Bayesian inference of impurity transport coefficient profiles

M.A. Chilenski, Y. Marzouk,* M. Greenwald,
N.T. Howard, J.E. Rice, and A.E. White

MIT PSFC/Alcator C-Mod
*MIT Aero/Astro, Uncertainty Quantification Group

May 1, 2015

Supported by USDoE award DE-FC02-99ER54512. Supported in part by a DOE SCGF fellowship, administered by ORISE-ORAU under contract DE-AC05-06OR23100.

Increasing confidence in validation studies through statistically rigorous inference of impurity transport coefficient profiles

Motivation

- Validation of simulations requires rigorous inference of the experimental quantities used for comparison.
- Current approaches to inferring impurity transport coefficients suffer from issues with:
  - Uniqueness of solution
  - Complete accounting of uncertainty

Outline

- Measuring impurity transport coefficients on Alcator C-Mod.
- Current approaches and their shortcomings.
- Use of MCMC to infer impurity transport coefficients.

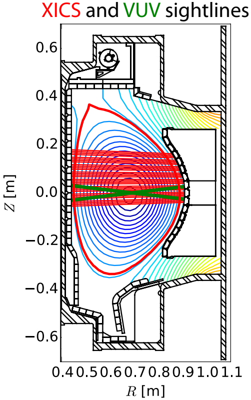
Alcator C-Mod is uniquely equipped to make detailed measurements of impurity transport

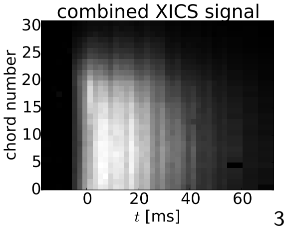
Multipulse laser blow-off impurity injector provides controlled impurity injections [1]

- Multiple injections per shot: up to 10 Hz
- Typically inject CaF_2: calcium is *non-intrinsic* and *non-recycling*

X-ray imaging crystal spectrometer [2] and VUV spectrometers [3] track the impurities

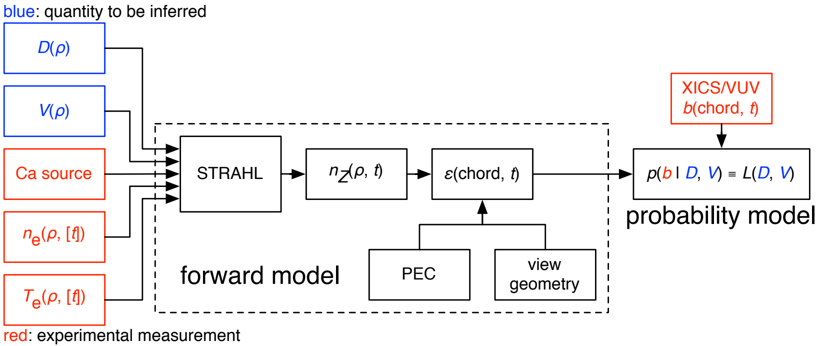
- XICS observes spatial profile of a *single charge state* (Ca^{18+}): more direct interpretation than unresolved soft x-rays
- Two single-chord VUV spectrometers measure Ca^{16+}, Ca^{17+}





[1] Howard et al. (2011), RSI [2] Ince-Cushman et al. (2008), RSI
[3] Reinke et al. (2010), RSI

Inferring impurity transport coefficients is a nonlinear inverse problem



- Objective is to find D, V profiles that best reproduce the observed brightnesses b on each of the diagnostics.
- Key issues are existence, uniqueness and stability of the solution.

Current approaches: maximum likelihood estimate (MLE)

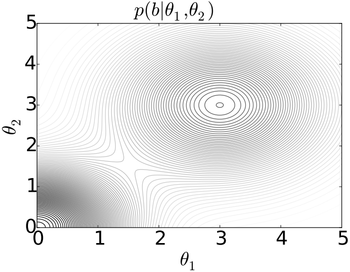
MLE is a standard approach to handle this problem...

$$\hat{D}, \hat{V} = \arg\max_{D,V} p(b|D, V)$$

- Pick D, V profiles which make the observations most likely.
- Use standard optimization techniques: assumption of Gaussian noise makes this a "least squares" problem.
- Need basis functions to represent the profiles with a finite number of variables: typically piecewise linear functions with fixed knots.

...but it has some potential shortcomings

- Point estimate:
  - Risk of underestimating uncertainty.
  - Not valid when there are multiple extrema.
- Propagation of uncertainty in n_e, T_e profiles requires an additional step.



Bayesian statistics provides a framework to overcome the shortcomings of MLE

- Use Bayes' rule to obtain the posterior distribution $p(D, V|b)$, including constraints/prior knowledge $p(D, V)$:

$$p(D, V|b) \propto p(b|D, V)p(D, V)$$

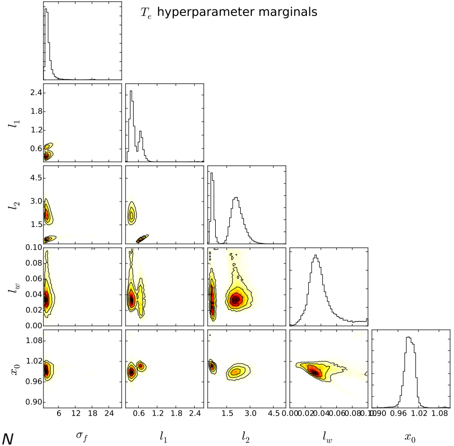
- $p(D, V|b)$ represents the state of knowledge about D, V after having accounted for the data b.
- Working with $p(D, V|b)$ avoids the issues of MLE.
- Can build a joint model that includes the effects of the n_e, T_e profiles explicitly:

$$p(D, V, n_e, T_e|b) \propto p(b|D, V, n_e, T_e)p(D, V)p(n_e)p(T_e)$$

$$p(D, V|b) = \int p(D, V, n_e, T_e|b)\,\mathrm{d}n_e\,\mathrm{d}T_e$$

Markov chain Monte Carlo (MCMC) sampling enables a complete accounting of uncertainty

- MCMC draws samples from unnormalized probability distribution such as $D^{(i)}, V^{(i)} \sim p(D, V|b) \propto p(b|D, V)p(D, V)$.
- Histogram to view $p(D, V|b)$ directly: nonuniqueness can be identified immediately.
- Allows for better point estimates, such as posterior mean and variance:



Chilenski et al. (2015), NF

$$\mathbb{E}[D|b] = \int D p(D|b)\,\mathrm{d}D \approx \frac{1}{N}\sum_{i=1}^{N} D^{(i)}$$

$$\mathrm{var}[D|b] = \int (D - \mathbb{E}[D|b])^2 p(D|b)\,\mathrm{d}D \approx \frac{1}{N-1}\sum_{i=1}^{N}(D^{(i)} - \mathbb{E}[D|b])^2$$

Analysis of C-Mod impurity transport data using these techniques is under way

- Preliminary results from new analysis *do not match previous results.*
  - Non-uniqueness of solution?
  - Poor choice of basis functions?
  - Model selection using information criteria (DIC) [1] is underway.
- Advanced techniques are being used to find "all" extrema [2].
  - Computationally expensive: 10 000+ CPU-hours.
  - Parallelizes well: theoretically linear up to ~ 5000 processors.
  - Code is being upgraded to use MPI, run on big clusters.

[1] Gelman et al. (2014), BDA3 [2] Vousden et al. (2015), arXiv:1501.05823

Application of Bayesian inference allows rigorous estimation of impurity transport coefficient profiles, better confidence in validation studies

- The combination of XICS and LBO enables detailed studies of impurity transport on Alcator C-Mod.
- Inferring impurity transport coefficient profiles using point estimates such as maximum likelihood suffer from issues with:
  - Uniqueness of solution
  - Complete accounting of uncertainty
- New approach under development: use MCMC to find “all” physically reasonable solutions to yield a complete accounting of uncertainty.

Backup slides

Introduction to Bayes' rule

Given a model with parameter vector $\boldsymbol{\theta}$ and observations $\boldsymbol{y}$, Bayes' rule is:

$$\underbrace{f(\boldsymbol{\theta}|\boldsymbol{y})}_{\text{posterior}} = \frac{\overbrace{f(\boldsymbol{y}|\boldsymbol{\theta})}^{\text{likelihood}}\ \overbrace{f(\boldsymbol{\theta})}^{\text{prior}}}{\underbrace{f(\boldsymbol{y})}_{\text{evidence}}}$$

- **Likelihood**: Probability of observing the data $\boldsymbol{y}$ given the parameters $\boldsymbol{\theta}$.
- **Prior**: Distribution encoding any prior assumptions about the parameters $\boldsymbol{\theta}$ (positivity, typical values, etc.)
- **Evidence**: Probability of the data under the model. Just a normalization constant for parameter estimation.
- **Posterior**: Probability distribution for the parameters $\boldsymbol{\theta}$ given the data $\boldsymbol{y}$: the end-goal of the inference.

Model selection with information criteria [1]

- Formalize the tradeoff between goodness of fit and complexity of model: picking the model which minimizes an information criterion is a way to avoid overfitting.
- Two common options:
  - Akaike information criterion (AIC):

$$AIC = -2 \ln p(b|\hat{D}, \hat{V}) + 2k$$

    - $\hat{D}, \hat{V} = \arg\max_{D,V} p(b|D, V)$
    - k is the number of free parameters.
    - Assumes posterior distribution is asymptotically normal.
  - Deviance information criterion (DIC):

$$DIC = -2 \ln p(b|\,\mathbb{E}[D|b], \mathbb{E}[V|b]) + 2p_{eff}$$

    - Effective number of parameters p_{eff} has two definitions:

$$p_{eff,1} = 2\left[\ln p(b|\,\mathbb{E}[D|b], \mathbb{E}[V|b]) - \mathbb{E}[\ln p(b|D, V)]\right]$$

$$p_{eff,2} = 2\,\mathrm{var}[\ln p(b|D, V)]$$

    - $p_{eff,1} = p_{eff,2} = k$ for linear models with flat priors.
    - Better accounts for the information in prior than AIC does.
    - Easier to compute from MCMC output.

[1] Gelman et al. (2014), BDA3