Concordance: In-flight Calibration of X-Ray Telescopes without Absolute References

Herman L. Marshall[1] , Yang Chen[2] , Jeremy J. Drake[3] , Matteo Guainazzi[4], Vinay L. Kashyap[3] , Xiao-Li Meng[5] ,
Paul P. Plucinsky[3] , Peter Ratzlaff[3], David A. van Dyk[6] , and Xufei Wang[5]

[1] Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA
hermanm@space.mit.edu
[2] University of Michigan, Ann Arbor, MI 48109, USA
[3] Harvard-Smithsonian Center for Astrophysics, 60 Garden St., Cambridge, MA 02138, USA
[4] ESTEC, Keplerlaan 1 2201AZ Noordwijk, The Netherlands
[5] Harvard University, Cambridge, MA 02138, USA
[6] Imperial College, London SW7 2AZ, UK

Abstract

We describe a process for cross-calibrating the effective areas of X-ray telescopes that observe common targets. The targets are not assumed to be "standard candles" in the classic sense, in that we assume that the source fluxes have well-defined, but a priori unknown values. Using a technique developed by Chen et al. that involves a statistical method called *shrinkage estimation*, we determine effective area correction factors for each instrument that bring estimated fluxes into the best agreement, consistent with prior knowledge of their effective areas. We expand the technique to allow unique priors on systematic uncertainties in effective areas for each X-ray astronomy instrument and to allow correlations between effective areas in different energy bands. We demonstrate the method with several data sets from various X-ray telescopes.

Unified Astronomy Thesaurus concepts: Flux calibration (544); Astronomical methods (1043); X-ray astronomy (1810); Calibration (2179)

Supporting material: machine-readable tables

1. Introduction

We address a perennial issue in instrument performance, when estimated fluxes using two or more instruments disagree. While many instruments can safely rely on calibration traceable to established standards, the performance of a space-based telescope usually cannot be recalibrated, as the instruments are not returned to the lab. Furthermore, space-based instruments may be affected by the physical rigors of launch into space and instrument performance can change with time due to gas leakage, filter deterioration, component failure, contamination buildup, and other reasons. Without absolute standards that may be observed while in space, astronomers generally resort to the use of secondary, astronomical standards that can be observed while operating. Nonvariable sources are generally chosen as secondary standards so that they can be reused by the telescope team and observed by others. For example, in the 2–8 keV energy range, the Crab Nebula was frequently used as a standard, especially the observation by Toor & Seward (1974); for a recent use of the Crab Nebula for cross-calibrating X-ray telescopes, see Kirsch et al. (2005) for a comparison between many missions and Madsen et al. (2017b) for a case of assessing the NuSTAR telescope effective model. However, most astronomical sources vary, even the Crab Nebula, at levels detectable by current instruments, necessitating another approach: joint in-flight observations for cross-calibration of instruments.

Our work is set upon the foundation established by the International Astronomical Consortium for High Energy Calibration (IACHEC). IACHEC was formed primarily to assist X-ray telescope teams who cross-calibrate instruments and to understand the sources used for this purpose. See recent IACHEC reports for summaries of recent activity (Madsen et al. 2019, 2020). All of the IACHEC working groups address issues of cross-calibration of X-ray telescopes and often find discrepancies between results for the same source. Examples of IACHEC work include observations of the supernova remnants (SNRs) G21.5 (Tsujimoto et al. 2011) and 1E 0102−7219 (Plucinsky et al. 2017), spectra of galaxy clusters (Nevalainen et al. 2010; Kettula et al. 2013; Schellenberger et al. 2015), spectra of white dwarfs and isolated neutron stars (Beuermann et al. 2006), and simultaneous observations of active galaxies such as 3C 273 and PKS 2155−304 (Ishida et al. 2011; Madsen et al. 2017a). The studies had a common problem: assessing how much any particular instrument's effective area should be adjusted so that the measurements might agree. Ordinary weighting of measurements based on photon-counting statistics would give the observations with large effective areas and exposure times the greatest influence on the result but without consideration of possible systematic errors. The overarching goal of IACHEC is thus to bring the competing adjustments to instrumental effective areas into concordance, while simultaneously including prior knowledge of possible systematic errors.

Here, we further develop the "shrinkage" method pioneered by Chen et al. (2019), hereafter referred to as Paper I, to compute objective corrections to effective areas of several high-energy instruments. See Section 2.1 for the model setup and notation. The method is extended to account for systematic uncertainties specific to each instrument and incorporate systematic correlations across passbands. We present the method here in some generality, without the details of the mathematical machinery presented in Paper I (Section 2.2),
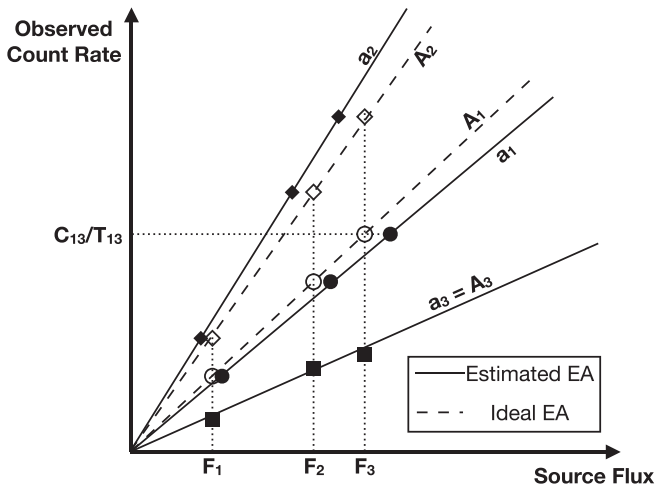
Figure 1. Goal of Concordance illustrated schematically. This schematic supposes three sources and three instruments, and plots expected count rates, C_{ij}/T_{ij}, on the vertical axis and plots true and estimated fluxes on the horizontal axis. The three instruments are represented by different symbols, with solid symbols representing the naive estimates of F_j, i.e., $f_{ij} = c_{ij}/(T_{ij}a_i)$, and open symbols representing the same, but with a_i replaced by the true effective area, A_i. The estimates are systematically biased because the a_i are inaccurate estimates of A_i; compare the solid and dashed lines. (In the plot EA is used as an abbreviation for effective area.) Our aim is to estimate the A_i values that yield best agreement between instruments for each source.

along with its extensions (Section 2.3), and apply it to a variety of data sets (Section 3). We present the results in Section 4, along with an assessment and simulation of the method, and a simulation study that aims to validate the method, and discuss the next steps in Section 5. We note that while we examine the case of X-ray telescopes in this paper, the method is extendable to most types of telescopes.

2. Method

2.1. Background

We start with an idealized calibration data set where M objects are observed by each of N instruments, obtaining photon counts, c_{ij}, where $j = 1,...,M$ indexes the sources and $i = 1,...,N$ indexes the instruments.[7] Each observation is described by a set of observational parameters (e.g., exposure time) that we encapsulate in a matrix $T = \{T_{ij}\}$.

Denoting the true effective area of instrument i by A_i and the true flux of source j by F_j, the expected counts C_{ij} for each object/detector combination is given by

$$C_{ij} = T_{ij} A_i F_j, \quad 1 \leqslant i \leqslant N, \quad 1 \leqslant j \leqslant M, \quad (1)$$

where T_{ij} has units of seconds × counts per photon, F_j has units of photons per unit area per second, and A_i has units of area. We assume for now that the true exposure factors have negligible error and equal the observed values, i.e., $T_{ij} = t_{ij}$. We follow the notation of Paper I by using lower case to indicate measured quantities and upper case to indicate the "true" values to be estimated. Note that, in fact, the multiplicative constant T_{ij} contains not only the exposure time but also other factors that can be calculated precisely for any given observation.

---

[7] We consider each unique combination of telescope and detector with any filters or gratings to be an "instrument."

We aim to estimate F_j using the calibration data, i.e., the observed counts c_{ij}, and an external (prior) estimated effective area, a_i for each instrument. A naive procedure simply substitutes each quantity in Equation (1) with its measured counterpart, to obtain the "estimating equation,"

$$c_{ij} = t_{ij} a_i f_{ij} \quad (2)$$

and solving Equation (2) for f_{ij} for each instrument–source combination, thus yielding N different estimates of each flux. The resulting estimator $f_{ij} = c_{ij}/(a_i t_{ij})$ is a *ratio estimator*, which is known in statistical literature to be both seriously biased and highly variable. More precisely, the variability in the denominator can cause large uncertainty (considering dividing by a value close to zero). Furthermore, the average, $\frac{1}{N}\sum_{i=1}^{N} f_{ij}$, is a biased estimator of F_j. We return to the issue associated with ratio estimators in Section 4.4.1.

We can do much better by analyzing all data together using more principled and sophisticated statistical methods, such as the one given in Paper I. We can then achieve our goal to obtain *better* estimates of the instrumental effective areas, A_i, that bring our flux estimates, f_{ij}, closest to the F_j in some strict statistical sense. Figure 1 gives a schematic representation of our goal.

In practice, estimating a flux from an observation is not as simple an operation as merely counting events. A strictly correct approach takes into consideration the response of the instrument, especially when attempting to measure the flux in a specific bandpass. It is often necessary to use a forward-folding method, such as implemented in xspec, that takes a count spectrum as input and effective area and response functions as externally (and accurately) provided. Here, we approximate this process using Equation (1) because the flux (as defined here) is often a robustly computed quantity for any given observation. In fact, we do not actually compute c_{ij} but a proxy for c_{ij} as described below.

We now carefully examine the flux measurement process to justify using the simplistic model represented by Equation (1). Source j is assumed to have a photon spectrum $n_E(\Theta_j)$ in units of photons per unit area per unit energy at energy E. Each Θ_j represents a vector of spectral parameters for a source, which we assume to include an overall normalization n_j with the same units as n_E, so $q(E; \theta_j) \equiv n_E(\Theta_j)/n_j$ defines the spectral shape as a function of the remaining spectral parameters, θ_j. In the case of a spectrometer measuring an emission line, the bandpass may be so narrow that q is well approximated by a delta function, in which case $F_j = n_j$. Otherwise, the flux in a band from E_1 to E_2 is

$$F_j = \int_{E_1}^{E_2} n_E(\Theta_j) dE = n_j \int_{E_1}^{E_2} q(E; \theta_j) dE \quad (3)$$

giving

$$n_j = \frac{F_j}{\int_{E_1}^{E_2} q(E; \theta_j) dE} \equiv \frac{F_j}{\tilde{q}_j \Delta E} \quad (4)$$

where $\Delta E = E_2 - E_1$.

For most X-ray telescopes, the instrumental effective area can be separated into two parts. There is the geometric area of the optics, i.e., A_i^g, with losses due to mechanical obscuration, reflections, and transmissions of the optics (including filters) given by $r_i(E)$. The other part is due to the quantum efficiency of the detector, $Q(E)$, which gives the probability of a detecting

a photon. We characterize the (true) effective area of instrument i as a function of energy in the bandpass of interest by $\tilde{A}_i(E) = A_i^g r_i(E) Q_i(E) \equiv A_i \alpha_i(E)$, where $\alpha_i(E)$ now gives the shape of the effective area in the band and is defined such that its integral over the band is unity, i.e.,

$$\alpha_i(E) = \frac{r_i(E) Q_i(E)}{\int_{E_1}^{E_2} r_i(E) Q_i(E) dE}. \tag{5}$$

Consequently,

$$A_i = A_i^g \cdot \int_{E_1}^{E_2} r_i(E) Q_i(E) dE \tag{6}$$

is the scale of the effective area and does not depend on E. We take this approach because the model of the effective area through the band is generally better known than its absolute value.

The detector provides counts in K channels that are related to the true photon energy, E, via a response function $\Phi_k(E)$, where

$$\sum_{k=1}^{K} \Phi_k(E) = 1. \tag{7}$$

A given observation has an accurately determined exposure time t_{ij} that sets the expected count in channel k:

$$C_{ijk} = t_{ij} \int \tilde{A}_i(E) n_E(\Theta_j) \Phi_k(E) dE \tag{8}$$

$$= t_{ij} A_i n_j \int \alpha_i(E) q(E; \theta_j) \Phi_k(E) dE \tag{9}$$

$$= t_{ij} A_i F_j \frac{\int \alpha_i(E) q(E; \theta_j) \Phi_k(E) dE}{\tilde{q}_j \Delta E}, \tag{10}$$

$$\equiv \tilde{T}_{ijk} A_i F_j, \tag{11}$$

where Equation (10) follows from Equation (4). It is important to point out that Equation (8) is well-known in the astronomy literature but Equation (10) is an innovative way of simplifying the expressions, which is essential for tackling the current problem. Finally, the counts that are associated with the energy band (E_1, E_2) are mostly in the range of channels (k_1, k_2), which is chosen so that we have a reliable estimate of flux in the band of interest, giving the expected count relevant to a particular flux measurement:

$$C_{ij} = \sum_{k=k_1}^{k_2} C_{ijk} = A_i F_j \sum_{k=k_1}^{k_2} \tilde{T}_{ijk} \equiv T_{ij} A_i F_j, \tag{12}$$

which defines T_{ij} in terms of t_{ij}, a normalized sum over the response function, and the two shape functions. This is the formal basis for Equation (1).

In actual data analysis, the observed counts, c_{ijk}, not the expected counts, C_{ijk}, are input into the iterative routine that estimates the fluxes, f_{ij}, and, again, the estimated effective area, a_i, must be used because the true effective area is unknown. Thus, in analogy to Equation (2), we replace C_{ij}, A_i, and F_j in Equation (12) with their observed counterparts to obtain the estimating equation

$$c_{ij} = \sum_{k=k_1}^{k_2} c_{ijk} = a_i f_{ij} T_{ij} = a_i f_{ij} \sum_{k=k_1}^{k_2} \tilde{T}_{ijk}. \tag{13}$$

Finally, naive, instrument-specific flux estimates are obtained by solving Equation (13),

$$f_{ij} = \frac{c_{ij}}{T_{ij} a_i}, \tag{14}$$

which is essentially a full derivation of the ratio estimator.

It is important to our analysis that the T_{ij} values be independent of A_i and F_j. There are two circumstances that may cause a problem in this regard: (1) when there is some nonlinearity of the detector response, such as pileup, where $Q_i(E)$ depends on the source flux, and (2) when $\alpha_i(E)$ is highly nonlocal and the bandpass of integration in Equation (8) is large, involving portions of the spectrum where $q(E; \theta_j)$ or $\alpha_i(E)$ are poorly determined. By choosing instrument data in the linear regime, avoiding pileup, and restricting the bandpasses of interest, we mitigate these potential issues.

The goal of our statistical modeling is to determine the best estimates of A_i and F_j consistent with the data, c_{ij} with uncertainties σ_{ij}. If all the effective areas were precisely correct, i.e., if $A_i = a_i$, then we could estimate the F_j via a relatively trivial regression model. However, we know that there are systematic errors in our estimated effective areas because observations show that the f_{ij} do not cluster about any given value within their individual uncertainties. We addressed this problem in Paper I by introducing estimates of the systematic errors on the estimated effective areas and applying a statistical method called *shrinkage estimation*.

2.2. Statistical Model and Shrinkage Estimates

Paper I proposes a linear regression model for the log count rates, denoted $y_{ij} \equiv \log(c_{ij}/T_{ij})$, such that

$$y_{ij} = B_i + G_j - \frac{\sigma_i^2}{2} + e_{ij}, \tag{15}$$

where the $B_i \equiv \log A_i$, and the $G_j \equiv \log F_j$ are the quantities to be estimated, and the e_{ij} are noise terms that are assumed normally distributed with mean 0 and (known) variance σ_i^2. The $-\frac{\sigma_i^2}{2}$ term in Equation (15) is a half-variance correction that is included to maintain the multiplicative mean modeling in Equation (12). This correction ensures that $E(c_{ij}) = C_{ij} = T_{ij} A_i F_j$ because if $\log x \sim N(\mu, \nu)$, then $E(x) = e^{0.5 \nu + \mu}$.

Paper I adopts a Bayesian hierarchical modeling approach to estimate the unknown quantities B_i, G_j, and σ_i^2. This involves setting independent Gaussian prior distributions for the B_i, with prior mean $b_i = \log a_i$ and prior variance τ_i^2, setting independent flat priors for the G_i, and setting independent conjugate priors for the σ_i^2. Paper I goes on to show how a Hamiltonian Monte Carlo algorithm can be used to obtain a sample from the joint posterior distribution of all unknown quantities in Model (15) and cross-checks this computation with a blocked Gibbs sampler. The resulting Monte Carlo sample can be transformed back to the effective areas and source fluxes on their original scale to obtain their posterior distributions, estimates, and error bars.

A Bayesian perspective allows us to combine multiple sources of information—in this case the information from the calibration data, y_{ij}, and from the prior estimates of the effective areas, a_i. By replacing the estimated effective areas with prior distributions that reflect their uncertainty, we are able to update the estimated effective areas and their error bars in light of the

Table 1
Effective Area Uncertainty Priors (τ_i)[a]

Instrument	Energy Bands (keV)								
	0.15–0.33	0.33–0.54	0.54–0.8	0.8–1.2	1.2–1.8	1.8–2.2	2.2–3.5	3.5–5.5	5.5–10
Astrosat SXT	⋯	15	15	10	10	10	10	10	10
Chandra ACIS	3	3	3	3	2.6	3.3	3.3	4.9	5
Chandra HETGS	⋯	⋯	10	5	4	4	4	5	7
Chandra LETGS	5	7	7	7	7	7	7	10	10
ROSAT PSPC	10	10	10	10	10	10	⋯	⋯	⋯
Suzaku XIS1	⋯	20	15	10	10	15	5	5	5
Suzaku XIS0,2,3	⋯	⋯	15	10	10	15	5	5	5
Swift PC/WT	⋯	15	10	7.5	7.5	10	5	5	5
XMM MOS1,2	20	10	6	6	6	6	6	6	10
XMM pn	2	2	2	2	2	2	2	2	3
XMM RGS	⋯	8	5	5	5	⋯	⋯	⋯	⋯

Note.
[a] The τ_i values are given as percentages. The ellipses indicate bandpasses where the instrument has an negligible effective area.

calibration data. This approach provides improved estimates (e.g., in terms of mean squared error) of the effective areas and the fluxes simultaneously. As a result, we obtain estimates of the sources' true fluxes that combine the instrument-specific estimates in a statistically principled manner.

The updated estimates of the effective areas are weighted averages of their priors and the best values based on the current calibration data. In the context of Model (15), we work on the log scale. The estimates of B_i and G_j are given by

$$\widehat{B}_i = W_i(\bar{y}_{i\cdot}' - \bar{G}_i) + (1 - W_i)b_i \quad \text{and}$$
$$\widehat{G}_j = \bar{y}_{\cdot j}' - \bar{B}_j, \qquad (16)$$

where

$$\bar{y}_{i\cdot}' = \frac{\sum_{j=1}^{M} y'_{ij}\sigma_i^{-2}}{\sum_{j=1}^{M}\sigma_i^{-2}}, \quad \bar{y}_{\cdot j}' = \frac{\sum_{i=1}^{N} y'_{ij}\sigma_i^{-2}}{\sum_{i=1}^{N}\sigma_i^{-2}},$$

$$\bar{G}_i = \frac{\sum_{j=1}^{M}\widehat{G}_j\sigma_i^{-2}}{\sum_{j=1}^{M}\sigma_i^{-2}}, \quad \bar{B}_j = \frac{\sum_{i=1}^{N}\widehat{B}_i\sigma_i^{-2}}{\sum_{i=1}^{N}\sigma_i^{-2}}, \qquad (17)$$

$y'_{ij} = y_{ij} + 0.5\sigma_i^2$, and the weights are given by $W_i = M\sigma_i^{-2}/(\tau_i^{-2} + M\sigma_i^{-2})$. If the prior estimate of the effective area of a particular instrument is very precise relative to its calibration data, i.e., $\tau_i^2 \ll \sigma_i^2/M$, then $W_i \approx 0$, and the updated estimate of that instrument's effective area is dominated by the prior estimate, resulting in $\widehat{B}_i \approx b_i$. In contrast, if the calibration data are much more precise, then the weights are near unity and the updated estimate of the effective area is dominated by the calibration data, giving $\widehat{B}_i \approx \bar{y}_{i\cdot}' - \bar{G}_i$.

The estimates $\widehat{B}_i$ in Equation (16) are often called "shrinkage estimates" due to their historical use for "shrinking" several estimates together toward a common prior mean (Efron & Morris 1972, 1973) when, for example, the b_i are all the same. Because the prior means, b_i, are different for different instruments, the $\widehat{B}_i$ are simply a sensible combination of prior knowledge captured by b_i and data represented by $\bar{y}_{i\cdot}'$, weighted by their respective precisions, which are the reciprocals of their variances (assuming the σ_i are known). This combination allows our model to weigh the prior estimate of the effective area for a given instrument against deviations between the observed fluxes of the same sources from different instruments

and ultimately to obtain the joint estimates of the true fluxes and effective areas that are most consistent with the calibration data and the priors on the effective areas.

Paper I further describes how to handle the case where all sources are not observed by all instruments and presents a robust version of Model (15) that allows for outliers among the measured fluxes (or source counts) by replacing the log normal error model with a log t error model.

2.3. Extensions of the Model

Paper I proposed modeling calibration data using Model (15) and its extensions, derived computational methods for fitting these models, and validated their statistical properties. However, application of Model (15) to real data requires relaxing some of its basic assumptions. Here, we illustrate how this is accomplished via two extensions to the method.

2.3.1. Heterogeneous Uncertainties in Effective Area Priors

IACHEC scientists recognize that the quality of ground-based calibrations varies significantly from instrument to instrument, resulting in perceived differences in the reliabilities of the estimated effective areas. The formalism laid out in Paper I allows for instrument-specific prior distributions for the B_i, as explained in Section 2.2, given by Gaussian distributions with instrument-specific variances τ_i^2. In the numerical examples in Paper I, however, the τ_i^2 were set to τ^2 for each i, essentially assuming that all modeled calibrations are equally uncertain in percentage terms. Here we allow for heterogeneous τ_i^2 values, as covered by the theory given in Paper I. At IACHEC meetings in 2017, 2018, and 2019 (Madsen et al. 2019, 2020), we asked instrument calibration scientists to specify values of τ_i for their instruments in each of a specific set of bandpasses. These values are given in Tables 1 and 2; instruments with significant effective area below 1 keV appear in Table 1 and other instruments appear in Table 2.

Of course, in practice it is difficult even for experts to quantify τ_i precisely. Thus, it is important that we examine the sensitivity of our results to the specified values. Often, however, there is a body of experience and expert knowledge on the reliability of ground-based standards that allows rough estimation of systematic errors.

Table 2
Effective Area Uncertainty Priors (τ_i)[a]

Instrument	Energy Bands (keV)						
	2.2–3.5	3.5–5.5	5.5–10	15–25	25–50	50–100	100–300
Astrosat CZTI	⋯	⋯	⋯	20	20	20	25
Astrosat LAXPC	⋯	15	15	15	15	20	⋯
INTEGRAL IBIS	⋯	⋯	⋯	⋯	8	15	20
INTEGRAL SPI	⋯		⋯	⋯	5	5	5
NuSTAR	⋯	4	3	3	15	20	⋯
RXTE PCA	5	10	3	3	10	50	⋯
RXTE HEXTE	⋯	⋯	⋯	5	5	5	⋯
Suzaku HXD	⋯	⋯	⋯	20	20	20	20
Swift BAT	⋯	⋯	⋯	15	4	4	12

Note.
[a] The τ_i values are given as percentages.

2.3.2. Prior Correlations among Effective Areas

The second extension allows correlations between the effective areas in different energy bands for each instrument. In Paper I, we treated different energy bands as separate (independent) instruments, while in reality their effective areas can be strongly correlated. Continuing to work on the log relative scale given in Equation (23), we denote the effective areas as a function of the energy band $\mathcal{E} = [E_1, E_2]$ by $B(\mathcal{E}, \boldsymbol{\xi}) = \log \int_{E_1}^{E_2} A(E, \boldsymbol{\xi})\, dE$, where $\boldsymbol{\xi}$ parameterizes the effective area and includes quantities such as the geometric area, filter thicknesses, and chemical compositions that are initially estimated during ground calibration. Uncertainties in $\boldsymbol{\xi}$ are quantified via the prior distribution, $p(\boldsymbol{\xi})$; examples generated and used for ACIS analyses can be found in Drake et al. (2006), Kashyap et al. (2008), Lee et al. (2011), and Xu et al. (2014). We suppress the subscript i throughout this section for notational simplicity because we are restricting consideration to an arbitrary instrument.

Following the discussion in Section 2.2, we specify the prior distribution on (the logarithm of) the effective areas of U energy bands, $\{B(\mathcal{E}_u, \boldsymbol{\xi})\}$ for $u = 1, \ldots, U$, as a multivariate Gaussian distribution with expected values, β_u, and variances, Λ_u, for each energy band u, and with correlations, ρ_{uv}, between all pairs of distinct energy bands, $u \neq v$. Among the b_u, Λ_u, and ρ_{uv}, only the correlations, ρ_{uv}, (or, more precisely, the Monte Carlo estimates of ρ_{uv}) are used in our data analyses. To distinguish the Monte Carlo estimates of the prior means and variances of the B_i obtained here from those we actually use, we introduce new notation for these quantities that differ from those used in Paper I and in Section 2.2. The current methods for computing Λ_u are still somewhat experimental, so we rely instead on the τ_i values from IACHEC scientists.

For a given instrument and energy band u, the expectation is

$$\beta_u = \int B(\mathcal{E}_u, \boldsymbol{\xi})\, p(\boldsymbol{\xi})\, d\boldsymbol{\xi}. \qquad (18)$$

In practice, calibration scientists set a_i to be the prior estimate of A_i (on the original scale) based on their best information and experience. Transforming to a logarithmic scale, we might set our prior estimate of B_i to $\log a_i$, which is the choice of prior mean we suggest in Section 2.2. Equation (18) can be used in the absence of such intuition or if we prefer to use a parameterized model for B. (This is particularly relevant for the correlations, since we have less intuition for them.) In this

case, a reasonable strategy is to proceed via Monte Carlo integration of Equation (18). More precisely, we obtain a *calibration sample*, $\{B^{(k)}(\mathcal{E}_u), k = 1, \ldots, K\}$, that quantifies prior uncertainty in $B(\mathcal{E}_u, \boldsymbol{\xi})$, for example by obtaining a sample of size K from $p(\boldsymbol{\xi})$ and computing $B(\mathcal{E}_u, \boldsymbol{\xi}^{(k)})$ for each sample, $\boldsymbol{\xi}^{(k)}$, from $p(\boldsymbol{\xi})$. The $B^{(k)}(\mathcal{E}_u)$ is expressed as $\log\left[\int_{E_1}^{E_2} A^{(k)}(E)\, dE\right]$ for each u (given all Monte Carlo samples $A^{(k)}(E)$ for $E \in \mathcal{E}_u$). The Monte Carlo version of Equation (18) is then

$$\hat{\beta}_u = \frac{1}{K} \sum_{k=1}^{K} B^{(k)}(\mathcal{E}_u). \qquad (19)$$

The prior variance of $B(\mathcal{E}_u, \boldsymbol{\xi})$ and its Monte Carlo estimate are

$$\Lambda_u = \int [B(\mathcal{E}_u, \boldsymbol{\xi}) - \beta_u]^2\, p(\boldsymbol{\xi})\, d\boldsymbol{\xi} \quad \text{and}$$

$$\hat{\Lambda}_u = \frac{1}{K-1} \sum_{k=1}^{K} [B^{(k)}(\mathcal{E}_u) - \hat{\beta}_u]^2, \qquad (20)$$

respectively. Finally, the prior correlation between $B(\mathcal{E}_u, \boldsymbol{\xi})$ and $B(\mathcal{E}_v, \boldsymbol{\xi})$ is the covariance normalized by the respective standard deviations,

$$\rho_{uv} = \frac{1}{\sqrt{\Lambda_u \Lambda_v}} \int [B(\mathcal{E}_u, \boldsymbol{\xi}) - \beta_u][B(\mathcal{E}_v, \boldsymbol{\xi}) - \beta_v]\, p(\boldsymbol{\xi})\, d\boldsymbol{\xi} \qquad (21)$$

with Monte Carlo estimate,

$$\hat{\rho}_{uv} = \frac{1}{(K-1)\sqrt{\hat{\Lambda}_u \hat{\Lambda}_v}} \sum_{k=1}^{K} [B^{(k)}(\mathcal{E}_u) - \hat{\beta}_u][B^{(k)}(\mathcal{E}_v) - \hat{\beta}_v]. \qquad (22)$$

2.4. Practical Implementation

This section discusses practical implementation of our methods, specifically in terms of normalization of observed counts/fluxes and computation of correlation matrices.

2.4.1. Normalization in Practice

Because X-ray data analysis packages such as xspec return the f_{ij} and their uncertainties, it is convenient to rewrite

Equation (15) in terms of

$$\tilde{y}_{ij} = \log \frac{f_{ij}}{\tilde{f}} = \log \frac{c_{ij}}{T_{ij} a_i \tilde{f}},$$

$$\tilde{B}_i = \log \frac{A_i}{a_i}, \quad \text{and} \quad \tilde{G}_j = \log \frac{F_j}{\tilde{f}} \qquad (23)$$

to obtain

$$\tilde{y}_{ij} = \tilde{B}_i + \tilde{G}_j - \frac{\sigma_i^2}{2} + e_{ij}, \qquad (24)$$

where $\tilde{f}$ is a fiducial flux (usually the maximum of the f_{ij}) used to normalize the data to the range [0, 1]. Model (24) is functionally equivalent to Model (15). This definition of $\tilde{G}_j$, normalized by $\tilde{f}$, which depends on data, is only introduced for computational convenience and does not affect the model or its interpretation. Technically, using a data-dependent "parameter," here $\tilde{G}_j$, implies a data-dependent prior distribution, which is generally not legitimate from a Bayesian viewpoint. However, because using a flat prior on G_j is the same as using a flat prior on $\tilde{G}_j = G_j - \log \tilde{f}$, there is no actual effect in our implementation. Thus, Model (24) has the same form as Model (15), so we can embed it into the Bayesian hierarchical model described in Paper I to obtain the full posterior distribution, estimates, and error bars for the $\tilde{B}_i$ and $\tilde{G}_j$. Paper I suggests setting $b_i = 0$ (because $\tilde{B}_i = 0$ implies that $A_i = a_i$).

2.4.2. Deriving Correlations in Practice

We proceed by computing numerous instances of instrument effective areas that are varied in controlled ways dictated by current knowledge of uncertainties in calibration. The basis of the method is to generate a so-called calibration sample of areas that represents the range of uncertainties of the effective area, *including all the correlations between different energies*. The approach to generating the calibration samples for the different instruments we study here is common to all instruments, with some additional complexity built into the Chandra samples. We describe the method in brief below and refer the reader to Drake et al. (2006) for a more complete description.

We devise a "perturbation function" that comprises piecewise cubic segments that stretch between the natural absorption edges of the different materials encountered along the optical path for a given instrument. This function varies about unity by random amounts but is constrained within fixed limits based on specified calibration uncertainties by the cubic function whose parameters are randomly drawn from a truncated Gaussian distribution. The perturbation function is applied as a multiplicative factor to the different subassembly component contributions to the effective area, which are considered on a case-by-case basis. For instance, in the case of Chandra/ACIS, six plausible mirror effective areas are used, uncertainties in the optical blocking filter and contamination and contamination transmittance are modeled by altering the optical depth of each chemical component within their known uncertainties and recomputing ensemble transmittance, and CCD quantum efficiencies are computed for different realizations of depletion depth and SiO_2 layers. Details of this process are given in Drake et al. (2006, see also J. J. Drake et al. 2021, in preparation).

Chandra/HETGS and Chandra/LETGS use a similar approach with additional perturbation functions applied for the transmission grating diffraction efficiencies. In the case of the other instruments considered here, the perturbation function approach alone is used.

Table 3
Heterogeneous τ_i Values for 1E0102 Analysis[a]

Instrument	Oxygen	Neon
XMM/RGS1	5	5
XMM/MOS1	6	6
XMM/MOS2	6	6
XMM/pn	2	2
ACIS-S3	3	3
ACIS-I3	3	3
ACIS/HETG	3	3
Suzaku/XIS0	15	10
Suzaku/XIS1	15	10
Suzaku/XIS2	15	10
Suzaku/XIS3	15	10
Swift/XRT-WT	10	7.5
Swift/XRT-PC	10	7.5

Note.
[a] Values for τ_i are in percentages for each combination of instrument and line complex, using τ_i values taken from Table 1.

3. Observations and Data Processing

Three data sets are considered in Paper I and we add a fourth in this paper. Here, we detail how the data sets are handled and the required data processing.

3.1. Supernova Remnant 1E0102.2-7219

As in Paper I, the fluxes of the emission line complexes of O and Ne in the X-ray spectra of SNR 1E0102.2−7219 are taken from the detailed comparison of 13 instruments by Plucinsky et al. (2017). Briefly, the spectra of each non-dispersive instrument are fit with a model with five free parameters: an overall normalization and four emission line fluxes of O VII, O VIII, Ne IX, and Ne X. Because the emission lines of the same element have similar energies, their effective areas are comparable and highly correlated, so we combined the O and Ne line fluxes to create two fluxes for each instrument. In our statistical analysis, these two fluxes are treated as "sources": one for O and another for Ne. The data are normalized to the O or Ne fluxes obtained by the XMM-Newton pn instrument. By requiring that each instrument analysis use the same model, except primarily for the strengths of the emission lines, the $q(E; \theta_j)$ values do not depend on the instrument, nor on the measured line fluxes.

Table 3 shows the τ_i values assigned to the effective areas for each instrument considered. The values were taken from Table 1 using the bandpass that covers the lines of interest: 0.54–0.80 keV for O and 0.8–1.2 keV for Ne. For the correlation matrix, there is only one off-diagonal term, which we set to 0.88 for ACIS instruments and 0.82 for XMM instruments, as derived as in Section 2.3.2.

3.2. Sources from the 2XMM Catalog

We select a sample of X-ray sources from the Second European Photon Imaging Camera (EPIC) Serendipitous Source (2XMM) Catalog (Watson et al. 2009). EPIC consists of three X-ray cameras with CCD sensors mounted on the ESA spacecraft XMM-Newton (Jansen et al. 2001); the EPIC-pn (Strüder et al. 2001) and two EPIC-MOS (Turner et al. 2001) cameras observe celestial sources quasi-simultaneously within their co-aligned fields of view. For this analysis, v14.0 of the Science Analysis System (SAS; Gabriel et al. 2004) is used, as well as the

Table 4
Heterogeneous τ Values for 2XMM and XCAL Analyses[a]

Instrument	Soft Band	Medium Band	Hard Band
pn	2	2	2.3
MOS1	6	6	7.3
MOS2	6	6	7.3

Note.
[a] Values for τ are percentages for each combination of instrument and line complex, using τ values from Table 1.

Table 5
Correlation Matrix for 2XMM Analyses

Band	Soft Band	Medium Band	Hard Band
Soft band	1	0.61	0.13
Medium band	0.61	1	0.53
Hard band	0.13	0.53	1

Table 6
Correlation Matrix for XCAL Analyses

Band	Soft Band	Medium Band	Hard Band
Soft band	1	0.63	0.20
Medium band	0.63	1	0.52
Hard band	0.20	0.52	1

calibration files as available in 2016. A description of the data reduction and spectral extraction procedure appears in Read et al. (2014). Soft, medium, and hard bands are defined to be the 0.5–1.5, 1.5–2.5, and 2.5–10 keV bands, respectively. Due to variability, different observations of the same source are treated as separate sources for a total of 41 observations of 35 distinct sources. The normalizing (maximum) fluxes in the soft, medium, and hard bands are 0.138, 0.000701, and 0.00223 photons cm^{-2} s^{-1}, the brightest sources in their respective lists. The data are provided in Table A1 in the Appendix. There are two primary features of the analysis procedure that make the results suitable for Concordance analysis: (1) the count spectra for the different instruments (pn, MOS1, MOS2) are fit simultaneously to power laws, so that the spectral slopes, $\theta_j = \Gamma_j$ (where $n_E[\Theta_j] = n_j E^{-\Gamma_j}$) do not depend on the instrument combination, and (2) the sources are faint enough that pileup is not an issue.

Table 4 gives the τ_i values assigned to the effective areas for the pn and MOS instruments and the three bandpasses. The values are taken from Table 1 using the 0.54–0.8 and 0.8–1.2 keV τ_i values for the soft band, the 1.2–2.2 keV τ_i value for the medium band, and an average of the 2.2–10 keV τ_i values for the hard band. Table 5 gives the correlation matrix values, ρ_{mn}, computed for the pn and MOS instruments and the three bandpasses used in the 2XMM catalog while Table 6 provides these values for the bandpasses used in the so-called "XMM-Newton Cross-Calibration" (XCAL) analysis. For simplicity, we assume that the ρ_{mn} are the same for each instrument.

3.3. Active Galaxies from the XCAL Sample

Another set of EPIC spectra used to validate the model is the XCAL sample.[8] This is a sample of radio-loud active galactic nuclei, primarily blazars, observed routinely by XMM-Newton in the framework of its in-flight calibration program (Guainazzi et al. 2015). As with the 2XMM sample described in Section 3.2, the sources are variable. In this case, there are more blazars and high signal observations, for a total of 108, 103, and 94 observations of 22 distinct sources in the soft,

medium, and hard bands that exceeded a flux limit criterion without highly discrepant fluxes between the three instruments. The normalizing (maximum) fluxes in the soft, medium, and hard bands are set to 0.126, 0.0156, and 0.0154 photons cm^{-2} s^{-1}, the brightest sources in their respective lists. Data are provided in Table A1 in the Appendix. As with the 2XMM sources, the pn, MOS1, and MOS2 data were fit simultaneously to power-law spectra so that Γ_j is the same for each instrument. However, compared to the 2XMM sources, the XCAL sources are bright, often exceeding the count rate threshold beyond which the fraction of events affected by pileup is no longer negligible (Jethwa et al. 2015). Spatial regions on the detector affected by pileup are removed by excising the core of the telescope point-spread function (PSF) up to an observation-dependent radius. This radius is determined on the basis of the ratio between non X-ray diagonal and standard X-ray "patterns" (measure of the event shape in the CCD) in EPIC-MOS (Jethwa et al. 2015), and by visual inspection of the pattern distribution curves in EPIC-pn using the SAS task `epatplot` in EPIC-pn. This "PSF core excising method," while unavoidable to retain the highest possible fidelity of event spectral calibration, may introduce excess variance in the f_{ij} via systematic uncertainty in the energy-dependent correction for the fraction of events scattered into or out of the annular spectral extraction region by the PSF (the so-called "encircled energy correction" fraction). Values for τ_i and ρ_{mn} are the same as for the 2XMM sample.

3.4. Active Binary Capella

Capella (α Aur AB; G1III+G8III; 13 pc) is a spectroscopic binary that is the brightest line-dominated source accessible to non-solar X-ray missions. It is remarkably steady for a coronal source, having never exhibited significant flaring. While it does vary over timescales of months, it does not show any evidence of flux variability over timescales of weeks or less. Consequently, it has often been used as a calibration target, in particular with Chandra. It has been observed several times with different detector and grating combinations in close proximity (see Table A2 in the Appendix). These observations allow us to carry out an assessment of the internal cross-calibration of the Chandra grating spectrometers.

We estimated the total fluxes in each of several strong lines: the highly-ionized lines of Fe XVII (at 15 and 17 Å) whose formation temperatures overlap the peak emission measures of Capella, and the hydrogenic lines of Ne X (12.13 Å) and O VIII (18.96 Å). For the purposes of this calculation, we treat each of the four emission lines as different sources. We then form 21 epoch groups, comprised of observations that are within 0.1 yr of each other,[9] giving a total of 84 sources. Similarly, +1 and

---

[8] Details of the XCAL processing are available in Section 4 of the XMM calibration memo XMM-SOC-CAL-TN-0052. The memo is available at https://xmmweb.esac.esa.int/docs/documents/CAL-TN-0052.ps.gz.

[9] Aug/Sep99, Mar00, Feb01, Apr02, Oct02, Sep03, Sep04, Mar05, Oct05, Apr06, Apr07, Apr08, Apr09, Nov09, Nov/Dec10, Dec11, Dec13, Dec14, Jul16, Sep16, and Dec18; see Table A2.

Table 7
Values of τ_i for Capella Analyses[a]

Line	Ne X λ12	Fe XVII λ15	Fe XVII λ17	O VIII λ19
HEG	4	5	5	10
MEG	4	5	5	10
LEG/A	4	5	5	10
LEG/H	7	7	7	7

Note.
[a] Values for τ are in percentages.

Table 8
ACIS Correlation Matrix Used for Capella

Line	Ne X λ12	Fe XVII λ15	Fe XVII λ17	O VIII λ19
Ne X λ12	1	0.96	0.92	0.89
Fe XVII λ15	0.96	1	0.99	0.97
Fe XVII λ17	0.92	0.99	1	0.99
O VIII λ19	0.89	0.97	0.99	1

-1 grating orders are treated distinctly for each of four grating/detector combinations, ACIS-S/HEG, ACIS-S/MEG, ACIS-S/LEG, and HRC-S/LEG, for a total of eight instruments. The fluxes were normalized to the maximum values for each emission line: 28.02, 60.23, 49.66, and 23.33×10^{-13} erg s^{-1} cm^{-2} for O VIII, Fe XVII 17 Å, Fe XVII 15 Å, and Ne X, respectively. We use *CIAO*v4.11 to extract the dispersed spectra, and compute the effective areas using the contamination corrections as in `CALDBv4.8.0.1`.

The values of τ_i are taken from Table 7 and the correlation matrices are given in Table 8 and 9.

4. Results

Here we present results from new measurements and extensions to the results in Paper I. In each case, we generate 10,000 Monte Carlo replicates from the respective posterior distributions as the basis for our statistical inferences.

4.1. 1E0102

These data provide a illustration of a case where there are many instruments that obtain data on the same source, shown in Figures 2 and 3. The effect of allowing heterogeneous τ values is apparent in both cases. Generally, when the prior distribution on an instrument's effective area is more uncertain than average, giving a relatively large value of τ_i, then the data for that instrument are given less weight, so the posterior estimate of its effective area is more likely to deviate from the prior estimate by comparison to when all instruments have equally uncertain prior estimates. In addition, the posterior range of the deviation is more likely to be large when τ_i is larger. The Suzaku results for the O lines all show this effect. When effective areas are correlated between the O and Ne data sets, the ACIS-I3 point is particularly affected due to the discrepant results obtained when Ne and and O data are considered independently.

4.2. XMM Samples

Figures 4 and 5 show the results of the Concordance analysis for the two XMM-Newton data. In this example, there are many sources and few instruments, in contrast with the 1E0102 data set. The 2XMM results show a high degree of consistency

Table 9
HRC Correlation Matrix Used for Capella

Line	Ne X λ12	Fe XVII λ15	Fe XVII λ17	O VIII λ19
Ne X λ12	1	0.84	0.71	0.62
Fe XVII λ15	0.84	1	0.83	0.74
Fe XVII λ17	0.71	0.83	1	0.91
O VIII λ19	0.62	0.74	0.91	1

between the instruments, consistently favoring 3%–5% increases to the effective areas of the MOS detectors across all bands and a corresponding slight decrease to the pn effective area. With the use of individualized τ values, the Concordance analysis drives the pn effective areas toward the prior, as one might expect due to the significantly smaller τ assigned to the pn compared to the MOS detectors.

The XCAL sample shows similar trends to that of the 2XMM sample with with a more significant indication that the MOS2 detector's effective area should be increased 2%–3% more than that of the MOS1 detector.

4.3. Capella Line Fluxes for Chandra Grating Spectrometers

Results from the Concordance analysis as applied to the Capella data are shown in Figure 6. There are several features of interest. First, the effective area corrections for the LETGS (ALEG and HLEG) are generally negative while those of the HETGS (HEG and MEG) are generally positive. These corrections are consistent with preliminary results on independent data where the instruments are cross-calibrated with alternating observations of Mk 421. Second, the $+1$ and -1 orders generally agree well for all instruments and wavelengths. Third, when the effective area correlations are included, the posterior effective areas for the longer wavelengths (O VIII and Fe XVII λ17) more strongly deviate from their priors.

4.4. Method Validation and Assessment

Paper I includes a series of numerical studies that explore the statistical properties of our method. For example, Figure 2 of Paper I illustrates that our posterior distributions of the effective areas cover the true effective areas in a simulation study. Figure 7 of Paper I then goes on to contrast the estimated 95% intervals for log-fluxes constructed using the standard instrument-specific estimates with the combined estimate based on our posterior distribution, illustrating how our Bayesian process achieves a single consistent estimate for each flux but with smaller errors than the standard estimates. Here we consider additional ways to evaluate how robust our method's results may be, supplementing the simulations performed in Paper I.

4.4.1. Simulation Studies

The method developed and applied in Paper I produces Bayesian posterior distributions for each estimated quantity. The main quantities of interest here are the fractional corrections to instrument effective areas, given by $\log A_i - \log a_i$, and the fractional corrections to the estimated fluxes of sources, given by $\log F_j - \log f_{ij}$.

We demonstrate how the Concordance method yields accurate and reliable estimates of A_i and F_j with a simulation study. The simulation involves 40 simulated sources observed by each of five instruments. We set the prior means of the effective areas of the instruments to differ from their actual
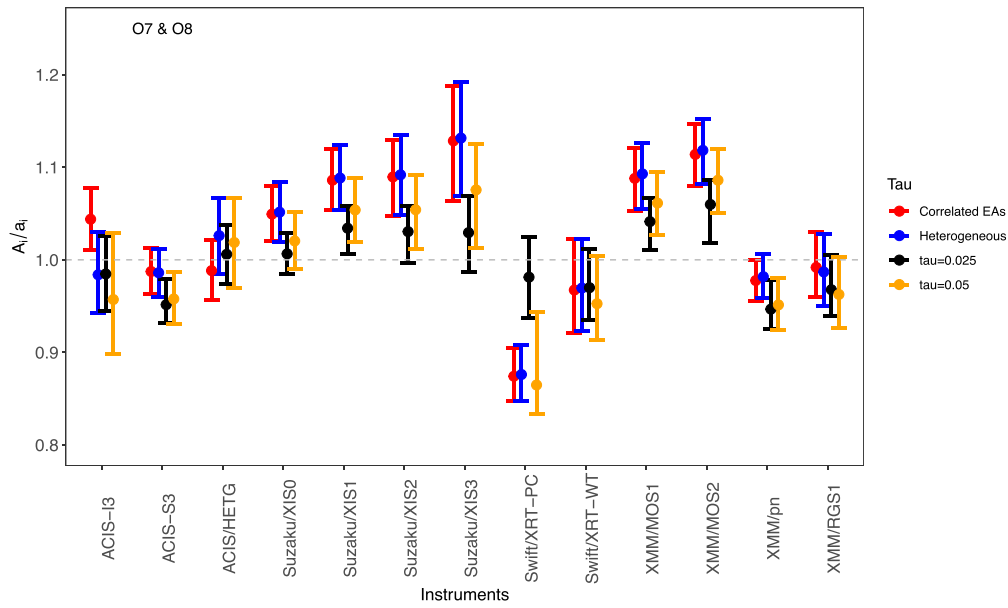
Figure 2. Results of the Concordance analysis for the data from SNR 1E0102 for the combination of fluxes of the lines of O VII and O VIII. The $\tau = 0.025$ (in black) and 0.05 (in yellow) results are the same as given in Paper I and are shown to elucidate the effects of including heterogeneous τ values (in blue) and adding effective area correlations (in red). The error bars represent the 90% (5%–95%) confidence regions on the posterior estimate of A_i/a_i, as defined in Section 2.2. When effective areas are correlated between the O and Ne data sets, the ACIS-I3 point is particularly affected due to the discrepant results obtained when Ne and and O data are considered independently.
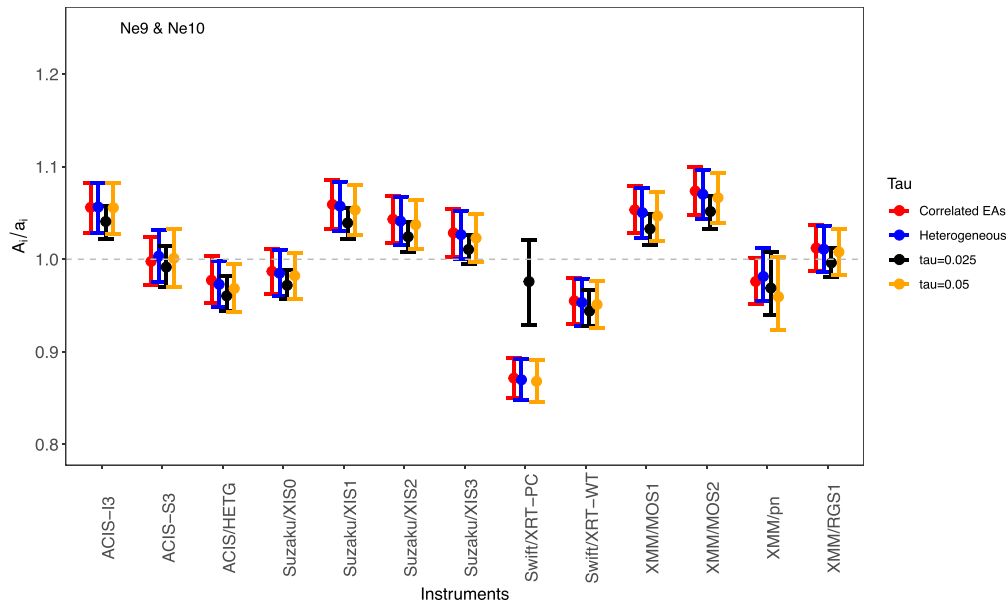
Figure 3. Same as Figure 2 except for the combination of fluxes of the lines of Ne IX and Ne X (see the text). Unlike the case for the O lines, the posterior estimates of the effective area tend to be more stable in this band and relatively independent of the uncertainties in the effective area priors.

effective areas by $\log A_i - \log a_i = [0, 1, -1, 2, -2]$, for $i = 1,...,5$, respectively. For example, for instrument 4, the true effective area is systematically higher than the prior mean by a factor of e^2, resulting in flux estimates that are systematically too high compared to the true values. The τ_i values are all 1.0 in this simulation, indicating large uncertainties in prior estimates of the effective areas, and the measurement uncertainties (i.e., σ_{ij}) are all set to 0.5 on the log scale, similarly indicating large uncertainties, except for instrument 5, for which $\sigma = 0.1$. This simulation setup is designed to test the robustness of the Concordance method to data from an instrument with high signal/noise but a systematically biased effective area. We replicated this entire set up 200 times and

processed each replicate with our Concordance method. A representative replicate is shown in Figure 7. The simulations demonstrate that the Concordance method provides source flux estimates that are substantially better than would be obtained by simply using the prior means as estimates of the effective areas. The replicate simulations indicate that the 95% equal-tailed posterior intervals cover the true values of the effective areas and source fluxes over 99% of the time. Thus, we not only obtain better estimates of the source fluxes, but also estimate the effective area corrections well. Note that the instrument-specific flux estimates can deviate substantially from the true values for any given source, so that when using a set of such estimates, the weighted flux estimator would be
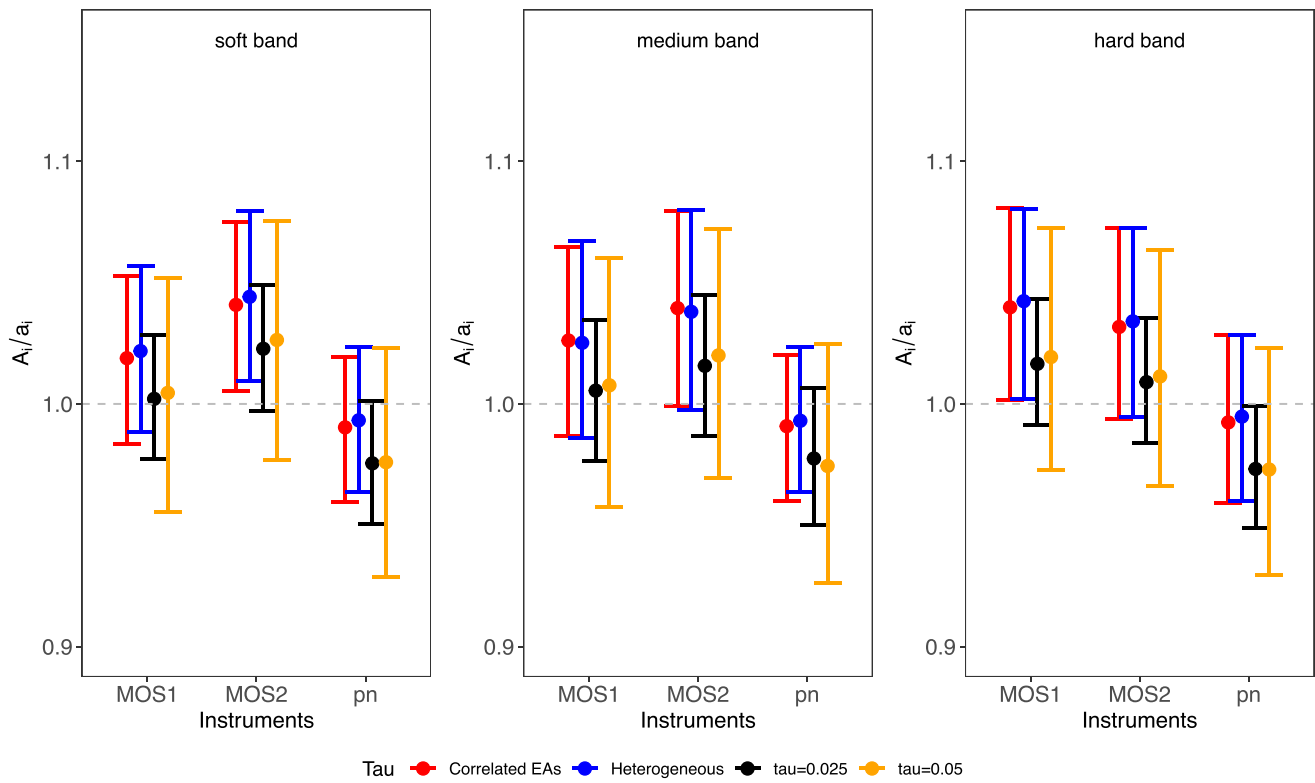
Figure 4. Concordance results for the 2XMM sample. Results are color-coded as in Figure 2. When the τ values are allowed to vary by instrument, the "heterogeneous" case, the posterior for the pn centers on the prior, due to the smaller value of τ than used for the MOS detectors. At the same time, higher effective areas for the MOS detectors are indicated across all bands.
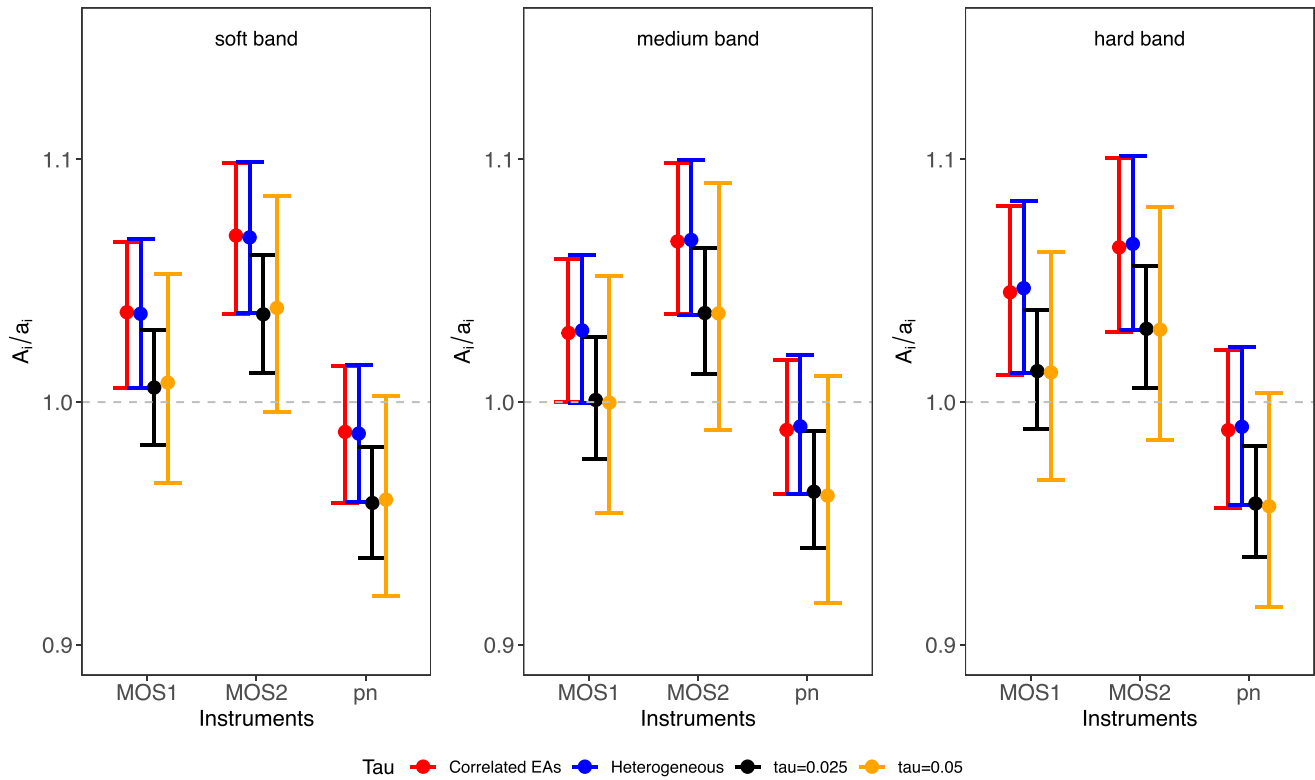

Figure 5. Same as Figure 4 except for the XCAL sample. These results are generally consistent with those from the 2XMM sample but with a somewhat stronger indication that the MOS2 effective area should be increased relative to MOS1.
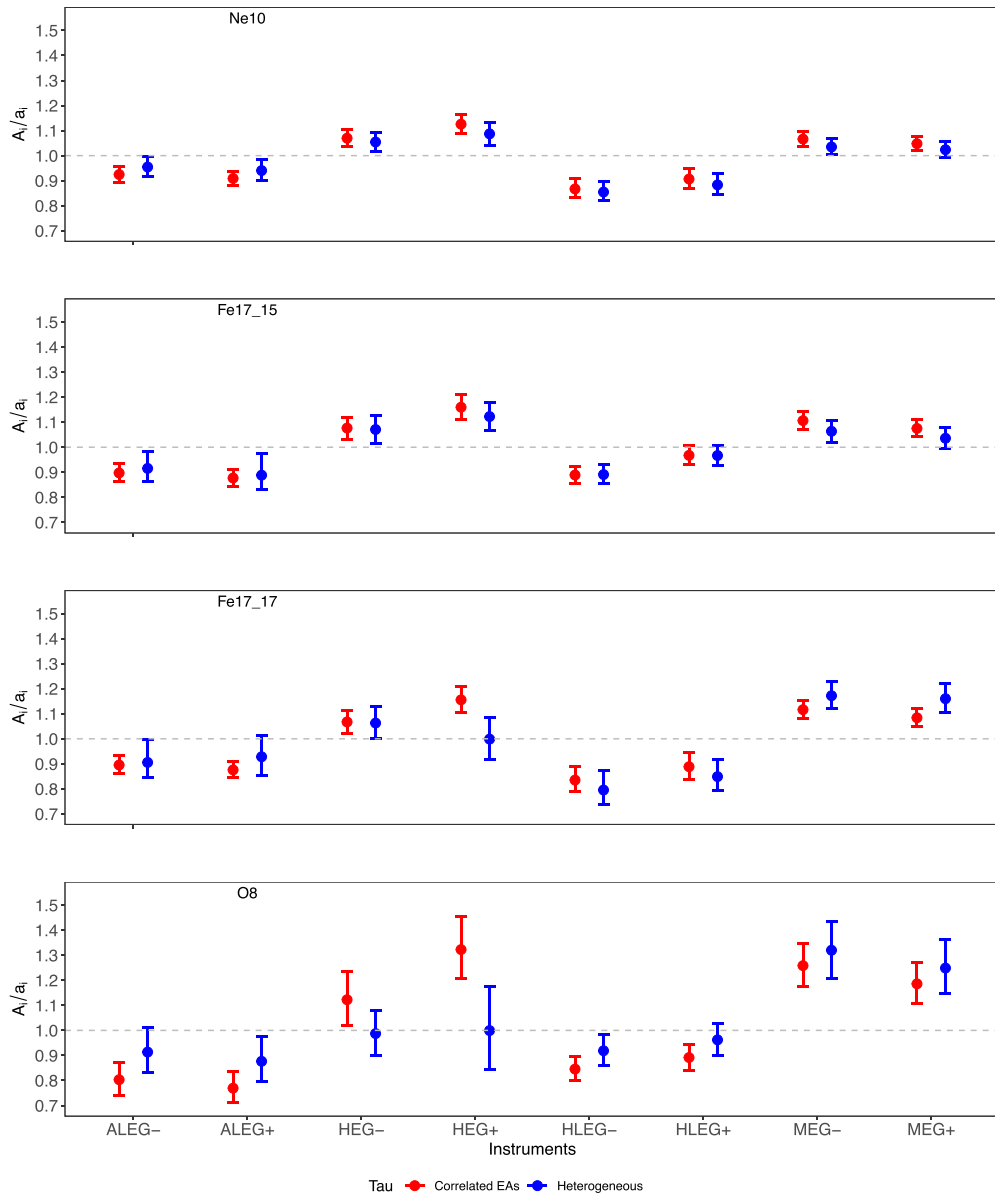
Figure 6. Concordance analysis using measurements of four emission lines using the the Chandra grating spectrometers. Emission lines used in the various panels are: (a) Ne X, (b) Fe XVII λ15, (c) Fe XVII λ17, (d) O VIII. Color coding of results are as in Figure 2. Features to note are (1) that the effective area corrections for the LETGS (ALEG and HLEG) are generally negative while those of the HETGS (HEG and MEG) are generally positive and (2) the $+1$ and -1 orders generally agree well.

generally biased toward the instrument with the highest signal/ noise, regardless of the accuracy of the instrument's effective area. Note that the instrument-specific flux estimates can deviate substantially from the true values for any given source, so that when using a set of such estimates, the *weighted average ratio estimators* for the flux (i.e., $[\sum_j f_{ij} \sigma_j^{-2}]/\sum_j \sigma_j^{-2}$) would be generally biased toward the instrument with the highest signal/noise, regardless of the accuracy of the instrument's effective area.

With an additional pair of simulations, we also quantified the improvement that can be obtained with the Concordance method. In the first case, we simulated $M = 3$ instruments with $A_i/a_i = [1, 1, 0.9]$ for $i = [1, 2, 3]$, $\tau_i = 0.05$, and $N = 20$ sources with true fluxes all equal to 1. There were 200 independent simulations and analyses for each setup. The sources were assumed to have good signal/noise as might be

expected for calibration observations: $\sigma_i = 0.03$ (i.e., the c_{ij} were drawn from a Poisson distribution with a mean of about 1100). The second case is the same as the first except there is a higher statistical precision for observations with instrument 3: $\sigma_3 = 0.003$. Source fluxes were estimated for each simulation using the Concordance method and also using the above-mentioned weighted average of ratio estimators: $(\sum_j f_{ij} \sigma_j^{-2})/\sum_j \sigma_j^{-2}$. The 95% uncertainty bounds on the flux estimates and the coverage fraction where the true flux is included within the uncertainty bounds are shown in Table 10. The Concordance estimator generated uncertainty intervals that were accurate and covered the ground truth, in contrast to the ratio estimator which, despite the widths of the confidence intervals being nominally smaller, generated biased estimates and confidence intervals that did not cover the true fluxes. The situation was worse for the second case, where the
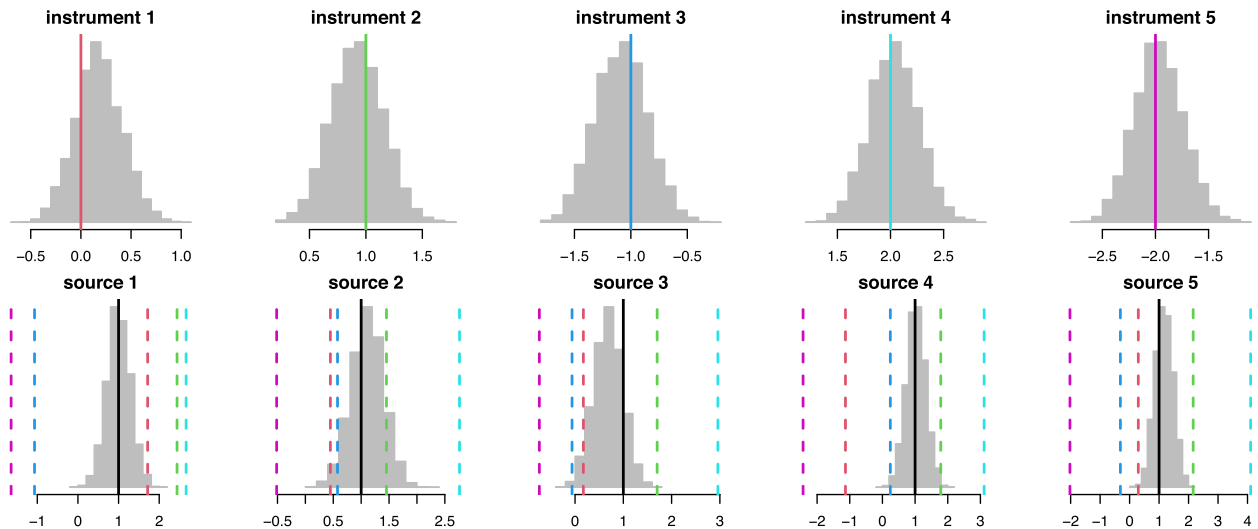
Figure 7. Posterior distributions of the logarithms of the effective area corrections (top row) and on the logarithms of the source fluxes (bottom row, for five sources) for one of 200 replicate simulations, each involving five instruments and 40 sources. In the simulation, all sources have a true value of $\log F_j = 1$ but the prior means of the instrument effective areas are off by factors of $\exp([0, 1, -1, 2, -2])$ for $i = 1,…,5$. The top row shows that the Concordance method generates posterior distributions of the effective area corrections that are well centered on the true values for each instrument (shown as vertical solid lines in various colors). The bottom row shows that the posterior distributions of the source fluxes are well centered on the true values (vertical solid black lines), even while the instrument-specific estimates based on the prior means of the effective areas (vertical dotted lines of colors corresponding to those of the instruments in the top row) can be individually erroneous by large factors.

<div style="text-align:center">

Table 10
Results from Two Concordance Simulations

</div>

Simulation Setup[a]	Flux Estimation Method	95% Flux Range[b]		r_{95}[c]
		F_{lo}	F_{hi}	
1	Concordance	0.903	1.033	0.964
1	Ratio estimator	0.927	0.994	0.372
2	Concordance	0.905	1.031	0.990
2	Ratio estimator	0.896	0.907	0.000

Notes.
[a] Setups 1 and 2 are the same except that instrument 3 (of 3) has 3% statistical errors for setup 1 and 0.3% statistical errors for setup 2. The prior for the effective area of instrument 3 is 10% higher than its true value in both setups. See Section 4.4.1 for details.
[b] Average 95% confidence intervals for source flux estimates; the true fluxes for all sources are set to 1.
[c] Fraction of flux estimates covering true fluxes at the 95% confidence level out of 200 simulations of 20 sources each.

Concordance intervals did not change appreciably, but the ratio estimators were biased low by ≈10%, reflecting the higher statistical weight given to an instrument with a biased estimate of its effective area. Indeed, in this latter case, there was only one source (of 20) in only one simulation (of 200) where the ratio estimator confidence interval included the true value. This pair of simulation setups illustrates the robustness of the Concordance method to erroneous effective area priors, showing that it is appropriate to use in calibration work where robustness and accuracy are highly valued.

4.4.2. Posterior Histograms

We have found that the posterior distributions of the effective area corrections are typically well described by Gaussians, as shown in Figure 8, parts (A)–(C). These three examples were randomly chosen among the dozens of such histograms generated in our analysis of the data from Sections 4.1–4.3.

Occasionally, however, there are histograms that are not obviously Gaussian, so we also show three "bad" examples. In one case, Figure 8, part (D), there is a distinct "notch" in a side of the distribution and in two cases (Figure 8, parts (E) and (F)), there is noticeable skew—tails to large fractional corrections. These three cases were quite rare but give warning that there may be inconsistencies in the underlying data. One known source of error that is not accounted for in our analysis is in the shape of the response function, $\Phi_k(E)$. For instruments like ACIS and the EPIC detectors, the low-energy response is somewhat uncertain and difficult to calibrate.

4.4.3. Sensitivity to Uncertainties in Priors

The specification of priors is typically under scrutiny for Bayesian analysis in practice. Typically researchers conduct sensitivity analysis to study the outcome sensitivity with respect to small perturbations of the priors. In our setting, the sensitivity of the results with respect to τ values is revealed by the comparison between heterogeneous τ values versus the two homogeneous τ value choices. However, for the correlation matrix in the prior distribution, we adopted Monte Carlo estimates, which is subject to random variations. Thus, we undertook a simple example of the test of sensitivity in this paper, but it is feasible for any user of the Concordance tools. Namely, for the Capella data, instead of adopting the full correlation matrix as given in Tables 8 and 9, we only keep the correlations between the two Fe bands. Again, we can test out different variations of the correlation matrix with the same procedure and similar analysis. Thorough sensitivity analysis requires extensive testing on a carefully designed set of variations of the prior distributions. See Figure 9 for the results of applying this variation to the Capella data. By comparing the results of Figure 9 with the original, Figure 6, we can reveal the sensitivity of the results as opposed to variations of correlations between Fe bands and others, and the correlations of others (Ne and O) within their own. We can see that in Figure 9, the O and
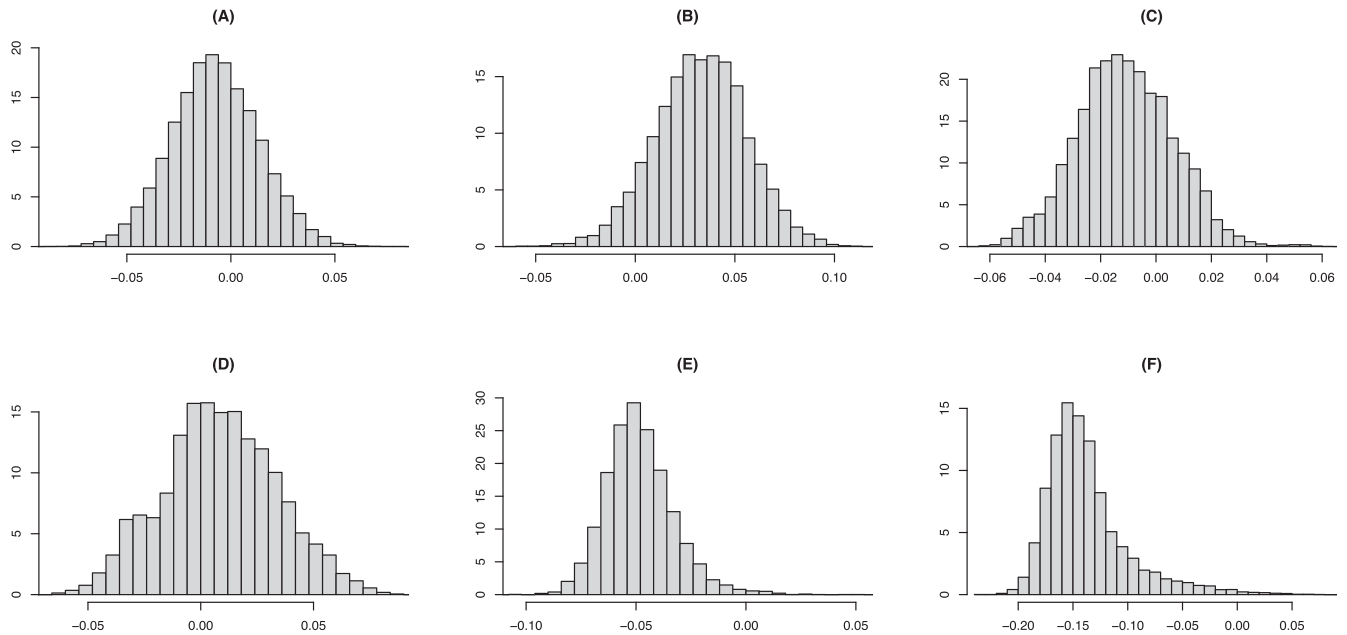
Figure 8. Posterior histograms for the fractional variation of the effective area. The data sets are (A) 2XMM data, hard band, XMM/pn, correlated τ values; (B) 2XMM data, hard band, XMM/MOS2, heterogeneous τ values; (C) XCAL data, medium band, XMM/pn, correlated τ values; (D) XCAL data, soft band, XMM/MOS1, $\tau = 0.05$ for all instruments; (E) 1E0102 data, O lines, Chandra/ACIS-S3, $\tau = 0.025$ for all instruments; (F) 1E0102 data, O lines, Swift XRT/PC, $\tau = 0.05$ for all instruments. Histograms (A)–(C) are typical, chosen randomly from several dozen; the distributions are well approximated as Gaussians. Histograms (D)–(F) are atypical, showing skew or other non-Gaussian shapes.

Ne show nicely aligned results under correlated effective areas versus uncorrelated effective areas. But this is not true for Figure 6, where O and Ne are still correlated with each other and with other bands, especially for HEG+. Furthermore, while the adjustments for the two Fe channels are similar across the two figures, the adjustments for Ne are very different. The resulting adjustments of effective areas not only deviate more significantly from zero but also have smaller error bars when correlations are taken into account. This makes intuitive sense because the benefit of accounting for correlations among effective areas is to obtain sharper or more informative estimates.

5. Conclusion and Directions for Development

These data sets provided an excellent foundation for the Concordance project, whose goal is to determine quantitative and objective evidence for making effective area adjustments in order to improve agreement between instrument measurements. The process applied here is available for use in studies such as we have undertaken. There are some avenues to explore for expanding this particular implementation of the Concordance analysis.

5.1. Correlations between Source Bandpass Fluxes

Several types of calibration sources have simple spectra, which is why they are often used in cross-calibration. Examples are isolated neutron stars with blackbody spectra, blazars with power-law spectra, and SNRs and clusters of galaxies with thermal spectra. To the extent that these spectra can be characterized by only a few parameters, such as a power-law slope, then the flux in one band is closely related to that in an adjacent band. Furthermore, many types of source have smoothly continuous spectra—their spectral fluxes are tightly correlated on small scales. Modeling many bandpasses of a blazar spectrum with a series of power laws with different slopes

would lead to unphysical discontinuities at bandpass boundaries. Thus, it would be advantageous to take advantage of this astrophysical knowledge and include spectral band correlations due to spectral continuity and simplicity. The XMM-Newton–Chandra blazar XCAL sample is an excellent data set to examine next, involving three Chandra configurations and all four XMM-Newton X-ray detectors and covering the energy range from 0.1 to 10 keV using simultaneous observations of active galaxies obtained over 20 yr of operation. Preliminary results have been reported at various IACHEC meetings.

5.2. Secular Variations of Instrument Responses

While many sources may well vary erratically, instrument behavior can often be subject to gradual degradation. With adequate modeling of many observations, one avenue to explore is how to link instrument effective areas over time within the Concordance framework.

5.3. Nonlocal Instrument Responses

There are definite difficulties that are encountered when the detector energy response $\Phi_k(E)$ has a non-Gaussian component, a broad asymmetry, or bimodality because systematic errors in the response function can appear in an apparently unrelated bandpass. Response function errors may be responsible for some of the artifacts in our posterior histograms (see Section 4.4.1). The response functions of solid state detectors have escape peaks that can generate events at a significantly different apparent energy than that of the incoming photon. Modeling the effects of systematic errors in response functions is possible in principle, especially with methods such as used to determine the effective correlation function (see Section 2.3.2). One approach for dealing with this issue would be to expand the Concordance mathematical model to include a term to account for variance of the T_{ij} values.
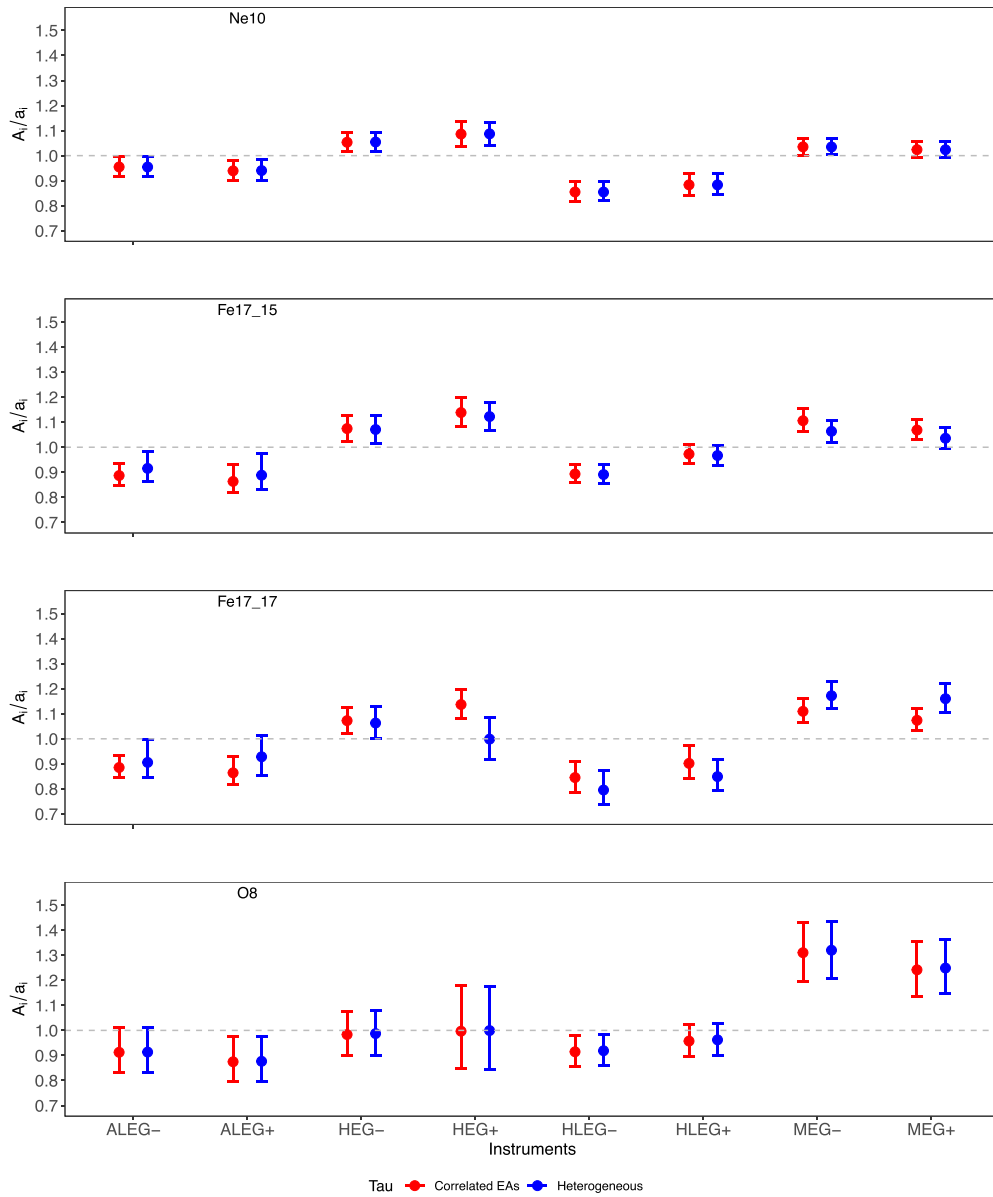
Figure 9. Same as Figure 6 except that only the correlations between the two Fe bands are kept while others are set to zero, for both ACIS and HRC instruments.

Appendix

A.1. Measured Fluxes

In Table A1 we list all the fluxes used in the calculations described above: for sources in the 2XMM catalog (see Section 3.2) in the soft, medium, and hard bands; and for active galaxies from the XCAL sample (Section 3.3) in the soft, medium, and hard bands. The table is published in its entirety in machine-readable format. The line fluxes measured during the various Capella grating observations with Chandra (Section 3.4) are provided in Table A2. Note that the data used for the analysis of SNR 1E0102.2-7219 (Section 3.1) are given in Chen et al. (2019).

Table A1
Concordance Fluxes

Target	Sample			pn		MOS1		MOS2	
	Sample	Band	Norm.[a]	f_{ij}	σ_{ij}	f_{ij}	σ_{ij}	f_{ij}	σ_{ij}
1127−145	2XMM	Soft	0.138	0.0075	0.0009	0.0074	0.0009	0.0077	0.0010
1E0919+515	2XMM	Soft	0.138	0.0068	0.0013	0.0081	0.0015	0.0083	0.0015
4C06.41	2XMM	Soft	0.138	0.0024	0.0001	0.0026	0.0001	0.0026	0.0001
APM 08279+5255	2XMM	Soft	0.138	0.0035	0.0006	0.0035	0.0006	0.0036	0.0006
1127−145	2XMM	Medium	0.000701	0.481	0.049	0.496	0.053	0.490	0.052
1E0919+515	2XMM	Medium	0.000701	0.053	0.053	0.069	0.066	0.068	0.065
4C06.41	2XMM	Medium	0.000701	0.131	0.015	0.142	0.017	0.143	0.018
APM 08279+5255	2XMM	Medium	0.000701	0.085	0.041	0.088	0.042	0.082	0.040
1127−145	2XMM	Hard	0.00223	0.173	0.007	0.186	0.008	0.179	0.008
1E0919+515	2XMM	Hard	0.00223	0.011	0.003	0.012	0.003	0.011	0.002
4C06.41	2XMM	Hard	0.00223	0.042	0.001	0.045	0.002	0.044	0.002
APM 08279+5255	2XMM	Hard	0.00223	0.016	0.001	0.017	0.002	0.016	0.001
1ES0414+009	XCAL	Soft	0.126	0.0992	0.0100	0.1058	0.0107	0.1092	0.0113
1ES1101-232	XCAL	Soft	0.126	0.2124	0.0058	0.2249	0.0063	0.2245	0.0064
1ES1553+113	XCAL	Soft	0.126	0.1097	0.0033	0.1053	0.0032	0.1059	0.0032
1H0414+009	XCAL	Soft	0.126	0.0459	0.0019	0.0477	0.0020	0.0491	0.0021
1ES0414+009	XCAL	Medium	0.0156	0.086	0.010	0.090	0.011	0.101	0.013
1ES1101-232	XCAL	Medium	0.0156	0.315	0.024	0.334	0.027	0.349	0.028
1ES1553+113	XCAL	Medium	0.0156	0.133	0.010	0.123	0.009	0.124	0.010
1H0414+009	XCAL	Medium	0.0156	0.044	0.006	0.045	0.006	0.046	0.007
1ES1101-232	XCAL	Hard	0.0154	0.1884	0.0062	0.1939	0.0070	0.2013	0.0077
1ES1553+113	XCAL	Hard	0.0154	0.0657	0.0025	0.0630	0.0022	0.0635	0.0022
1H0414+009	XCAL	Hard	0.0154	0.0152	0.0004	0.0159	0.0005	0.0155	0.0005
1H1219+301	XCAL	Hard	0.0154	0.1409	0.0145	0.1499	0.0162	0.1586	0.0176

Note.
[a] Fluxes are normalized by this value in units of photons cm^{-2} s^{-1}.

(This table is available in its entirety in machine-readable form.)

Table A2
Capella Grating Observations with Chandra

Epoch	ObsID	Exposure (ks)	Detector	Grating Arm	Line Fluxes (10^{-13} erg s^{-1} cm^{-2})			
					Ne X $\lambda 12$	Fe XVII $\lambda 15$	Fe XVII $\lambda 17$	O VIII $\lambda 19$
1999 Aug	1099	14.57	ACIS-S	HEG+1	21.1 ± 2.0	57.8 ± 4.7	...	...
				HEG-1	21.8 ± 1.7	63.3 ± 4.8	64.3 ± 6.4	...
				MEG+1	21.0 ± 1.0	47.5 ± 2.1	68.9 ± 3.3	29.3 ± 2.8
				MEG-1	20.0 ± 1.0	48.7 ± 1.5	69.4 ± 2.1	34.4 ± 1.7
1999 Aug	1235	14.57	ACIS-S	HEG+1	22.4 ± 2.1	59.1 ± 4.7	...	...
				HEG-1	20.5 ± 1.6	47.2 ± 4.2	66.3 ± 6.4	31.3 ± 14.4
				MEG+1	20.5 ± 1.0	50.7 ± 2.2	68.0 ± 3.3	30.7 ± 2.9
				MEG-1	20.6 ± 1.0	51.7 ± 1.6	69.9 ± 2.1	33.3 ± 1.7
1999 Aug	1100	14.57	ACIS-S	HEG+1	23.8 ± 2.1	54.1 ± 4.5	...	...
				HEG-1	22.3 ± 1.7	64.4 ± 4.9	61.6 ± 6.2	46.0 ± 17.4
				MEG+1	22.1 ± 1.1	52.7 ± 2.3	74.9 ± 3.4	35.1 ± 3.1
				MEG-1	21.7 ± 1.0	49.5 ± 1.6	73.3 ± 2.2	32.7 ± 1.7
1999 Aug	1236	14.57	ACIS-S	HEG+1	26.4 ± 2.3	57.1 ± 4.6	...	...
				HEG-1	21.4 ± 1.6	54.4 ± 4.5	63.9 ± 6.3	31.5 ± 14.1
				MEG+1	20.6 ± 1.0	49.2 ± 2.2	69.3 ± 3.3	31.2 ± 2.9
				MEG-1	21.8 ± 1.0	47.8 ± 1.5	71.9 ± 2.1	33.2 ± 1.7
1999 Aug	1101	14.57	ACIS-S	HEG+1	26.1 ± 2.2	50.4 ± 4.3	...	...
				HEG-1	23.9 ± 1.7	54.8 ± 4.5	71.5 ± 6.7	22.4 ± 11.2
				MEG+1	20.8 ± 1.0	48.0 ± 2.1	60.9 ± 3.1	26.5 ± 2.7
				MEG-1	21.2 ± 1.0	48.9 ± 1.5	71.0 ± 2.1	35.2 ± 1.7
1999 Aug	1237	14.57	ACIS-S	HEG+1	24.0 ± 2.2	57.2 ± 4.6	...	...
				HEG-1	22.5 ± 1.7	44.2 ± 4.0	76.0 ± 6.9	38.0 ± 14.4
				MEG+1	21.3 ± 1.0	46.7 ± 2.1	69.4 ± 3.3	26.0 ± 2.6
				MEG-1	21.9 ± 1.0	48.7 ± 1.5	70.6 ± 2.1	29.6 ± 1.6
1999 Sep	62435	32.38	HRC-S	LEG+1	19.8 ± 0.8	45.0 ± 1.1	56.9 ± 1.3	26.1 ± 0.8
				LEG-1	18.9 ± 0.8	46.1 ± 1.1	59.8 ± 1.3	25.4 ± 0.8
1999 Sep	1167	15.14	HRC-S	LEG+1	19.7 ± 1.2	46.5 ± 1.7	58.4 ± 1.9	25.5 ± 1.2
				LEG-1	16.5 ± 1.1	44.1 ± 1.6	56.2 ± 1.9	25.7 ± 1.2
1999 Sep	1244	12.12	HRC-S	LEG+1	19.7 ± 1.4	42.9 ± 1.8	52.7 ± 2.1	23.9 ± 1.3
				LEG-1	18.2 ± 1.3	49.7 ± 1.9	58.8 ± 2.1	27.4 ± 1.3
1999 Sep	62410	11.23	HRC-S	LEG+1	22.5 ± 1.5	44.0 ± 1.9	51.1 ± 2.1	25.9 ± 1.4
				LEG-1	19.5 ± 1.4	43.9 ± 1.9	58.9 ± 2.2	27.2 ± 1.4
1999 Sep	1246	14.60	HRC-S	LEG+1	17.7 ± 1.2	45.7 ± 1.7	55.9 ± 1.9	26.7 ± 1.2
				LEG-1	18.6 ± 1.2	45.3 ± 1.7	56.8 ± 1.9	25.3 ± 1.2
1999 Sep	62422	11.29	HRC-S	LEG+1	19.6 ± 1.4	47.0 ± 2.0	55.8 ± 2.2	25.8 ± 1.4
				LEG-1	18.2 ± 1.4	47.6 ± 1.9	65.2 ± 2.3	26.2 ± 1.4
1999 Sep	62423	14.56	HRC-S	LEG+1	18.8 ± 1.2	46.1 ± 1.7	60.7 ± 2.0	26.0 ± 1.2
				LEG-1	19.5 ± 1.2	47.2 ± 1.7	57.9 ± 1.9	25.6 ± 1.2
1999 Sep	1103	40.52	ACIS-S	HEG+1	25.8 ± 1.4	60.4 ± 2.9	...	...
				HEG-1	26.9 ± 1.1	60.9 ± 2.8	81.6 ± 4.4	19.7 ± 8.4

(This table is available in its entirety in machine-readable form.)

ORCID iDs

Herman L. Marshall ![ORCID] https://orcid.org/0000-0002-6492-1293
Yang Chen ![ORCID] https://orcid.org/0000-0002-9516-8134
Jeremy J. Drake ![ORCID] https://orcid.org/0000-0002-0210-2276
Vinay L. Kashyap ![ORCID] https://orcid.org/0000-0002-3869-7996
Xiao-Li Meng ![ORCID] https://orcid.org/0000-0003-3687-0385
Paul P. Plucinsky ![ORCID] https://orcid.org/0000-0003-1415-5823
David A. van Dyk ![ORCID] https://orcid.org/0000-0002-0816-331X

References

Beuermann, K., Burwitz, V., & Rauch, T. 2006, A&A, 458, 541
Chen, Y., Meng, X.-L., Wang, X., et al. 2019, J. Am. Stat. Assoc., 114, 1018
Drake, J. J., Ratzlaff, P., Kashyap, V., et al. 2006, Proc. SPIE, 6270, 62701I
Efron, B., & Morris, C. 1972, Biometrika, 59, 335
Efron, B., & Morris, C. 1973, J. Am. Stat. Assoc., 68, 117
Fruscione, A., McDowell, J. C., Allen, G. E., et al. 2006, Proc. SPIE, 6270, 62701V
Gabriel, C., Denby, M., Fyfe, D. J., et al. 2004, in ASP Conf. Ser. 314, Astronomical Data Analysis Software and Systems (ADASS) XIII, ed. F. Ochsenbein, M. G. Allen, & D. Egret (San Francisco, CA: ASP), 759
Guainazzi, M., David, L., Grant, C. E., et al. 2015, JATIS, 1, 047001

Ishida, M., Tsujimoto, M., Kohmura, T., et al. 2011, PASJ, 63, S657
Jansen, F., Lumb, D., Altieri, B., et al. 2001, A&A, 365, L1
Jethwa, P., Saxton, R., Guainazzi, M., Rodriguez-Pascual, P., & Stuhlinger, M. 2015, A&A, 581, A104
Kashyap, V., & Drake, J. J. 2000, BASI, 28, 475
Kashyap, V. L., Lee, H., Siemiginowska, A., et al. 2008, Proc. SPIE, 7016, 70160P
Kettula, K., Nevalainen, J., & Miller, E. D. 2013, A&A, 552, A47
Kirsch, M. G., Briel, U. G., Burrows, D., et al. 2005, Proc. SPIE, 5898, 22
Lee, H., Kashyap, V. L., van Dyk, D. A., et al. 2011, ApJ, 731, 126
Madsen, K. K., Beardmore, A. P., Forster, K., et al. 2017a, AJ, 153, 2
Madsen, K. K., Forster, K., Grefenstette, B. W., Harrison, F. A., & Stern, D. 2017b, ApJ, 841, 56
Madsen, K. K., Natalucci, L., Belanger, G., et al. 2019, arXiv:1901.00934
Madsen, K. K., Terada, Y., Burwitz, V., et al. 2020, arXiv:2001.11117
Nevalainen, J., David, L., & Guainazzi, M. 2010, A&A, 523, A22
Plucinsky, P. P., Beardmore, A. P., Foster, A., et al. 2017, A&A, 597, A35
Read, A. M., Guainazzi, M., & Sembay, S. 2014, A&A, 564, A75
Schellenberger, G., Reiprich, T. H., Lovisari, L., Nevalainen, J., & David, L. 2015, A&A, 575, A30
Strüder, L., Briel, U., Dennerl, K., et al. 2001, A&A, 365, L18
Toor, A., & Seward, F. D. 1974, AJ, 79, 995
Tsujimoto, M., Guainazzi, M., Plucinsky, P. P., et al. 2011, A&A, 525, A25
Turner, M. J. L., Abbey, A., Arnaud, M., et al. 2001, A&A, 365, L27
Watson, M. G., Schröder, A. C., Fyfe, D., et al. 2009, A&A, 493, 339
Xu, J., van Dyk, D. A., Kashyap, V. L., et al. 2014, ApJ, 794, 97