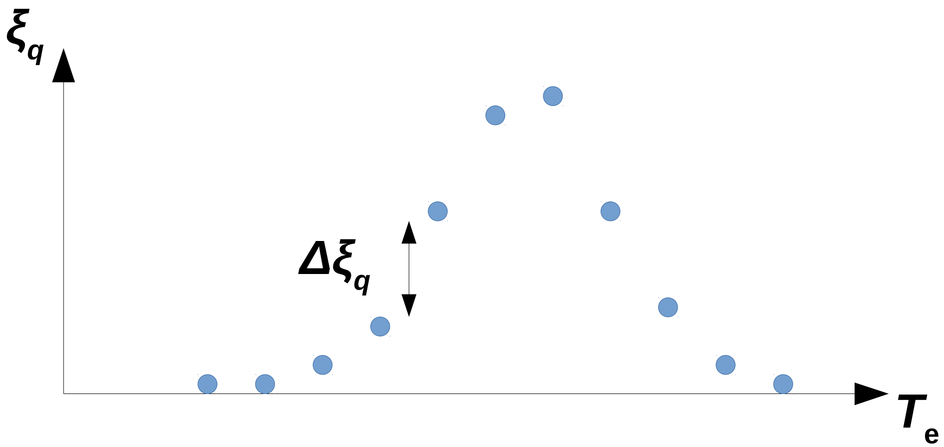
Prior (assumption)

$$p(\xi_q|\lambda_\xi) = \mathcal{N}(\Delta\xi_q|0, \lambda_\xi)$$

How strong we assume.

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$$

Choose how much we should assume from data



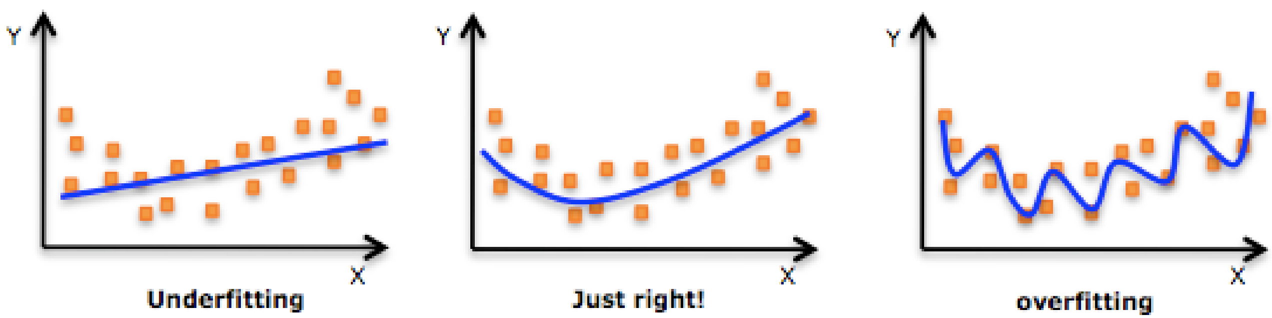
Prior (assumption)

$$p(\xi_q|\lambda_\xi) = \mathcal{N}(\Delta\xi_q|0, \lambda_\xi)$$

How strong we assume.

Too strong assumption.

Too weak assumption.



small λ ← λ → Large λ

$$p(\xi_q|\lambda_\xi) = \mathcal{N}(\Delta\xi_q|0, \lambda_\xi)$$

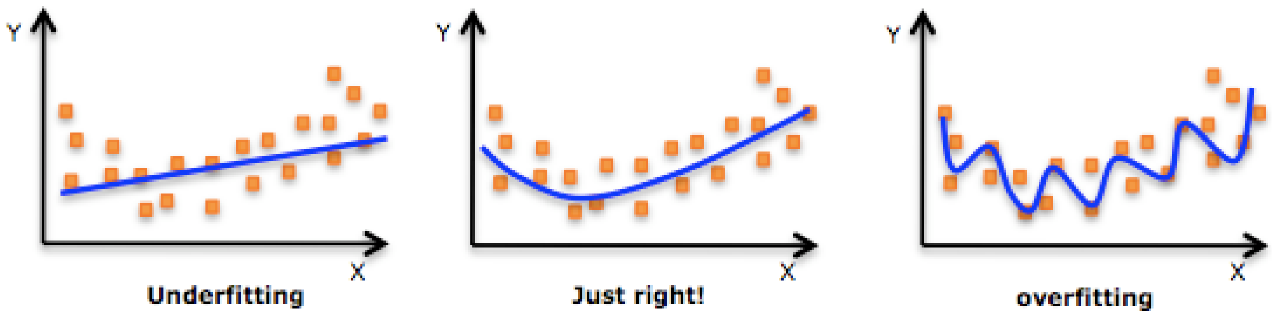
Image from http://pingax.com/regularization-implementation-r/

Opt. Eng. Lab.

Choose how much we should assume from data.

Too strong assumption.

Too weak assumption.



small λ ← λ → Large λ

$$p(\xi_q|\lambda_\xi) = \mathcal{N}(\Delta\xi_q|0, \lambda_\xi)$$

How should we remove the dependence on λ_ξ?

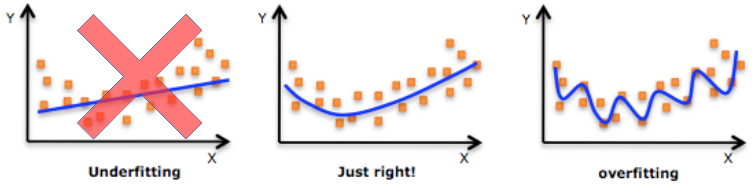
Marginalization (apply prior for λ_ξ and integrate out)

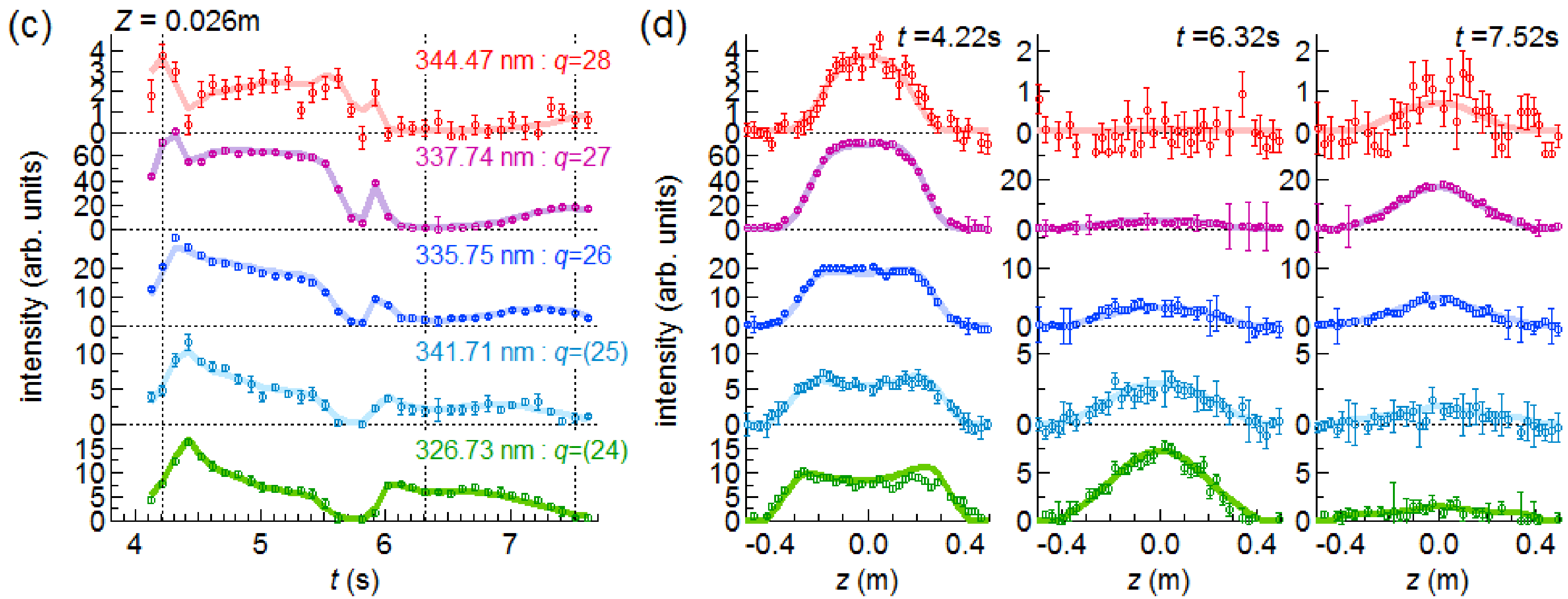
$$p(\xi_q|d) \propto \int p(d|\xi_q)p(\xi_q|\lambda_\xi)p(\lambda_\xi)\mathrm{d}\lambda_\xi$$

This avoids the under and over-fitting.

Opt. Eng. Lab.

Result





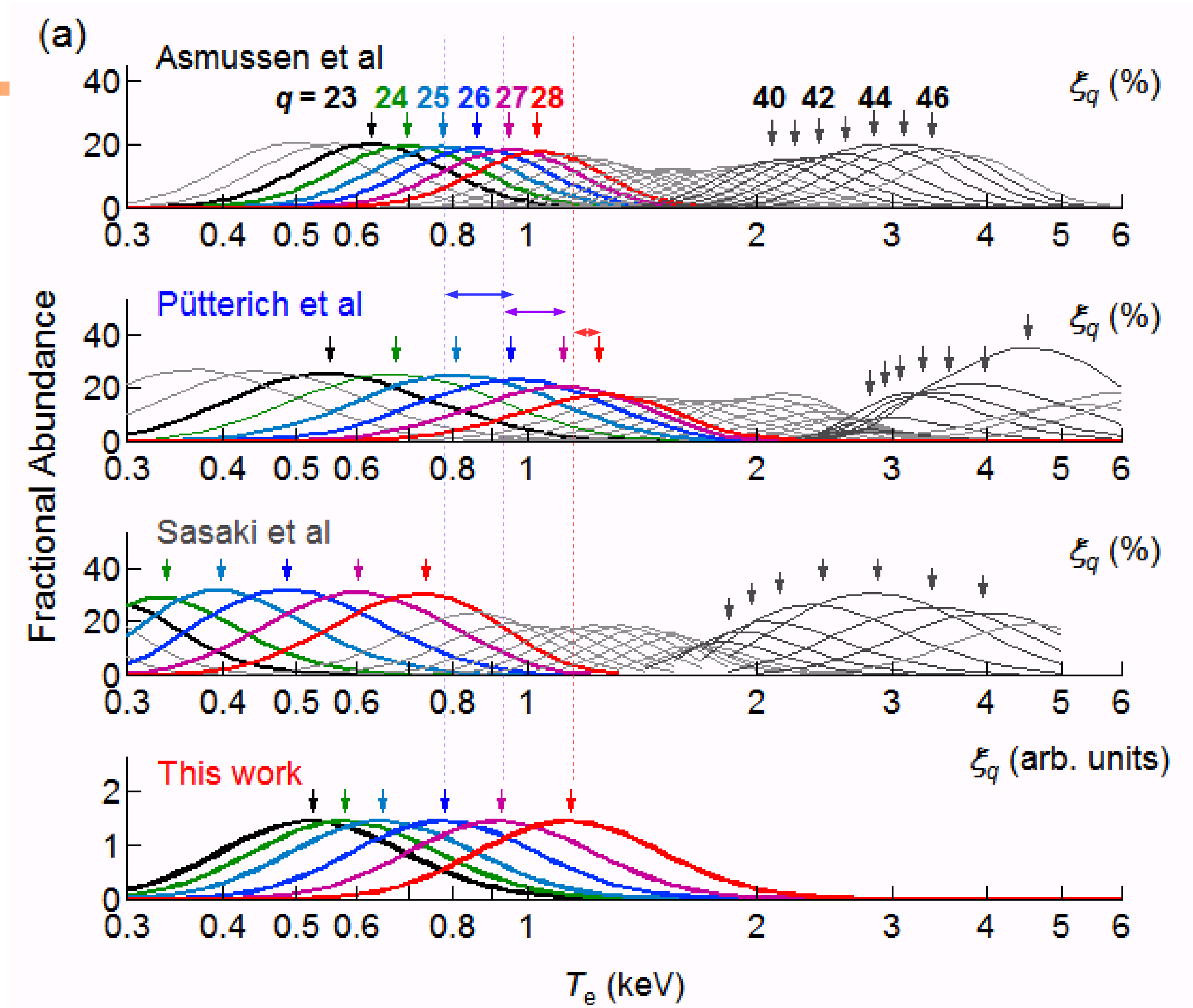
Our model well represents the measured data.

No under-fitting

Result

Inferred ξ_q profiles are smooth enough.

No over-fitting



Our results are close to those by Putterich et al,
but our peak positions locate at the smaller T_e side.

Our results may be used as benchmark for future theoretical works.

Outline

- Brief introduction
- Evaluation of fractional abundance data for W
  - Avoiding over and under fitting

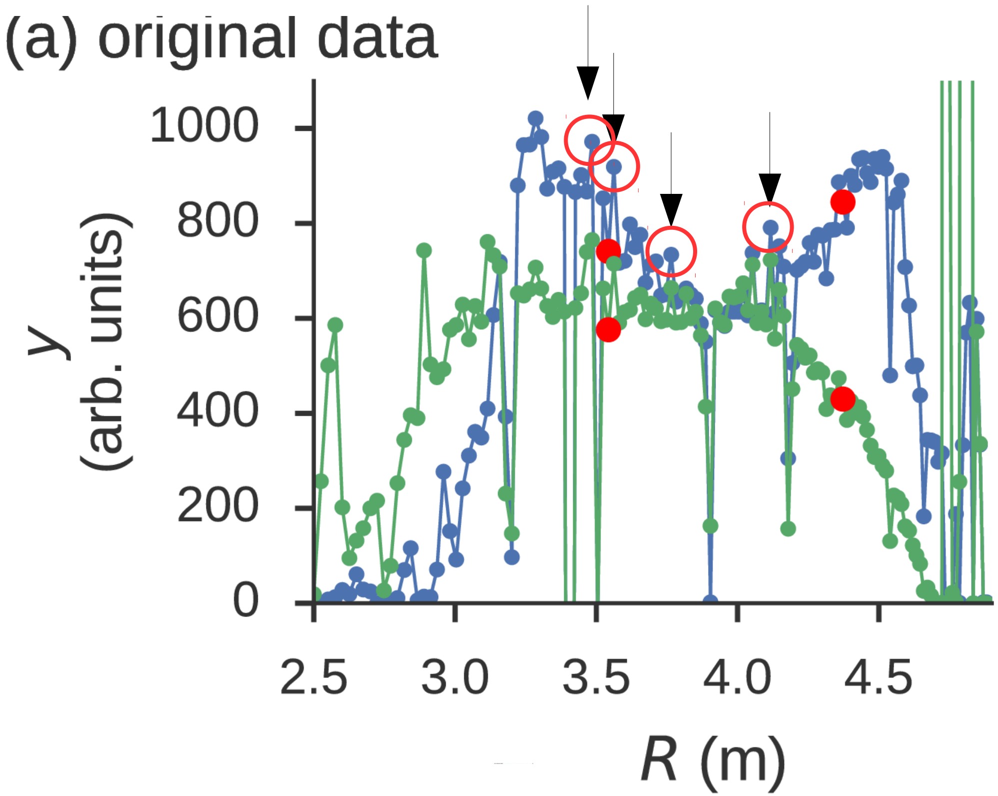
    -model selection-
- Evaluation of systematic noise of

  LHD-TS diagnostic system.

  arXiv:1607.05380
- Summary

Opt. Eng. Lab.

Systematic noise in LHD-TS system



Significant dependent noise due to mis-calibration.

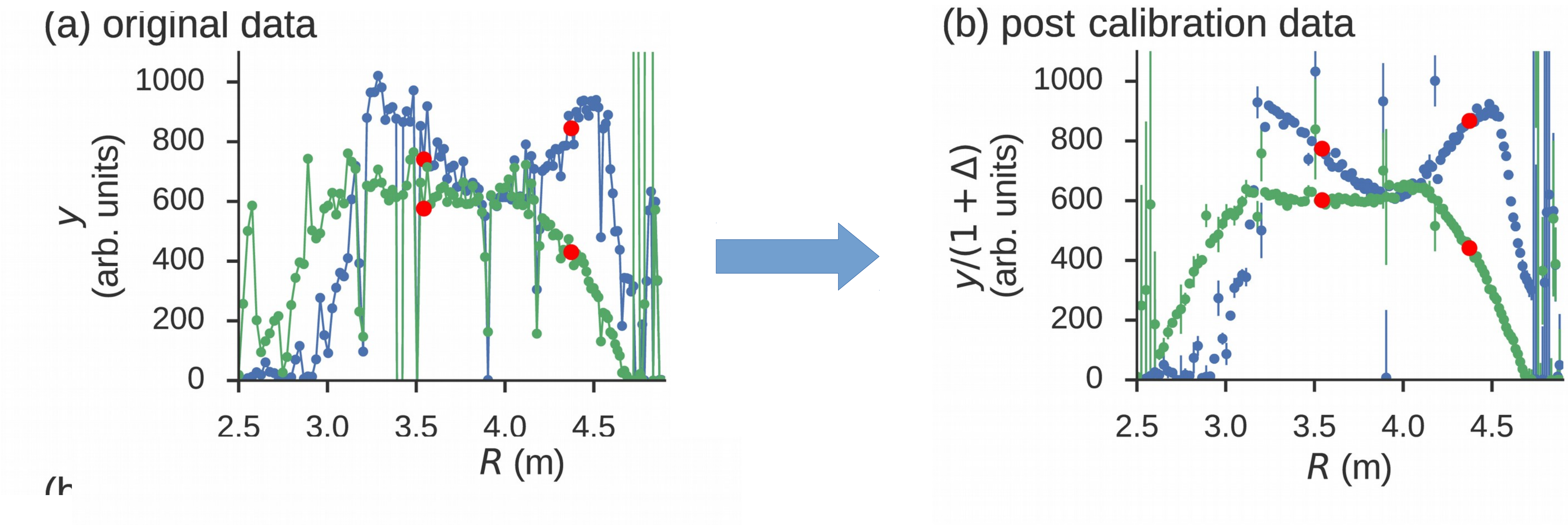
Random noise： Varies randomly.
Thermal noise, shot noise

Can be analyzed by legacy statistic.

Systematic noise： Has large correlation.
Inaccurate calibration, model, ...

Bayesian statistics

Objective:
Machine learning of the mis-calibration noise



Systematic noise model

Current calibration factor for channel *i*

$$R_i = R_i^0(1 + \Delta_i.)$$

True calibration factor for channel *i*

Mis-calibration noise (to be estimated)

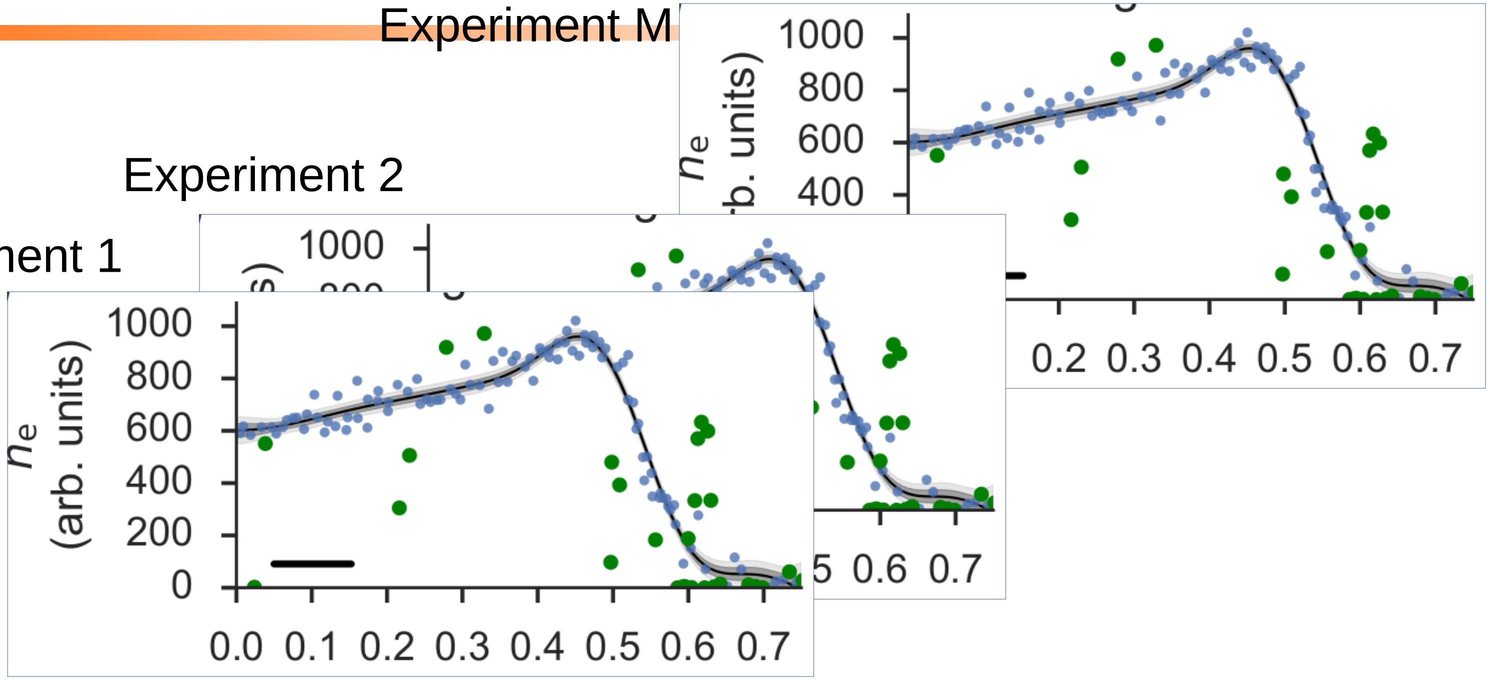
Opt. Eng. Lab.

Probabilistic modeling

Experiment M

Experiment 2

Experiment 1



data — true values — Random noise — systematic noise

$$\mathbf{y}_1 = \mathbf{f}_1 + \mathbf{n}_1 + \mathbf{n}_\Delta$$

$$\mathbf{y}_2 = \mathbf{f}_2 + \mathbf{n}_2 + \mathbf{n}_\Delta$$

$$\vdots$$

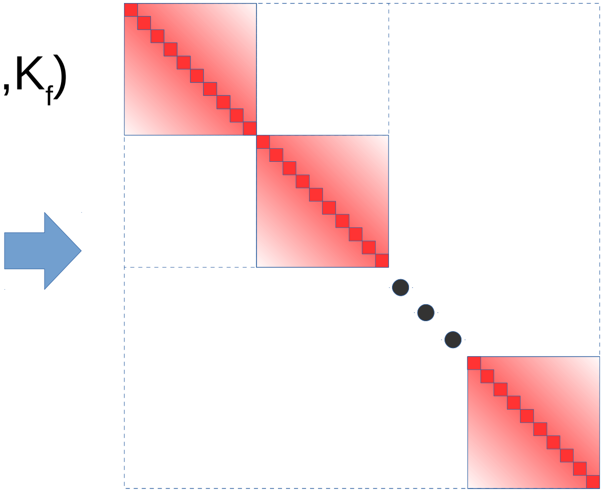
$$\mathbf{y}_M = \mathbf{f}_M + \mathbf{n}_M + \mathbf{n}_\Delta$$

Gaussian Process for multiple frame data

$$
\left.
\begin{aligned}
\mathbf{y}_1 &= \mathbf{f}_1 + \mathbf{n}_1 + \mathbf{n}_\Delta \\
\mathbf{y}_2 &= \mathbf{f}_2 + \mathbf{n}_2 + \mathbf{n}_\Delta \\
&\vdots \\
\mathbf{y}_M &= \mathbf{f}_M + \mathbf{n}_M + \mathbf{n}_\Delta
\end{aligned}
\right\}
$$

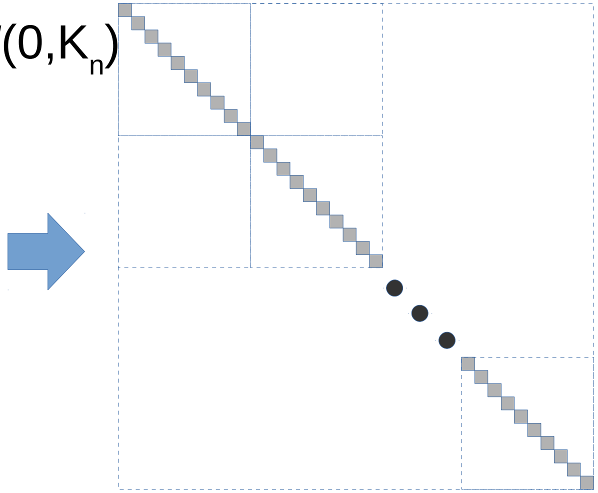
prior
$p(\mathbf{f}) = N(0, K_f)$

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_M \end{bmatrix}$$



prior
$p(\mathbf{n}) = N(0, K_n)$

$$\begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_M \end{bmatrix}$$



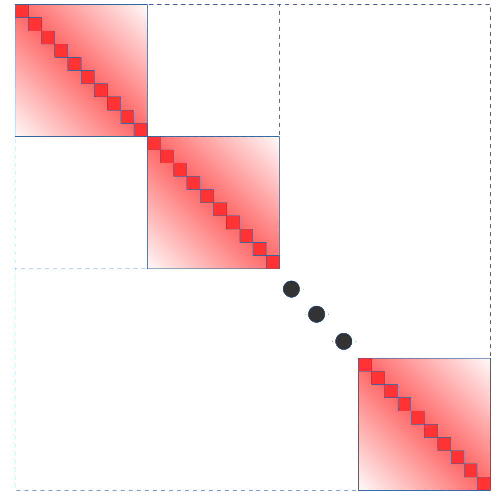
Opt. Eng. Lab.

Gaussian Process for multiple frame data

$$
\left.
\begin{aligned}
\mathbf{y}_1 &= \mathbf{f}_1 + \mathbf{n}_1 + \mathbf{n}_\Delta \\
\mathbf{y}_2 &= \mathbf{f}_2 + \mathbf{n}_2 + \mathbf{n}_\Delta \\
&\vdots \\
\mathbf{y}_M &= \mathbf{f}_M + \mathbf{n}_M + \mathbf{n}_\Delta
\end{aligned}
\right\}
$$

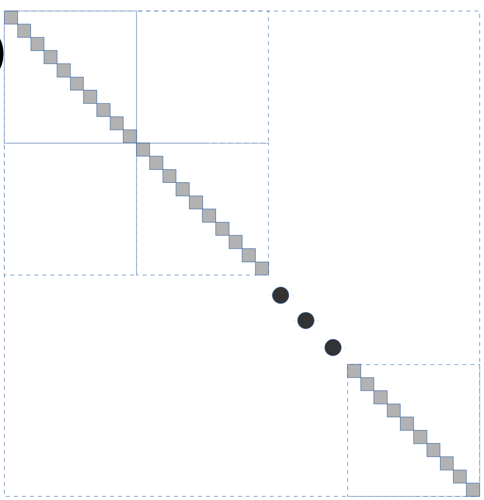
prior
$p(\mathbf{f}) = N(0, K_f)$

$$\begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_M \end{bmatrix} \Rightarrow$$



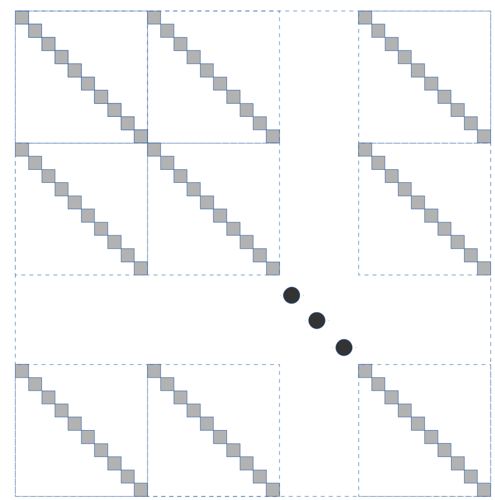
prior
$p(\mathbf{n}) = N(0, K_n)$

$$\begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_M \end{bmatrix} \Rightarrow$$



prior
$p(\mathbf{\Delta}) = N(0, K_\Delta)$

$$\begin{bmatrix} \mathbf{n}_\Delta \\ \mathbf{n}_\Delta \\ \vdots \\ \mathbf{n}_\Delta \end{bmatrix} \Rightarrow$$



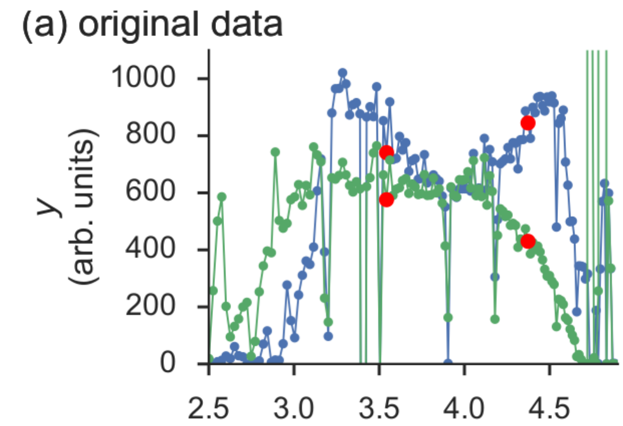
Results

$$R_i = R_i^0(1 + \Delta_i.)$$

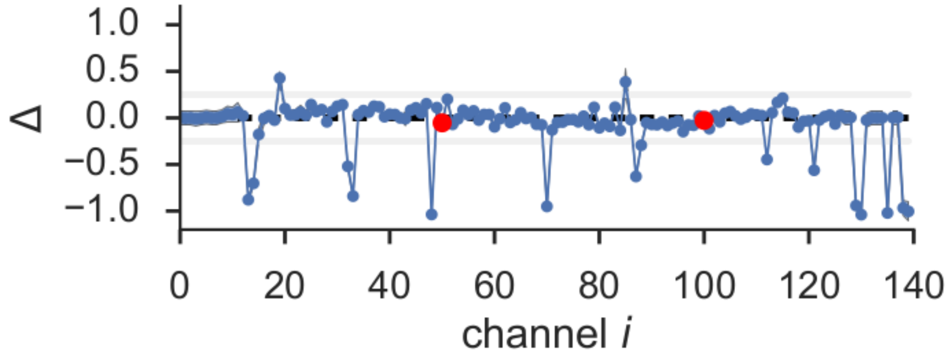
Current calibration factor for channel *i*

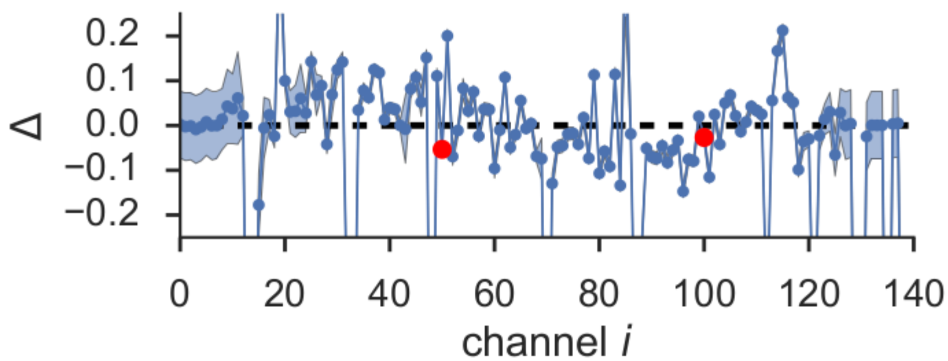
True calibration factor for channel *i*

Mis-calibration noise (to be estimated)

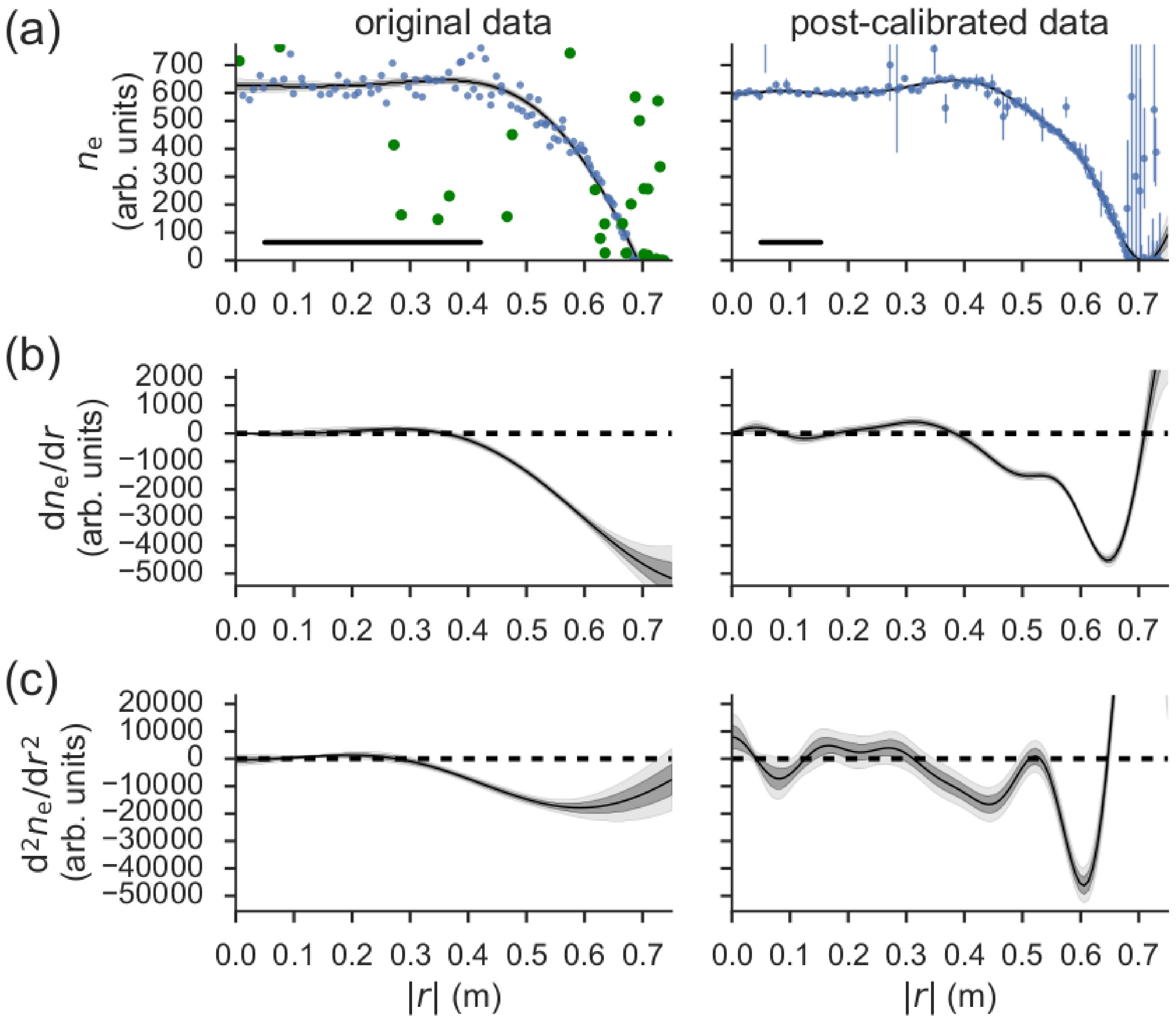


210 frames of the ne observation data by LHD Thomson system.





Application to the derivative inference.



Opt. Eng. Lab.

Outline

- Brief introduction
- Evaluation of fractional abundance data for W
  - Avoiding over and under fitting
    -model selection-
- Evaluation of systematic noise of
  LHD-TS diagnostic system.
- Summary

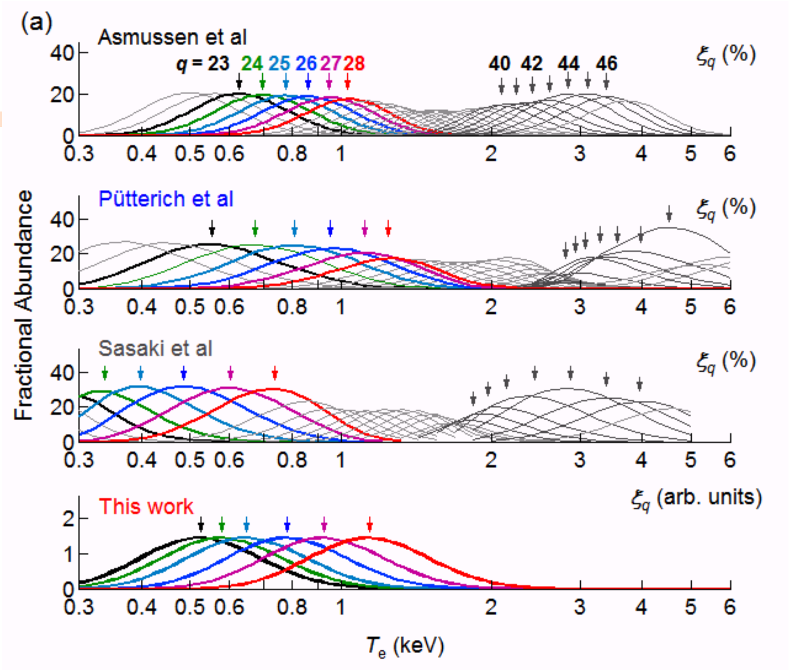
Opt.
Eng. Lab.

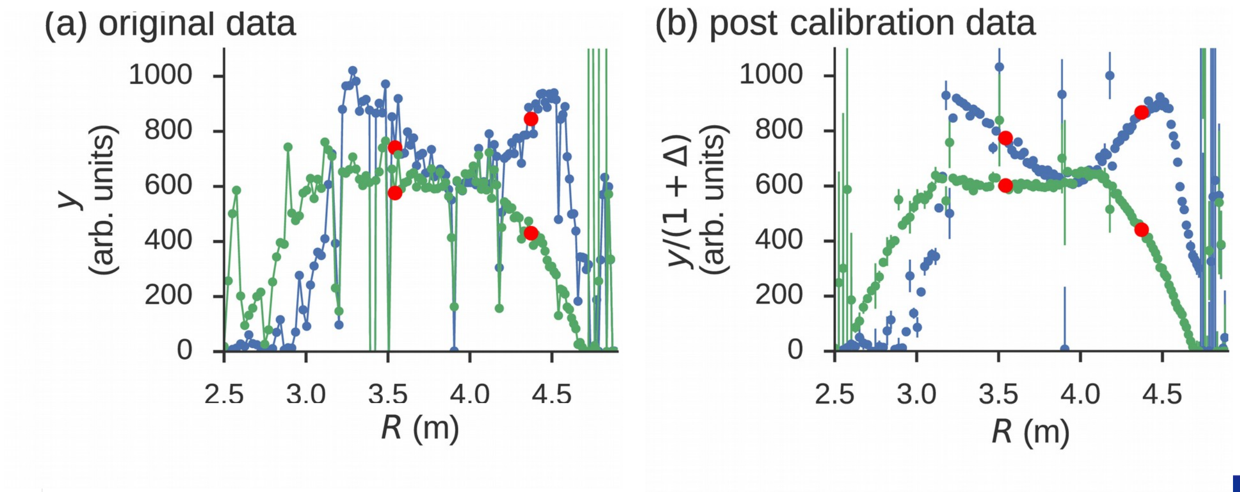
Summary1

We inferred

- the fractional abundance of W ions from LHD experimental data
- The systematic noise for the LHD-TS system

by applying the Bayesian inference.





Summary2

Posterior
Full knowledge of θ
Incl. mean and standard deviation

Likelihood
How data (or noise) behaves.
e.g.
Gaussian with mean θ

Prior
Our assumption on data.

$$p(\theta|d) = \frac{p(d|\theta)p(\theta)}{p(d)}$$

Bayesian statistics states
"the importance of the assumption".

The main challenge in Bayesian statistics is **how we quantify our assumption**. → **Probabilistic modeling**

My message

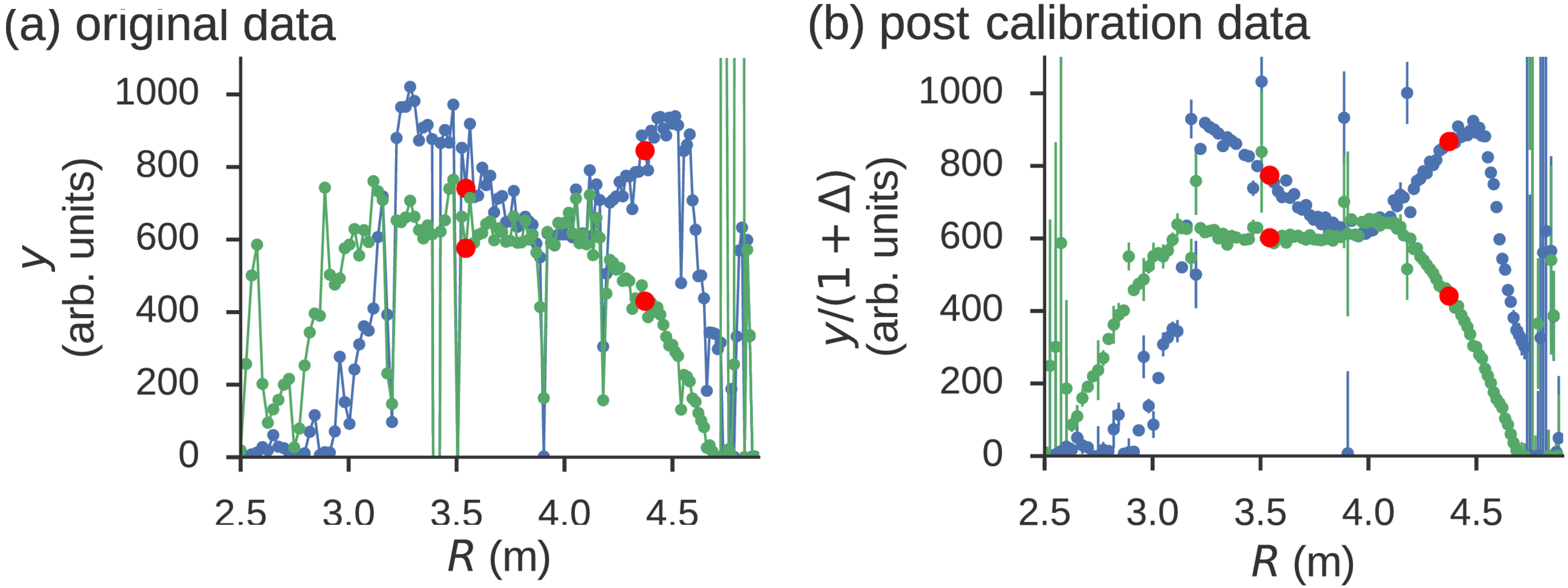
There is **no super-tool** that is used for all the purposes.
We A.M. data unit may need to develop our own statistical models to
- model the theoretical results
- update the data with experimental data

Opt. Eng. Lab.

Summary2

Our second attempt is to infer the systematic noise for LHD-TS system from a large amount of LHD experiment data.
(data-driven science)

Revealed more detailed structure of n_e.



Details 1: Additive approximation

$$R_i = R_i^0(1 + \Delta_i.)$$

True calibration factor for channel *i*

Current calibration factor for channel *i*

Mis-calibration noise (to be estimated)

$$\mathbf{y} = \mathbf{f} + \mathbf{n} + \mathbf{n}_{\Delta}$$

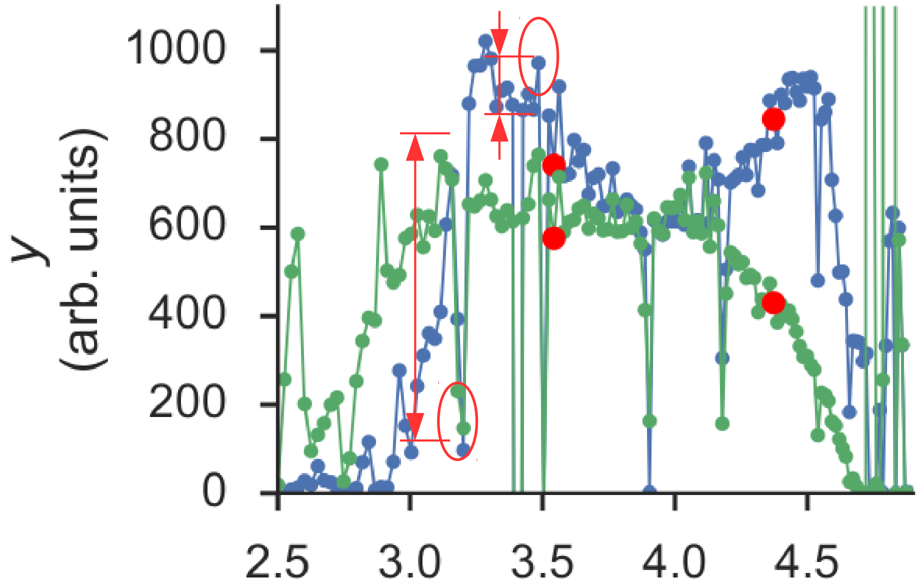
Mis-calibration noise is not additive.

$$\mathbf{y}_j = \mathbf{f}_j + \mathbf{n}_j + \mathbf{f}_j\boldsymbol{\Delta}$$

Additive approximation with iteration.

Opt. Eng. Lab.

Details 2:
Non Gaussian prior



The distribution of **Δ** may not be Gaussian.

We adopt a Cauchy distribution for **Δ**.

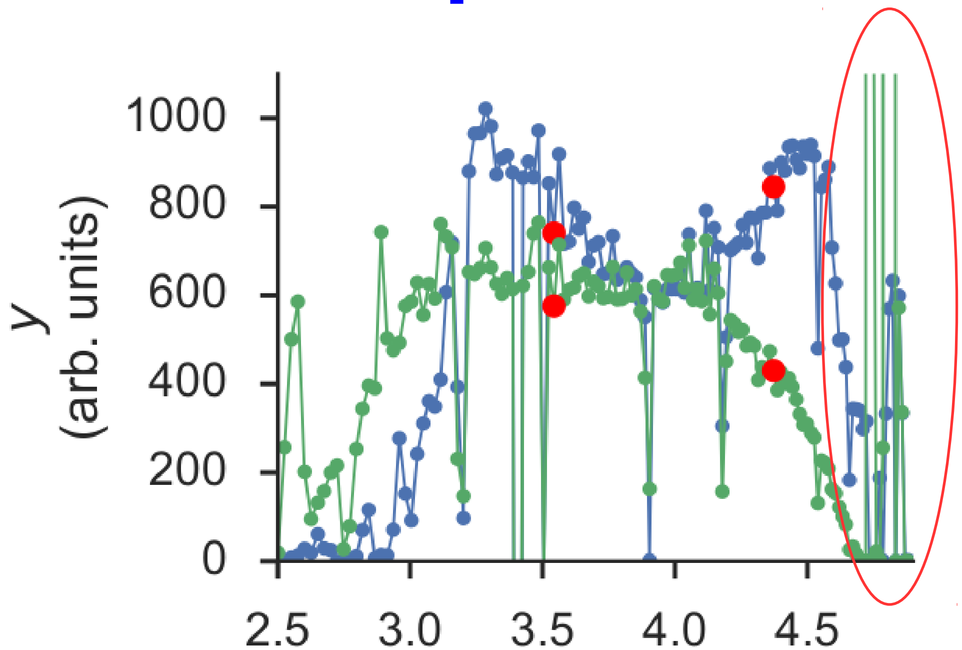
Hierarchical model

$$p(\Delta_i) = \mathcal{N}(0, \sigma_i^2)$$

$$p(\sigma_i^2) = \mathcal{IG}\left(\frac{1}{2}, \frac{\sigma_\Delta^2}{2}\right)$$

Opt. Eng. Lab.

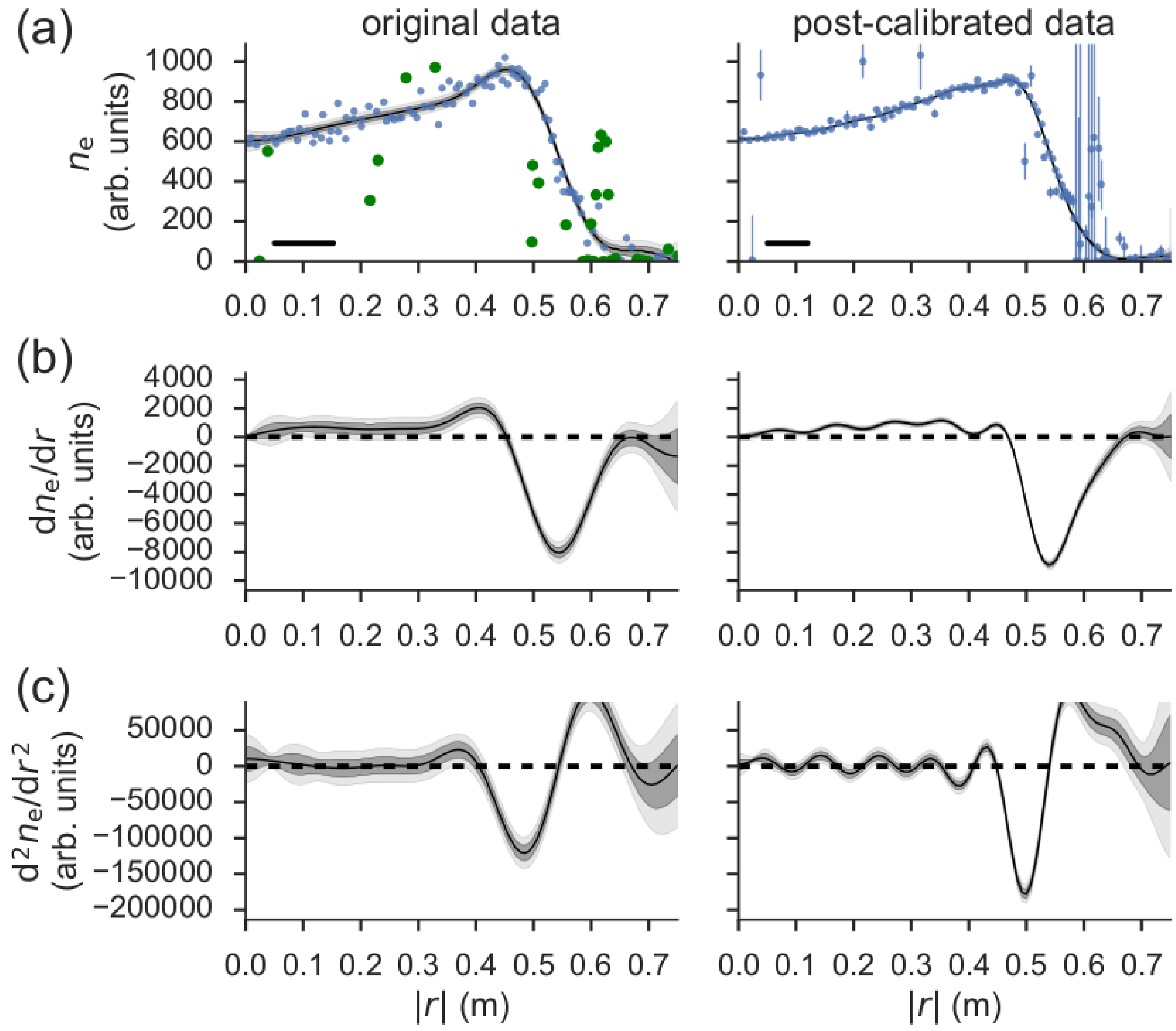
Details 2:
Non Gaussian prior



There are some outliers.
The distribution of **n** may not be Gaussian.

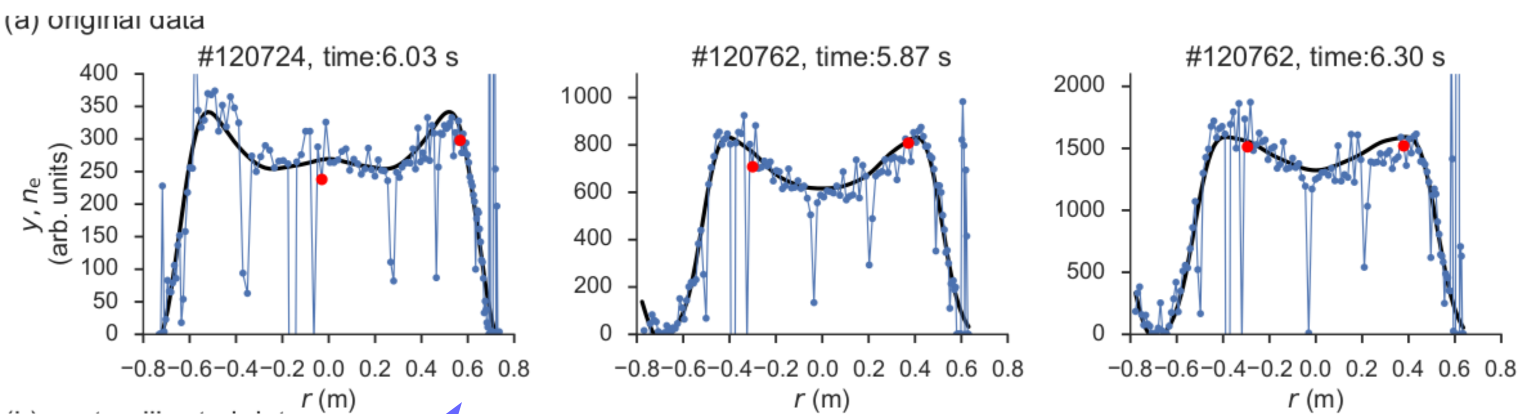
We adopt a Cauchy distribution also for **n**.

Application to the derivative inference.



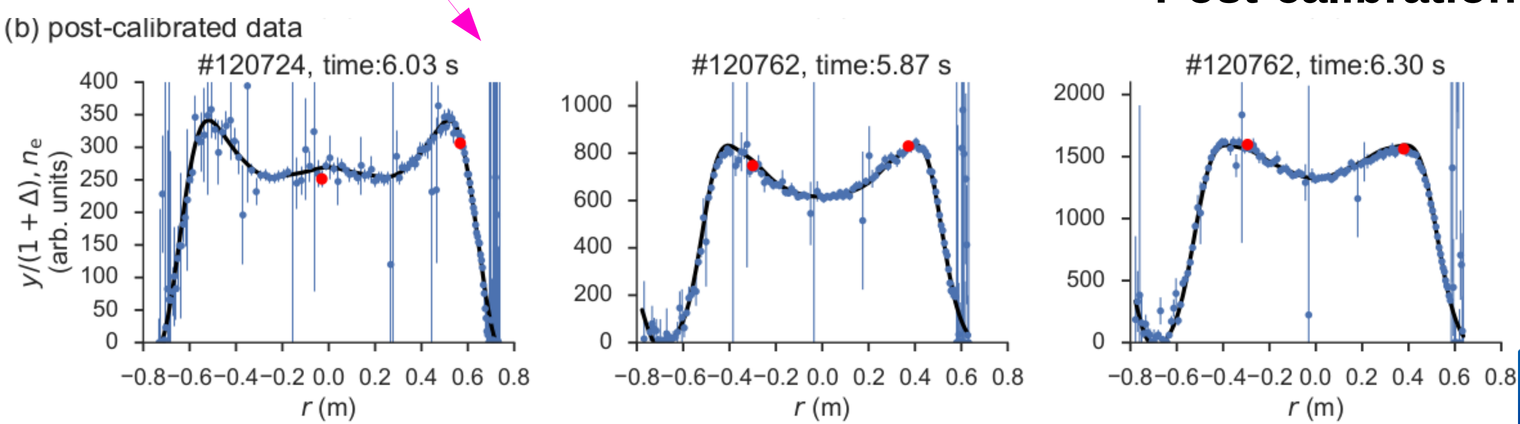
Inference for the training data

Original data



$$\frac{y_{i,j}}{1+\Delta_i} = f_{i,j} + \frac{n_{\mathrm{D}i,j} + n_{\mathrm{P}i,j}}{1+\Delta_i}$$

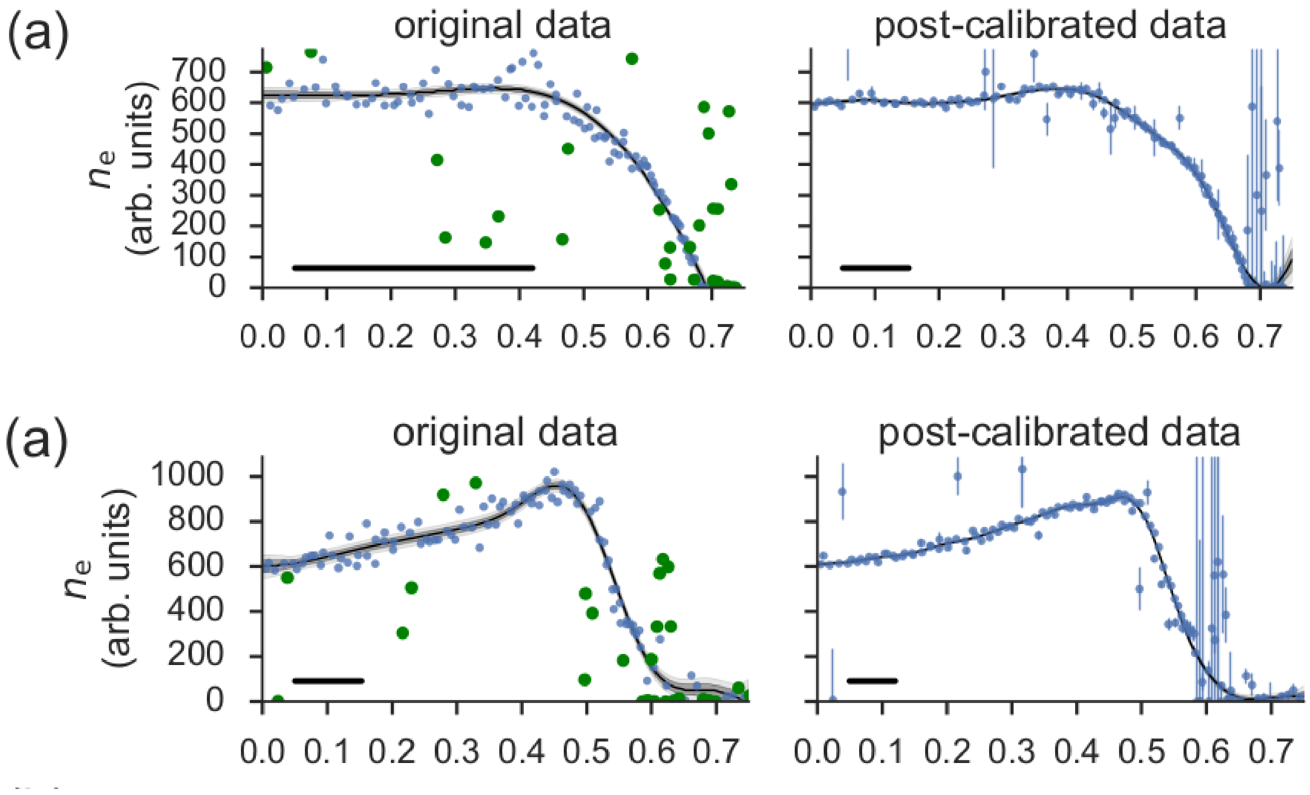
Post-calibration data



Opt. Eng. Lab.

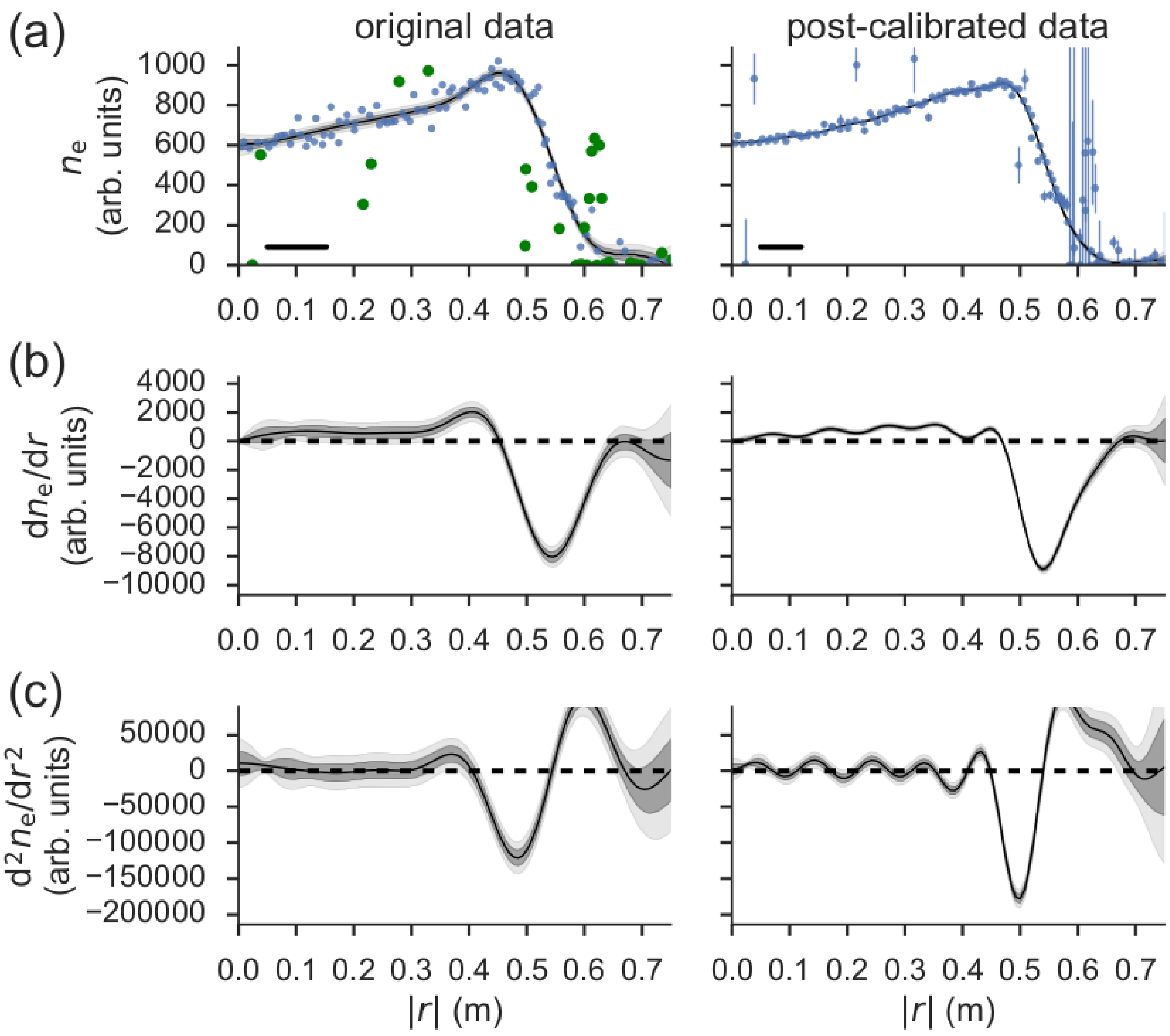
Inference for the test data

We made this post-calibration for **test data** that are NOT used for the **Δ** inference.



Detailed structures become apparent, suggesting no over-fitting.

Application to the derivative inference.



Opt. Eng. Lab.

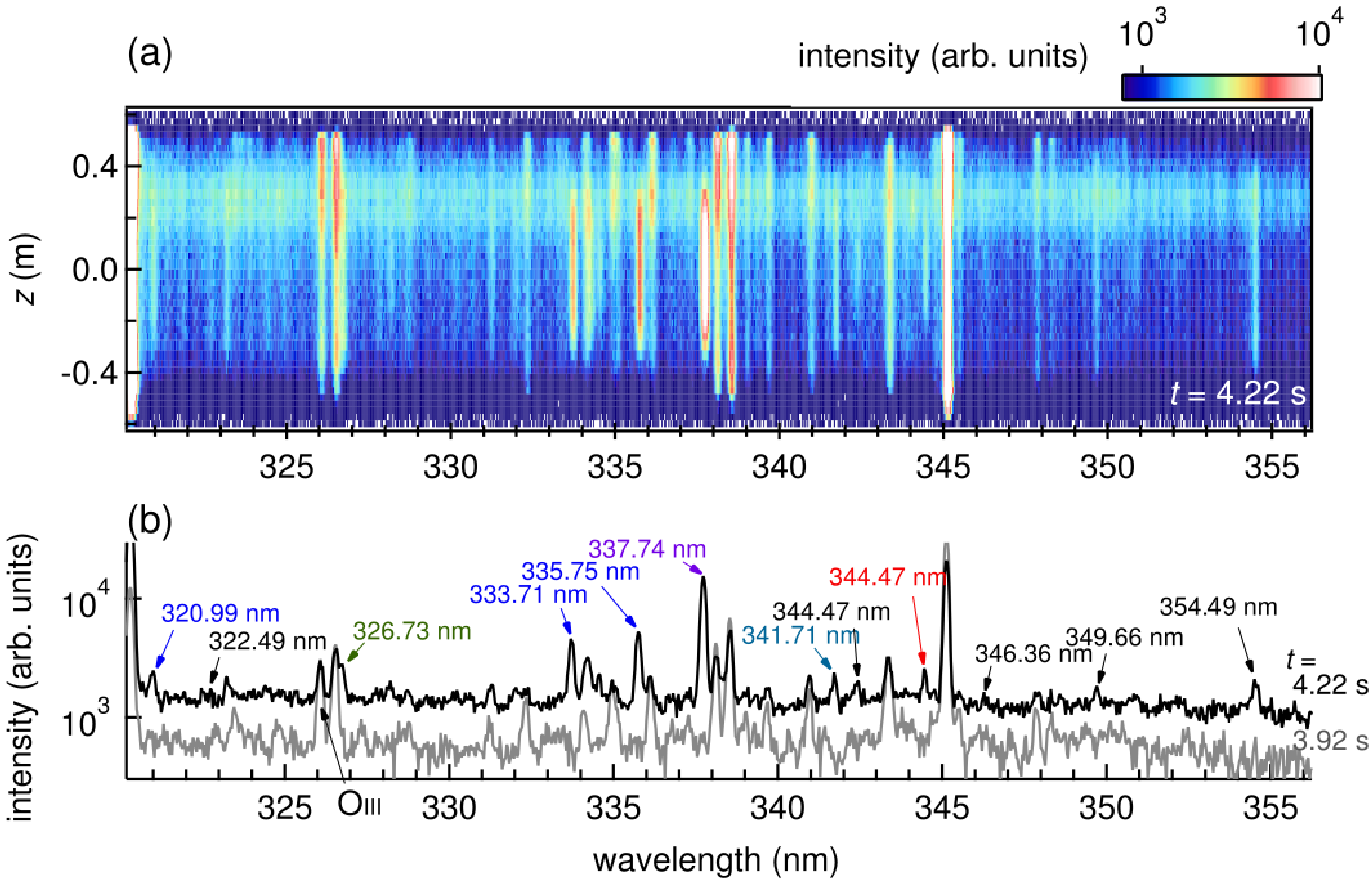


FIG. 5. (a) Two-dimensional image of the spectrum observed for the discharge #121534 at t= 4.22 s as a function of the wavelength (horizontal axis), height z (vertical axis) and intensity (by false color). (b) The spectrum observed at t= 4.22 s for the LOS with z = 0.026 m. The central wavelengths for the highly charged tungsten ion emission lines are indicated in the figure.

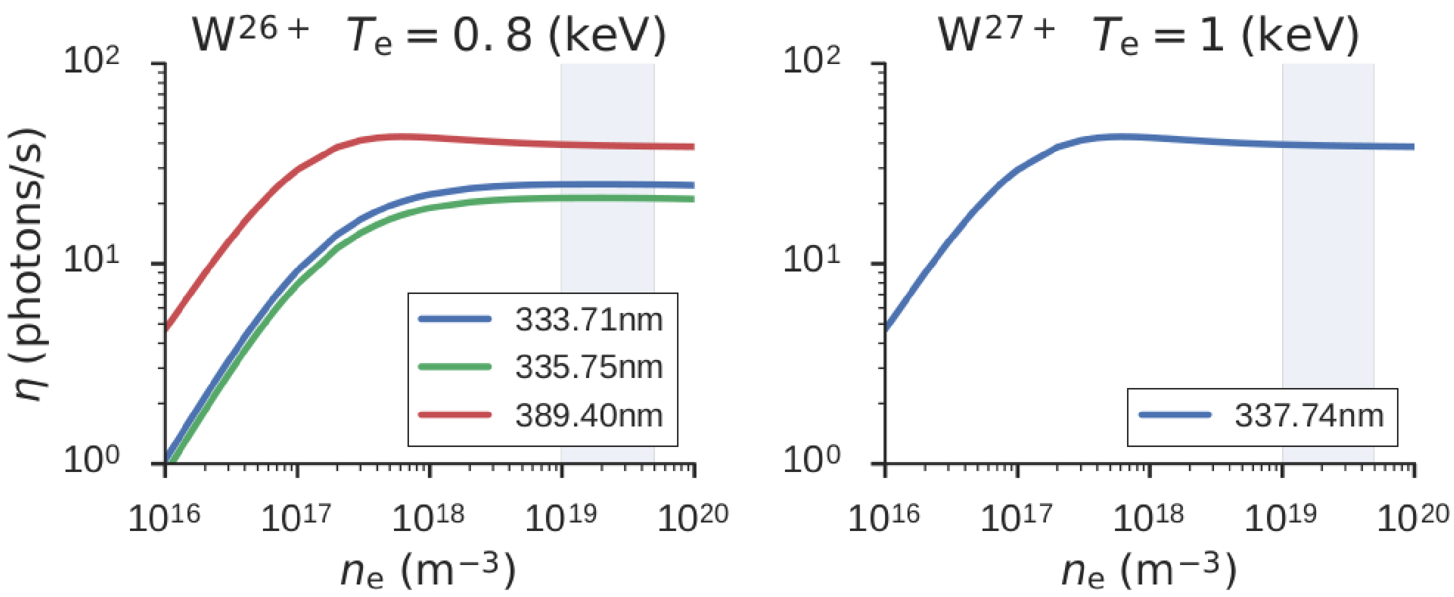


FIG. A.1. n_e dependence of η_i values for 333.71-, 335,75-, 389.40- and 337.74-nm lines estimated by collisional-radiative model [30]. The calculations were made with the assumption of $T_e = 0.8$ keV for the $q = 26$ lines, while $T_e = 1.0$ keV is assumed for the $q = 27$ lines. η_i linearly increases in $n_e < 10^{17}$ m^{-3}, while it becomes saturated in $n_e > 10^{18}$ m^{-3}. The n_e range considered in this work ($n_e = 1 - 5 \times 10^{19}$ m^{-3}) are indicated by shadows. Note that this calculation does not contain the ion-collision effect.

Eng. Lab.