Accounting for Calibration Uncertainty:

High Energy Astrophysics and the PCG Sampler

David A. van Dyk[1]
Vinay Kashyap[2] Taeyoung Park[3] Jin Xu[4]
Imperial-California-Harvard AstroStatistics Collaboration

[1]Statistics Section, Imperial College London
[2]Smithsonian Astrophysical Observatory
[3]Yonsei University
[4]Department of Statistics, University of California, Irvine

Imperial College London, March 2012

Imperial College London

Outline

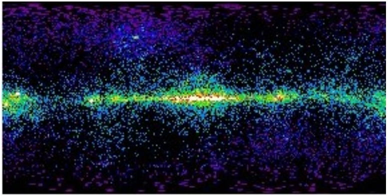
Imperial College London

Outline

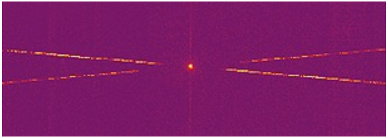
Imperial College London

High-Energy Astrophysics

- Produced by multi-million degree matter, e.g., magnetic fields, extreme gravity, or explosive forces.
- Provide understanding into the hot turbulent regions of the universe.
- X-ray and γ-ray detectors typically count a *small number of photons* in each of a *large number of pixels*.



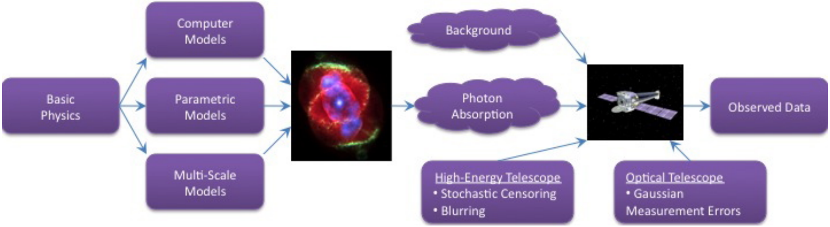
EGERT γ-ray counts >1GeV (entire sky and mission life).



Dispersion grating spectrum of an Active Galactic Nucleus; emission from matter accreting onto a massive Black Hole.

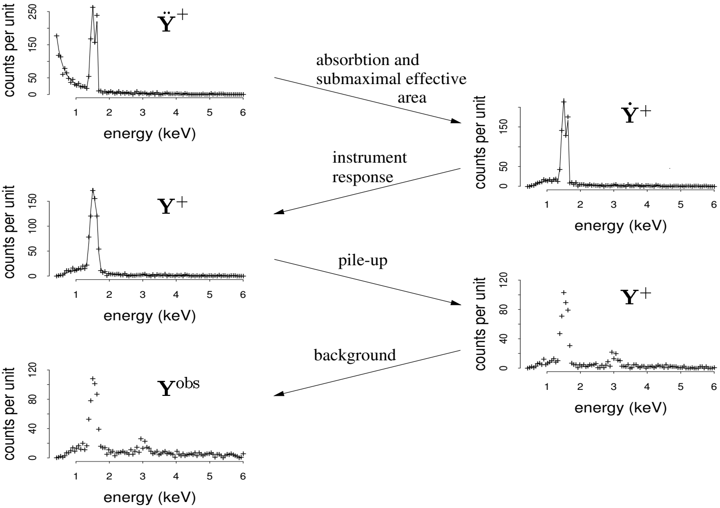
Imperial College London

The Basic Statistical Model



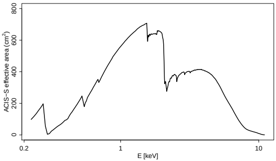
- Aim to formulate models in terms of specific questions of scientific interest.
- Must account for complexities of data generation.
- Embed complex physics-based and/or instrumental models into multi-level statistical models.
- State of the art data and computational techniques enable us to fit the resulting complex model.

Degradation of the Photon Counts

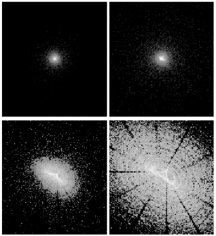


Calibration Products

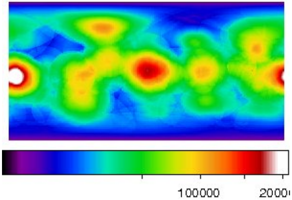
- Analysis is highly dependent on *Calibration Products:*
  - Effective area records sensitivity as a function of energy
  - Energy redistribution matrix can vary with energy/location
  - Point Spread Functions can vary with energy and location
  - Exposure Map shows how effective area varies in an image
- In this talk we focus on uncertainty in the effective area.



A *CHANDRA* effective area.



Sample Chandra psf's
(Karovska et al., ADASS X)

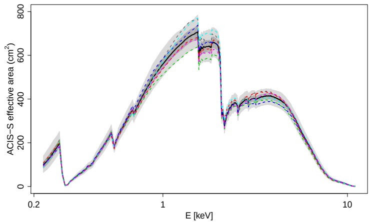


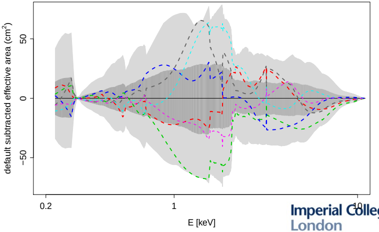
EGERT exposure map
(area × time)

Imperial College
London

Derivation of Calibration Products

- Prelaunch ground-based and post-launch space-based empirical assessments.
- Aim to capture deterioration of detectors over time.
- Complex computer models of subassembly components.
- Calibration scientists provide a sample representing uncertainty
- *Calibration Sample* is typically of size $M \approx 1000$.





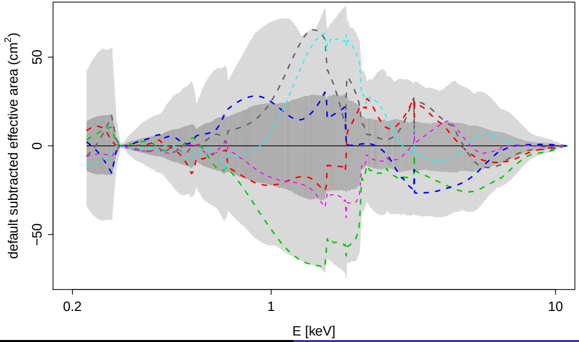
Imperial College London

Outline

Imperial College London

Complex Variability

Using a computer model generated calibration sample requires storage of a large number of high dimensional calibration products.



Imperial College London

Simple Emmulation of Complex Variability

We use *Principal Component Analysis* to represent uncertainly:

$$A \sim A_0 + \bar{\delta} + \sum_{j=1}^{m} e_j r_j \boldsymbol{v}_j,$$

A_0: default effective area,

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

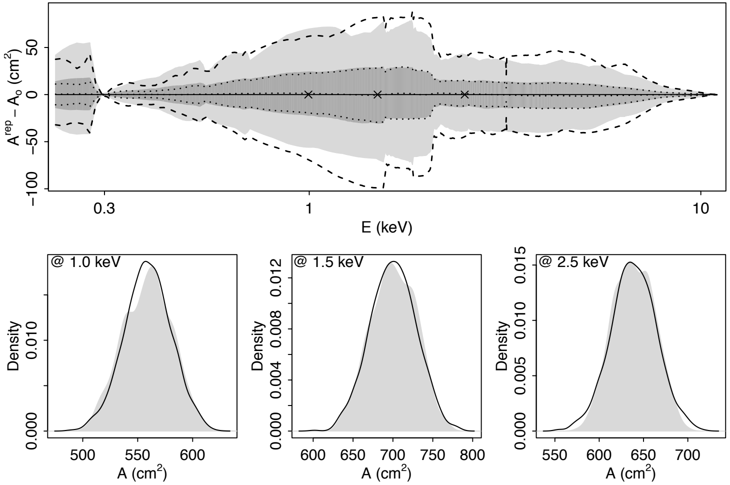
r_j and $\boldsymbol{v}_j$: first m principle component eigenvalues & vectors,

e_j: independent standard normal deviations.

Capture 95% of variability with $m = 6 - 9$.

Imperial College London

Accounting for Uncertainty



Imperial College London

Using Monte Carlo to Account for Uncertainty

1. Drake et al. (2006) propose a bootstrap-like method:
   - Simulate M spectra under fit model & default effective area.
   - Fit each spectra with effective area from calibration sample.
2. A simpler solution involves Multiple Imputation:
   - Treat $m \ll M$ effective areas from calibration sample as imputations and fit the model m times.
   - Use MI *Combining Rules* to compute estimates and errors.
3. When using MCMC in a Bayesian setting we can:
   - Sample a different effective area from calibration sample at each iteration according to its conditional distribution.
   - Effectively average over the calibration uncertainty.

Two Possible Target Distributions

We consider inference under:

A PRAGMATIC BAYESIAN TARGET: $\pi_0(A, \theta) = p(A)p(\theta|A, Y)$.

THE FULLY BAYESIAN POSTERIOR: $\pi(A, \theta) = p(A|Y)p(\theta|A, Y)$.

Concerns:

Statistical Fully Bayesian target is “correct”.

Cultural Astronomers have concerns about letting the current data influence calibration products.

Computational Both targets pose challenges, but pragmatic Bayesian target is easier to sample.

Practical How different are $p(A)$ and $p(A|Y)$?

Imperial College London

Outline

The Partially Collapsed Gibbs Sampler

A Gibbs sampler:

STEP 1: $\psi_1 \sim p(\psi_1|\psi_2)$
STEP 2: $\psi_2 \sim p(\psi_2|\psi_1)$

A Partially Collapsed Gibbs (PCG) Sampler:

STEP 1: $\psi_1 \sim p(\psi_1|g(\psi_2))$
STEP 2: $\psi_2 \sim p(\psi_2|h(\psi_1))$

- g and/or h are non-invertible functions.
- Generalizes blocking & collapsing, involves incompatibility.
- Step order can effect stationary distribution.
- Improves convergence rate (van Dyk & Park, 2008, JASA).
- Spectral analysis, time series, and multiple imputation (van Dyk & Park; 2011 MCMC Hndbk; 2009 JCGS; 2008 ApJ).

Metropolis-Hastings within Gibbs & PCG Samplers

An MH within Gibbs sampler:

STEP 1: $\psi_1 \sim \mathcal{K}(\psi_1|\psi)$ via MH with limiting dist. $\underline{p(\psi_1|\psi_2)}$

STEP 2: $\psi_2 \sim p(\psi_2|\psi_1)$

Using MH within the Partially Collapsed Gibbs Sampler:

STEP 1: $\psi_1 \sim \mathcal{K}(\psi_1|\psi)$ via MH with limiting dist. $\underline{p(\psi_1)}$

STEP 2: $\psi_2 \sim p(\psi_2|\psi_1)$

- If MH is unnecessary, obtain i.i.d. draws from $p(\psi_1, \psi_2)$.
- With MH we must verify the stationary distribution.
- Improved convergence if ψ_1 and ψ_2 are highly correlated.
- Need only evaluate $p(\psi_1) = p(\psi_1, \psi_2)/p(\psi_2|\psi_1)$.

Imperial College London

But... Be Careful!

Another MH within Gibbs Sampler:

STEP 1: $\psi_1 \sim p(\psi_1|\psi_2)$

STEP 2: $\psi_2 \sim \mathcal{K}(\psi_2|\psi_1)$ via MH with limiting dist. $p(\psi_2|\psi_1)$

A *naive* Sampler:

STEP 1: $\psi_1 \sim p(\psi_1)$

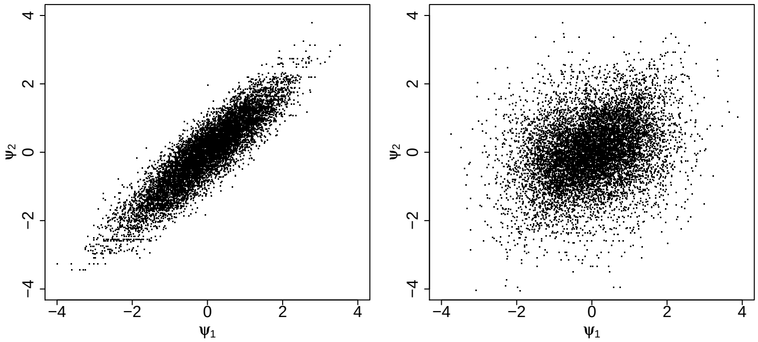
STEP 2: $\psi_2 \sim \mathcal{K}(\psi_2|\psi_1)$ via MH with limiting dist. $p(\psi_2|\psi_1)$

Simulation Study:

- Suppose $\binom{\psi_1}{\psi_2} \sim \mathrm{N}_2\left[\binom{0}{0}, \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}\right]$
- MH: a Gaussian jumping rule centered at previous draw.

Be Careful When Combining MH and PCG Sampling

MH within Gibbs Sampler | The *naive* Sampler



Imperial College London

What Goes Wrong

The *naive* Sampler:

STEP 1: $\psi_1^{(t)} \sim p(\psi_1)$

STEP 2: $\psi_2^{(t)} \sim \mathcal{K}(\psi_2|\psi_1^{(t)}, \psi_2^{(t-1)})$ via Metropolis Hastings

The update of ψ_2 depends on both $\psi_1^{(t)}$ and $\psi_2^{(t-1)}$:

- The limiting distribution of the MH step is $p(\psi_2|\psi_1^{(t)})$.
- If the proposal is rejected, ψ_2 is set to $\psi_2^{(t-1)}$.

BUT: $\psi_1^{(t)} \sim p(\psi_1)$—independent of $\psi_2^{(t-1)}$ *at every iteration*.

STEP 2 *will never produce samples from* $p(\psi_2|\psi_1)$.

Constructing a Legitimate MH within PCG Sampler

1. <u>Marginalzing</u>

$p(\psi_1|\psi_2')$
$\mathcal{K}(\psi_2|\psi')$ w/ limit $p(\psi_2|\psi_1)$ $\longrightarrow$ $p(\psi_1|\psi_2')$
$\mathcal{K}(\psi_1,\psi_2|\psi')$ w/ limit $p(\psi_1,\psi_2)$

Move quantities from the right to the left of the conditioning sign. This does not alter the stationary dist'n, but improves the rate of convergence.

2. <u>Permuting</u>

$\longrightarrow$ $\mathcal{K}(\psi_1,\psi_2|\psi')$ w/ limit $p(\psi_1,\psi_2)$
$p(\psi_1|\psi_2')$

Permute the order of the steps. This can have minor effects on the rate of convergence, but does not affect the stationary distribution.

3. <u>Trimming</u>

$\longrightarrow$ $\mathcal{K}(\psi_2|\psi')$ w/ limit $p(\psi_2)$
$p(\psi_1|\psi_2')$

Remove quantities that are not part of the transition kernel. This does not effect the stochastic mapping or the rate of convergence.

Imperial College London

Back to Calibration Uncertainty

We consider inference under:

A PRAGMATIC BAYESIAN TARGET: $\pi_0(A, \theta) = p(A)p(\theta|A, Y)$.

THE FULLY BAYESIAN POSTERIOR: $\pi(A, \theta) = p(A|Y)p(\theta|A, Y)$.

Sampling either $\pi_0(A|\theta)$ or $\pi(A|\theta)$ is complicated.

- PCA emulator: degenerate normal approximation to $p(A)$.
- $\pi(A|\theta) = p(A|\theta, Y)$ can be evaluated, suggesting MH update for Fully Bayesian.
- $\pi_0(A|\theta) \propto p(A, \theta|Y)p(Y)/p(Y|A)$ is difficult to evaluate.

Sample of $\pi_0(A) = p(A)$ suggests PCG sampler for Pragmatic Bayes.

Pragmatic Bayes: PCG Sampler

A simple Markov Chain Monte Carlo Procedure

- Sample effective area uniformly from calibration sample: $A \sim p(A)$.
- Sample model parameters in the usual way, conditioning on the current sample of the effective area: $\theta \sim p(\theta|A, Y)$.

To obtain independent draws from $\pi_0(A, \theta)$.

This strategy in effect replaces a posterior draw with a prior draw when updating the effective area.

Imperial College London

Pragmatic Bayes: MH within PCG Sampler

Unfortunately, update of θ uses MH (pyBLoCXS in Sherpa) with limiting distribution $p(\theta|A, Y)$.

The *naive* Sampler Revisited:

STEP 1: $A^{(t)} \sim p(A)$

STEP 2: $\theta^{(t)} \sim \mathcal{K}(\theta|A^{(t)}, \theta^{(t-1)})$ via Metropolis Hastings.

A simple solution is a PCG (Simple Collapsed) Gibbs Sampler:

STEP 1: $A^{(t)} \sim p(A)$

STEP 2: Iteratively sample $\theta^{(t-1+k/K)} \sim \mathcal{K}(\theta|A^{(t)}, \theta^{(t-1)})$ via MH to obtain $\theta^{(t)} \sim p(\theta|A^{(t)})$.

In practice a moderate value of K (<10) is sufficient.

Pragmatic Bayes: Proposed Procedures

Calibration team provide users with:

- $\boldsymbol{a}_0$ and $\bar{\delta}$
- $m < 10$ PCA eigen values and vectors

Analysis will proceed in one of two ways:

1. When using MCMC, each iteration includes:
   - Sample effective area via $\boldsymbol{a} \sim \boldsymbol{a}_0 + \bar{\delta} + \sum_{j=1}^{m} e_j r_j \boldsymbol{v}_j$.
   - Iteratively sample model parameters in the usual way, conditioning on the current sample of the effective area.
2. When not using MCMC:
   - Sample m effective areas via $\boldsymbol{a} \sim \boldsymbol{a}_0 + \bar{\delta} + \sum_{j=1}^{m} e_j r_j \boldsymbol{v}_j$.
   - Fit model with each of the sampled effective areas.
   - Compute fitted values and errors using MI combining rules.

Imperial College London

Sampling the Full Posterior Distribution

- Sampling $\pi(A, \theta) = p(A, \theta | Y)$ is complicated because we only have a computer-model generated sample of $p(A)$ rather than an analytic form.
- But PCA gives a *degenerate normal* approximation:

$$A \sim A_0 + \bar{\delta} + \sum_{j=1}^{m} e_j r_j \boldsymbol{v}_j,$$

  where e_j are independent standard normals.
- PCA represents A as deterministic function of $e = (e_1, \ldots, e_m)$.
- We can construct an MCMC sampler of $p(e, \theta | Y)$.

Imperial College London

Implementing the Fully Bayesian Sampler

An MH within Gibbs Sampler:

STEP 1: $e \sim \mathcal{K}(e|e', \theta')$ via MH with limiting dist'n $p(e|\theta, Y)$

STEP 2: $\theta \sim \mathcal{K}(\theta|e', \theta')$ via MH with limiting dist'n $p(\theta|e, Y)$

- We use a mixture of two jumping rules in each step:
  - STEP 1: Gaussian Metropolis jump centered at e' and MH jump from the prior.
  - STEP 2: T Metropolis jump centered at θ' and MH jump from an approximation to the posterior.
- Six tuning parameters: three scale parameters, two proportions (M vs MH jump), and *df* for T.
- Tuning parameters must be adjusted.

Imperial College London

Outline

The Simulation Studies

Simulated Spectra

- Spectra were sampled using an absorbed power law,

$$f(E_j) = \alpha e^{-N_H x(E_j)} E_j^{-\Gamma},$$

accounting for instrumental effects; E_j is the energy of bin j.

- Parameters (Γ and N_H) and sample size/exposure times:

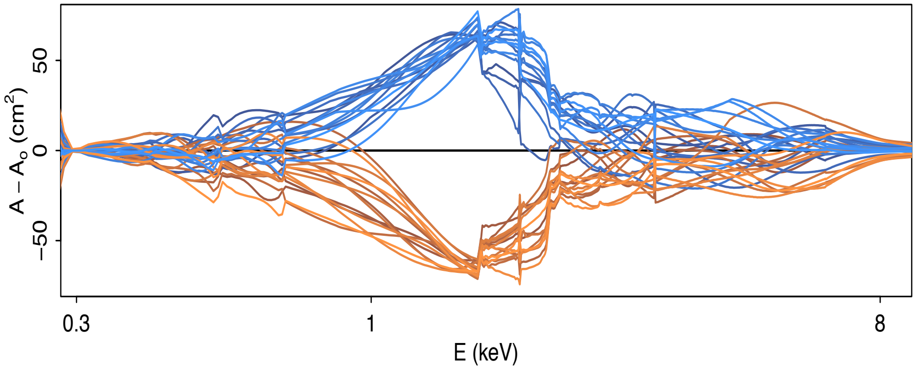
	Effective Area		Nominal Counts		Spectal Model	
	Default	Extreme	10^5	10^4	Hard†	Soft‡
SIM 1	X		X		X	
SIM 2	X		X			X
SIM 3	X			X	X	

†An absorbed powerlaw with $\Gamma = 2$, $N_{\rm H} = 10^{23}/{\rm cm}^2$

‡An absorbed powerlaw with $\Gamma = 1$, $N_{\rm H} = 10^{21}/{\rm cm}^2$

Imperial College London

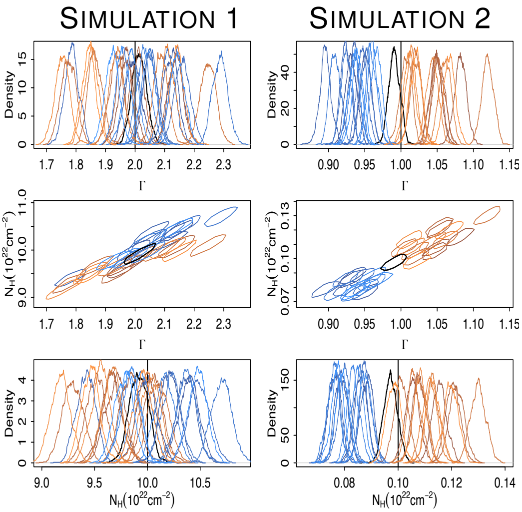
30 Most Extreme Effective Areas in Calibration Sample



15 largest and 15 smallest determined by maximum value

Imperial College London

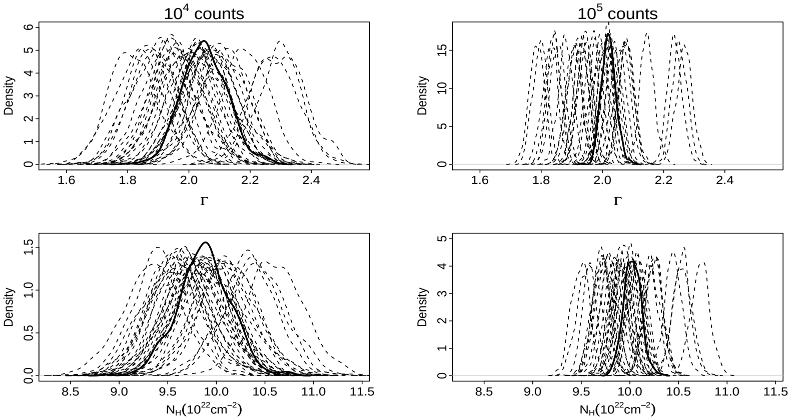
The Effect of Calibration Uncertainty



- Columns represent two simulated spectra.
- True parameters are horizontal lines.
- Posterior under default calibration is plotted in black.
- The posterior is highly sensitive to the choice of effective area!

Imperial College London

The Effect of Sample Size



The effect of Calibration Uncertainty is more pronounced with larger sample sizes.

Expanded Simulation for Pragmatic Bayes

Simulated Spectra

- Spectra were sampled using an absorbed power law,

$$f(E_j) = \alpha e^{-N_H x(E_j)} E_j^{-\Gamma},$$

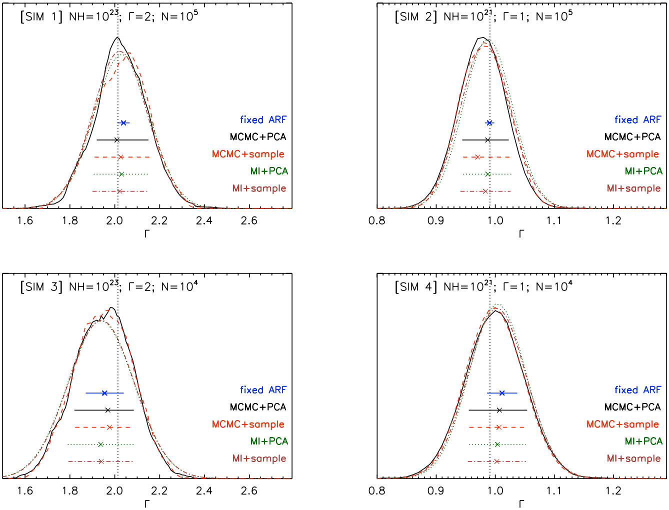
	Effective Area		Nominal Counts		Spectal Model	
	Default	Extreme	10^5	10^4	Hard†	Soft‡
SIMULATION 1	X		X		X	
SIMULATION 2	X		X			X
SIMULATION 3	X			X	X	
SIMULATION 4	X			X		X
SIMULATION 5		X	X		X	
SIMULATION 6		X	X			X
SIMULATION 7		X		X	X	
SIMULATION 8		X		X		X

† An absorbed powerlaw with $\Gamma = 2$, $N_{\rm H} = 10^{23}/{\rm cm}^2$

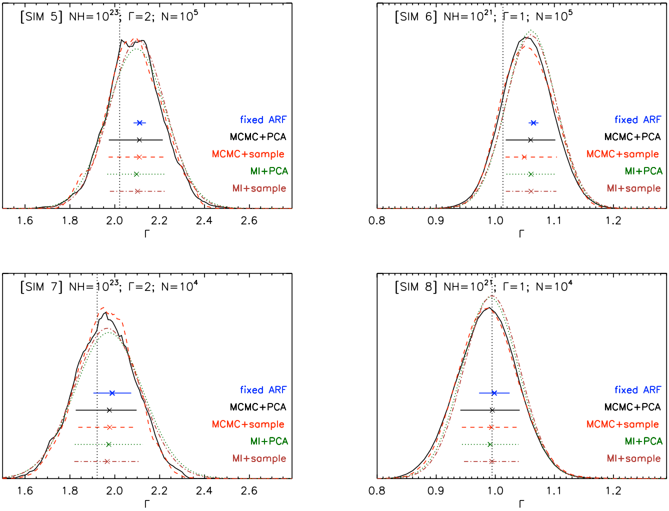
‡ An absorbed powerlaw with $\Gamma = 1$, $N_{\rm H} = 10^{21}/{\rm cm}^2$

Imperial College London

Pragmatic Bayes: Higher Variance Than Default



Pragmatic Bayes: Better Coverage Than Default



Imperial College London

A Simple Simulation for the Fully Bayesian Sampler

A Simple Simulation.

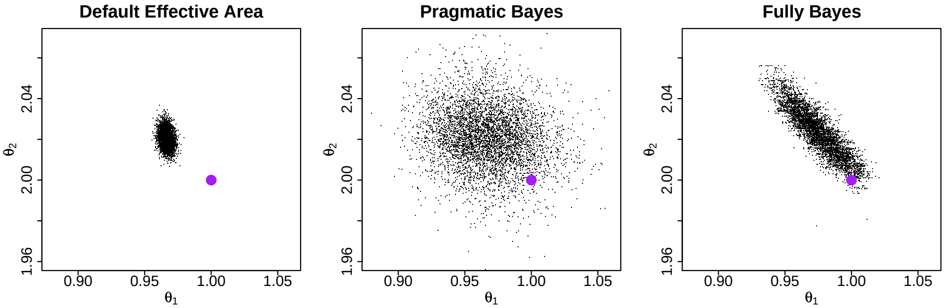
- Sampled 10^5 counts from a power law spectrum:

$$f(E_j) = \theta_1 e^{-\theta_3 x(E_j)} E_j^{-\theta_2}$$

- No energy blurring or backgraound contamination.
- Effective area used in the simulation differed from default:

  A_{true} is 1.5σ from the center of the calibration sample.

Sampling From the Full Posterior

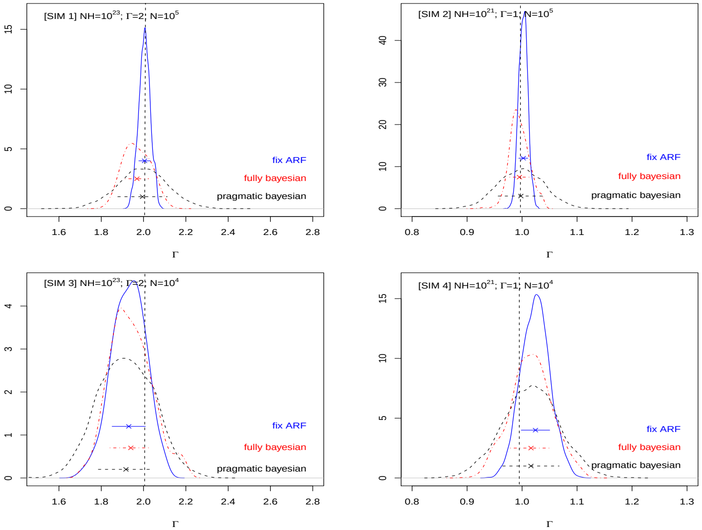


θ_1 = normalization, θ_2 = power law parameter
purple bullet = truth

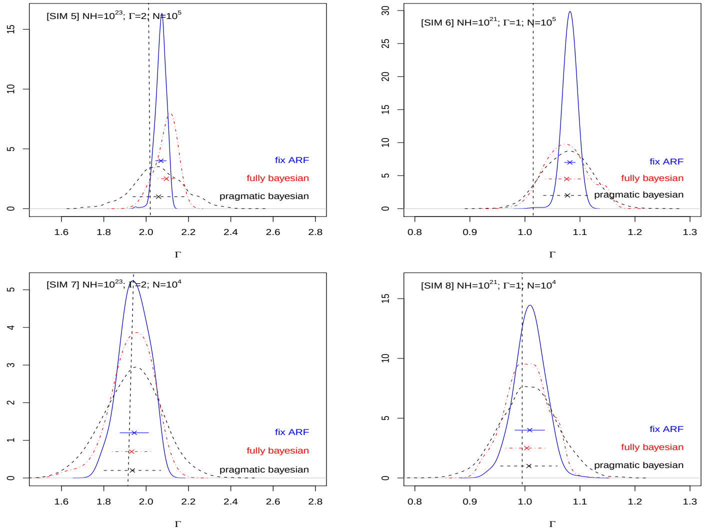
Pragmatic Bayes is clearly better than current practice, but a Fully Bayesian Method is the ultimate goal.

Imperial College London

Fully Bayesian: Less Variance that Pragmatic



Fully Bayesian: Better Coverage than Default



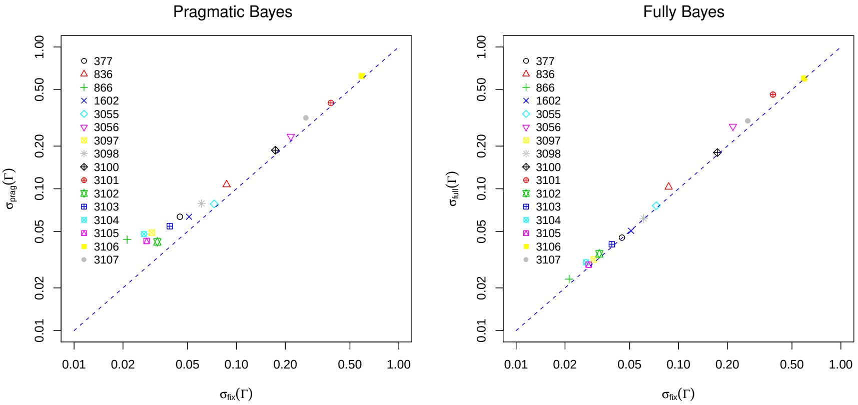
Imperial College London

The Effect of Sample Size Redux

A Set of Radio Loud Quasar Spectra

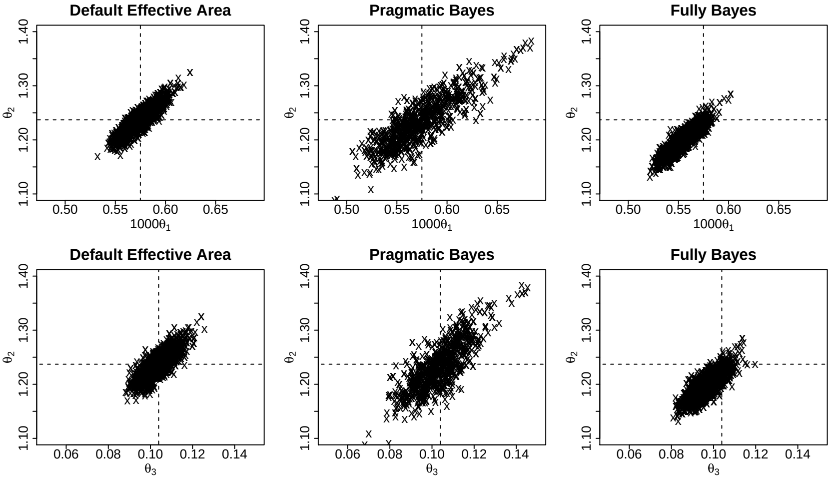
- Pragmatic and Fully Bayesian Methods were applied to a set of Quasars.
- Quasars are among the most distant distinguishable astronomical objects.
- The sixteen Quasar observations varied is size from 20 to over 10,000 photon counts.

Results



For large spectra calibration uncertainty swamps statistical error.
In large spectra fully Bayes identifies A and reduces uncertainty.

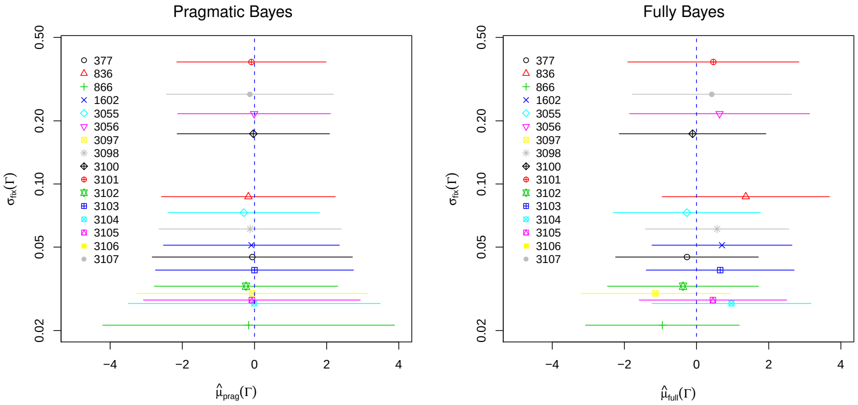
Quasar 866



Fully Bayes Shifts Posterior Without Increasing SD.

Imperial College London

Results: 95% Intervals Standardized by Standard Fit



For large spectra calibration uncertainty swamps statistical error.
In large spectra fully Bayes identifies A and shifts interval.

For Further Reading I

- Lee, H., Kashyap, V., van Dyk, D., Connors, A., Drake, J., Izem, R., Min, S., Park, T., Ratzlaff, P., Siemiginowska, A., and Zezas, A.
  Accounting for Calibration Uncertainties in X-ray Analysis: Effective Area in Spectral Fitting.
  The Astrophysical Journal, **731**, 126–144, 2011.
- van Dyk, D. A. and Park, T.
  Partially Collapsed Gibbs Sampling & Path-Adaptive Metropolis-Hastings in
  High-Energy Astrophysics. In *Handbook of Markov Chain Monte Carlo* (Editors: S. Brooks, A. Gelman, G. Jones and X.-L. Meng), Chapman & Hall/CRC Press, 383-399, 2011.
- Park, T. and van Dyk, D. A.
  Partially Collapsed Gibbs Samplers: Illustrations and Applications.
  Journal of Computational and Graphical Statistics, **18**, 283-305, 2009.
- van Dyk, D. A. and Park, T.
  Partially Collapsed Gibbs Samplers: Theory and Methods.
  Journal of the American Statistical Association, **103**, 790–796, 2008.
- Park, T., van Dyk, D. A., and Siemiginowska, A.
  Searching for Narrow Emission Lines in X-ray Spectra: Computation and Methods.
  The Astrophysical Journal, **688**, 807–825, 2008.

Imperial College London