Improved analysis of impurity transport coefficient profiles

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

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

APS-DPP
CP12.00009
Savannah, GA
November 16, 2015

Careful analysis is needed to properly measure impurity transport coefficient profiles

- Validation of simulations requires rigorous inference of the experimental quantities used for comparison.
- Measuring impurity transport coefficients is a very challenging nonlinear inverse problem.
- In particular, it is not sufficient to compute merely one reasonable value for the transport coefficients – there can be multiple, dramatically different solutions that describe the data equally well.
- This lack of uniqueness in the solution *must* be taken into account when computing the uncertainty in the solution and comparing to simulations.

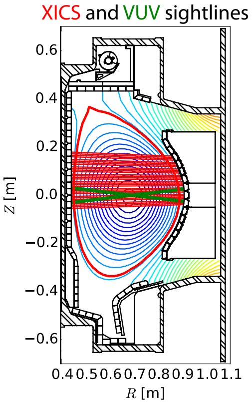
Alcator C-Mod is uniquely equipped to make detailed measurements of impurity transport

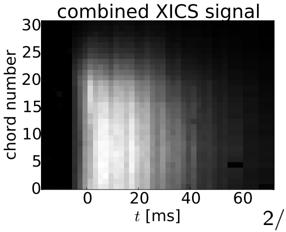
Multipulse laser blow-off impurity injector provides controlled impurity injections [1]

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

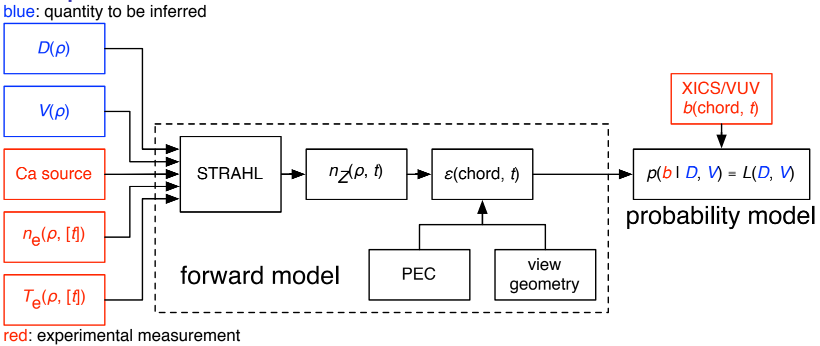
X-ray imaging crystal spectrometer [2] and VUV spectrometers [3] track the impurities

- 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+}





Inferring impurity transport coefficients is a nonlinear inverse problem



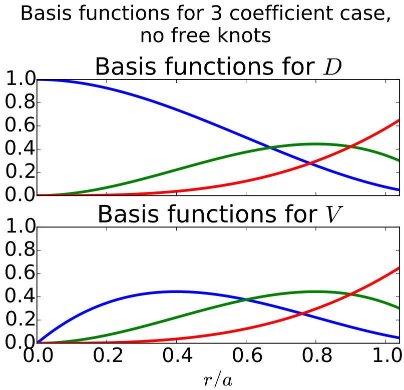
- **This is an inverse problem: given the forward model $b = F(D, V)$, the objective is to find D, V profiles that best reproduce the observed brightness b on each of the diagnostics.**
- Key issues are existence, **uniqueness** and stability of the solution.

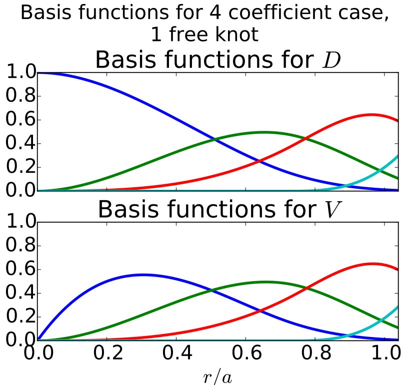
The forward model is built around the STRAHL code

- Assume impurity flux of the form $\Gamma_Z = -D\nabla n_Z + V n_z$.
- STRAHL [4] computes the temporal evolution of the impurity charge state densities $n_{Z,j}(\rho, t)$ for given profiles of D, V.
- Photon emission coefficients from ADAS [5] are used to convert $n_{Z,j}(\rho, t)$ to spectral line emissivity profiles.
- TRIPPy tomography code [6] is used to perform line integrations to obtain $b(\text{chord}, t)$.
- Uncertainty in data is taken to be Gaussian:

$$p(b_{\text{obs}}|D, V, n_e, T_e) = \prod_i \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left\{-\frac{[b_{\text{obs},i} - b_i(D, V, n_e, T_e)]^2}{2\sigma_i^2}\right\}$$

B-spline basis functions are used to obtain a smooth profile, impose constraints





- $\mathrm{d}D/\mathrm{d}r = 0$ at $r/a = 0$
- $D \geq 0$ everywhere
- $V(0) = 0$

There are several challenges when solving this inverse problem

- The forward model is somewhat expensive to evaluate: $\sim 1\,\mathrm{s}$ per run on a typical workstation.
  - This reduces the practicality of sampling-based inference techniques, which may require $\sim 10^7$ samples to find the mode(s) of the posterior distribution and fully characterize the parameter space.
- The nonlinear relationship between the inputs D, V and the outputs $b(\mathrm{chord}, t)$ introduces the possibility that there are multiple profiles of D, V which describe the data equally well.
  - In statistical terms, this means that the posterior distribution may be multimodal.
  - **Failure to account for multiple modes can lead to a dramatic underestimation of the uncertainty in D, V.**

Current approaches: maximum likelihood estimate (MLE)

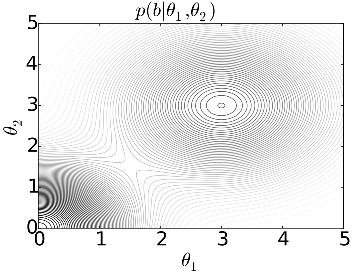
MLE is a standard approach to handle this problem...

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

- 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 provide a framework to overcome the shortcomings of MLE

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

$$\underbrace{f(D, V|b)}_{\text{posterior}} = \frac{\overbrace{f(b|D, V)}^{\text{likelihood}}\overbrace{f(D, V)}^{\text{prior}}}{\underbrace{f(b)}_{\text{evidence}}}$$

- **Likelihood**: Probability of observing the data b given D, V, assumed to be Gaussian.
- **Prior**: Distribution encoding any prior assumptions about D, V (positivity, typical values, etc.)
- **Evidence**: Probability of the data under the model. Just a normalization constant for parameter estimation.
- **Posterior**: Probability distribution for D, V given the data b: **contains all information which is known about D, V.**

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:

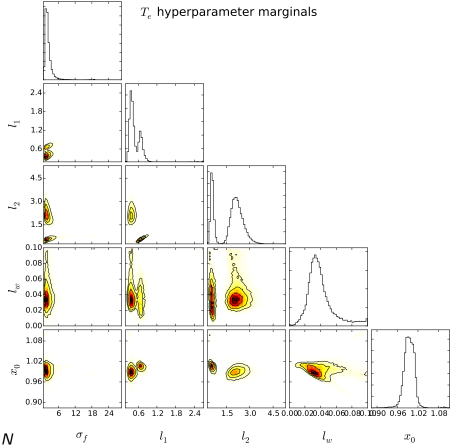


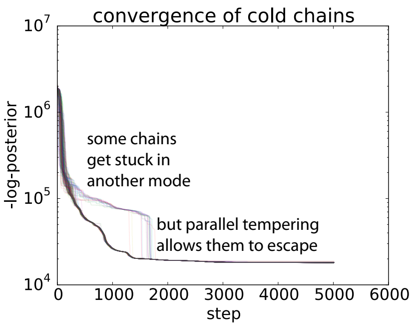
Figure from [7]

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

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

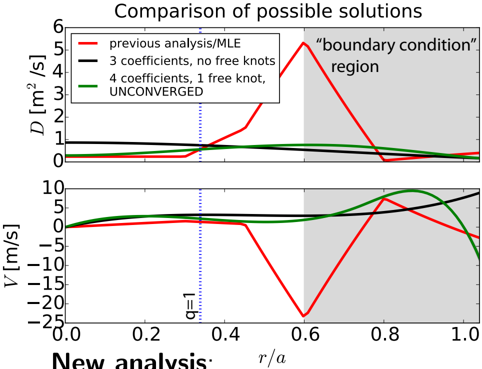
Multimodal posterior necessitates advanced MCMC

- Affine-invariant ensemble sampler (ES) [8, 9]
  - Eliminates need to tune proposal distribution.
  - But, cannot efficiently sample distributions with well-separated modes.
- Parallel tempering (PT) [10]
  - Sample from $p(b|D, V)^{1/T} p(D, V)$ for multiple values of $1 \leq T \leq \infty$.
  - Exchange of information between adjacent T lets chains move between modes.
- Adaptive parallel tempering (APT) [11]
  - Automatically tune T ladder.



- APT with ES in each temperature.
- 200 walkers per temperature, 25 temperatures.
- Plot shows $-\ln p(D, V|b)$ on a log scale: lower value = better fit.

Preliminary results do not match previous analysis



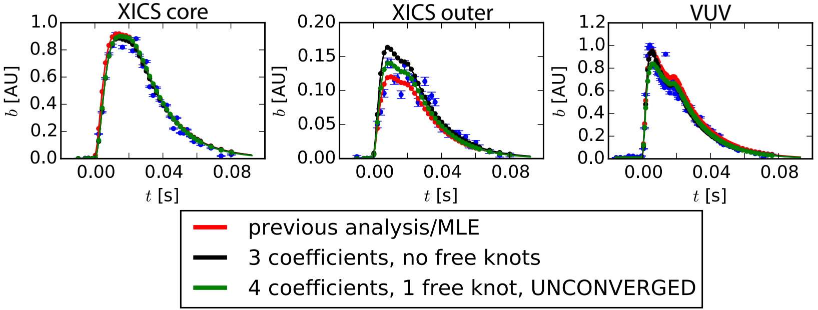
Previous analysis [12]:

- Piecewise linear basis functions.
- MLE without estimate of width of posterior distribution.
- Behavior in $r/a > 0.6$ thought to be only weakly constrained.
- But, uncertainty there too small to be consistent with this.

New analysis:

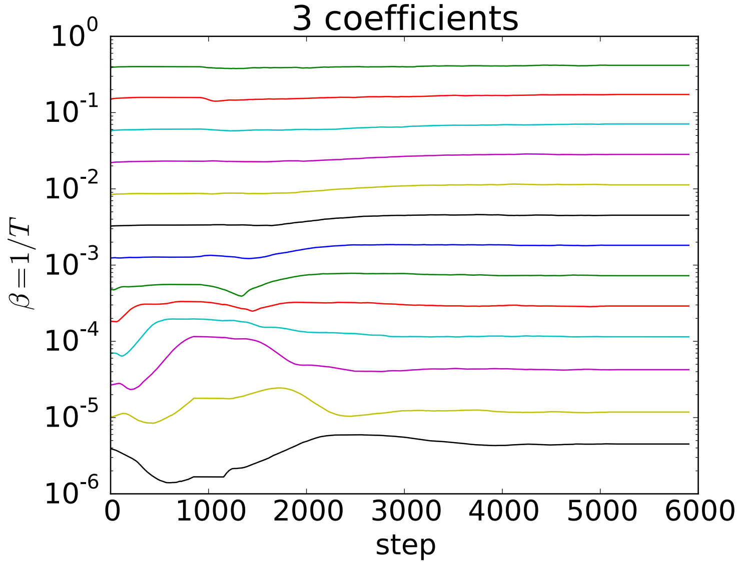
- Cubic B-spline basis functions.
- APT to handle multiple maxima, width of posterior distribution.
- Uncertainty estimate in $r/a > 0.6$ still too small to be consistent with assumed lack of knowledge there.
- *Cases shown are likely overconstrained.*
- Models with more free parameters failed to burn in, *even after many thousands of CPU-hours*.

Predicted brightnesses are similar between all three cases



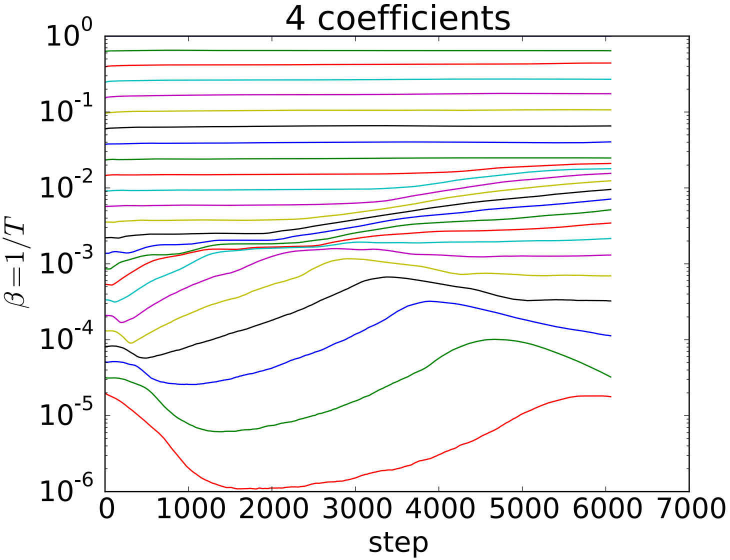
- Agreement on core XICS chords is good in all cases.
- Agreement on outer XICS chords shows widest variation – 4 coefficient case seems to do best job.
- Agreement on VUV spectrometer is reasonable in all cases.
- *This shows the importance of accounting for the possibility of multiple solutions.*

Temperature ladder adaptation for 3 coefficient case



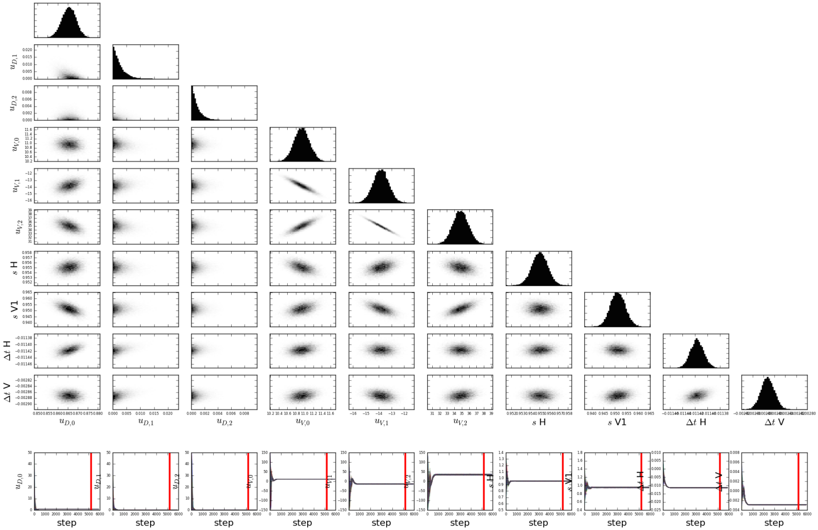
Appears to have settled down after about 5000 steps.

Temperature ladder adaptation for 4 coefficient case



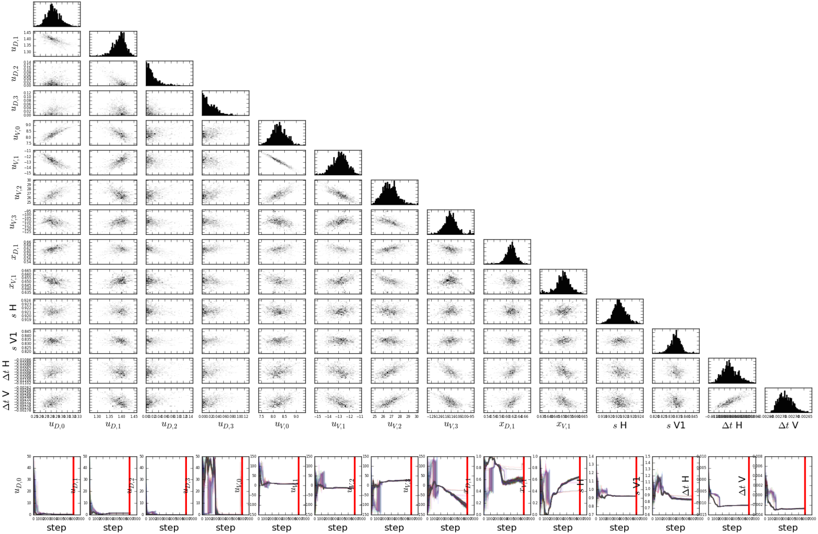
Has not settled down even after 6000 steps.

Posterior distribution for 3 coefficient case



Appears to be unimodal and is fully burned-in.

Posterior distribution for 4 coefficient case



Appears to be unimodal *but has not burned in even after 6000 steps.*

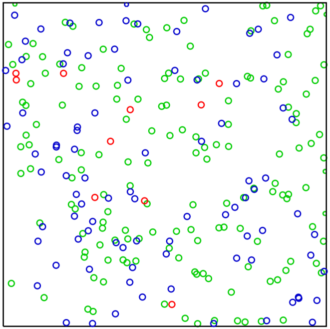
Convergence of MCMC/APT depends strongly on quality of starting conditions

- 4 coefficient case above used 3×10^7 calls to STRAHL – about 9000 CPU-hours.
- More flexible basis functions take even longer to burn in.
- Ideally, the sampler will be initialized with most walkers already near the posterior modes.
- This necessitates the use of global optimization techniques to find "all" of the posterior modes:
    - PaGMO/PyGMO [13] enables parallelization of genetic algorithms-based global optimizers through use of the Generalized Island Model [14].
    - Also includes tools for efficiently searching a parameter space for local extrema.

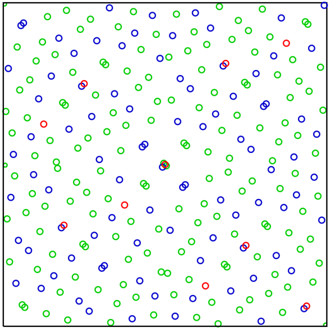
The Sobol sequence provides a systematic, efficient way of exploring the parameter space

Sobol sequence efficiently fills the parameter space, leaving fewer holes than pseudorandom sampling, better statistical properties than a uniform grid [15, 16].

Pseudorandom sequence:



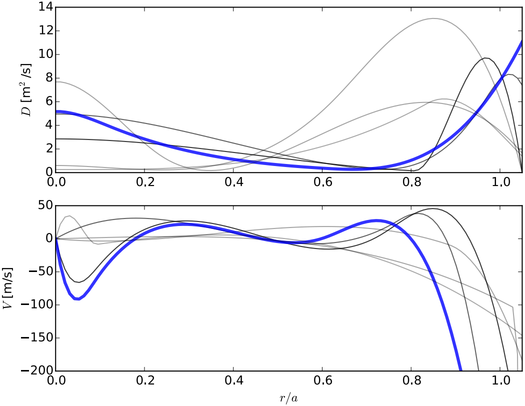
Sobol quasirandom sequence:



Images from [17, 18]

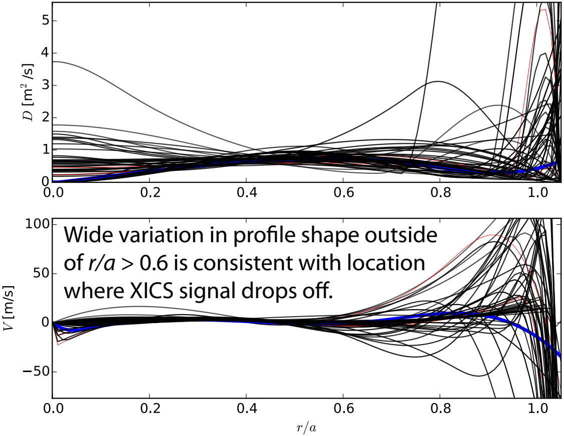
Systematic exploration of the parameter space using local optimizers is underway

- Sampled parameter space using 5×10^5 point Sobol sequence then started local optimizers at the 100 best points found.
- Discarded solutions which ended up too close to bounds.
- This left 6 possible solutions, best solution is shown in blue, remainder are shaded according to χ^2.

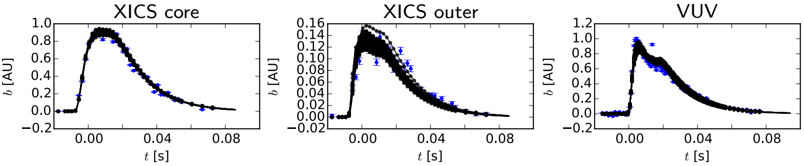


A more brute-force approach was also attempted

- Launched local optimizers at $\sim$1000 points, again using a Sobol sequence to efficiently sample the space.
- Repeatedly restarted optimizers from the previous solutions, periodically pruning bad/stuck solutions.
- Ended up with 46 solutions.



Despite dramatic differences in profile shape, the solutions from the brute-force local extrema search all provide reasonable fits to the brightness data



- Fit to core XICS, VUV chords is good in all cases.
- Fit to outer XICS chords again shows widest variation.
- **This indicates that the solution is not unique. Failing to account for the multiple possible solutions leads to an underestimation of the uncertainty in D, V.**

Choice of local optimizer has a critical effect on the quality of the solution obtained

algorithm	result
Nelder-Mead	17967
Subplex	19390
compass search	46328
COBYLA	60674
BOBYQA	84250

- Result shown is the negative log-posterior ($\propto \chi^2$): lower is better.
- Each algorithm started with the same initial guess.
- Limited each algorithm to 1000 iterations.
- Repeated 10 times, since this (surprisingly) seemed to deliver better performance than just running for 10000 iterations.

Next step: include uncertainty in n_e, T_e profiles

Form joint posterior distribution, now also conditional on the profile measurements d:

$$p(D, V, n_e, T_e|b, d) = p(D, V|n_e, T_e, b, d)p(n_e, T_e|b, d)$$

Use Gaussian processes for n_e, T_e [7]:

$$p(n_e|d) = \mathcal{N}(m(\rho), k(\rho, \rho))$$

Reduce dimension of parameter space by approximating this with truncated eigendecomposition:

$$n_e = Q\Lambda^{1/2}u + m(\rho), \quad u \sim \mathcal{N}(0, I), \quad k(\rho, \rho) = Q\Lambda Q^{-1}$$

Find marginal posterior distribution for D, V using MCMC:

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

Conclusions

- Rigorous quantification of the uncertainties in impurity transport coefficients is essential for validation of multichannel transport simulations.
- The computational expense, nonlinearity of the forward model make the problem difficult to solve and susceptible to multiple extrema.
- Work is underway to combine advanced optimization and inference tools to overcome these issues.

Future Work

- Deployment of more advanced optimizers to more efficiently identify local extrema.
- MCMC sampling using results from local extrema search.
- Incorporation of SXR data.
- Accounting for uncertainties in n_e, T_e data.

References

[1] N. T. Howard, M. Greenwald, and J. E. Rice, Review of Scientific Instruments **82**, 033512 (2011).

[2] A. Ince-Cushman et al., Review of Scientific Instruments **79**, 10E302 (2008).

[3] M. L. Reinke et al., Review of Scientific Instruments **81**, 10D736 (2010).

[4] R. Dux, STRAHL user manual, Technical Report IPP 10/30, Max-Planck-Institut für Plasmaphysik, 2006.

[5] H. P. Summers, The ADAS user manual, version 2.6, http://www.adas.ac.uk, 2004.

[6] I. C. Faust, Toroidal Radiation Inversion Protocol (Python), https://github.com/icfaust/TRIPPy.

[7] M. A. Chilenski et al., Nuclear Fusion **55**, 023012 (2015).

[8] J. Goodman and J. Weare, Communications in Applied Mathematics and Computational Science **5**, 65 (2010).

[9] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publications of the Astronomical Society of the Pacific **125**, 306 (2013).

[10] D. J. Earl and M. W. Deem, Physical Chemistry Chemical Physics **7**, 3910 (2005).

[11] W. Vousden, W. M. Farr, and I. Mandel, Dynamic temperature selection for parallel-tempering in Markov chain Monte Carlo simulations, http://arxiv.org/abs/1501.05823, 2015.

[12] N. T. Howard et al., Nuclear Fusion **52**, 063002 (2012).

[13] D. Izzo, PyGMO and PyKEP: Open source tools for massively parallel optimization in astrodynamics (the case of interplanetary trajectory optimization), in *5th ICATT: International Conference on Astrodynamics Tools and Techniques, ESTEC/ESA, The Netherlands, 29 May – 1 June 2012*, 2012.

[14] D. Izzo, M. Ruciński, and F. Biscani, *Parallel Architectures and Bioinspired Algorithms*, volume 415 of *Studies in Computational Intelligence*, chapter The Generalized Island Model, pages 151–169, Springer-Verlag, Berlin Heidelberg, 2012.

[15] I. M. Sobol, USSR Computational Mathematics and Mathematical Physics **7**, 86 (1967).

[16] C. Lemieux, *Monte Carlo and Quasi-Monte Carlo Sampling*, Springer Series in Statistics, Springer Science+Business Media, New York, NY, 2009.

[17] J. Heald, Pseudorandom sequence 2d, https://en.wikipedia.org/wiki/File:Pseudorandom_sequence_2D.svg, 2011.

[18] J. Heald, Sobol sequence 2d, https://en.wikipedia.org/wiki/File:Sobol_sequence_2D.svg, 2011.

Acknowledgements

- This material is based upon work conducted using the Alcator C-Mod tokamak, a DOE Office of Science user facility.
- This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Fusion Energy Sciences under Award Number DE-FC02-99ER54512.
- The XEUS and LoWEUS spectrometers were developed at the LLNL EBIT lab. Work at LLNL was performed under the auspices of the US DOE under contract DE-AC52-07NA-27344.
- This material is based upon work supported in part by the U.S. Department of Energy Office of Science Graduate Research Fellowship Program (DOE SCGF), made possible in part by the American Recovery and Reinvestment Act of 2009, administered by ORISE-ORAU under contract number DE-AC05-06OR23100.
- Computations using STRAHL were carried out on the MIT PSFC parallel AMD Opteron/Infiniband cluster Loki.