#41.15, Astrostatistics, AAS-HEAD, LA, 2008

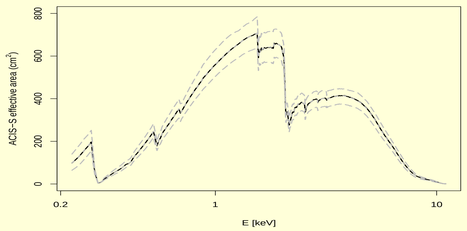
Incorporating Effective Area Uncertainties into Spectral Fitting

H.Lee*, V.Kashyap, J.Drake, A.Connors, R.Izem, T.Park, P.Ratzlaff, A.Siemiginowska, D.vanDyk, A.Zezas

H. Lee*
Harvard-Smithsonian
Center for Astrophysics
hlee@cfa.harvard.edu

Effective area is not that well known

Be aware that your effective area curve (arf) has errors and these errors affect your analysis. The plot below shows the coverage of a sample of 1000 ACIS-S arfs generated by Drake et al. (2006, 2007) and the default arf ($\mathbf{a}_o$) is in a black line.

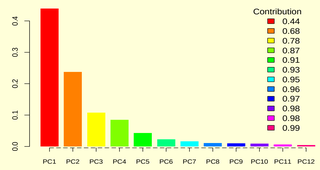


This scatter must be included in the analysis. By devising MCMC algorithms, we propose Bayesian hierarchical modeling for spectral fitting.

Summarizing the black box of arfs ($\mathcal{A}$)

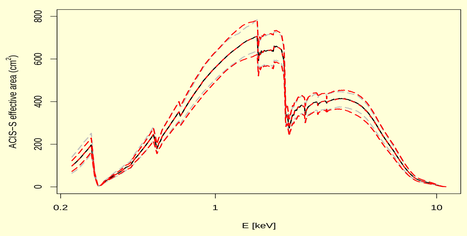
Do we need to carry this entire sample into the analysis? NO! Principal Component Analysis (PCA) reduces dimensionality and summarizes the arf set $\mathcal{A}$ with a small number of principal components (PCs). Let $\mathbf{a}_{j=\{1,...,M\}} \in \mathcal{A}$ be given arfs by calibration scientists on which we perform PCA.

Scree plot: 8 PCs explain 96% of total variation; 12 PCs explain 99%. No need for 1000 arfs. We will use the first 8 PCs ($\mathbf{v}_n$) and 8 coefficients (r_n) to simulate arfs. An arf $\mathbf{a}_{(j^*)}$ with 8 PCs is generated via:



$$\mathbf{a}_{(j^*)} = \mathbf{a}_o^* + \delta_\mathbf{a} + \sum_{n=1}^{8} e_n r_n \mathbf{v}_n, \quad e_n \sim N(0,1) \qquad (1)$$

where $\mathbf{a}_o^*$ is the supplied default arf, $\mathbf{a}_o$ is the default arf, $\bar{\mathbf{a}}$ the mean of $\mathbf{a}_j$s, and $\delta_\mathbf{a} = \bar{\mathbf{a}} - \mathbf{a}_o$ (Kashyap et al.,2007). The figure below shows simulated 1000 arfs in red lines. The 8 PCs are sufficient to match the arf uncertainty represented by gray lines.



Gibbs sampler for fitting spectra

For the best fits and their uncertainties, we simulate the posterior distribution of parameters of interest θ given photon counts $\mathbf{y}$ and an instrument response $\mathbf{a}$ via BLoCXS (van Dyk et al. 2001) that implements MCMC (Gibbs sampler) for spectral fitting in the Poisson regime. A Gibbs sampler iterates between these two steps:

Data Augmentation: $\mathbf{Y}^{(k)} \sim p(y|\theta^{(k-1)}, \mathbf{a})$
Draw Parameters: $\theta^{(k)} \sim p(\theta|\mathbf{y}^{(k)}, \mathbf{a})$

The posterior distribution offers best fits and uncertainties automatically. Averaging $p(\theta|\mathbf{y}, \mathbf{a})$ over $\mathcal{A}$ (a set of arfs) provides the posterior of θ given photon counts $\mathbf{y}$, $p(\theta|\mathbf{y})$ that includes calibration uncertainty.

Marginalizing over arfs

$$p(\theta|\mathbf{y}) = \int_\mathcal{A} p(\theta|\mathbf{y}, \mathbf{a}) p(\mathbf{a}) d\mathbf{a} = \frac{1}{M} \sum_{j=1}^{M} p(\theta|\mathbf{y}, \mathbf{a}_j)$$

How to marginalize over arfs?

Drake et al (2006, 2007) proposed a strategy [B.0] using standard packages (e.g., XSPEC) and we propose three algorithms [B.1-B.3] with BLoCXS. [B.0] tends to be tedious and time consuming depending on the size of the arf library, whose Bayesian counterpart is [B.1]. To speed up [B.1], we introduce [B.2] by selecting arfs randomly from the arf library. To improve computational efficiency, we introduce [B.3]. Given the observed spectrum, [B.0-B.3] work as follows:

[B.0] Fit with XSPEC

Require: M ARFs and spectral fitting engines;
  for $j = 1, ..., M$ **do**
    Set a new arf and fit the spectrum, best fit $\hat{\theta}_j$.
  end for
  Compute mean and variance of $\{\hat{\theta}_j\}_{j=1,...,M}$.

Repeat fitting procedures as many times as the size of the arf library instead of the supplied default arf. Very tedious!!!

[B.1] Fit with Gibbs sampler

Require: M ARFs and Bayesian spectral fitting engines;
  Set initial values including priors
  for $j = 1, ..., M$ **do**
    repeat
      Augment data $y_{k|j}$ given $\theta_{k-1|j}$ and $\mathbf{a}_j$
      Draw $\theta_{k|j}$ from $p(\theta|\mathbf{y}_{k|j}, \mathbf{a}_j)$
    until the chain $\{\theta_{k|j}\}$ is stable, $k = 1, ..., n_j$.
    Drop n_b draws of a burn-in period.
  end for
  Compute mean and variance of $\{\theta_{k|j}\}_{j=1,...,M}$ for calibration.

Extra-tedious! The individual gibbs sequence $\{\theta_{k|j}\}$ offers a statistical error that varies depending on arfs. See the plot on the lower right.

[B.2] Fit with randomized arfs

Require: M ARFs and Bayesian spectral fitting engines;
  Set initial values including priors
  repeat
    Choose $\mathbf{a}_{(j)}$ randomly among M ARFs.
    Augment data $y_{k(j)}$ given $\theta_{k-1(j-1)}$ and $\mathbf{a}_{(j)}$
    Draw $\theta_{k(j)}$ from $p(\theta|\mathbf{y}_{k(j)}, \mathbf{a}_{(j)})$
  until the chain $\{\theta_{k(j)}\}$ is stable, $k = 1, ..., n$.
  Drop n_b draws of a burn-in period.
  Compute mean and variance of $\{\theta_{k(j)}\}$.

Randomizing arfs saves the *for* loop in [B.1].

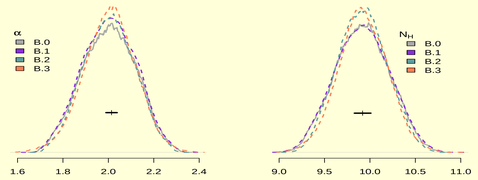
[B.3] Fit with PC simulated arfs

Require: PCs ($\mathbf{v}_n$), coefficients (r_n), and spectral fitting engines;
  Set initial values including priors
  repeat
    Simulate $\mathbf{a}_{(j^*)}$ based on PCs. (see eq.(1).)
    Augment data $y_{k(j^*)}$ given $\theta_{k-1(j^*-1)}$ and $\mathbf{a}_{(j^*)}$
    Draw $\theta_{k(j^*)}$ from $p(\theta|\mathbf{y}_{k(j^*)}, \mathbf{a}_{(j^*)})$
  until the chain $\{\theta_{k(j^*)}\}$ is stable, $k = 1, ..., n$.
  Drop n_b draws of a burn-in period.
  Compute mean and variance of $\{\theta_{k(j^*)}\}$.

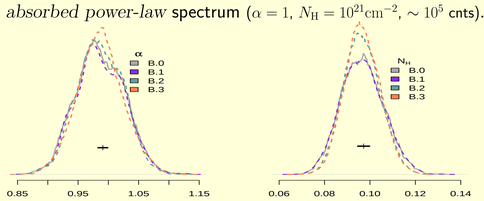
We distinguish (j^*), PC simulation from (j), randomization.

Comparison across algorithms

Results from these algorithms work very similarly as shown below but [B.3] is most efficient. One histogram of best fits [B.0] and three posterior density profiles [B.1-B.3] from fitting an *absorbed power-law* spectrum of photon index $\alpha = 2$, column density $N_\mathrm{H} = 10^{23}\mathrm{cm}^{-2}$, and total counts $\sim 10^5$ are shown. The black bar indicates a best fit$\pm\hat{\sigma}$ only with the default arf. The widths of posterior densities represent errors including calibration uncertainty.



Another *absorbed power-law* spectrum ($\alpha = 1$, $N_\mathrm{H} = 10^{21}\mathrm{cm}^{-2}$, $\sim 10^5$ cnts).



How many arfs?

PCs and coefficients depend on the arf sample provided by calibration scientists but they indicate that a relatively small number of arfs is sufficient to incorporate calibration uncertainty instead of thousands.

Law of Total Variance (LTV)

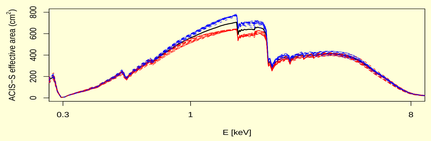
LTV explains the complexity of the error decomposition. A best fit depends on arfs and its uncertainty has two components, statistical error and calibration error which are not independent. LTV indicates that the calibration error is dominant with high count data where the statistical error becomes minuscule. This law also explains that [B.3] of 8 PCs (96% calibration error) tends to be slightly narrower than other algorithms.

$$V[\theta] = V[E[\theta|\mathbf{a}]] + E[V[\theta|\mathbf{a}]]$$

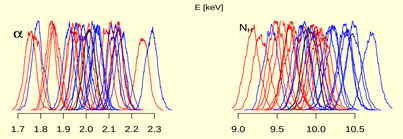
Behaviors of calibration and statistical errors

Depending on the model used, these two errors may not be separable. In the plot below, two groups of 15 similar arfs are colored and the histograms of gibbs sequences are colored according to the arf colors (default arf in black). The shifting pattern of posteriors does not match between these two spectra. This figure clearly shows that best fit values change with arfs and that calibration uncertainty must be incorporated into spectral fitting.

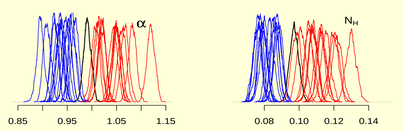
an *absorbed power-law* spectrum



$\alpha = 2$, $N_\mathrm{H} = 10^{23}\mathrm{cm}^{-2}$, $\sim 10^5$ counts

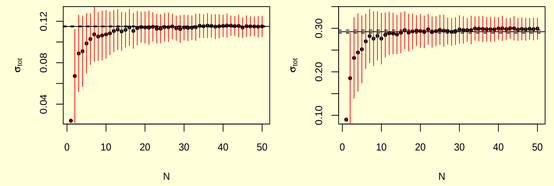


$\alpha = 1$, $N_\mathrm{H} = 10^{21}\mathrm{cm}^{-2}$, $\sim 10^5$ counts



Asymptotics of calibration error

In the figure below, the horizontal solid line represents the average uncertainty derived from [B.0] and the dashed lines represent the range in this uncertainty obtained from 20 simulations ($\alpha = 2$, $N_\mathrm{H} = 10^{23}\mathrm{cm}^{-2}$, $\sim 10^5$ counts). Also shown are the results obtained from combining posterior pdfs by using different numbers of arfs. Dots represent the mean uncertainty and vertical bars denote errors on the means; in other words, N arfs from 1000 are randomly chosen to get the uncertainty of $\frac{1}{N}\sum_{(j)} p(\theta|\mathbf{y}, \mathbf{a}_{(j)})$ for 200 times, and the means and rms errors of these uncertainties are the dots and bars. This figure shows that after N≈25, the estimated uncertainty is stabilized and therefore, ∼25 fits with different arfs are sufficient to account for calibration uncertainty provided that the full posterior pdf on the parameters is obtained.



Analyzing Quasar Spectra: α and N_H

Best fit and (90% confidence interval or credible region)
Subscripts next to obs id indicate total counts

α

Source id	XSPEC	BLoCXS	[B.3]
obs 3102$_{(2867)}$	1.892 (1.797, 1.935)	1.812 (1.770, 1.859)	1.813 (1.761, 1.874)
obs 3103$_{(4169)}$	1.879 (1.851, 1.992)	1.827 (1.778, 1.877)	1.824 (1.767, 1.887)
obs 3104$_{(10728)}$	1.842 (1.802, 1.886)	1.767 (1.735, 1.801)	1.764 (1.714, 1.815)

N_H ($10^{20}\mathrm{cm}^{-2}$)

Source id	XSPEC	BLoCXS	[B.3]
obs 3102$_{(2867)}$	0.910 (0.000, 1.823)	0.228 (0.041, 0.671)	0.243 (0.042, 0.703)
obs 3103$_{(4169)}$	1.839 (1.184, 3.510)	1.206 (0.588, 1.867)	1.188 (0.547, 1.872)
obs 3104$_{(10728)}$	6.026 (5.178, 7.173)	4.363 (3.796, 4.959)	4.326 (3.650, 5.035)

It is important to note that incorporating the arf uncertainty shows wider intervals than the Bayesian fitting without arf uncertainty.

Summary

We have developed a fast, robust, and general method to incorporate effective area calibration uncertainties in model fitting of low-resolution spectra. Because such uncertainties are ignored during spectral fits, the error bars derived for model parameters are generally underestimated. Incorporating them directly into spectral analysis with existing analysis packages is not possible without extensive case-specific simulations, but it is possible to do so in a generalized manner in a Markov chain Monte Carlo framework. We describe our implementation of this method here, in the context of recently codified Chandra effective area uncertainties. We develop our method and apply it to both simulated as well as actual Chandra ACIS-S data. We estimate the posterior probability densities of absorbed power-law model parameters that include the effects of such uncertainties. Overall, a single run of the Bayesian MCMC spectral fitting algorithm incorporates calibration uncertainty effectively.

References

Drake, J. et al. (2006) Proc. SPIE, 6270, p.49
Drake, J. et al. (2007) 2007.1 CCW, Huntsville, AL
Kashyap, V. et al. (2007) 2007.2, CCW, Huntsville, AL
van Dyk, D. et al (2001) ApJ, 548(1), p.224

Acknowledgment

This work was supported by NASA/AISRP grant NNG06GF17G and NAS8-39073 to the Chandra X-Ray Center.