

# A Fully Bayesian Method for Jointly Fitting Instrumental Calibration and X-ray Spectral Models

Jin Xu<sup>1</sup>, David A. van Dyk<sup>2</sup>, Vinay L. Kashyap<sup>3</sup>, Aneta Siemiginowska<sup>3</sup>, Alanna Connors<sup>4</sup>,  
Jeremy Drake<sup>3</sup>, Xiao-Li Meng<sup>5</sup>, Pete Ratzlaff<sup>3</sup>, and Yaming Yu<sup>1</sup>,

<sup>1</sup> Department of Statistics, University of California,  
Irvine, CA 92697-1250, USA

`jinx@uci.edu`

`yamingy@ics.uci.edu`

<sup>2</sup>Statistics Section, Imperial College London  
Huxley Building, South Kensington Campus, London SW7 2AZ, UK

`dvandyk@imperial.ac.uk`

<sup>3</sup>Smithsonian Astrophysical Observatory,  
60 Garden Street, Cambridge, MA 02138, USA

`vkashyap@cfa.harvard.edu`

`asiemiginowska@cfa.harvard.edu`

`jdrake@cfa.harvard.edu`

`pratzlaff@cfa.harvard.edu`

<sup>4</sup>deceased

<sup>5</sup>Department of Statistics, Harvard University  
One Oxford Street, Cambridge, MA 02138, USA

`meng@stat.harvard.edu`

Received \_\_\_\_\_; accepted \_\_\_\_\_

**June 9, 2014**

## ABSTRACT

Owing to a lack of robust principled methods, systematic instrumental uncertainties have generally been ignored in astrophysical data analysis despite wide recognition of the importance of including them. Ignoring calibration uncertainty can cause bias in the estimation of source model parameters and can lead to underestimation of the variance of these estimates. We previously introduced a pragmatic Bayesian method to address this problem. The method is “pragmatic” in that it introduced an ad hoc technique that simplified computation by neglecting the potential information in the data for narrowing the uncertainty for the calibration product. Following that work we use a principal component analysis to efficiently represent the uncertainty of the effective area of an X-ray (or  $\gamma$ -ray) telescope. Here, however, we leverage this representation to enable a principled, fully Bayesian method that coherently accounts for the calibration uncertainty in high-energy spectral analysis. In this setting, the method is compared with standard analysis techniques and the pragmatic Bayesian method. The advantage of the fully Bayesian method is that it allows the data to provide information not only for estimation of the source parameters but also for the calibration product—here the effective area, conditional on the adopted spectral model. In this way, it can yield more accurate and efficient estimates of the source parameters along with valid estimates of their uncertainty. Provided that the source spectrum can be accurately described by a parameterized model, this method allows rigorous inference about the effective area by quantifying which possible curves are most consistent with the data.

*Subject headings:* X-rays: general, methods: data analysis, methods: statistical, techniques: miscellaneous, standards

## 1. Background

### 1.1. Calibration Uncertainty in High-Energy Astrophysics

Observed data are always the result of the interaction of the incident spectrum with an instrument such as a telescope and detector assembly. These are described by *instrument calibration products* such as effective area curves, energy redistribution matrices, and point-spread functions. The careful specification of these calibration products is critical both for parameter fitting and for properly accounting for the statistical errors of these fits. It is only through instrument calibration that we can transform measured signals into physically meaningful quantities and have any hope of interpreting data analyses in a useful manner. Misspecification of calibration products can lead to serious bias in the fitted parameters, unreliable statistical errors, and uninterpretable results.

In practice it is well-known that instrumental properties (e.g., the quantum efficiency of a CCD detector, point-spread functions, etc.) are measured with error. Unfortunately, typical analyses only account for nominal estimates of calibration products without regard for their errors and/or their possible misspecification. This can seriously degrade fitted parameters and their error bars. In spectral analysis, for example, Drake *et al.* (2006) demonstrated that ignoring calibration uncertainty with good quality high signal-to-noise data can result in error bars that are underestimated by a factor of as much as five; see their Figure 5. We show that ignoring these errors not only is detrimental to error bars, but more importantly can seriously bias the fitted values themselves.

Efforts have been made to develop methods that account for calibration uncertainty in high-energy astrophysics and such methods exist both in other areas of astrophysics and in related fields such as particle physics (Heinrich and Lyons 2007) and observational cosmology (Bridle *et al.* 2002); see Lee *et al.* (2011) for a review. The nature of the errors in high-energy calibration products includes their complex correlations for which existing methods generally do not provide reliable results. Modern instruments such as the *Chandra X-ray Observatory* are calibrated using data from particularly well-understood

astronomical sources or from labs, and comparing these data with theoretical predictions. These measurements are not typically used directly, but rather they are used to tune sophisticated physics-based computer codes that model the instrument as a whole. These codes can be used to derive both nominal estimates of calibration products as well as measures of their uncertainty. The calibration products are high-dimensional and exhibit complex and large-scale correlation structures among their components. Accounting for this complex uncertainty is further complicated in high-energy astrophysics by the non-Gaussian nature of both the source photon counts and the instrument response. The non-Gaussian data along with the complex correlations in the calibration uncertainty mean that existing methods are not by and large applicable in this setting. The general method of combining measurement and calibration errors in quadrature (e.g., Bevington and Robinson 1992), for example, assumes uncorrelated Gaussian errors and a one-to-one relationship between calibration errors and data points. This is not appropriate in the context of the complex correlations exhibited by the wide variety of calibration products computed and recommended by calibration scientists. But a given product, taken to be error-free, can introduce serious biases into the final fitted values. Thus, calibration uncertainty must be folded into analyses in order to provide statistically principled and scientifically sound results. As we shall see, doing so can allow the data to inform the choice among the possible calibration products.

To address these complex correlations, Drake *et al.* (2006) suggested a bootstrap-like method that relies on the availability of a large representative sample of possible calibration products capturing various calibration uncertainties; we refer to this sample as the *calibration library*. In particular, they generated a replicate data set for each calibration product in the library and fit it in the usual way. The variability among the resulting fitted model parameters then estimates the effect of calibration uncertainty on the fitted parameters. While this is a useful method to demonstrate the scale of the effect of calibration uncertainty on error bars for the model parameters, generating replicated data sets requires knowledge of typically unknown model parameters. From a practical point

of view, the method’s reliance on a large calibration library is also problematic, especially considering that calibration products of space-based detectors degrade over time, and hence different calibration libraries are required for different observations.

Lee *et al.* (2011, henceforth Paper I<sup>1</sup>) developed this approach further by, first, replacing the large calibration library with a low-dimensional model for the calibration uncertainty and, second, by embedding the model for calibration uncertainty into a Bayesian procedure that simultaneously fits the model parameters and accounts for calibration uncertainty. As suggested by Drake *et al.* (2006), Paper I modeled the calibration library using a principal component analysis (PCA), but did so in a way that avoided the need for observation-specific calibration libraries. This strategy effectively embeds the instrument-modeling code via the calibration library as an integral part of the statistical computing techniques. This approach, however, supposes purely for simplicity that the observed photon counts and the calibration product are conditionally independent, that is, that the data provide no information for narrowing the calibration uncertainty. An advantage of this independence assumption is that it significantly simplifies the computing, making the algorithm easy to implement. For this reason in Paper I, we call this approach a *pragmatic Bayesian method*. The independence assumption of the pragmatic Bayesian method also implies that the choice of calibration product is determined by calibration scientists, calibration experiments, and calibration simulations, rather than the data from a particular observation. This may be viewed as an advantage by some researchers, particularly if they do not trust the post-observation model assumptions.

---

<sup>1</sup>We refer to Lee *et al.* (2011) as Paper I simply for easy reference; both the current paper and Lee *et al.* rely on earlier work that characterize calibration uncertainty (Drake *et al.* 2006) and lay out the framework of MCMC-based Bayesian spectral analysis (van Dyk *et al.* 2001). In contrast to van Dyk *et al.*, who use Gibbs-based MCMC, both Paper I and this work adopt Metropolis and Metropolis-Hastings type MCMC for the sake of flexibility in astrophysical model specification.

The primary objective of this article is to remove the questionable independence assumption of the pragmatic Bayesian approach and allow the data to narrow the calibration uncertainty. This is a more principled approach from a statistical perspective. If a subset of the calibration products are plausible before seeing the data but inconsistent with the data once it is observed, this subset should not play a role in the final analysis. We call this a *fully Bayesian method* because it fully follows the principles of Bayesian analysis. This approach has other advantages. For example, in addition to the calibration uncertainty quantified through idealized experiments, calibration products are subject to errors stemming from differences between these idealized settings and the variety of actual settings in which the products are used. Indeed, suspected systematic errors cannot be fully understood without taking into account the actual data in any particular observation and/or cross-instrument comparisons that may be made (e.g., Nevalainen *et al.* 2010). Our fully Bayesian method allows the data to inform our choice of possible calibration products. In practice, we find that relatively large data sets ( $\gg 10^4$  counts) are needed to obtain appreciable power in narrowing calibration uncertainty.

Like the pragmatic Bayesian method, the fully Bayesian approach embeds a model derived from a PCA of the calibration library into a larger statistical model. Unlike the pragmatic Bayesian method, however, it then marginalizes over calibration uncertainty *while conditioning on the observed data*, whereas the pragmatic method did so *without conditioning on the data*. In this regard, the fully Bayesian method is in line with a methods proposed by Bridle *et al.* (2002) and Heinrich and Lyons (2007) for handling systematic errors in cosmology and particle physics, respectively. These proposals, however, use a parameterized form for the systematics under which marginalization can be achieved analytically. While these specific proposals are not applicable in our setting, they share our emphasis on the general principle of building a joint model that incorporates all sources of uncertainty and then marginalizing over nuisance parameters while conditioning on the observed data.

The statistical framework we present is quite general, but in this paper we focus on

X-ray spectral analysis with uncertainty in the effective area curve. We begin by outlining the necessary background on Bayesian spectral analysis, the PCA-based calibration model proposed in Paper I, and the pragmatic Bayesian method. Section 2 lays out our fully Bayesian method and uses a simple numerical example to illustrate its advantage over both the typical strategy of ignoring calibration uncertainty and the pragmatic Bayesian method. Section 3 validates our proposal using a set of simulation studies that include a comprehensive frequency evaluation. We find striking improvement in the statistical properties of estimates and error bars, when large-count spectra ( $\gg 10^4$  counts) are generated with an effective area curve consistent with the pre-specified calibration uncertainty but the spectra are fit with a misspecified default curve. Section 4 illustrates how the fully Bayesian method works for several data from the *Chandra* telescope: a collection of QSO’s observed near the aimpoint of ACIS-S and described with absorbed power-law models; a bright O star system at a large off-axis location on ACIS-S2 and modeled as absorbed multi-thermal spectra; and co-added long-duration grating observation of an isolated neutron star modeled as a blackbody spectrum. Finally, we discuss the consequences and future directions of our method in Section 5. An appendix details the Markov chain Monte Carlo (MCMC) algorithms we designed to implement our method; they rely on the `pyBLoCXS` module (Siemiginowska *et al.* 2011) in *Sherpa* and existing algorithms for the pragmatic Bayesian method.<sup>2</sup> A glossary of the symbols we use appears in Table 1.

---

<sup>2</sup>A version of our software that implements the pragmatic Bayesian fitting of Paper I is currently available within *CIAO/Sherpa*. A newer version that includes both the pragmatic and fully Bayesian algorithms will soon be made publicly available at <http://github.com/astrostat/pyblocxs/> and eventually made available within *CIAO/Sherpa*.

Table 1: Glossary of symbols used in the text

Symbol	Description
$A$	an effective area (ARF) curve
$A_0$	the default effective area curve on which the calibration library is based.
$A_0^*$	the observation specific effective area curve.
$A_l$	effective area curve $l$ in the calibration library
$\bar{A}$	the average of the effective area curves in the calibration library
$A^{\text{prop}}$	a proposed effective area curve in an MH sampler
$A_{\text{pB}}^{(t+1)}, A_{\text{fB}}^{(t+1)}$	an effective area curve simulated with the pragmatic and fully Bayesian sampler
$\mathcal{A}$	a set of effective areas, the calibration library
$\mathcal{CI}$	a confidence interval
$E$	energy of incident photon
$E^*$	Detector channel at which the detector registers the incident photon
$e$	the low-dimensional PCA representation of $A$ , see Equation 5
$e^{\text{prop}}$	value of $e$ proposed in an MH sampler
$e_{\text{fB}}^{(t+1)}, e_{\text{pB}}^{(t+1)}$	value of $e$ simulated with the pragmatic and fully Bayesian samplers
$g$	generic proposal distribution in a MH sampler
$I$	number of inner iterations in MCMC loop, typically 10
$i$	inner iteration number or index
$\mathcal{I}$	information obtained prior to data acquisition, for example by calibration scientists
$J$	number of components used in PCA analysis, here 8
$j$	principal component number or index
$\mathcal{K}$	an MCMC kernel
$\mathcal{K}_{\text{pyB}}$	the MCMC kernel used in pyBLocXS
$L$	number of replicate effective area curves in calibration library
$l$	replicate effective area number or index
$\mathcal{L}$	the likelihood function
$M$	number of replicated draws of $\theta$ per draw of $A$ in Iterated MH within PCG sampler
$p$	a generic prior or posterior distribution
$p_{\text{std}}, p_{\text{pB}}, p_{\text{fB}}$	standard, pragmatic Bayesian, and fully Bayesian posterior distributions
$R$	energy redistribution matrix (RMF)
$r_j^2$	eigenvalue or PC coefficient of component $l$ in the PCA representation
$T$	number of MCMC iterations
$t$	main MCMC iteration number or index
$(t+\dots)$	the superscript indicates the running index of random draws
$u$	a uniformly distributed random number between zero and one
$v_j$	eigen- or feature-vector for component $l$ in the PCA representation
$Y$	data, typically used here as counts spectra in detector PI bins
$\hat{\alpha}$	correlation used to validate the choice of $I$ in the MH within PCG sampler
$\rho$	the acceptance probability in an MH sampler
$\theta$	spectral model parameters
$\hat{\theta}$	estimate of $\theta$
$\theta^{\text{prop}}$	a value of $\theta$ proposed in an MH sampler
$\theta_{\text{std}}^{(t+\dots)}, \theta_{\text{fB}}^{(t+\dots)}, \theta_{\text{pB}}^{(t+\dots)}$	value of $\theta$ simulated with standard, pragmatic Bayesian, and fully Bayesian samplers
$\psi$	generic notation for unknown quantities in analysis, viz., $\psi = \theta$ or $\psi = (A, \theta)$
$\psi^{\text{prop}}$	a value of $\psi$ proposed in an MH sampler
$\hat{\sigma}_{\text{std}}, \hat{\sigma}_{\text{pB}}, \hat{\sigma}_{\text{fB}}$	error bars under the standard, pragmatic, and fully Bayesian methods



## 1.2. Bayesian Spectral Analysis

To fix ideas, we focus attention on a general spectral analysis problem in which the observed photon count in channel  $E^*$  is modeled according to the Poisson distribution,<sup>3</sup>

$$Y(E^*) \sim \text{POISSON} \left( \sum_E \Lambda(E; \theta) A(E) R(E^*; E) + B(E^*) \right), \quad (1)$$

where  $\Lambda(E; \theta)$  is the source spectral intensity at energy  $E$ ,  $\theta$  is the (vector) spectral parameter,  $A(E)$  is the effective area curve at energy  $E$ ,  $R(E^*; E)$  is the energy redistribution matrix of the detector, and  $B(E^*)$  is the background intensity in channel  $E^*$ . The photon counts in each channel,  $Y(E^*)$  are independent Poisson variables. For notational simplicity, we represent the collection of observed photon counts by  $Y = \{Y(E^*)\}$ , the effective area by  $A = \{A(E)\}$ , and the photon redistribution matrix by  $R = \{R(E^*; E)\}$ . We focus on methods that treat  $A$  as an unknown quantity and allow its uncertainty to affect both the fit and the error bars of  $\theta$ . Although our methods can be employed to account for uncertainty in  $R$  or in both  $A$  and  $R$ , we do not pursue these possibilities in this article.

To fit the spectral parameters,  $\theta$ , given the observed photon counts,  $Y$  while accounting for calibration uncertainty, we adopt a Bayesian framework. In particular, we quantify our state of knowledge before having seen the data using a so-called *prior distribution* and that after having seen the data using a *posterior distribution*. Bayes' Theorem allows us to transform the prior distribution into the posterior distribution by conditioning on the observed counts. In particular, suppose  $\psi$  represents the quantities that are to be estimated, and  $\mathcal{I}$  is the information we have before seeing the data, including that used to estimate the calibration products and/or their uncertainty. We compare methods that treat only  $\theta$  as unknown with others that treat both  $\theta$  and  $A$  as unknown, so either  $\psi = \theta$  or  $\psi = (\theta, A)$ ,

---

<sup>3</sup>The notation  $Y \sim \text{DISTRIBUTION}$  is a standard way in statistics literature to indicate that a variable (e.g., the observed counts) is modeled using the specified distribution (e.g., Poisson).

while  $R$  (and sometimes  $A$ ) is part of  $\mathcal{I}$ . In this setting Bayes' Theorem states that the posterior distribution of  $\psi$  given  $Y$  and  $\mathcal{I}$  is:

$$p(\psi|Y, \mathcal{I}) = \frac{\mathcal{L}(Y|\psi, \mathcal{I}) p(\psi|\mathcal{I})}{p(Y|\mathcal{I})}, \quad (2)$$

where  $p(\psi|\mathcal{I})$  is the prior distribution of  $\psi$ ,  $\mathcal{L}(Y|\psi, \mathcal{I})$  is the likelihood of  $Y$  given  $\psi$ , and  $p(Y|\mathcal{I}) = \int p(Y|\psi, \mathcal{I})p(\psi|\mathcal{I})d\psi$  is the normalizing constant that ensures  $p(\psi|Y, \mathcal{I})$  integrates to one. We include  $\mathcal{I}$  here to emphasize that analyses *always* rely on some information external to  $Y$ . To simplify notation, however, we assume all probability distributions are conditional on  $\mathcal{I}$  and omit it from our notation for the remainder of the article. We also implicitly assume that the adopted spectral model is appropriate. Its misspecification can lead to biases in addition to those caused by misspecification of the calibration products.

Substituting  $\psi = (\theta, A)$  into Equation 2 and assuming the prior distributions for  $\theta$  and  $A$  are independent, we can write the posterior distribution as

$$p(\theta, A|Y) \propto \mathcal{L}(Y|\theta, A) p(\theta) p(A), \quad (3)$$

where we have omitted the denominator of Equation 2 because it is determined by the numerator. We typically use a diffuse prior distribution on  $\theta$  representing prior ignorance and an informative prior distribution on  $A$  representing the information obtained from calibration studies.

Whereas our primary goal is to consider methods for joint inference for  $\theta$  and  $A$  using Equation 3, we also compare such methods with the standard approach that treats  $A$  as fixed and known. For clarity we refer to this approach as the *standard method*. For example, in a Bayesian analysis (e.g., van Dyk *et al.* 2001), the standard method involves estimating  $\theta$  using its posterior distribution given the observation,  $Y$ , and the nominal effective area curve associated with this observation,  $A_0^*$ , that is, using

$$p_{\text{std}}(\theta | Y, A_0^*) \propto \mathcal{L}(Y | \theta, A_0^*) p(\theta). \quad (4)$$

Because this approach assumes that  $A = A_0^*$ , it does not account for calibration uncertainty. Paper I illustrates that this can lead to misleading estimates of  $\theta$  and can significantly

underestimate the error bars associated with these estimates. Nevertheless, because this is the standard approach in practice, we treat it as a baseline in our numerical comparisons.

### 1.3. Quantifying Calibration Uncertainty

The specification of the posterior distribution in Equation 3 requires that we formulate a prior distribution on  $A$  that encapsulates the calibration uncertainty. Although they were not working in a Bayesian setting, Drake *et al.* (2006) generated a library of ACIS effective area curves. This was accomplished by explicitly including uncertainties in each of the subsystems of the telescope (UV/ion shield transmittance, CCD quantum efficiency, and the telescope mirror reflectivity) using truncated Gaussian distributions for the parameters of different instrument models and by modifying a spline correction curve that multiplies a default curve. More recently, we compiled a second calibration library to represent uncertainty in the LETGS+HRC-S grating/detector system (Drake et al., in preparation). This includes corrections applied to the telescope mirror reflectivity, grating obscuration and efficiency, UV/ion shield transmittance, MCP QE and uniformity, etc. Additionally, spline knots were set at all prominent spectral edges due to materials that were used in the construction of the telescope, grating, and detector. In our numerical studies, we illustrate how either of these libraries can be incorporated into a fully Bayesian analysis. Both consist of  $L = 1000$  simulated effective area curves, each of length 1078 in the ACIS library and each of length 16,384 in the LETGS+HRC-S library. The former is used in Sections 4.1 and 4.2, and the latter in Section 4.3. In our general notation, we represent a calibration library by  $\mathcal{A} = \{A_1, A_2, \dots, A_L\}$ , define  $\bar{A}$  to be the arithmetic mean curve of the calibration library, and let  $A_0$  denote the default effective area curve associated with the library;  $\bar{A}$  and  $A_0$  are similar but not necessarily equal.

In practice, the calibration library must be large to fully represent the uncertainty in high-dimensional calibration products. To summarize this sample into a concise and usable form, Paper I implemented a Principal Component Analysis (PCA) on the mean-subtracted

calibration sample,  $\{A_1 - \bar{A}, \dots, A_L - \bar{A}\}$ . PCA is a mathematical procedure that uses orthogonal transformations to convert a set of possibly correlated variables into a set of linearly uncorrelated variables called principal components. Approximately 8 (20) principal components (out of 1000) account for 97% (99%) of the variability in the ACIS calibration library.

As in Paper I, we conduct a Bayesian analysis that treats  $A$  and  $\theta$  as unknown. We use the PCA summary of the calibration library to formulate the prior distribution for  $A$ ,  $p(A)$ . In particular, we suppose that under  $p(A)$ ,

$$A(e) = A_0^* + (\bar{A} - A_0) + \sum_{j=1}^J e_j r_j v_j, \quad (5)$$

where  $A_0^*$  is the user-generated observation-specific effective area curve,  $r_j^2$  and  $v_j$  are the principal component eigenvalues and eigenvectors, and  $e_j$  are independent standard normal deviations.<sup>4</sup> Since  $\bar{A} \approx A_0$ , we can view Equation 5 as starting with the user-specified effective area,  $A_0^*$ , and adding the random term  $\sum_{j=1}^J e_j r_j v_j$  to account for uncertainty;  $\bar{A} - A_0$  adjusts for the necessary mean-subtraction of  $\mathcal{A}$  when conducting PCA. To simulate replicate effective area curves under the prior distribution given in Equation 5, we only need to draw  $J$  independent standard normal deviations,  $(e_1, \dots, e_J)$ , and evaluate Equation 5. We treat  $A(e)$  as the generic notation for the effective area curve and continue to simply write  $A$  when its explicit dependence on  $e$  is not pertinent.

Using Equation 5 to summarize the calibration library involves several assumptions. First we assume that the uncertainty in  $A$  can be described by a multivariate normal distribution. The similarity of the effective area curves in  $\mathcal{A}$  means that most of the correlations among the components of this distribution are very strong (i.e. near one).

---

<sup>4</sup>An additional residual term,  $\xi = \sum_{j=J+1}^L r_j v_j$ , may also be included in Equation 5. Adding  $e_{J+1}\xi$  can help to account for the full range of calibration uncertainty when  $J$  is small, or for components that contribute significantly over small energy ranges, yet make up a small fraction of the overall variance of  $A$ .

Equation 5 stipulates that the distributions associated with calibration uncertainty for observation-specific effective area curves differ only in their means and that they have the same variance. This means we can use the variance of the calibration library of Drake *et al.* (2006) to represent the variance of any observation-specific effective area. In practice, this procedure avoids generating a calibration library for each observation while still allowing us to account for calibration uncertainty in practical manner.

Figure 1 illustrates the performance of PCA in summarizing the structure of individual effective area curves in  $\mathcal{A}$ . It shows that with  $J = 8$ , the re-constructed effective area curve nicely captures the structure of the original  $A \in \mathcal{A}$ . This means that we can capture the vector  $A$  of length 1078 with the vector  $e$  of length 8, by using  $A(e)$ . We use  $J = 8$  in all of our numerical studies.

This PCA-based representation,  $A(e)$ , is critical both for the pragmatic Bayesian method of Paper I and the fully Bayesian method described here. It not only provides a simple way to quantify calibration uncertainty, but also allows us to evaluate both  $p(A, \theta | Y)$  and  $p(A)$  using Equations 3 and 5. The evaluation of both of these distributions is necessary for our MCMC sampler.

#### 1.4. A Pragmatic Bayesian Method

As noted above (see Equation 4), standard analyses assume that the effective area is fixed. That is, the parameters are estimated *conditional* on  $A_0^*$ . Here, we aim to eliminate this conditioning. Mathematically, this involves treating  $A$  as unknown rather than conditioning on its value, and expressing Equation 3 as

$$p_{\text{fB}}(\theta, A | Y) = p(\theta | A, Y) p(A | Y); \tag{6}$$

here the subscript fB indicates that this is the fully Bayesian posterior distribution. Paper I made the “pragmatic assumption” that  $p(A | Y) = p(A)$ . This assumption says that the observed data and calibration are independent, that is, the data provide no information for

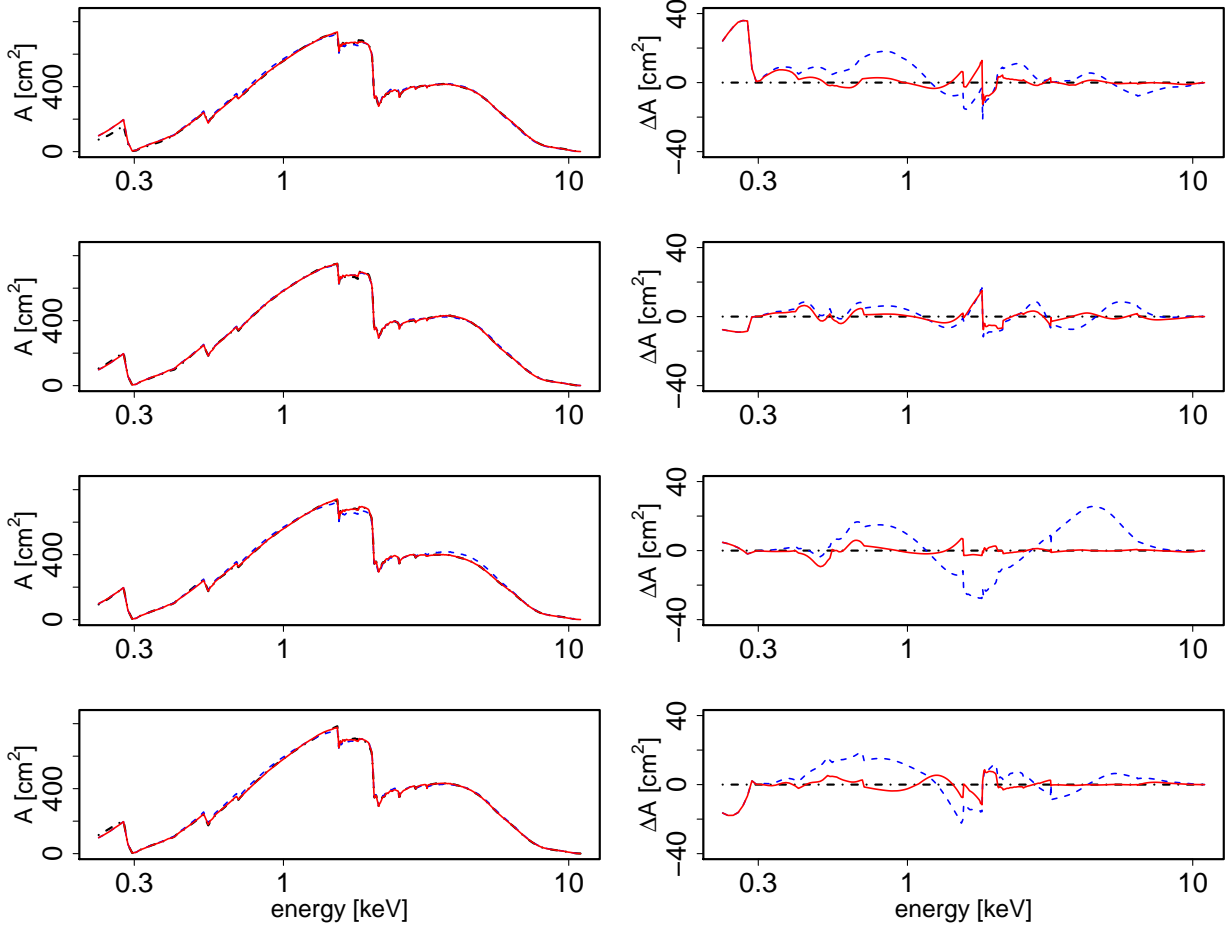


Fig. 1.— The PCA representation of several effective area curves from the calibration library. The left column plots four randomly selected  $A_l \in \mathcal{A}$ , one in each row, along with their PCA representation,  $A_l(e)$ , for several values of  $J$ . The original curves are plotted in dot-dashed black and the PCA representations are plotted as dashed blue and solid red for  $J = 1$  and 8, respectively. The right column is constructed in the same manner, but plots the error in our approximation,  $\Delta A_l = A_l(e) - A_l$ . Although  $A_l(e)$  deviates from  $A_l$ , even with  $J = 8$ , the right column shows that the scale of this deviation is quite small and that overall using  $J = 8$  concisely captures the structure of each of the effective area curves.

narrowing the uncertainty in the choice of effective area curve. Under this assumption, the posterior distribution in Equation 6 can be written as

$$p_{\text{pB}}(\theta, A | Y) = p(\theta | A, Y) p(A) \quad (7)$$

where  $p(\theta | A, Y)$  is given in Equation 4 with  $A_0^*$  replaced by the generic  $A$ . We use the subscript pB in Equation 7 to emphasize that this is the posterior distribution under the pragmatic assumption of Paper I. Under the model in Equation 7, inference for  $\theta$  is based on its marginal posterior distribution,

$$p_{\text{pB}}(\theta | Y) = \int p(\theta | A, Y) p(A) dA. \quad (8)$$

The pragmatic Bayesian method accounts for calibration uncertainty in a conservative manner. The assumption that  $p(A | Y) = p(A)$  ignores information in the data that may reduce uncertainty in  $A$  and hence in  $\theta$ . We now consider methods that allow  $Y$  to narrow the uncertainty of  $A$ .

## 2. A Fully Bayesian Solution

### 2.1. Motivation and Theory

Under the fully Bayesian posterior distribution, the marginal distribution of  $\theta$  given in Equation 8 is replaced with

$$p_{\text{fB}}(\theta | Y) = \int p_{\text{fB}}(\theta, A | Y) dA, \quad (9)$$

where  $p_{\text{fB}}(\theta, A | Y) = p(\theta | A, Y) p(A | Y)$ . The fully Bayesian posterior distribution of  $\theta$  can be viewed as a weighted version of that under the pragmatic posterior:

$$\begin{aligned} p_{\text{fB}}(\theta | Y) &= \int p(\theta | A, Y) \frac{p(A | Y)}{p(A)} p(A) dA \\ &\approx \frac{1}{T} \sum_{t=1}^T p(\theta | A^{(t)}, Y) \frac{p(A^{(t)} | Y)}{p(A^{(t)})}, \end{aligned} \quad (10)$$

where  $A^{(t)}$  is drawn from  $p(A)$ . In principle, we could implement the fully Bayesian method by obtaining a sample  $\{A^{(t)}, t = 1, \dots, T\}$  from  $p(A)$  and evaluating Equation 10. (This technique is known as importance sampling.) Unfortunately, evaluating  $p(A | Y) = \int p(\theta, A | Y) d\theta$  is extremely challenging in this setting. Alternative computational techniques therefore are needed; we outline our strategy in Section 2.3 and give details in Appendix A.

The posterior distribution of  $A$  under the pragmatic method,

$$p_{\text{pB}}(A | Y) = \int p_{\text{pB}}(\theta, A | Y) d\theta = \int p(\theta | A, Y) p(A) d\theta = p(A),$$

is simply equal to the prior distribution of  $A$ . The fully Bayesian posterior for  $A$ ,

$$p_{\text{fB}}(A | Y) = \int p_{\text{fB}}(\theta, A | Y) d\theta = \int p(\theta | A, Y) p(A | Y) d\theta = p(A | Y),$$

however, can be used to learn what effective area curves are more or less consistent with the observed data. We illustrate how this is done in Section 3.

## 2.2. The Advantage of a Fully Bayesian Analysis

To illustrate the advantage of the fully Bayesian method over the pragmatic Bayesian method, we compared their performance in a simulation study. In Section 3 we reproduce part of the simulation study of Paper I but this time including the fully Bayesian method. Here we give detailed results under the simulation setting called Simulation II in Section 3. In particular, we simulate an absorbed power-law source model with three parameters (power-law index  $\Gamma$ , absorption column density  $N_{\text{H}}$ , and normalization) using the `fake_phi` routine in Sherpa (Freeman *et al.* 2001). The data set was simulated without background contamination, using the XSPEC model `wabs*powerlaw` and a default photon redistribution matrix (RMF) for ACIS-S. We consider the energy range from 0.3 keV to 7 keV, which is divided evenly into approximately one thousand bins. Here we give detailed results for Simulation II which set  $\Gamma = 1$ ,  $N_{\text{H}} = 10^{21} \text{cm}^{-2}$ , generated  $10^5$  counts, and used an extremal effective area curve from the calibration library of Drake *et al.* (2006). (We use the same



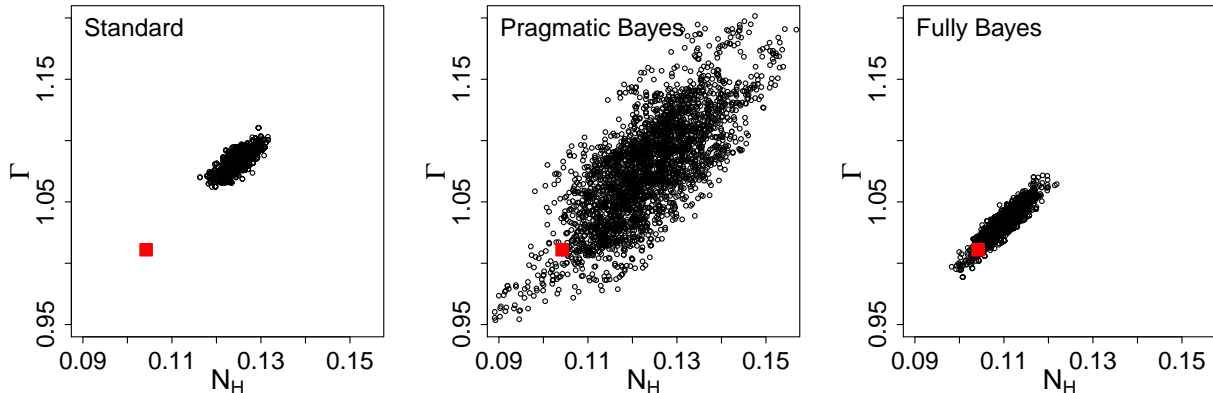


Fig. 2.— Comparison of the standard (left), pragmatic Bayesian (middle), and fully Bayesian (right) methods. Each panel compares a complete Monte Carlo posterior sample of values of  $\theta = (\Gamma, N_{\text{H}})$  to its true value, marked with a red square;  $N_{\text{H}}$  is shown in units of  $10^{22} \text{ cm}^{-2}$ . When fit using an incorrect effective area curve, the standard method can result in misleading estimates of  $\theta$  (see left panel). The pragmatic Bayesian method, on the other hand, averages over all *a priori* possible effective area curves and significantly enlarges the posterior variance for  $\theta$  (see middle panel). Although the centers of the posterior distributions under the standard and pragmatic Bayesian methods are similar, the larger error bars computed with the latter allows them to include the true values of  $\theta$ . Finally, the posterior distribution under the fully Bayesian method shifts toward the true value of  $\theta$ , allowing it to cover the true value while avoiding the large error bars of the pragmatic Bayesian method (see right panel).

extremal effective area curve as in Paper I and label it  $A_{\text{ext}}$ ; it is the curve with the highest value of the effective area in the calibration library.)

Figure 2 plots a Monte Carlo sample from the posterior distribution under the standard method,  $p_{\text{std}}(\theta | A_0, Y)$ , the pragmatic Bayesian method  $p_{\text{pB}}(\theta | Y)$ , and the fully Bayesian method,  $p_{\text{fB}}(\theta | Y)$ , where  $A_0$  is the default under the calibration library of Drake *et al.* (2006) and here  $\theta = (\Gamma, N_{\text{H}})$ . The red square in each panel gives the true value of  $\theta$ . The MCMC samplers used in Figure 2 are described in Appendix A. Although the error bars computed with the standard method are the smallest, in this simulation the method misses the true value of  $\theta$ . *The results are precise but inaccurate.* This is not unusual when the

default effective area curve,  $A_0$ , is misspecified, as it is in this case because the data were generated under a different curve—one that is nonetheless plausible given the calibration uncertainty. The pragmatic method accounts for calibration uncertainty by averaging over all *a priori* possible effective area curves, resulting in much larger error bars that capture the true value of  $\theta$ —the method is *imprecise but accurate*. As pointed out in Paper I, this is a clear advantage over the standard method. Finally, the fully Bayesian method accounts for calibration uncertainty by averaging over those effective area curves that are consistent with the observed data. The resulting error bars are only slightly larger than those produced with a fixed effective area curve, but the fitted values for  $\theta$  have shifted enough that the error bars still capture the true value. This example clearly illustrates the benefits of the fully Bayesian method: we can characterize the performance of the standard method as “*precise but inaccurate*,” of the pragmatic method as “*imprecise but accurate*,” and of the fully Bayesian method as both “*precise and accurate*.”

### 2.3. Fitting the Fully Bayesian Model

We use a Metropolis-Hastings (MH) MCMC sampler to obtain a correlated sample,  $\{(\theta_{\text{fB}}^{(t)}, A_{\text{fB}}^{(t)}), t = 1, \dots, T\}$ , from  $p_{\text{fB}}(\theta, A | Y)$ . An introduction to MCMC, MH samplers, and the other computational methods that we employ appears in Appendix A.1. We use a type of MH sampler that is known as an *independence sampler*. It uses an approximation,  $g$ , to  $p_{\text{fB}}(\theta, A | Y)$  that does not depend on  $(\theta_{\text{fB}}^{(t)}, A_{\text{fB}}^{(t)})$  as its proposal distribution when sampling  $(\theta_{\text{fB}}^{(t+1)}, A_{\text{fB}}^{(t+1)})$ .

For such a sampler to work efficiently, any range of values that has appreciable probability under  $p_{\text{fB}}(\theta, A | Y)$  must also have appreciable probability under  $g$ . This is because the only values that can be simulated with an independence sampler are those values that can be simulated with  $g$ . If there is a range of values that has negligible probability under  $g$ , values in this range are unlikely to be simulated. To be sure that no important values of  $(\theta, A)$  are missed, it is critical that  $g$  be an *over-dispersed* approximation

to  $p_{\text{fB}}(\theta, A | Y)$ . That is  $g$  must be a reasonable approximation to  $p_{\text{fB}}(\theta, A | Y)$ , but with more proclivity to extreme values.

Fortunately, the pragmatic posterior distribution,  $p_{\text{pB}}(\theta, A | Y)$ , provides just such an over-dispersed approximation to  $p_{\text{fB}}(\theta, A | Y)$ . This is illustrated numerically in Figure 2 in which the posterior distribution under the pragmatic Bayesian model (middle panel) is clearly an over-dispersed approximation to that under the fully Bayesian model (right panel). More generally, the pragmatic Bayesian approach includes all values of  $A$  that are possible under  $p(A)$ , whereas the fully Bayesian approach more heavily weighs those values that are more consistent with  $Y$ . Thus, the fully Bayesian posterior distribution focuses on a narrow range of  $A$ . This in turn leads to a narrower range of values of  $\theta$  that are plausible under  $p_{\text{fB}}(\theta, A | Y)$  than under  $p_{\text{pB}}(\theta, A | Y)$ .

Unfortunately, although  $p_{\text{pB}}(\theta, A | Y)$  is an ideally suited over-dispersed approximation to  $p_{\text{fB}}(\theta, A | Y)$ , it cannot be used directly as the proposal distribution,  $g$ , because it is difficult to evaluate. (We need to evaluate  $g$  in order to compute  $\rho$  given below in Equation 11.) In Appendix A.3 we discuss this difficulty and how we construct a simple approximation to  $p_{\text{pB}}(\theta, A | Y)$  that is a suitable choice for  $g$ .

Because  $g$  is based on the pragmatic posterior distribution, we call it a *pragmatic proposal*. Using Equation 5,  $A$  is completely represented by the low-dimensional  $e$ , so we need only sample  $e$  and  $\theta$  and can compute  $A(e)$ . Thus, we write our approximation as  $g(\theta, e)$ . Specifically, our sampler for the fully Bayesian model is

### Pragmatic Proposal Sampler

For  $t = 0, 1, \dots, T$ ,

**Step 1:** Draw  $(\theta^{\text{prop}}, e^{\text{prop}}) \sim g(\theta, e)$ , set  $A^{\text{prop}} = A_0^* + (\bar{A} - A_0) + \sum_{j=1}^J e_j^{\text{prop}} r_j v_j$ , and compute

$$\rho = \frac{p_{\text{fB}}(\theta^{\text{prop}}, A^{\text{prop}} | Y) g(\theta_{\text{fB}}^{(t)}, e_{\text{fB}}^{(t)})}{p_{\text{fB}}(\theta_{\text{fB}}^{(t)}, A_{\text{fB}}^{(t)} | Y) g(\theta^{\text{prop}}, e^{\text{prop}})}. \quad (11)$$

**Step 2:** Let  $u$  be a uniformly distributed random number between zero and one and set

$$(\theta_{\text{fB}}^{(t+1)}, e_{\text{fB}}^{(t+1)}, A_{\text{fB}}^{(t+1)}) = \begin{cases} (\theta^{\text{prop}}, e^{\text{prop}}, A^{\text{prop}}) & \text{if } u < \rho \\ (\theta_{\text{fB}}^{(t)}, e_{\text{fB}}^{(t)}, A_{\text{fB}}^{(t)}) & \text{otherwise} \end{cases}.$$

In our numerical studies, we make one final adaptation. At each iteration we either use the update described in the Pragmatic Proposal Sampler with  $g$  set to the approximation to  $p_{\text{pB}}(\theta, A | Y)$  described in Appendix A.3, or we conduct a random walk update to  $(\theta, e)$  by replacing Step 1 of the sampler with

**Alternative Step 1:** Draw  $e_j^{\text{prop}} = e_j^{(t)} + \mathcal{N}(0, \sigma)$  for  $j = 1, \dots, J$  and  $\theta^{\text{prop}} \sim g(\theta | e^{\text{prop}})$ , set  $A^{\text{prop}} = A_0^* + (\bar{A} - A_0) + \sum_{j=1}^J e_j^{\text{prop}} r_j v_j$ , and compute

$$\rho = \frac{p_{\text{fB}}(\theta^{\text{prop}}, A^{\text{prop}} | Y) g(\theta_{\text{fB}}^{(t)} | e_{\text{fB}}^{(t)})}{p_{\text{fB}}(\theta_{\text{fB}}^{(t)}, A_{\text{fB}}^{(t)} | Y) g(\theta^{\text{prop}} | e^{\text{prop}})}.$$

The choice between the two versions of Step 1 is made randomly at each iteration of the sampler. Mixing proposal distributions in this way tends to improve the convergence of MCMC samplers. We use this *Mixed Pragmatic Proposal Sampler* in all of our numerical studies.

### 3. Simulation Studies

In this section we use a series of simulation studies to demonstrate the circumstances under which the fully Bayesian method is advantageous. In Section 4 we illustrate the method in real data analyses.

#### 3.1. Replicating the Simulation Study of Paper I

We begin by replicating four of the eight of the simulation studies of Paper I, but this time using the fully Bayesian method. The four simulations are conducted exactly as the

Table 2: The four simulations used to compare the standard, pragmatic Bayesian, and fully Bayesian methods.

	Nominal Counts		Spectral Model	
	$10^5$	$10^4$	Hard <sup>†</sup>	Soft <sup>‡</sup>
SIMULATION I	X		X	
SIMULATION II	X			X
SIMULATION III		X	X	
SIMULATION IV		X		X

<sup>†</sup>An absorbed powerlaw with  $\Gamma = 2$ ,  $N_{\text{H}} = 10^{23} \text{ cm}^{-2}$

<sup>‡</sup>An absorbed powerlaw with  $\Gamma = 1$ ,  $N_{\text{H}} = 10^{21} \text{ cm}^{-2}$

one in Section 2.2 but data were generated with different spectral models and different nominal counts (see Table 2). Specifically, the simulation represents a  $2 \times 2$  design with the two factors being

1. either  $10^5$  or  $10^4$  photon counts are simulated, and
2. data are generated under either a hard spectral powerlaw model ( $\Gamma = 2$ ,  $N_{\text{H}} = 10^{23} \text{ cm}^{-2}$ ) or a soft spectral powerlaw model ( $\Gamma = 1$ ,  $N_{\text{H}} = 10^{21} \text{ cm}^{-2}$ ).

Following Paper I, all four simulated data sets were generated with the extremal effective area curve,  $A_{\text{ext}}$ , from the calibration library of Drake *et al.* (2006). This curve differs substantially from the default curve used in the standard method and is extreme within the calibration library, but is nonetheless consistent with the uncertainty described by the library. We also examined four additional simulations in Paper I where the simulated data were generated using the default effective area curve. We exclude discussion of these simulations here because, first, it is unrealistic to assume that the true effective area is known, and second, if such an assumption is made, it only shows, as expected, the standard method performs the best; see Xu (2014) for details. In practice, the effective area curve is specified with uncertainty and it is not feasible to correctly specify its value. The four

simulations we conduct (Simulations I–IV) are numbered Simulations 5–8 in Paper I.

We fit an absorbed power law model to each of the four simulated data sets, using the standard, pragmatic Bayes, and fully Bayesian methods. In all cases the standard method is run using the (misspecified) default effective area. This enables us to investigate the effect of misspecification of  $A_0$ , where the degree of misspecification is consistent with the range of variability of the calibration library. Posterior distributions, intervals, and fitted values for  $\Gamma$ , computed with each of three methods, run on each simulation, appear in Figure 3. (We use posterior means and central 68% posterior intervals for fitted values and error bars throughout the paper.) The joint posterior distribution of  $\Gamma$  and  $N_H$  under Simulation II for each of the three methods is discussed in Section 2.2 and appears in Figure 2.

In all four simulations, the standard method produces significantly narrower error bars than the other methods, especially in the large count Simulations I and II. Unfortunately, however, as expected its intervals miss the true value of  $\Gamma$  by a large margin. The pragmatic Bayesian method, by contrast, exhibits similar fitted values but much wider error bars that reflect the variability in the fits resulting from different choices of  $A$  in the range of  $\mathcal{A}$ . Finally, the fully Bayesian method uses the data to exclude some  $A$  in the range of  $\mathcal{A}$  that are inconsistent with the observed spectra. The result is optimal in that the fitted values shift toward the true value of  $\Gamma$  and the width of the error bars narrow relative to those produced with the pragmatic Bayesian method. Thus, like the pragmatic Bayesian method, the fully Bayesian method provides error bars that contain the true value at nearly the correct statistical rate (68.2% here), but these error bars are narrower than those provided by the pragmatic Bayesian method—and can be much narrower with large counts. We conduct a large-scale simulation study in Section 3.2 to better quantify these trends.

The results of the simulations can be understood by considering the statistical errors (due to Poisson fluctuations in the counts) and the systematic errors (due to misspecification of the effective area curve). The standard method ignores the latter sources of error and thus, not surprisingly, it provides misleadingly narrow errors and can exhibit significant bias if the misspecification of  $A$  is substantial. Because it only considers statistical errors,

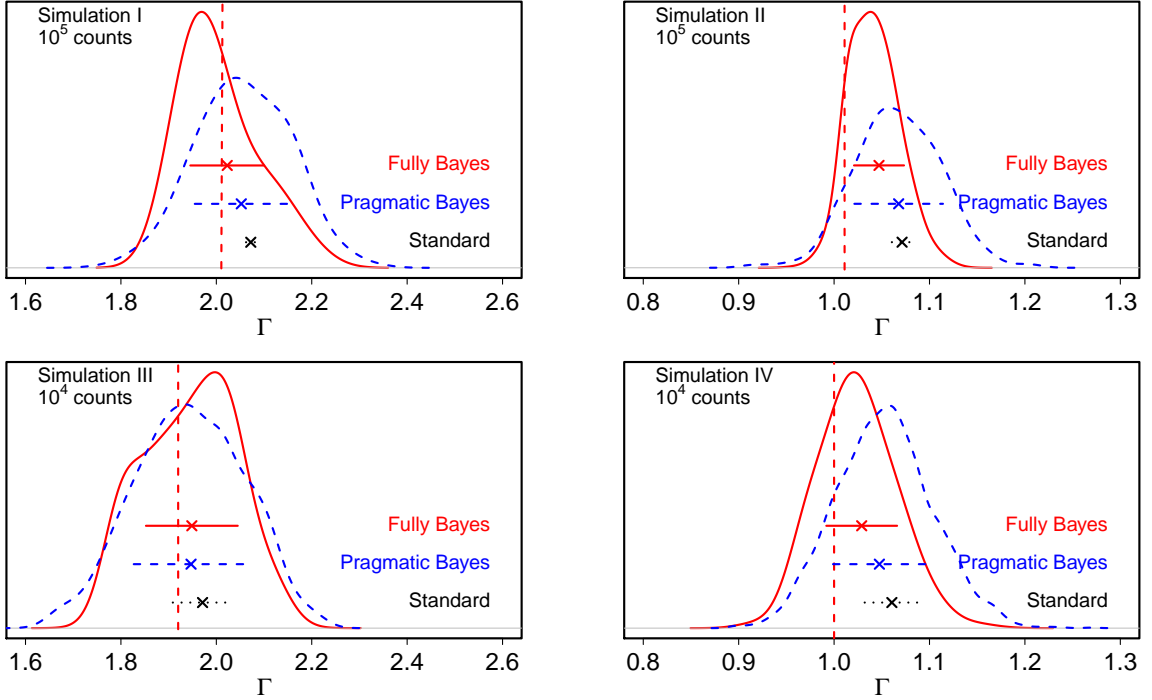


Fig. 3.— Results for Simulations I–IV. The panels show the posterior distributions (curves), fitted values ( $\times$ ), and  $1\sigma$  error bars (horizontal bars) for the spectral power-law index,  $\Gamma$ . Results for the standard, pragmatic Bayesian, and fully Bayesian methods are plotted in dotted black, dashed blue, and solid red, respectively. (The posterior distribution under the standard method is omitted to enhance the comparison of the other two methods. To smooth the plotted pragmatic Bayesian posterior distributions, we use the MH within PCG Sampler of Paper I, see Appendix A.2.) The true value of  $\Gamma$  is given by the red broken vertical line. In these simulations, we consider the situation in which the default effective area curve is misspecified to a degree that is consistent with the variability of the calibration library. Because the standard method uses this misspecified curve, it performs poorly. Both the pragmatic and the fully Bayesian methods avoid assuming that  $A$  is known without error allowing them to perform better. Of these two, the fully Bayesian method provides both estimates of  $\Gamma$  that are closer to its true values and narrower error bars.

the standard method underestimates the total error for large-count spectra. The pragmatic Bayesian method, on the other hand, incorporates both statistical and systematic errors, resulting in significantly larger error bars. Because systematic errors do not dissipate as the sample size grows, the error bars produced by the pragmatic Bayesian method are not particularly sensitive to the number of counts. In Paper I, we illustrate as the photon counts grow and the statistical errors become negligible, the error bars produced by the pragmatic methods will be entirely due to calibration uncertainty.

The power of the fully Bayesian method is that it actually uses the data to measure the systematic error. Put another way, it handles systematic error in the same way one would handle statistical error. Thus, its error bars are wider than those provided by the standard method because they incorporate both sources of error, but their width decreases as the number of counts grows because the data are able to narrow the calibration uncertainty. This in turn allows us to go one step further and actually estimate  $A$  using its posterior distribution,  $p_{\text{fB}}(A | Y)$ .

We illustrate this in Figure 4 which shows that  $A_0$  is inconsistent with the  $p_{\text{fB}}(A | Y)$  fit under Simulation II. The shaded light blue area corresponds to a 90% pointwise posterior region<sup>5</sup> for  $A$  and contains the true effective area curve,  $A_{\text{ext}}$  (plotted in red). The posterior region, however, is shifted toward  $A_0$  (plotted in black), which serves as the prior mean for  $A$ . Thus, the prior distribution on  $A$  has a clear influence on its posterior distribution. A similar pattern can be seen in Figure 3. The fitted values of  $\Gamma$  are pulled from the true value toward the best value assuming  $A = A_0$  as computed with the standard method. We emphasize, however, that the prior distribution on  $A$  used by the fully Bayesian method is in fact making a much *weaker* assumption than what is commonly made in practice: the assumption that  $A$  is exactly equal to  $A_0$ . The standard method makes this assumption and as we demonstrate in Section 3.2, its fitted values exhibit significant bias when  $A_0$  is

---

<sup>5</sup>In this case a 90% pointwise region is comprised of intervals for each energy, all of which have 90% (central) posterior probability.



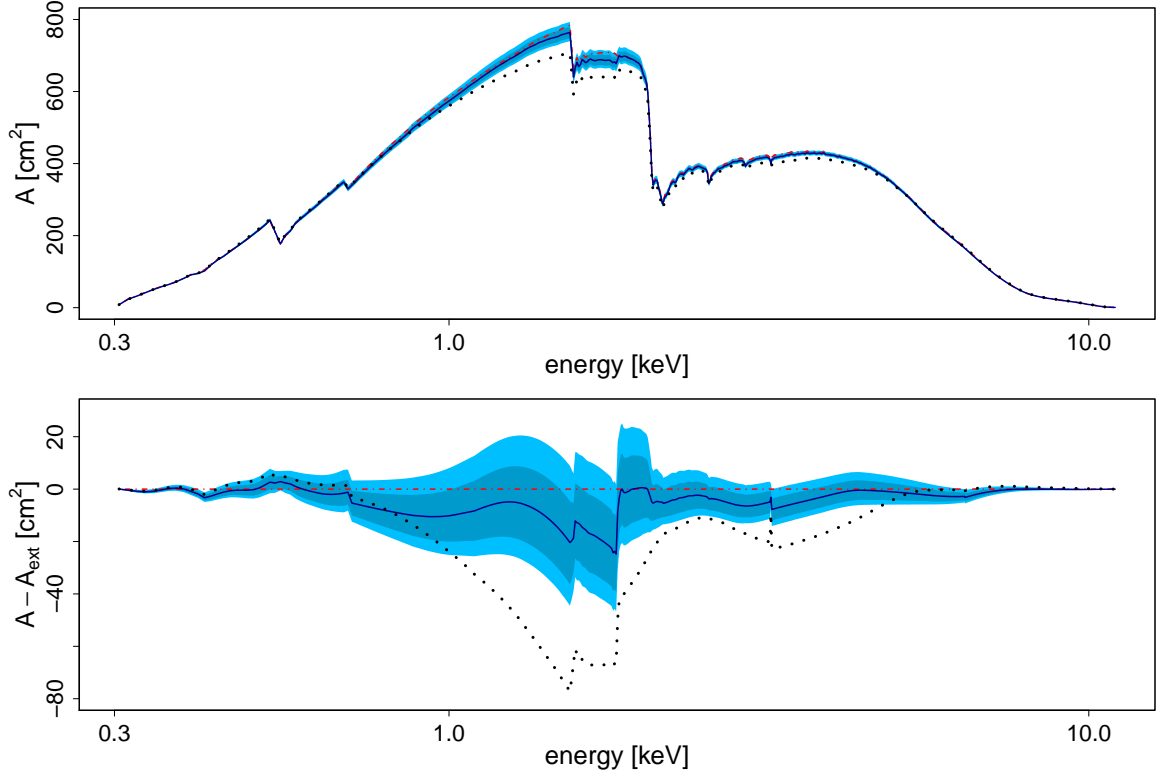


Fig. 4.— Estimating the range of effective area curves that are consistent with data. The plots summarize  $p_{\text{FB}}(A | Y)$  for Simulation II. The true (extremal) and nominal (default) effective area curves are plotted in dot-dashed red and dotted black. The pointwise posterior distribution of  $A$  is plotted in blue, where the dark (light) blue area corresponds to the central 68.3% (90%) region of  $p_{\text{FB}}(A | Y)$  and the solid blue curve plots the posterior mean. The top panel is on the effective area scale and the bottom panel subtracts off the true effective area curve,  $A_{\text{ext}}$ , to highlight differences. The true curve is contained entirely in the blue region. The posterior mean (solid blue curve) shifts from the prior mean (dotted black) toward the true effective area curve (dot-dashed red).

misspecified.

### 3.2. A frequency evaluation of the methods

The  $2 \times 2$  simulation study described in Section 3.1 generated only one data set for each of the four simulation studies. In this section we generate fifty spectra for each of the four simulation settings described in Table 2 and fit each of the resulting  $50 \times 4$  simulated spectra with the standard, pragmatic Bayesian, and fully Bayesian methods. The fitted values for  $\Gamma$  and their 68.2% error bars computed using the first ten spectra generated in each simulation appear in Figure 5. Numerical summaries of the entire simulation study appear in Table 3. The frequency analyses confirm the trends we highlighted in Section 3.1:

- In all four simulations, the standard method on average exhibits the narrowest error bars (mean error bars), the pragmatic Bayesian method has the widest, and those of the fully Bayesian method are in between. (This also holds when  $A$  is known with certainty, see Xu 2014.)
- When  $A$  is correctly specified all three methods perform well, but the standard method performs best because it uses precise, but in practice unattainable, knowledge of  $A$ . (These results are not shown, see Xu 2014.)
- In the more realistic situation in which there is uncertainty in  $A$  (i.e., Simulations I–IV) both the standard and pragmatic Bayesian methods exhibit substantial bias<sup>6</sup>. The fully Bayesian method reduces this bias, resulting in the overall smallest mean square error.

---

<sup>6</sup>By “bias”, we mean statistical bias, which if defined as the difference between the expected value of a fitted parameter upon repeated replication of the data and the true value of that parameter.

- The advantage of the fully Bayesian method is most striking when  $A$  is misspecified and the data set is large (Simulations I and II). In this case the estimates produced with the fully Bayesian method have much lower bias and root mean square error than those of the other two methods. Fully Bayesian intervals are much more likely to include the true value of  $\Gamma$  (coverage) than intervals based on a fixed effective area fit and the fully Bayesian error bars can be much narrower than the pragmatic Bayesian error bars. These effects dissipate with smaller data sets because substantial data are required to narrow the calibration uncertainty.

## 4. Data Analyses

Here we apply the three methods to real data to demonstrate the practical effects of the fully Bayesian method. We consider power-law and multi-thermal sources observed with ACIS-S, and a nominal blackbody source observed with Chandra HRC-S/LETG.

### 4.1. A Sample of Radio-Loud Quasars

In radio loud quasars, X-ray emission originates in close vicinity to a supermassive black hole and is believed to be caused by either an accretion disk or a relativistic jet. This emission can be modeled with a Compton scattering process and the X-ray spectrum described using an absorbed power law:

$$\Lambda(E; \theta) = N E^{-\Gamma} e^{-\sigma(E)N_{\text{H}}} \text{ photons cm}^{-2} \text{ sec}^{-1} \text{ keV}^{-1}, \quad (12)$$

where  $\sigma(E)$  is the absorption cross-section, and the three model parameters are  $\theta = (N, \Gamma, N_{\text{H}})$  with  $N$  the normalization at 1 keV,  $\Gamma$  the photon index of the power law, and  $N_{\text{H}}$  the absorption column. We consider a small sample of radio loud quasars whose spectra were observed with the *Chandra* X-ray Observatory in 2002 (Siemiginowska *et al.* 2008; Lee *et al.* 2011). Standard data processing including source extraction and calibration was preformed using the CIAO software (*Chandra* Interactive Analysis of Observations).

Table 3: Statistical summaries of the frequency simulation study.

	Standard					Pragmatic Bayesian					Fully Bayesian				
	cover- age <sup>1</sup>	error <sup>2</sup>	bias <sup>3</sup>	std error <sup>4</sup>	root mse <sup>5</sup>	cover- age	error	bias	std error	root mse	cover- age	error	bias	std error	root mse
Sim I <sup>†</sup>	<b>0.00</b>	<b>0.030</b>	<b>0.092</b>	<b>0.029</b>	<b>0.097</b>	<b>0.62</b>	<b>0.094</b>	<b>0.082</b>	<b>0.030</b>	<b>0.088</b>	<b>0.76</b>	<b>0.065</b>	<b>0.035</b>	<b>0.040</b>	<b>0.052</b>
Sim II <sup>†</sup>	<b>0.00</b>	<b>0.010</b>	<b>0.056</b>	<b>0.007</b>	<b>0.056</b>	<b>0.80</b>	<b>0.042</b>	<b>0.053</b>	<b>0.008</b>	<b>0.054</b>	<b>0.78</b>	<b>0.025</b>	<b>0.002</b>	<b>0.017</b>	<b>0.016</b>
Sim III	0.40	0.076	0.099	0.078	0.125	0.68	0.123	0.083	0.079	0.113	0.66	0.084	0.068	0.078	0.103
Sim IV	0.12	0.026	0.057	0.027	0.063	0.48	0.049	0.053	0.028	0.059	0.56	0.040	0.040	0.036	0.053

<sup>1</sup>The proportion of 68% ( $1\sigma$ ) intervals that contain the true value of  $\Gamma$ .

<sup>2</sup>The mean length of the  $1\sigma$  error bars.

<sup>3</sup>The mean of the fitted values of  $\Gamma$  minus its true value.

<sup>4</sup>The standard deviation of the fitted values of  $\Gamma$ .

<sup>5</sup>The square root of the mean of the squared deviations between the fitted value and the true value of  $\Gamma$ .

<sup>†</sup>Results for Simulations I and II are highlighted in bold because they show the largest gain in performance for the fully Bayesian method. Specifically, when the effective area curve is misspecified with a high-count spectra.

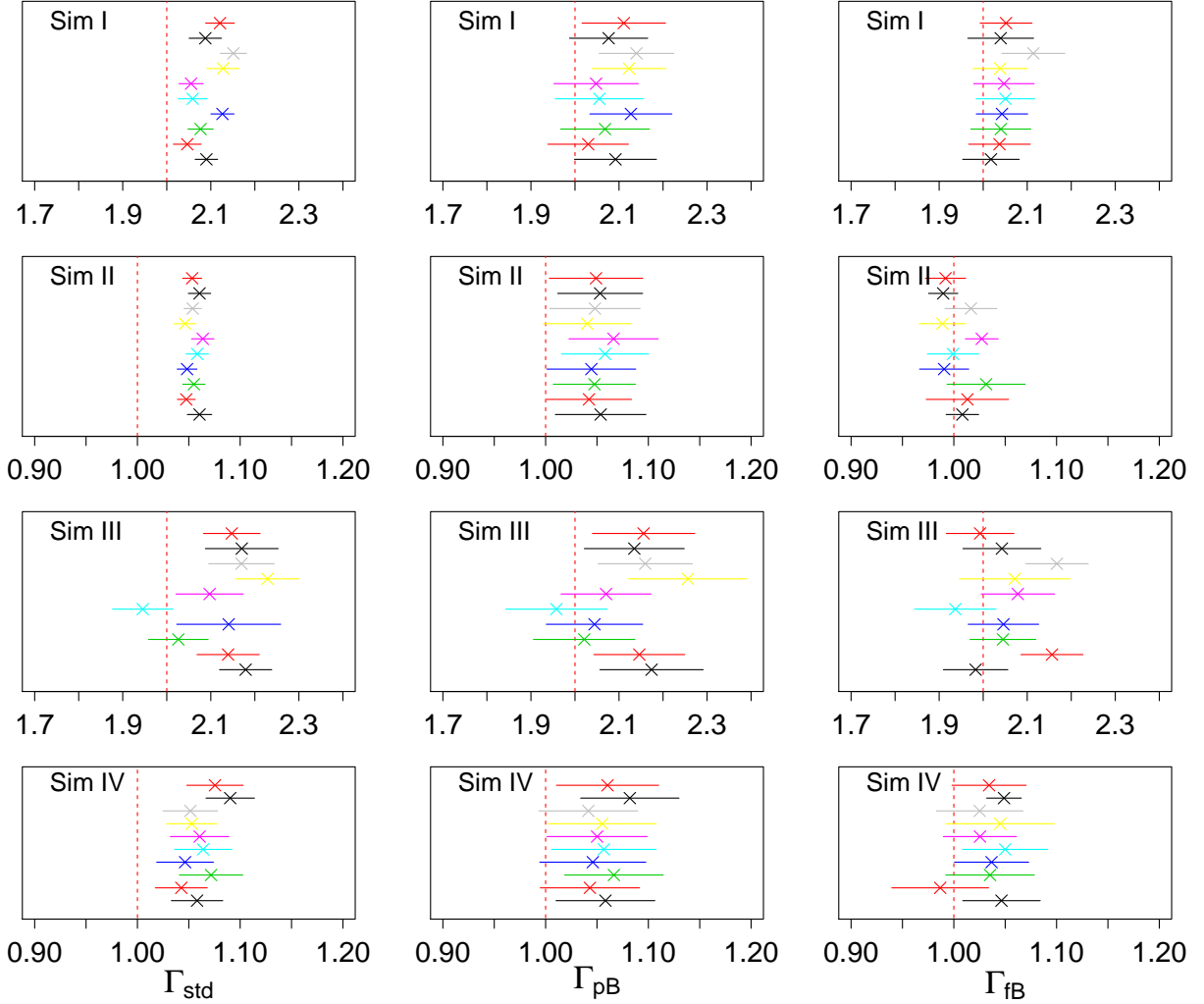


Fig. 5.— Frequency analysis for Simulations I–IV. Each panel gives the fitted values and  $1\sigma$  error bars for  $\Gamma$  resulting from ten replicate simulations. The red broken vertical line in each panel indicated the true value of  $\Gamma$ . The four rows correspond to the four simulation settings and the three columns correspond to the standard, pragmatic Bayesian, and fully Bayesian methods, respectively. Owing to the misspecification of  $A_0$ , the standard and pragmatic Bayesian methods both exhibit significant bias. The pragmatic Bayesian method adjusts for this bias with wider error bars, while the fully Bayesian method reduces the bias. Overall the fully Bayesian method is able to cover the true value of  $\Gamma$  more often with narrow errors bars than either of the other methods, especially with large-count spectra (i.e., Simulations I and II).

The number of counts in the observed spectra varies between 8 and 5500. As in Paper I, we excluded two of the spectra (ObsID 3099 which had 8 counts, and ObsID 836 which is better described by a weak thermal spectrum) and reanalyzed the remaining 15 using the standard, pragmatic Bayesian, and fully Bayesian methods. We account for background contamination using a background spectrum extracted over large annuli surrounding each source and a highly-structured background model that was originally fit to the blank-sky data provided by the *Chandra* X-ray Center, see Paper I for details. Only the normalization of the background model was fit in the individual spectral analyses. This approach was used for all but the two lowest-count spectra ( $< 45$ , both with short 5 ksec exposures), for which background was ignored. The sample of quasars was originally analyzed by Siemiginowska *et al.* (2008) who did not account for calibration uncertainty. A followup analysis accounted for calibration uncertainty using the pragmatic Bayesian method and resulted in substantially larger error bars for the high-count datasets (Paper I). As illustrated in Section 3, systematic errors due to calibration uncertainty swamp statistical errors for large data sets. For small data sets, however, the statistical errors may be much larger and the relative effect of calibration uncertainty is therefore less important. Here we reanalyze the same spectra with the fully Bayesian method and illustrate how it is able to deliver low-bias parameter estimates with smaller error bars than in the pragmatic Bayesian method.

We fit each spectrum in three ways:

1. with the standard method,
2. with the pragmatic Bayesian method using the Iterated MH within PCG Sampler (see Section A.2), and
3. with the fully Bayesian method using the Mixed Pragmatic Proposal Sampler.

With each of the three methods, we use the fifteen observation-specific default effective area curves,  $A_0^*$ , corresponding to each spectra. For the two Bayesian methods, we use  $J = 8$

in the PCA summary of the calibration library along with the  $A_0^*$  in Equation 5. When running the Iterated MH within PCG Sampler, we set  $I = 10$  and  $M = 10$ ; that is at each iteration we run `pyBLocXS`  $M + I - 1$  times, discarding the output of the first  $I - 1$  runs and keeping the output of the final  $M$  runs. In the Mixed Pragmatic Proposal Sampler, the two proposal rules were used in equal proportion and with the random-walk proposal we set  $\sigma = 0.1$ .

Results appear in Figures 6 and 7. Error bars for  $\Gamma$  computed under the pragmatic Bayesian ( $\hat{\sigma}_{\text{pB}}$ ) and fully Bayesian ( $\hat{\sigma}_{\text{fB}}$ ) methods are compared with those computed under the standard method ( $\hat{\sigma}_{\text{std}}$ ) in Figure 6. The left panel replicates results reported in Paper I and shows that for large data sets, for which  $\hat{\sigma}_{\text{std}}$  is small, the pragmatic Bayesian method produces much larger error bars than the standard method;  $\hat{\sigma}_{\text{pB}}$  accounts for systematic and statistical errors, whereas  $\hat{\sigma}_{\text{std}}$  only accounts for statistical errors. For the largest datasets  $\hat{\sigma}_{\text{pB}}$  is twice as big as  $\hat{\sigma}_{\text{std}}$ . The right panel of Figures 6 shows that the fully Bayesian method produces error bars more in line with  $\hat{\sigma}_{\text{std}}$ ; although for the largest data sets  $\hat{\sigma}_{\text{fB}}$  is bigger than  $\hat{\sigma}_{\text{std}}$ , it is not as big as  $\hat{\sigma}_{\text{pB}}$ .

Figure 7 compares the  $1\sigma$  intervals for  $\Gamma$  produced by the three methods. Consider an interval computed under the pragmatic Bayesian method:  $\mathcal{CI}_{\text{pB}}(\Gamma) = \{\hat{\Gamma} \pm \hat{\sigma}_{\text{pB}}(\Gamma)\}$ , where  $\hat{\Gamma}$  is the estimate of  $\Gamma$  under this method. To compare this interval with that computed under the standard method, we subtract  $\hat{\Gamma}_{\text{std}}$  from  $\mathcal{CI}_{\text{pB}}(\Gamma)$  and divide by  $\hat{\sigma}_{\text{std}}(\Gamma)$ . The result is an interval that extends from  $-1$  to  $1$  if  $\mathcal{CI}_{\text{pB}}(\Gamma)$  and  $\mathcal{CI}_{\text{std}}(\Gamma)$  are identical, is wider if  $\hat{\sigma}_{\text{pB}} > \hat{\sigma}_{\text{std}}$ , and shifts to the left or right if  $\hat{\Gamma}_{\text{pB}}$  and  $\hat{\Gamma}_{\text{std}}$  differ. These adjusted intervals are plotted in the left panel of Figure 7. There is little shifting to the left or right because the pragmatic Bayesian and standard methods produce similar estimates. For smaller count data sets the adjusted intervals are nearly  $\{-1, 1\}$ , but for larger count data sets the adjusted intervals are as much as twice as wide. The right panel of Figure 7 illustrates intervals computed under the fully Bayesian method,  $\mathcal{CI}_{\text{fB}}(\Gamma)$ , that are adjusted in the same manner. For the smaller data sets  $\mathcal{CI}_{\text{fB}}(\Gamma)$  and  $\mathcal{CI}_{\text{std}}(\Gamma)$  are similar, but for larger data sets, they differ: the adjusted intervals tend to shift to the left or right but are

not generally much more than two units wide. This means that the fully Bayesian method tends to adjust the fitted value and to increase error bars only moderately. In two cases (ObsID 3097 and 866) the fully Bayesian method shifts the fitted value of  $\Gamma$  by more than  $\hat{\sigma}_{\text{std}}$ . This constitutes a significant shift in the scientific inference for these observations when calibration uncertainty is accounted for in a principled Bayesian manner.

Guidelines as to how many counts are needed for the fully Bayesian fitted values and error bars to differ from those computed with the standard method would be very useful. In practice, however, this depends on a number of key factors including the instrument in question (i.e., the calibration library), the energy range over which fits are carried out, the fitted spectral model, and the true underlying source spectrum. Thus, there are no useful generic guidelines and the establishment of such guidelines in specific situations is an important area of future research. This example provides a single illustration of this possibility, where we see that observations with more than about 5000 counts produced noticeably different inference under the two methods, see Figure 7.

#### 4.2. Fitting a multi-thermal spectral model

Our analyses thus far have been carried out using a simple power-law spectral model (in Paper I, in the simulations in Section 3, and for the observed data in Section 4.1). Here we illustrate the pragmatic and fully Bayesian methods using a more complex multi-component thermal model. To do this, we analyze one of the strongest sources in the Chandra Source Catalog,  $\zeta$  Ori, a young (<12 Myr) binary system comprised of an X-ray bright O9 supergiant and a weaker B0 subgiant with about a  $3''$  separation. The source is observed (ObsID 1878) at 15 arcmin off-axis, situated on the ACIS-S2 chip, and is detected with a count rate of  $1.33 \text{ ct s}^{-1}$ , with  $> 10^5$  counts in 75.46 ksec. Because of the large off-axis location, the point spread function is broad, and the binary cannot be spatially resolved. Furthermore, the source is spread out over more than 20000 pixels, with maximum fluence at  $<0.0017 \text{ ct s}^{-1} \text{ pix}^{-1}$  so that CCD pileup effects may be ignored.



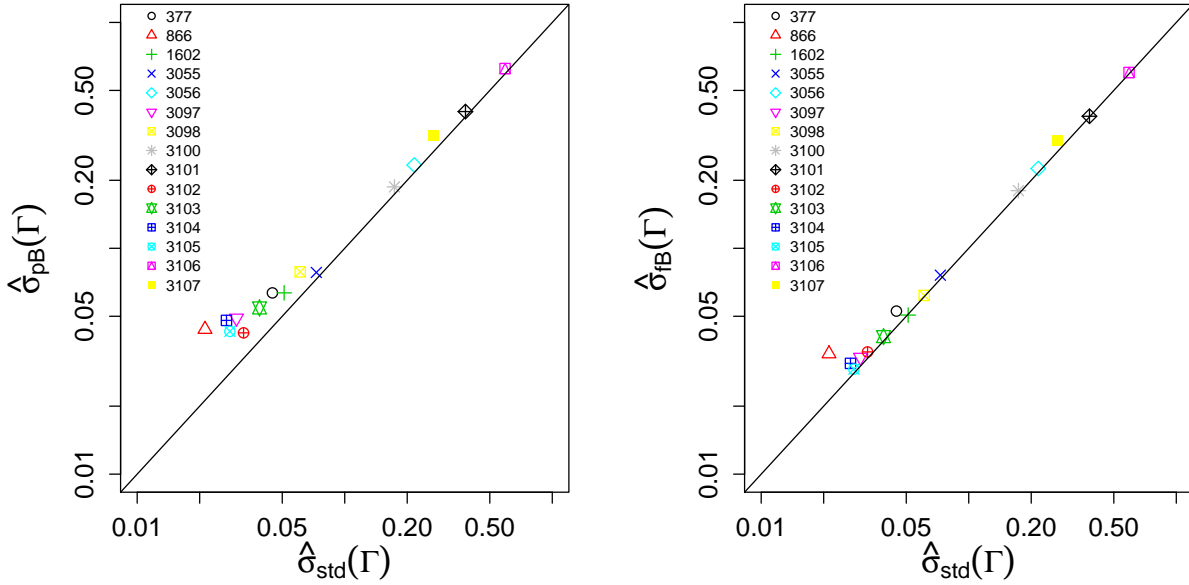


Fig. 6.— Comparing the error bars for  $\Gamma$  computed under the standard ( $\hat{\sigma}_{\text{std}}$ ), pragmatic Bayesian ( $\hat{\sigma}_{\text{pB}}$ ), and fully Bayesian ( $\hat{\sigma}_{\text{fB}}$ ) methods using spectra from each of fifteen radio-loud quasars. Smaller values of  $\hat{\sigma}_{\text{std}}$  correspond to data sets with more counts. For high-count spectra,  $\hat{\sigma}_{\text{pB}}$  tends to be substantially larger than  $\hat{\sigma}_{\text{std}}$  while  $\hat{\sigma}_{\text{fB}}$  is only moderately larger than  $\hat{\sigma}_{\text{std}}$ . Thus, the fully Bayesian methods is able to provide a principled accounting for calibration uncertainty with only a moderate increase in the final error bars.

Our objective is not to model the spectrum in detail, but rather, to consider the effect calibration uncertainty has on spectral fitting. (The X-ray emission is thermal, generated from shocked plasma deep in the wind; see Waldron and Cassinelli 2001, Pollock 2007, Raassen *et al.* 2008, Herve *et al.* 2013, for various models designed to account for X-ray emission from massive stars.) We construct a variable-abundance absorbed 2-temperature APEC spectral model and fit it to the data. This roughly mimics previous attempts to model spectra of  $\zeta$  Ori obtained with other telescopes such as ASCA (Yamauchi *et al.* 2000) and XMM-Newton (Raassen *et al.* 2008). For reference, Yamauchi *et al.* find two temperature components at  $T_1 = 0.2$ ,  $T_2 = 0.6$  keV, at an absorption column fixed at  $N_{\text{H}} = 2.6 \cdot 10^{20} \text{ cm}^{-2}$ ; Raassen *et al.* find three temperature components at  $T_1 = 0.55$ ,  $T_2 = 0.2$ ,  $T_3 = 0.07$  keV,  $N_{\text{H}} = 5 \cdot 10^{20} \text{ cm}^{-2}$ , with abundances of C, N, and O, being close

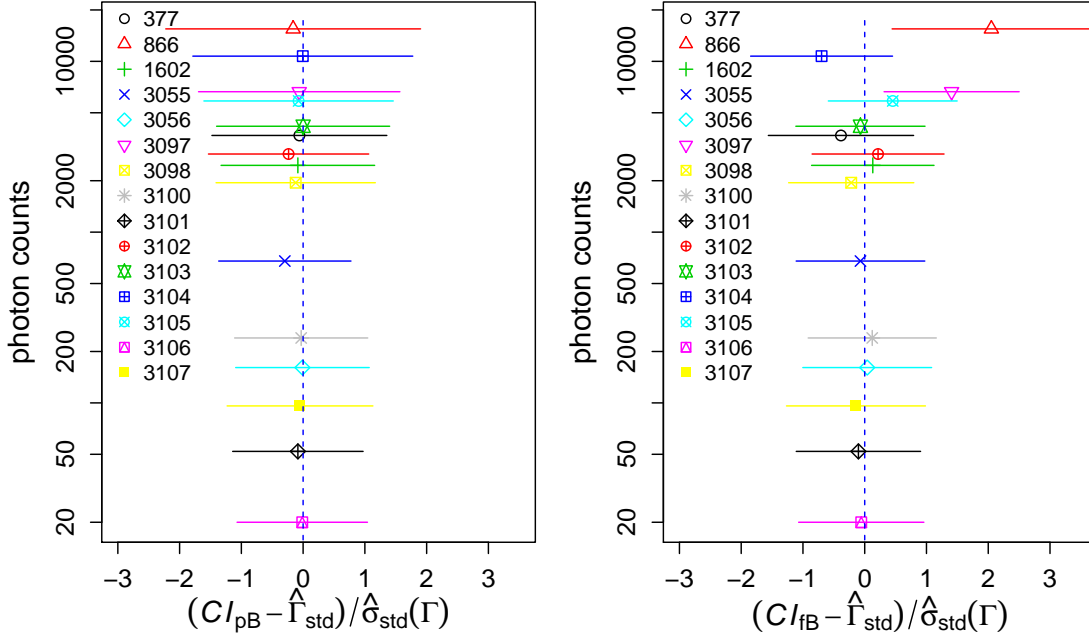


Fig. 7.— Standardized pragmatic (left) and fully Bayesian (right) intervals. Using spectra from the fifteen quasars, the left panel plots  $1\sigma$  confidence intervals computed with the pragmatic Bayesian method, but re-centered and re-scaled using the estimate and error bars of  $\Gamma$  computed under the standard method:  $(CI_{\text{pB}} - \hat{\Gamma}_{\text{std}}) / \hat{\sigma}_{\text{std}}(\Gamma)$ . If the pragmatic and standard methods return the same estimates and error bars, the plotted intervals would equal the interval  $(-1, 1)$ . In fact the plotted intervals are as much as twice this wide for large-count datasets, indicating that  $\hat{\sigma}_{\text{pB}}$  can be substantially larger than  $\hat{\sigma}_{\text{std}}$ . The right panel plots  $1\sigma$  confidence intervals computed with the fully Bayesian method, re-centered and re-scaled in the same manner. The plotted intervals shift right and left because the fitted values under the standard and fully Bayesian methods differ. The widths of the fully Bayesian intervals, however, are only moderately larger than the standard intervals.

Table 4: Fitted parameters for  $\zeta$  Ori

Model	Standard	Pragmatic Bayesian	Fully Bayesian
Parameters	Analysis	Analysis	Analysis
$N_{\text{H}}$ [ $10^{20}$ cm $^{-2}$ ]	14.8 $\pm$ 0.90	15.0 $\pm$ 0.92	15.1 $\pm$ 0.87
$T_1$ [keV]	0.179 $\pm$ 0.0016	0.179 $\pm$ 0.0022	0.181 $\pm$ 0.0019
$T_2$ [keV]	0.474 $\pm$ 0.0063	0.474 $\pm$ 0.0069	0.471 $\pm$ 0.0067
[C, N, O] $^{\dagger}$	0.23 $\pm$ 0.018	0.23 $\pm$ 0.024	0.21 $\pm$ 0.026
[Ne] $^{\dagger}$	0.48 $\pm$ 0.031	0.48 $\pm$ 0.033	0.47 $\pm$ 0.034
[Ni, Mg, Si, Ca, Fe] $^{\dagger}$	0.41 $\pm$ 0.028	0.41 $\pm$ 0.032	0.40 $\pm$ 0.031
Norm $_1$ [ $10^{12}$ cm $^{-5}$ ]	6.3 $\pm$ 0.65	6.4 $\pm$ 0.78	5.9 $\pm$ 0.56
Norm $_2$ [ $10^{12}$ cm $^{-5}$ ]	1.05 $\pm$ 0.057	1.07 $\pm$ 0.079	1.00 $\pm$ 0.059

$^{\dagger}$  abundances relative to solar (Anders and Grevesse 1989)

to solar photospheric (represented by the compilation of Anders and Grevesse 1989), and those of Ne, Mg, Si, and Fe being elevated. The results of applying the standard, pragmatic Bayesian, and fully Bayesian methods to the data are given in Table 4. We find significantly higher absorption columns, and lower temperatures and abundances, with our estimated abundance consistent with the low metallicities measured for the nearby NGC 2023 star cluster (López-García *et al.* 2013). These results are stable with respect to calibration uncertainty, with all three methods producing similar best-fit values.

The fits are also summarized in Figures 8 and 9. They show that relative to the standard method, the pragmatic Bayesian method delivers similar fitted values for  $T_1$  and  $T_2$ , but accounts for calibration uncertainty by inflating their error bars. The fully Bayesian method shifts the fitted values by a small amount and generally requires a smaller increase in the error bars. In this case, the posterior correlation of  $T_1$  and  $T_2$  decreases under the fully Bayesian method, see Figure 9.

The second row of Figure 8 shows time series plots of the Markov chains of  $T_1$  used to simulate the three posterior distributions. (The samplers were run with

$J = 8, I = 10, M = 10$ , and  $\sigma = 0.1$ .) While all three converge reasonably well, the fully Bayesian chain occasionally “sticks” at a particular value of the parameter for a number of iterations. This indicates that the pragmatic proposal distribution may attribute relatively little probability to some regions of the parameter space with appreciable probability under  $p_{\text{FB}}(\theta, A | Y)$ . This can also be seen in Figure 9 where the 90% contours of (normal approximations to) the pragmatic and fully Bayesian posterior distributions are plotted in green and blue, respectively. The fact that the fully Bayesian contour extends outside the pragmatic Bayesian contour indicates that we may have trouble exploring parts of  $p_{\text{FB}}(\theta, A | Y)$  using the pragmatic proposal distribution. (Recall that we require the jumping rule of the independence sampler to be an over-dispersed approximation to  $p_{\text{FB}}(\theta, A | Y)$ .) The mixed pragmatic proposal sampler, however, mixes the pragmatic proposal with a random walk update. This second component allows simulation of parameter values with relatively low probability under the pragmatic proposal rule. In this case the sampler was computationally costly. Owing to its “sticking” we ran the fully Bayesian sampler for 30,000 iterations. (The pragmatic sampler was run for 3000.) This combined with complexity of the multi-thermal spectral model fit with its eight parameters resulted in a larger computational burden ( $\gg 20\times$ ) than the other analyses carried out here (see Sections 4.1, 4.3 – those runs typically took 60-90 min on a 64-bit 2.2GHz CPU with 16GB of RAM).

In the fully Bayesian run, the data prove to be informative about the effective areas. The subset of  $\mathcal{A}$  that is consistent with the data and the adopted model suggests that the nominal effective area is underestimated, see Figure 10. This is not surprising, since effective areas at large off-axis angles are not as well calibrated as those near the aimpoint. Naturally, this result is conditional on the validity of the adopted spectral model of an absorbed two-temperature thermal emission with variable abundances. Our analysis suggests that at off-axis angles  $> 15$  arcmin the *Chandra* effective areas should be increased by  $\approx 10\%$  over the 0.5-2 keV range and by  $\approx 5\%$  at high energies  $> 6$  keV.

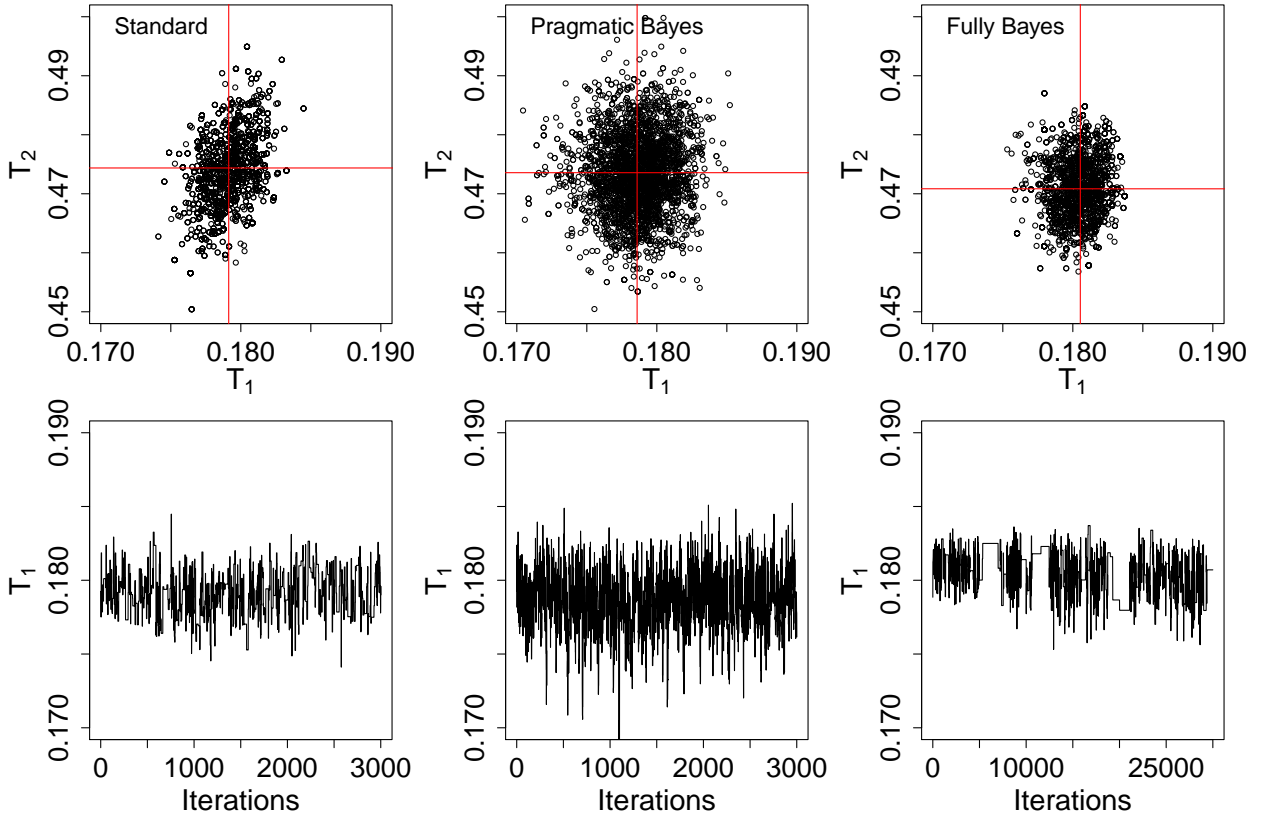


Fig. 8.— The standard, pragmatic Bayesian, and fully Bayesian fits of the  $T_1$  and  $T_2$  parameter in the spectrum of  $\zeta$  Ori. The first row plots a complete Monte Carlo simulation from each of the three posterior distributions; the solid red lines correspond to the posterior means of  $T_1$  and  $T_2$  under each fit. The fitted values (posterior means) for the two parameters are indistinguishable under the standard and pragmatic Bayesian methods, but shift noticeably under the fully Bayesian method. Error bars are quantified by the spread of the simulated parameter values under each of the methods. The error bars under the pragmatic Bayesian method are noticeably larger than those computed with the other two methods. Thus, the fully Bayesian method again is able to account for calibration uncertainty by shifting the fitted values rather than by increasing their error bars. The second row presents time series plots of the Markov chains for  $T_1$  used to generate the three Monte Carlo simulations. While all three chains are fairly well behaved, the fully Bayesian chain occasionally “sticks” at the same parameter value for a number of iterations. This is an indication that the pragmatic proposal distribution attributes relatively little probability to some regions of the parameter space with appreciable probability under  $p_{\text{fB}}(\theta, A | Y)$ . Nonetheless, the algorithm preforms well enough for valid inference. (Here we only plot every tenth iteration of the fully Bayesian sampler.)

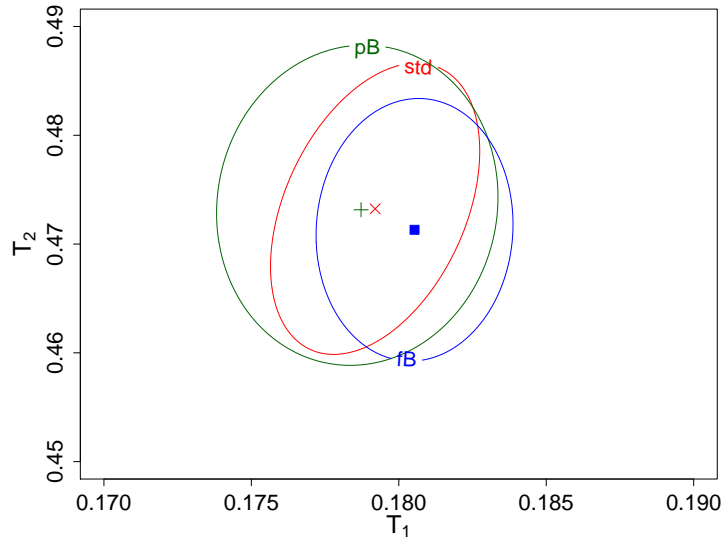


Fig. 9.— The plot superimposes the 90% contours of the three posterior distributions simulated in the first row of Figure 8. The contours are computed using a normal approximation to the posterior distribution computed under the standard (red), pragmatic Bayesian (green), and fully Bayesian (blue) methods. The posterior means under the three methods are plotted, respectively, as a red  $\times$ , a green  $+$ , and a filled blue square. Relative to the standard method, the pragmatic Bayesian method delivers similar fitted values and larger error bars, while the fully Bayesian method shifts the fitted value but only moderately inflates the error bars. In this example the correlation of the two parameters decreases under the fully Bayesian method.

### 4.3. Fitting a blackbody model to a grating spectrum

As an additional example of the versatility of our approach, we consider data from an entirely different detector, fitted with a blackbody spectral model. We analyze a high-resolution grating spectrum of an isolated neutron star, RX J1856.5-3754 (RXJ1856). This source has been observed with the LETGS+HRC-S grating/detector combination numerous times over the *Chandra* mission, and has accumulated 617.735 ks of exposure. RXJ1856 is an intrinsically interesting object: it was originally classified as an isolated neutron star, but was later suspected to be a quark star (Drake *et al.* 2002). The X-ray data were fit as a blackbody spectrum with a temperature of  $T = 61.1 \pm 0.3$  eV by Drake

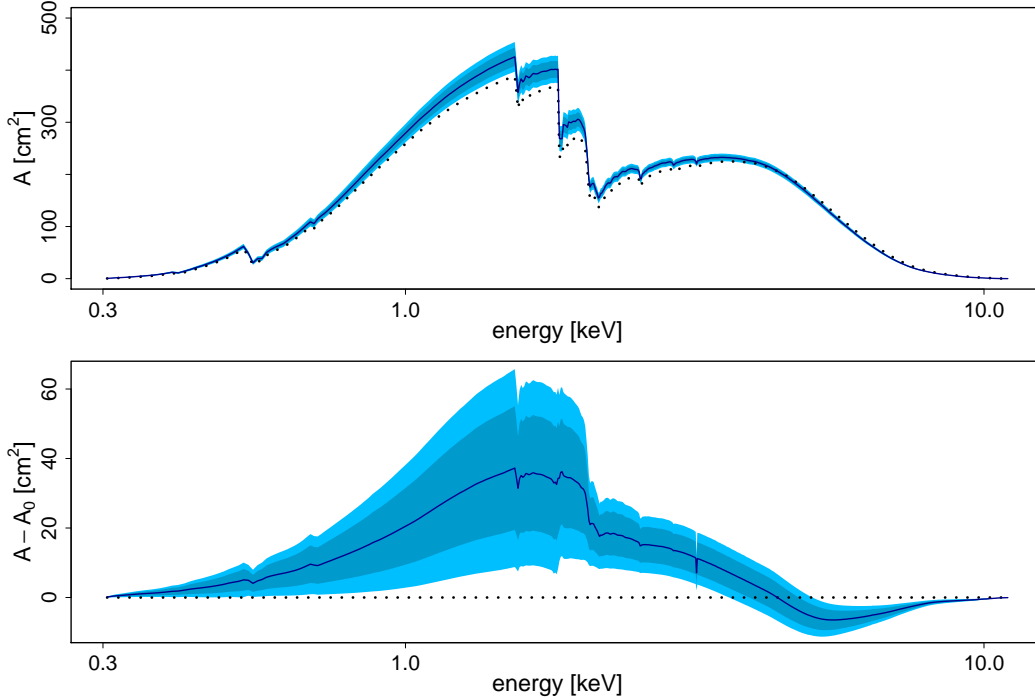


Fig. 10.— Estimating the range of effective area curves that are consistent with the spectrum of  $\zeta$  Ori. The plots summarize  $p_{\text{fB}}(A | Y)$ . The pointwise posterior distribution of  $A$  is plotted in blue, where the dark (light) blue area corresponds to the central 68.3% (90%) region of  $p_{\text{fB}}(A | Y)$ . The pointwise posterior mean of  $A$  and its default value,  $A_0$  are plotted as solid blue and dotted black lines, respectively. The top panel shows the full effective areas and the bottom panel subtracts off the default effective area curve,  $A_0$ , to highlight the differences. (We cannot subtract off the true curve as in Figure 4 because it is unknown.) The data suggest that  $A_0$  underestimates the true effective area by  $\approx 10\%$  over the 0.5-2 keV range and by  $\approx 5\%$  at high energies  $> 6$  keV, and the vignetting correction must be reduced for large off-axis angles for *Chandra* imaging observations.

*et al.*, resulting in a radius estimation of  $\approx 4\text{--}8$  km. But the optical data are inconsistent with the X-ray predictions, and require fitting by a more complex magnetic Hydrogen atmosphere model, with a temperature of  $T_\infty \approx 37$  eV and a radius  $R_\infty \approx 17$  km consistent with a conventional neutron star core (Ho *et al.* 2007). In the X-ray regime itself, these two models cannot be statistically distinguished (Ho *et al.*).

Here, for the sake of simplicity, we adopt an absorbed Blackbody spectrum model to fit the soft X-ray data. This spectral model was fit using exactly the same methods and algorithms as in Section 4.1 ( $J = 8, I = 10, M = 10, \sigma = 0.1.$ ), except that the background was modeled as a fixed eighth-degree polynomial whose coefficients were determined via a standard fit to the background spectrum, and was then incorporated into the source model with a variable normalization.

As was done for the *Chandra*/ACIS-S, we generate a calibration library based on constrained spline curve modifications of known subsystem uncertainties in the LETGS+HRC-S system (Drake *et al.*, in preparation). Figure 11 shows the top five principal components (in color), which together account for  $> 95\%$  of the variance in the library. Also shown, in grey, are the summed contributions of the remaining components, which account for  $< 5\%$  of the variance. Notice that there are wavelength regions where this residual could be a significant factor. The two wavelength ranges over which RXJ1856 data are informative are shown with vertical dashed lines; the residual components do not affect the analysis over these ranges. Note that in all of our analyses, we use  $J = 8$  principal components.

We limit our analysis to the wavelength ranges  $+ [25:59.5, 68:80]$  Å, with the gap centered on the HRC-S chip gap. The chip gap is excluded because, even though the nominal effective area  $A_0$  includes the effect of dither and corrects for the drop across the gap, it is subject to additional systematic errors due to deformations in the locations of the active regions on the chip, and these are not included in the calibration library. We co-add the spectra from all positive dispersion datasets of RXJ1856. We exclude the negative-order dispersion data for the sake of simplicity. It has a chip gap in a different location and



the uncertainty in its effective area likely exhibits qualitatively different characteristics compared to the positive-order, and thus using it would make interpretation of the analysis results difficult. There are  $\approx 129$  kct in the spectrum over the chosen wavelength range, of which  $\approx 43.8$  kct are estimated to be due to the background. Although the large number of net counts ( $\approx 85$  kct) puts this dataset in the range where calibration uncertainty will likely affect the total error bars, the large fraction of expected background counts makes this an inefficient dataset to place constraints on the calibration library. As expected, the application of the pragmatic Bayesian method increases the error bars on the best-fit model parameters, see Table 5. For instance, the temperature estimate remains stable at  $T = 62.4$  eV, but the uncertainty increases from  $\pm 0.6$  eV for standard analysis, to  $\pm 1.05$  eV for pragmatic Bayesian, and decreases slightly to  $\pm 0.93$  eV for fully Bayesian analysis, see also Figure 12.

Unlike the case with  $\zeta$  Ori (Section 4.2), there is no significant effect on the range of effective area curves in the calibration library that are consistent with the spectrum of RXJ1856 (see Figure 13). We attribute this lack of sensitivity partly to the high background that contaminates the dataset, and also to the short wavelength range over which the dataset is informative. As we see from the Principal Components displayed in Figure 11, there are long-range correlations present in the library which will be selected for when a source with a suitably long wavelength range is analyzed. In the current analysis, there is a suggestive small curvature bias of  $\sim 1\text{-}2\%$  in the HRC-S/LETG effective area over the  $25 - 80$  Å range. This bias, however, is fully contained within the nominal  $1\sigma$  range of the effective area curves.

## 5. Discussion

In this article we demonstrate the advantage of a fully Bayesian accounting of calibration uncertainty. Relative to the pragmatic Bayesian method the fully Bayesian method delivers estimates with smaller bias and smaller error bars. As with the pragmatic

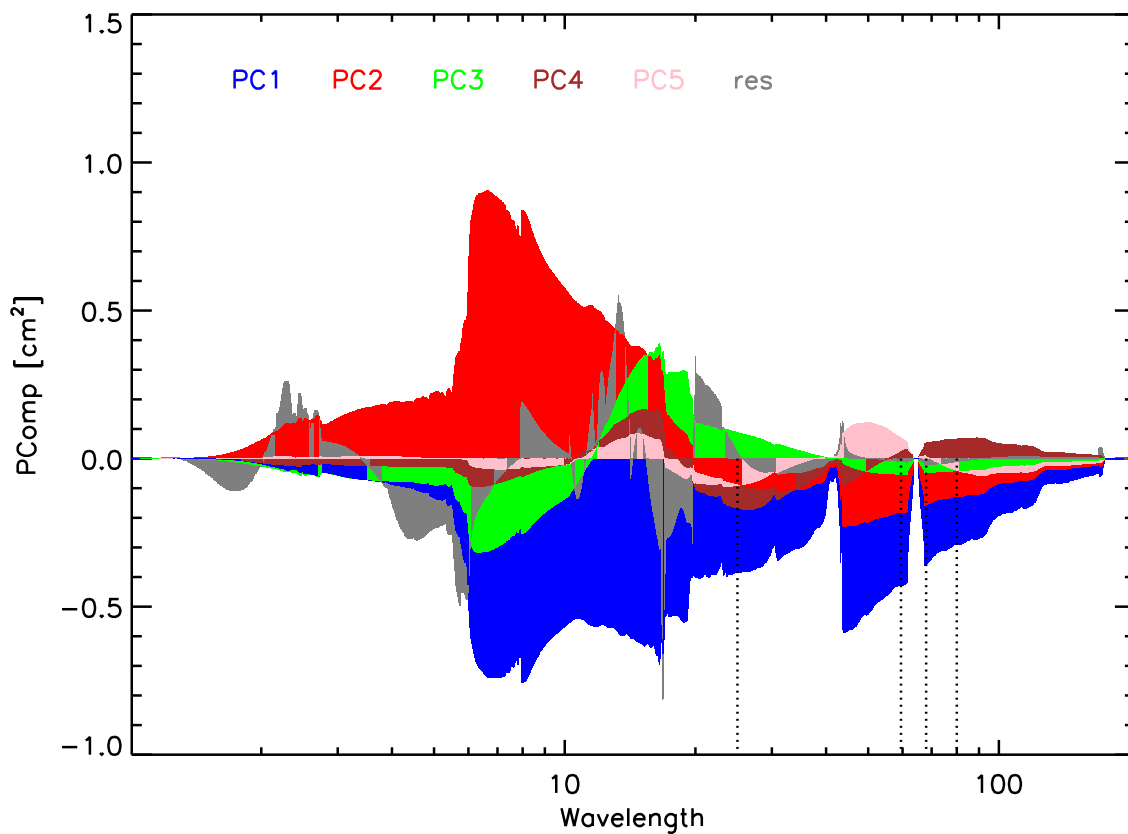


Fig. 11.— Top five principal components of the *Chandra* LETGS+HRC-S effective area calibration library. The 1<sup>st</sup> (blue), 2<sup>nd</sup> (red), 3<sup>rd</sup> (green), 4<sup>th</sup> (brown), and 5<sup>th</sup> (pink) components, weighted by the square-root of their eigenvalues, are shown as colored regions. Together, they account for 95.5% of the variance in the LETGS+HRC-S calibration library. The sum of the similarly weighted contributions from the remaining components is shown as grey. The vertical dashed lines indicate the wavelength range over which the data analysis is carried out.

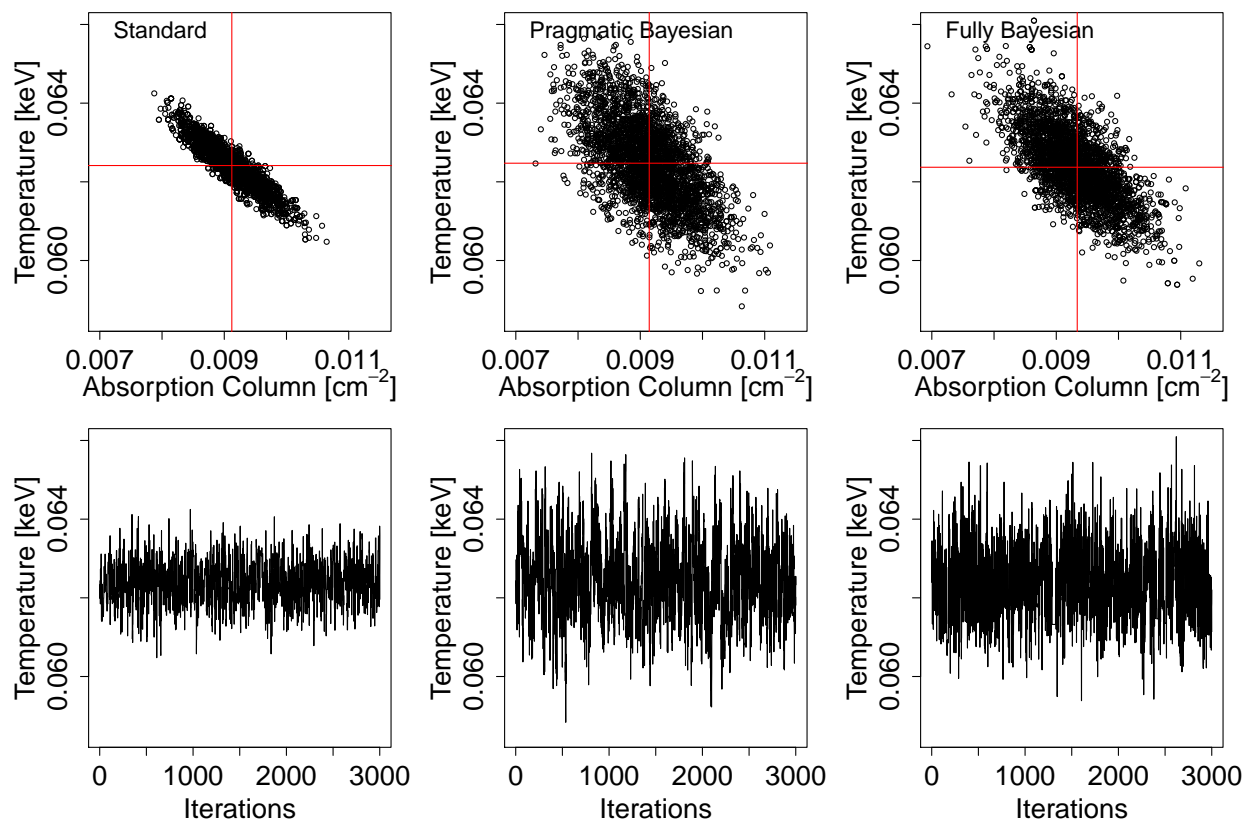


Fig. 12.— The standard, pragmatic Bayesian, and fully Bayesian fits of the temperature and absorption column in the spectrum of RX J1856.5-3754. The plots are layered as in Figure 8. In this case, however, there is only a small shift of the fitted values (posterior means given by the solid red lines). There is a noticeably less joint uncertainty in the two parameters under the fully Bayesian fit than under the pragmatic fit. There is no noticeable sticking for any of the MCMC samplers; compare the second row with that of Figure 8.

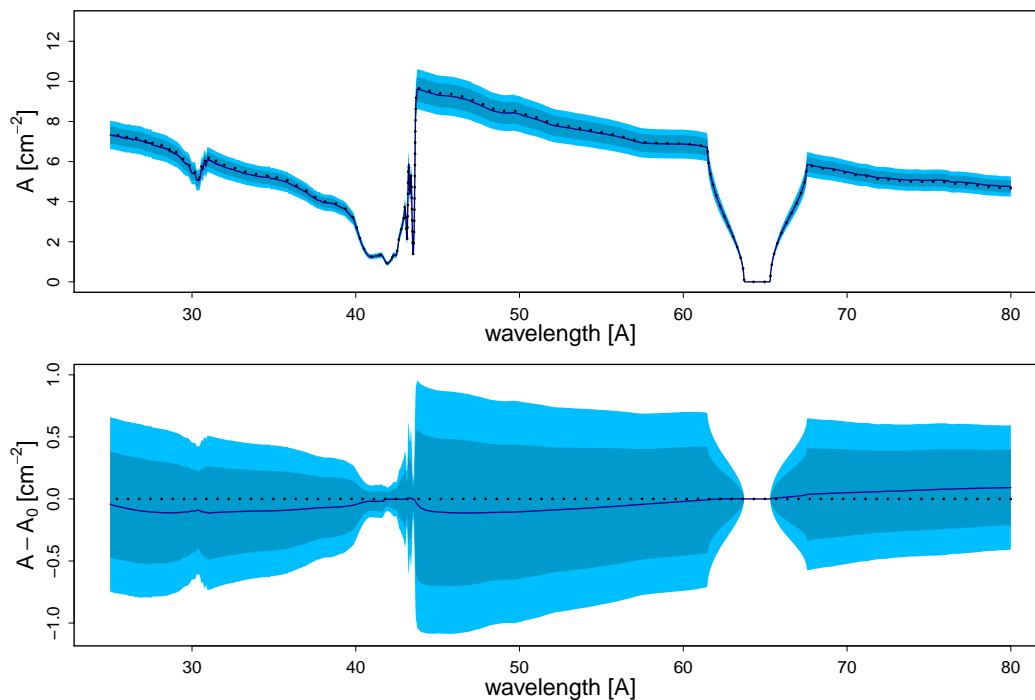


Fig. 13.— Estimating the range of effective area curves that are consistent with the spectrum of RX J1856.5-3754. The plots summarize  $p_{\text{FB}}(A | Y)$ . The pointwise posterior distribution of  $A$  is plotted in blue, where the dark (light) blue area corresponds to the central 68.3% (90%) region of  $p_{\text{FB}}(A | Y)$ . The pointwise posterior mean of  $A$  and its default value,  $A_0$  are plotted as solid blue and dotted black lines, respectively. The top panel displays the effective areas and the bottom panel shows the same data with the default effective area,  $A_0$ , subtracted off to highlight the differences. (We cannot subtract off the true curve as in Figure 4 because it is unknown.) Here the fitted (posterior mean) and default (dotted black curve) are very similar.

Table 5: Fit parameters for RX J1856.5-3754

Model	Standard		Pragmatic Bayesian		Fully Bayesian	
Parameters	Analysis		Analysis		Analysis	
$N_{\text{H}}$ [ $10^{20}$ $\text{cm}^{-2}$ ]	0.91	$\pm 0.043$	0.91	$\pm 0.058$	0.93	$\pm 0.056$
$T$ [eV]	62.4	$\pm 0.58$	62	$\pm 1.1$	62	$\pm 0.9$
Norm [ $\frac{L_{36}}{D_{10}^2}$ † ergs $\text{s}^{-1}$ $\text{cm}^{-2}$ ]	0.311	$\pm 0.0056$	0.31	$\pm 0.023$	0.32	$\pm 0.022$
Background scale	66	$\pm 1.5$	66	$\pm 1.8$	65	$\pm 1.7$

†  $L_{36}$  is the source luminosity in units of  $10^{36}$  ergs  $\text{s}^{-1}$  and  $D_{10}$  is the distance in units of 10 kpc.

method, the fully Bayesian method requires large-counts data sets to deliver significant gains over the standard method. In low-counts data sets, uncertainty stemming from random fluctuations in the counts swamps that due to calibration. The advantage of the fully Bayesian method stems from its ability to handle systematic errors due to calibration uncertainty in the same way one would handle statistical errors. In this way it accounts for calibration uncertainty by shifting the fitted values of spectral parameters. From a scientific perspective, this is preferable to increasing their reported error bars, the mechanism by which the pragmatic Bayesian method accounts for calibration uncertainty.

Fitting a spectral model under the fully Bayesian method poses significant computational challenges. We illustrate how we deal with this problem by leveraging the pragmatic Bayesian fit to deliver parameter values simulated under this fully Bayesian posterior distribution. This strategy allows us to simultaneously fit the effective area curve and the spectral parameters. Thus, we are able to use information in large-count observed spectra to narrow the uncertainty for the calibration product.

We focus on the application of the fully Bayesian methods to account for uncertainty in the effective area curve in X-ray spectral analysis. The general techniques we employ, however, have broad applicability in handling systematic errors. Obvious extensions include accounting for uncertainty in other calibration products, such as photon redistribution matrices (RMF), exposure maps, and point spread functions. Because all of these calibration

products exhibit more complex structure than an effective area curve—they are represented by matrices rather than vectors—more sophisticated methods will be needed to summarize their calibration libraries into concise and useable forms.

Less obviously, our methods can be used to account for other sources of systematic errors. In our fit of the radio-loud quasars in Section 4.1 and of the background-dominated isolated neutron star RXJ1856 in Section 4.3, for example, we used highly-structured background models that were originally fit to the blank-sky data provided by the *Chandra* X-ray Center or to the locally measured background from a spatially offset region. Only the normalization of this background model was fit to the individual sources. Just like an effective area curve, this background model is a vector that can only be specified with uncertainty. Ignoring this uncertainty can lead to biases and systematic errors. Similarly, the comprehensive atomic line emissivity database, AtomDB (Foster *et al.* 2012), is often used to specify a spectral models for X-ray data. While this database has been compiled by carefully combining empirical observations with theoretical calculations, its entries are not known exactly. Like a calibration product, its errors exhibit complex high-dimensional correlations that cannot be summarized with simple error bars for each entry. A better strategy is to compile an “AtomDB library” akin to a calibration library that can then be modeled to fully integrate uncertainty in AtomDB into individual spectral analyses. In principle, we can image fitting models that simultaneously account for uncertainty both in multiple calibration products and in multiple model components. In practice, this will involve significant modeling challenges such as accounting for correlations between calibration products and/or model components. Sophisticated computation methods will also be required, and large data sets needed to narrow uncertainty on multiple sources of systematic error. Although such work will involve substantial effort, it is likely to pay significant dividends in reducing bias stemming from calibration and model misspecification, at least when large-count data sets are available.

**Acknowledgements:** We dedicate this paper to the memory of Alanna Connors. Her commitment to disseminating principled statistical methods among astronomers and to

demonstrating their practical benefits in real problems not only influenced this work but the burgeoning discipline of astrostatistics as a whole. She was also an enthusiastic educator of statisticians, helping them to grapple with the subtleties of astrophysical data and models and to become actively involved in solving statistical challenges in astronomy. Although the impact of her efforts remains clearly felt, she is dearly missed.

This work was supported by NASA contract NAS8-03060 to the Chandra X-ray Center (VLK, AS, JD, and PR), NSF grants DMS 09-07522, DMS-09-07185, DMS-12-08791 and DMS-12-09232 (DvD, XLM, YY, AC, JX). In addition, DvD was supported in part by a British Royal Society Wolfson Research Merit Award, by a European Commission Marie-Curie Career Integration Grant, and by the STFC (UK).

## A. Statistical Computing

In this appendix we outline our computational strategy for fitting the fully Bayesian model. We begin in Section A.1 with a brief review of how Bayesian models can be fit using MCMC. As we discuss in Section 2.3, our fully Bayesian samplers are built upon a pragmatic Bayesian sampler. Thus, in Section A.2 we describe how we can improve the efficiency of the computational method presented in Paper I for fitting the pragmatic Bayesian model. Section A.3 describes the simple approximation to  $p_{\text{pB}}(\theta, A | Y)$  that we use in our fully Bayesian sampler. Additional details of our algorithms, samplers, and numerical methods can be found in Xu (2014).

### A.1. Fitting Bayesian Models Using MCMC

To compute fitted parameters and their error bars in a Bayesian analysis, we use quantities such as the mean, variance and percentiles of the posterior distribution. Although in some simple cases, the posterior distribution is a well-known distribution and these quantities can be derived analytically, numerical methods are typically used in practice.

One very popular numerical method is Monte Carlo which involves simulating replicate values,  $\{\psi^{(1)}, \dots, \psi^{(T)}\}$ , from the posterior distribution and using sample means, variances and percentiles of the replicate values to compute fitted parameters and error bars.

The most common Monte Carlo methods involve MCMC. (Brooks *et al.* (2011) offers a general reference for MCMC while van Dyk *et al.* (2001) and Park *et al.* (2008) reviews how it can be used in X-ray spectral analysis.) A Markov chain is constructed as an ordered sequence  $\{\psi^{(t)}, t = 1, 2, \dots\}$ , each value of which only depends on previous value in the chain. Specifically, suppose we were able to obtain a simulated value,  $\psi^{(0)}$ , from the posterior distribution, and using it we could randomly generated another value:

$$\psi^{(1)} \sim \mathcal{K}(\psi|\psi^{(0)}), \tag{A1}$$

so that  $\psi^{(1)}$  is also a simulated value from the posterior distribution. In Equation A1,  $\mathcal{K}$  is called the *kernel* of the MCMC sampler; it describes a random distribution that depends only on the previous value  $\psi^{(t-1)}$  and that we use to simulate a next value  $\psi^{(t)}$ . Iterating Equation A1 in this way results in a chain of values that is called a Markov chain, and that delivers a *correlated sample* from the posterior distribution. In practice we must start the chain at a typically arbitrary value,  $\psi^{(0)}$ , and let it run until it *converges* to the posterior distribution. Convergence can be checked by running multiple chains and waiting until they appear to sample from the same distribution (e.g., van Dyk *et al.* 2001; Gelman and Shirley 2011).

Deriving kernels that are easy to use and result in MCMC samplers that are fast to converge is key in practice. Two popular methods for deriving MCMC samplers are the Gibbs sampler (e.g., Casella and George 1992) and the Metropolis-Hastings (MH) algorithm (e.g., Hastings 1970). The Gibbs sampler proceeds by sampling each of several components of  $\psi$  for their conditional posterior distributions given the other components of  $\psi$ . Specifically, again letting  $\psi = (\theta, A)$ ,

**Two-Step Gibbs Sampler:**



For  $t = 0, 1, 2, \dots, T$ ,

**Step 1:** Draw  $A^{(t+1)} \sim p(A \mid \theta^{(t)}, Y)$ .

**Step 2:** Draw  $\theta^{(t+1)} \sim p(\theta \mid A^{(t+1)}, Y)$ .

The Partially Collapsed Gibbs (PCG) sampler reduces the conditioning in one or more steps of a Gibbs Sampler while maintaining the validity of the resulting algorithm (van Dyk and Park 2008). For example, if we draw  $A$  in Step 1 of the two-step Gibbs sampler without conditioning on  $\theta$ , i.e.,  $A^{(t+1)} \sim p(A \mid Y)$  the resulting sampler would be a PCG sampler. PCG samplers typically converge faster than their parent Gibbs samplers.

The MH algorithm draws from a proposal distribution,  $g(\psi \mid \psi^{(t)})$ , that is different from the posterior distribution, but can be corrected using a rejection step. Specifically,

**MH Algorithm:**

For  $t = 0, 1, 2, \dots, T$ ,

**Step 1:** Draw  $\psi^{\text{prop}} \sim g(\psi \mid \psi^{(t)})$  and compute

$$\rho = \frac{p(\psi^{\text{prop}} \mid Y)g(\psi^{(t)} \mid \psi^{\text{prop}})}{p(\psi^{(t)} \mid Y)g(\psi^{\text{prop}} \mid \psi^{(t)})}.$$

**Step 2:** Let  $u$  be a uniformly distributed random number between zero and one and set

$$\psi^{(t+1)} = \begin{cases} \psi^{\text{prop}} & \text{if } u < \rho \\ \psi^{(t)} & \text{otherwise} \end{cases}.$$

For a given  $\psi^{(t)}$ , the Gibbs sampler and the MH algorithm both provide a mechanism to randomly generate  $\psi^{(t+1)}$ . We refer to this random mechanism as the kernel, denoted by  $\mathcal{K}$ .

The Sherpa module, `pyBLoCXS` (Bayesian Low Count X-ray Spectral analysis in Python) is adopted from the method of van Dyk *et al.* (2001) and uses MH algorithms

for Bayesian spectral analysis in the Sherpa environment. `pyBLoCXS` provides a convenient and efficient MCMC sampler for drawing the spectral parameter,  $\theta$ , from its posterior distribution given  $A$ , that is from  $p(\theta | Y, A)$ ; see Siemiginowska *et al.* (2011) and <http://hea-www.harvard.edu/astrostat/pyblocxs/> for details of `pyBLoCXS`.

In this article, we use Gibbs-type samplers to draw  $\theta$  and  $A$  from their joint posterior distribution. We accomplish Step 2 of the Gibbs sampler using `pyBLoCXS`. That is, `pyBLoCXS` is incorporated as a component of an overall algorithm that simulates from the full posterior distribution,  $p(\theta, A|Y)$ . To clarify its roll in the overall algorithm, we refer to the `pyBLoCXS` kernel as  $\mathcal{K}_{\text{pyB}}$ . Using an MH algorithm for one or more of the component steps of a (partially collapsed) Gibbs sampler is a common hybrid strategy known as MH within (partially collapsed) Gibbs sampling (e.g., van Dyk and Jiao 2014). Our algorithm for implementing the fully Bayesian method builds on the algorithm for the pragmatic Bayesian method. For computational efficiency, we have modified and improved the latter compared with that given in Paper I. Here, we first review the algorithm from Paper I and then detail our improvements to it.

## A.2. Fitting the Pragmatic Bayesian Model

Fitting the pragmatic Bayesian model involves simulating a sequence of effective area curves from  $p(A)$  and running `pyBLoCXS` for each simulated curve. Specifically, in Section 4.2.2 of Paper I, we proposed

### MH within PCG Sampler of Paper I

For  $t = 0, 1, 2, \dots, T$ ,

**Step 1:** Draw  $e_j \sim \mathcal{N}(0, 1)$  for  $j = 1, \dots, J$  and set

$$e_{\text{pB}}^{(t+1)} = (e_1^{(t+1)}, \dots, e_J^{(t+1)}) \quad \text{and} \quad A_{\text{pB}}^{(t+1)} = A_0^* + (\bar{A} - A_0) + \sum_{j=1}^J e_j^{(t+1)} r_j v_j.$$

**Step 2 (Inner Loop):** For  $i = 1, \dots, I$ , simulate  $\theta_{\text{pB}}^{(t+i/I)} \sim \mathcal{K}_{\text{pyB}}(\theta | \theta^{(t+(i-1)/I)}; Y, A^{(t+1)})$ .

The iteration of `pyBLoCXS` in Step 2 aims to eliminate the dependence of its final output,  $\theta_{\text{pB}}^{(t+1)}$ , on its starting value,  $\theta_{\text{pB}}^{(t)}$ . Only  $\theta_{\text{pB}}^{(t+1)}$  is retained; the intermediate draws from the inner loop,  $\theta_{\text{pB}}^{(t+1/I)}, \dots, \theta_{\text{pB}}^{(t+(I-1)/I)}$  are discarded. (Notice that the superscripts on the discarded draws of  $\theta$  are fractional; we only retain the draws of  $\theta$  with integer superscripts.) This is necessary because  $A_{\text{pB}}^{(t+1)}$  is simulated from its marginal distribution in Step 1 and is independent of  $\theta_{\text{pB}}^{(t)}$ . The simulated value  $\theta_{\text{pB}}^{(t+1)}$  should be correlated with  $A_{\text{pB}}^{(t+1)}$ , insofar as  $A$  is informative for  $\theta$  and thus  $A$  and  $\theta$  are correlated under  $p_{\text{pB}}(\theta, A | Y)$ . Likewise, because  $A_{\text{pB}}^{(t)}$  and  $A_{\text{pB}}^{(t+1)}$  are independent,  $\theta_{\text{pB}}^{(t)}$  and  $\theta_{\text{pB}}^{(t+1)}$  should also be independent. The number of inner iterations,  $I$ , is determined by examining the empirical autocorrelation function of  $\mathcal{K}_{\text{pyB}}$  for a given data analysis. Its value should be set to ensure that  $\theta_{\text{pB}}^{(t)}$  and  $\theta_{\text{pB}}^{(t+1)}$  are approximately independent, see Paper I and van Dyk and Jiao (2014) for discussion.

While this MH within PCG sampler effectively delivers fitted values and error bars that account for calibration uncertainty under  $p_{\text{pB}}(\theta, A | Y)$ , it can be very slow in terms of its required computational time. This is owing to the inner workings of `pyBLoCXS` which approximates  $p(\theta | A, Y)$  by fitting the spectral model via maximum likelihood (i.e., using the Cash statistic) in Sherpa. This is of little consequence when `pyBLoCXS` is used with fixed  $A$ , because the Sherpa fit only needs to be preformed once. In the context of the MH within PCG Sampler, however, `pyBLoCXS` must be reinitiated with a new Sherpa fit at every iteration because  $A_{\text{pB}}^{(t+1)}$  is updated at every iteration. Because each Sherpa fit requires about 6-8 seconds of CPU time, a run of 3000 MCMC iterations requires 5-7 hours on Sherpa fits.

Below, we describe an alternate implementation of the MH within PCG Sampler that is dramatically faster. In particular, it fully leverages each of the effective-area specific Sherpa fits by continuing to iterate as in Step 2 after the dependence on  $\theta_{\text{pB}}^{(t)}$  has worn off in order to obtain several simulated values of  $\theta$  per iteration. Specifically, we propose

### Iterated MH within PCG Sampler

For  $t = 0, 1, \dots, T$ ,

**Step 1:** Draw  $e_j^{(tM+1)} \sim \mathcal{N}(0, 1)$  for  $j = 1, \dots, J$  and set

$$e_{\text{pB}}^{(tM+1)} = (e_1^{(tM+1)}, \dots, e_J^{(tM+1)}) \quad \text{and} \quad A_{\text{pB}}^{(tM+1)} = A_0^* + (\bar{A} - A_0) + \sum_{j=1}^J e_j^{(tM+1)} r_j v_j.$$

**Step 2 (Inner Loop):** For  $i = 1, \dots, I$ , simulate  $\theta_{\text{pB}}^{(Mt+i/I)} \sim \mathcal{K}_{\text{pyB}}(\theta \mid \theta^{(t+(i-1)/I)}; Y, A^{(t+1)})$ .

**Step 3:** For  $i = 2, \dots, M$ , simulate  $\theta_{\text{pB}}^{(Mt+i)} \sim \mathcal{K}_{\text{pyB}}(\theta \mid \theta^{(Mt+i-1)}; Y, A^{(t+1)})$ .

At iteration  $t$ , we only retain  $e_{\text{pB}}^{(tM+1)}$ ,  $A_{\text{pB}}^{(tM+1)}$ , and  $\theta_{\text{pB}}^{(tM+1)}, \dots, \theta_{\text{pB}}^{(Mt+M)}$  and discard the intermediate draws of  $\theta$  obtained internally within the inner loop of Step 2. That is, we retain draws of  $\theta$  with integer valued superscripts and discard those with fractional superscripts. Notice that this strategy results in  $M$  times more simulated values of  $\theta$  than of  $A$ . The advantage is that the extra  $T(M - 1)$  simulations of  $\theta$  require relatively little computational time. The disadvantage is that these simulated values are correlated whereas the simulated values of the MH within PCG sampler of Paper I are essentially independent.

We illustrate this trade-off using the simulation study described in Section 2.2. The left and right columns of Figure 14 show the performance of the MH within PCG Sampler of Paper I and our Iterated Sampler (run with  $M = 10$ ), respectively. The first row is a time-series plot of the first 300 consecutive draws of  $\Gamma^{(t)}$  obtained with the two samplers. The blocky nature of the draws obtained with the Iterated Sampler is a result of its infrequent updating of  $A$ ; each block is drawn with a common effective area curve. The second row presents the same time-series plots, but as a function of CPU time (in second) rather than iteration number. The relative speed of the Iterated Sampler is apparent. The final row shows scatterplots of the simulated values of  $(N_{\text{H}}^{(t)}, \Gamma^{(t)})$  obtained in about one hour. Again the advantage of the Iterated Sampler is clear, despite the blocky nature of its simulated values.

Running the Iterated Sampler requires deciding on the values of both  $I$  and  $M$ , that is, the number of initial simulated values of  $\theta$  to discard ( $I - 1$ ) at each iteration and the number of additional extra simulated values to keep ( $M$ ). The initial iteration in Step 2 is designed to mitigate the dependence of the simulated values on the input value of  $\theta$  at each iteration, e.g., the dependence of  $\theta^{(Mt+1)}$  on  $\theta^{(Mt)}$ . To measure this effect, we can compute

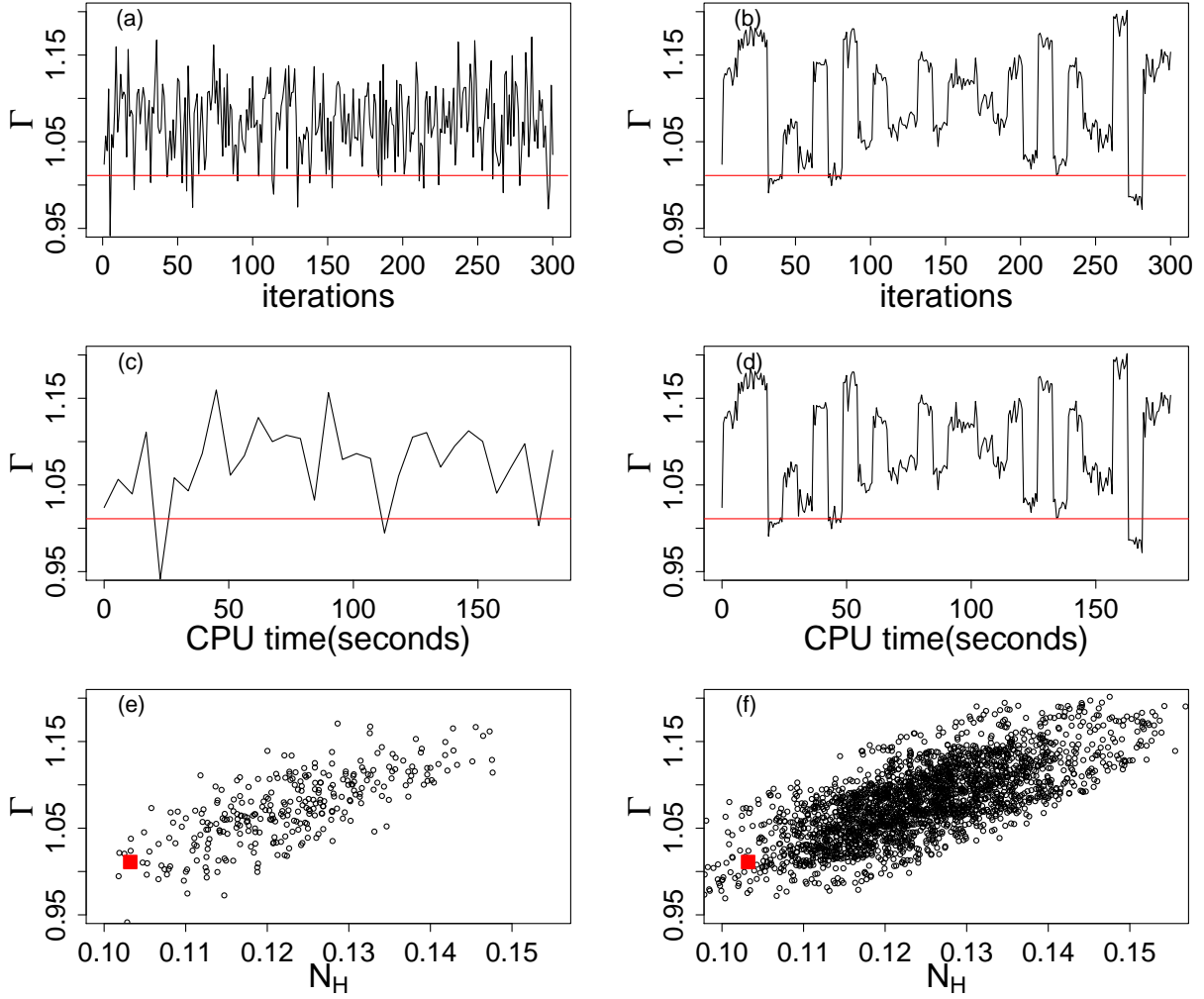


Fig. 14.— The improved speed of the iterated MH within PCG sampler. Using the simulation study described in Section 2.2, panels (a)-(b) plot  $\Gamma^{(t)}$  as a function of iteration number using the sampler of Paper I and the iterated MH within PCG sampler, respectively. Panels (c)-(d) plot the same, but as a function of CPU time rather than of iteration. The relative speed of the iterated method is apparent. Panels (e)-(f) are scatterplots of  $(N_H^{(t)}, \Gamma^{(t)})$  obtained in a run of about one hour for each of the two samplers. The red lines in panels (a)-(d) and squares in panels (e)-(f) indicate the true parameter values. Compare panels (e) and (f): the iterated sampler obtains far more draws per unit time. Thus, despite the blocky nature of the chains in panels (b) and (d), the advantage of the iterated sampler in terms of computational speed is clear.

the correlation of  $\theta^{(Mt)}$  and  $\theta^{(Mt+1)}$ ,

$$\hat{\alpha} \approx \frac{\sum_{t=1}^T (\theta^{(Mt)} - \hat{\theta}_{\text{pB}})(\theta^{(Mt+1)} - \hat{\theta}_{\text{pB}})}{\sum_{t=1}^T (\theta^{(Mt)} - \hat{\theta}_{\text{pB}})^2}, \quad (\text{A2})$$

where  $\hat{\theta}_{\text{pB}} = \frac{1}{TM} \sum_{t=1}^{TM} \theta_{\text{pB}}^{(t)}$ . A statistical hypothesis test of the correlation being zero can be rejected if  $T\hat{\alpha}^2/(1 - \hat{\alpha}^2)$  is greater than about two (for large  $T$ , e.g.  $> 100$ ). If this threshold is met, the sampler should be rerun with a larger value of  $I$ . Van Dyk and Jiao (2014) discuss other methods for choosing and validating  $I$  in the general MH within PCG sampler. (The choice of  $I$  is less important when implementing the fully Bayesian method, see Section A.3.)

An optimal value of  $M$  can be obtained by minimizing the Monte Carlo variance of the estimated posterior expectation of  $\theta$ . This quantity is approximated with  $\hat{\theta}_{\text{pB}}$  and is a common choice for the fitted value of  $\theta$ . Xu (2014) shows how the Monte Carlo variance of  $\hat{\theta}_{\text{pB}}$  depends on  $M$  and how it can be approximately minimized on the fly within the Iterated Sampler. Xu (2014) also discusses various choices of error bars for  $\hat{\theta}_{\text{pB}}$  that may be less affected by the blocky nature of the chain generated by the Iterated Sampler. We do not pursue this topic here, however, because the Iterated Sampler is only an intermediate step toward our goal of an algorithm for fitting the fully Bayesian model.

### A.3. Approximating the Pragmatic Posterior Distribution

As discussed in Section 2.3, although  $p_{\text{pB}}(\theta, A | Y)$  is ideally suited as an over-dispersed approximation to  $p_{\text{fB}}(\theta, A | Y)$ , it cannot be used as the proposal distribution,  $g$ , in the Pragmatic Proposal Sampler for the simple reason that we cannot directly evaluate  $p_{\text{pB}}(\theta, A | Y)$ . (We need to evaluate  $g$  in order to compute  $\rho$  given by Equation 11.) The problem becomes evident if we write, recalling that  $\theta$  and  $A$  are a priori independent under the pragmatic Bayesian assumption,

$$p_{\text{pB}}(\theta, A | Y) = p(\theta | A, Y)\pi(A) = \frac{\mathcal{L}(Y | \theta, A)\pi(\theta)}{p(Y | A)}\pi(A).$$

Evaluating the denominator,  $p(Y | A) = \int p(Y | \theta, A)\pi(\theta)d\theta$ , would require a possibly high-dimensional numerical integration at each iteration of the Pragmatic Proposal Sampler. To avoid this difficulty, we propose a simple approximation to  $p_{\text{pB}}(\theta, A | Y)$  that serves equally well as an over-dispersed approximation of  $p_{\text{fB}}(\theta, A | Y)$ .

Although  $p_{\text{pB}}(\theta, A | Y)$  is difficult to evaluate it has a relatively simple form. In particular, its marginal distribution for  $A$  can be represented by a  $J \times 1$  vector,  $e$ , the components of which are independent  $\mathcal{N}(0, 1)$  variables. (Recall that under the pragmatic Bayesian model,  $A$  is simulated according to its prior distribution, see Equation 5). Thus, we only need to approximate  $p_{\text{pB}}(\theta | A, Y)$ . To do this we run the Iterated Sampler described in Section A.2 to obtain  $\{e_{\text{pB}}^{(tM+1)}, \theta_{\text{pB}}^{(tM+1)}, \dots, \theta_{\text{pB}}^{(tM+M)}\}$ , for  $t = 0, \dots, T$  and regress the  $M$  replicates of  $\theta$ , (i.e.,  $\theta_{\text{pB}}^{(tM+1)}, \dots, \theta_{\text{pB}}^{(tM+M)}$ ) on each  $e_{\text{pB}}^{(tM+1)}$  using multivariate Gaussian regression (details are given in Xu 2014). This results in a multivariate normal approximation to  $p_{\text{pB}}(\theta | A, Y)$  that we combine with a multivariate standard normal distribution for  $e$  to form the  $g(\theta, e)$  used in the Pragmatic Proposal Sampler. In particular, this choice of  $g$  serves as the needed over-dispersed approximation to  $p_{\text{fB}}(\theta, A | Y)$ . In this context we run the Iterated MH within PCG Sampler only to obtain the approximation,  $g$ . Because  $g$  does not need to exactly match  $p_{\text{pB}}(\theta, A | Y)$ , the choice of  $I$  in this sampler is less critical. The choice of  $M$ , on the other hand, controls how quickly the sampler runs, and thus remains important for efficient computation.

## REFERENCES

- Anders, E. and Grevesse, N. (1989). Abundances of the elements - Meteoritic and solar. *Geochimica et Cosmochimica Acta* **53**, 197–214.
- Bevington, P. and Robinson, D. (1992). Testing the fit. *Data Reduction and Error Analysis for the Physical Sciences* 194–198.
- Bridle, S., Crittenden, R., Melchiorri, A., Hobson, M., Kneissl, R., and Lasenby, A. (2002). Analytic marginalization over CMB calibration and beam uncertainty. *Monthly Notices of the Royal Astronomical Society* **335**, 4, 1193–1200.
- Brooks, S., Gelman, A., Jones, G. J., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. Chapman & Hall / CRC, London.
- Casella, G. and George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician* **46**, 3, 167–174.
- Drake, J. J., Marshall, H. L., Dreizler, S., Freeman, P. E., Fruscione, A., Juda, M., Kashyap, V., Nicastro, F., Pease, D. O., Wargelin, B. J., and Werner, K. (2002). Is RX J1856.5-3754 a quark star? *The Astrophysical Journal* **572**, 996–1001.
- Drake, J. J., Ratzlaff, P., Kashyap, V., Edgar, R., Izem, R., Jerius, D., Siemiginowska, A., and Vikhlinin, A. (2006). Monte Carlo processes for including Chandra instrument response uncertainties in parameter estimation studies. In *Astronomical Telescopes and Instrumentation*, 62701I–62701I. International Society for Optics and Photonics.
- Foster, A. R., Ji, L., Smith, R. K., and Brickhouse, N. S. (2012). Updated atomic data and calculations for X-ray spectroscopy. *The Astrophysical Journal* **756**, 2, 128.
- Freeman, P., Doe, S., and Siemiginowska, A. (2001). Sherpa: a mission-independent data analysis application. In J.-L. Starck and F. D. Murtagh, eds., *Astronomical Data Analysis*, vol. 4477 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 76–87.



- Gelman, A. and Shirley, K. (2011). Inference from simulations and monitoring convergence. In *Handbook of Markov Chain Monte Carlo* (Editors: S. Brooks, A. Gelman, G. Jones, and X.-L. Meng), 163–174. Chapman & Hall/CRC Press.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 1, 97–109.
- Heinrich, J. and Lyons, L. (2007). Systematic errors. *Annual Review of Nuclear and Particle Science* **57**, 1, 145.
- Herve, A., Rauw, G., and Naze, Y. (2013). Hot gas distribution in the wind of  $\zeta$  pup and  $\zeta$  ori. In *Massive Stars: From  $\alpha$  to  $\Omega$ , held 10-14 June 2013*.
- Ho, W. C., Kaplan, D. L., Chang, P., van Adelsberg, M., and Potekhin, Alexander, Y. (2007). Magnetic hydrogen atmosphere models and the neutron star RX J1856.5-3754. *Monthly Notices of the Royal Astronomical Society* **375**, 821–830.
- Lee, H., Kashyap, V. L., van Dyk, D. A., Connors, A., Drake, J. J., Izem, R., Meng, X. L., Min, S., Park, T., Ratzlaff, P., Siemiginowska, A., and Zezas, A. (2011). Accounting for calibration uncertainties in X-ray analysis: Effective areas in spectral fitting. *The Astrophysical Journal* **731**, 126–144 (Paper I).
- López-García, M. A., López-Santiago, J., Albacete-Colombo, J. F., Pérez-González, P. G., and de Castro, E. (2013). A comprehensive study of NGC 2023 with XMM-Newton and Spitzer. *Monthly Notices of the Royal Astronomical Society* **429**, 775–791.
- Nevalainen, J., David, L., and Guainazzi, M. (2010). Cross-calibrating X-ray detectors with clusters of galaxies: an IACHEC study. *arXiv preprint arXiv:1008.2102* .
- Park, T., van Dyk, D. A., and Siemiginowska, A. (2008). Searching for narrow emission lines in X-ray spectra: Computation and methods. *The Astrophysical Journal* **688**, 807–825.

- Pollock, A. (2007). A new paradigm for the x-ray emission of O stars from XMM-Newton observations of the O9.7 supergiant  $\zeta$  Orionis. *Astronomy & Astrophysics* **463**, 1111–1123.
- Raassen, A. J. J., van der Hucht, K. A., Miller, N. A., and Cassinelli, J. P. (2008). Xmm-newton observations of  $\zeta$  Orionis (O9.7 Ib): A collisional ionization equilibrium model. *Astronomy & Astrophysics* **478**, 513–520.
- Siemiginowska, A., Kashyap, V., Refsdal, B., van Dyk, D., Connors, A., and Park, T. (2011). pyBLoCXS: Bayesian low-counts X-ray spectral analysis in Sherpa. In *Astronomical Data Analysis Software and Systems XX (Editors: I. N. Evans, A. Accomazzi, D. J. Mink, and A. H. Rots )*, vol. 442, 439–442.
- Siemiginowska, A., LaMassa, S., Aldcroft, T. L., Bechtold, J., and Elvis, M. (2008). X-ray properties of the gigahertz peaked and compact steep spectrum sources. *The Astrophysical Journal* **684**, 811–.
- van Dyk, D. A., Connors, A., Kashyap, V., and Siemiginowska, A. (2001). Analysis of energy spectra with low photon counts via Bayesian posterior simulation. *The Astrophysical Journal* **548**, 224–243.
- van Dyk, D. A. and Jiao, X. (2014). Metropolis-Hastings within partially collapsed Gibbs samplers. *Journal of Computational and Graphical Statistical (to appear)* .
- van Dyk, D. A. and Park, T. (2008). Partially collapsed Gibbs samplers: Theory and methods. *Journal of the American Statistical Association* **103**, 790–796.
- Waldron, W. L. and Cassinelli, J. P. (2001). Chandra discovers a very high density x-ray plasma on the O star  $\zeta$  Orionis. *The Astrophysical Journal* **548**, L45–L48.
- Xu, J. (2014). *Accounting for Calibration Uncertainty in Detectors for High-Energy Astrophysics*. Ph.D. thesis, Department of Statistics, University of California, Irvine.

Yamauchi, S., Kamimura, R., and Koyama, K. (2000). Asca x-ray observations of the NGC 2023 and NGC 2024 regions. *Publications of the Astronomical Society of Japan* **52**, 1087–1096.