# Adventures in Astrostatistics

Chandra X-ray Center

Images and videos courtesy of NASA/*Chandra/HST* unless otherwise noted

**Aneta Siemiginowska**

CENTER FOR ASTROPHYSICS

HARVARD & SMITHSONIAN

# Adventures in Astrostatistics

## High Energy Astrophysics

Chandra X-ray Center

**Aneta Siemiginowska**

CENTER FOR **ASTROPHYSICS**
HARVARD & SMITHSONIAN

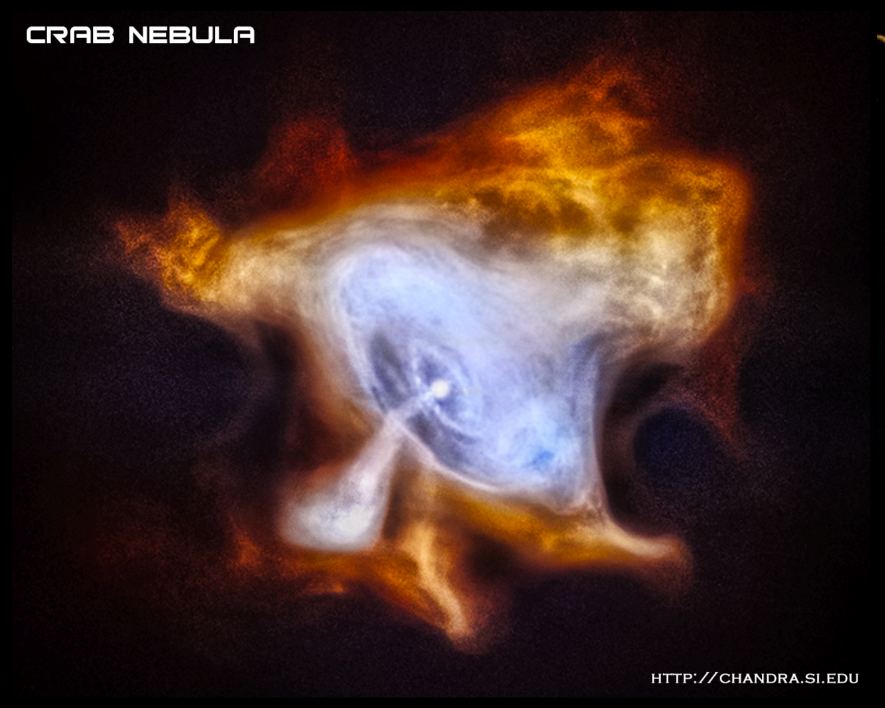
Images and videos courtesy of NASA/*Chandra/HST* unless otherwise noted

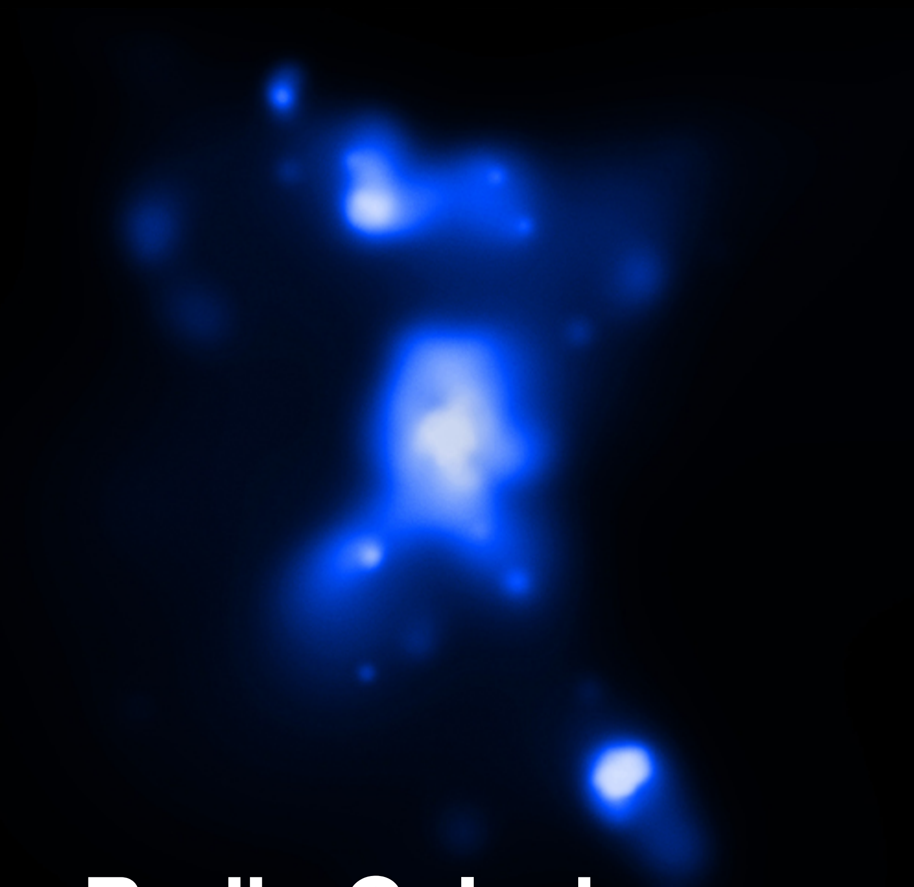
# X-ray Universe

**Solar System**

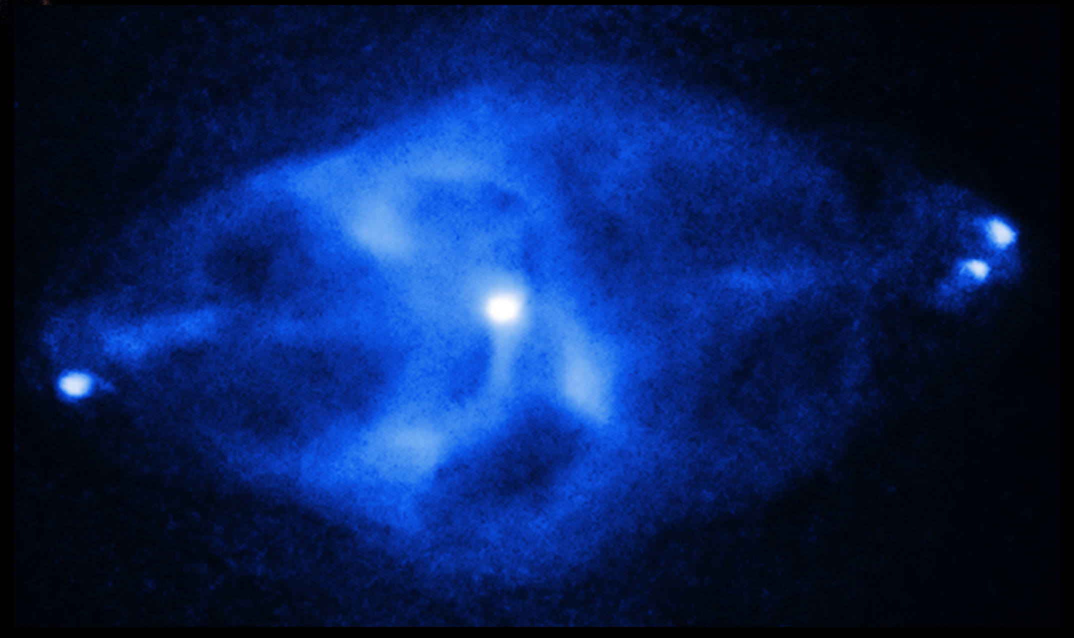
Hot gas > $10^5$ K
Energetic particles

**Supernova Remnants**

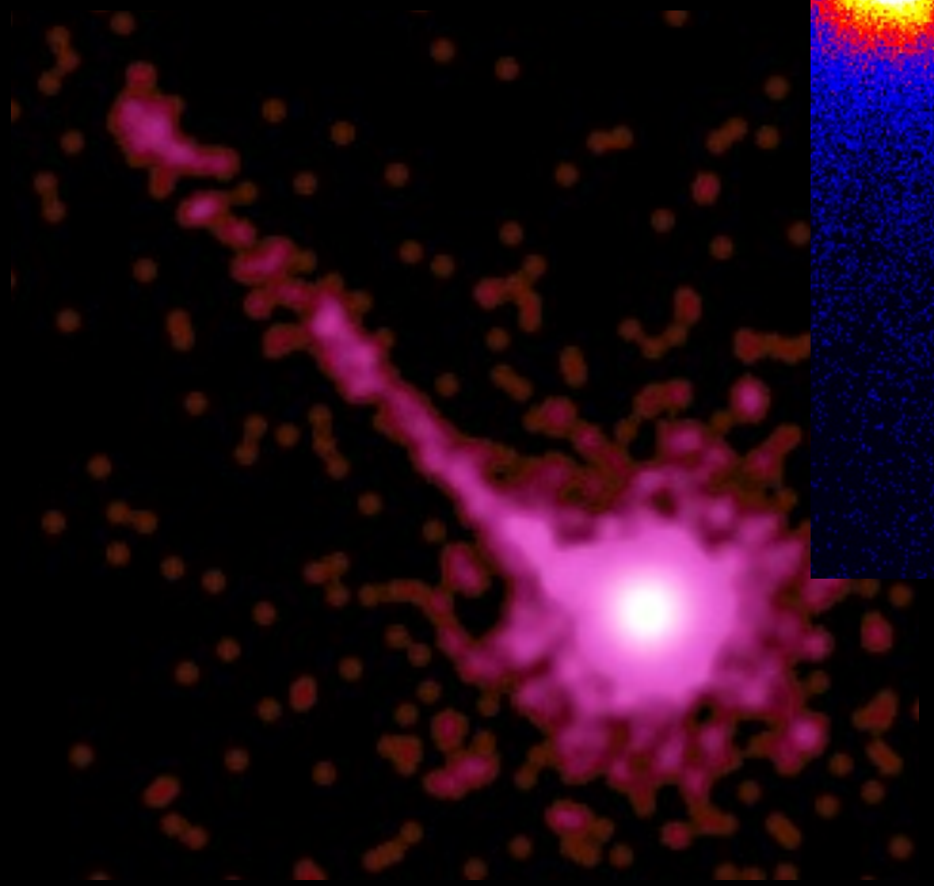


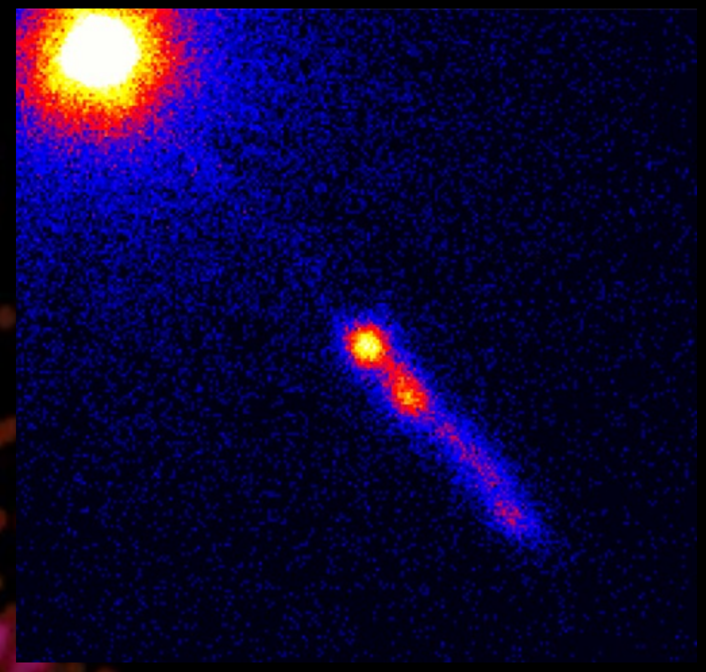



**Radio Galaxies**

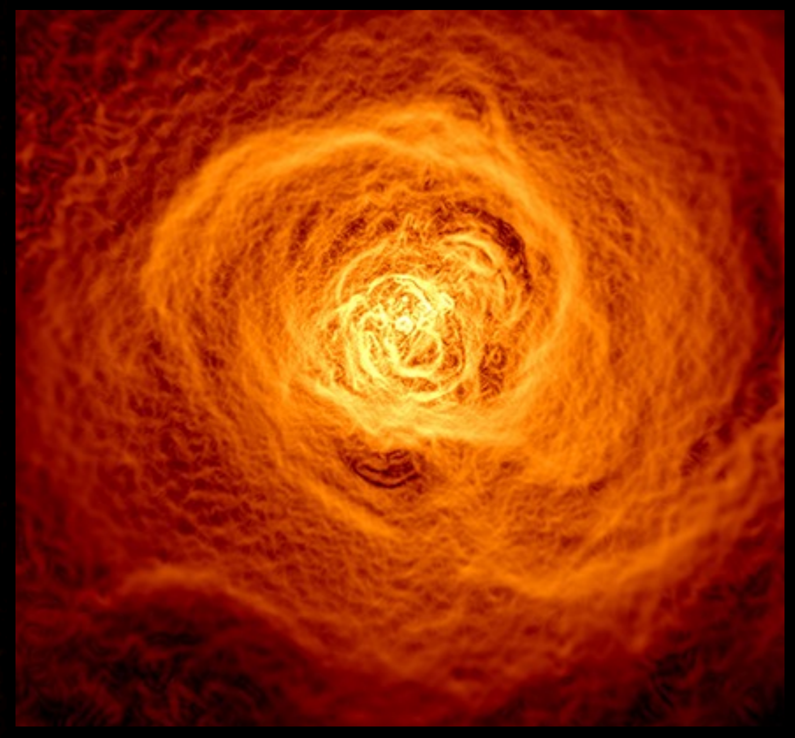


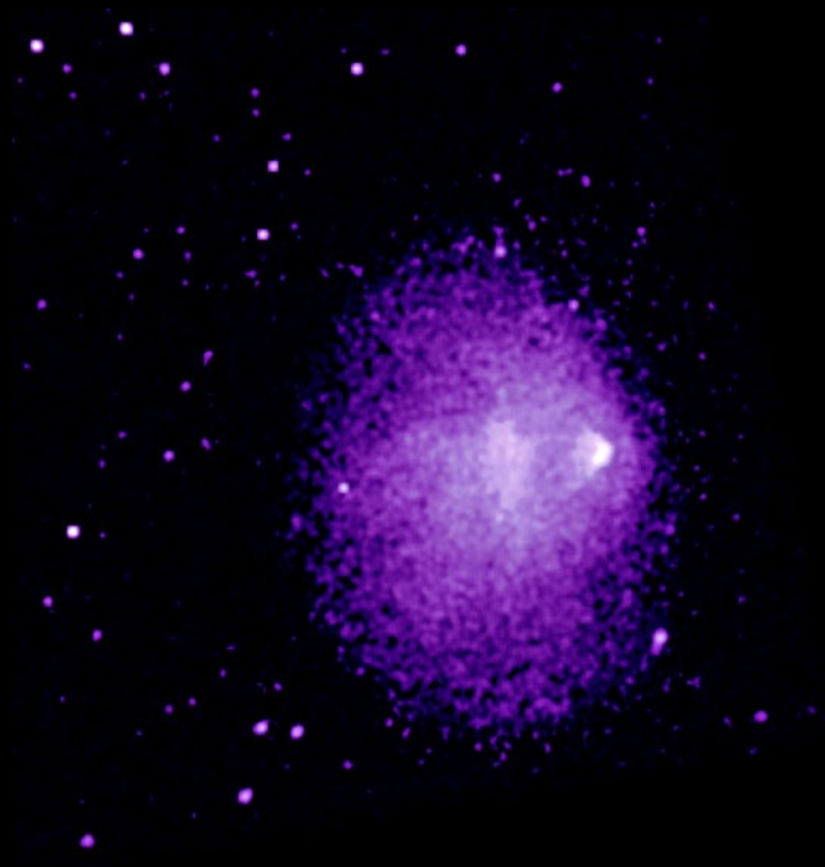
**Quasar Jets**





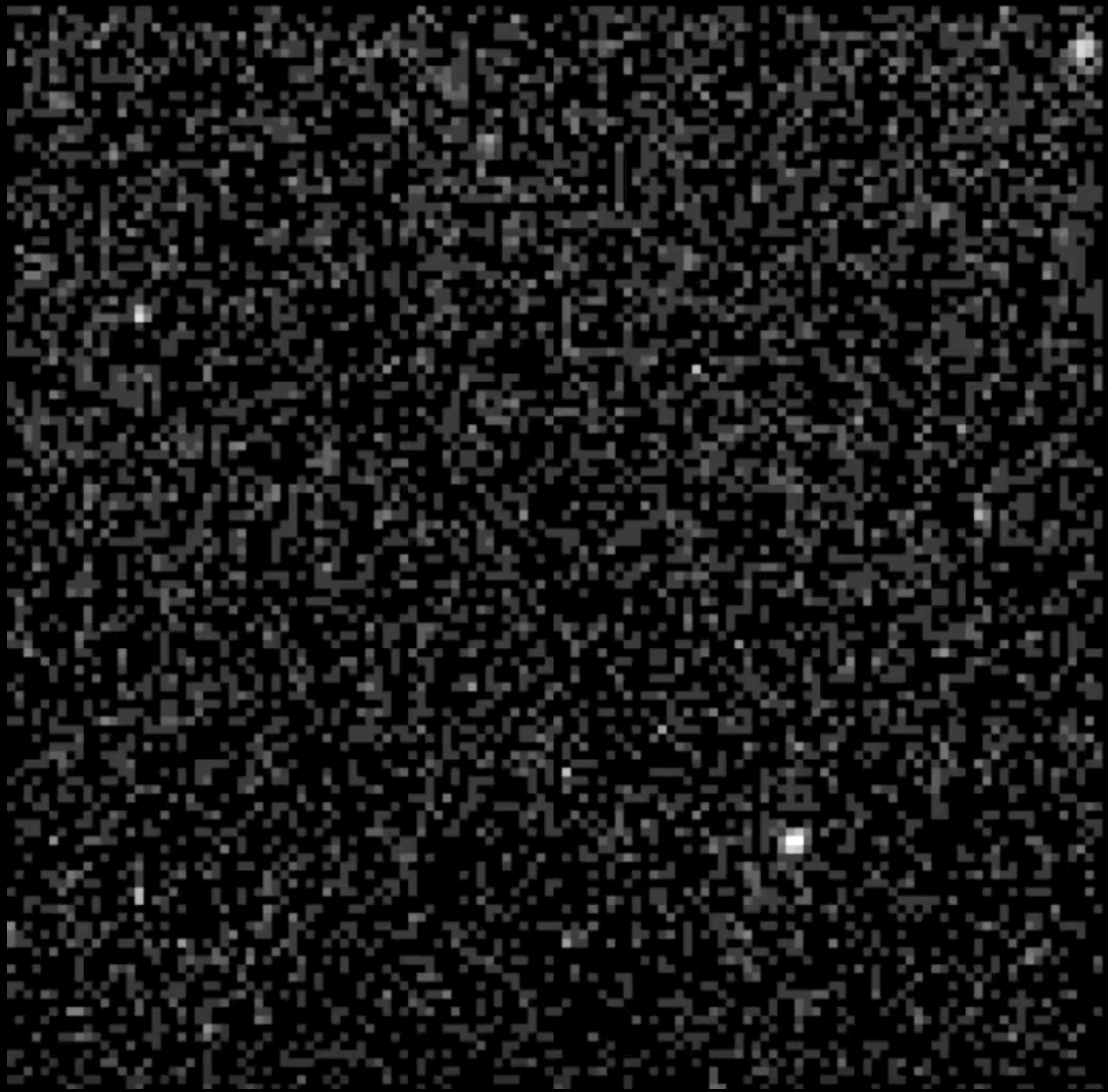
**Clusters of Galaxies**





# CHANDRA FIRST LIGHT... AUGUST 12 1999

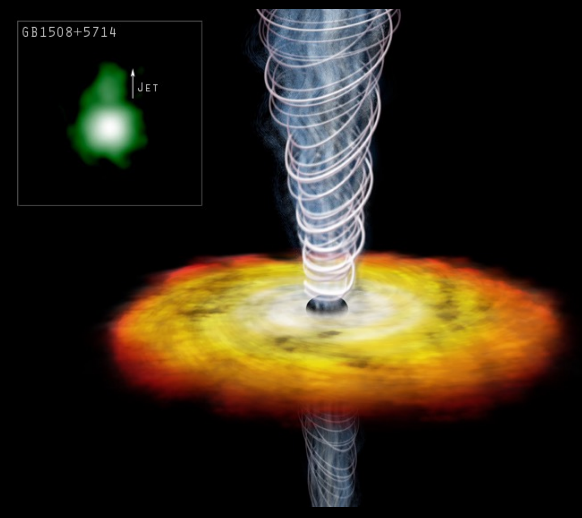
Launch of Chandra
(23rd July 1999)



# Outline

- Beginning
- Adventures:
    - A1 Bayesian Framework:
        - likelihood, priors, marginalization, MCMC, calibration uncertainties Why?
    - A2 Model Selection - hypothesis testing, Protassov et al, Park et al - detecting spectral lines
    - A3 Hardness Ratio - Upper limits, detection significance
    - A4 High resolution images - reconstruction, source boundaries, significance
- Emerging Methodology and Future

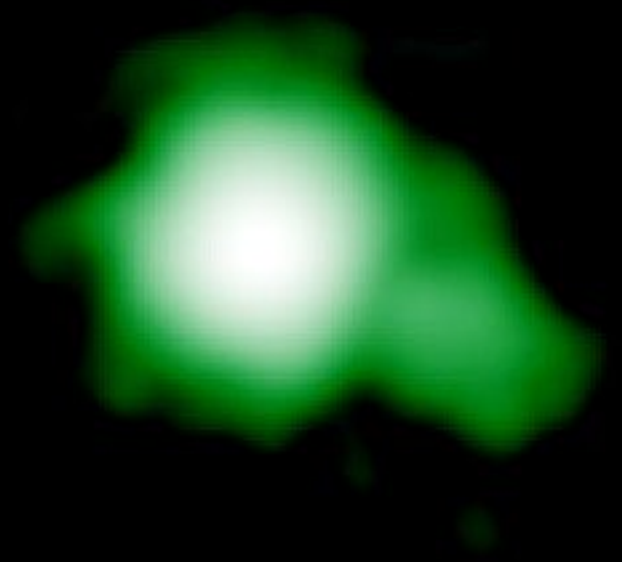
# Scientific Experiment



Define the Experiment:
- idea, proposal for observations, simulations, estimating required telescope time

Observations and Data Collection

Data Preparation:
- standard processing, instrument specific software, calibration



Data Analysis and Scientific Measurements

Final Conclusion

Statistics

# Beginning

- Astro-Statistics started in the ancient times with the statistical methodology developed and applied to astronomical data over thousands years.

  Note: modern statistics is much more than calculating mean and std!

- X-ray astronomy started in the ‘60s

- Methodology based on techniques developed in the past is not directly applicable to X-ray data -  several issues and potential approaches were noted in the early papers

- Collaborations with Statisticians!


Energy Spectra of X-ray Clusters of Galaxies
Avni, Y.
1976, ApJ, 210, 642

Parameter Estimation in Astronomy through Application of the Likelihood Ratio
Cash, W.
1979, ApJ, 228, 939

Generation of Confidence Intervals for Model Parameters in X-ray Astronomy
Cash, W.
1976, A&A, 52, 307

Parameter Estimation in X-ray Astronomy
Lampton, M.; Margon, B.; Bowyer, S.
1976, ApJ, 208, 177

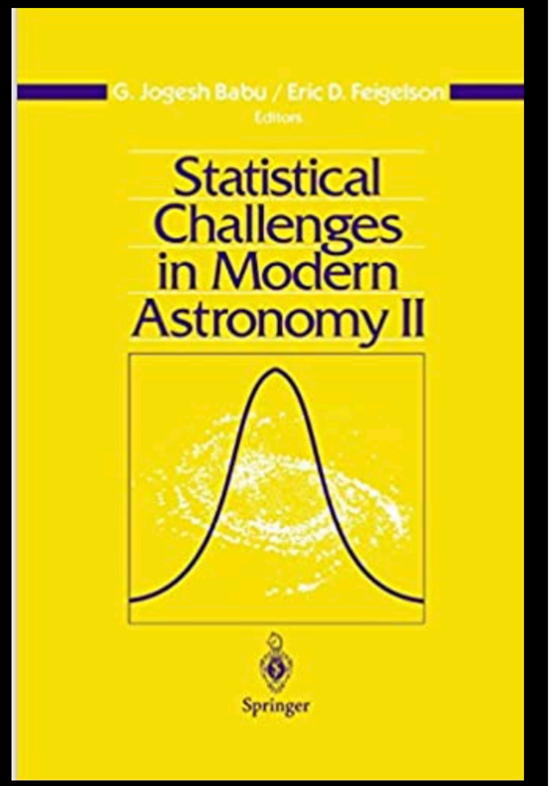
Parameter Estimation in X-ray Astronomy using Maximum Likelihood
Wachter, K.; Leach, R.; Kellogg, E.
1979, ApJ, 230, 274

Chi-squared and C Statistic Minimization for Low Count per Bin Data
Nousek, John A.; Shue, David R.
1989, ApJ, 342, 1207

Determination of Confidence Limits for Experiments with Low Numbers of Counts
Kraft, R.P.; Burrows, D.N. Nousek, John, A.
1991, ApJ, 374, 344

# Beginning

- SCIMA 1996 - Chandra Data Challenges
- Collaboration with the Harvard Department of Statistics
- Workshops at the AAS HEAD meetings




## AXAF Data Analysis Challenges

Aneta Siemiginowska[1], Martin Elvis[1], Alanna Connors[2], Peter Freeman[3], Vinay Kashyap[3], and Eric Feigelson[4]

ABSTRACT The high quality of the AXAF X-ray data provides new challenges for the X-ray data analysis. It is clear that an "old" approach is not enough to fully exploit the capabilities of the AXAF instruments. We describe a few of the statistical and computational problems that we have so far identified. Some of them appear to be theoretically solvable but computationally challenging, while others state problems for theoretical statistics which, so far as we know, are unsolved.The problems divide, from an astronomical point of view, into: Modeling the Data (e.g. nonlinear parameter estimation, uncertainties in the model, weighting the data, correlated residuals), Source Detection (events in N-space, use of wavelets, significance of detected structures) and Instrument Related Issues (pile-up in AXAF ACIS, overlapping orders in grating spectra).

**CHASC web site: http://hea-www.harvard.edu/AstroStat/**

## International CHASC Astro-Statistics Collaboration

This page lists resources of specific interest to astronomers. For detailed descriptions and reports of C-BAS/ICHASC activities, see www2.imperial.ac.uk/~dvandyk/astrostat.php

Software | Activities | Bibliography | Astro jargon | Stat jargon | People | Mailing-List | Internal

astrostat-announce GoogleGroup | GoogleCalendar | AstroStat Slog Archive

# People

## Faculty/Researchers/Postdocs

Statisticians

- David van Dyk, Imperial College London
- Thomas Lee, University of California, Davis
- Xiao-Li Meng, Harvard
- Yaming Yu, University of California, Irvine
- David Jones, Texas A&M University
- Hyungsuk Tak, Penn State University
- David Stenning, Simon Fraser University
- Yang Chen, University of Michigan
- Raymond Wong, Texas A&M University
- Sara Algeri, University of Minnesota
- Alex Geringer-Sameth, Imperial College London
- Jianing Zhang, Imperial College London
- Bernhard Klingenberg, Williams College

Astronomers

- Andreas Zezas, Crete and Center for Astrophysics | Harvard & Smithsonian
- Aneta Siemiginowska, Center for Astrophysics | Harvard & Smithsonian
- Axel Donath, Center for Astrophysics | Harvard & Smithsonian
- Hans Moritz Gunther, MIT
- Herman Marshall, MIT
- Kaisey Mandel, Cambridge University
- Saku Vrtilek, Center for Astrophysics | Harvard & Smithsonian
- Vinay Kashyap, Center for Astrophysics | Harvard & Smithsonian

## Current Students (Undergrad/Masters/PhD)

Antoine Meyer (Impereial)
Di Zhang (UC Irvine)
Josh Ingram (New College of Florida)
Jue Wang (UC Davis)
Karthik Reddy (UMaryland)
Maximilian Autenrieth (Imperial)
Youwei Yan (Simon Fraser)
Xiangyu Zhang (Minnesota)

**http://hea-www.harvard.edu/AstroStat/people.html**

## Former Faculty / Post Docs / Researchers / Associates

Adam Roy, ReVision Optics
Alanna Connors
Alex Young, NASA-GSFC
Brandon Kelly, Goodyear
Eric Kolaczyk, Boston University
Hyunsook Lee, Lam Research
James Chiang, Stanford
Jason Kramer, UC Irvine
Jeremy Drake, Center for Astrophysics | Harvard & Smithsonian
Margarita Karovska, Center for Astrophysics | Harvard & Smithsonian
Pavlos Protopapas, Harvard
Peter Freeman, Carnegie Mellon
Rima Izem
Taeyoung Park, Yonsei
Yue Wu, University of Washington

## Former Students

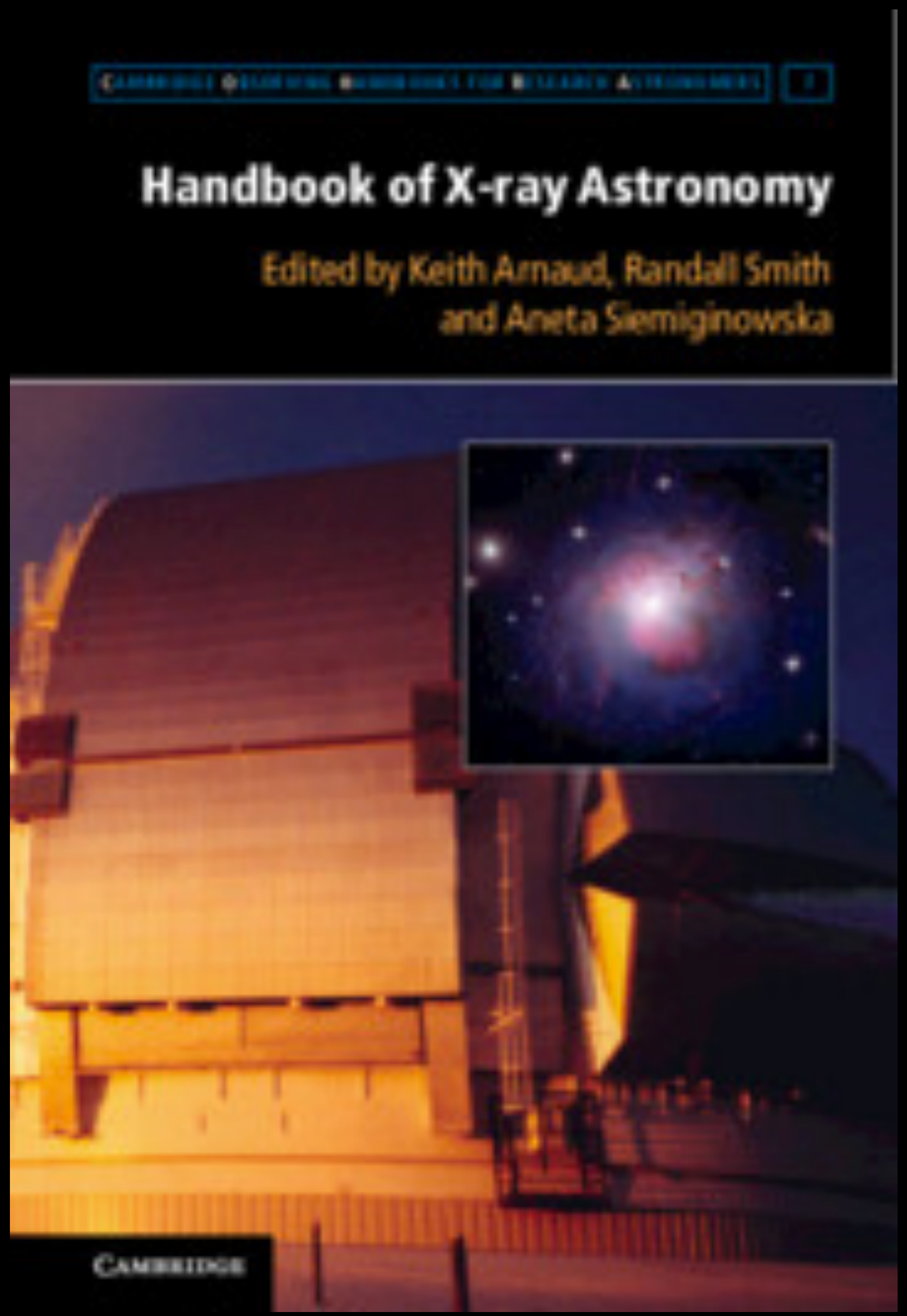
CJ Zijin Shen (Harvard 2000), Head of Options Trading, Virtu Financial, NY
Chris Hans (Harvard AB 2001), Associate Professor, Ohio State University
Rostislav S. Protassov (Harvard 2003) Director, Citigroup, NY
David Esch (Harvard 2003), Director of Research, New Frontier Advisors, MA
Hosung Kang (Harvard 2005), Quantitative Researcher, quantPort, UK
Yaming Yu (Harvard 2005), Associate Professor, University of California, Irvine
Taeyoung Park (Harvard 2006), Professor, Yonsei University, Korea
Jingchen Liu (Harvard 2008), Associate Professor, Columbia University
Alan Burton Lenarcic (Harvard 2009), Manhattan Securities and Exchange Commission
Paul David Baines (Harvard 2010), Checkr, CA
Xianchao Xie (Harvard 2011), Two Sigma Investments
Li Zhu (Harvard 2012), Marshall Wace, UK
Victoria Liublinska (Harvard 2013), Data Scientist, Harvard Business School
Nathan Stein (Harvard 2013), Spotify, NY
Alex Blocker (Harvard 2013), Foresite Labs, CA
Jin Xu (UC Irvine 2014), Adobe, CA
Shandong Zhao (UC Irvine 2014), Google, CA
Raymond Wong (UC Davis 2014), Assistant Professor, Texas A&M University
Irina Udaltsova (UC Davis 2014), Pratt & Whitney, CT
David Stenning (UC Irvine 2015), Assistant Professor, Simon Fraser University
Lazhi Wang (Harvard 2015), Two Sigma, NY
David Jones (Harvard 2016), Assistant Professor, Texas A&M
Hyungsuk Tak (Harvard 2016), Assistant Professor, Penn State University
Xiyun Jiao (Imperial 2016), Postdoctoral Fellow, University College London
Yang Chen (Harvard 2017), Assistant Professor, University of Michigan
Xufei Wang (Harvard 2017), Two Sigma, NY
Minjie Fan (UC Davis 2017), Quantitative Analyst at Google Research, CA
Qi Gao (UC Davis 2017), Data Scientist at Stitch Fix, CA
Vasileios Stampoulis (Imperial 2018), Algorithmic Trader, UBS
Sara Algeri (Imperial 2018), Assistant Professor, University of Minnesota
Chunzhe Zhang (UC Davis 2018), Data Scientist at LinkedIn, CA
Shijing Si (Imperial 2018), Duke University, NC
Luis Campos (Harvard 2019), Etsy, NY
Shihao Yang (Harvard 2019)
Katy McKeough (Harvard 2020), Red Sox, MA
Andrea Sottosanti (Imperial 2020), Padua, Italy
Chang Goh (Imperial 2020)
Xixi Yu (Imperial 2021), Citibank, UK
Cong Xu (UC Davis 2021) , Columbus, OH

# Astro-Stat Language

## International CHASC Astro-Statistics Collaboration

pecific interest to astronomers. For detailed descriptions and reports of C-BAS/ICHASC activities, see www2.imperial.a

Software | Activities | Bibliography | Astro jargon | Stat jargon | People | Mailing-List | Internal




We briefly describe here a number of terms normally used by statisticians, with translations where appropriate into the terminology used in X-ray astrophysics; this information is taken from the CHASC jargon page at http://hea-www.harvard.edu/AstroStat/statjargon.html.

**Background marginalization** is integration of a background probability over uninteresting parameters.

**Bias** is a systematic difference between an estimated and a true value of a parameter.

**Biased sample** is a sample of objects selected from a population such that some objects are more likely to be included than others.

**Bootstrap** is a method for estimating parameter variance or other properties using an approximation to a distribution created by resampling the observed data themselves.

**Cash statistic** is a formulation of Poisson likelihood for a parametric model in X-ray astronomy.

**Chi-square statistic** is a statistic applied in X-ray astronomy which provides a measure of the goodness-of-fit. The name comes from the $\chi^2$ distribution, however many of the "chi-square statistic" expressions do not follow the $\chi^2$ distribution. Here are the most common expressions used:

- **model variance** $\chi^2 = (D - M)^2/M$
- **data variance** $\chi^2 = (D - M)^2/D$
- **iterative Primini approximation** $\chi_i^2 = (D - M_i)^2/M_{i-1}$, where $i$ is the iteration fitting step.

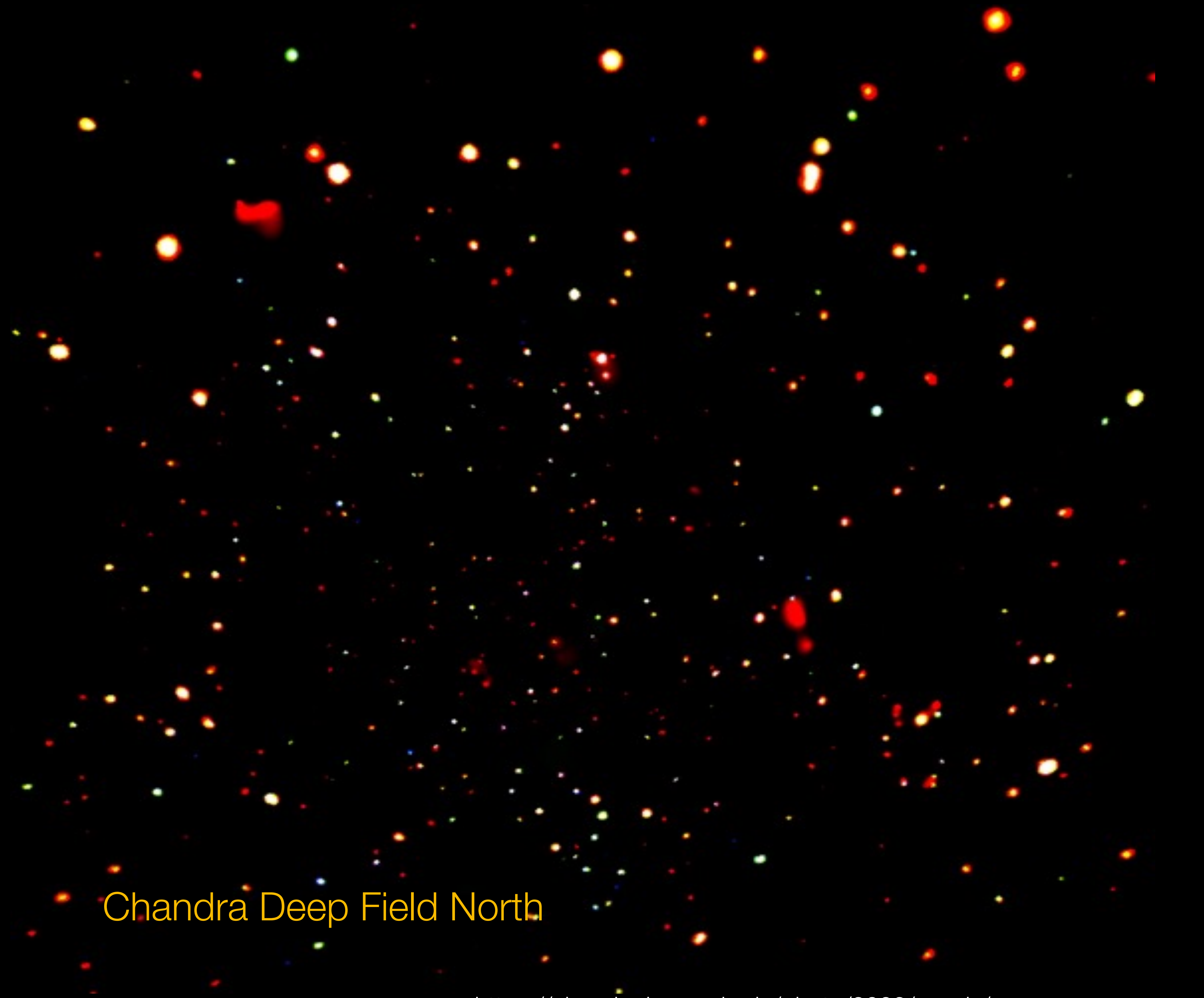
**Conditional distribution** (or probability density) is the probability distri-

# X-ray data

- Collecting X-ray data means counting arriving photons (Poisson counts) - different from optical data

- For each photon location on the sky, arrival time and energy are recorded (x,y,t,E) - events

- X-ray observations take a long time - a short observation with Chandra last ~10 ksec (~3 hours) while typical observations take a day or more.  Chandra Deep Field observations took about 23 days.

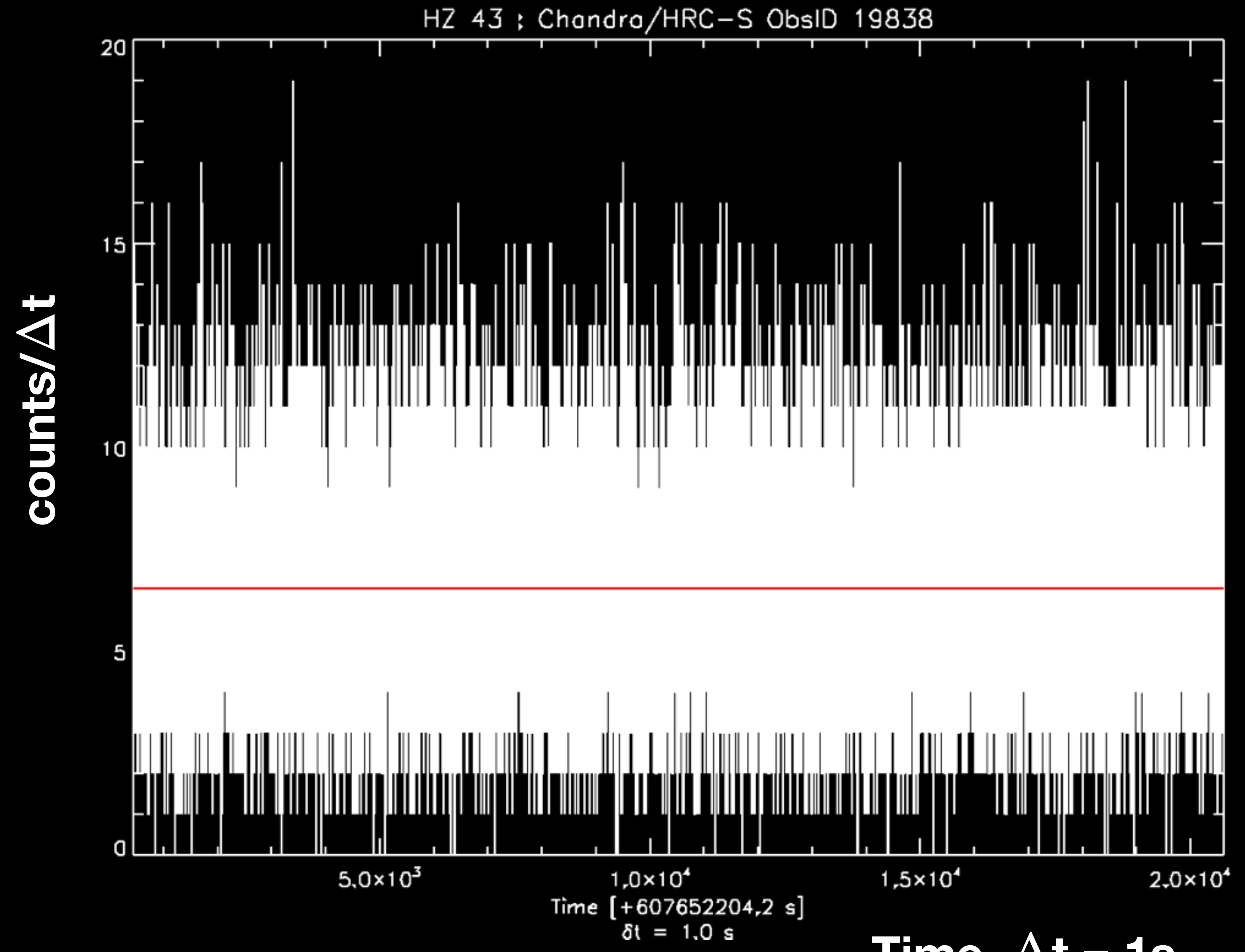



https://chandra.harvard.edu/photo/2003/goods/

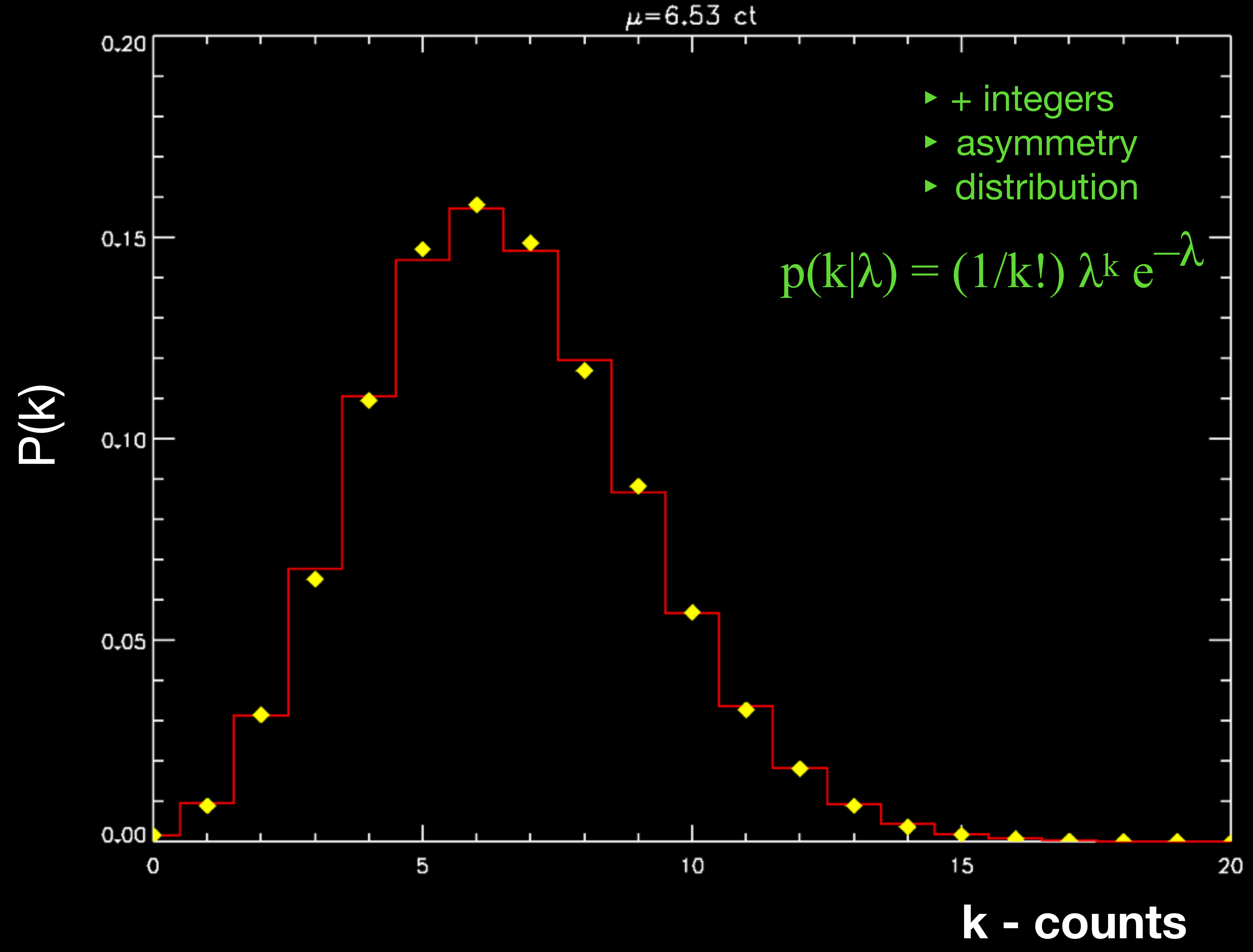
The faintest sources - one X-ray photon every 4 days!

# Poisson Counts

Light curve of a steady source HZ 43 binned in 1 sec time bins
Notice asymmetry scatter around the mean



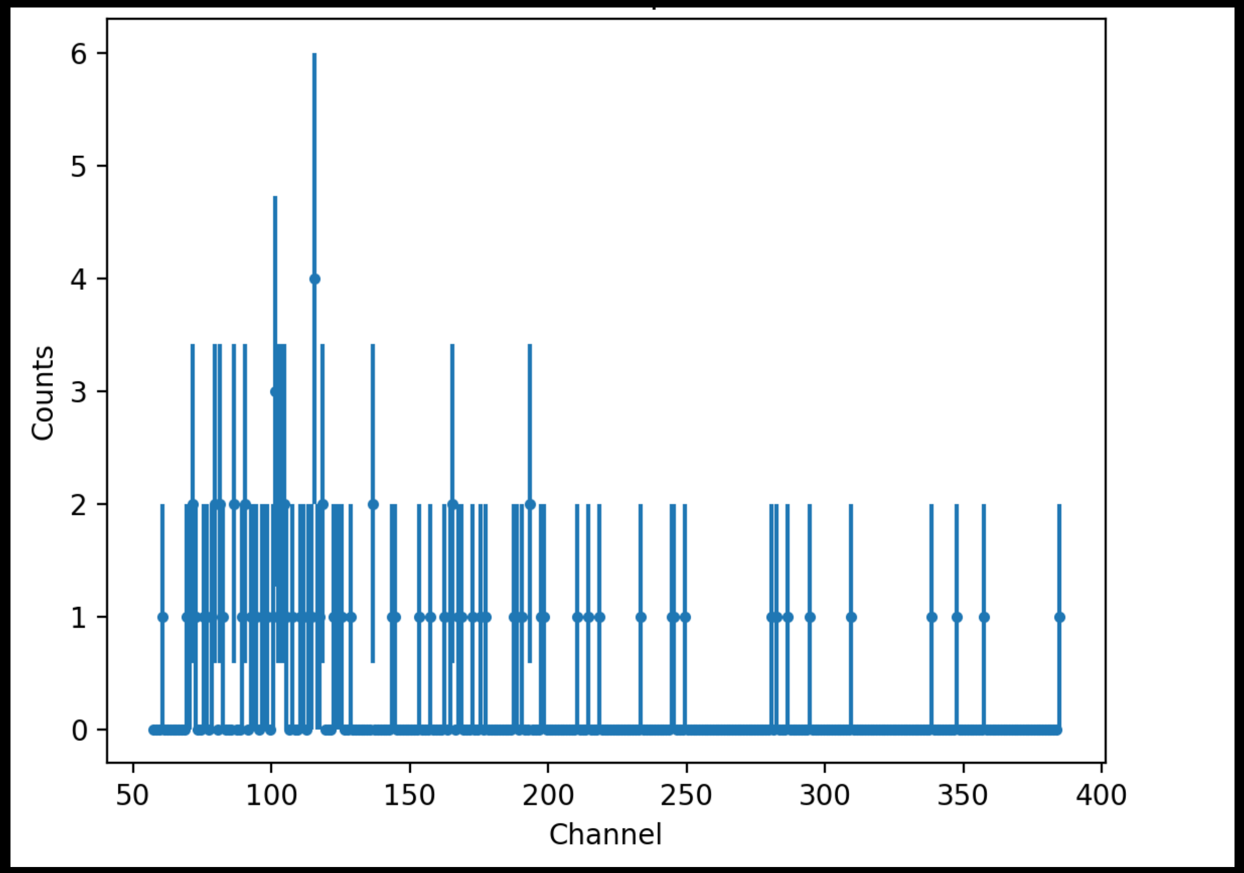

Distribution of counts (k) in a light curve with
Poisson rate $\lambda = 6.53$ ct/s




$$p(k|\lambda) = (1/k!)\, \lambda^k e^{-\lambda}$$

# Adventure 1
# Issues in Modeling low counts X-ray Spectra

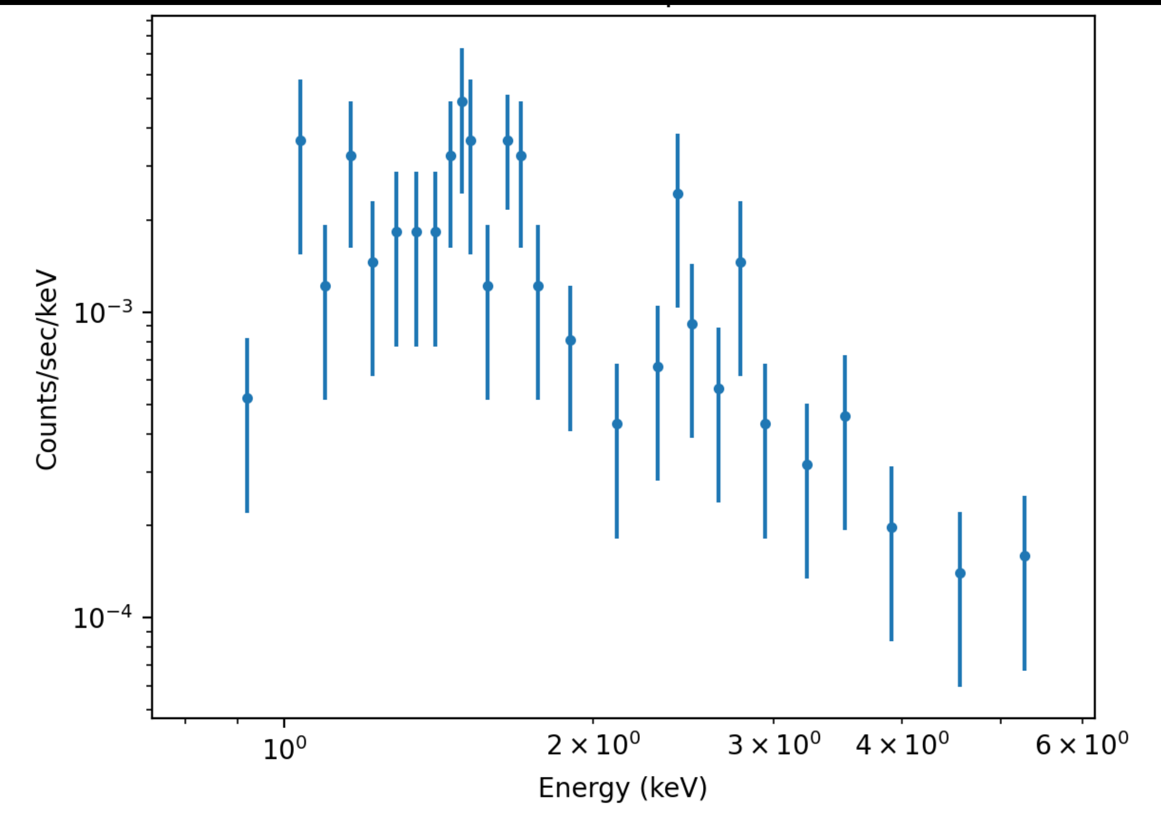
- Spectral fitting:
  - Includes instrument response directly -> calibration impact on the results  Counts(i)= Int [arf (E) * rmf( E,i ) * Model( p, E) dE]
  - Non-linear astrophysical models, computer generated models
  - Appropriate fit statistics, no binning/grouping data, no background subtraction
  - Modification to the fit statistics (weighted chi2) still not good for low number of counts, e.g. Gehrels (1986) $\sigma_X \approx \sqrt{X + 0.75} + 1,$
  - Formulations for the Poisson likelihood - Cash (1979), cstat, wstat
- Why important?
  - **bias,** negative data if subtracting background or false spectral features, loss of information with binning, optimization with high number of parameters (e.g. finding the best-fit)




channel

keV

convert matrix (RMF)




log( Energy ) [keV]

# Fit Statistics and Bias

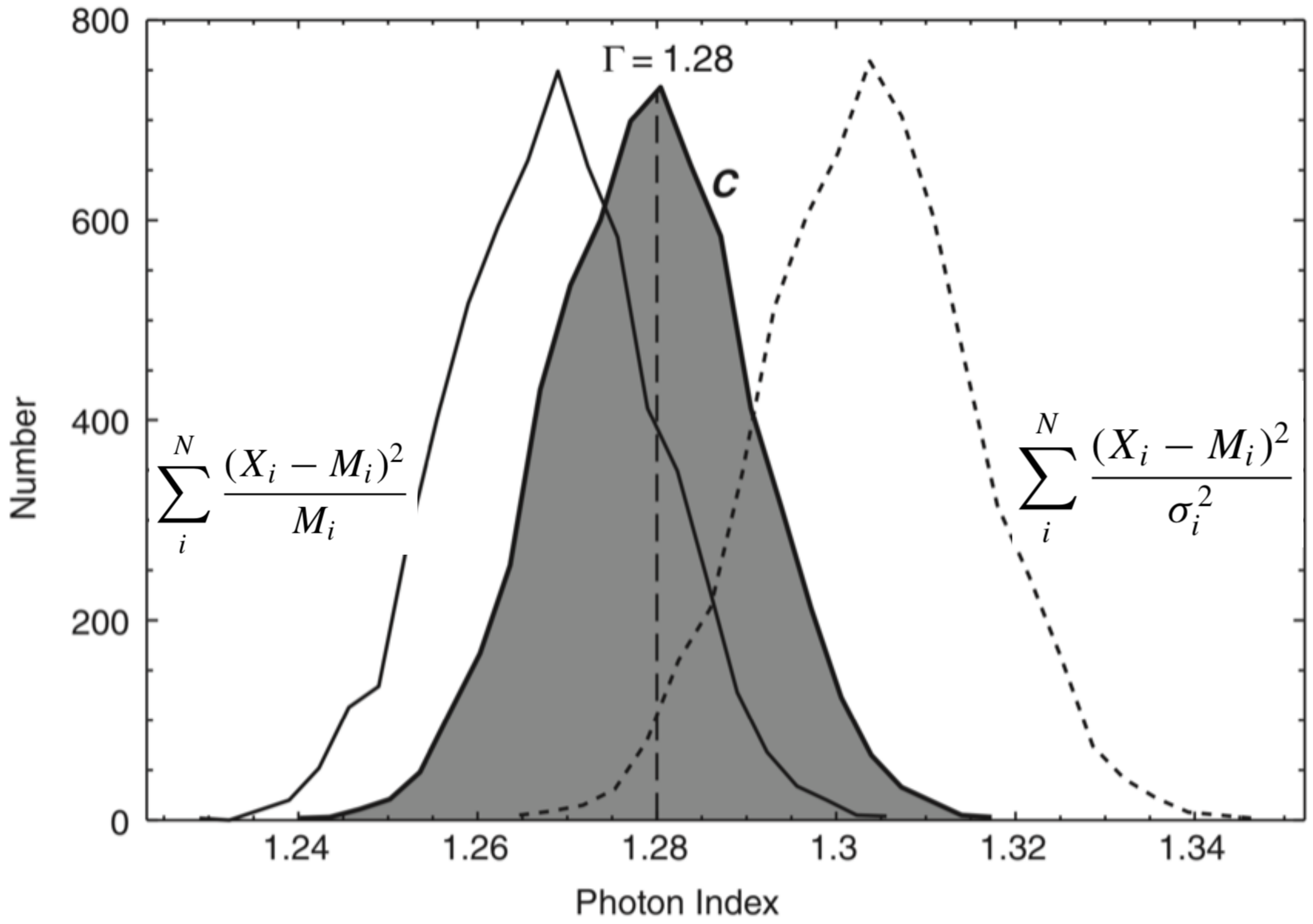



Fig. 7.3 Distributions of a photon index parameter $\gamma$ obtained by fitting simulated X-ray spectra with 60 000 counts and using the three different statistics: $S^2_{\rm Pearson}$, $S^2$ and $C$ (i.e. the Poisson likelihood) statistics. The true value of the simulated photon index is marked with a dashed line and it was set at $\gamma = 1.28$

**Simulations:**

Distribution of photon index parameter obtained by fitting simulated X-ray spectra using different fit statistics:
Gaussian likelihood ($\chi^2$ data, $\chi^2$ model)
and Poisson likelihood (Cash).
The assumed photon index =1.28 is marked.

$X_i$ - data
$M_i$ - model

Chapter 7 on Statistics

Handbook of X-ray Astronomy (Arnaud, Smith, Siemiginowska): https://doi.org/10.1017/CBO9781139034234.008

# Adventure 1
# Bayesian Model for Low Counts X-ray Spectra

Poisson likelihood and application of Markov Chain Monte Carlo (MCMC) and Gibbs Sampler

The Astrophysical Journal, 548:224–243, 2001 February 10

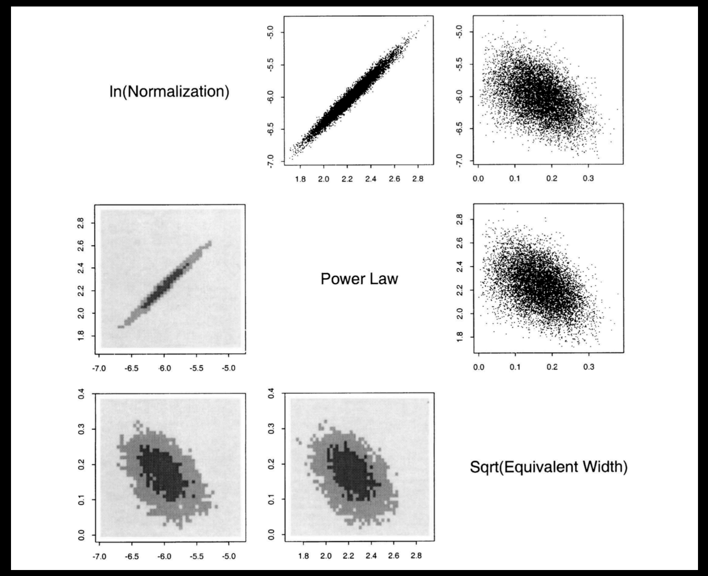

**Van Dyk et al 2001**

ANALYSIS OF ENERGY SPECTRA WITH LOW PHOTON COUNTS VIA BAYESIAN POSTERIOR SIMULATION

David A. van Dyk
Department of Statistics, Harvard University, 1 Oxford Street, Cambridge, MA 02138; vandyk@stat.harvard.edu
Alanna Connors[1]
Department of Astronomy, Whitin Observatory, Wellesley College, Wellesley, MA 02481; connors@frances.astro.wellesley.edu
and
Vinay L. Kashyap and Aneta Siemiginowska
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138; kashyap@head-cfa.harvard.edu, aneta@head-cfa.harvard.edu
*Received 2000 March 9; accepted 2000 August 25*

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

posterior distribution — likelihood — prior on model parameters




**Bayesian framework:**

- probability of the model given the data
- takes into account all information (data, instruments, model etc.)
- no binning of data
- background as part of the statistical model
- non-biased results
- full information on the posterior distribution (probability of the model given the observation)
- can take into account calibration uncertainties (see Lee et al 2011, Xu et al 2014, Marshall et al 2021)

pyBLoXCS

Note:
Included in Sherpa https://cxc.harvard.edu/sherpa/
Applied in processing of the Chandra Source Catalog (Evans et al 2010)

# Adventure 1
# Bayesian Model for Low Counts X-ray Spectra

## Systematic (Calibration) Uncertainties

**non-linear uncertainties - non-additive**

posterior distribution | likelihood | model prior

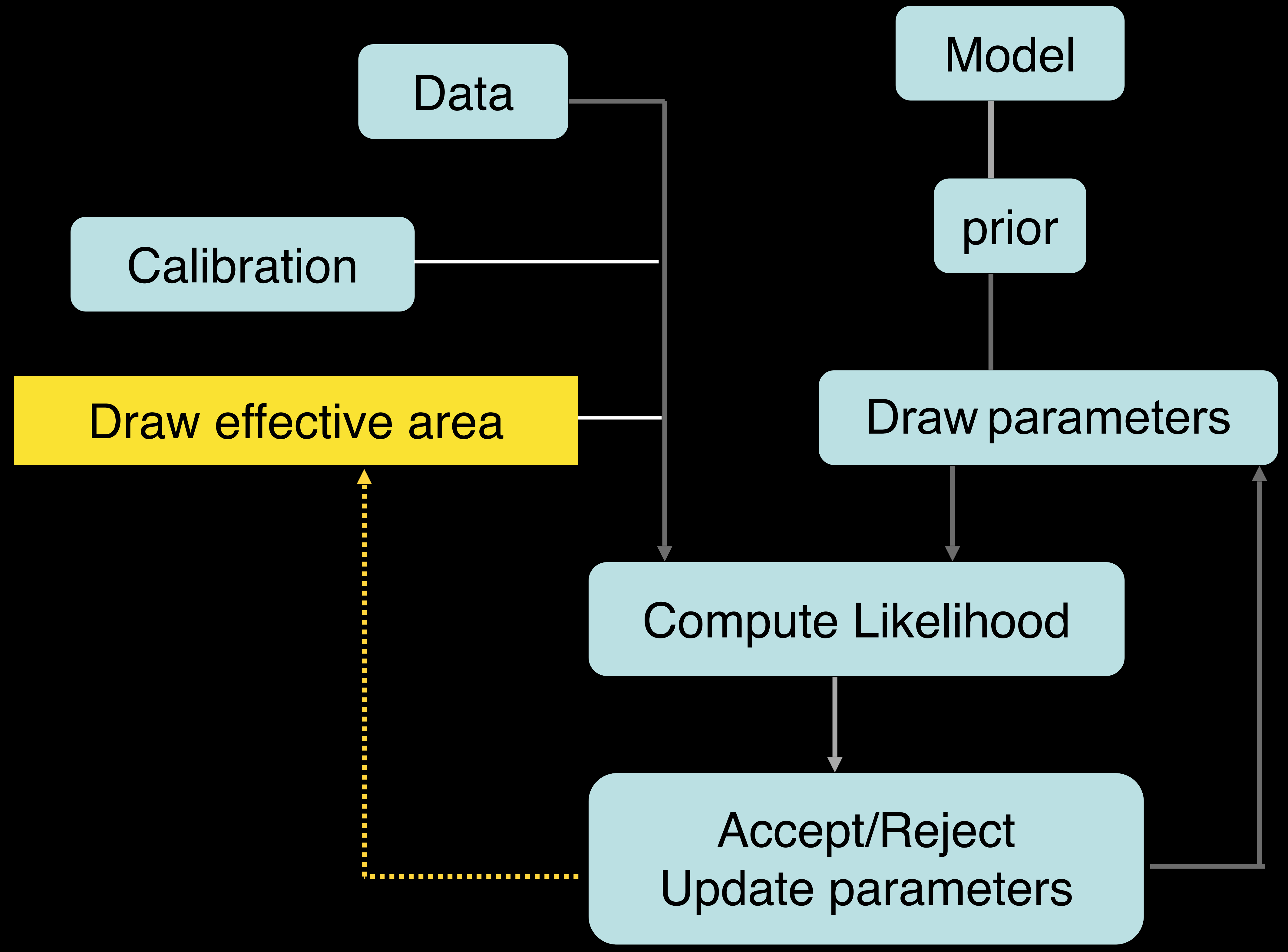
$$p_{\mathrm{std}}(\theta \mid Y, A_0^*) \propto \mathcal{L}(Y \mid \theta, A_0^*)\ p(\theta).$$

Standard
Fixed ARF

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

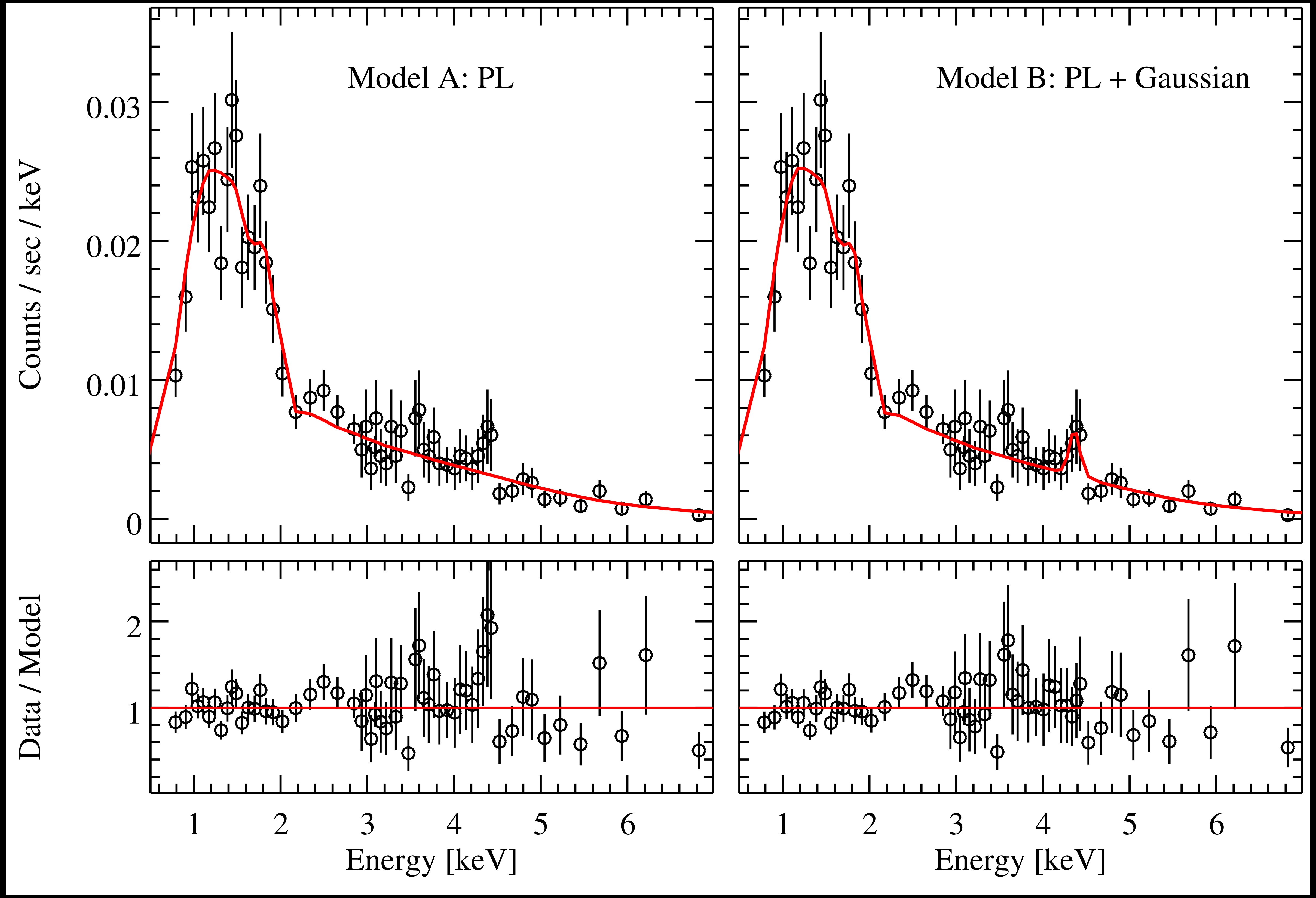
Full Bayes
ARF prior

Lee+ 2010, Xu+ 2014



# Adventure 2
# Hypothesis testing and Model Selection

- Spectral features - line detections
- Additional model components




Siemiginowska+ 2016

# Adventure 2
# Hypothesis testing and Model Selection

THE ASTROPHYSICAL JOURNAL, 571:545–559, 2002 May 20


Protassov et. al.2002

## STATISTICS, HANDLE WITH CARE: DETECTING MULTIPLE MODEL COMPONENTS WITH THE LIKELIHOOD RATIO TEST

ROSTISLAV PROTASSOV AND DAVID A. VAN DYK
Department of Statistics, Harvard University, 1 Oxford Street, Cambridge, MA 02138; protasso@stat.harvard.edu, vandyk@stat.harvard.edu

ALANNA CONNORS
Eureka Scientific, 2452 Delmer Street, Suite 100, Oakland, CA 94602-3017; connors@frances.astro.wellesley.edu

AND

VINAY L. KASHYAP AND ANETA SIEMIGINOWSKA
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138;
kashyap@head-cfa.harvard.edu, aneta@head-cfa.harvard.edu
*Received 2001 June 1; accepted 2002 January 25*

## ABSTRACT

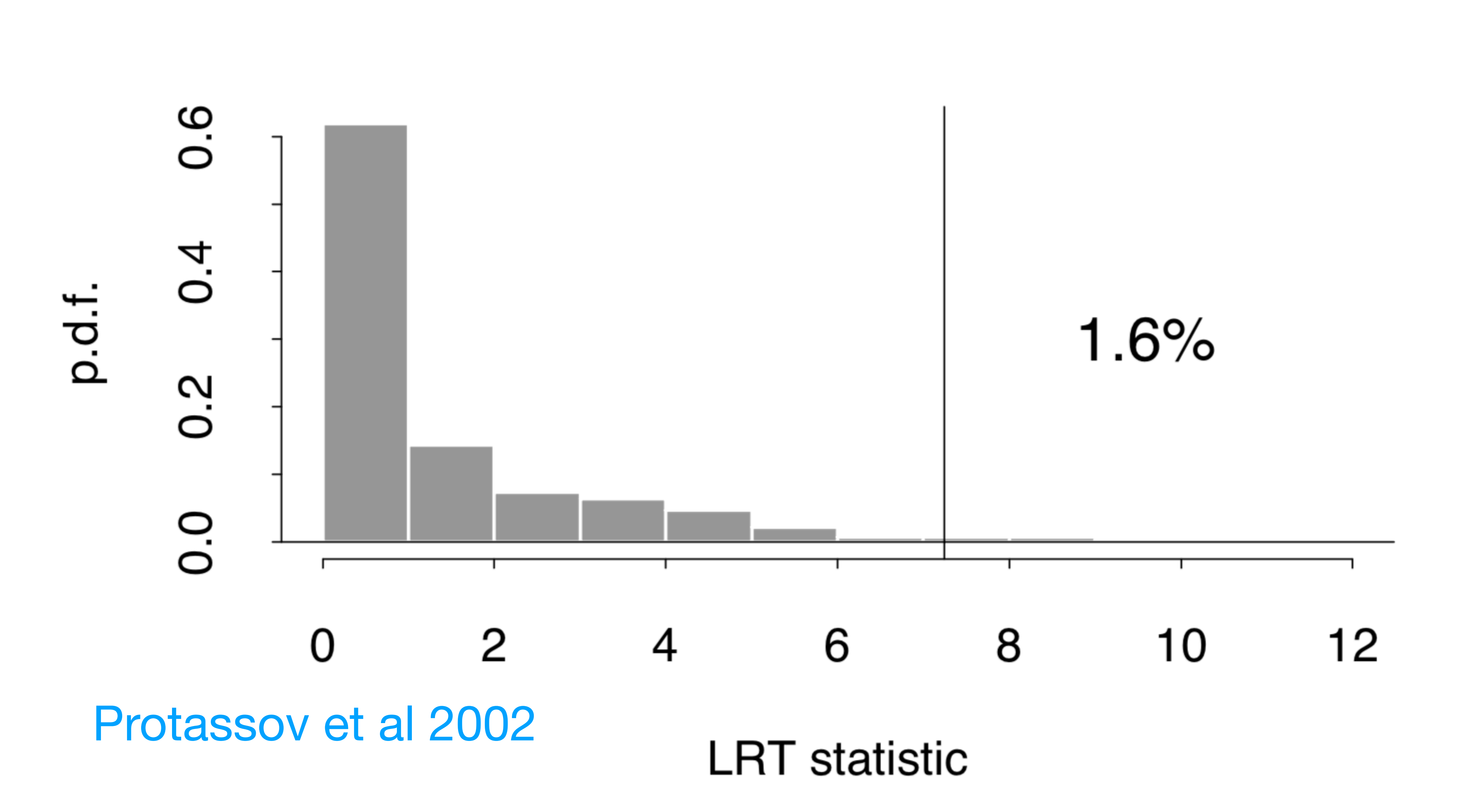
The likelihood ratio test (LRT) and the related *F*-test, popularized in astrophysics by Eadie and coworkers in 1971, Bevington in 1969, Lampton, Margon, & Bowyer, in 1976, Cash in 1979, and Avni in 1978, do not (even asymptotically) adhere to their nominal $\chi^2$ and *F*-distributions in many statistical tests common in astrophysics, thereby casting many marginal line or source detections and nondetections into doubt. Although the above authors illustrate the many legitimate uses of these statistics, in some important cases it can be impossible to compute the correct false positive rate. For example, it has become common practice to use the LRT or the *F*-test to detect a line in a spectral model or a source above background despite the lack of certain required regularity conditions. (These applications were not originally suggested by Cash or by Bevington.) In these and other settings that involve testing a hypothesis that is on the boundary of the parameter space, *contrary to common practice, the nominal $\chi^2$ distribution for the LRT or the F-distribution for the F-test should not be used*. In this paper, we characterize an important class of problems in which the LRT and the *F*-test fail and illustrate this nonstandard behavior. We briefly sketch several possible acceptable alternatives, focusing on Bayesian posterior predictive probability values. We present this method in some detail since it is a simple, robust, and intuitive approach. This alternative method is illustrated using the gamma-ray burst of

- Testing for a presence of an emission line
- Standard LRT and F-test does not apply
- Simulations needed to calibrate test statistics
- Posterior Predictive p-values

  Protassov+2002 Park+2008

# Adventure 2
# Hypothesis testing and Model Selection

**Bayesian Posterior Predictive P-values**




Simulations:

1 Simulate L data sets under null model (H0) and compute the test statistic for each of the L data sets fit with null (H0) and the complex model (H1)

2 A histogram of the simulated test statistics approximates the sampling distribution of the test statistic.

3 Compute the p-value for the observed value of the test statistics

Computing the p-value:

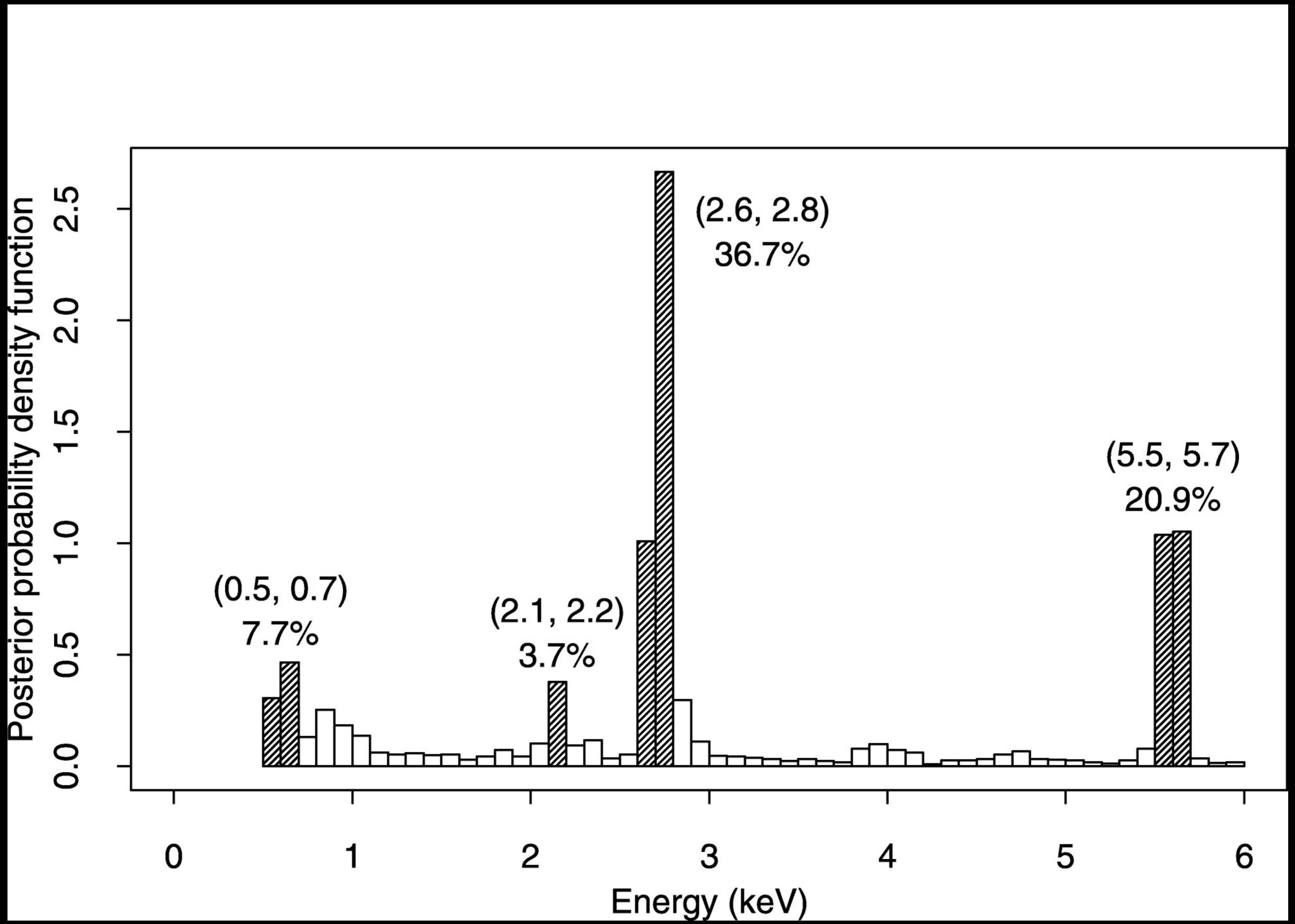
the proportion of simulated test statistics LRT values larger (more extreme) than the observed LRT

Note:
Included in Sherpa https://cxc.harvard.edu/sherpa/

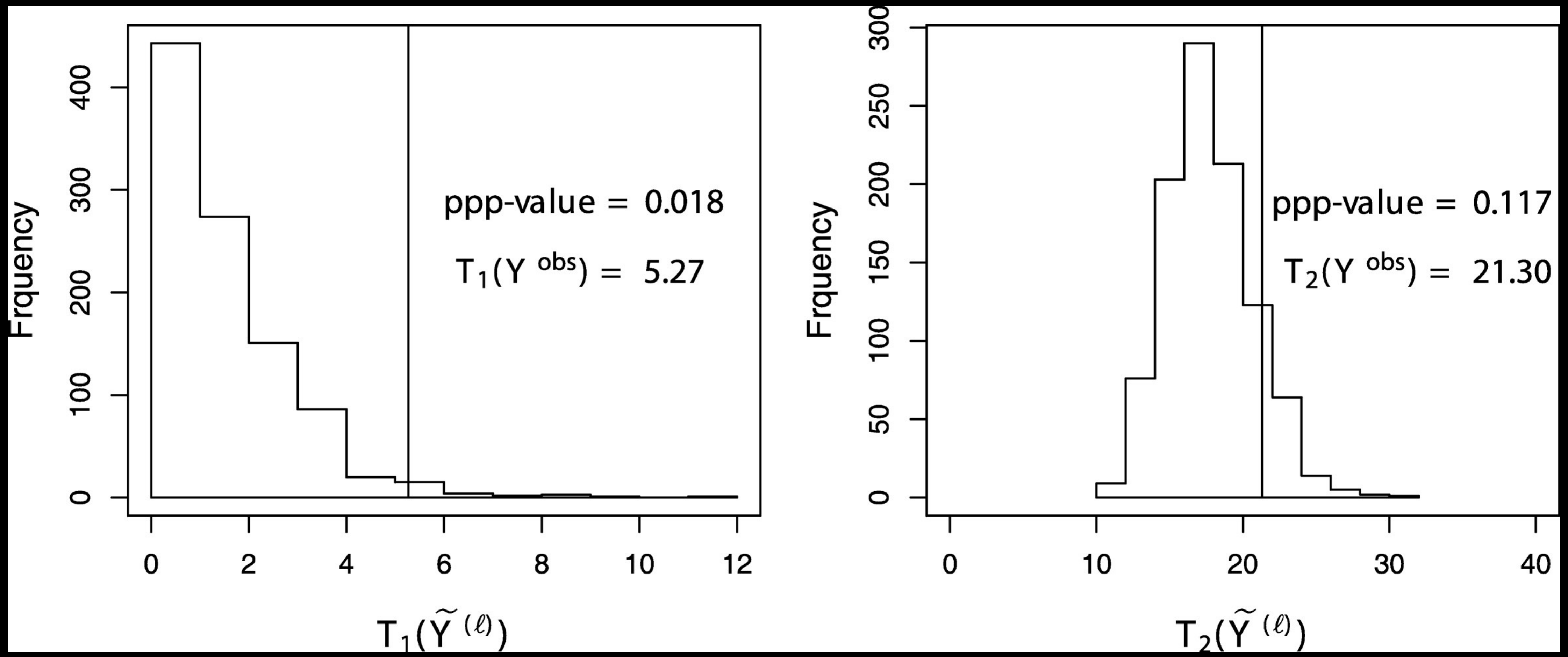
# Adventure 2
# Hypothesis testing and Model Selection

Searching for lines in low counts low resolution spectrum:
- line locations and intensity

**High Posterior Density (HPD) - most likely line locations**




**Evidence for the line**



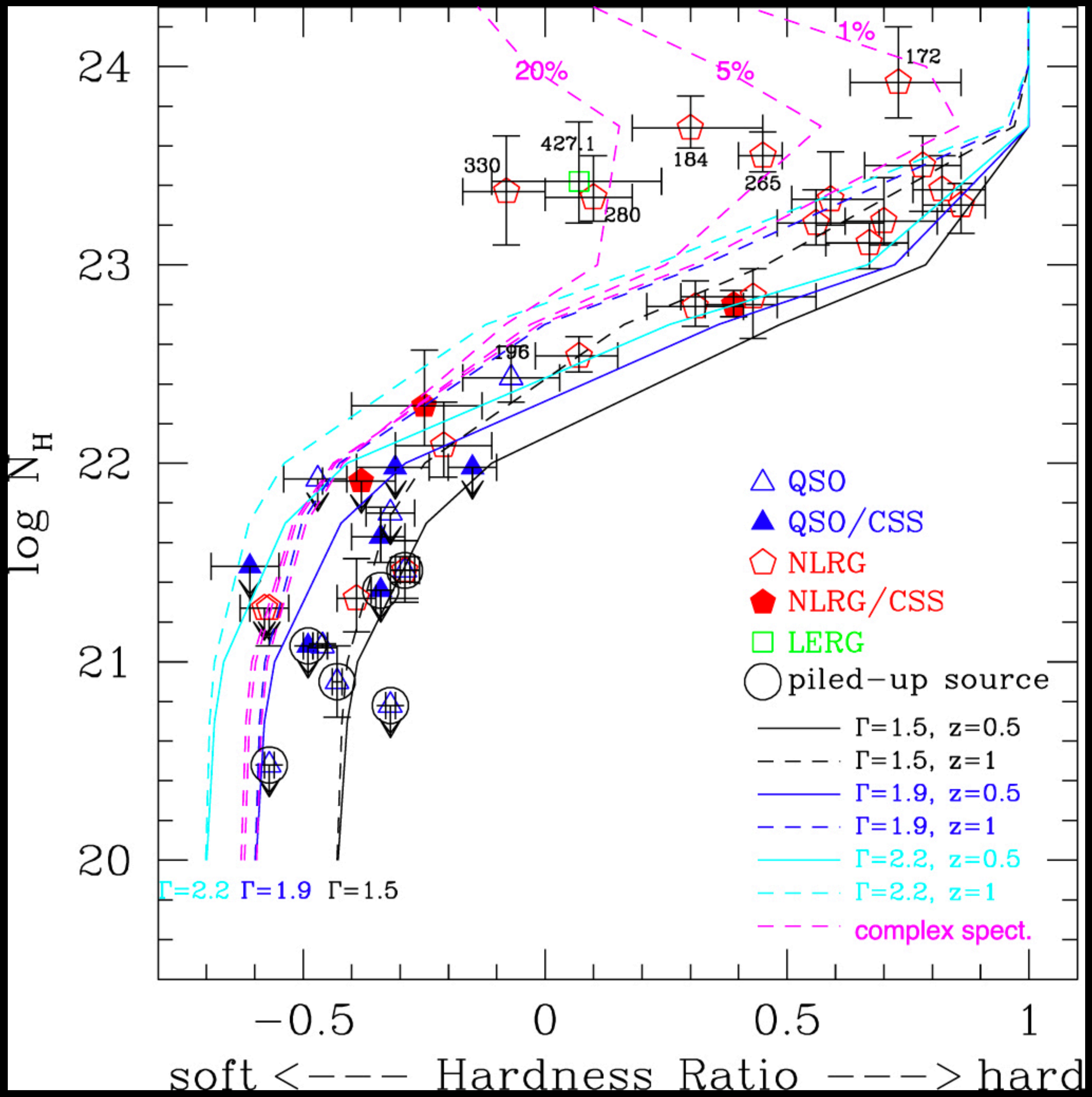

fixed line location at 2.84 keV

location unknown

Park et al 2008

# Adventure 3
## Analysis of Faint Sources

- Upper limits in the source detection? Kashyap et al 2010
- For faint sources - not enough counts for spectral modeling
- Hardness Ratio calculations
- What are the errors on the hardness ratios? BEHR - Park et al 2006




Kuraszkiewicz et al 2021

# Adventure 3
# Hardness Ratios

...YSICAL JOURNAL, 652:610–628, 2006 November 20


**Park et al 2006**

BAYESIAN ESTIMATION OF HARDNESS RATIOS: MODELING AND COMPUTATIONS

TAEYOUNG PARK,[1] VINAY L. KASHYAP,[2] ANETA SIEMIGINOWSKA,[2] DAVID A. VAN DYK,[3] ANDREAS ZEZAS,[2] CRAIG HEINKE,[4] AND BRADFORD J. WARGELIN[2]
Received 2005 December 24; accepted 2006 June 14

**BEHR**

$$\text{simple ratio,}\quad \mathcal{R} \equiv \frac{S}{H},$$
$$\text{color,}\quad C \equiv \log_{10}\left(\frac{S}{H}\right),$$
$$\text{fractional difference,}\quad \mathcal{HR} \equiv \frac{H-S}{H+S},$$

$$\mathcal{R} = \frac{S - B_S/r}{H - B_H/r},$$
$$C = \log_{10}\left(\frac{S - B_S/r}{H - B_H/r}\right),$$
$$\mathcal{HR} = \frac{(H - B_H/r) - (S - B_S/r)}{(H - B_H/r) + (S - B_S/r)}.$$

## Classical Approach

$$\sigma_{\mathcal{R}} = \frac{S - B_S/r}{H - B_H/r}\sqrt{\frac{\sigma_S^2 + \sigma_{B_S}^2/r^2}{(S - B_S/r)^2} + \frac{\sigma_H^2 + \sigma_{B_H}^2/r^2}{(H - B_H/r)^2}},$$
$$\sigma_C = \frac{1}{\ln(10)}\sqrt{\frac{\sigma_S^2 + \sigma_{B_S}^2/r^2}{(S - B_S/r)^2} + \frac{\sigma_H^2 + \sigma_{B_H}^2/r^2}{(H - B_H/r)^2}},$$
$$\sigma_{\mathcal{HR}} = 2\Big[(H - B_H/r)^2(\sigma_S^2 + \sigma_{B_S}^2/r^2) + (S - B_S/r)^2(\sigma_H^2 + \sigma_{B_H}^2/r^2)\Big]^{1/2} \times \Big[(H - B_H/r) + (S - B_S/r)\Big]^{-2},$$

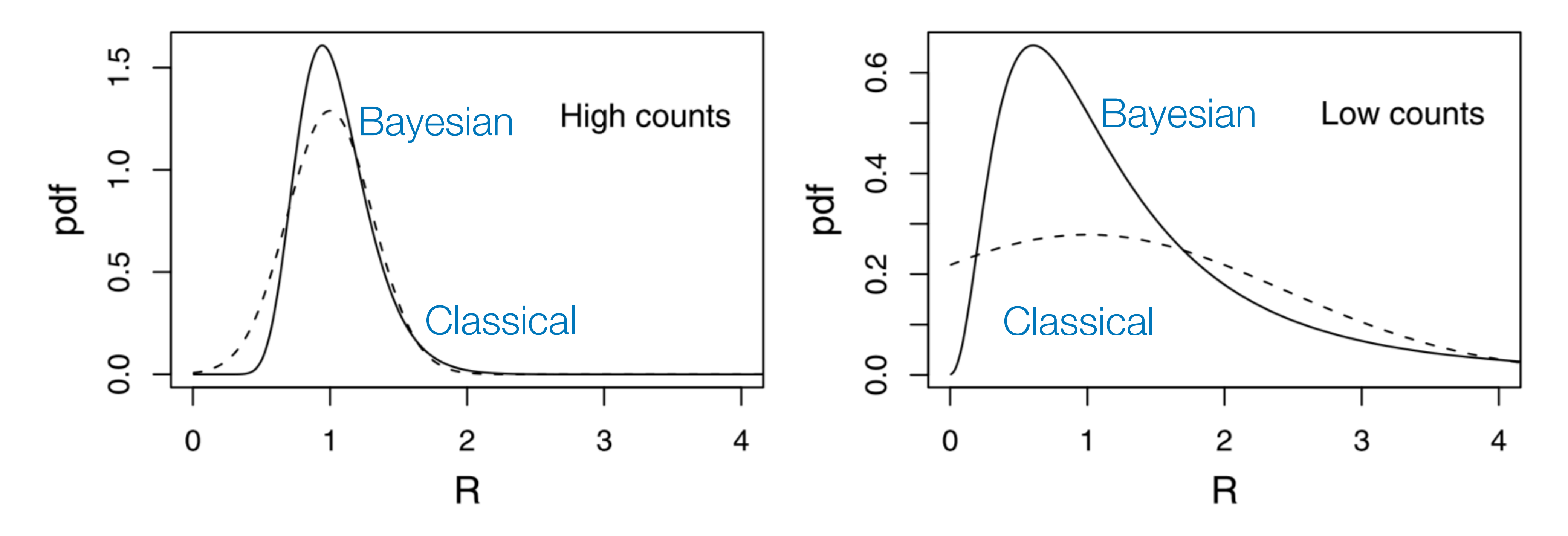
with Gehrels errors on measured counts

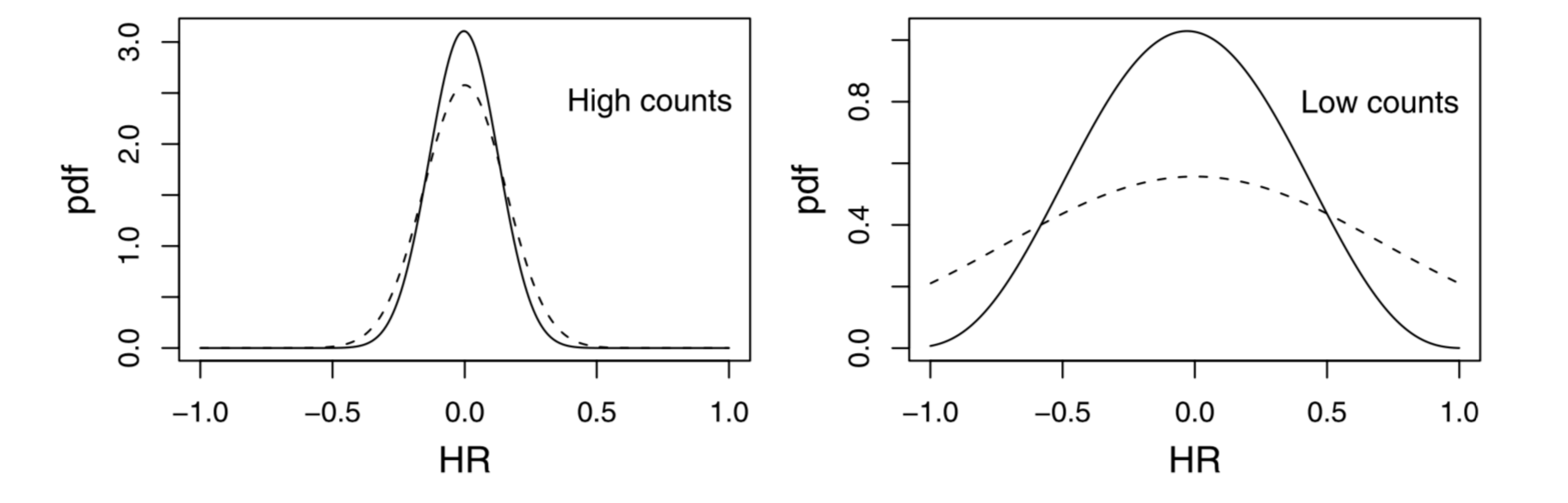
$$\sigma_X \approx \sqrt{X + 0.75} + 1,$$

**Issues with Classical Method:**

Background subtraction
R is positive - probability distribution skewed
HR is within [-1,+1]
C - asymmetric errors

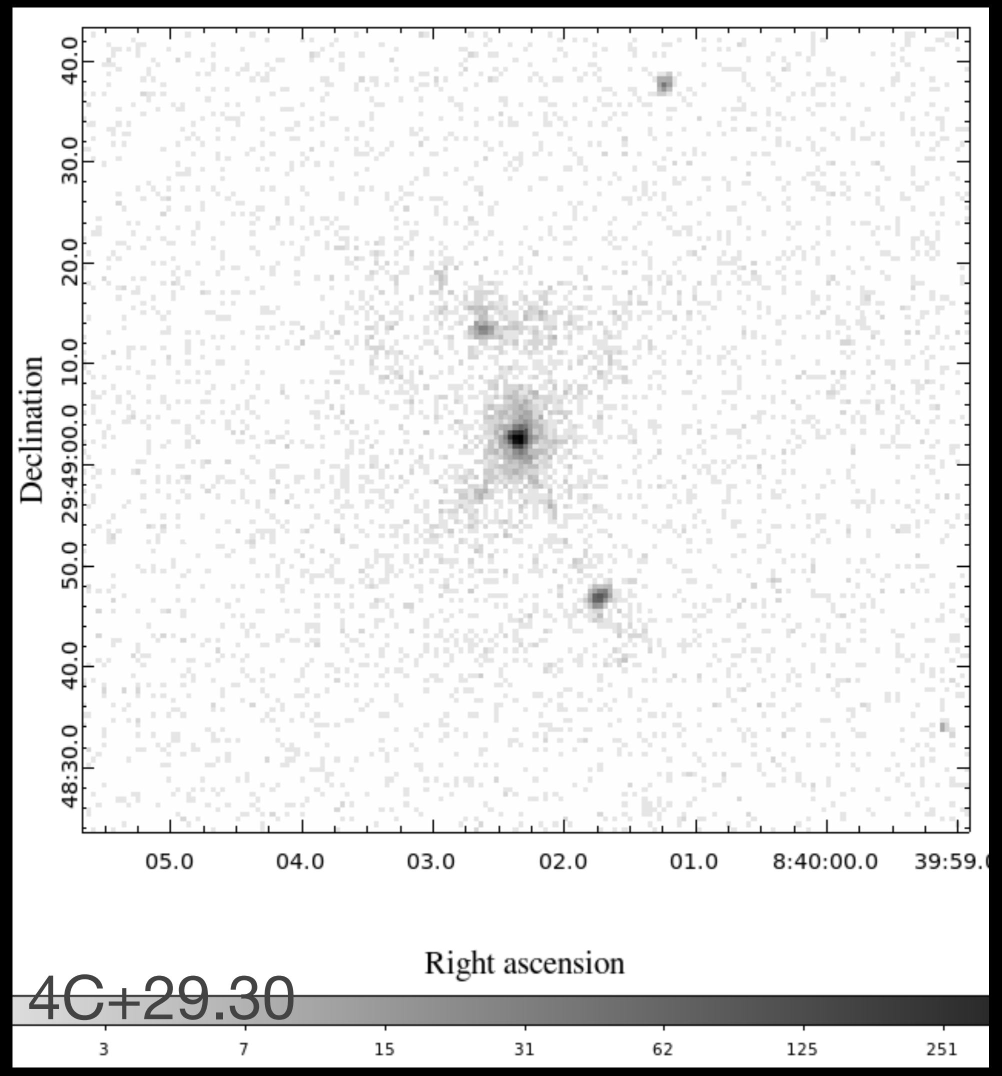
PARK ET AL.



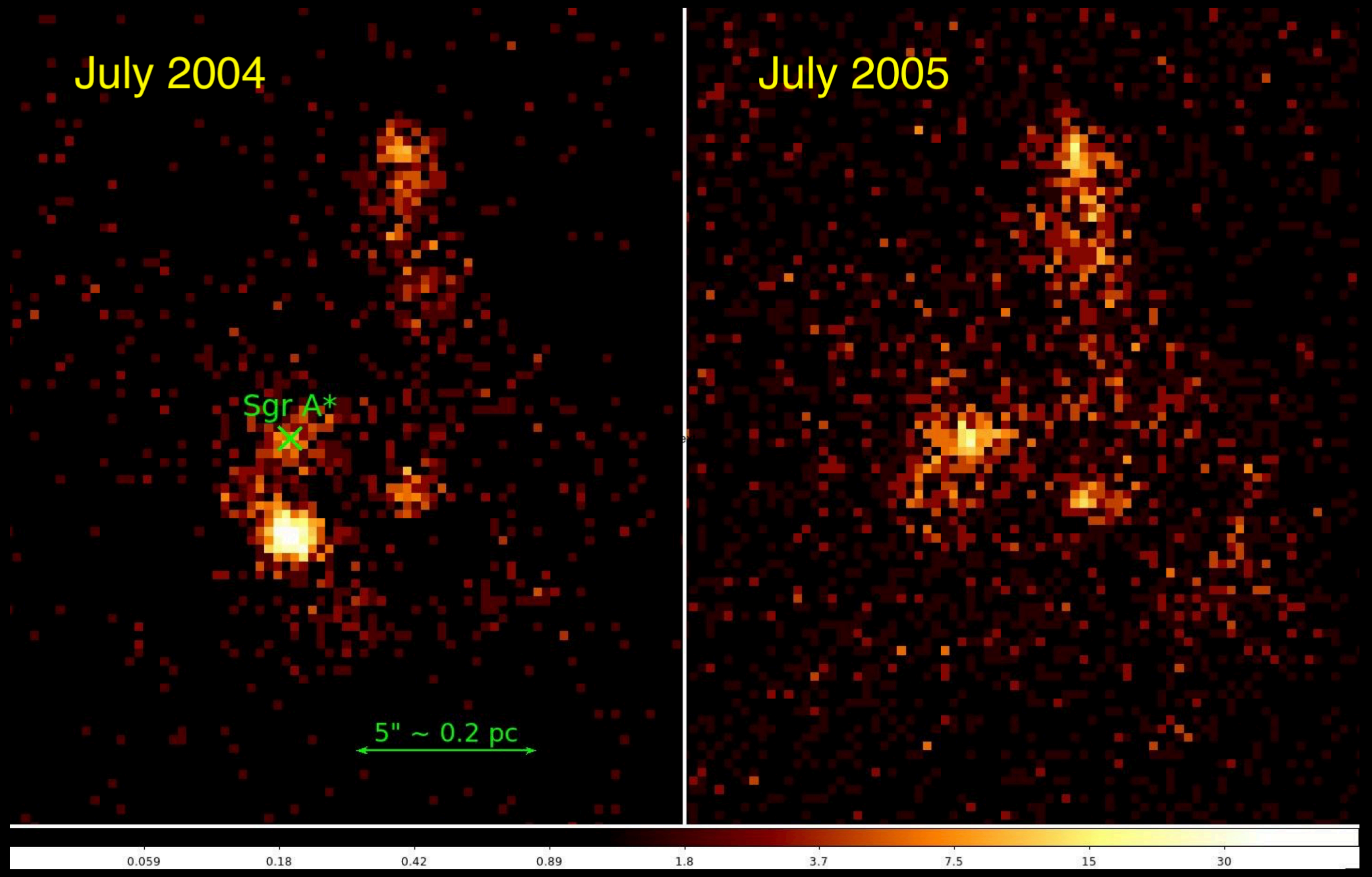




Note:
C-code available https://hea-www.harvard.edu/AstroStat/**BEHR**/
Applied in processing of the Chandra Source Catalog (Evans et al 2010, Primini et al. 2011)

# Adventure 4
# High Resolution X-ray Images




- Chandra takes the highest resolution X-ray images of the Universe
- Poisson counts - sparse images, with many empty pixels
- PSF variable across the images cannot be described in an analytical form, the PSF image is a simulation from the computer model of the Chandra mirrors with calibration measurements
- Some issues:
  - detection of features and upper limits
  - detecting and identifying low surface brightness structures
  - resolving source in crowded fields - overlapping sources, diffuse emission
  - finding source boundaries
  - PSF uncertainties

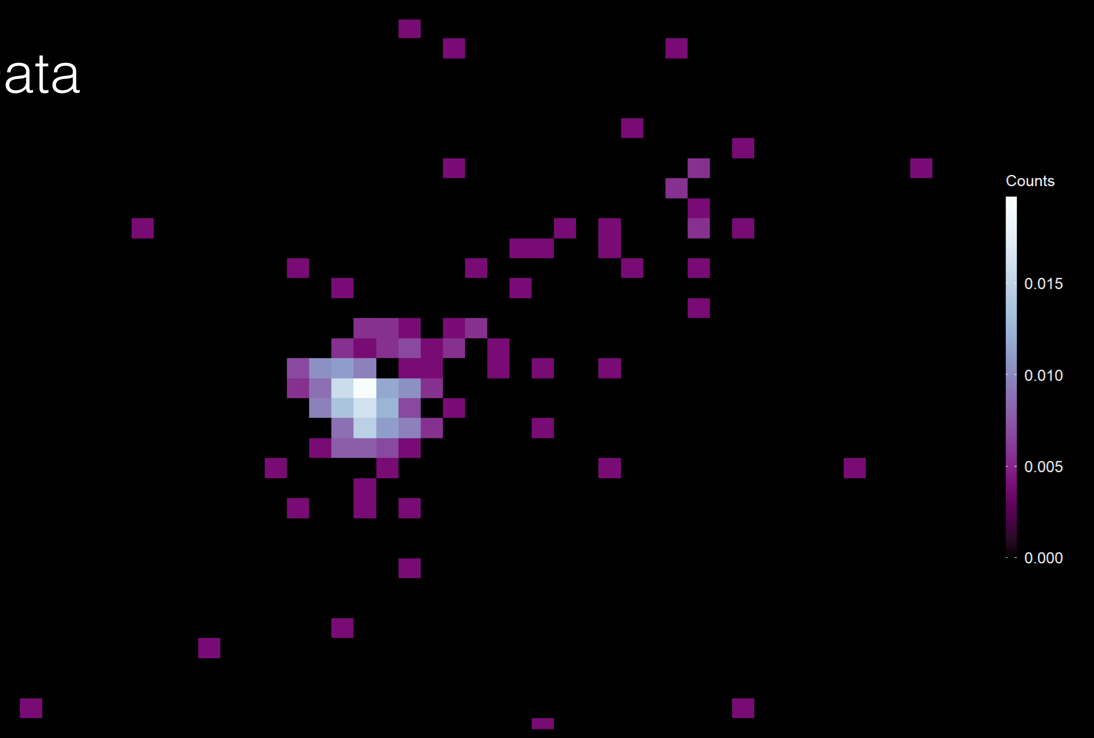



Chandra Image of the Galactic Center

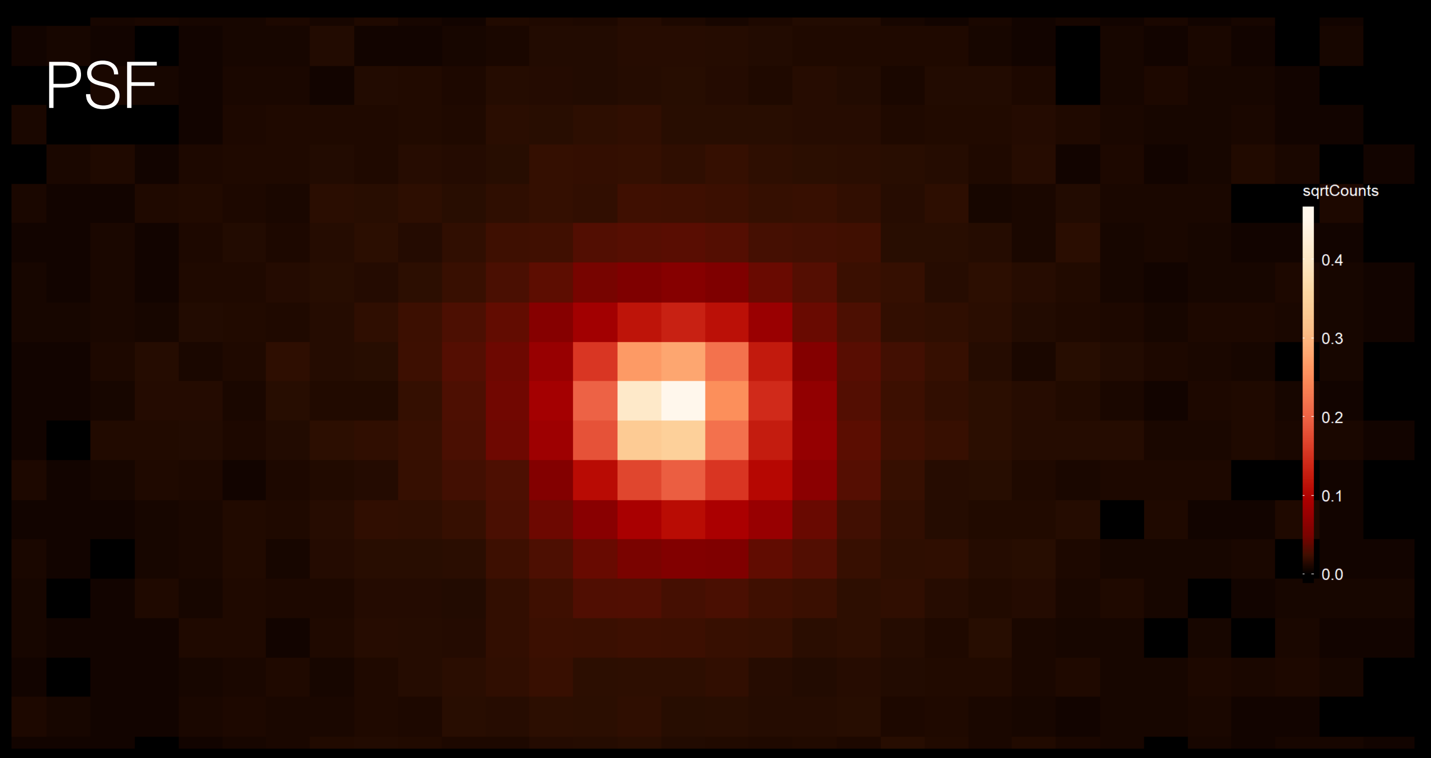
# Adventure 4
# High Resolution X-ray Images

LIRA - Low-Counts Image Reconstruction and Analysis
Bayesian Hierarchical Model

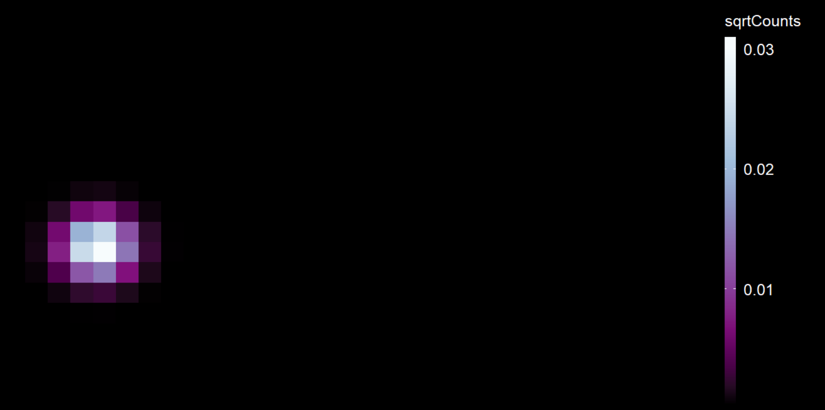
Chandra Data



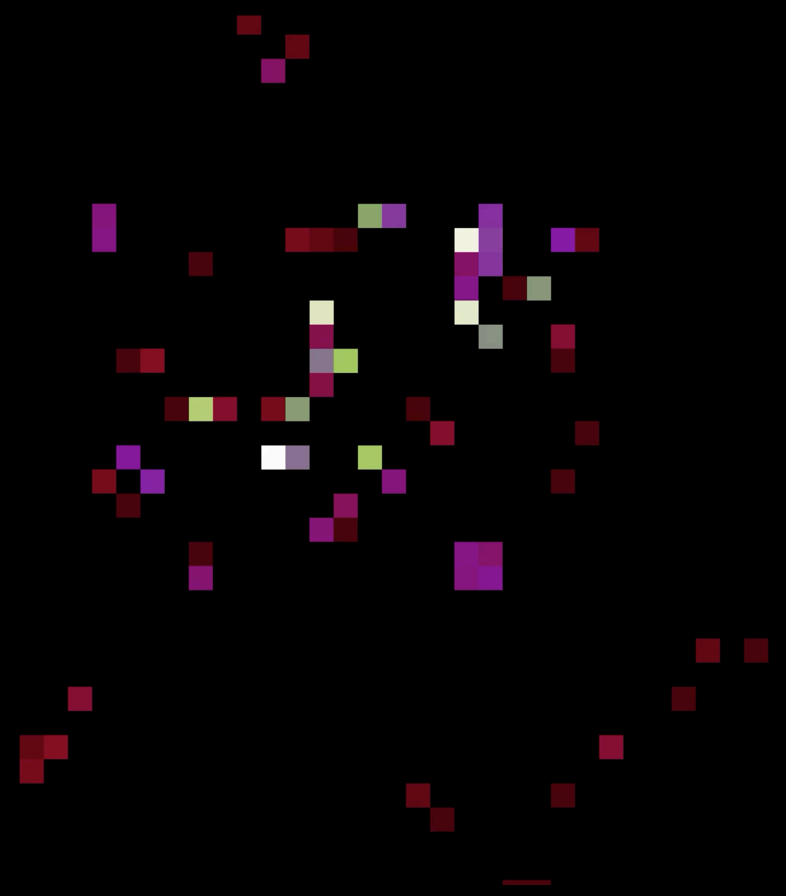

PSF



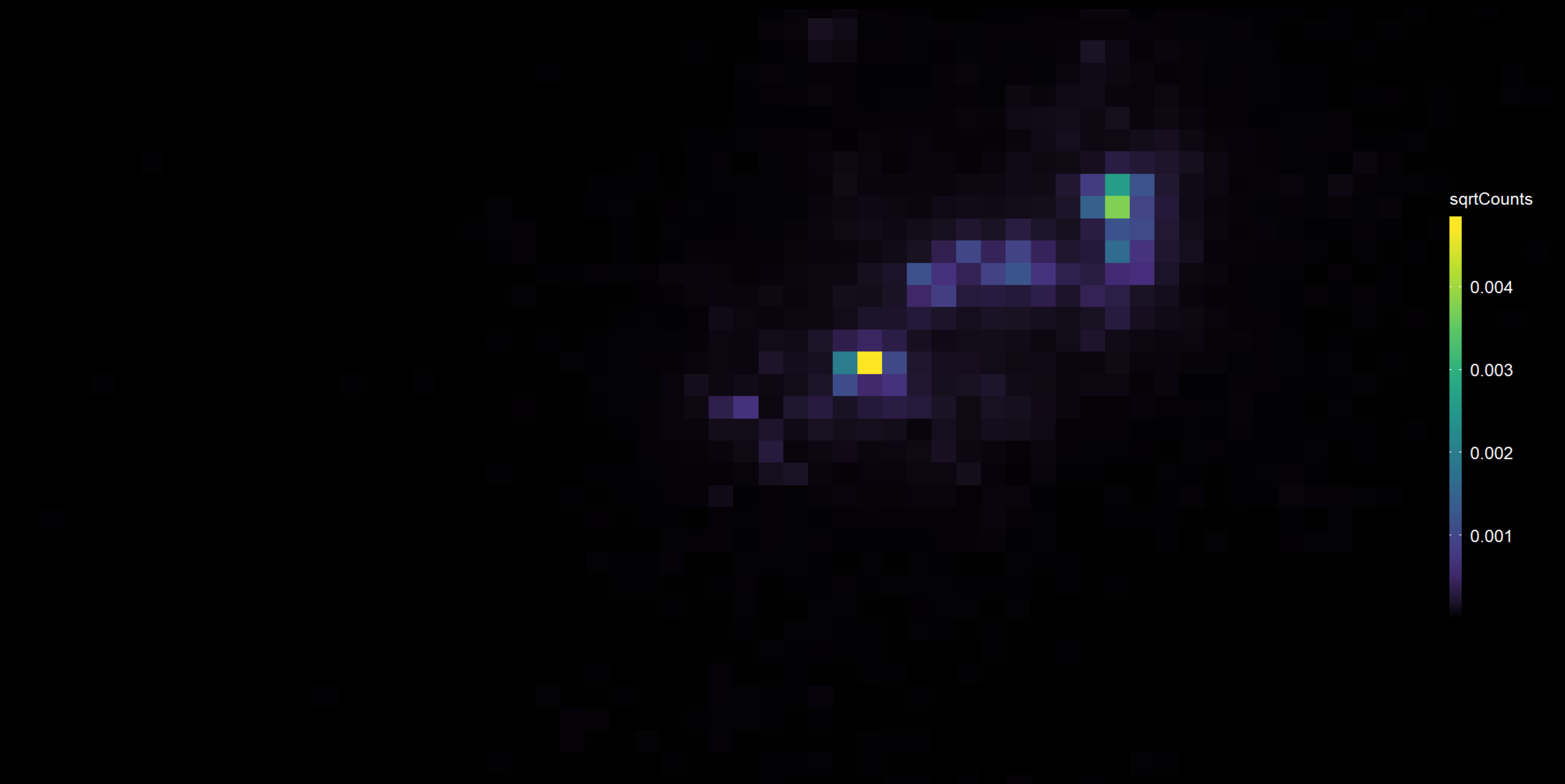

Point Source + Bkg
Baseline Model




McKeough et al 2016



Posterior Draws with MCMC
Expected photon counts in each pixel
given the observed counts




McKeough et al 2016

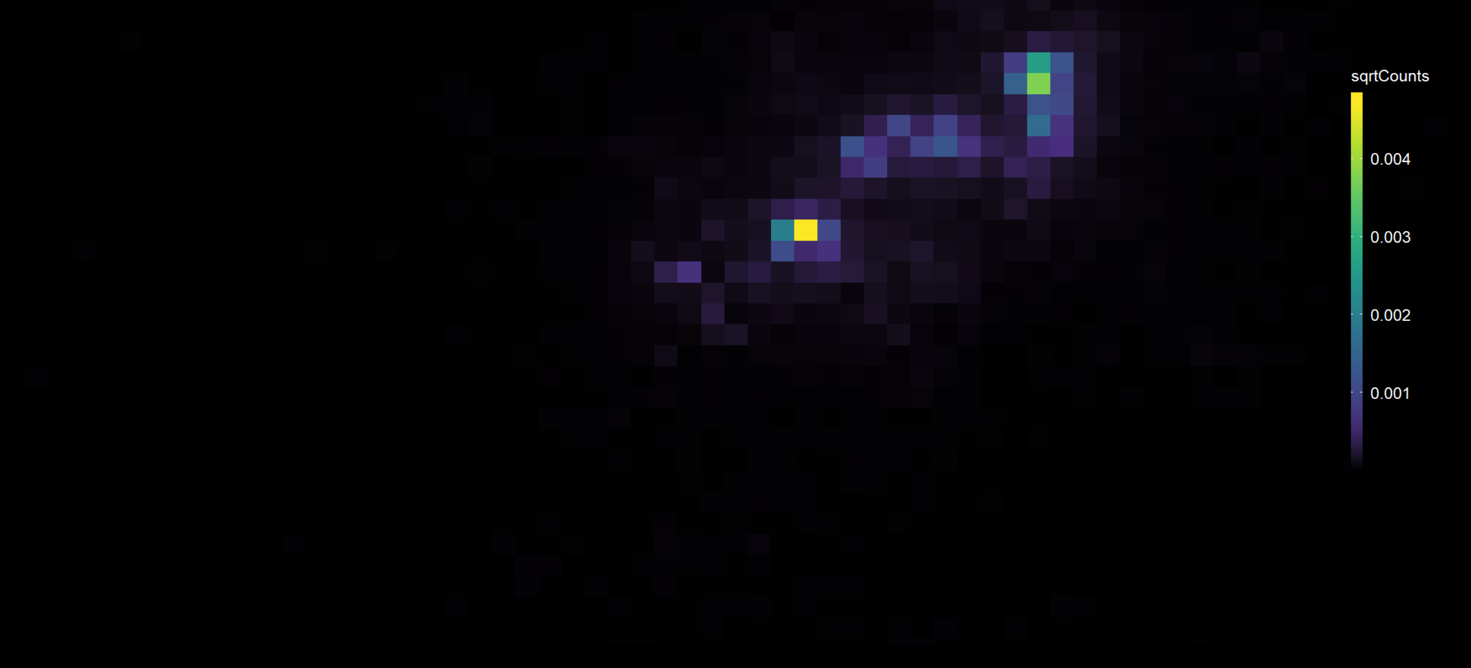
Posterior Mean

Note:
Code available: https://github.com/astrostat/LIRA
Esch et al 2004, Connors & Van Dyk 2007, Stein et al. 2015

# Adventure 4
# High Resolution X-ray Images

## Finding the source boundary

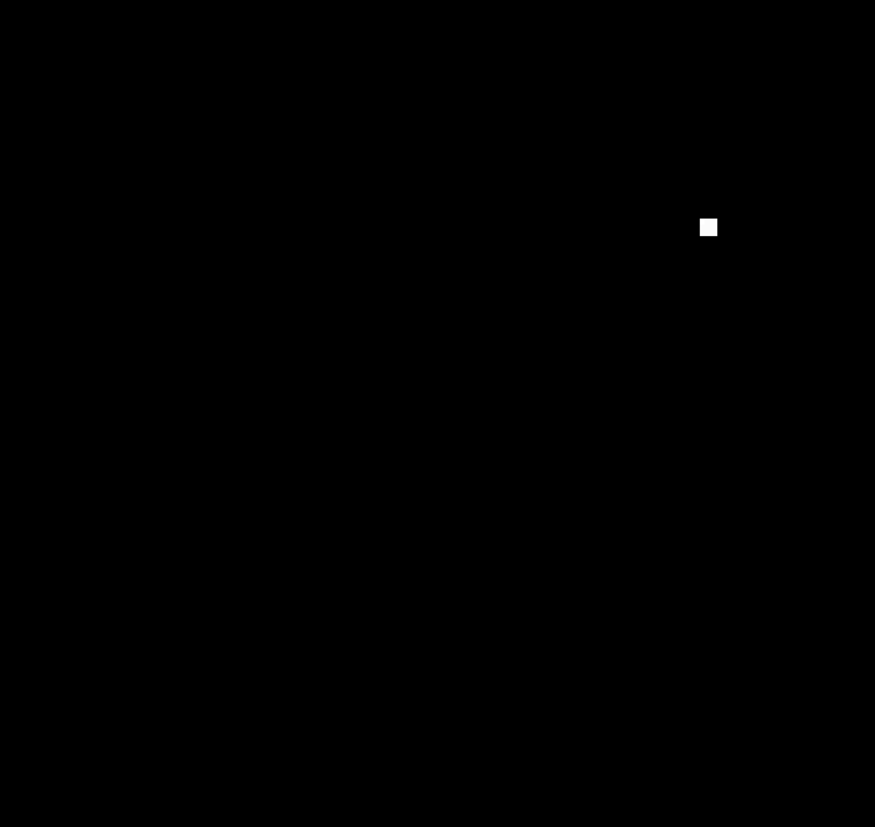
Posterior Mean




Posterior Draws with MCMC
probability distribution of pixel assignments

ISING Prior
Correlation between neighboring pixels

Optimal Boundary



Boundary with maximum probability
given LIRA-Ising posterior

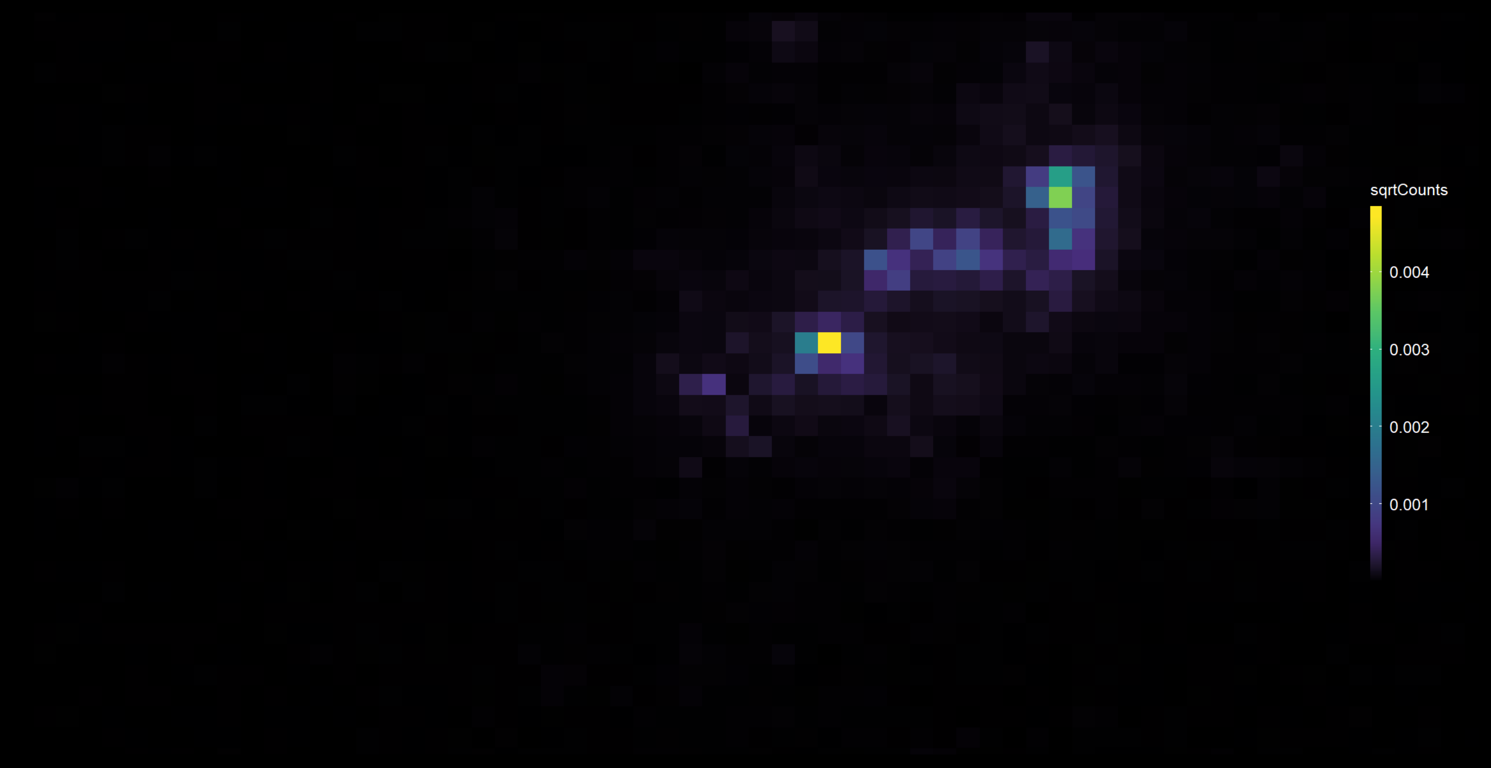
McKeough et al (in prep)

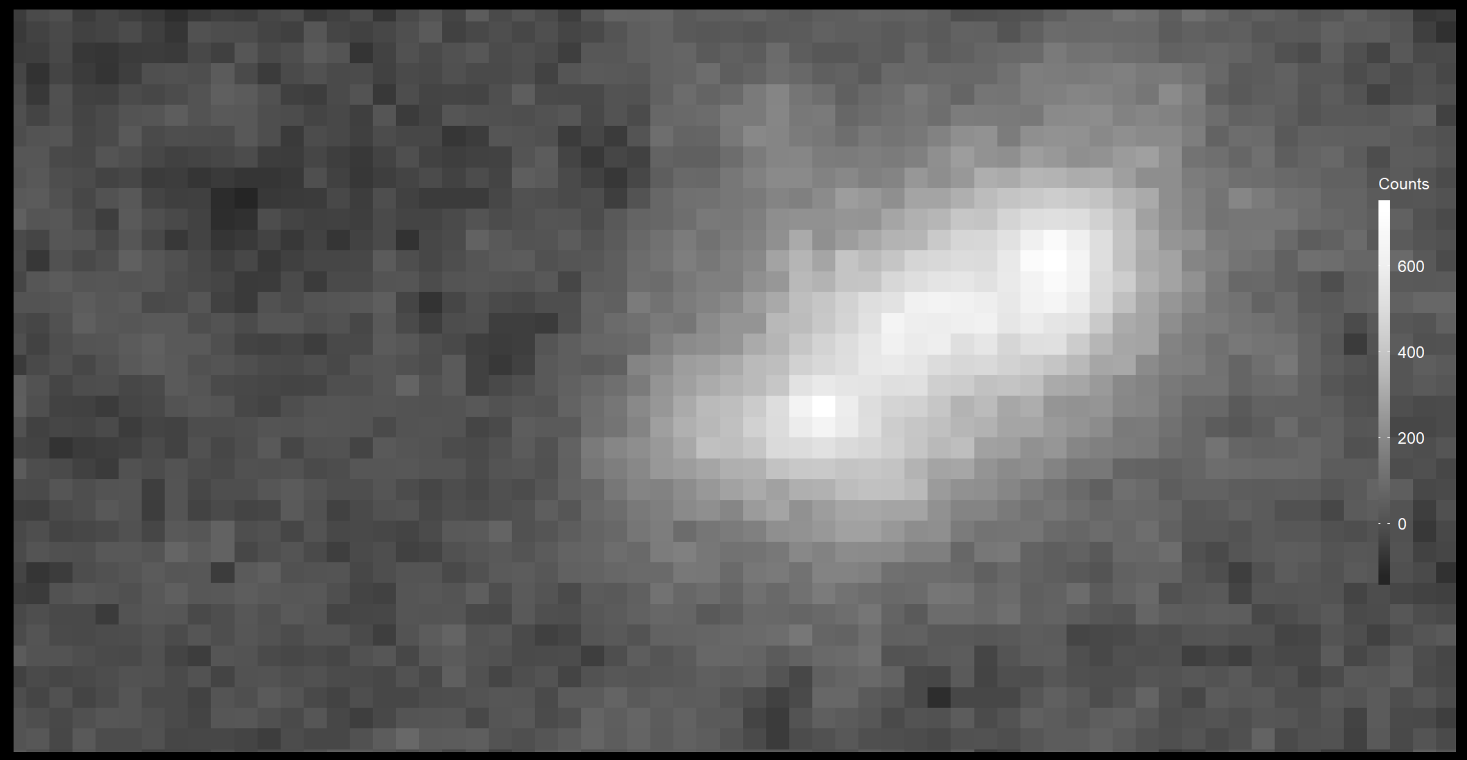
# Adventure 4
# High Resolution X-ray Images

## Finding the source boundary

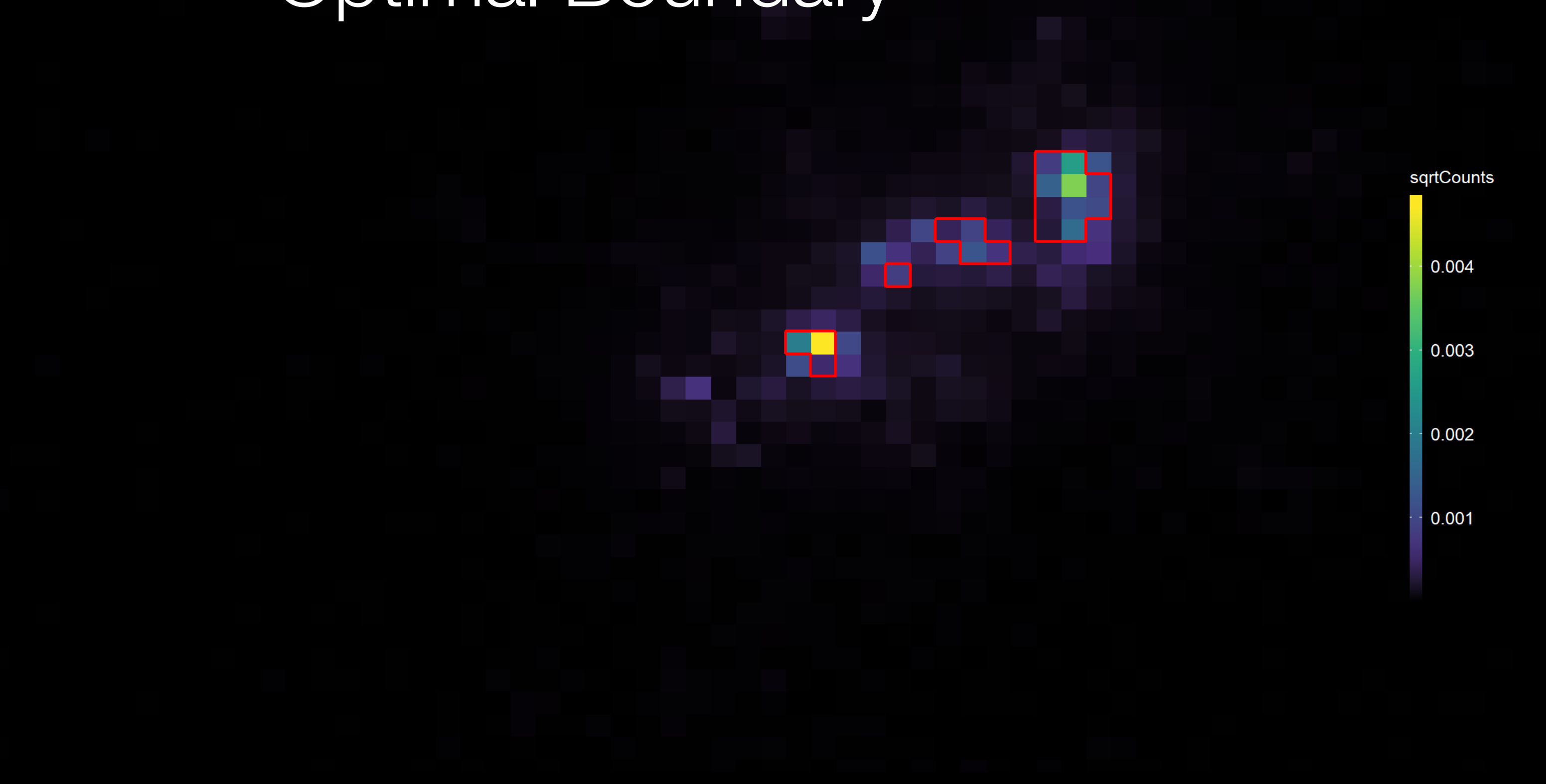
Posterior Draws with MCMC
probability distribution of pixel assignments

Posterior Mean




Optimal Boundary




ISING Prior
Correlation between neighboring pixels

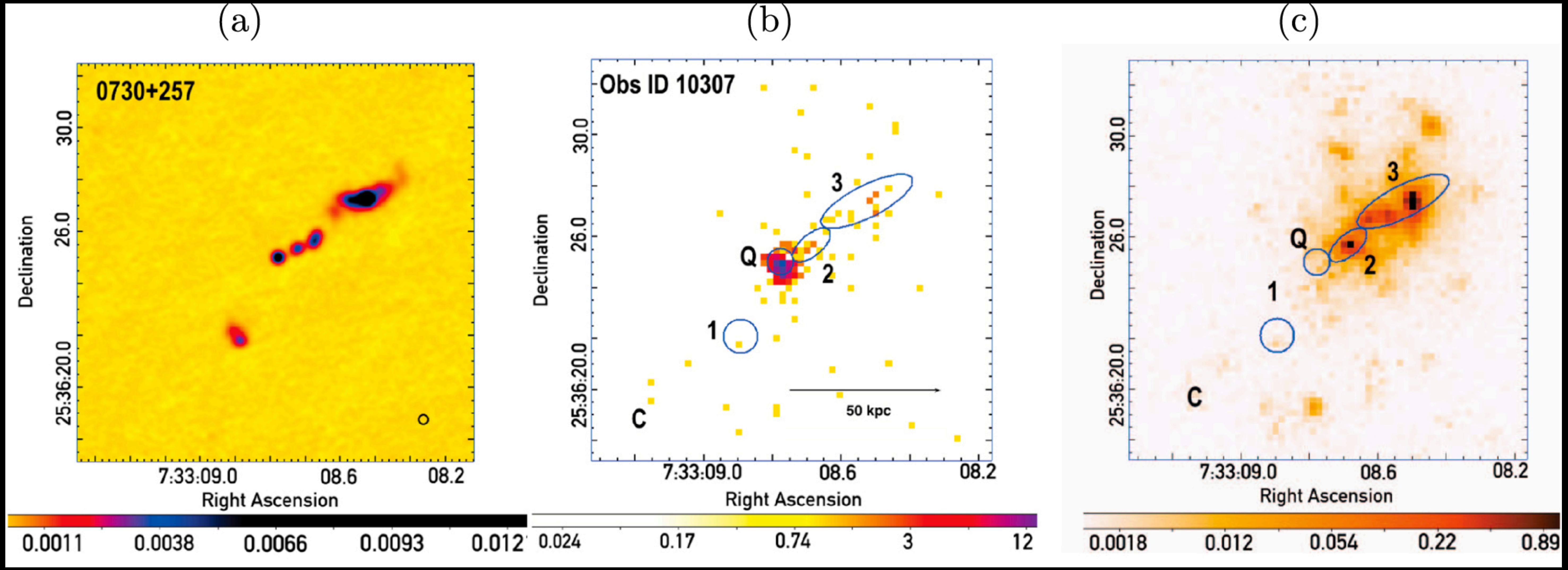
Boundary with maximum probability
given LIRA-Ising posterior

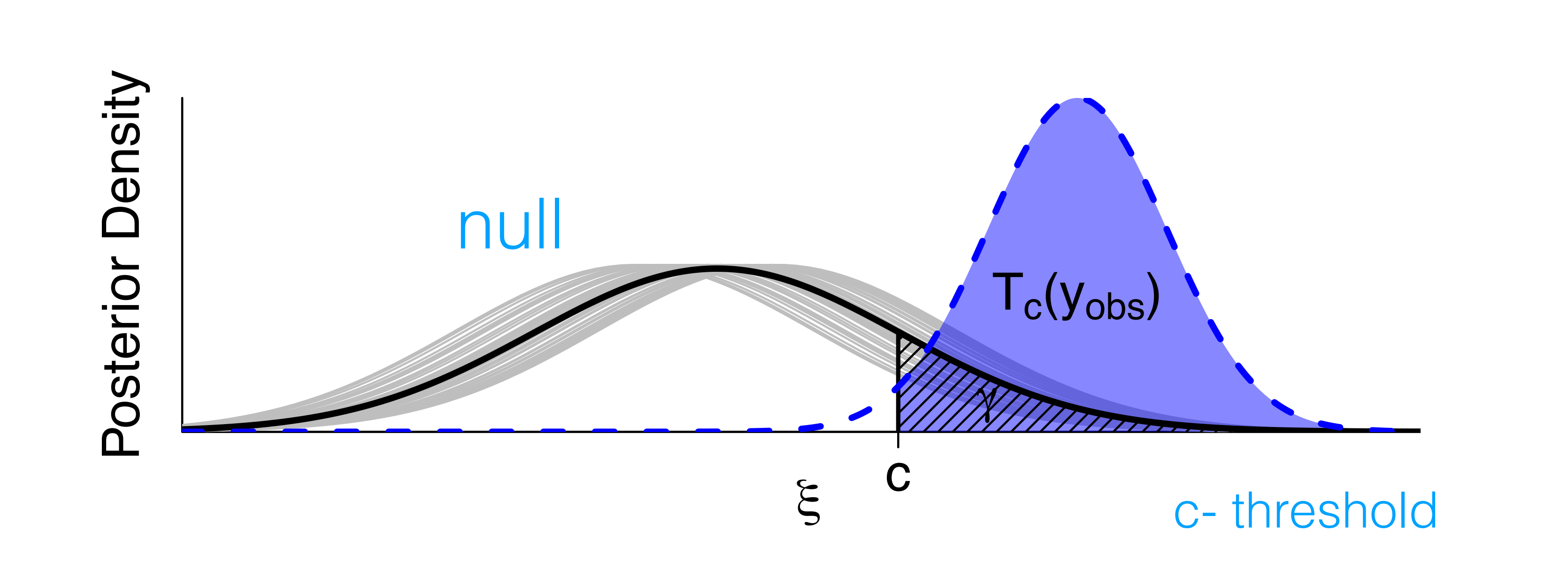
McKeough et al (in prep)

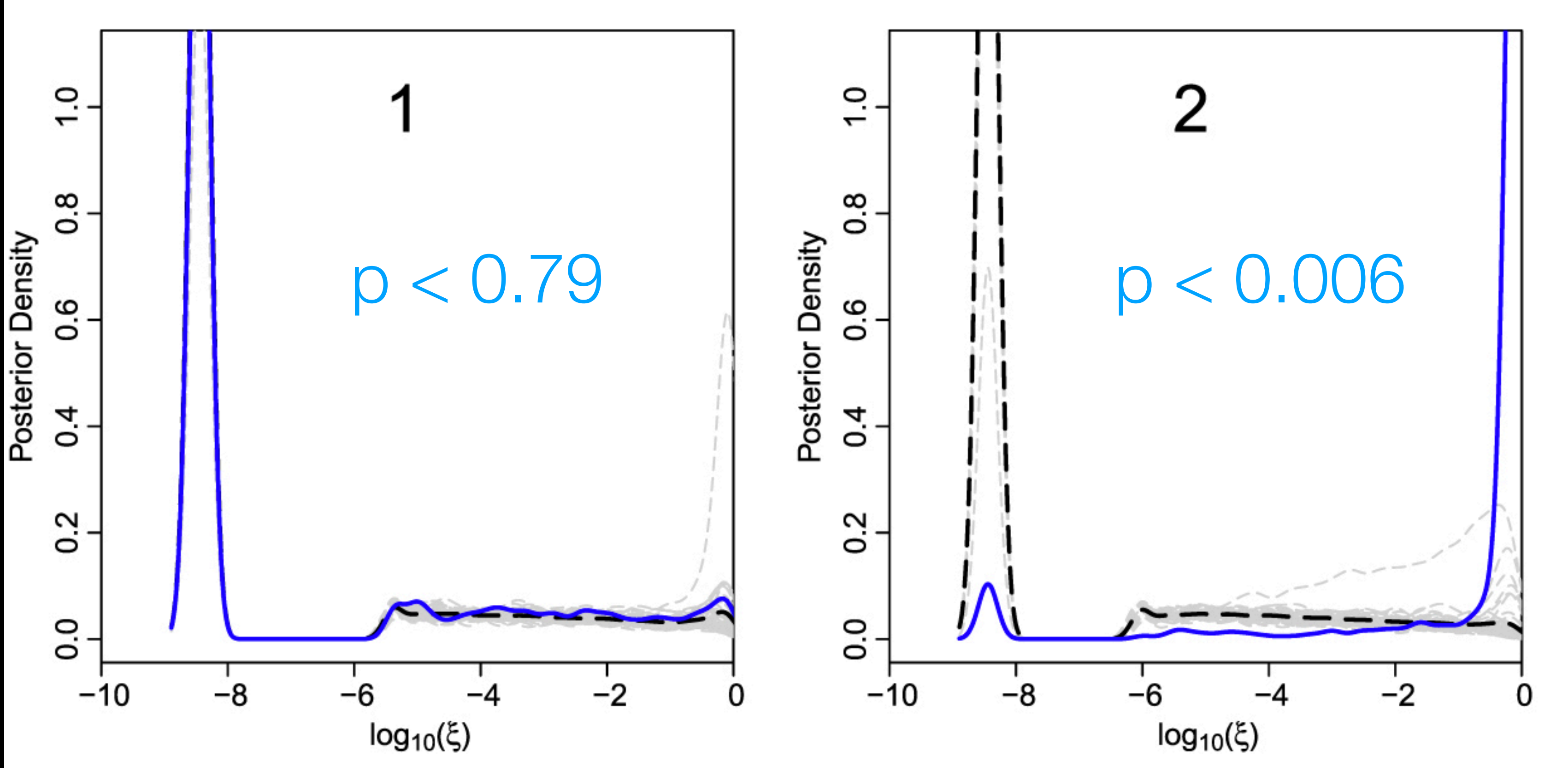
# Adventure 4
# High Resolution X-ray Images

**Evaluate Significance of feature over pre-specified region**

**Radio Band** **Chandra** **LIRA**




Stein et al. 2015, McKeough et al. 2016

# Adventures
so far….

**Low Counts Spectra**
- fitting complex spectral models
- line detection
- hardness ratio

**Sparse Poisson Images**
- source detection and upper limits
- structures in high resolution Poisson images
- source boundaries
- significance

# X-ray Analysis Standard Domains

- X-ray image is made by binning events into images, e.g. accumulating photons in a selected energy band and fixed exposure time:

  - *no spectral or temporal information*

  - *analysis require a point spread function*

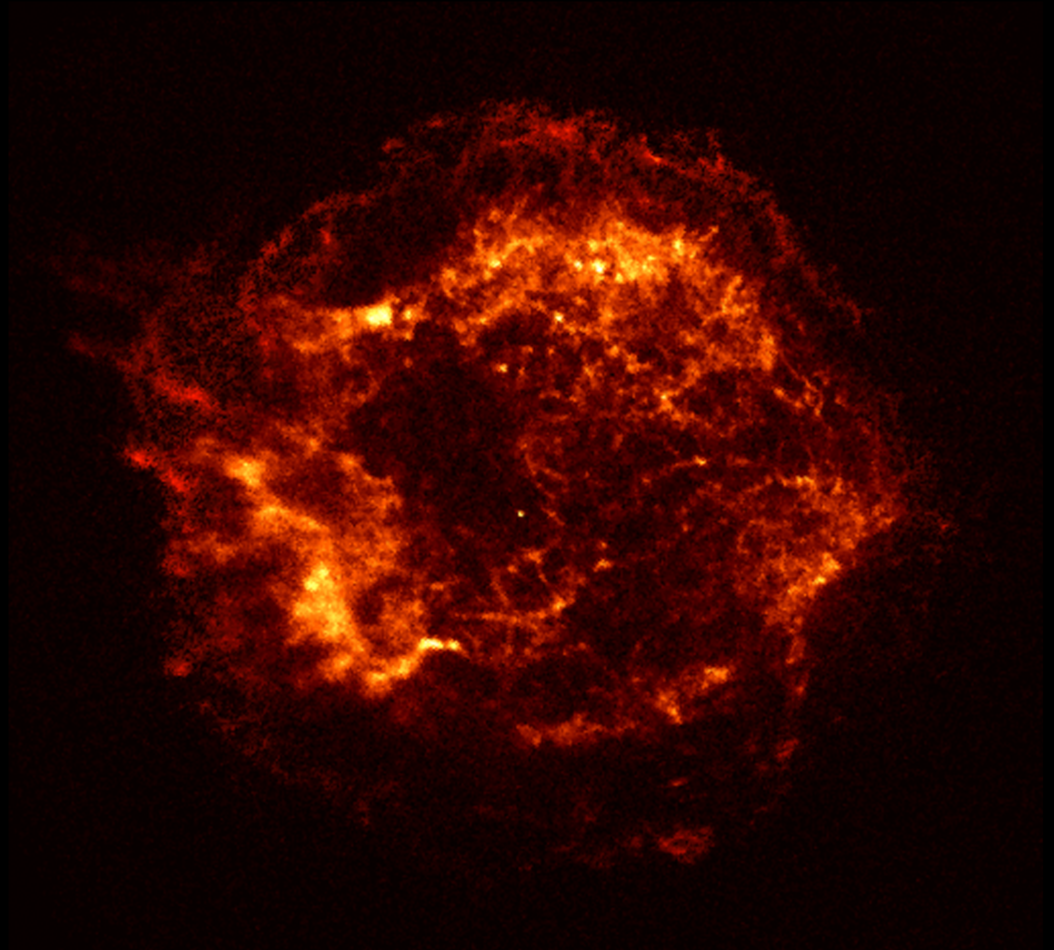
- Spectra for selected regions are generated by binning the events in energy:

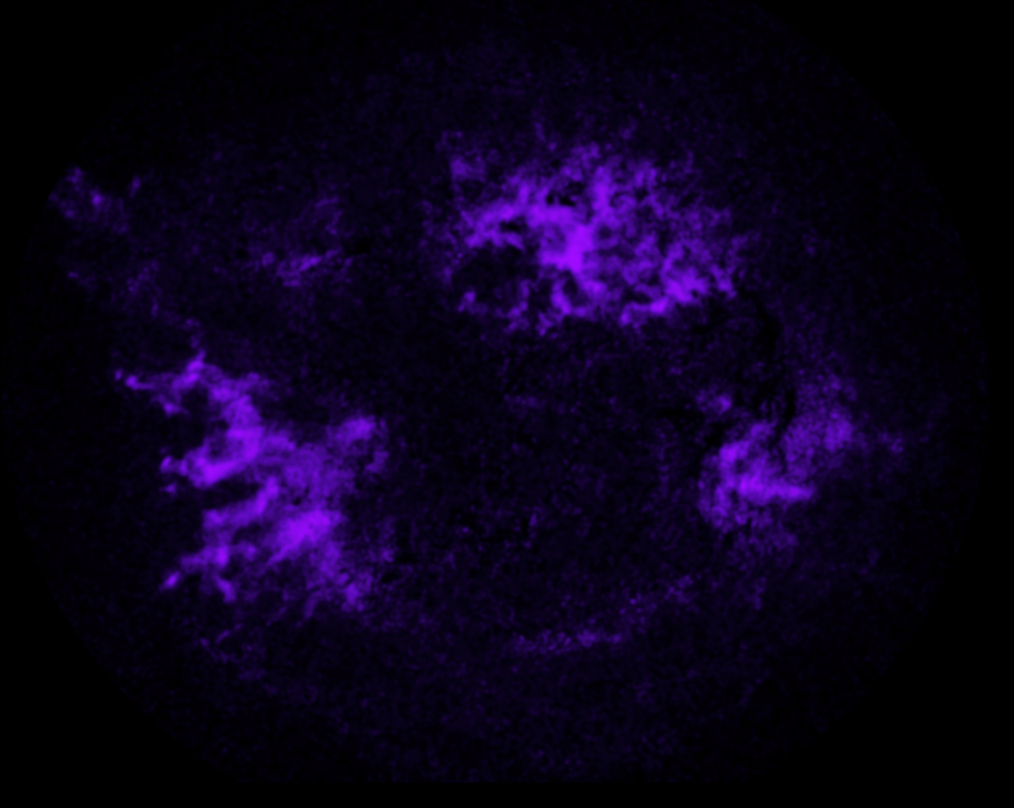
  - *no spatial or temporal information*

  - *require additional calibration files*

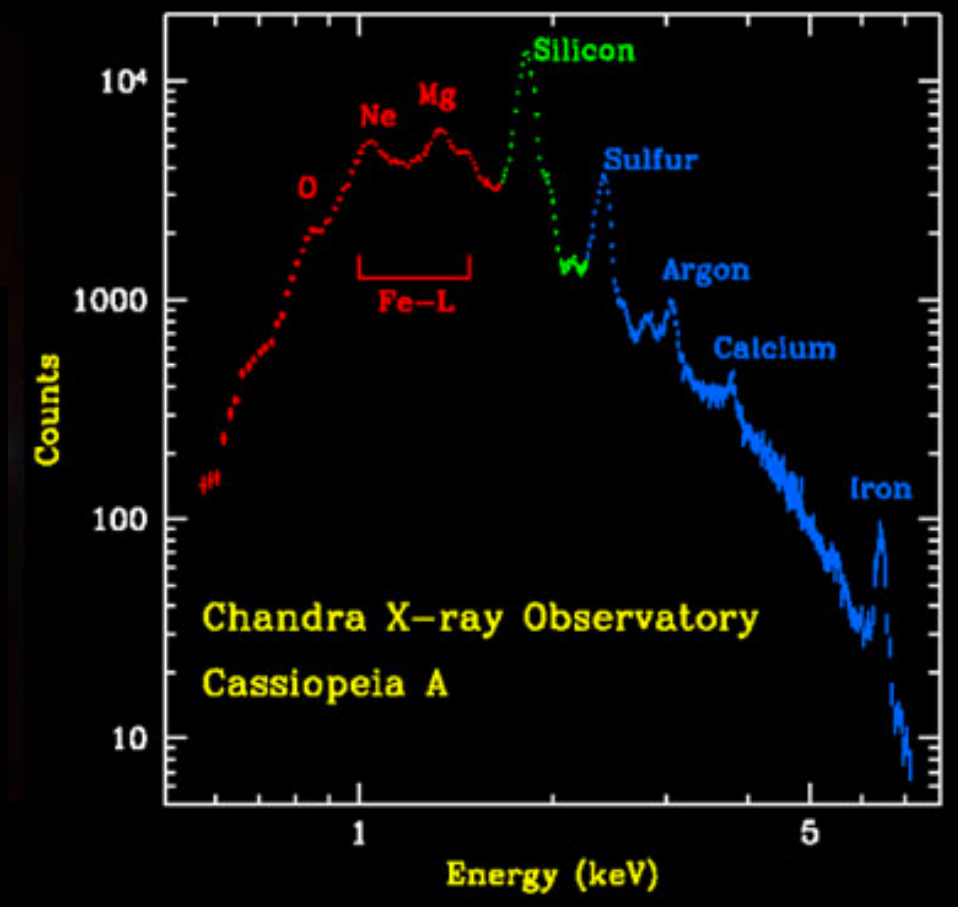
- Lightcurves for selected region and energy band binning the events in time:

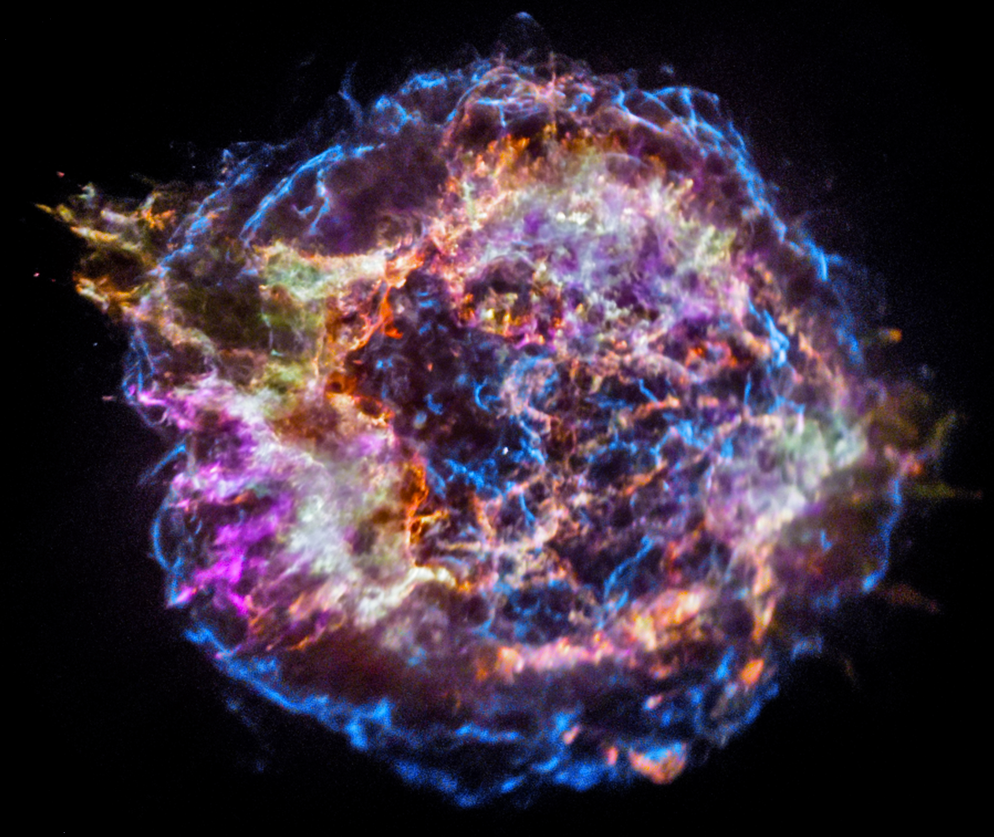
  - *no spatial or energy information*



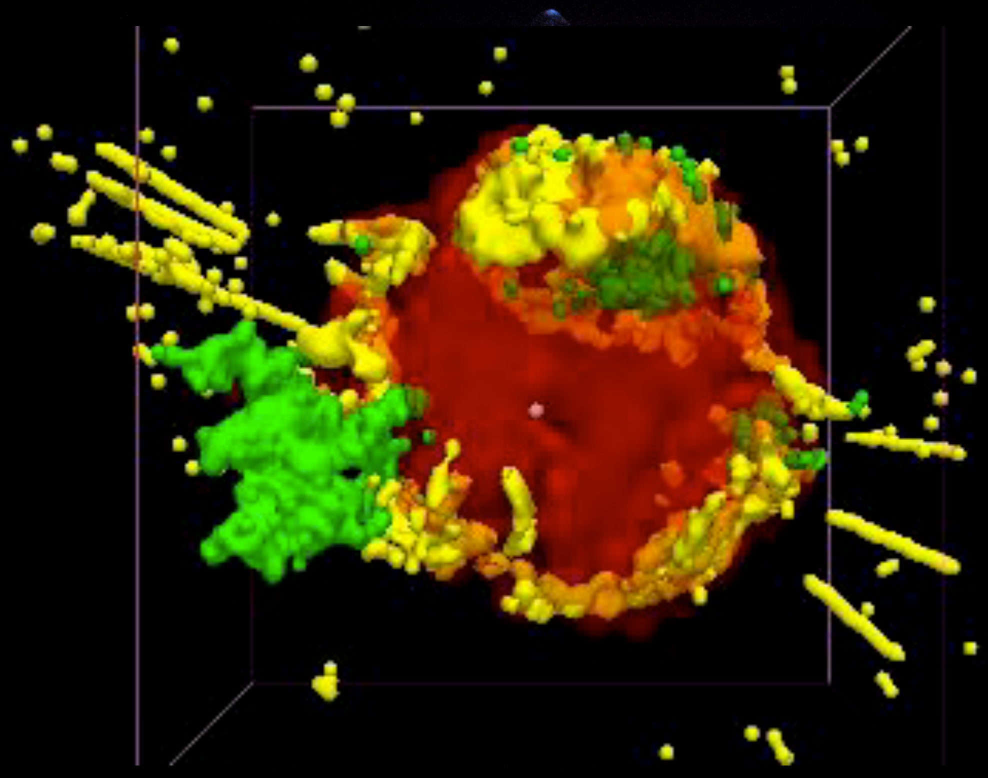


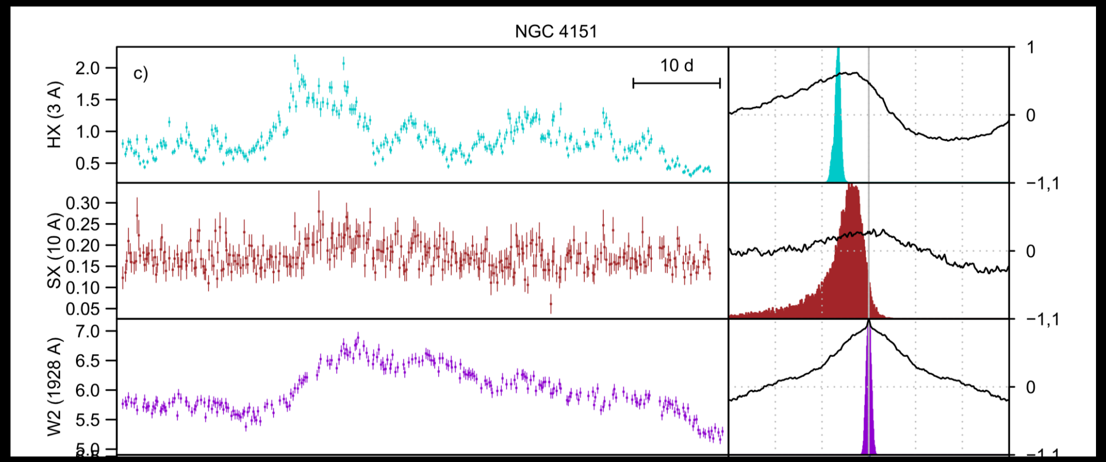
Cassiopeia A Supernova Remnant

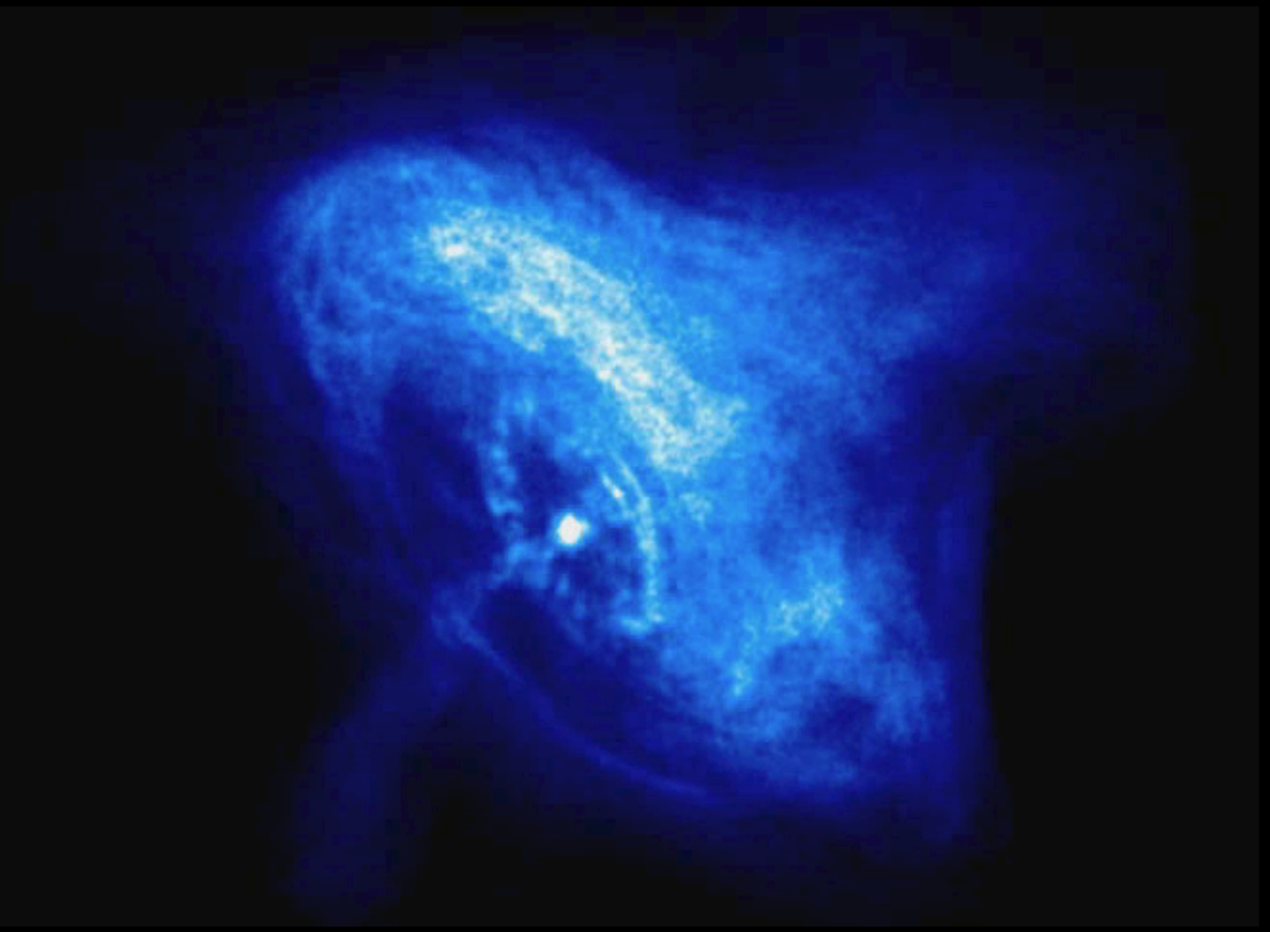


# Emerging Multi-Domain Analysis





| Analysis | Description | Current Method | Challenges | Emerging Methods |
|---|---|---|---|---|
| Spectral-Image | loss of time | source detection (VTP), spectral-image model, project, deproject in clusters, SNR | multi-spectra, averaging over image, overlapping sources, transients | BASCS |
| Spectral-Time | loss of location | multi-spectra, inter-band correlation | low counts spectra, non-even sampling, different apertures, multi-components | cross-spectrum, ABC, JAVELIN, Auto-Mark |
| Image-Time | loss of energy | image difference, source detection | spectral information, evolving boundaries, PSF, averaging | eBASCS 4D-automark |

# Emerging Multi-Domain Analysis

**Full information: Image-Spectral-Time**

**Examples:**
Probabilistic separation of  photons
from two close sources with eBASCS
using location, spectrum and time (Meyer+ 2021)

Change-points and Image Segmentation
for Time series of Images - 4D_Automark  (Xu+ 2021)

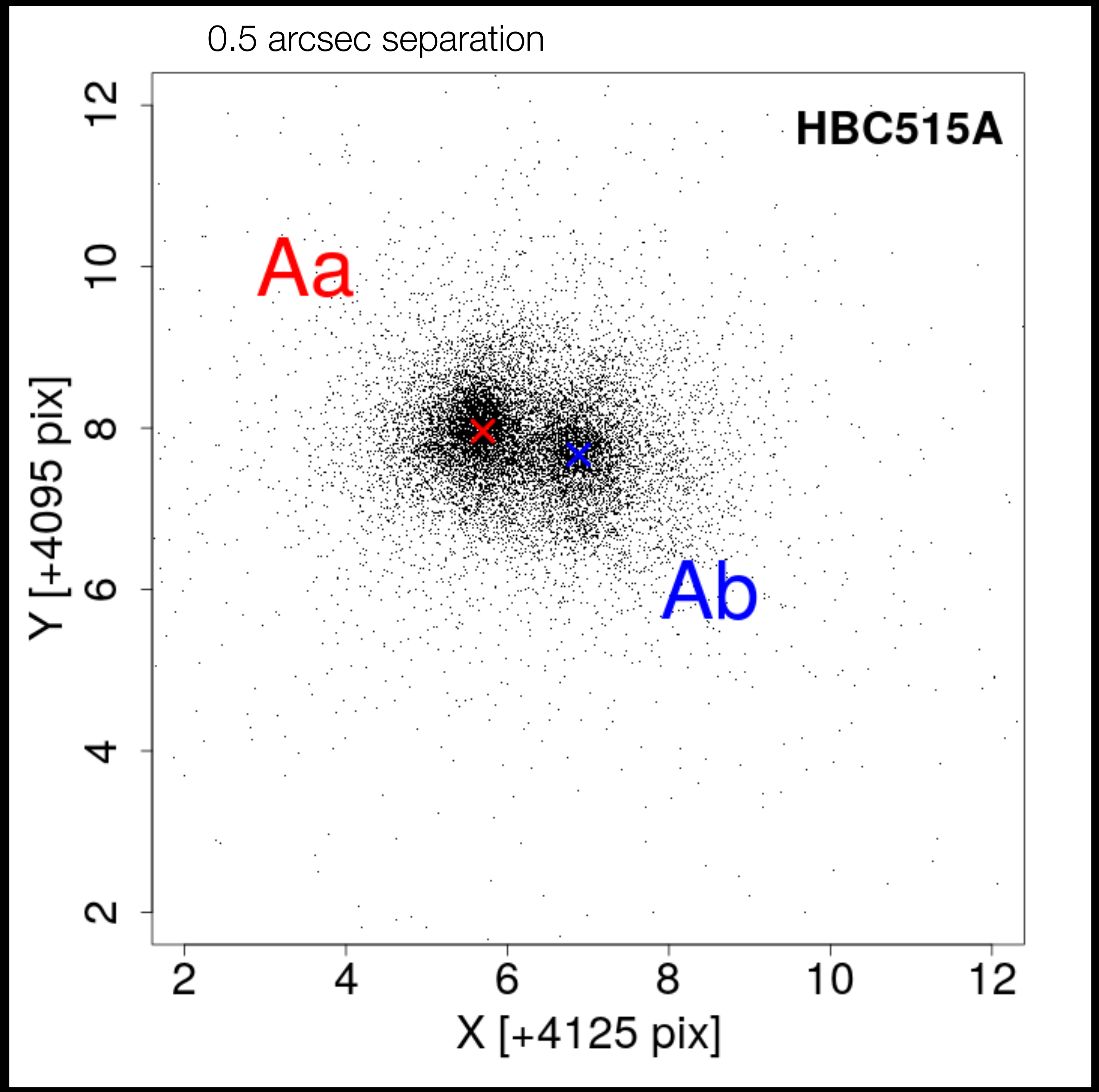
Chandra X-ray Image of Orion Nebula

Credit: NASA/CXC/Penn State/E.Feigelson & K.Getman et al.

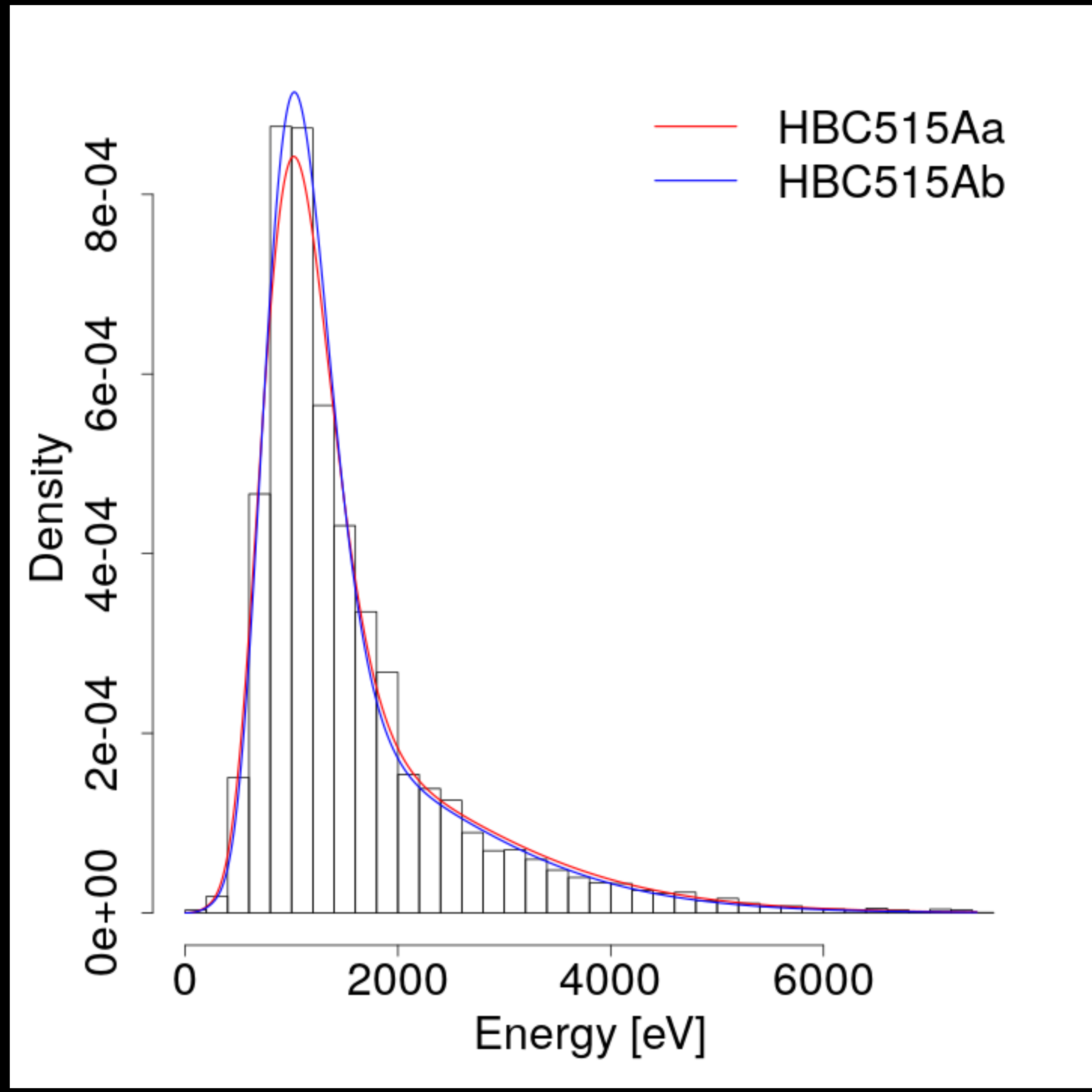
# Emerging Multi-Domain Analysis

**Full information: Image-Spectral-Time**

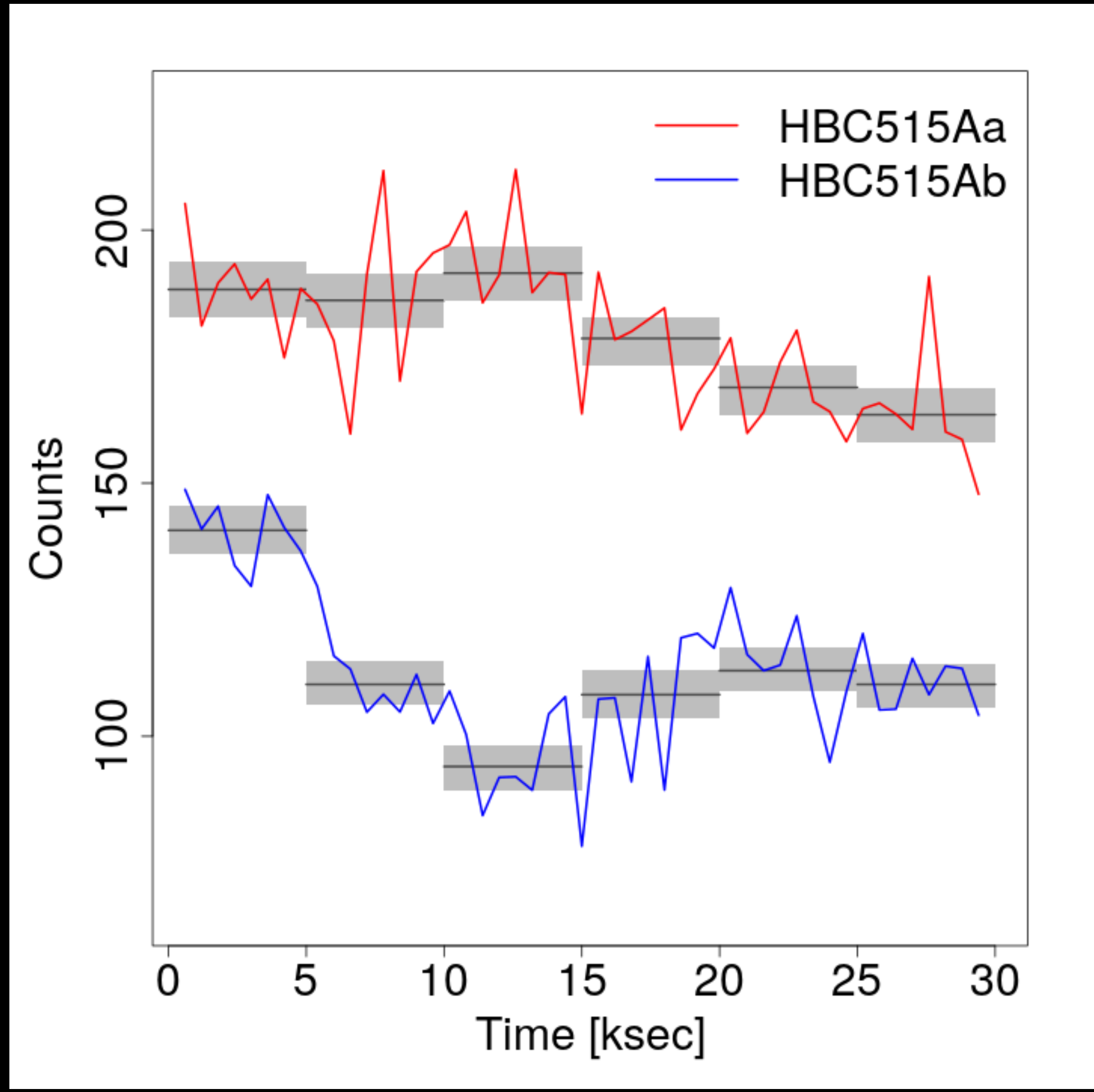
**Example:**
Probabilistic separation of photons from two close sources with eBASCS using location, spectrum and time




locations of the events
posterior mean of the locations of Aa and Bb with BASCS




spectra for each star with eBASCS



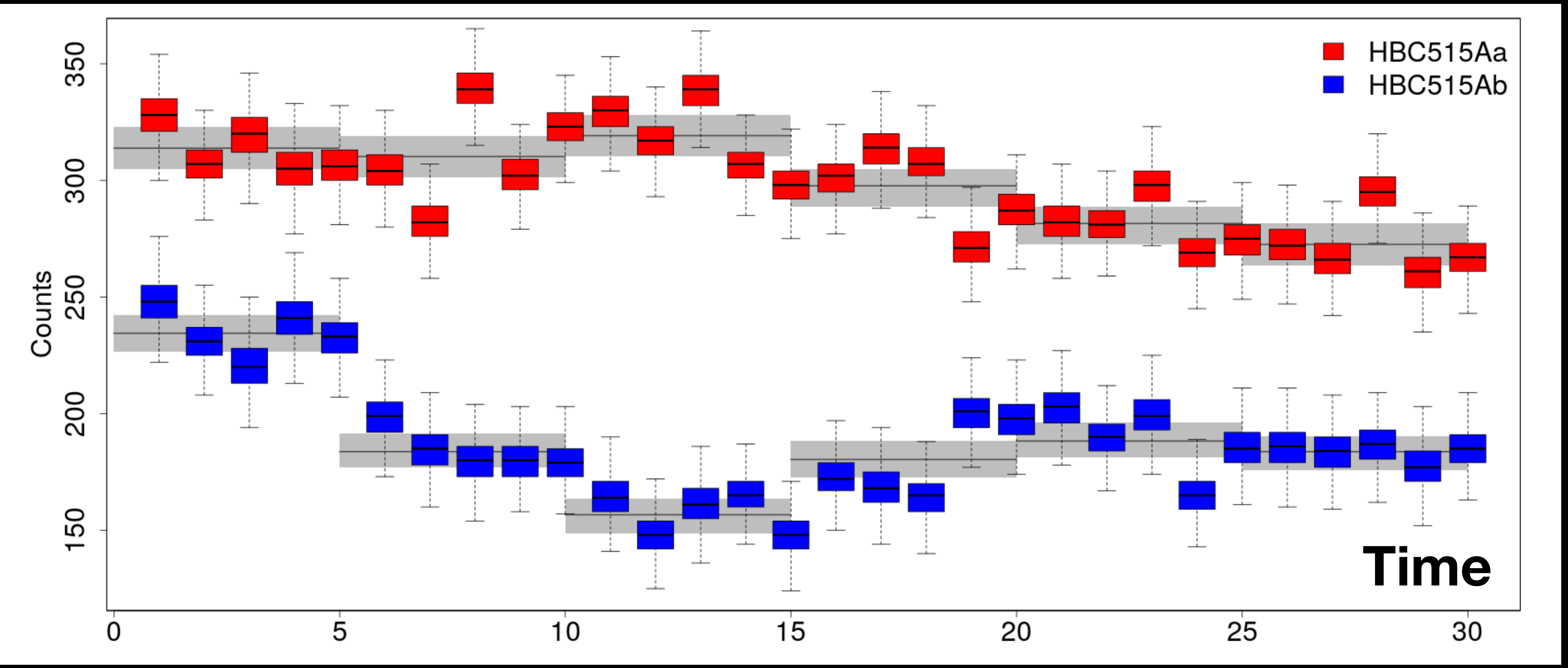

light curves of each component eBASCS

Meyer et al 2021

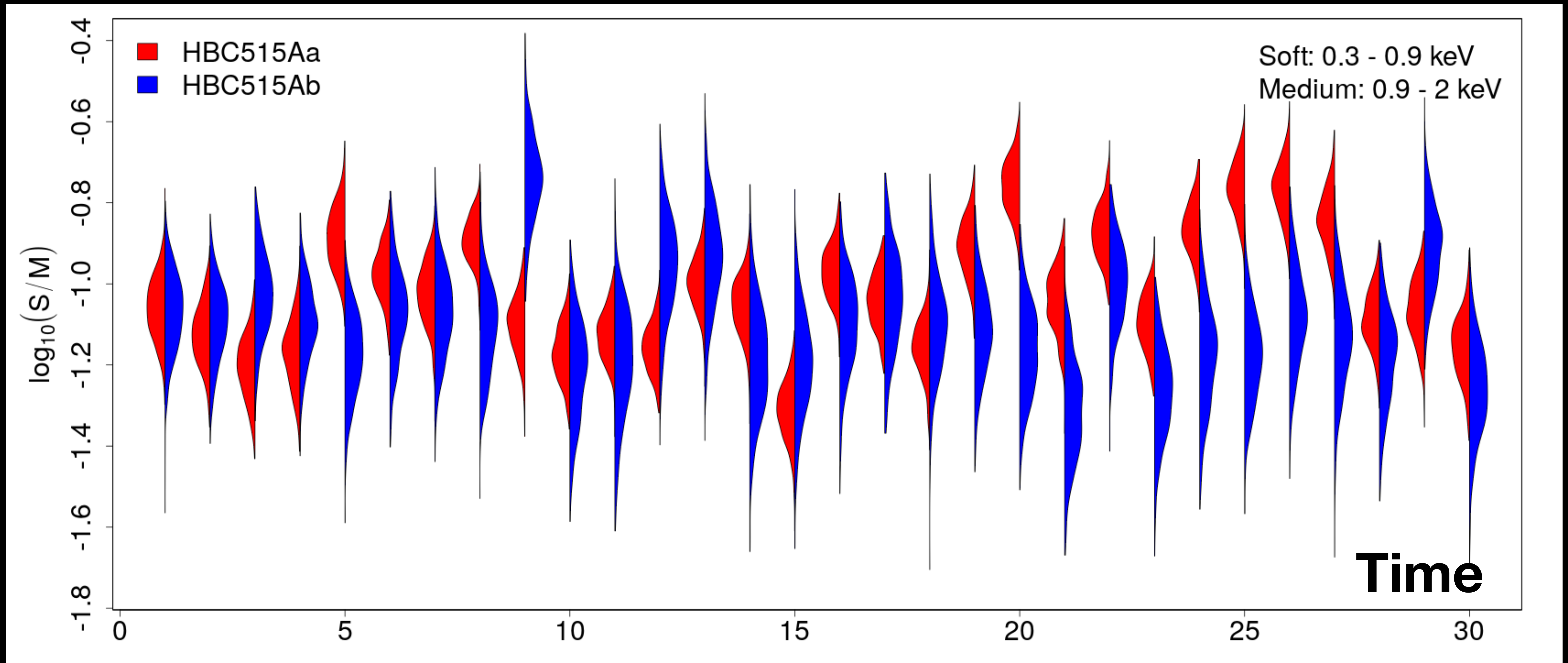
# Emerging Multi-Domain Analysis

**Full information: Image-Spectral-Time**

Counts




Ratio
Soft / Medium



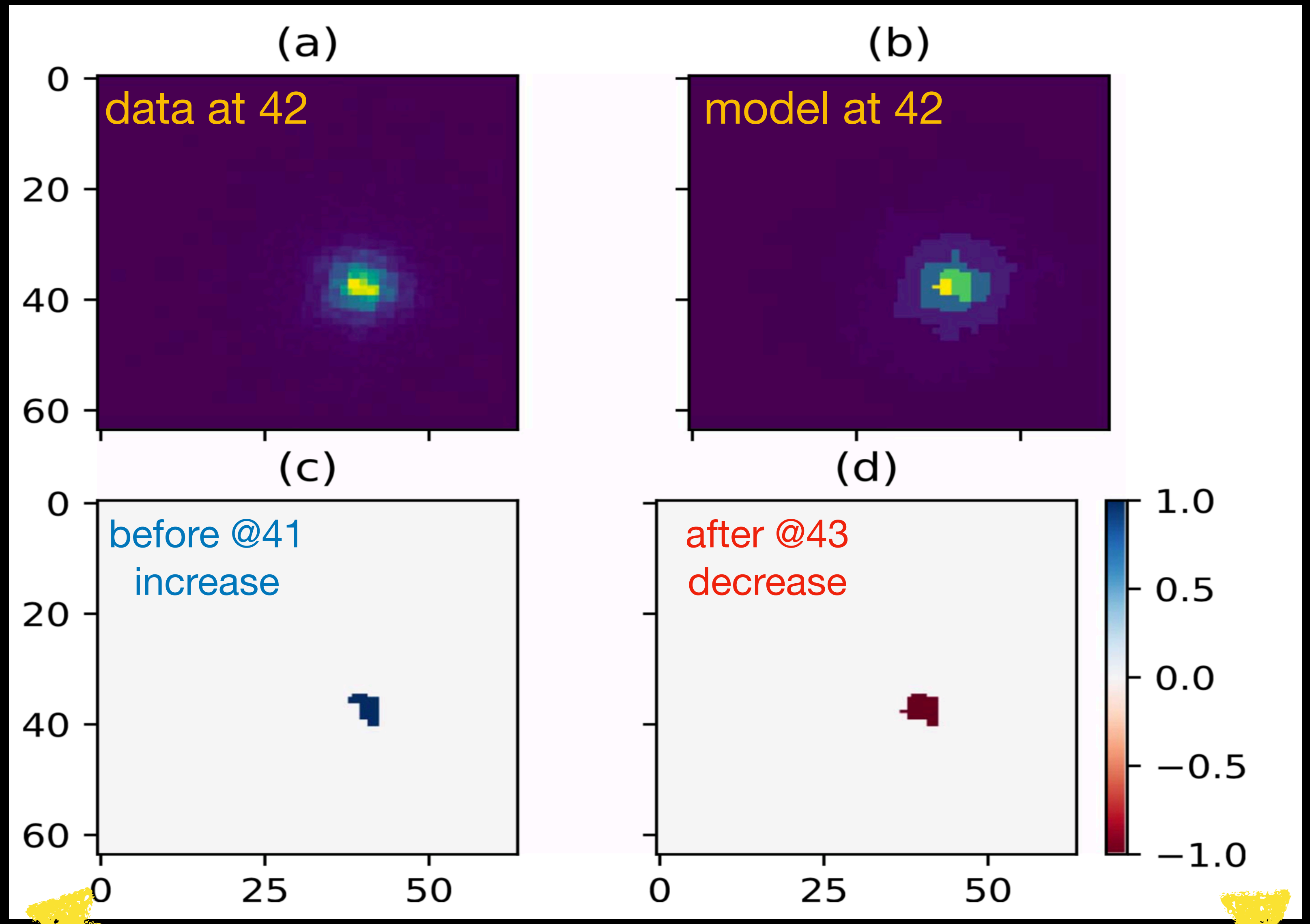

**eBASCS:**
**Bayesian model to separate events from each star using energy, timing and location to mark X-ray photons assigned to each star to calculate intensity and hardness ratio variation in time.**

Meyer et al 2021

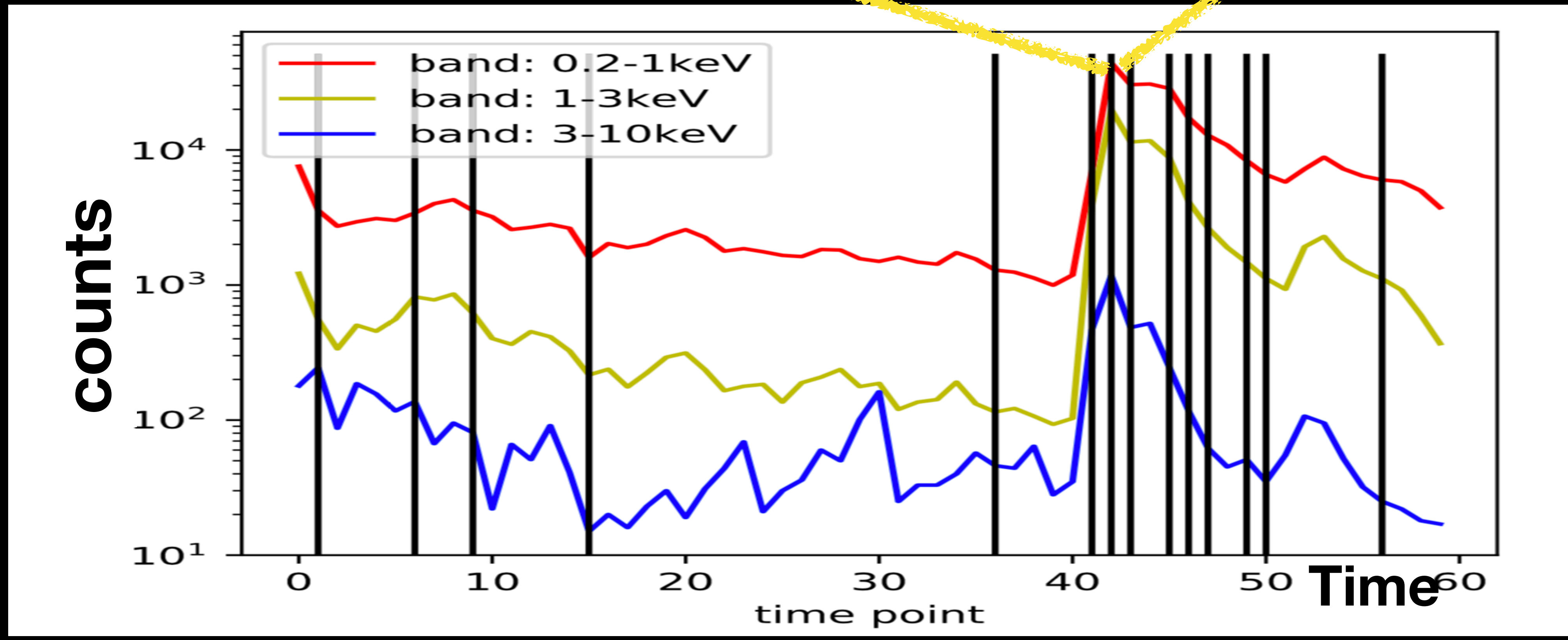
# Emerging Multi-Domain Analysis

**Full information: Image-Spectral-Time**

XMM data
Proxima Centauri




**Change-points and Image Segmentation for Time Series Images - 4D-Automark**




Xu+ 2021

# Future Full Multi-Domain Analysis

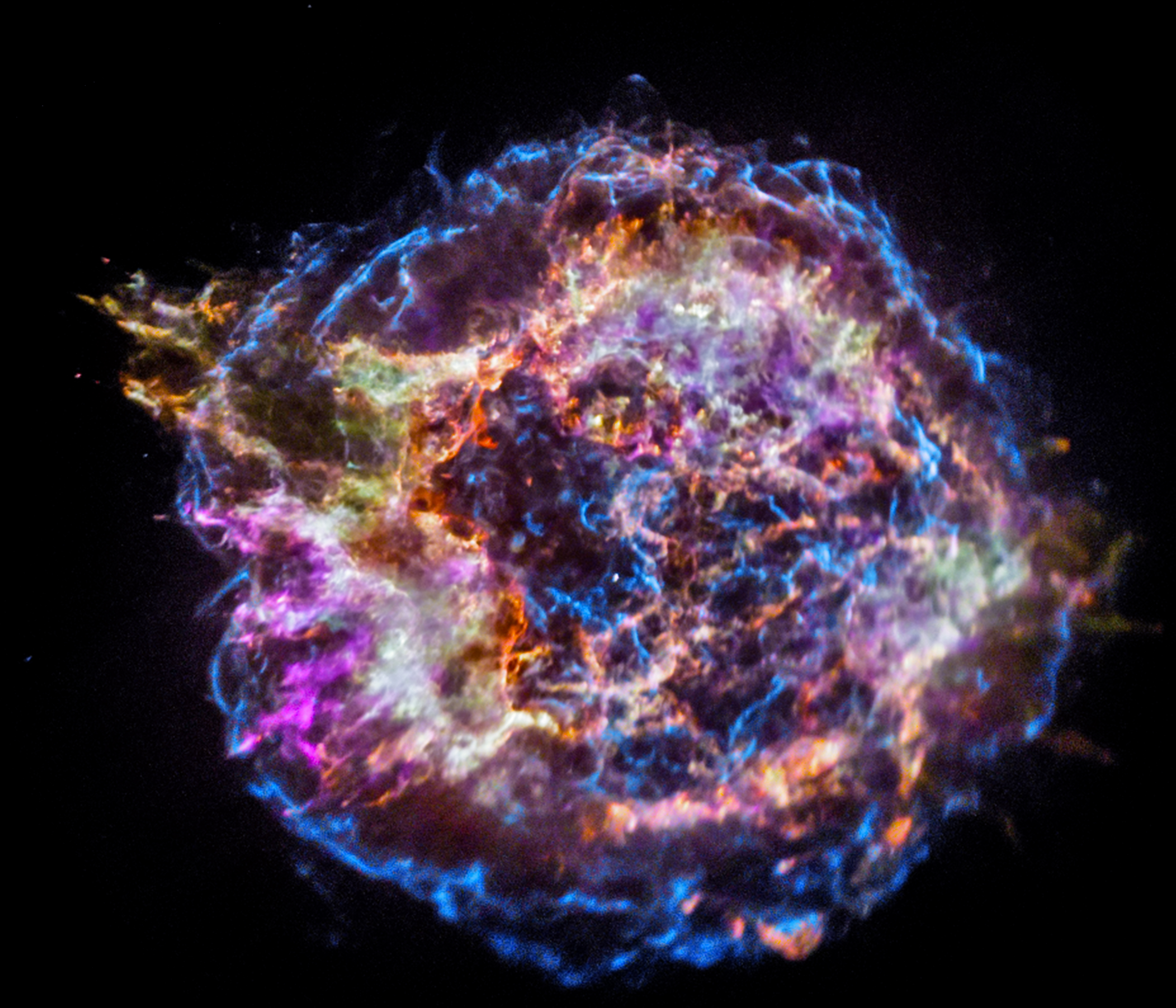
| Analysis | Description | Current Methods | Challenges | Emerging Methodology |
|---|---|---|---|---|
| spectral-image-time | use energy, location, time | multi-band images in several time bins | non-binned events instrument response, background | eBASCS, 4D-automark |
| polarimetry | new domain | simultaneous 3D spectral modeling | no energy information, correlation between Stokes vectors | |

# Adventures Summary

## Past

Low Counts X-ray Spectra
- fitting complex spectral models
- line detection
- hardness ratio

Sparse X-ray Images
- source detection and upper limits
- structures in high resolution Poisson images
- source boundaries
- significance



## Future

- **full multi-domain analysis**
- **rising interest in methodology**
- **likelihood free simulation based methods**
- **application of Machine Learning and Artificial Inteligence methods**

## Astro2020 White Papers

Siemiginowska et al (2019) AAS WGAA *The Next Decade of Astroinformatics and Astrostatistics*
https://ui.adsabs.harvard.edu/abs/2019BAAS...51c.355S/abstract

Eadie et al (2019) AAS WGAA *Realizing the potential of astrostatistics and astroinformatics*
https://ui.adsabs.harvard.edu/abs/2019BAAS...51g.233E/abstract

Peek et al (2019) *Robust Archives Maximize Scientific Accessibility*
https://ui.adsabs.harvard.edu/abs/2019BAAS...51g.105P/abstract

Fabbiano et al (2019) *Increasing the Discovery Space in Astrophysics - A Collation of Six Submitted White Papers*
https://ui.adsabs.harvard.edu/abs/2019arXiv190306634F/abstract

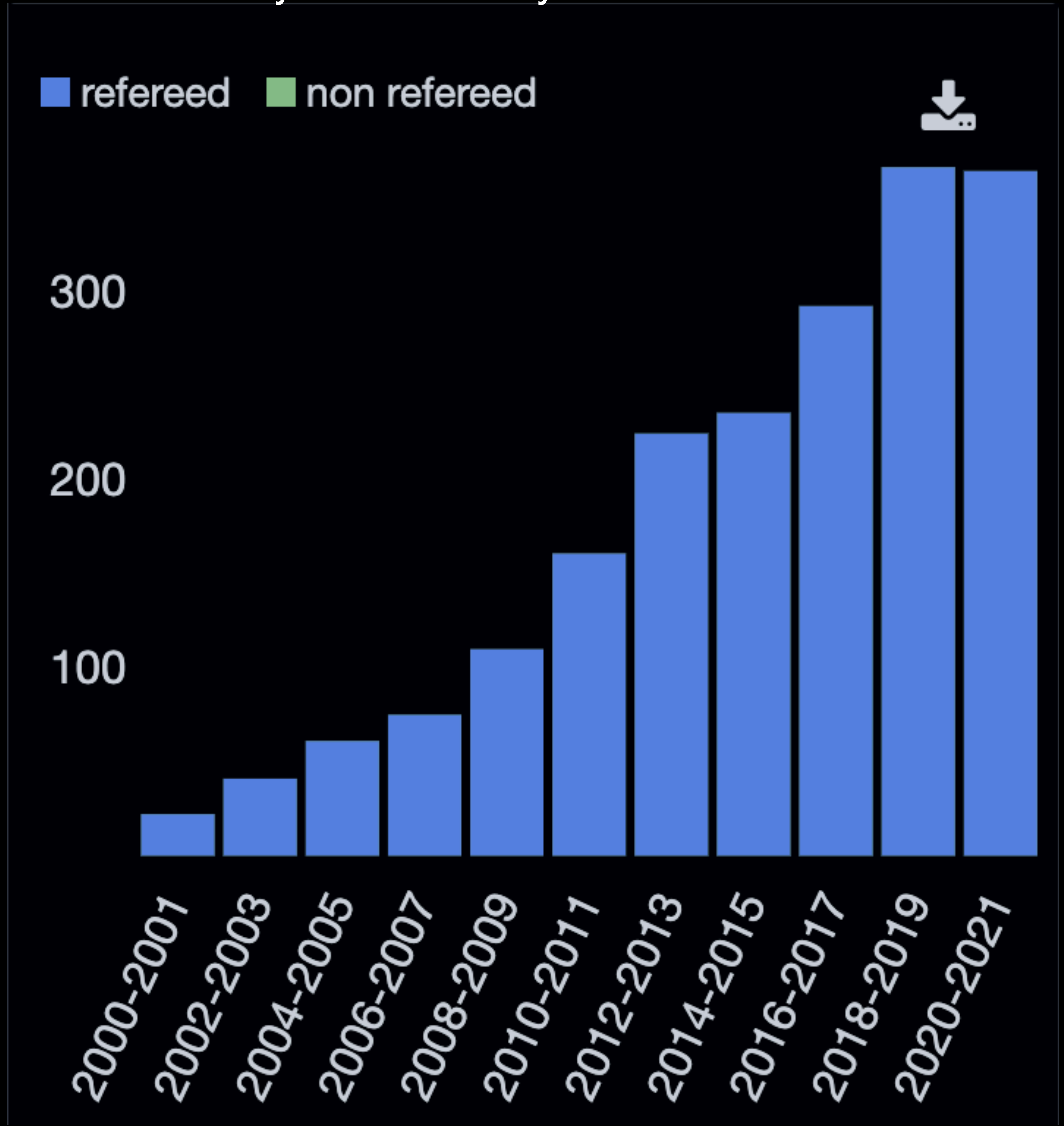
Kurtz et al (2020) *Enabling Synergy: Improving the Information Infrastructure for Planetary Science*
https://ui.adsabs.harvard.edu/abs/2020arXiv200914323K/abstract

# Future

- Steady growth of the field
- Collaboration with statisticians and computer scientists necessary
- Funding to support research in astrostatistics
- Education and curriculum
- Faculty jobs in Astrostatistics

Note: 3 job postings on the AAS Job register

1957 refereed research papers in ApJ, MNRAS, A&A with words “Bayesian & X-rays”



# Astrostatistics Organizations

- AAS Working Group on Astroinformatics and Astrostatistics (WGAA) - 2012
  https://aas.org/comms/working-group-astroinformatics-and-astrostatistics-wgaa
- ASA Astrostatistics Interest Group - 2014 https://astrostat.org/join.html
- IAU Commision B3 on Astroinformatics and Astrostatistics - 2015
  https://www.iau.org/science/scientific_bodies/commissions/B3/
- IAA International Astrostatistics Association - 2012 http://iaa.mi.oa-brera.inaf.it/IAA/home.html
- ISI International Statistical Institute - 2010 Astrostatistics Network

**AIG Student Paper Competition 2022**
**deadline Dec.13, 2021** https://astrostat.org/competition/

**Activities:**

**Newsletters**
**Seminars**

IAU-IAA https://sites.google.com/view/iau-iaa-seminar/home

CHASC http://hea-www.harvard.edu/AstroStat/CHASC_2122/

## references

Meyer, A. D. +2021, *"eBASCS: Disentangling overlapping astronomical sources II, using spatial, spectral, and temporal information"*, doi:10.1093/mnras/stab1456.

Xu, C., Günther, H. M., Kashyap, +2021 *"Change-point Detection and Image Segmentation for Time Series of Astrophysical Images"*, doi:10.3847/1538-3881/abe0b6.

McKeough+ 2016, *"Detecting Relativistic X-Ray Jets in High-redshift Quasars"*, doi:10.3847/1538-4357/833/1/123.

Stein, N. M., van Dyk, D. A., Kashyap, V. L., and Siemiginowska, A., 2015, *"Detecting Unspecified Structure in Low-count Images"*. doi:10.1088/0004-637X/813/1/66.

Jones, D. E., Kashyap, V. L., and van Dyk, D. A., 2015 *"Disentangling Overlapping Astronomical Sources Using Spatial and Spectral Information"*, doi:10.1088/0004-637X/808/2/137.

Xu, J.+2014, *"A Fully Bayesian Method for Jointly Fitting Instrumental Calibration and X-Ray Spectral Models"*, doi:10.1088/0004-637X/794/2/97.

Lee, H.+2011, *"Accounting for Calibration Uncertainties in X-ray Analysis: Effective Areas in Spectral Fitting"*, doi:10.1088/0004-637X/731/2/126.

Kashyap+2010, *"On Computing Upper Limits to Source Intensities"*, doi:10.1088/0004-637X/719/1/900.

Park, T., +2006, *"Bayesian Estimation of Hardness Ratios: Modeling and Computations"*, doi:10.1086/507406.

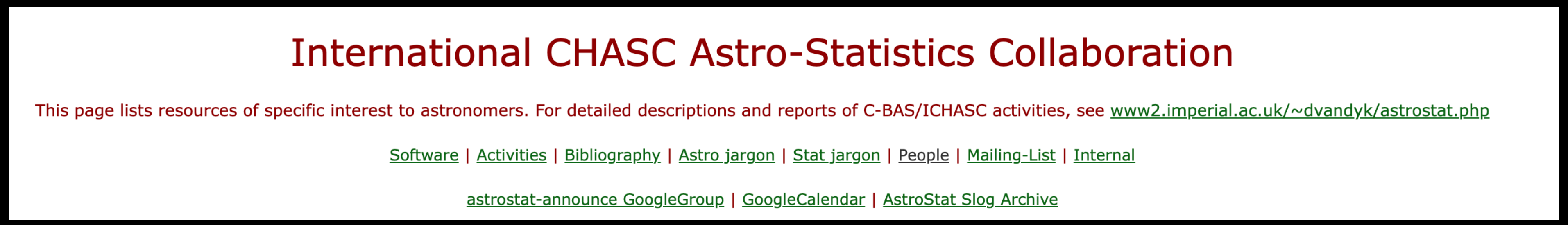
Park, T., van Dyk, D. A., and Siemiginowska, A., 2008, *"Searching for Narrow Emission Lines in X-ray Spectra: Computation and Methods"*, doi:10.1086/591631

Protassov, R., van Dyk, D. A., Connors, A., Kashyap, V. L., and Siemiginowska, A. 2002 *"Statistics, Handle with Care: Detecting Multiple Model Components with the Likelihood Ratio Test"*, doi:10.1086/339856.

van Dyk, D. A., Connors, A., Kashyap, V. L., and Siemiginowska, A., *"Analysis of Energy Spectra with Low Photon Counts via Bayesian Posterior Simulation"*, doi:10.1086/318656.

# Thank you CHASC!


David Van Dyk   Imperial College London
Xiao-Li Meng   Harvard Statistics
Vinay Kashyap   CfA



International CHASC Astro-Statistics Collaboration

This page lists resources of specific interest to astronomers. For detailed descriptions and reports of C-BAS/ICHASC activities, see www2.imperial.ac.uk/~dvandyk/astrostat.php

Software | Activities | Bibliography | Astro jargon | Stat jargon | People | Mailing-List | Internal

astrostat-announce GoogleGroup | GoogleCalendar | AstroStat Slog Archive


- **Chandra X-ray Center** NAS8-03060
- **NSF** Division of Mathematical Sciences 15-3492 15-3484  15-3546 18-11308 18-11083 18-11661
- **NASA** APRA 80-NSSC21-K0285