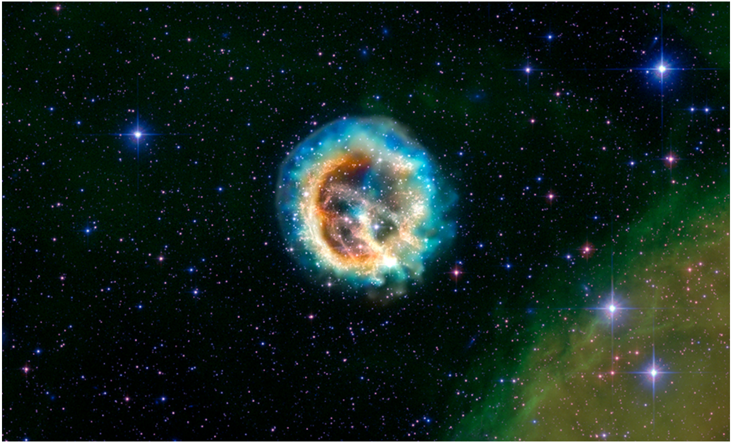
Calibration Concordance for Astronomical Instruments

Yang Chen

Joint work with X.-L. Meng (Harvard University), X. Wang (Two Sigma Inc.), D. van Dyk (Imperial College London), V. Kashyap (Center for Astronomy), H. Marshall (MIT)

September 8, 2020

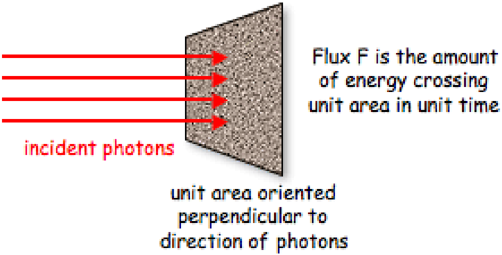
Calibration Concordance Problem (Example: E0102)

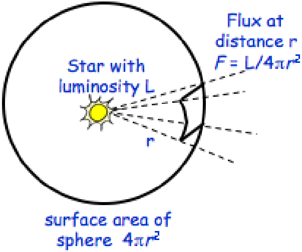


- Supernova remnant E0102
- Four sources correspond to four spectral lines in E0102

Measurements

Flux is the total amount of energy that crosses a unit area per unit time.

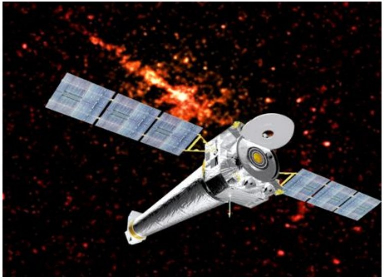




The flux of an astronomical source (F) depends on the luminosity of the object (L) and its distance from the Earth (r), $F = L/4\pi r^2$.

Observatory and Instruments

Current X-ray Observatory



USA: Chandra X-ray Observatory

High angular resolution (~0.5")

Europe: XMM-Newton

High throughput (large effective area)

And
- Rossi X-ray Timing Explorer
- Swift
- INTEGRAL etc.

9
/72

Observatory and Instruments

CHANDRA X-RAY OBSERVATORY

CXC Home | Proposer | Archive | Data Analysis | Instruments & Calibration | For the Public

Chandra Instruments and Calibration

The Chandra X-ray Observatory (CXO) is designed for high resolution (≤ 1/2 arcsec) X-ray imaging and spectroscopy. The High Resolution Mirror Assembly (HRMA) focuses X-rays onto one of two instruments, ACIS or HRC. Only one detector (HRC or ACIS) is in the focal plane at any given time. Two grating spectrometers (LETG or HETG) can be placed in the optical path behind the HRMA. The dispersed spectrum is read out by either ACIS or HRC. A high level overview of the instruments on-board the Chandra X-ray Observatory can be found on the About Chandra pages and a more detailed description can be found in the Proposers' Observatory Guide.

Current calibration data products for use in CIAO and other analysis systems can be found in the CALDB pages. A complete listing of all calibration products in the CALDB and a brief description of these products can be found in the Calibration Data Products.

- Calibration Status Summary
- ACIS
- HRC
- HETG
- LETG
- HRMA
- Calibration Database (CALDB)
- Cross-Calibration with other X-Ray Telescopes
- Aspect Information
- Calibration Workshops and Reviews
- SPIE Proceedings
- Science and Calibration Requirements

Advanced CCD Imaging Spectrometer (ACIS)

The ACIS has two arrays of CCDs, one (ACIS-I) optimized for imaging wide fields (16x16 arc minutes) the other (ACIS-S) optimized as a readout for the HETG transmission grating. One chip of the ACIS-S (S3) can also be used for on-axis (8x8 arc minutes) imaging and offers the best energy resolution of the ACIS system.

High Resolution Camera (HRC)

The HRC comprises two micro-channel plate imaging detectors, and offers the highest spatial (<0.5 arc second) and temporal (16 msec) resolutions. The HRC-I has the largest field-of-view (31x31 arc minutes) available on Chandra. The HRC-S is most commonly used to read out the dispersed spectrum from the LETG.

High Energy Transmission Grating (HETG)

The HETG is optimized for high-resolution spectroscopy of bright sources over the energy band 0.4-10 keV. It is most commonly used with ACIS-S. The resolving power (E/ΔE) varies from ~800 at 1.5 keV to ~200 at 6 keV.

Low Energy Transmission Grating (LETG)

The LETG provides the highest spectral resolving power (E/ΔE > 1000) on Chandra at low energies (0.07 - 0.2 keV). The LETG/HRC-S combination is used extensively for high resolution spectroscopy of bright, soft sources such as stellar coronae, white dwarf atmospheres and cataclysmic variables.

Observatory and Instruments

CHANDRA X-RAY OBSERVATORY

CXC Home | Proposer | Archive | Data Analysis | Instruments & Calibration | For the Public

Chandra Instruments and Calibration

The Chandra X-ray Observatory (CXO) is designed for high resolution (≤ 1/2 arcsec) X-ray imaging and spectroscopy. The High Resolution Mirror Assembly (HRMA) focuses X-rays onto one of two instruments, ACIS or HRC. Only one detector (HRC or ACIS) is in the focal plane at any given time. Two grating spectrometers (LETG or HETG) can be placed in the optical path behind the HRMA. The dispersed spectrum is read out by either ACIS or HRC. A high level overview of the instruments on-board the Chandra X-ray Observatory can be found on the About Chandra pages and a more detailed description can be found in the Proposers' Observatory Guide.

Current calibration data products for use in CIAO and other analysis systems can be found in the CALDB pages. A complete listing of all calibration products in the CALDB and a brief description of these products can be found in the Calibration Data Products.

- Calibration Status Summary
- ACIS
- HRC
- HETG
- LETG
- HRMA
- Calibration Database (CALDB)
- Cross-Calibration with other X-Ray Telescopes
- Aspect Information
- Calibration Workshops and Reviews
- SPIE Proceedings
- Science and Calibration Requirements

Advanced CCD Imaging Spectrometer (ACIS)

The ACIS has two arrays of CCDs, one (ACIS-I) optimized for imaging wide fields (16x16 arc minutes) the other (ACIS-S) optimized as a readout for the HETG transmission grating. One chip of the ACIS-S (S3) can also be used for on-axis (8x8 arc minutes) imaging and offers the best energy resolution of the ACIS system.

High Resolution Camera (HRC)

The HRC comprises two micro-channel plate imaging detectors, and offers the highest spatial (<0.5 arc second) and temporal (16 msec) resolutions. The HRC-I has the largest field-of-view (31x31 arc minutes) available on Chandra. The HRC-S is most commonly used to read out the dispersed spectrum from the LETG.

High Energy Transmission Grating (HETG)

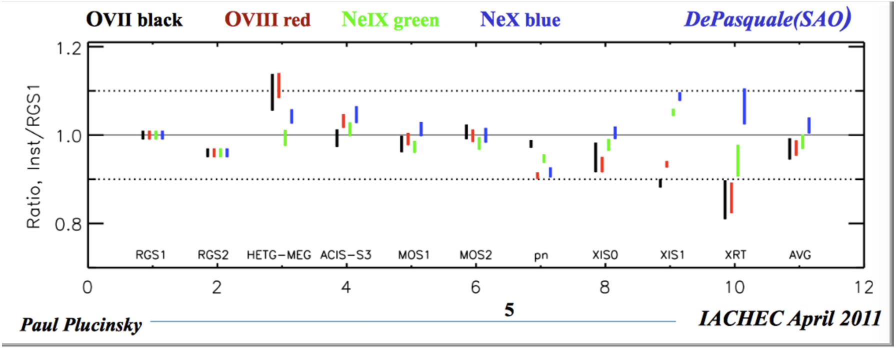
The HETG is optimized for high-resolution spectroscopy of bright sources over the energy band 0.4-10 keV. It is most commonly used with ACIS-S. The resolving power (E/ΔE) varies from ~800 at 1.5 keV to ~200 at 6 keV.

Low Energy Transmission Grating (LETG)

The LETG provides the highest spectral resolving power (E/ΔE > 1000) on Chandra at low energies (0.07 - 0.2 keV). The LETG/HRC-S combination is used extensively for high resolution spectroscopy of bright, soft sources such as stellar coronae, white dwarf atmospheres and cataclysmic variables.

Each of these instruments has a different photon collection efficiency – Effective Area. Reflectivity and vignetting, among other effects, cause the geometric area of a telescope to be reduced to a smaller "effective area".

Calibration Concordance Problem (Example: E0102)



- Four spectral lines observed with 11 X-ray detectors
- Main challenge – the data/instruments do not agree

Outline

Notation

- N Instruments with true effective area A_i, $1 \leq i \leq N$.
  - For each instrument i, we know estimated a_i ($\approx A_i$) but not A_i.

Notation

- N Instruments with true effective area A_i, $1 \leq i \leq N$.
  - For each instrument i, we know estimated a_i ($\approx A_i$) but not A_i.
- M Sources with fluxes F_j, $1 \leq j \leq M$.
  - For each source j, F_j is unknown.

Notation

- N Instruments with true effective area A_i, $1 \leq i \leq N$.
  - For each instrument i, we know estimated a_i ($\approx A_i$) but not A_i.
- M Sources with fluxes F_j, $1 \leq j \leq M$.
  - For each source j, F_j is unknown.
- Photon counts c_{ij}: from measuring flux F_j with instrument i.

Notation

- N Instruments with true effective area A_i, $1 \leq i \leq N$.
  - For each instrument i, we know estimated a_i ($\approx A_i$) but not A_i.
- M Sources with fluxes F_j, $1 \leq j \leq M$.
  - For each source j, F_j is unknown.
- Photon counts c_{ij}: from measuring flux F_j with instrument i.
- Lower cases: data / estimators.
- Upper cases: parameter / estimand.

Calibration Concordance Problem

1. Astronomers' Dilemma:

$$\frac{c_{ij}}{a_i} \neq \frac{c_{i'j}}{a_{i'}} \text{ for } i \neq i'.$$

Different instruments give different estimated flux of the same object!

Calibration Concordance Problem

1. Astronomers' Dilemma:

$$\frac{c_{ij}}{a_i} \neq \frac{c_{i'j}}{a_{i'}} \text{ for } i \neq i'.$$

Different instruments give different estimated flux of the same object!

2. Scientific Question:
   - Are there systematic errors in 'known' effective areas?
   - Can we derive properly adjusted effective areas?
   - Can we unify estimates of the same flux with different instruments?

Scientific and Statistical Models

Scientific Model

Multiplicative in original scale and additive on the log scale.

$$\text{Counts} = \text{Exposure} \times \text{Effective Area} \times \text{Flux},$$
$$C_{ij} = T_{ij} A_i F_j, \quad \Leftrightarrow \quad \log C_{ij} = B_i + G_j,$$

where log area $= B_i = \log A_i$, log flux $= G_j = \log F_j$; let $T_{ij} = 1$.

Scientific and Statistical Models

Scientific Model

Multiplicative in original scale and additive on the log scale.

$$\text{Counts} = \text{Exposure} \times \text{Effective Area} \times \text{Flux},$$
$$C_{ij} = T_{ij} A_i F_j, \quad \Leftrightarrow \quad \log C_{ij} = B_i + G_j,$$

where log area $= B_i = \log A_i$, log flux $= G_j = \log F_j$; let $T_{ij} = 1$.

Statistical Model

$$\text{log counts } y_{ij} = \log c_{ij} - \alpha_{ij} = B_i + G_j + e_{ij}, \quad e_{ij} \overset{indep}{\sim} \mathcal{N}(0, \sigma_{ij}^2);$$

where $\alpha_{ij} = -0.5\sigma_{ij}^2$ to ensure $E(c_{ij}) = C_{ij} = A_i F_j$.

- **Known Variances**: σ_{ij} known.
- **Unknown Variances**: $\sigma_{ij} = \sigma_i$ unknown.

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

$$\text{log counts } | \textit{area} \,\&\, \textit{flux} \,\&\, \textit{variance} \overset{\text{indep}}{\sim} \text{Gaussian distribution},$$

$$y_{ij} \mid B_i,\ G_j,\ \sigma_i^2 \overset{\text{indep}}{\sim} \mathcal{N}\left(B_i + G_j,\ \sigma_i^2\right),$$

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

$$\begin{aligned}
\text{log counts } | \textit{area \& flux \& variance} &\overset{\text{indep}}{\sim} \text{Gaussian distribution}, \\
y_{ij} \mid B_i,\ G_j,\ \sigma_i^2 &\overset{\text{indep}}{\sim} \mathcal{N}\left(B_i + G_j,\ \sigma_i^2\right), \\
B_i &\overset{\text{indep}}{\sim} N(b_i,\ \tau_i^2), \\
G_j &\overset{\text{indep}}{\sim} \text{flat prior},
\end{aligned}$$

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

$$
\begin{aligned}
\text{log counts } |\textit{area \& flux \& variance} &\overset{\text{indep}}{\sim} \text{Gaussian distribution},\\
y_{ij} \mid B_i,\ G_j,\ \sigma_i^2 &\overset{\text{indep}}{\sim} \mathcal{N}\left(B_i + G_j,\ \sigma_i^2\right),\\
B_i &\overset{\text{indep}}{\sim} N(b_i,\ \tau_i^2),\\
G_j &\overset{\text{indep}}{\sim} \text{flat prior},\\
\text{If variance unknown: } \sigma_i^2 &\overset{\text{indep}}{\sim} \text{Inv-Gamma}(df_g,\ \beta_g).
\end{aligned}
$$

Setting the prior parameters.

1. $b_i = \log a_i$, τ_i are given by astronomers.

Bayesian Hierarchical Model

Log-Normal Hierarchical Model.

$$\begin{aligned}
\text{log counts } | \textit{area} \,\&\, \textit{flux} \,\&\, \textit{variance} &\overset{\text{indep}}{\sim} \text{Gaussian distribution}, \\
y_{ij} \mid B_i,\ G_j,\ \sigma_i^2 &\overset{\text{indep}}{\sim} \mathcal{N}\left(B_i + G_j,\ \sigma_i^2\right), \\
B_i &\overset{\text{indep}}{\sim} N(b_i,\ \tau_i^2), \\
G_j &\overset{\text{indep}}{\sim} \text{flat prior}, \\
\text{If variance unknown: } \sigma_i^2 &\overset{\text{indep}}{\sim} \text{Inv-Gamma}(df_g,\ \beta_g).
\end{aligned}$$

Setting the prior parameters.

1. $b_i = \log a_i$, τ_i are given by astronomers.
2. df_g, β_g are given based on the variability in data.

Posterior Propriety and Identifiability

Posterior Propriety. The posterior is proper if each source is measured by at least one instrument, i.e., $|I_j| \geq 1$ for all $1 \leq j \leq M$.

Posterior Propriety and Identifiability

Posterior Propriety. The posterior is proper if each source is measured by at least one instrument, i.e., $|I_j| \geq 1$ for all $1 \leq j \leq M$.

Identifiability

- $\tau_i^2 = \infty$: same posteriors with $\{B_i, G_j\}$ and $\{B_i + \delta, G_j - \delta\}$;

Posterior Propriety and Identifiability

Posterior Propriety. The posterior is proper if each source is measured by at least one instrument, i.e., $|I_j| \geq 1$ for all $1 \leq j \leq M$.

Identifiability

- $\tau_i^2 = \infty$: same posteriors with $\{B_i, G_j\}$ and $\{B_i + \delta, G_j - \delta\}$;
- the *condition number* of $\boldsymbol{\Omega}(\boldsymbol{\sigma}^2)$ (conditional variance of $\boldsymbol{B}, \boldsymbol{G}$) is

$$\frac{\lambda_{\max}(\boldsymbol{\Omega}(\boldsymbol{\sigma}^2))}{\lambda_{\min}(\boldsymbol{\Omega}(\boldsymbol{\sigma}^2))} \geq \frac{u^\top \boldsymbol{\Omega}(\boldsymbol{\sigma}^2)\ u}{v^\top \boldsymbol{\Omega}(\boldsymbol{\sigma}^2)\ v} = 1 + \frac{4\sum_{i=1}^N |J_i| \sigma_i^{-2}}{\sum_{i=1}^N \tau_i^{-2}}, \tag{1}$$

where $u = (\mathbf{1}_N, \mathbf{1}_M)^\top$ and $v = (\mathbf{1}_N, -\mathbf{1}_M)^\top$.

Posterior Propriety and Identifiability

Posterior Propriety. The posterior is proper if each source is measured by at least one instrument, i.e., $|I_j| \geq 1$ for all $1 \leq j \leq M$.

Identifiability

- $\tau_i^2 = \infty$: same posteriors with $\{B_i, G_j\}$ and $\{B_i + \delta, G_j - \delta\}$;
- the *condition number* of $\boldsymbol{\Omega}(\boldsymbol{\sigma}^2)$ (conditional variance of $\boldsymbol{B}, \boldsymbol{G}$) is

$$\frac{\lambda_{\max}(\boldsymbol{\Omega}(\boldsymbol{\sigma}^2))}{\lambda_{\min}(\boldsymbol{\Omega}(\boldsymbol{\sigma}^2))} \geq \frac{u^\top \boldsymbol{\Omega}(\boldsymbol{\sigma}^2)\ u}{v^\top \boldsymbol{\Omega}(\boldsymbol{\sigma}^2)\ v} = 1 + \frac{4\sum_{i=1}^N |J_i| \sigma_i^{-2}}{\sum_{i=1}^N \tau_i^{-2}}, \tag{1}$$

  where $u = (\mathbf{1}_N, \mathbf{1}_M)^\top$ and $v = (\mathbf{1}_N, -\mathbf{1}_M)^\top$.
- $\{\tau_i^2\} >> \{\sigma_i^2\}$: elongated posterior contours.

Posterior Propriety and Identifiability

Posterior Propriety. The posterior is proper if each source is measured by at least one instrument, i.e., $|I_j| \geq 1$ for all $1 \leq j \leq M$.

Identifiability

- $\tau_i^2 = \infty$: same posteriors with $\{B_i, G_j\}$ and $\{B_i + \delta, G_j - \delta\}$;
- the *condition number* of $\boldsymbol{\Omega}(\boldsymbol{\sigma}^2)$ (conditional variance of $\boldsymbol{B}, \boldsymbol{G}$) is

$$\frac{\lambda_{\max}(\boldsymbol{\Omega}(\boldsymbol{\sigma}^2))}{\lambda_{\min}(\boldsymbol{\Omega}(\boldsymbol{\sigma}^2))} \geq \frac{u^\top \boldsymbol{\Omega}(\boldsymbol{\sigma}^2)\ u}{v^\top \boldsymbol{\Omega}(\boldsymbol{\sigma}^2)\ v} = 1 + \frac{4 \sum_{i=1}^N |J_i| \sigma_i^{-2}}{\sum_{i=1}^N \tau_i^{-2}}, \tag{1}$$

  where $u = (\mathbf{1}_N, \mathbf{1}_M)^\top$ and $v = (\mathbf{1}_N, -\mathbf{1}_M)^\top$.
- $\{\tau_i^2\} >> \{\sigma_i^2\}$: elongated posterior contours.

Alternative: setting $B_1 = 0$ or $\tau_1 = 0$.

Bayesian Computation (Unknown Variances)

Markov Chain Monte Carlo (MCMC) algorithms.

Bayesian Computation (Unknown Variances)

Markov Chain Monte Carlo (MCMC) algorithms.

- Gibbs Sampling: update parameters one-at-a-time.

Bayesian Computation (Unknown Variances)

Markov Chain Monte Carlo (MCMC) algorithms.

- Gibbs Sampling: update parameters one-at-a-time.
- Block Gibbs Sampling: update vectors of parameters.

Bayesian Computation (Unknown Variances)

Markov Chain Monte Carlo (MCMC) algorithms.

- Gibbs Sampling: update parameters one-at-a-time.
- Block Gibbs Sampling: update vectors of parameters.
    - The joint distribution of the B_i and G_j is Gaussian.

Bayesian Computation (Unknown Variances)

Markov Chain Monte Carlo (MCMC) algorithms.

- Gibbs Sampling: update parameters one-at-a-time.
- Block Gibbs Sampling: update vectors of parameters.
  - The joint distribution of the B_i and G_j is Gaussian.
- Hamiltonian Monte Carlo (HMC) – Stan package.

Bayesian Computation (Unknown Variances)

Markov Chain Monte Carlo (MCMC) algorithms.

- Gibbs Sampling: update parameters one-at-a-time.
- Block Gibbs Sampling: update vectors of parameters.
  - The joint distribution of the B_i and G_j is Gaussian.
- Hamiltonian Monte Carlo (HMC) – Stan package.
  - Highly correlated parameters, high-dim parameter space.

Shrinkage Estimators: Known Fluxes and Errors

Hierarchical model $\Rightarrow$ Shrinkage estimators (weighted averages of evidence from 'Prior' and evidence from 'Data').

Shrinkage Estimators: Known Fluxes and Errors

Hierarchical model $\Rightarrow$ Shrinkage estimators (weighted averages of evidence from 'Prior' and evidence from 'Data').

(1) When fluxes and variances are known,

Original Scale

$$\hat{A}_i = a_i^{W_i}\left[(\tilde{c}_{i\cdot}\tilde{f}^{-1})e^{\sigma_i^2/2}\right]^{1-W_i},$$

where

$$\tilde{c}_{i\cdot} = \prod_j c_{ij}^{1/M},\ \ \tilde{f} = \prod_j f_j^{1/M}$$

are geometric means.

Log-Scale

$$\hat{B}_i = W_i b_i + (1 - W_i)(\bar{y}_{i\cdot} - \bar{G}),$$

where

$$\bar{G} = \frac{\sum_j g_j}{M}, \bar{y}_{i\cdot} = \frac{\sum_j y_{ij}}{M}$$

are arithmatic means.

The 'weights', $W_i = \frac{\tau_i^{-2}}{\tau_i^{-2}+M\sigma_i^{-2}}$, represents the direct information in b_i relative to indirect information in fluxes.

Shrinkage Estimators: Known Errors

(2) When fluxes are unknown and variances are known,

$$\hat{B}_i = W_i b_i + (1 - W_i)(\bar{y}_{i\cdot} - \bar{G}_i), \quad \hat{G}_j = \bar{y}_{\cdot j} - \bar{B},$$

where $\bar{G}_i = \frac{\sum_j \hat{G}_j}{M}, \quad \bar{B} = \frac{\sum_i \hat{B}_i \sigma_i^{-2}}{\sum_i \sigma_i^{-2}}, \; \bar{y}_{i\cdot} = \frac{\sum_j y_{ij}}{M}, \quad \bar{y}_{\cdot j} = \frac{\sum_i y_{ij}\sigma_i^{-2}}{\sum_i \sigma_i^{-2}}.$

Shrinkage Estimators: Known Errors

(2) When fluxes are unknown and variances are known,

$$\hat{B}_i = W_i b_i + (1 - W_i)(\bar{y}_{i\cdot} - \bar{G}_i), \quad \hat{G}_j = \bar{y}_{\cdot j} - \bar{B},$$

where $\bar{G}_i = \frac{\sum_j \hat{G}_j}{M}, \quad \bar{B} = \frac{\sum_i \hat{B}_i \sigma_i^{-2}}{\sum_i \sigma_i^{-2}}, \; \bar{y}_{i\cdot} = \frac{\sum_j y_{ij}}{M}, \quad \bar{y}_{\cdot j} = \frac{\sum_i y_{ij}\sigma_i^{-2}}{\sum_i \sigma_i^{-2}}.$

(3) When variances are unknown, shrinkage estimator of variance,

$$\hat{\sigma}_i^2 = \frac{2}{1 + \sqrt{1 + S_{y,i}^2}}\, S_{y,i}^2, \quad S_{y,i}^2 = \frac{1}{|J_i| + \alpha}\left[\sum_{j \in J_i} (y_{ij} - \hat{B}_i - \hat{G}_j)^2 + \beta\right]$$

Benefits of Fitting σ_i^2

- Tolerance to model/error model misspecification.

Benefits of Fitting σ_i^2

- Tolerance to model/error model misspecification.
- Pitfalls of assuming 'known' variances:
  - Overly optimistic 'known variances'

    $\Rightarrow$ overly narrow confidence intervals

    $\Rightarrow$ possible false discoveries

Benefits of Fitting σ_i^2

- Tolerance to model/error model misspecification.
- Pitfalls of assuming 'known' variances:
  - Overly optimistic 'known variances'

    $\Rightarrow$ overly narrow confidence intervals

    $\Rightarrow$ possible false discoveries
  - 'known variances' $\geq$ true variability

    $\Rightarrow$ noninformative results

Extentions: Log-t Model

Question: Outliers? Less restrictions on the variances?

Extentions: Log-t Model

Question: Outliers? Less restrictions on the variances?

$$y_{ij} \mid B_i,\ G_j,\ \xi_{ij} = -\frac{\sigma^2}{2\xi_{ij}} + B_i + G_j + \frac{Z_{ij}}{\sqrt{\xi_{ij}}},$$

$$Z_{ij} \overset{\text{indep}}{\sim} N(0, \sigma^2),$$

$$B_i \overset{\text{indep}}{\sim} N(b_i, \tau_i^2).$$

Extentions: Log-t Model

Question: Outliers? Less restrictions on the variances?

$$
\begin{aligned}
y_{ij} \mid B_i,\ G_j,\ \xi_{ij} \quad &= \quad -\frac{\sigma^2}{2\xi_{ij}} + B_i + G_j + \frac{Z_{ij}}{\sqrt{\xi_{ij}}}, \\
Z_{ij} \quad &\overset{\text{indep}}{\sim} \quad N(0, \sigma^2), \\
B_i \quad &\overset{\text{indep}}{\sim} \quad N(b_i, \tau_i^2).
\end{aligned}
$$

If $\xi_{ij} \overset{\text{indep}}{\sim} \chi^2_\nu$, i.e. independent chi-squared distributions, the error term

$Z_{ij}/\sqrt{\xi_{ij}}$ follows independent student-t distributions, i.e. $\frac{Z_{ij}}{\sqrt{\xi_{ij}}} \overset{\text{indep}}{\sim} \frac{\sigma}{\sqrt{\nu}} \mathrm{t}_\nu$.

A Numerical Example with Outliers

Simulation: $N = 10, M = 40$, $G_1 = -1$ and $G_j = 3, j > 1$.
Asymptotic variance of log-counts: $e^{-B_i - G_j} \Rightarrow$ outliers.

A Numerical Example with Outliers

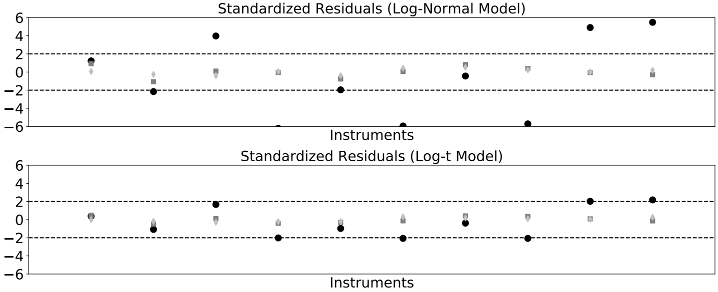
Simulation: $N = 10, M = 40,\ G_1 = -1$ and $G_j = 3, j > 1.$
Asymptotic variance of log-counts: $e^{-B_i - G_j} \Rightarrow$ outliers.

$$\hat{\mathcal{R}}_{ij} = \frac{y_{ij} - \hat{B}_i - \hat{G}_j + 0.5 \times \hat{\sigma}_i^2}{\hat{\sigma}_i}, \hat{\mathcal{R}}_{ij} = \frac{y_{ij} - \hat{B}_i - \hat{G}_j + 0.5 \times \kappa^2/\hat{\xi}_{ij}}{\kappa/\hat{\xi}_{ij}^{1/2}}$$

A Numerical Example with Outliers

Simulation: $N = 10, M = 40,\ G_1 = -1$ and $G_j = 3, j > 1$.
Asymptotic variance of log-counts: $e^{-B_i - G_j} \Rightarrow$ outliers.

$$\hat{\mathcal{R}}_{ij} = \frac{y_{ij} - \hat{B}_i - \hat{G}_j + 0.5 \times \hat{\sigma}_i^2}{\hat{\sigma}_i}, \hat{\mathcal{R}}_{ij} = \frac{y_{ij} - \hat{B}_i - \hat{G}_j + 0.5 \times \kappa^2/\hat{\xi}_{ij}}{\kappa/\hat{\xi}_{ij}^{1/2}}$$



Coverage Properties With Outliers, Misspecification

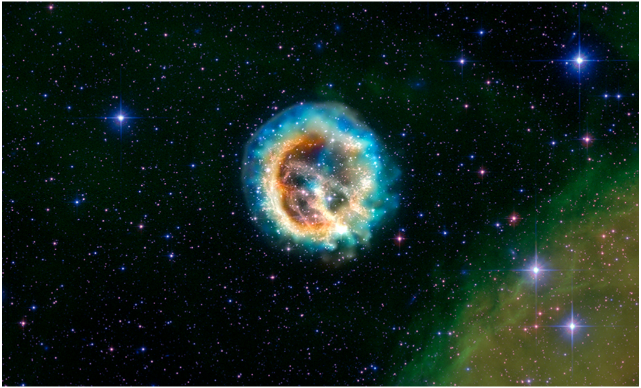
Poisson Model	Para.	Coverage Probability		Length of Interval	
		log-Normal	log-t	log-Normal	log-t
$N = 10$	$\boldsymbol{B}$	[0.941, 0.959]	[0.971, 0.975]	0.067$\pm$0.005	0.073 $\pm$ 0.002
$N = 10$	G_1	*0.399*	*0.700*	*0.090$\pm$ 0.015*	*0.182$\pm$0.045*
$N = 10$	$G_{2:M}$	[0.967, 0.977]	[0.996, 0.999]	0.077$\pm$0.003	0.104$\pm$0.002
$N = 40$	$\boldsymbol{B}$	[0.953, 0.969]	[0.993, 0.998]	0.041$\pm$0.007	0.050$\pm$0.001
$N = 40$	G_1	*0.398*	*0.686*	*0.045$\pm$0.003*	*0.093$\pm$0.013*
$N = 40$	$G_{2:M}$	[0.965,0.977]	[0.996,0.999]	0.038$\pm$0.001	0.051$\pm$0.001

Table 1: $M = 40$. Coverage of nominal 95% posterior intervals calculated from 2000 datasets simulated under a Poisson model. The intervals in columns 3 and 4 give the smallest and largest coverage observed for the corresponding parameter. The last two columns give the lengths of nominal 95% intervals in the format: mean $\pm$ standard deviation.

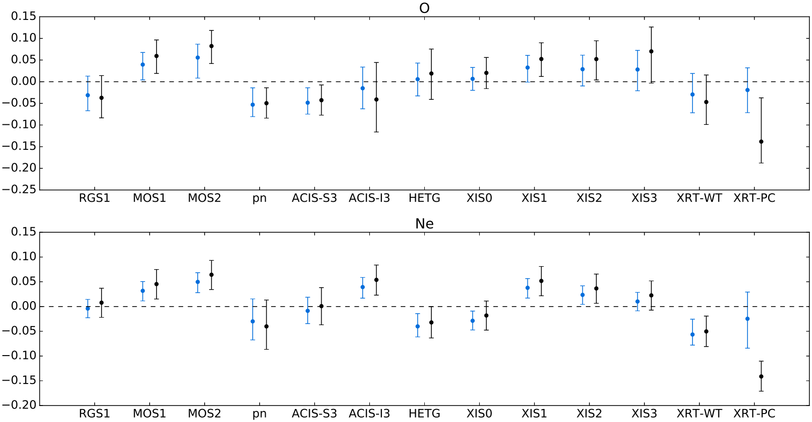
Numerical Results (E0102)

Recap: Supernova remnant E0102.

Four sources are four spectral lines in E0102.



Estimates of $B_i = \log A_i$ ($M = 2$ each panel)



- Adjusted so that default effective area, $b_i = \log a_i = 0$.
- 95% posterior intervals (black:$\tau = 0.05$; blue: $\tau = 0.025$).
- Some instruments systematically high, others low.

Prior Influence

Instrument	Oxygen		Neon	
	$\tau = 0.025$	$\tau = 0.05$	$\tau = 0.025$	$\tau = 0.05$
RGS1	0.570	0.205	0.063	0.016
MOS1	0.279	0.077	0.075	0.019
MOS2	0.355	0.065	0.077	0.017
pn	0.250	0.041	0.620	0.218
ACIS-S3	0.218	0.040	0.270	0.088
ACIS-I3	0.906	0.640	0.099	0.026
HETG	0.648	0.341	0.129	0.034
XIS0	0.180	0.051	0.069	0.018
XIS1	0.298	0.078	0.071	0.019
XIS2	0.463	0.140	0.063	0.016
XIS3	0.772	0.364	0.062	0.018
XRT-WT	0.726	0.278	0.154	0.026
XRT-PC	0.934	0.235	0.906	0.017

Table 2: Proportion of prior influence, as defined by $1 - W_i$, for E0102 data.

Numerical Results (2XMM)

- 2XMM catalog: used to generate large, well-defined samples of various types of astrophysical objects; collected with the XMM-Newton European Photon Imaging Cameras (EPIC).

Numerical Results (2XMM)

- 2XMM catalog: used to generate large, well-defined samples of various types of astrophysical objects; collected with the XMM-Newton European Photon Imaging Cameras (EPIC).
- Three EPIC instruments: the EPIC-pn, and the two EPIC-MOS detectors (pn, MOS1, and MOS2).

Numerical Results (2XMM)

- 2XMM catalog: used to generate large, well-defined samples of various types of astrophysical objects; collected with the XMM-Newton European Photon Imaging Cameras (EPIC).
- Three EPIC instruments: the EPIC-pn, and the two EPIC-MOS detectors (pn, MOS1, and MOS2).
- Three datasets: hard (2.5 - 10.0 keV), medium (1.5 - 2.5 keV) and soft (0.5 - 1.5 keV) energy bands. The three instruments (pn, MOS1 and MOS2) measured 41, 41, and 42 sources respectively in hard, medium, and soft bands. Faint sources.

Numerical Results (2XMM)

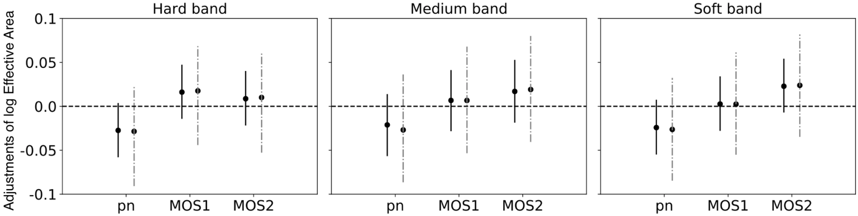


Figure 1: Adjustments of the log-scale Effective Areas for hard band (left), medium band (middle) and soft band (right) of the 2XMM datasets.

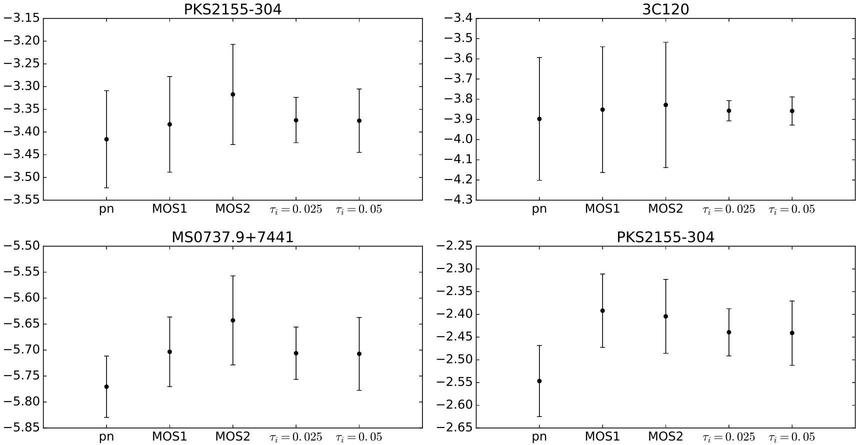
Numerical Results (XCAL)

- **XCAL data**: Bright active galactic nuclei from the XMM-Newton cross-calibration sample.
  - Observed in hard ($n = 94$), medium ($n = 103$), soft ($n = 108$) bands.
- **Pileup**: Image data are clipped to eliminate the regions affected by pileup, determined using `epatplot`.
- **Three detectors**: MOS1, MOS2 and pn.
- We fit our model and show results on

Sources: M=103 (in medium band).

The hard and soft bands data are fitted similarly – treating hard/medium/soft band as three different data sets.

Numerical Results (XCAL): Calibration Concordance



4 out of 103 Sources in medium band. y-axis: G (log flux); vertical bars (left 3 in each panel): mean $\pm$ 2 s.d. based on observed fluxes, vertical bars (right 2 in each panel): 95% posterior intervals based on our model.

Prior Influence

Data Name	$\tau_i = 0.025$			$\tau_i = 0.05$		
	pn	mos1	mos2	pn	mos1	mos2
hard band 2XMM	0.093	0.075	0.082	0.025	0.020	0.022
medium band 2XMM	0.250	0.216	0.222	0.076	0.065	0.067
soft band 2XMM	0.093	0.075	0.069	0.025	0.020	0.018
hard band XCAL	0.010	0.019	0.031	0.003	0.005	0.008
medium band XCAL	0.023	0.016	0.028	0.006	0.004	0.007
soft band XCAL	0.021	0.011	0.007	0.005	0.003	0.002

Table 3: Proportion of prior influence.

Summary

Statistics

1. *Multiplicative* mean modeling:

log-Normal hierarchical model.

Summary

Statistics

1. *Multiplicative* mean modeling:

   log-Normal hierarchical model.

2. Shrinkage estimators.

Summary

Statistics

1. *Multiplicative* mean modeling:

   log-Normal hierarchical model.

2. Shrinkage estimators.
3. Bayesian computation: MCMC & Stan.

Summary

Statistics

1. *Multiplicative* mean modeling:

   log-Normal hierarchical model.

2. Shrinkage estimators.
3. Bayesian computation: MCMC & Stan.
4. The potential pitfalls of assuming 'known' variances.

Summary

Statistics

1. *Multiplicative* mean modeling:

   log-Normal hierarchical model.

2. Shrinkage estimators.
3. Bayesian computation: MCMC & Stan.
4. The potential pitfalls of assuming 'known' variances.

Astronomy

1. Adjustments of effective areas of each instrument.

Summary

Statistics

1. *Multiplicative* mean modeling:

   log-Normal hierarchical model.

2. Shrinkage estimators.
3. Bayesian computation: MCMC & Stan.
4. The potential pitfalls of assuming 'known' variances.

Astronomy

1. Adjustments of effective areas of each instrument.
2. Calibration concordance.

Discussions: Ongoing and Future Work

- Correlations among instruments.
  - Estimated correlations based on theoretical simulations.
  - Prior? Extra data? Uncertainty?

Discussions: Ongoing and Future Work

- Correlations among instruments.
  - Estimated correlations based on theoretical simulations.
  - Prior? Extra data? Uncertainty?
- Robustness $\Rightarrow$ Misspecified models.

Discussions: Ongoing and Future Work

- Correlations among instruments.
  - Estimated correlations based on theoretical simulations.
  - Prior? Extra data? Uncertainty?
- Robustness $\Rightarrow$ Misspecified models.
- Better quantification of prior influence.

Discussions: Ongoing and Future Work

- Correlations among instruments.
    - Estimated correlations based on theoretical simulations.
    - Prior? Extra data? Uncertainty?
- Robustness $\Rightarrow$ Misspecified models.
- Better quantification of prior influence.
- Coverage properties when outliers exist.

Acknowledgement

Yang Chen (UMich), Xufei Wang (Two Sigma), Xiao-Li Meng (Harvard), David van Dyk (ICL), Herman Marshall (MIT) & Vinay Kashyap (cfA)