Using PyBLoCXS for a Fully Bayesian Analysis of Calibration Uncertainty in High Energy Spectral Analysis

David A. van Dyk[1], Jin Xu[1], Vinay Kashyap[2], Aneta Siemiginowska[2], Alanna Connors[3], Brian Refsdal[2]

[1]University of California, Irvine, dvd@ics.uci.edu, jinx@uci.edu

[2]Smithsonian Astrophysical Observatory, vkashyap@cfa.harvard.edu, asiemiginowska@cfa.harvard.edu, brefsdal@cfa.harvard.edu

[3]Eureka Scientific, aconnors@eurekabayes.com

Abstract

Systematic instrumental uncertainties in astronomical analyses have been generally ignored due to the lack of robust principled method, though the importance of incorporating instrumental calibration uncertainty is widely realized by users and instrument builders. Ignoring calibration uncertainty can cause bias in the estimate of source model parameters and underestimate their variance. In this poster, we focus on incorporating uncertainty for the effective area curve into a principled fully Bayesian spectral analysis. A principle component analyses is explored to efficiently represent the variability of the effective area curve, enabling a fully Bayesian analysis of calibration uncertainty in spectral analysis of high-energy Chandra data. The method is compared with standard analysis techniques and the so-called"pragmatic" Bayesian method of Lee et al(2011, ApJ). The advantage of the fully Bayesian method is that the data itself can provide information for both the source parameters and for the effective area curve. It is verified that implementing our fully Bayesian method can result in more accurate and efficient estimation of source parameters, and valid estimates of uncertainty.

Calibration Products

Analysis highly depende on Calibration Products:

- Effective Area records sensitivity as a function of energy
- Energy Redistribution Matrix
- Point Spread Functions
- Exposure Map

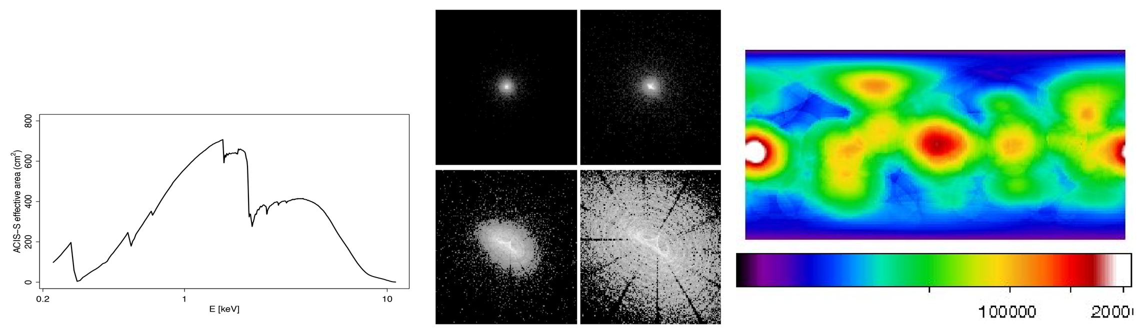

FIGURE 1: The left image is an example of a CHANDRA effective area curve. The middle one is four samples of Chandra psf's (Karovska et al., ADASS X). The right one shows an EGERT exposure map (area × time).

PyBLoCXS

Our methods are carried out using PyBLoCXS, a sophisticated Markov chain Monte Carlo (MCMC) based algorithm designed to carry out Bayesian Low-Count X-ray Spectral (BLoCXS) analysis in the Sherpa environment. PyBLoCXS use a mixture of a random walk Metropolis sampler and a Metropolis Hastings Independence sampler to draw parameters from their Bayesian posterior distribution. In principle, It can be implemented in combination with all the models and methods available in Sherpa. More details about PyBLoCXS can be found in the following link:
http://hea-www.harvard.edu/AstroStat/pyBLoCXS/

The Calibration Sample

Drake et al. (2006), suggests generating calibration samples of effective area curves to represent the uncertainty. We denote Calibration Sample by $\{A_1, A_2, A_3, ..., A_L\}$.

To parameterize calibration sample, we use Principle Component Analysis(PCA).

$$A = A_0 + \bar{\delta} + \sum_{j=1}^{m} e_j r_j v_j$$

A_0 : default effective area,

$\bar{\delta}$: mean deviation from A_0,

r_j,v_j : first m PCA eigenvalues & vectors,

e_j : independent standard normal deviations.

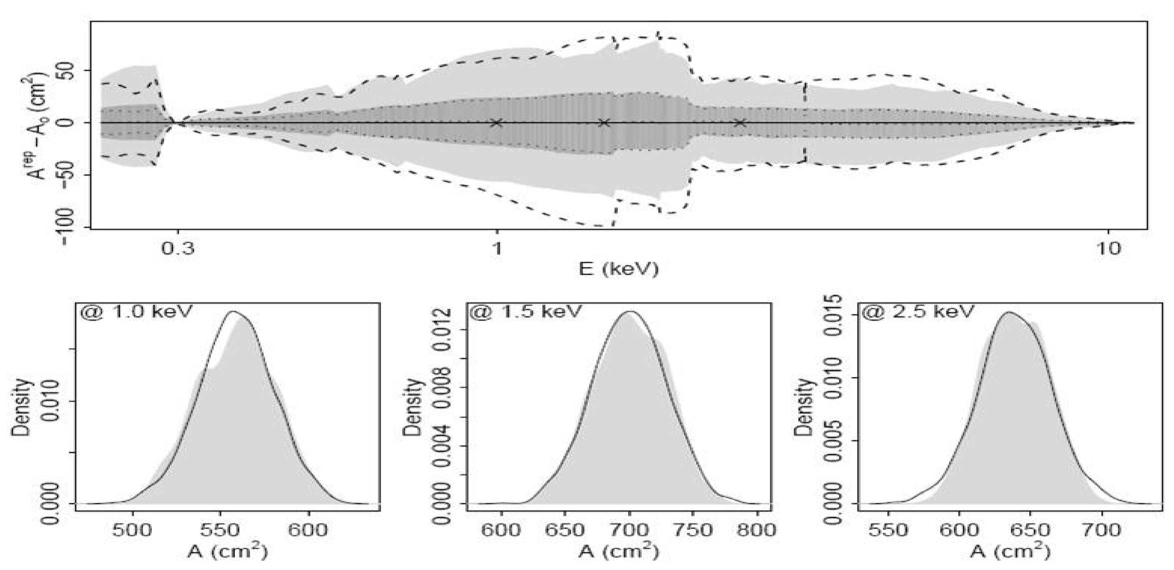

FIGURE 2: The grey regions in the upper panel give intervals for each energy bin that contain 100% and 68.3% of the calibration sample. The dashed and dotted lines outline intervals for each energy bin containing 100% and 68.3% of 1000 PCA replicates of the effective area. The bottom panel details the comparisons between the calibration sample and PCA summary for the particular energy bins.

A Fully Bayesian Analysis

To incorporate calibration uncertainty into proper source parameter estimates,we develop a Fully Bayesian approach that works in conjunction with Markov chain Monte Carlo based fitting using PyBLoCXS. Fully Bayesian analysis allows the current data to influence the choice of calibration product, which improves estimation and error bars relative to the fixed effective area sampling scheme and the "pragmatic" Bayesian sampling scheme.

Let θ be the source parameter vector, Y the observed data, and $L(\cdot)$ the likelihood function, $p(\theta)$ is the prior distribution of θ derived from PCA method.

Scheme One: Fixed Effective Area Curved. We assume $A = A_0$, where A_0 is the default effective area curve. (But not necessarily the true one.)
Sampling Step:

Sample θ from $p(\theta|Y, A_0) \propto L(Y|\theta, A_0)p(\theta)$.
Using PyBLoCXS.

Scheme Two: Pragmatic Bayesian. This scheme incorporates the maximum uncertainty due to the calibration uncertainty. By sampling from the prior distribution, $p(A)$, rather than $p(A|Y)$, it assumes that the correct data is uninformative for the choice of effective area curve.

Step One:

Sample A from $p(A)$.

Step Two:

Sample θ from $p(\theta|Y, A) \propto L(Y|\theta, A)p(\theta)$.
Using PyBLoCXS.

Scheme Three: Fully Bayesian. This scheme uses a fully Bayesian principled approach. It samples A and θ jointly from their posterior distribution using a Gibbs sampler. (The update for θ is accomplished via the mixture of Metropolis-Hastings kernel used in PyBLoCXS.) In this way the data is allowed to influence the choice of effective area curve.

Step One:

Sample A from $p(A|Y, \theta) \propto L(Y|\theta, A)p(A)$.

Step Two:

Sample θ from $p(\theta|Y, A) \propto L(Y|\theta, A)p(\theta)$.

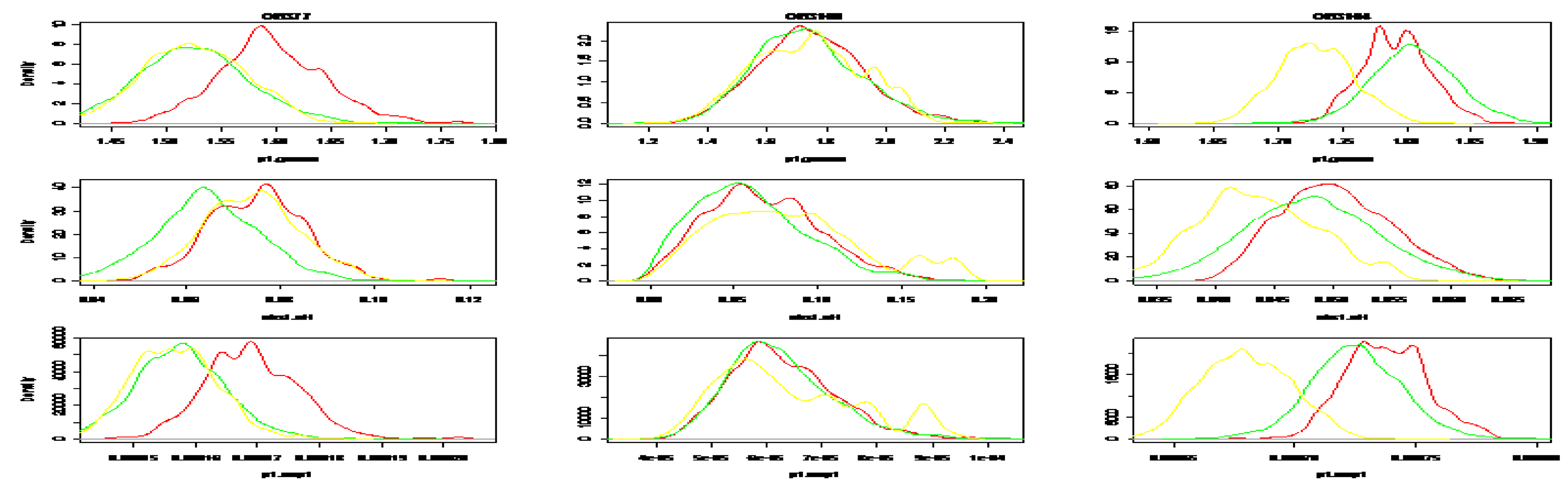

Figure 3: Comparison of three sampling schemes. From left to right are separately source parameters' sampling for Chandra datasets ob377, ob3100, and ob3104. Red line, green line and yellow line represent the histogram density results using Fixed Effective Area, Pragmatic Bayesian and Fully Bayesian.

We apply the three schemes to three data from Chandra observations of three quasars (see Figure 3). For dataset ob377 (first column), the result from the Pragmatic Bayesian scheme differs from the results with a fixed effective area curve, while the result of the Fully Bayesian scheme acts differently with different parameters. p1.gamma, p1.amp1 from Fully Bayesian scheme are similar to Pragmatic Bayesian scheme, and abs1.nH is similar to the result with a fixed effective area curve. For dataset ob3100 (column 2), the results from the three sampling schemes are quite similar. For dataset ob3104 (column 3), the results of the Fully Bayesian scheme is significant different from the the other two. From a Fully Bayesian perspective, the dataset is in better agreement with a more extreme curve from the calibration sample. This has a clear effect on the best fit for the spectral parameters and their error bars.

Conclusions

- Fully Bayesian method owns the advantage that the data itself can provide information for both the source parameters and for the effective area curve.
- Implementing fully Bayesian model can result in more accurate and efficient estimation of source parameters, and valid estimate of uncertainty.

This work was supported in part by NSF grant numbers DMS 09-07522 and 09-07185, CXC NASA contract NAS8-39073 and NASA AISRP grant NNG06GF17G.

References

[1] Lee, H., Kashyap, V., van Dyk, D., Connors, A., Drake, J., Izem, R., Min, S., Park, T., Ratzlaff, P., Siemiginowska, A., and Zezas, A. Accounting for Calibration Uncertainties in X-ray Analysis: Effective Area in Spectral Fitting. The Astrophysical Journal, Volume 731, Issue 2, article id. 126 (2011).

[2] van Dyk, D. A. and Park, T. Partially Collapsed Gibbs Sampling Path-Adaptive Metropolis-Hastings in High-Energy Astrophysics. In Handbook of Markov Chain Monte Carlo (Editors: S. Brooks, A. Gelman, G. Jones and X.-L. Meng), Chapman Hall/CRC Press.