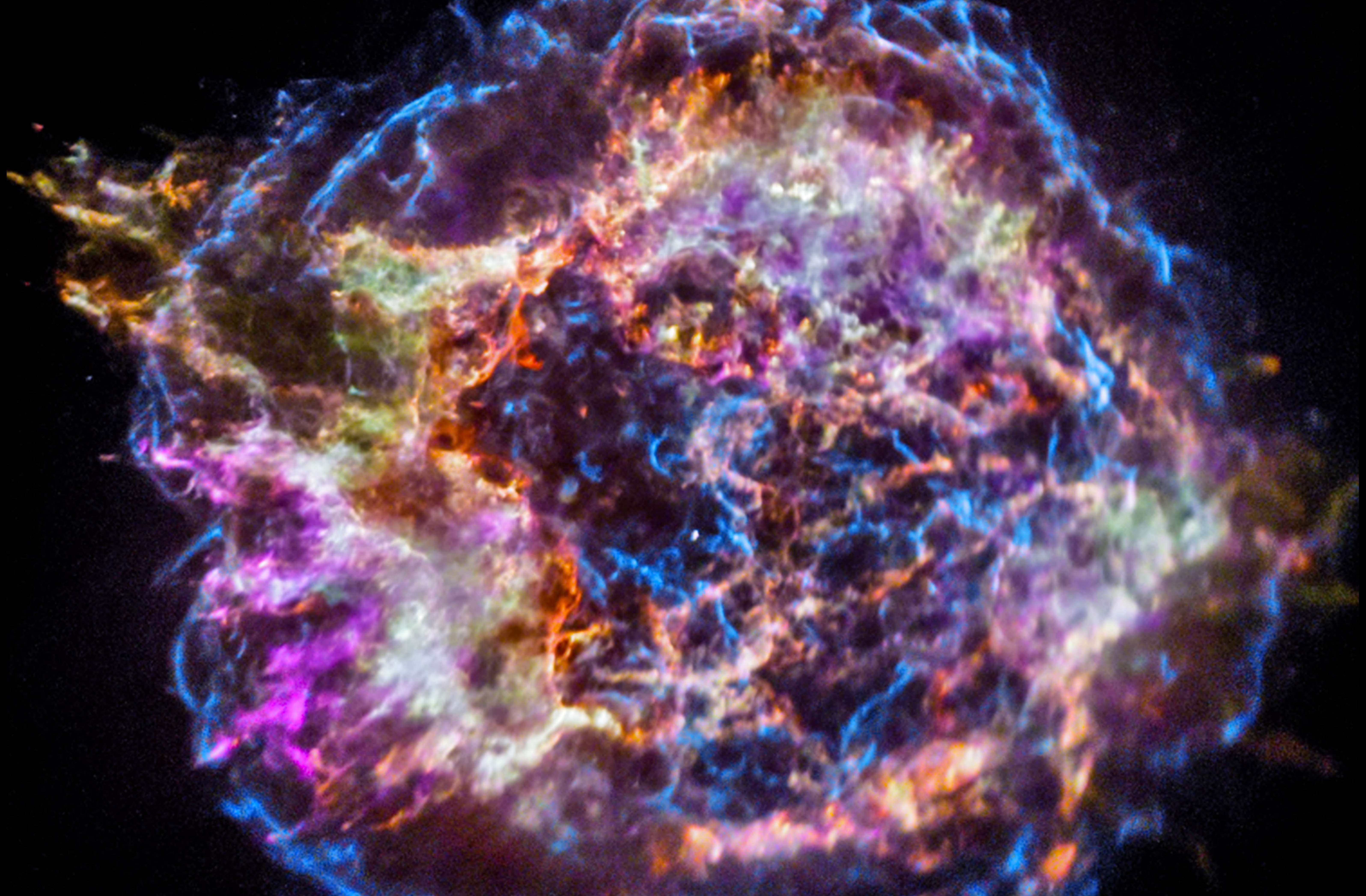
CHASC



Statistical Methodology for High-Energy Astronomical Datasets

Aneta Siemiginowska

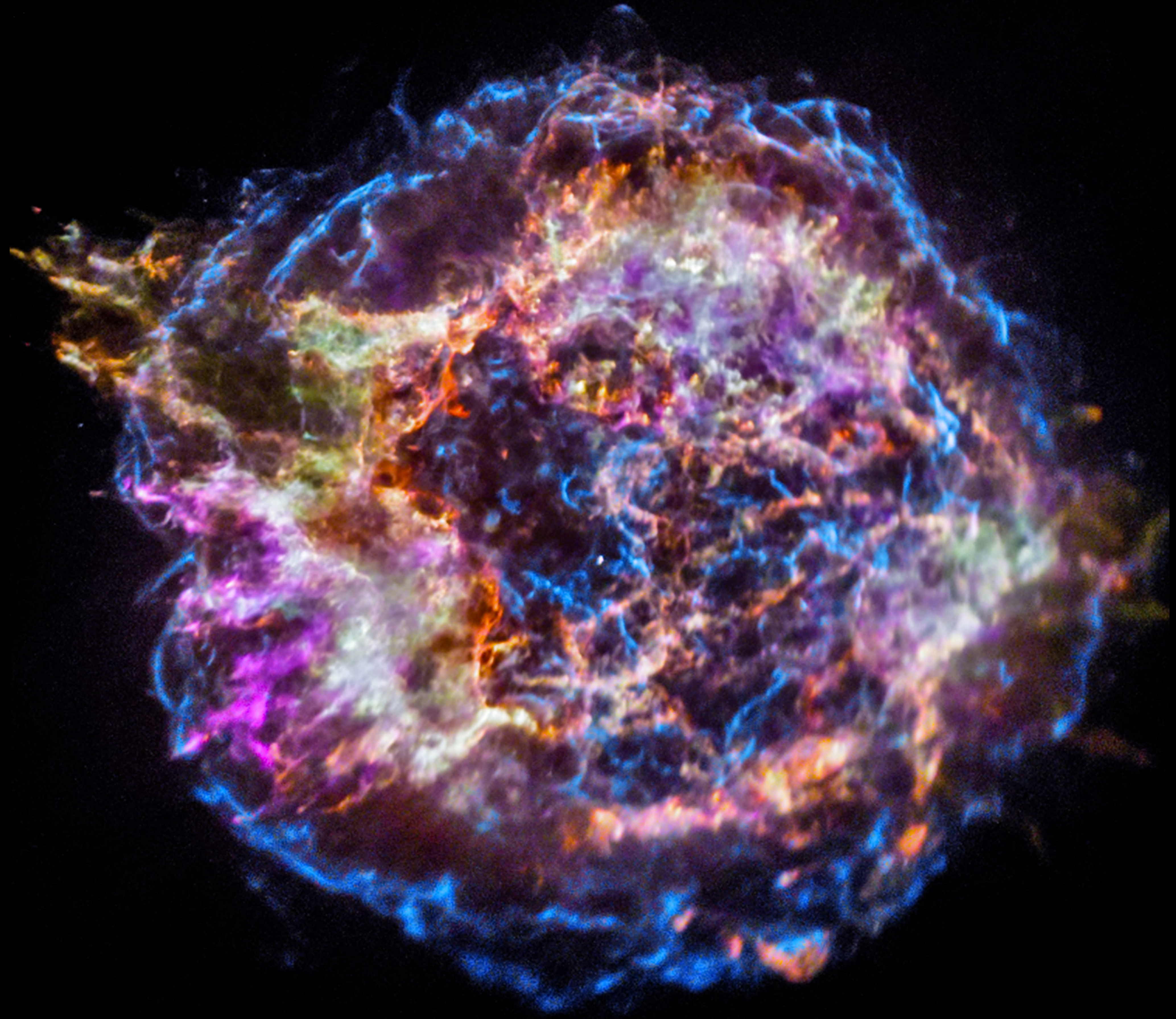
Chandra X-ray Center

CENTER FOR ASTROPHYSICS
HARVARD & SMITHSONIAN

Supernova Remnant Cassiopeia A

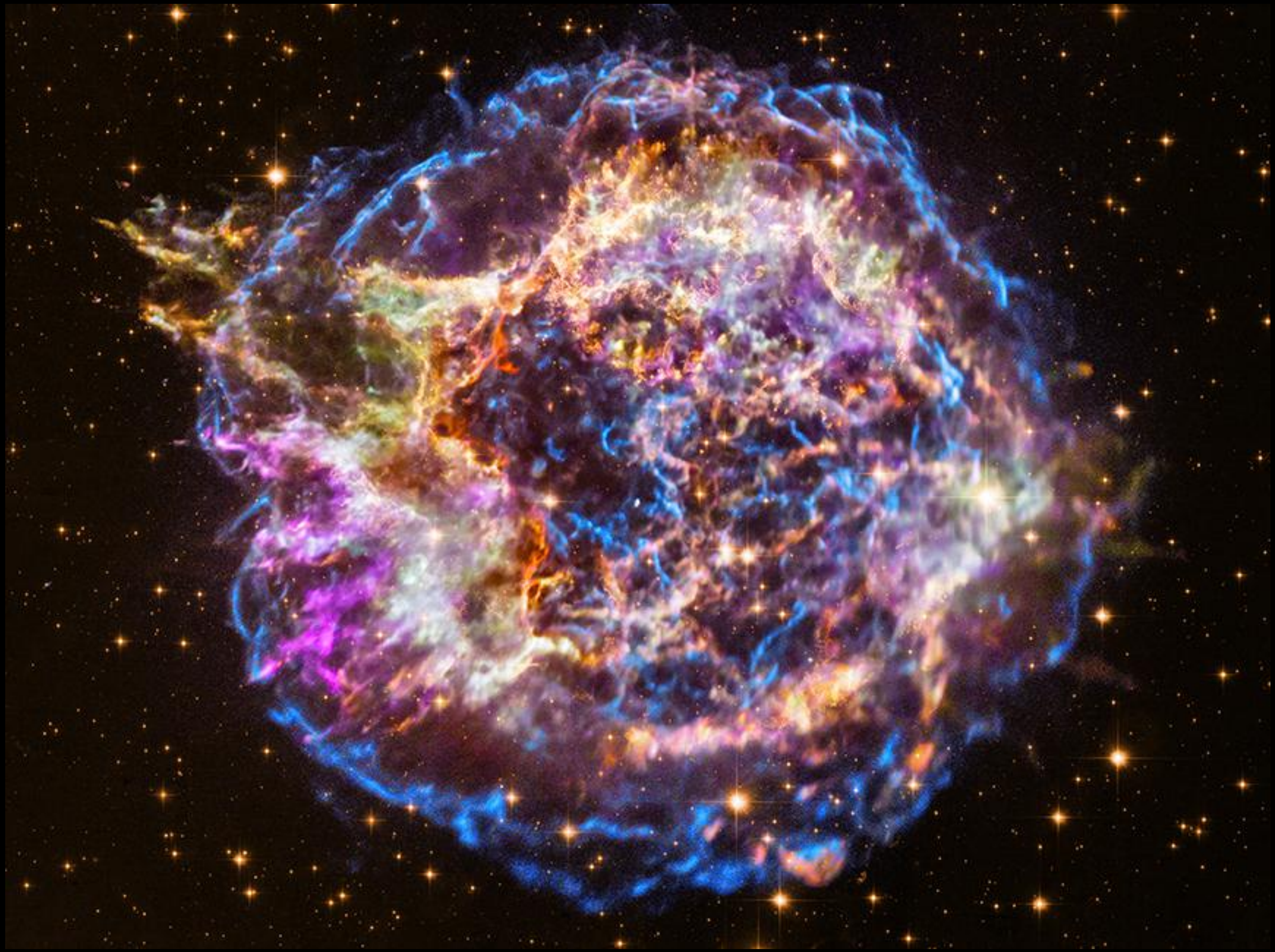
Visible Optical Light

'Non-visible' X-ray Light



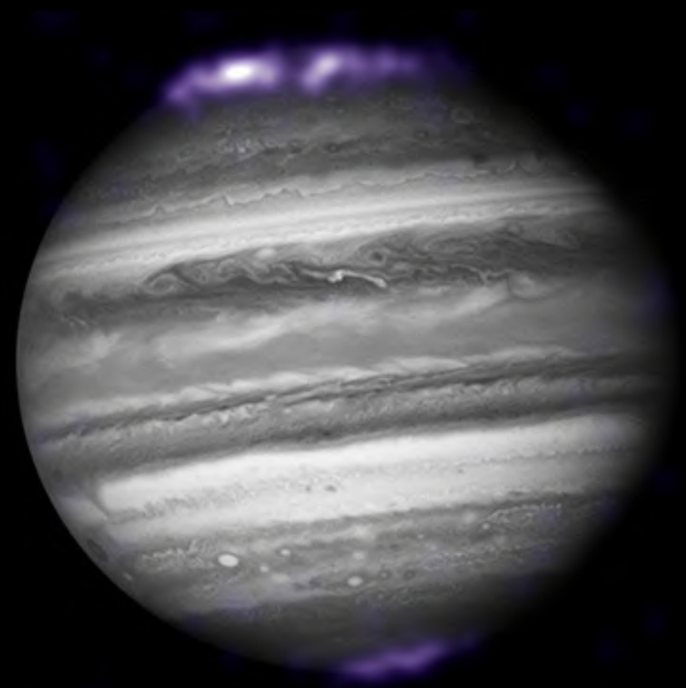
Supernova Remnant Cassiopeia A

Optical and X-ray Light



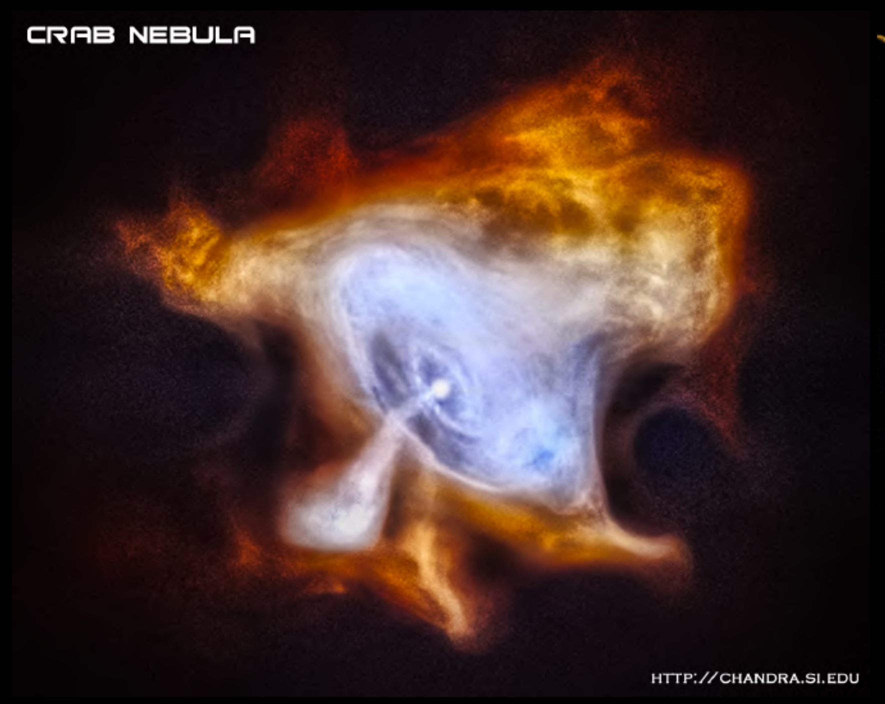
X-ray Universe

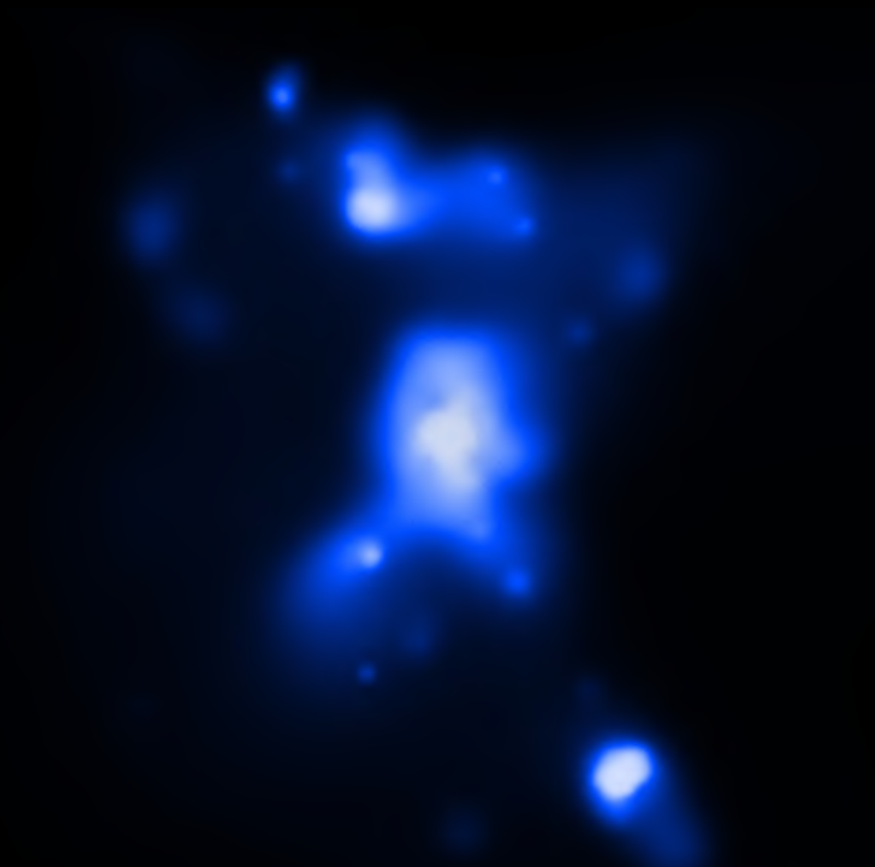
Solar System



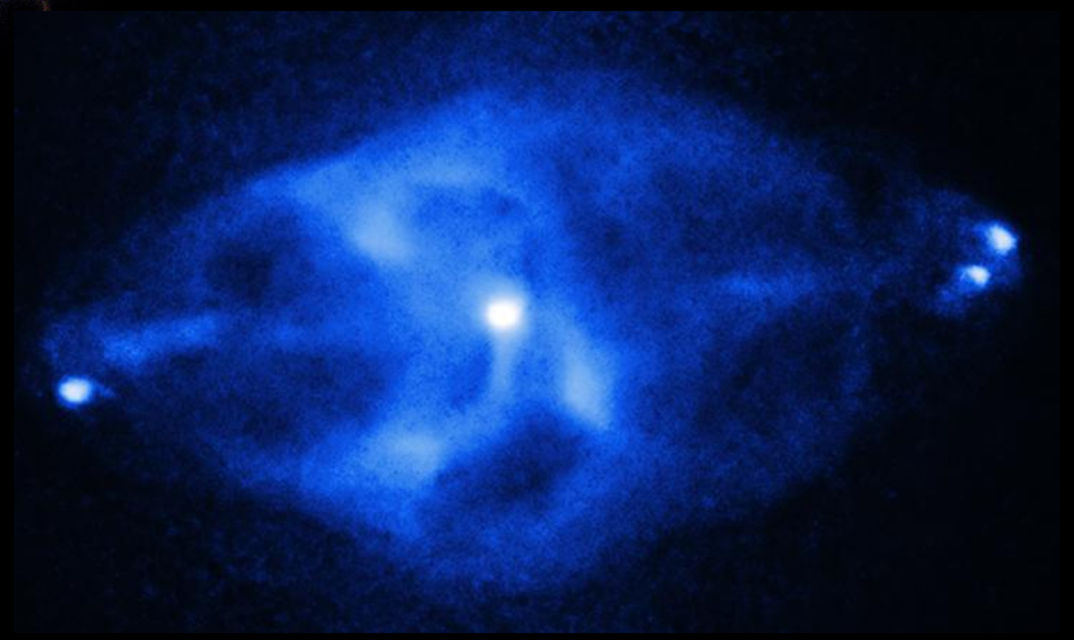
Hot gas > 10^5 K
Energetic particles

Supernova Remnants

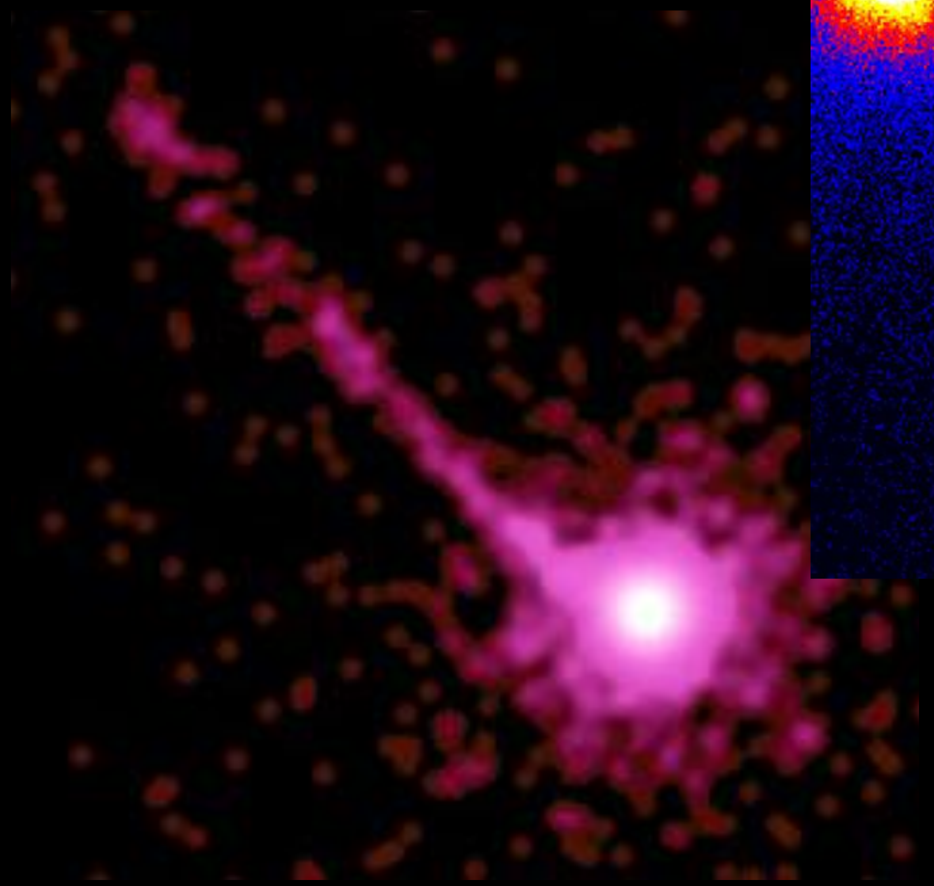


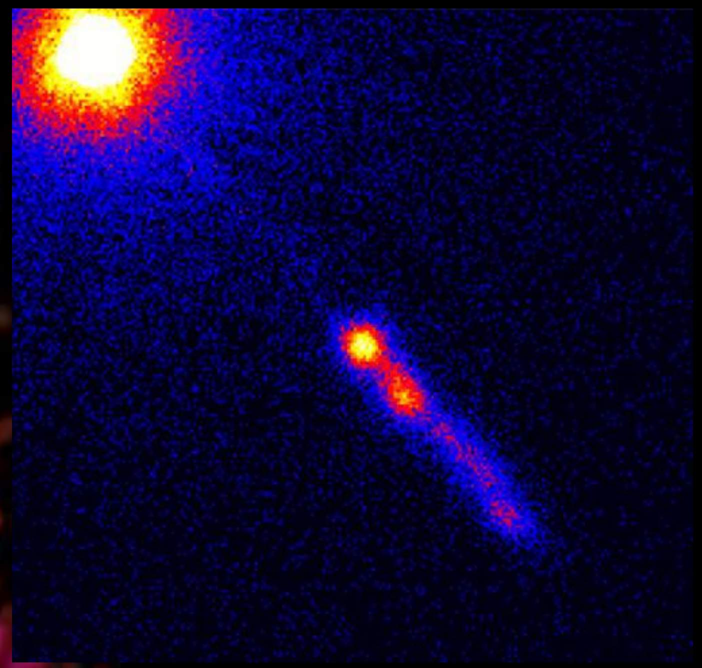


Radio Galaxies



Quasar Jets





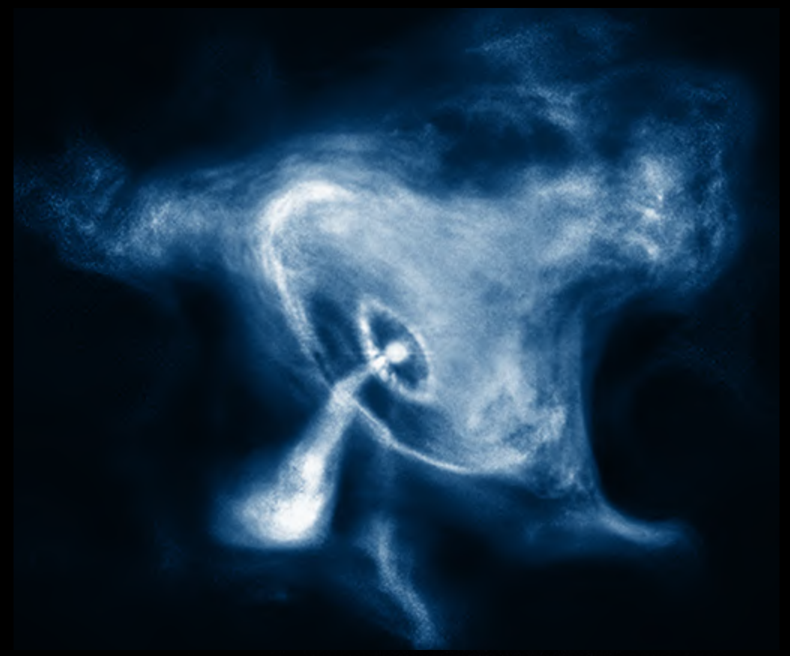
Clusters of Galaxies

X-ray Images obtained with the Chandra X-ray Observatory
(False Color)

Outline

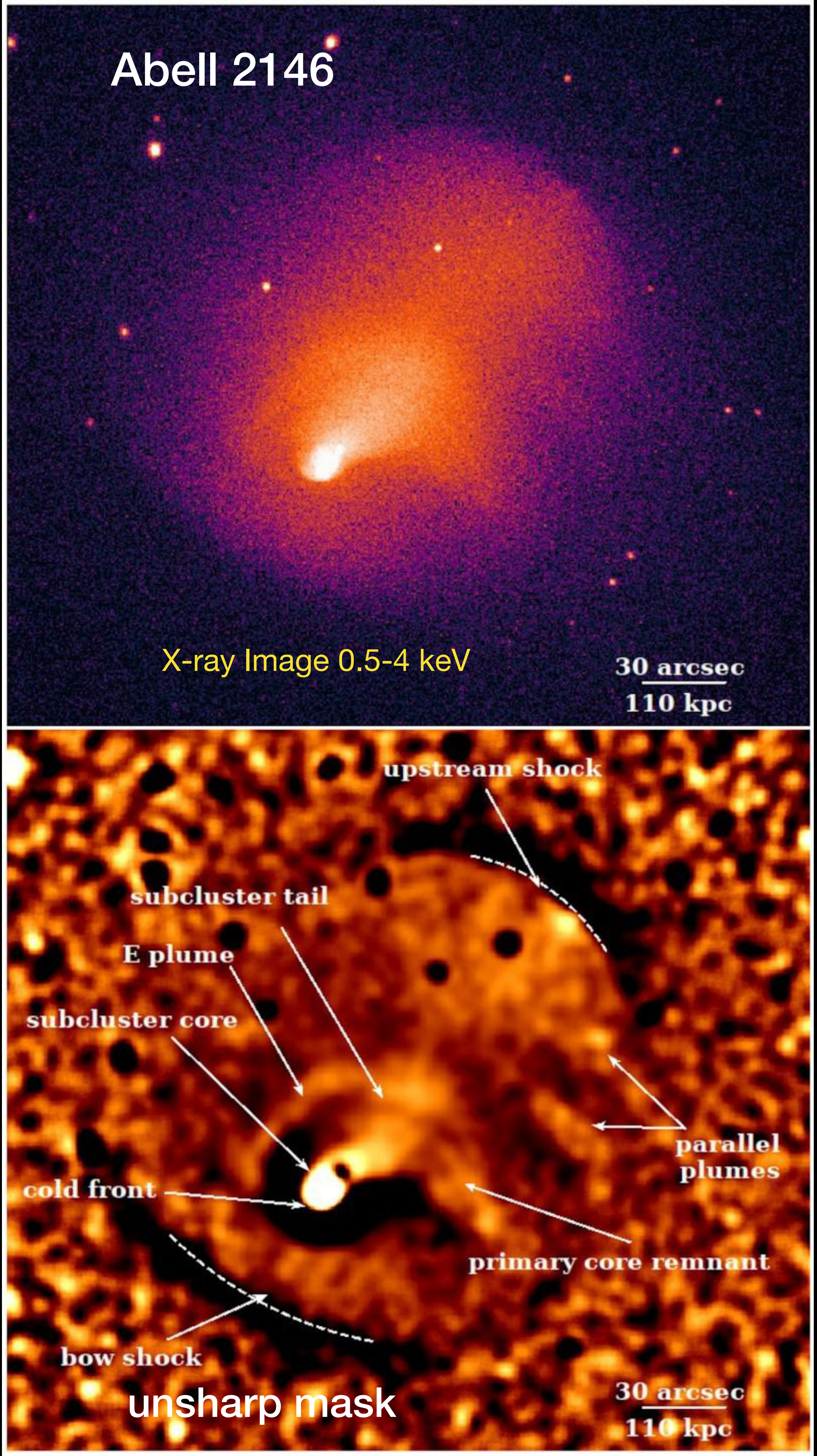
- Scientific measurements and X-ray Data
- Single Domain Methods
- Multi-Domain Methods

Scientific Measurements

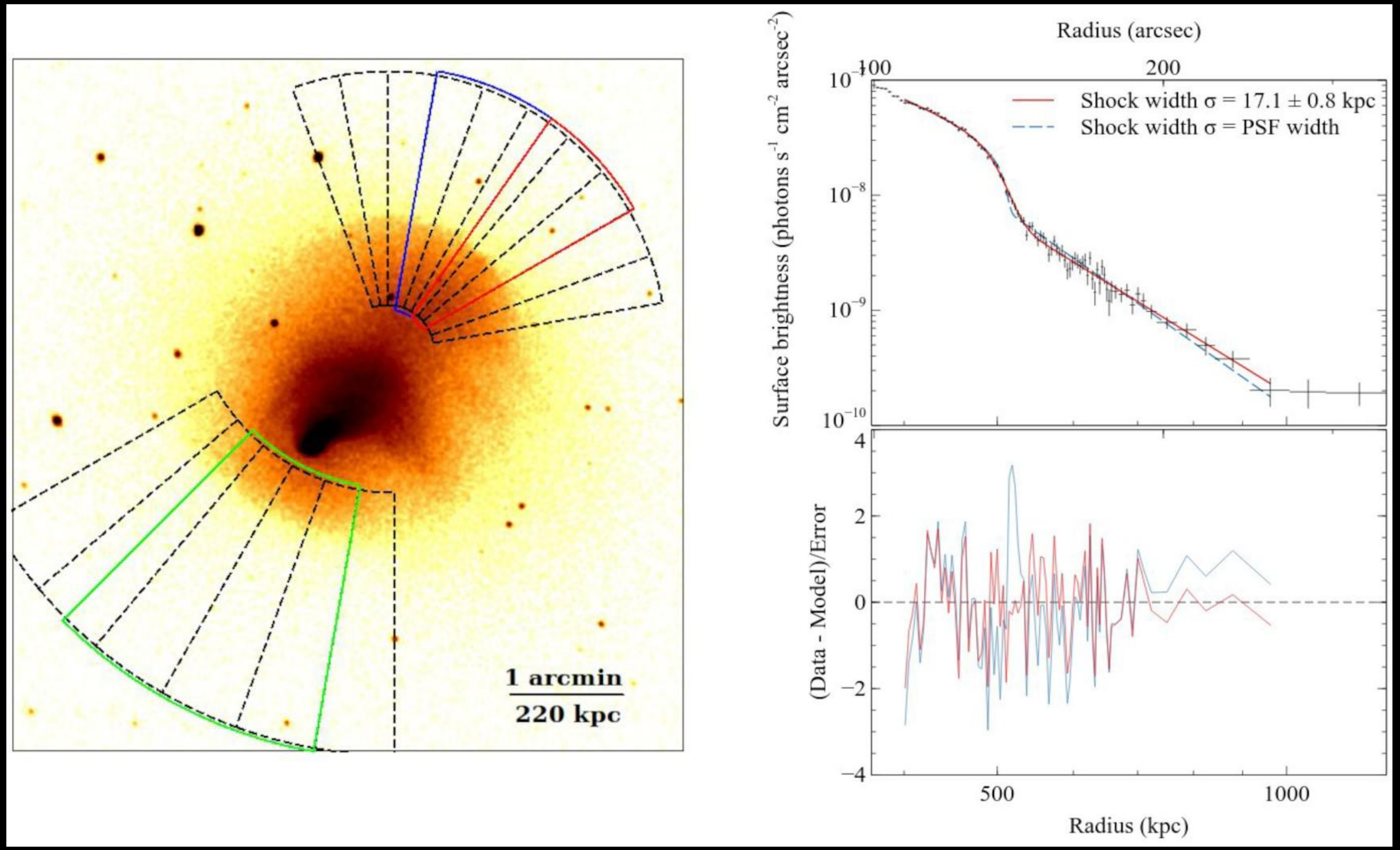


Measurements	Examples	Current Methods	Limitations
Morphology	point source, diffuse structures, filaments	detection algorithm, smoothing, unsharp mask, deconvolution	sparse images, defining source boundaries, upper bounds, separate sources in crowded field, background uncertainties
Scale and Size	emission features, boundary, clusters, unresolved structures, mass	surface brightness profiles, extent, deconvolution, variability timescales, correlation between different bands	resolution, source boundaries, projection, low counts, domain specific, background features
Source Properties	flux, luminosity, temperature, abundance, density, obscuration, age	model fit, aperture photometry	averaging regions, boundaries, instrumental effects (e.g.pileup, dead time)
Population	members, intensity, identification, flux distributions	detection algorithms, hardness ratios, catalog matches, spectral modeling	uncertainties, sparse Poisson images, overlapping sources, background, no energy/time resolution

Scientific Measurements



Russell et al 2022



Morphology
Scale and Size
Source Properties
Populations

NOTE on some Data Issues and Source of Uncertainties:

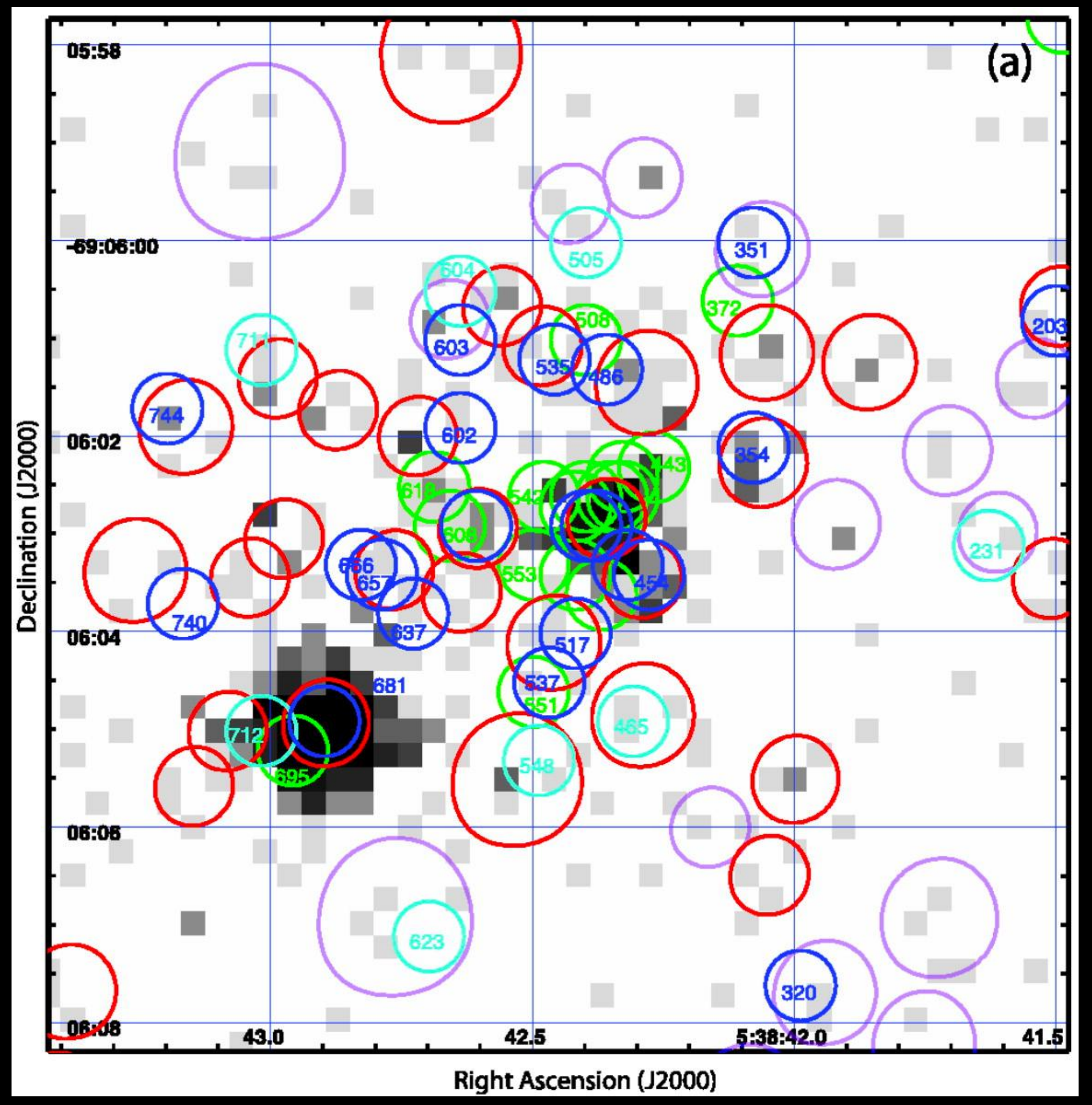
- combined multiple observations
- background level
- region selection
- model fit to the surface brightness profile
- PSF (blurring) impact on the measurements

Scientific Measurements

Star Cluster

Chandra X-ray image
Red: 0.5-2 keV
Blue: 2-7 keV

Townsley et al 2006



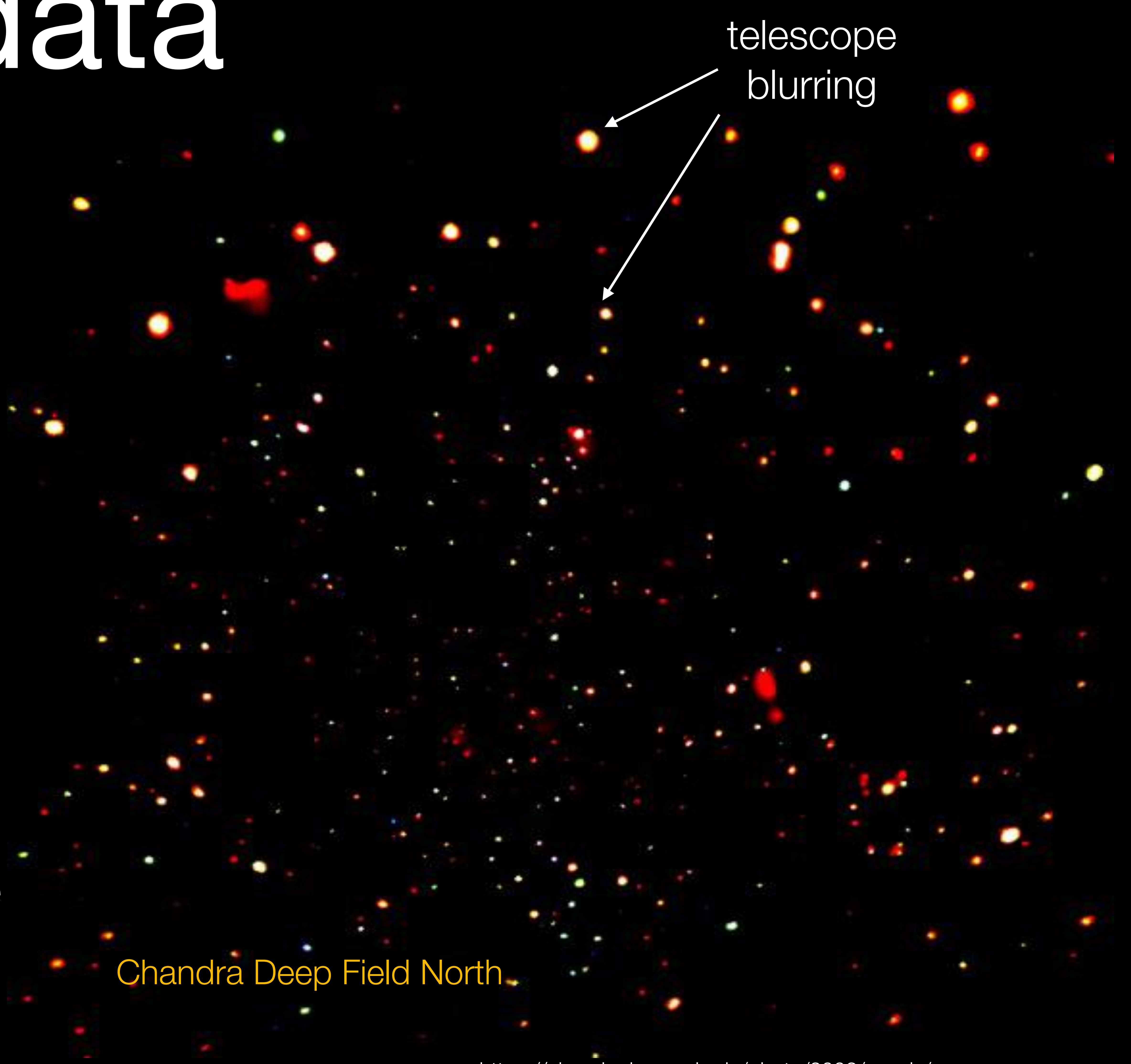
Morphology
Scale and Size
Source Properties
Populations

NOTE on some Data Issues and Source of Uncertainties:

- combined multiple observations
- background level
- region selection
  - PSF (blurring) impact on source detection
  - overlapping PSFs for source counts measurements

X-ray data

- Counting arriving photons (Poisson counts) - different from optical data
- For each photon location on the sky (x,y), arrival time (t) and energy (E) are recorded (x,y,t,E) - events
- X-ray observations take a long time - a short observation with Chandra X-ray Observatory lasts ~10 ksec (~3 hours) while typical observations take a day or more. The Chandra Deep Field observations took about **23 days**.



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

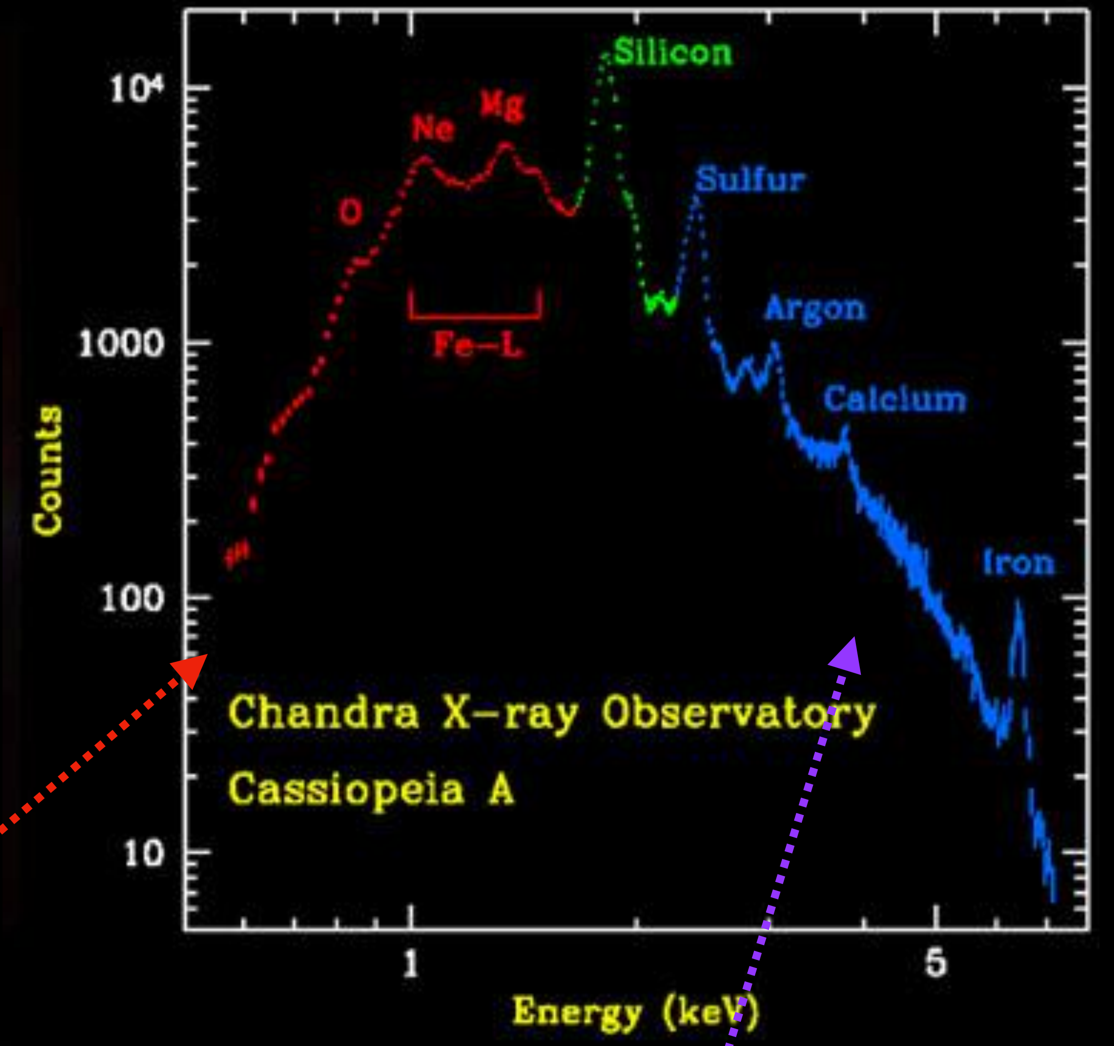
The faintest sources - one X-ray photon every 4 days!

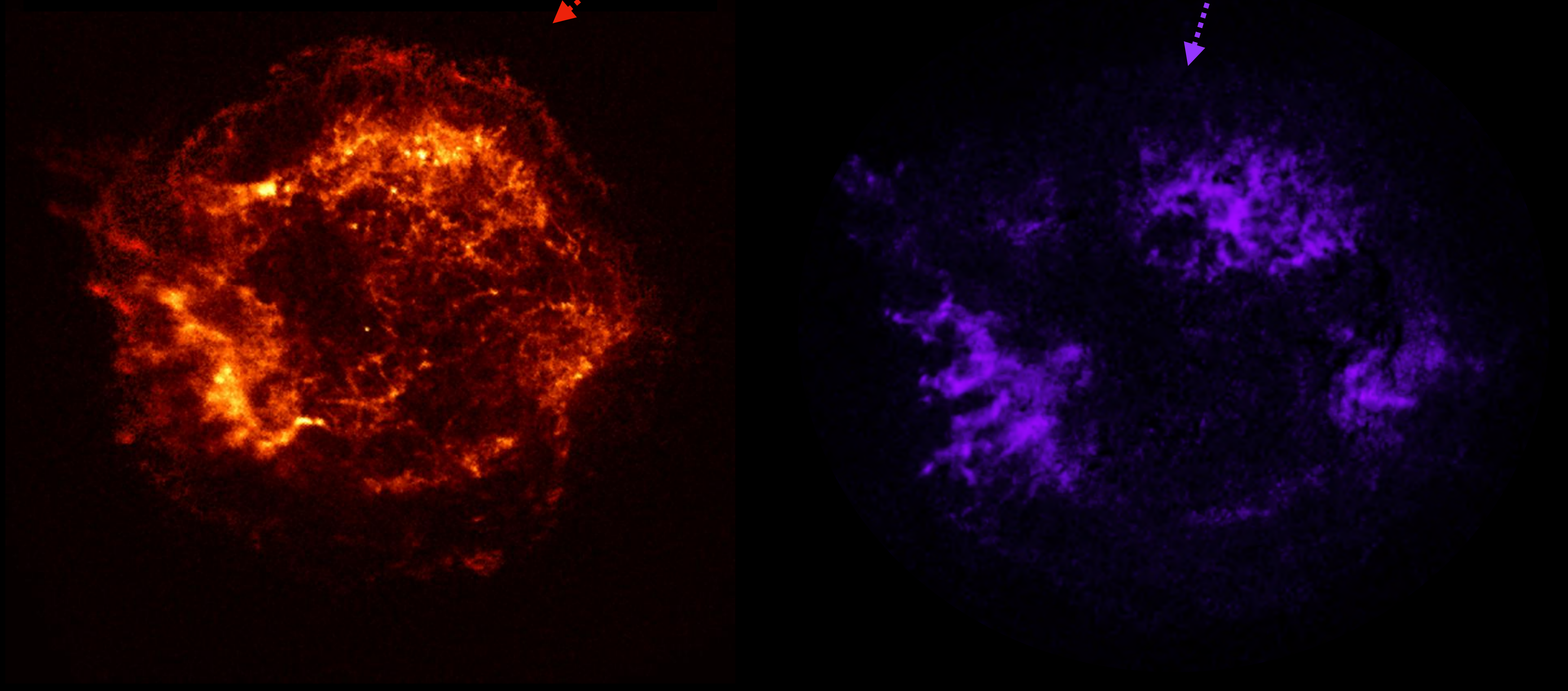
X-ray Analysis Single Domains

Event $e_i = (x_i, y_i, t_i, E_i)$

- X-ray image is made by binning events into images, e.g. accumulating photons in a selected energy band and fixed exposure time: $e_i(x,y) = \int e(x,y,t,E)\, dE\, dt$
  - no spectral or temporal information
  - analysis require a point spread function
- Energy Spectrum for selected regions are generated by binning the events in energy: $e_i(E) = \int e(x,y,t,E)\, d(x,y)\, dt$
  - no spatial or temporal information
  - require additional calibration files
- Lightcurve - time series for selected region and energy band binning the events in time: $e_i(t) = \int e(x,y,t,E) d(x,y) d(E)$
  - no spatial or energy information

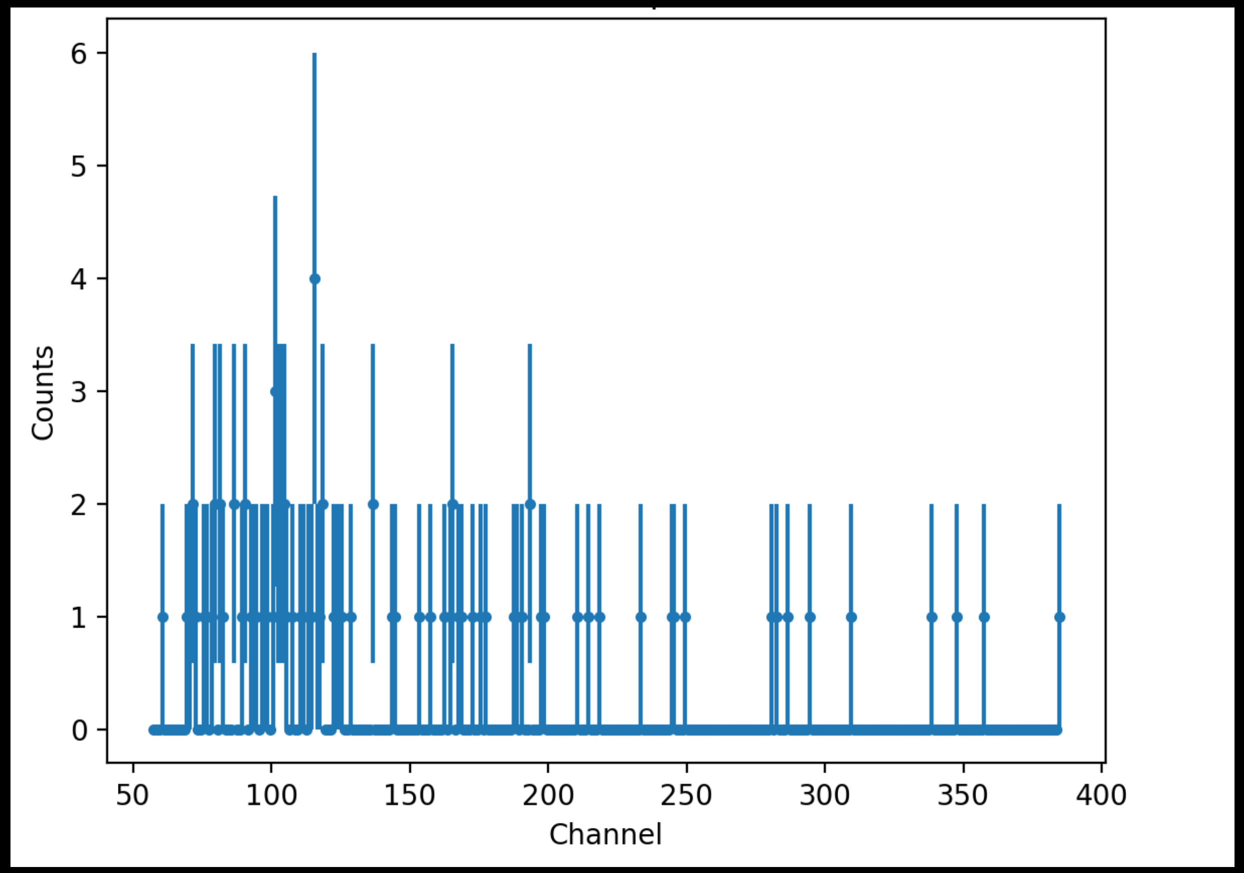
Cassiopeia A Supernova Remnant





X-ray Energy Spectra

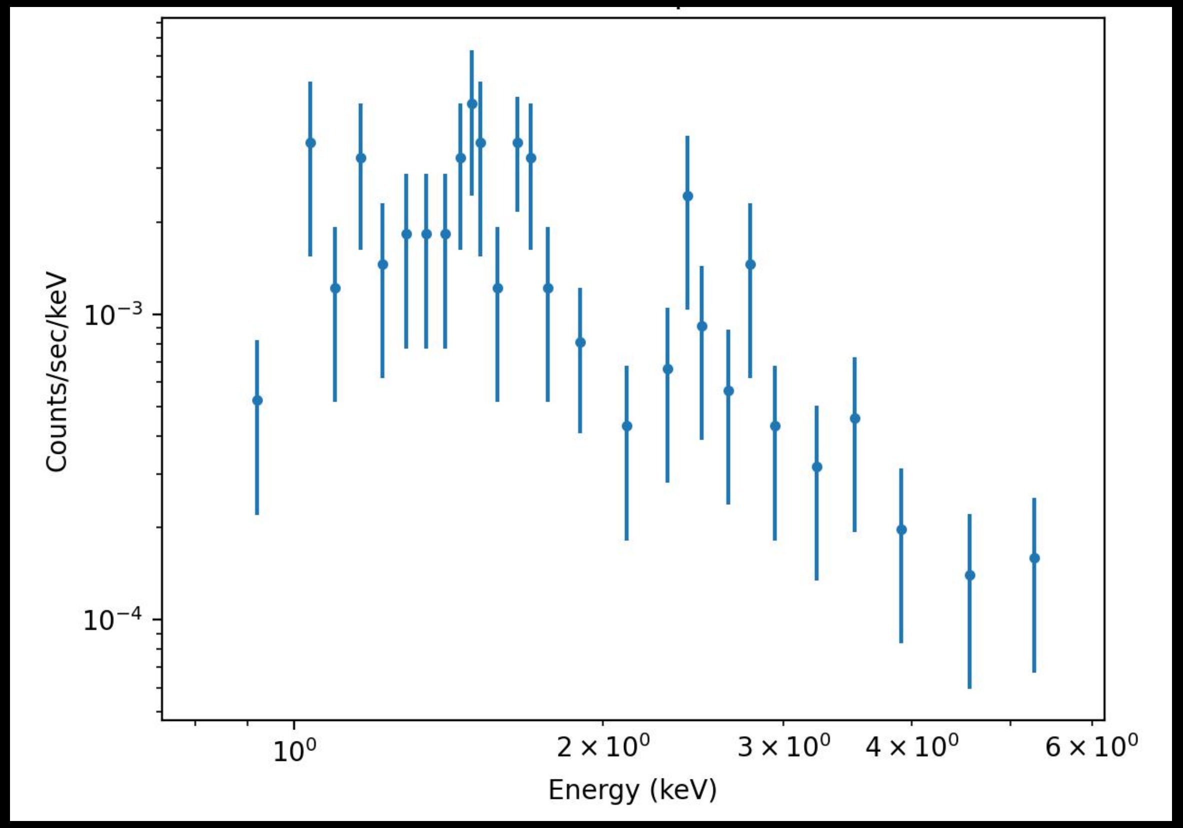
- Model fitting:
  - Includes instrument response directly -> calibration impact on the results $\text{Counts}\,(i) = \int R(i,E)\,A(E)\;M(E)\,dE$
  - Non-linear astrophysical models, computer generated models
  - Appropriate fit statistics, no binning/grouping data, no background subtraction
  - Modification 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
- Issues:
  - **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)
  - see Humphrey et al 2009, Siemiginowska 2011, Kastra 2017, Bonamente 2023



channel

keV

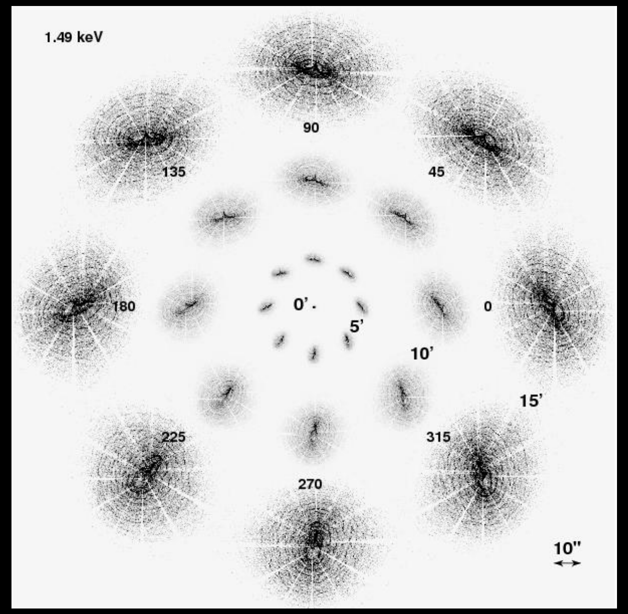
convert matrix (RMF)



log(Energy) [keV]

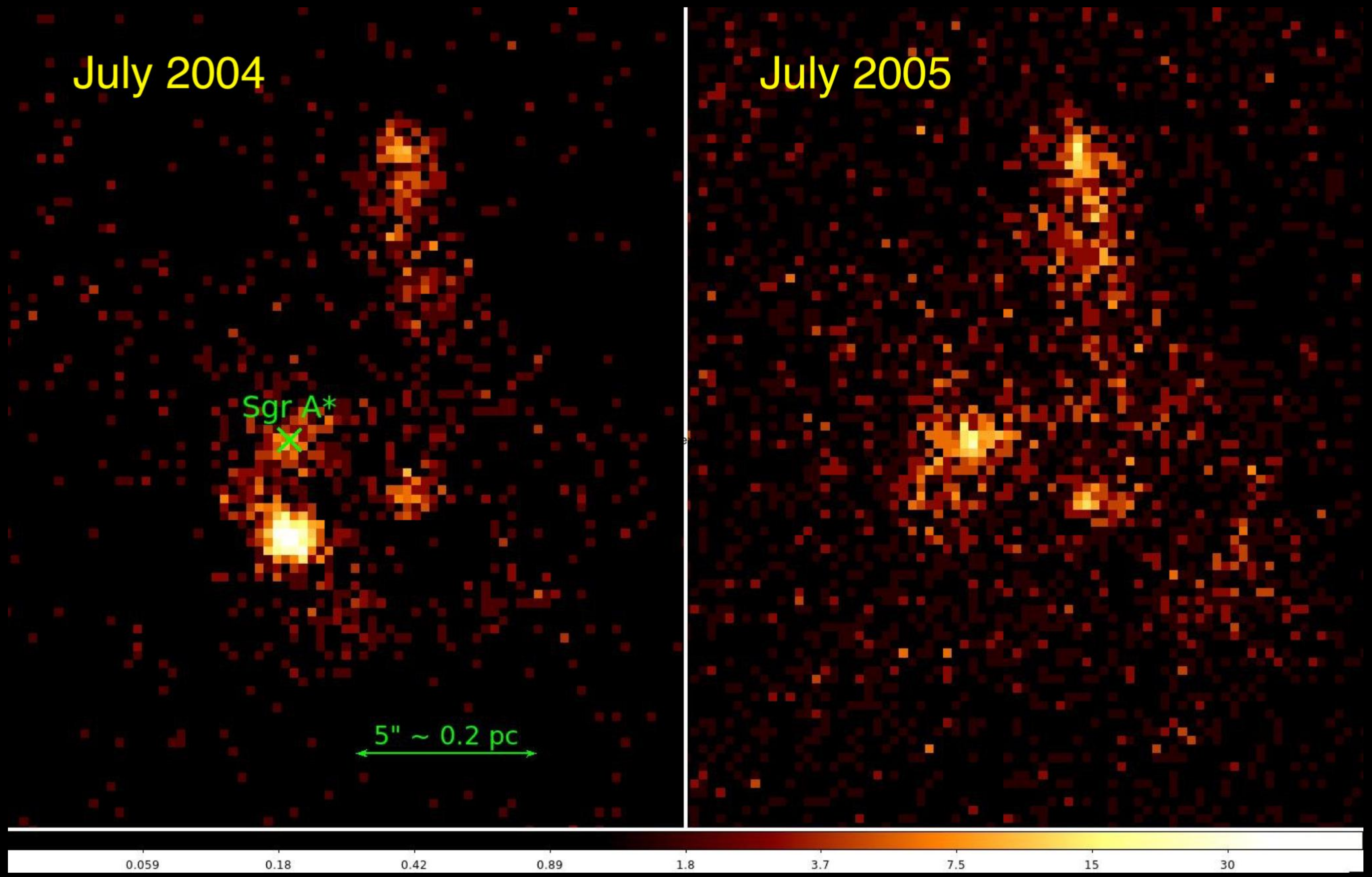
X-ray Images

- Chandra X-ray Observatory takes the highest angular 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

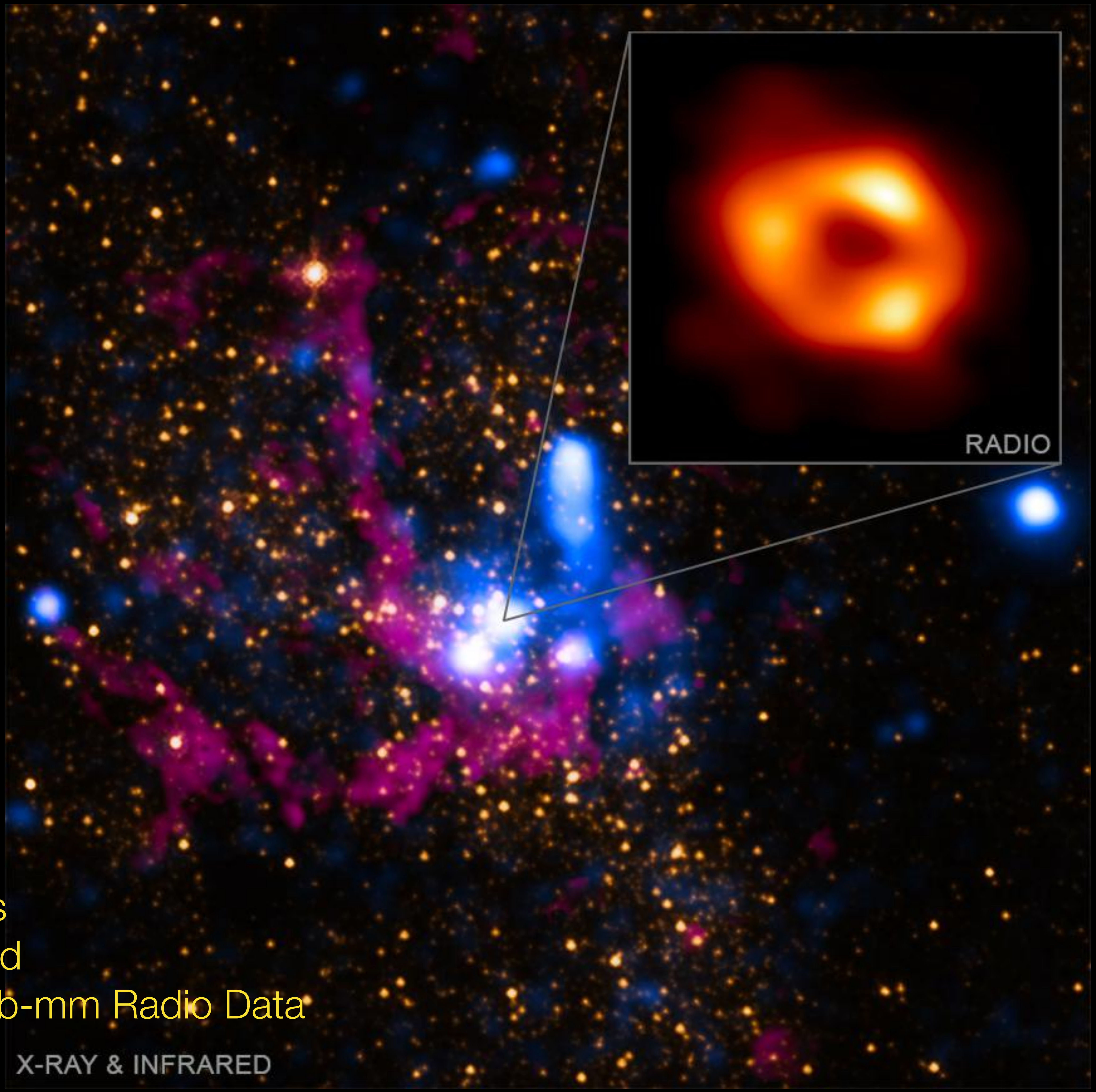


- Some issues:
  - detection of features and upper limits
  - detecting and identifying low surface brightness structures
  - resolving source in crowded fields - overlapping sources, diffuse emission
  - finding source boundaries
  - PSF uncertainties

X-ray Images of the Galactic Center

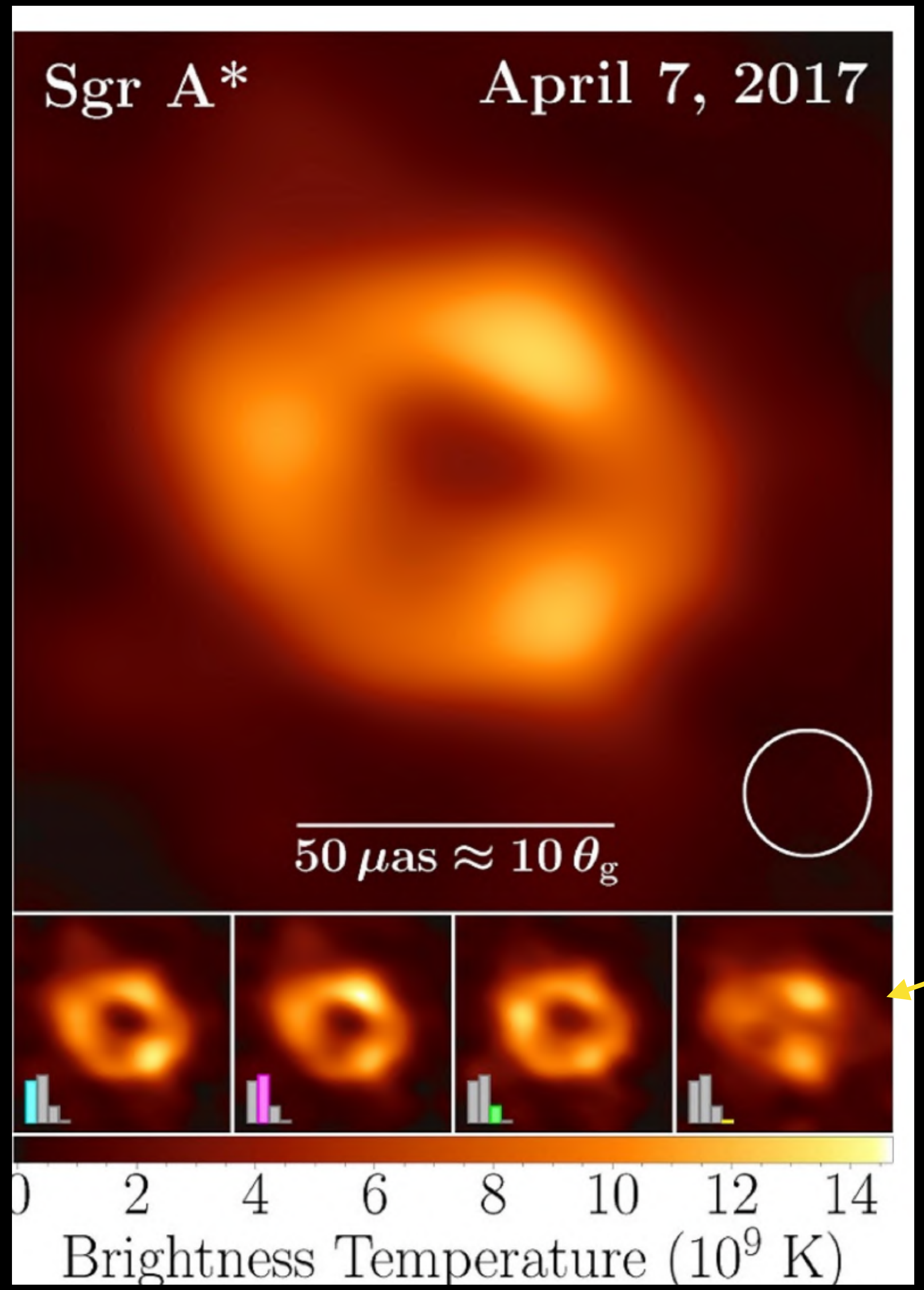


Multi-Band View of the Galactic Center



Blue: X-rays
Red: Infrared
Orange: Sub-mm Radio Data

BH Event Horizon



different solutions

Bower and EHT collaboration, ApJ 2022

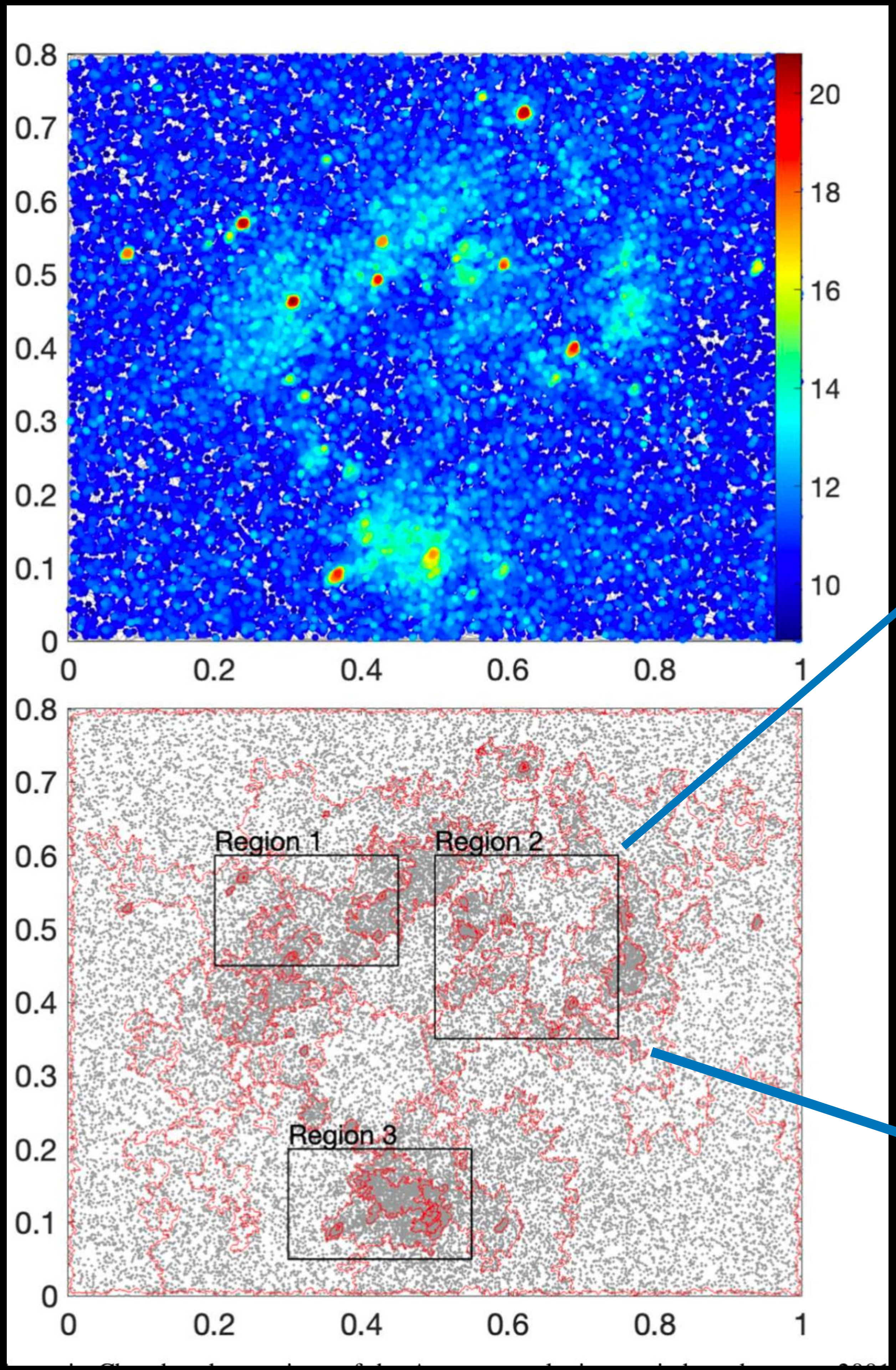
Single Domain Analysis

Analysis Domain	Description	Standard Methods	Challenges	Modern Methods
Spectra $e(E) = \int e(x, y, t, E)dx\,dy\,dt$	only energy, loss of time and morphology	forward fitting, multi-spectra, Poisson likelihood, model and instrument uncertainties	non-linear complex models, high resolution spectra, uncertainties in physical process & models	Bayesian Methods, Simulations, bootstrap, Likelihood free modeling, hierarchical Baysian models, model selection via ppp, BIC, AIC, ML
Image $e(x, y) = \int e(x, y, t, E)dE\,dt$	only location and morphology, loss of energy and time	source detections, morphology, contours, image reconstruction, deconvolution	faint structures, source boundaries, upper limits, crowded sources, background	Bayesian reconstruction, simulation for upper bounds, image segmentation
Time variability $e(t) = \int e(x, y, t, E)dx\,dy\,dE$	only time, loss of energy and source morphology	differences image/spectra, power spectra, periodogram, Bayesian Blocks, flares	S/N limitation on time resolution, break points, uneven sampling, non-detections	direct modeling of light curves (O-U, CARMA), periodograms, cross-spectra, flare detection

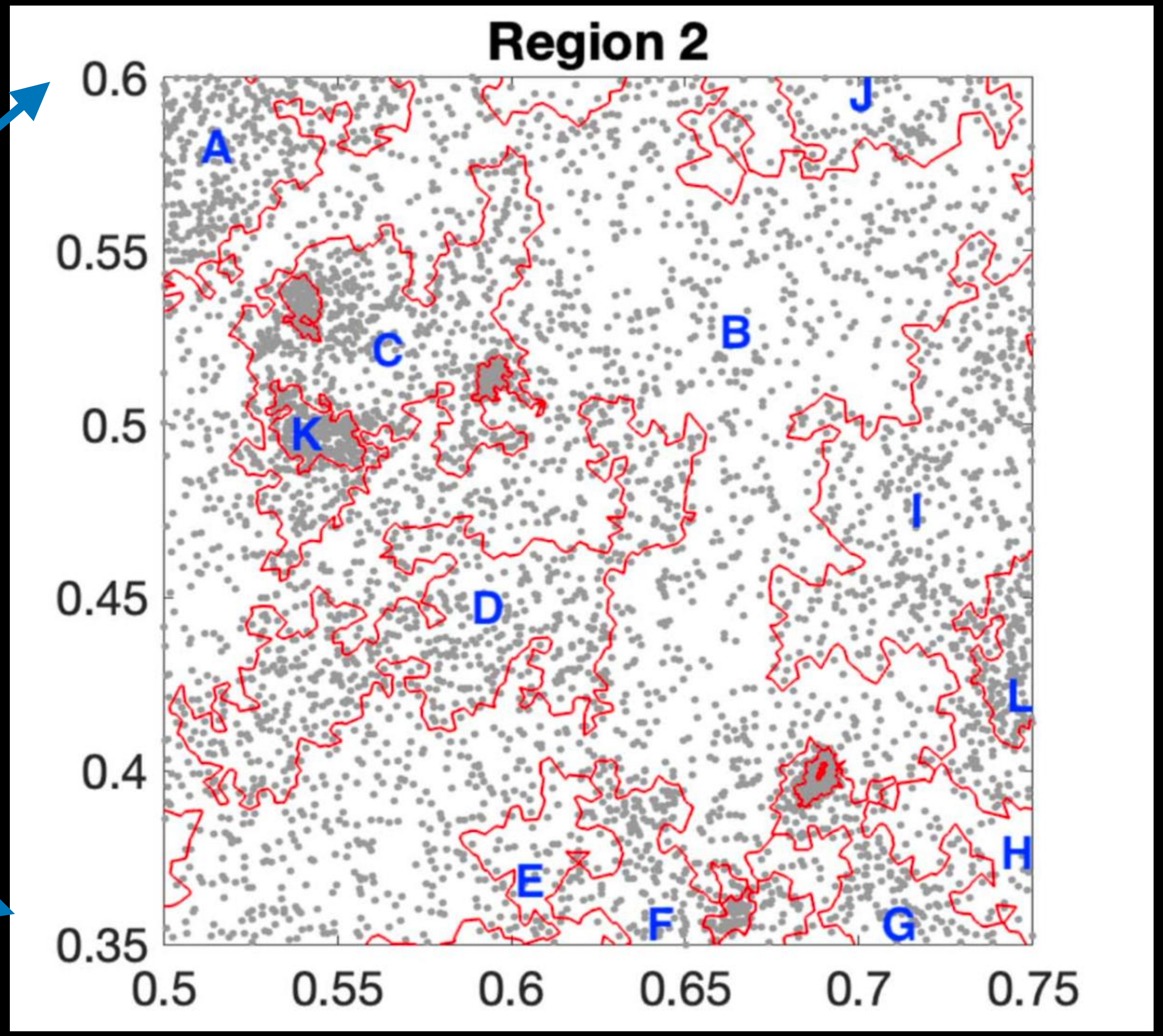
New Methods for Single Domain Analysis

Crowded Fields - Finding structures of diffuse emission

Large scale > PSF



Chandra data
Antennae galaxies



SRGonG
Region Growing on Graphs

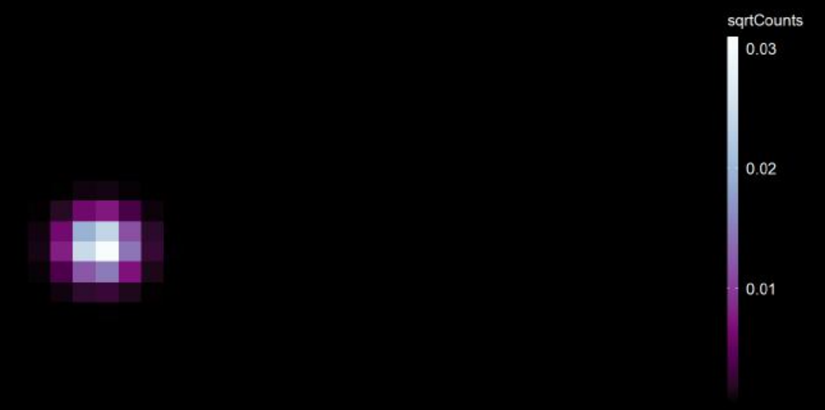
Fan et al 2023

- Non-binned images - direct use of photons
- Voronoi tessellation of the photon locations
- seeded region growing to grow segments
- over-segmented regions coalesce using greedy algorithm:
  - adjacent segments are merged
  - to minimize a model comparison statistic
  - Bayesian Information Criteria

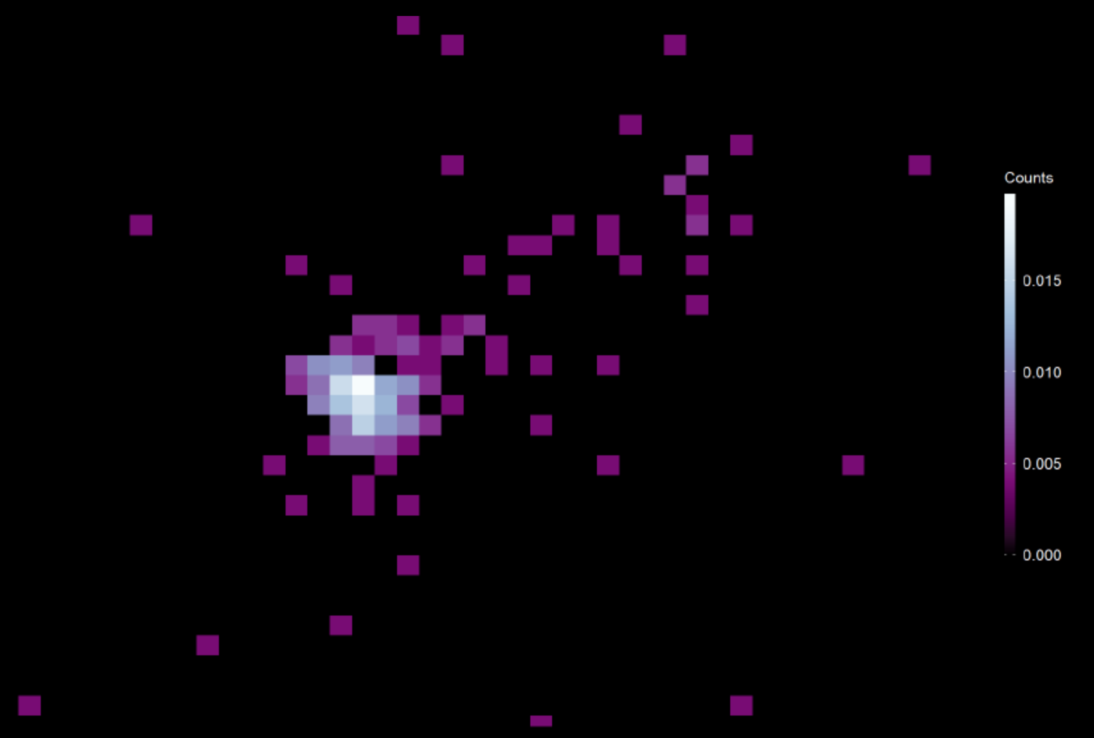
New Methods for Single Domain Analysis

Low counts X-ray Image

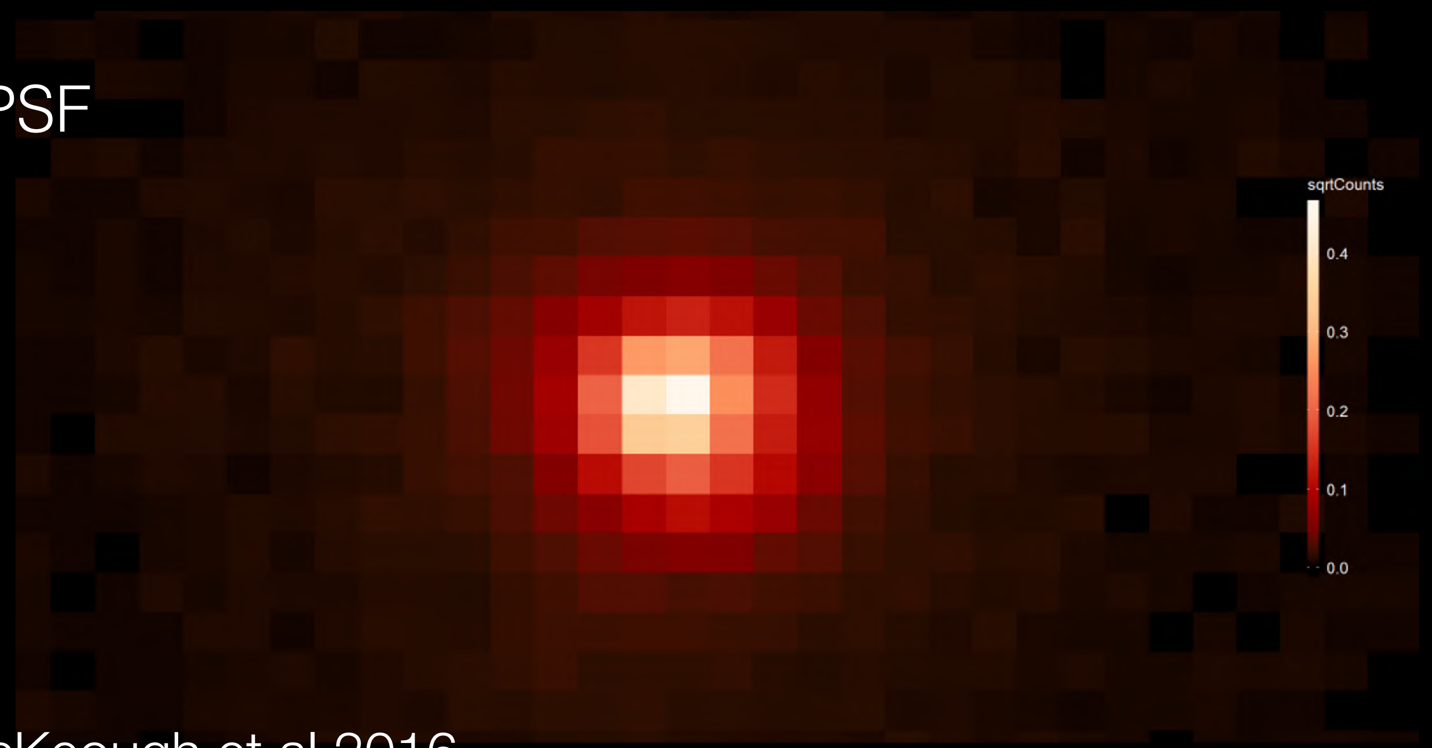
Point Source + Bkg
Baseline Model



Chandra Data



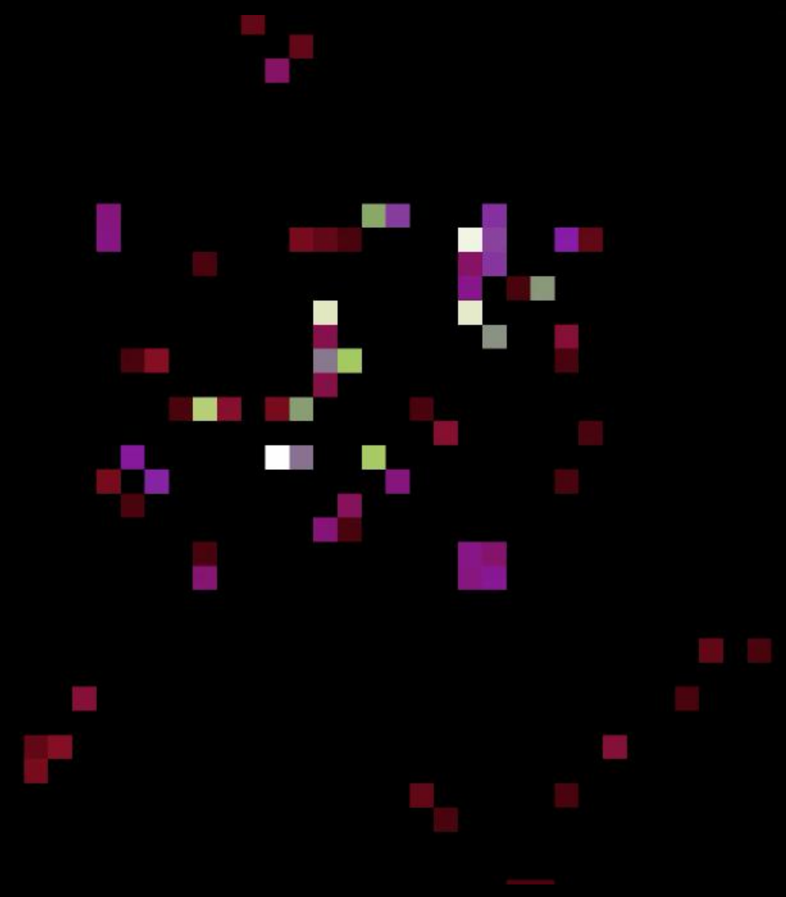
PSF



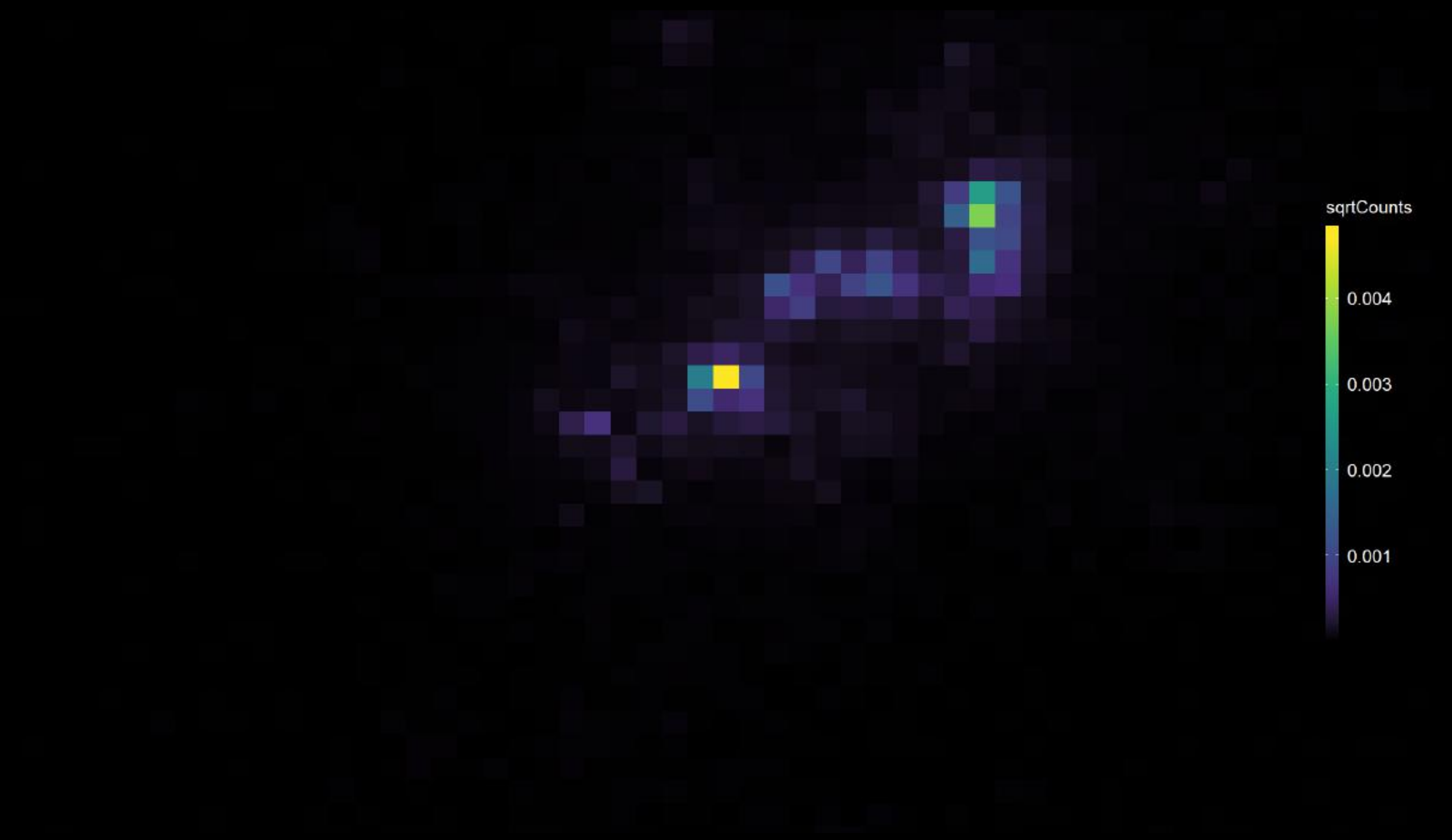
McKeough et al 2016

Posterior Draws with MCMC

Expected photon counts in each pixel
given the observed counts



Posterior Mean



LIRA - Low-counts Reconstruction and Analysis

Esch et al 2004; Connors & Van Dyk 2007;
Stein et al 2015; McKeough et al. 2016; Donath et al. 2022

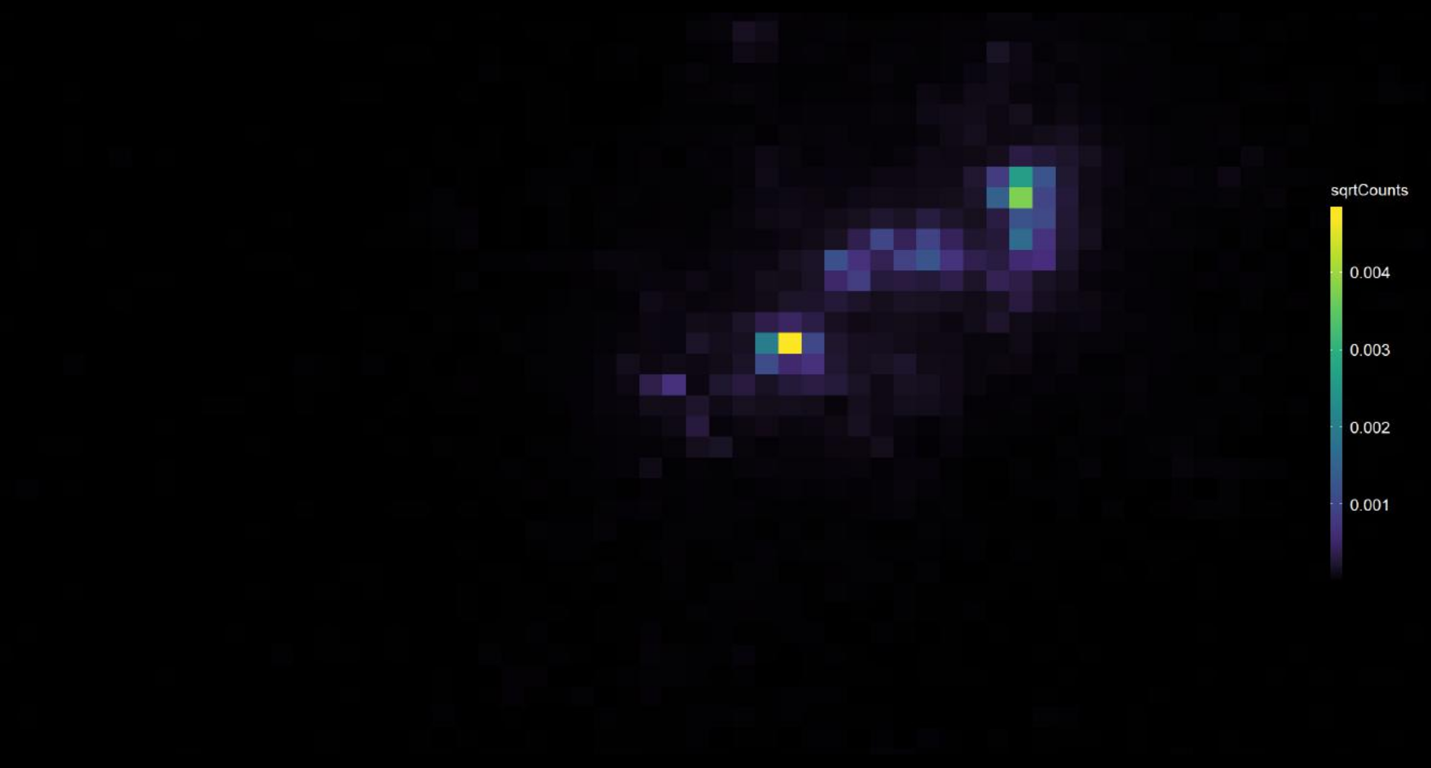
https://github.com/astrostat/LIRA
https://github.com/astrostat/pylira

New Methods for Single Domain Analysis

Finding the source boundary

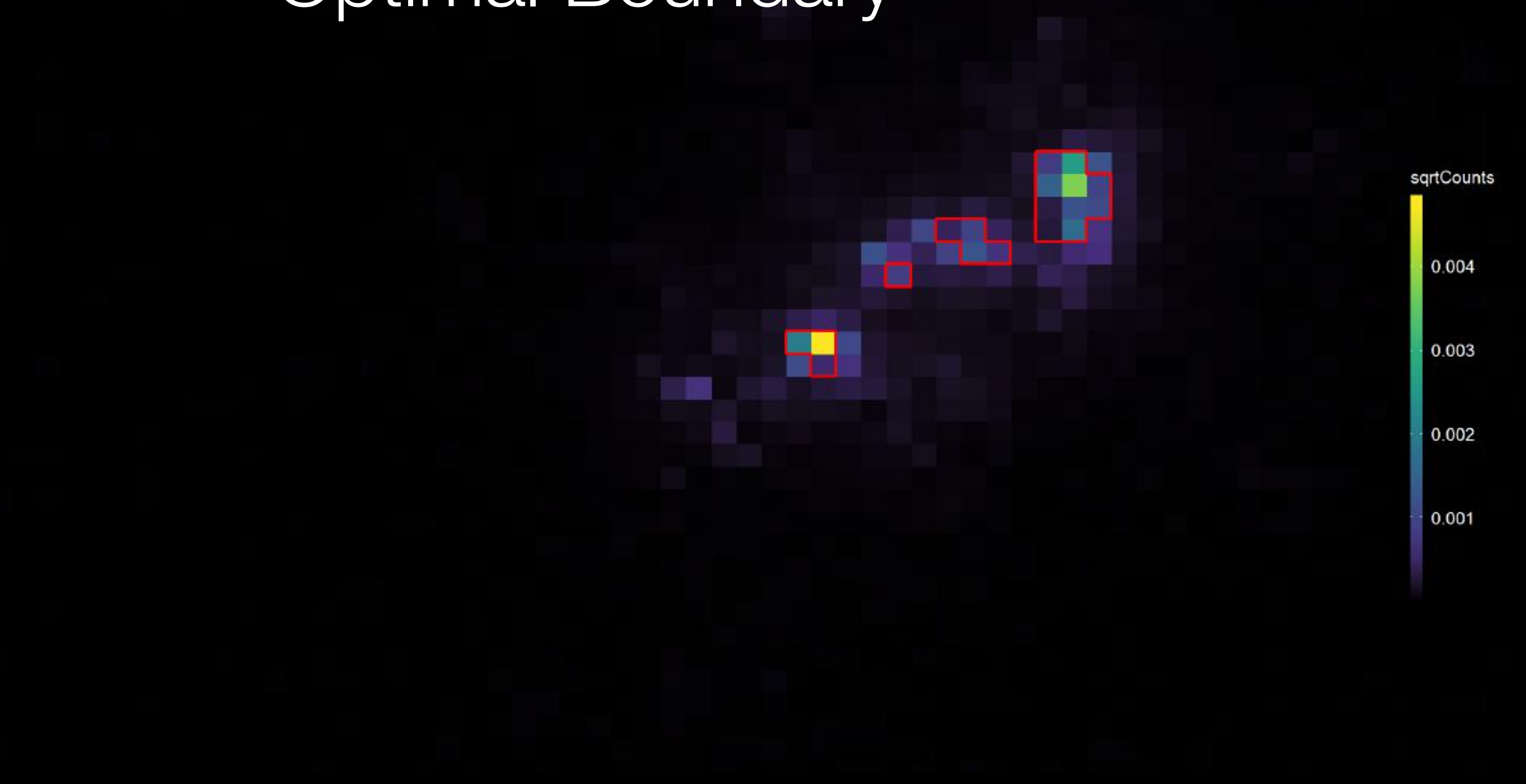
Posterior Draws with MCMC
probability distribution of pixel assignments

Posterior Mean



ISING Prior
Correlation between neighboring pixels

Optimal Boundary



Boundary with maximum probability
given LIRA-Ising posterior

Katy McKeough PhD Thesis

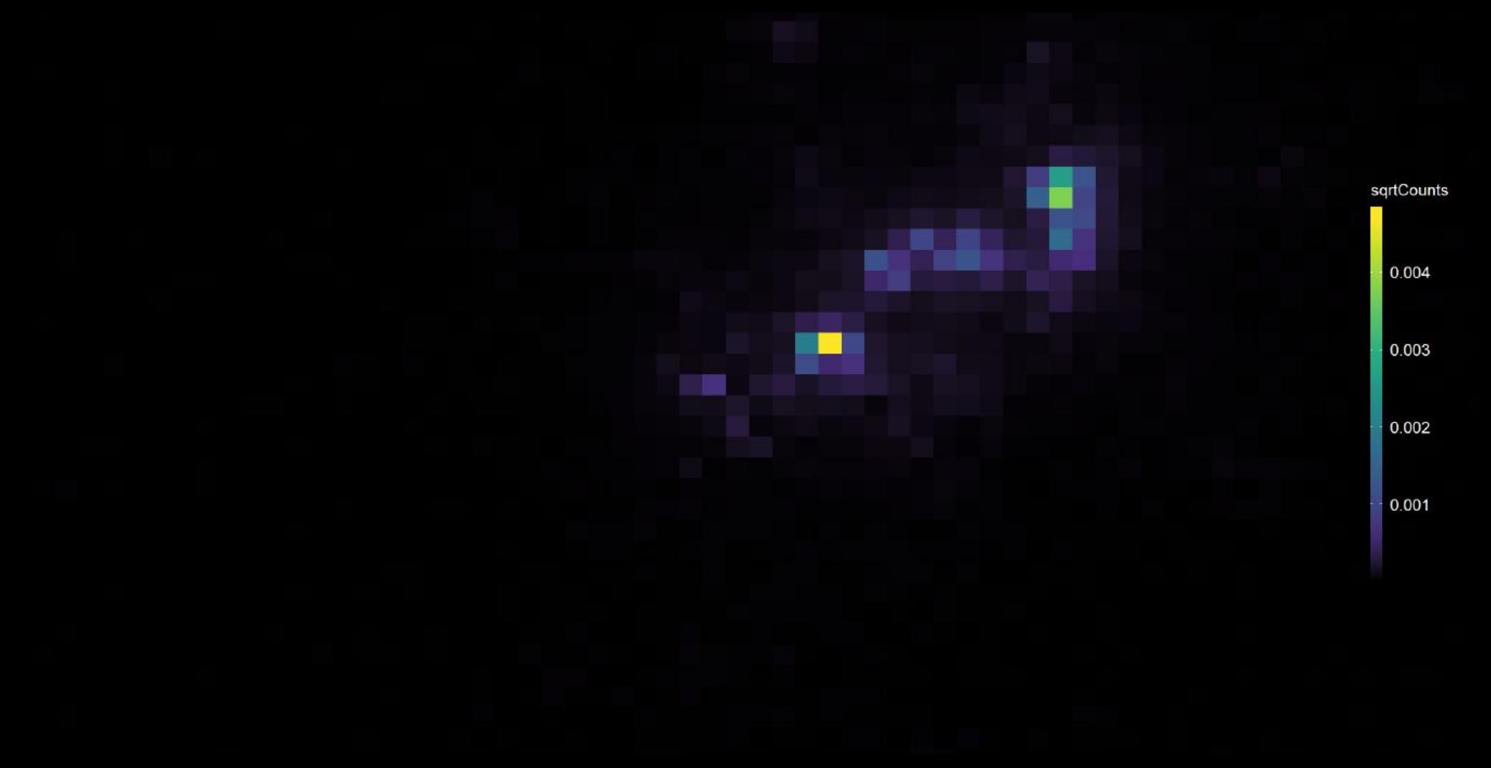
Talk at CHASC: https://hea-www.harvard.edu/astrostat/CHASC_2021/

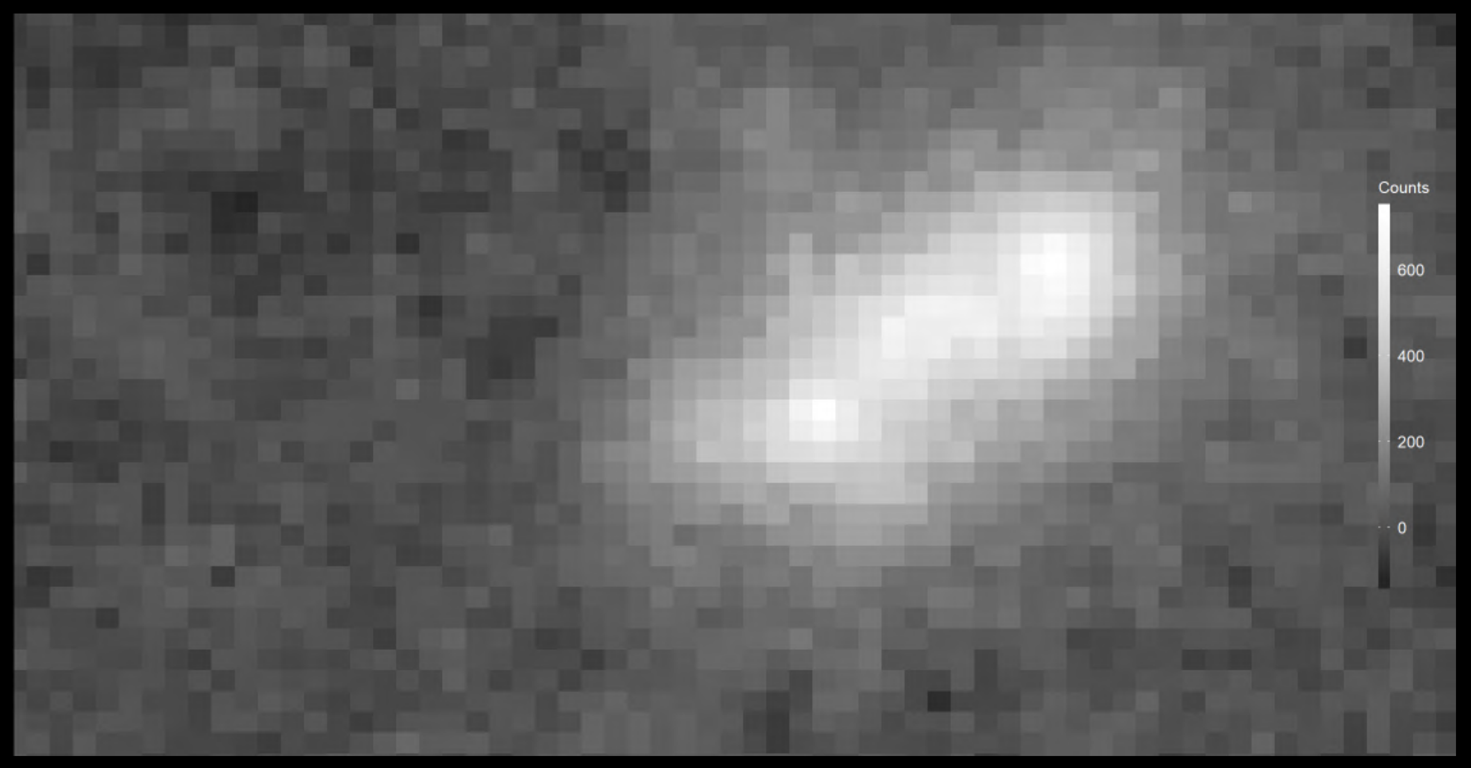
New Methods for Single Domain Analysis

Finding the source boundary

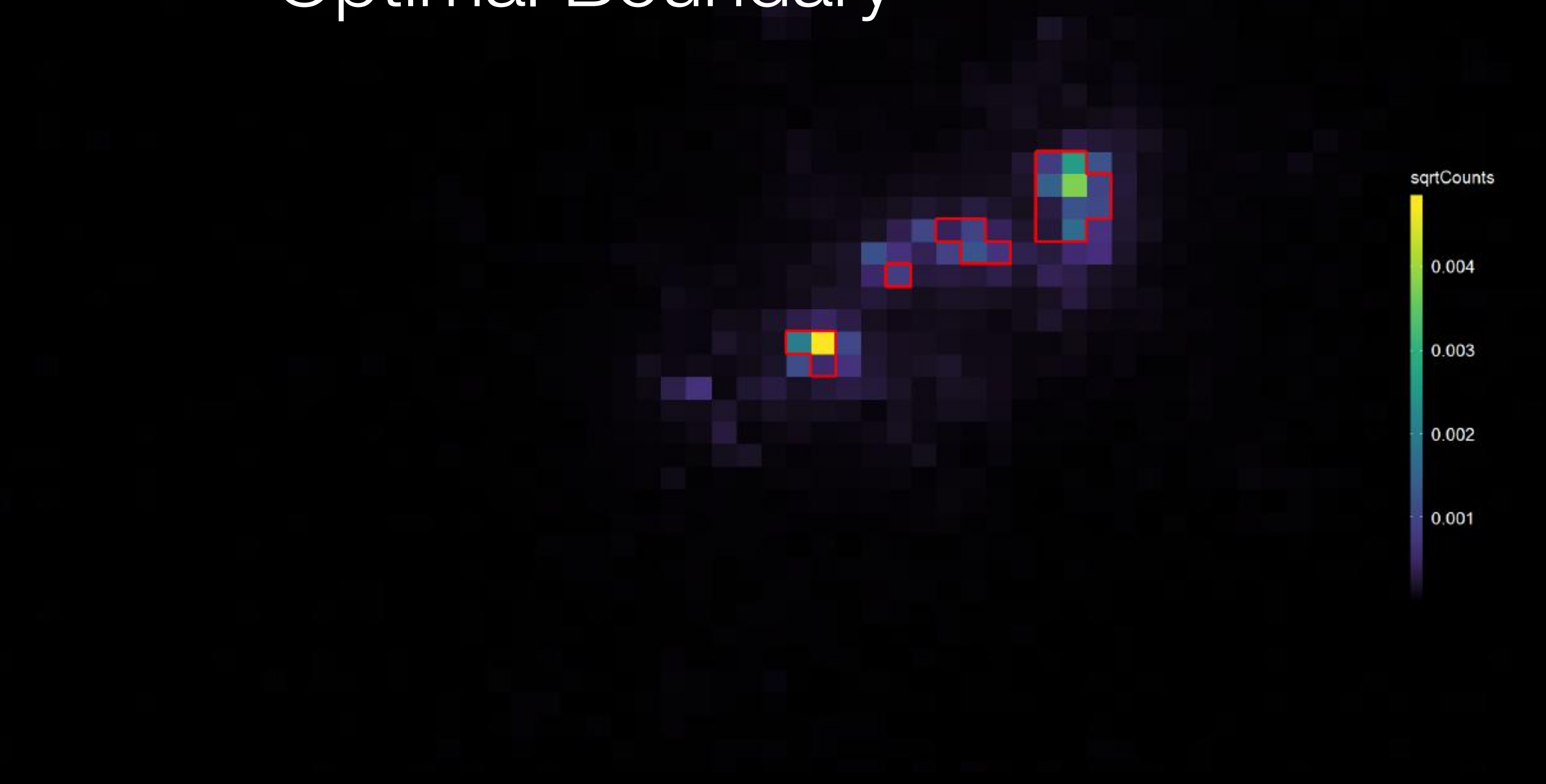
Posterior Draws with MCMC
probability distribution of pixel assignments

Posterior Mean





Optimal Boundary



ISING Prior
Correlation between neighboring pixels

Boundary with maximum probability
given LIRA-Ising posterior

Katy McKeough PhD Thesis

Talk at CHASC: https://hea-www.harvard.edu/astrostat/CHASC_2021/

Single Domain Analysis

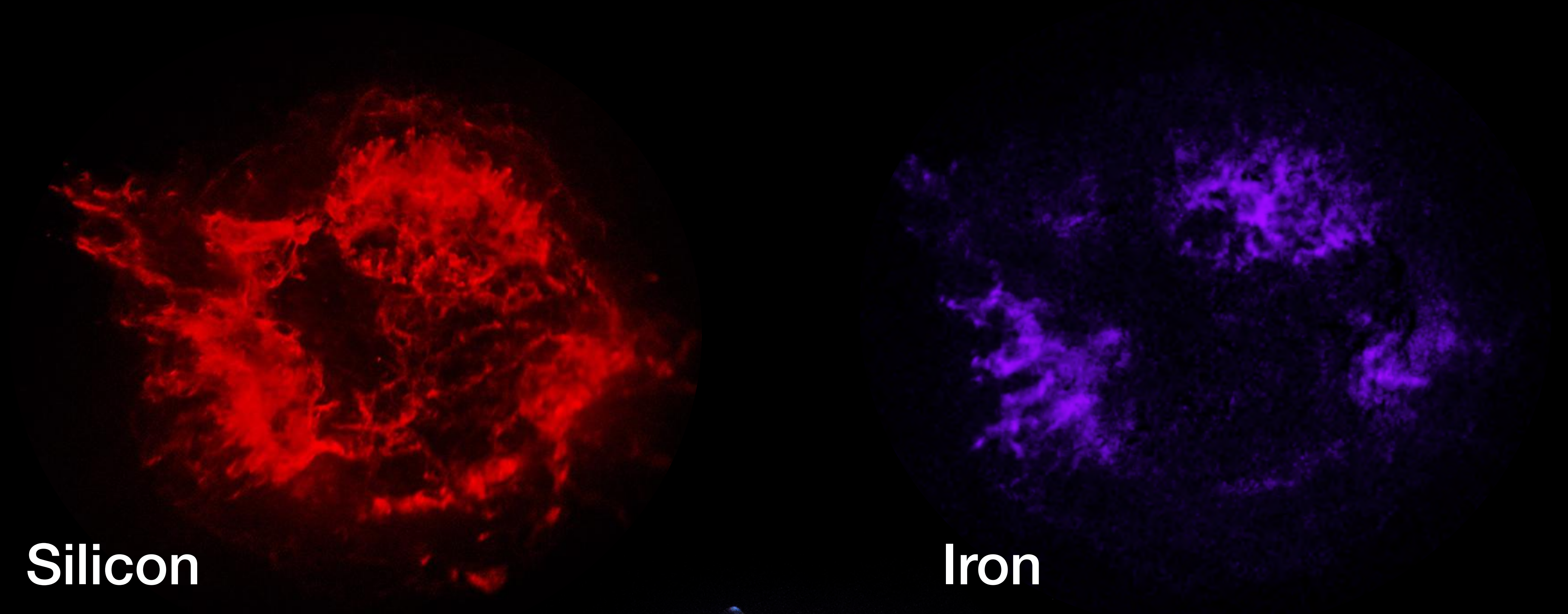
Analysis Domain	Description	Standard Methods	Challenges	Modern Methods
Spectra $e(E) = \int e(x, y, t, E) dx\, dy\, dt$	only energy, loss of time and morphology	forward fitting, multi-spectra, Poisson likelihood, model and instrument uncertainties	non-linear complex models, high resolution spectra, uncertainties in physical process & models	Bayesian Methods, Simulations, bootstrap, Likelihood free modeling, hierarchical Baysian models, model selection via ppp, BIC, AIC, ML
Image $e(x, y) = \int e(x, y, t, E) dE\, dt$	only location and morphology, loss of energy and time	source detections, morphology, contours, image reconstruction, deconvolution	faint structures, source boundaries, upper limits, crowded sources, background	Bayesian reconstruction, simulation for upper bounds, image segmentation
Time variability $e(t) = \int e(x, y, t, E) dx\, dy\, dE$	only time, loss of energy and source morphology	differences image/spectra, power spectra, periodogram, Bayesian Blocks, flares	S/N limitation on time resolution, break points, uneven sampling, non-detections	direct modeling of light curves (O-U, CARMA), periodograms, cross-spectra, flare detection

Multi-Domain Analysis

Spectra-Image

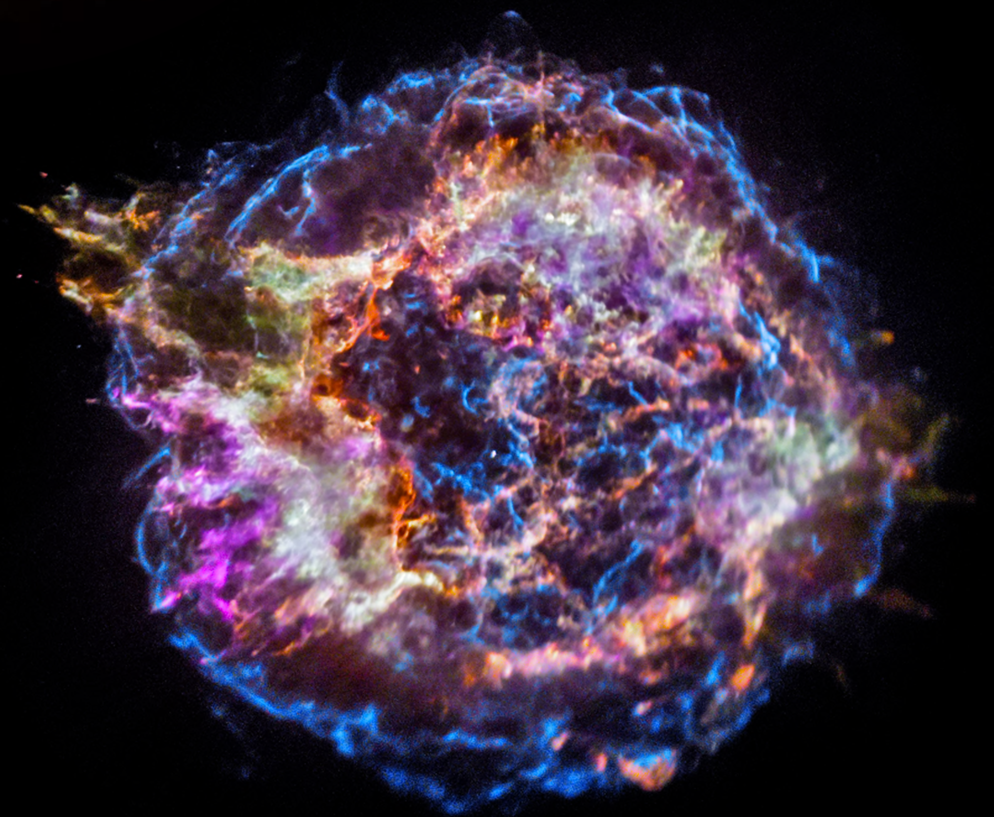
Example - 3D Model of SNR Cassiopeia A

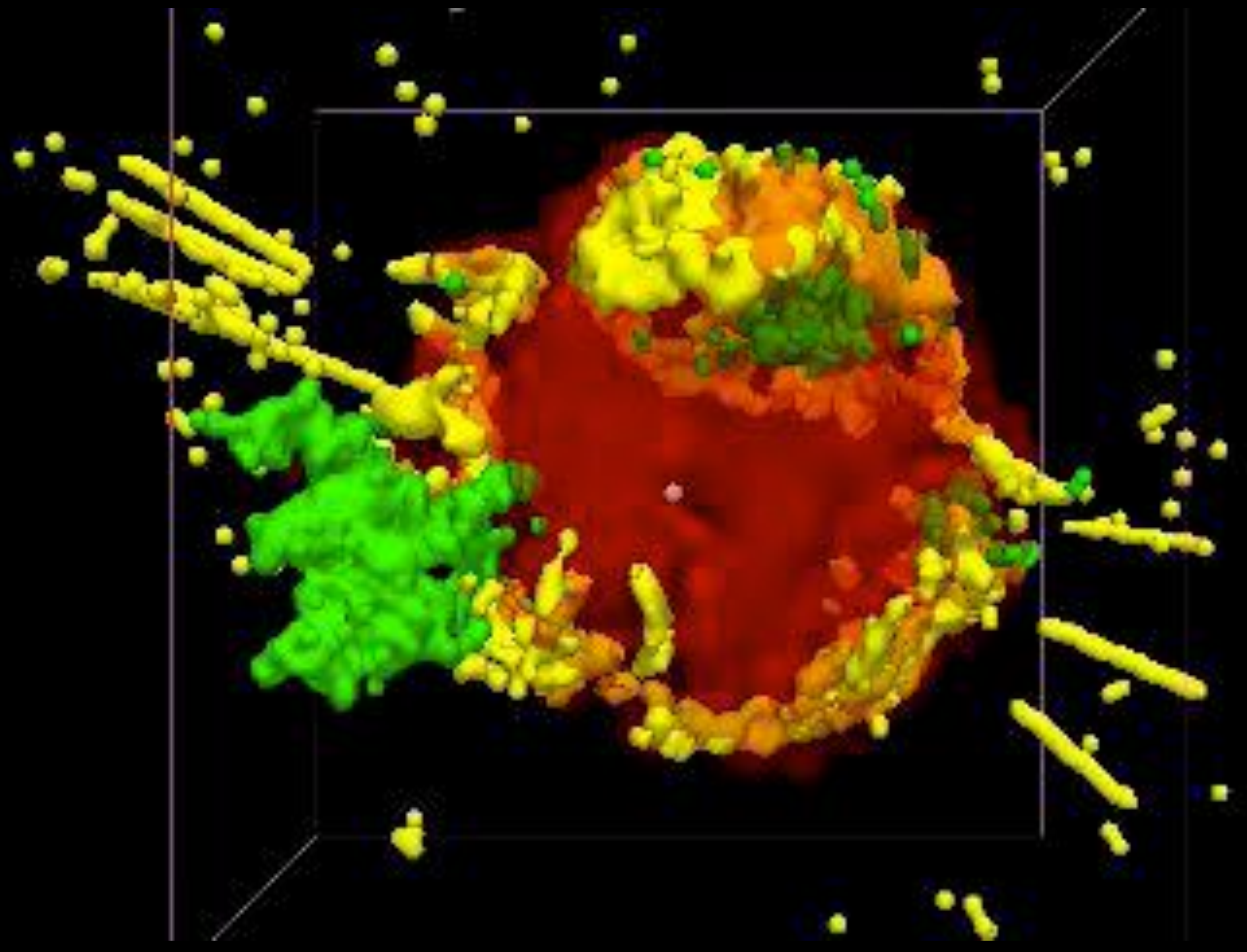
Spatial-Spectral distribution of specific elements



Silicon

Iron



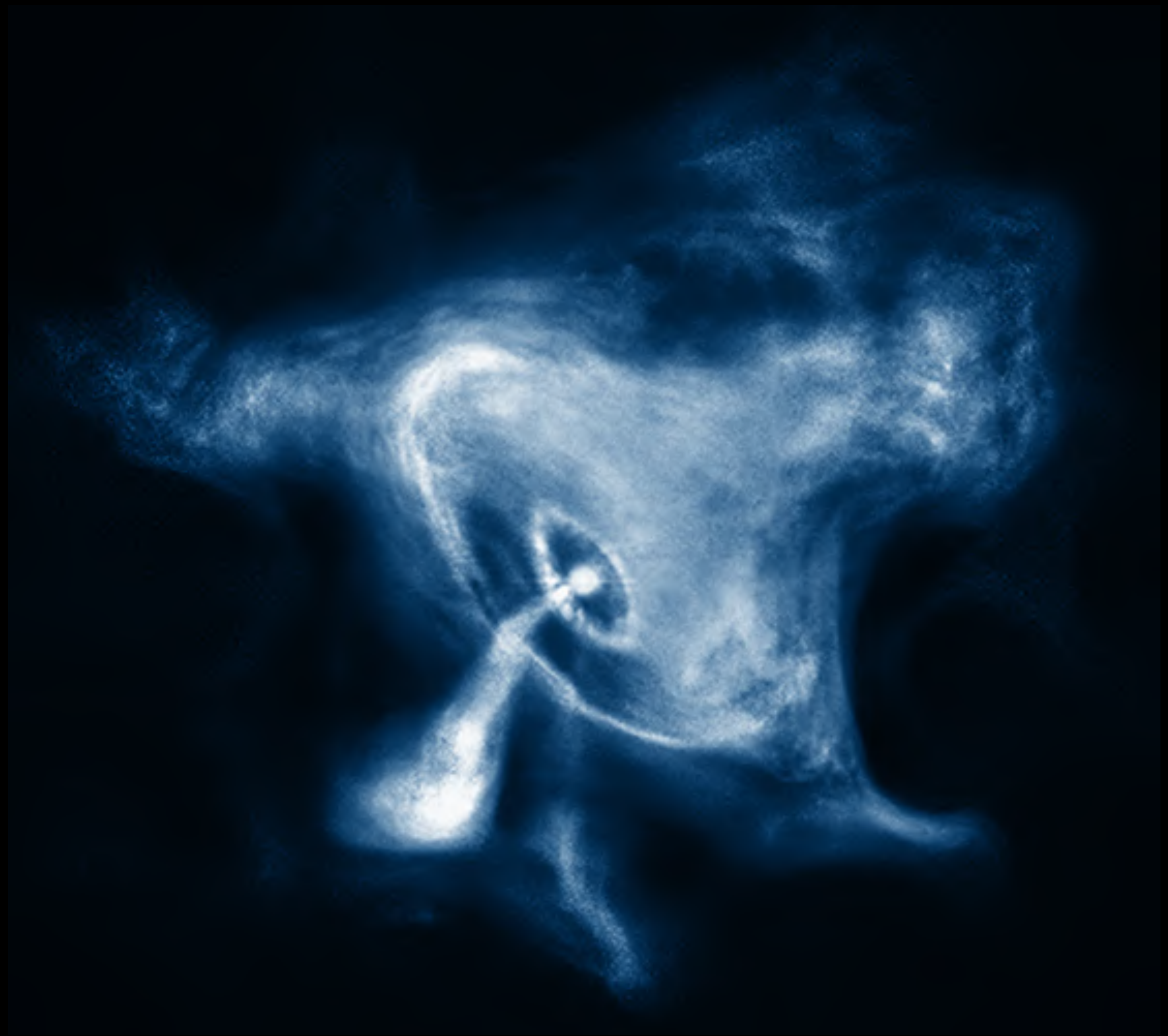


NASA/CXC/MIT Delaney et al 2009

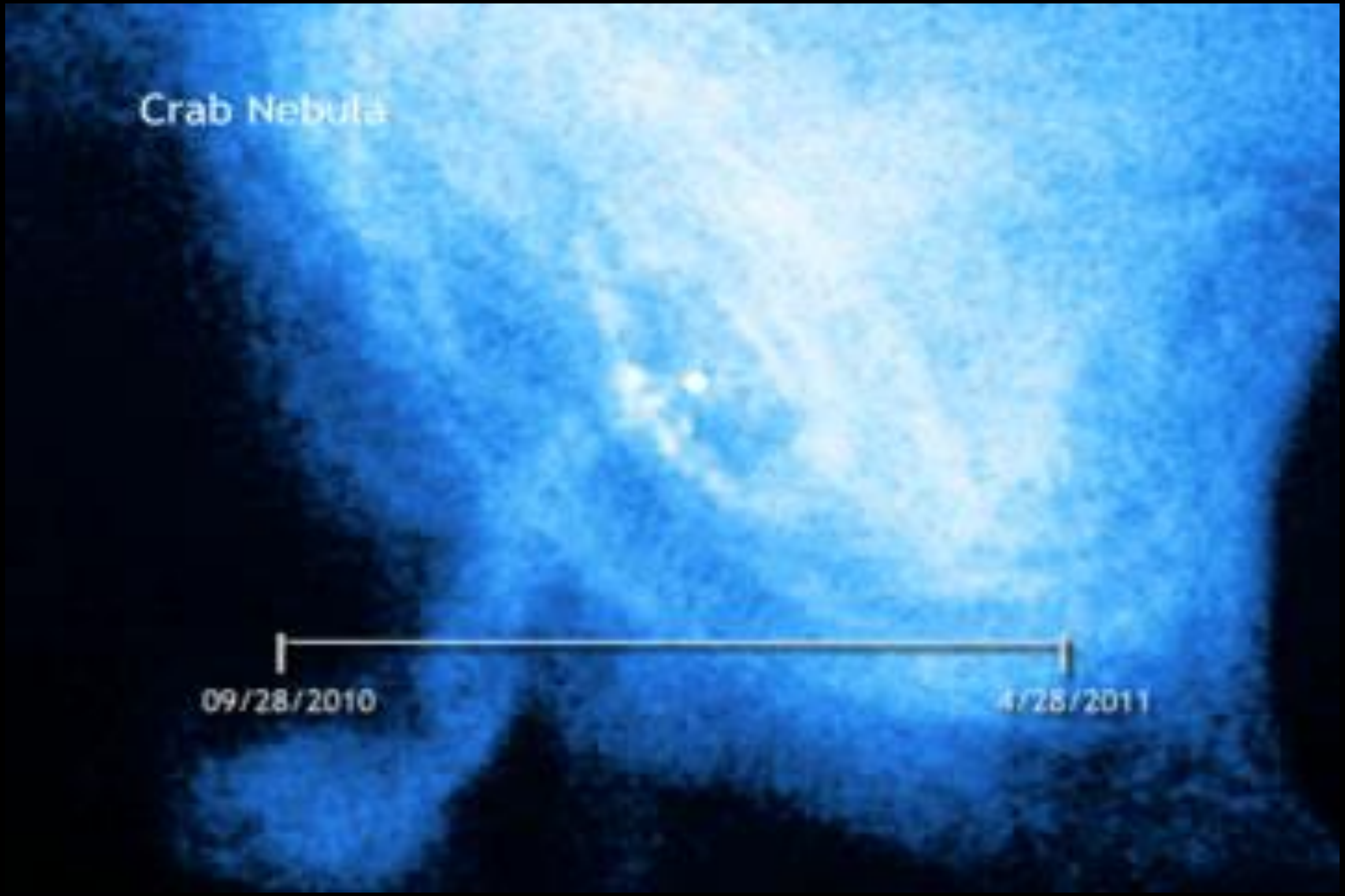
Multi-Domain Analysis

Image-Time

Example - Dynamical Evolution of a SNR



Chandra X-ray Image of Crab Nebula
Ring diameter ~ 1 lyr



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 (Xu+ 2021)

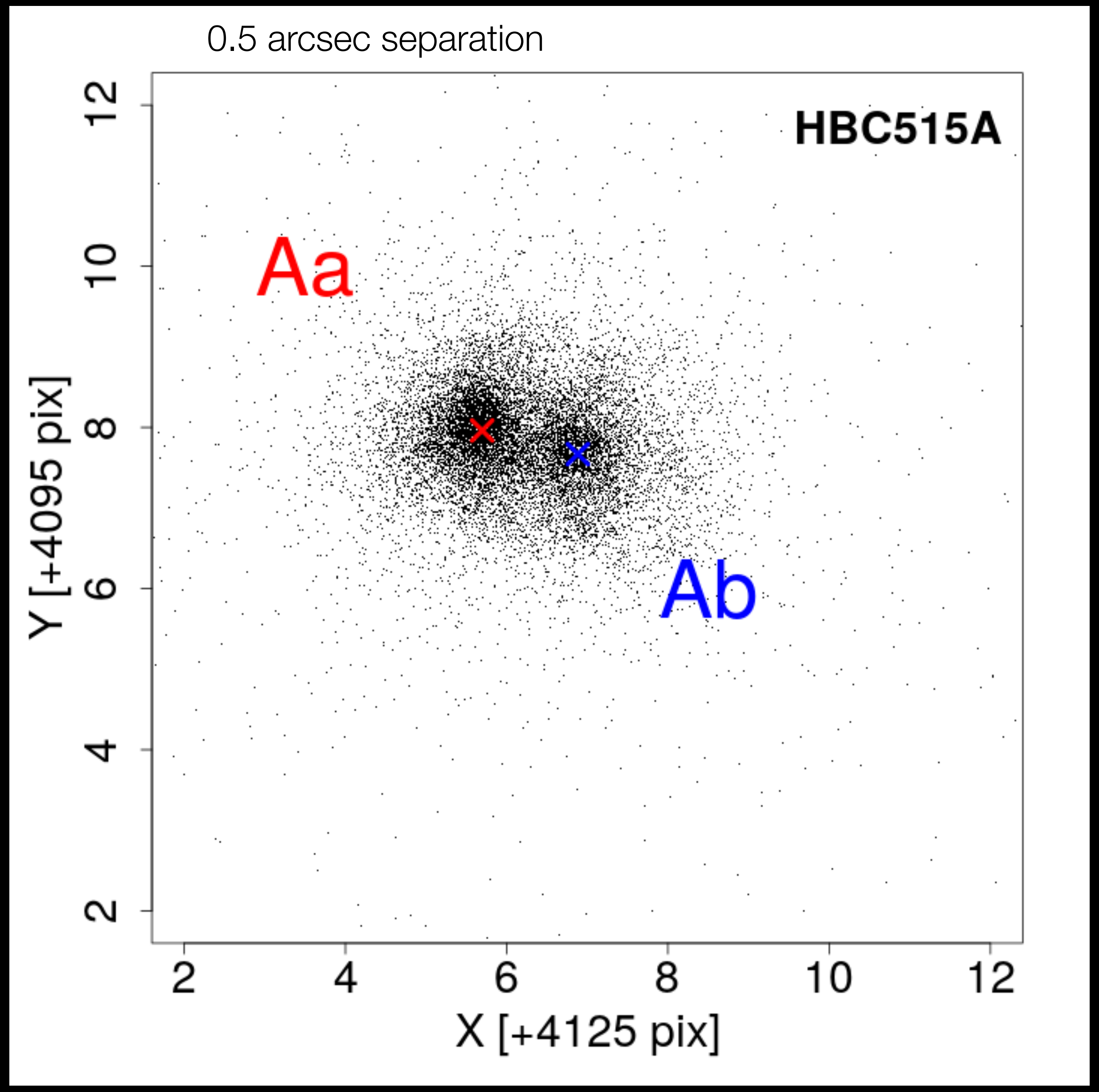
Chandra X-ray Image of Orion Nebula

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

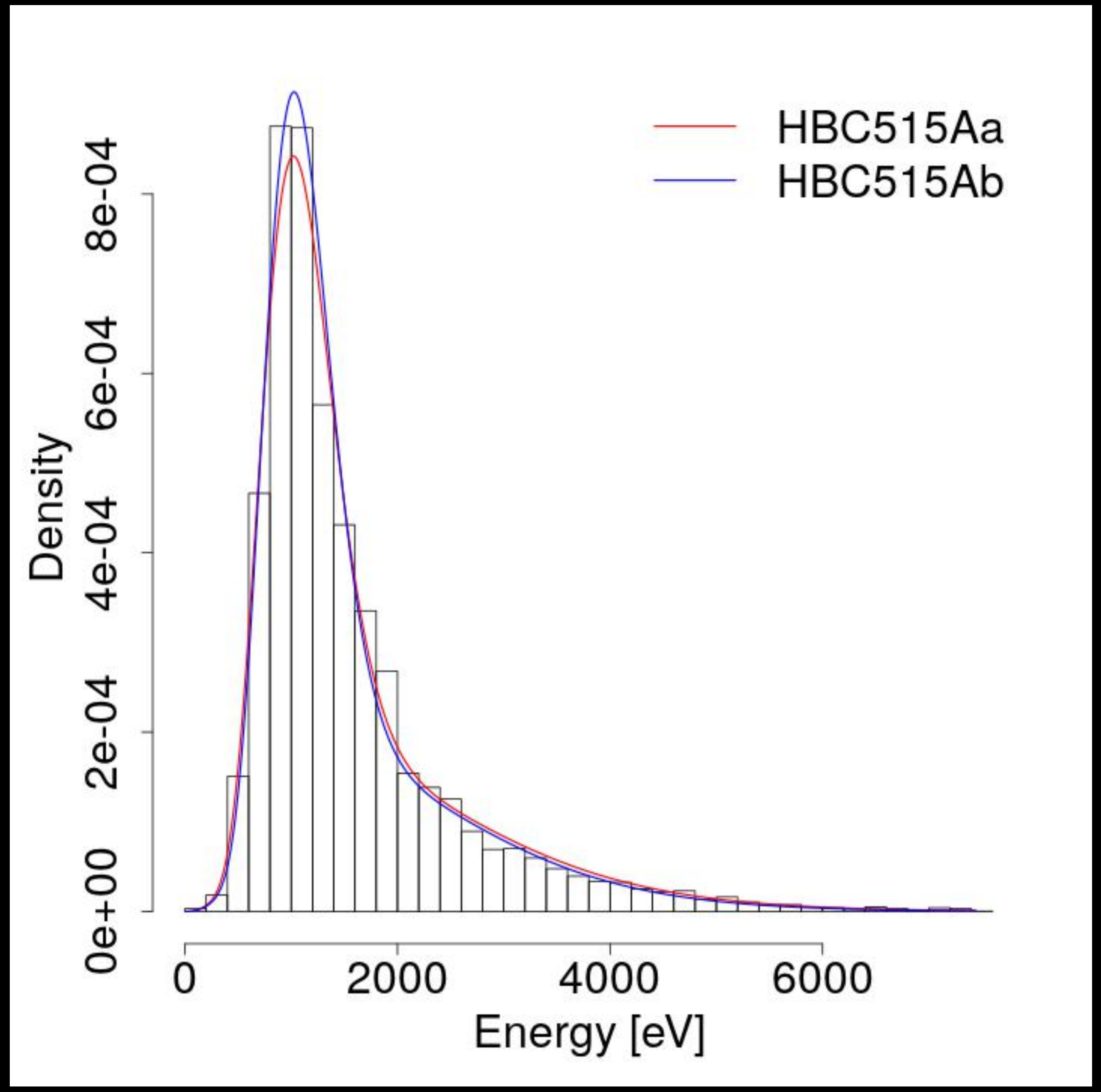
Emerging Multi-Domain Analysis

Full information: Image-Spectral-Time

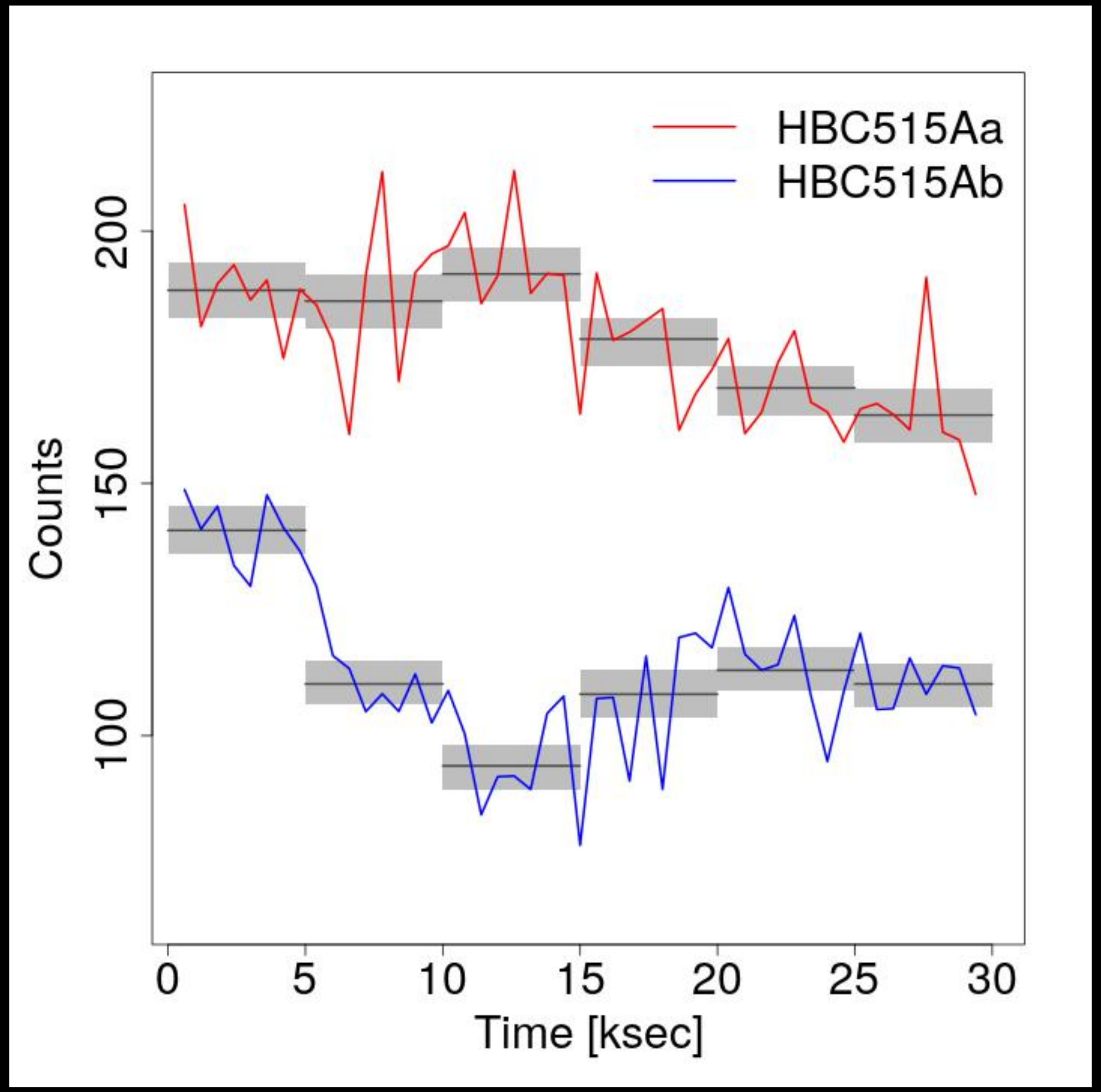
Example:
Probabilistic separation of photons from two close sources with eBASCS using location, spectrum and time



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



spectra for each star with eBASCS



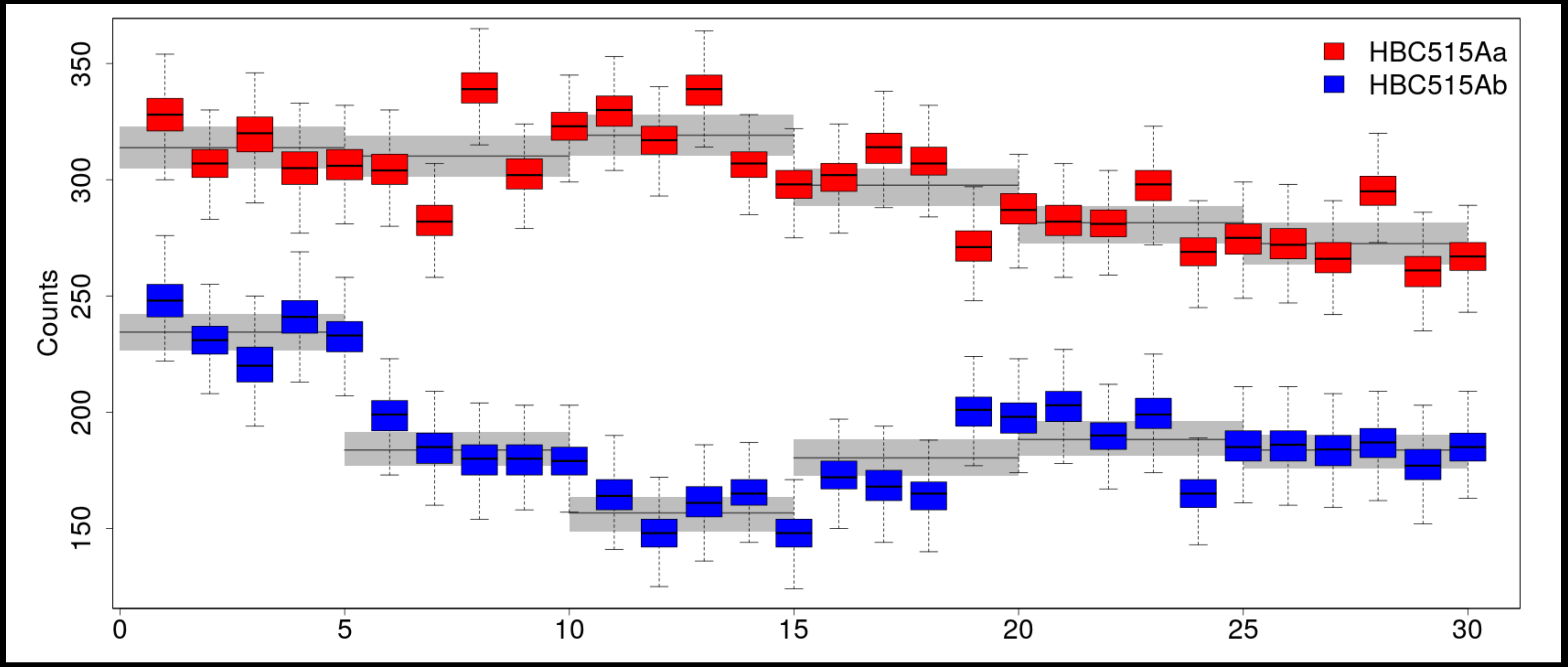
light curves of each component eBASCS

Meyer et al 2021

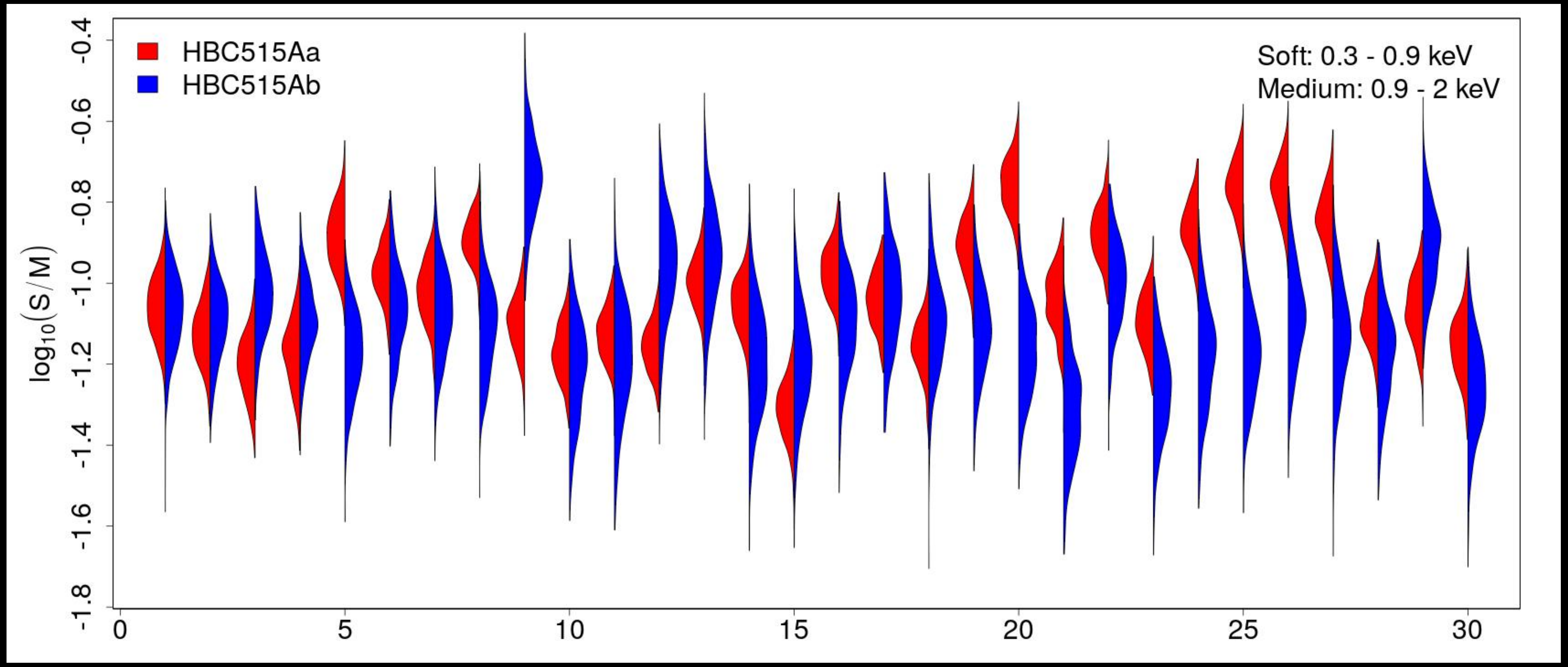
Emerging Multi-Domain Analysis

Full information: Image-Spectral-Time

Counts



Ratio S/M



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

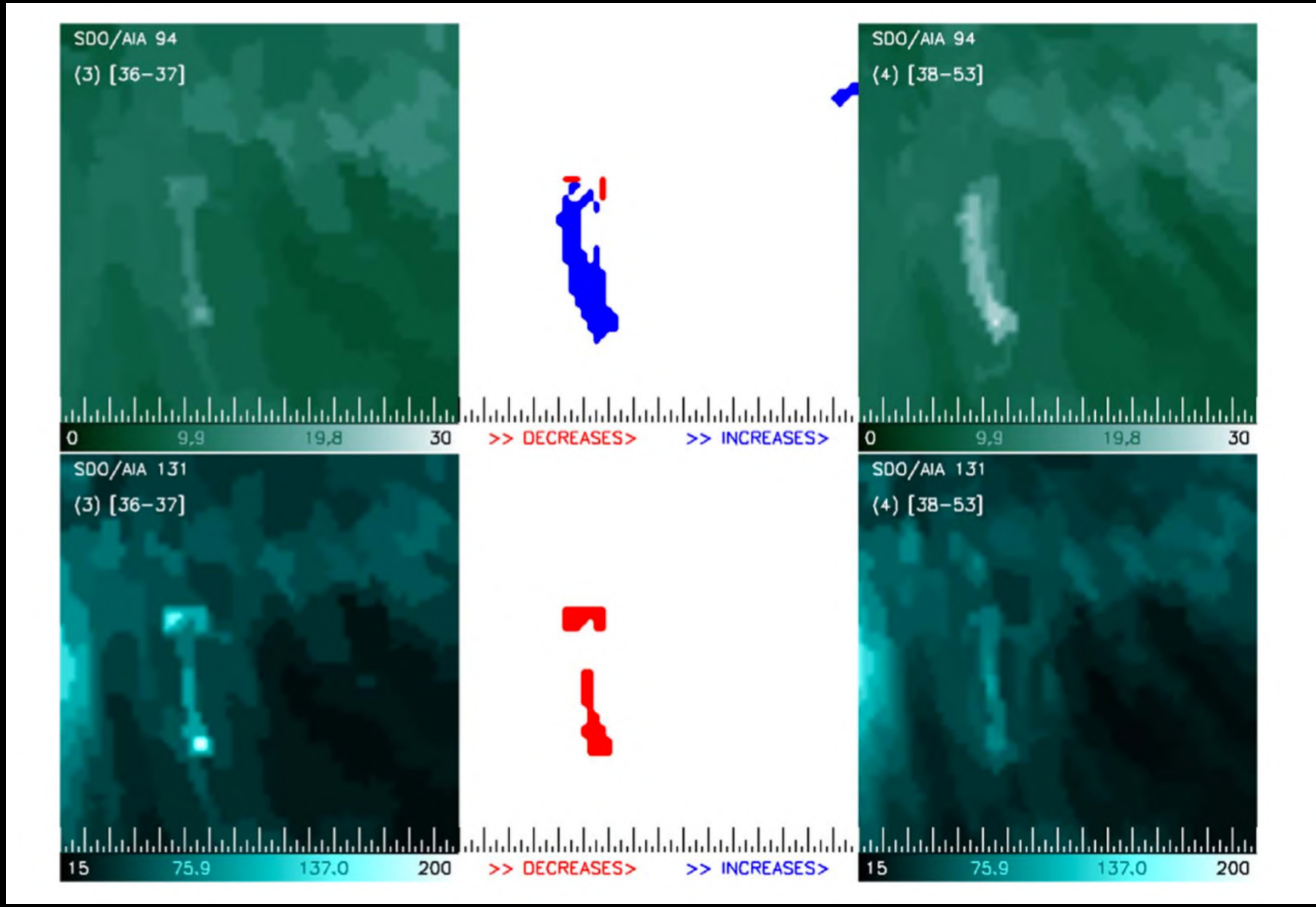
Note: need to include the instrument response in modeling spectra

Meyer et al 2021

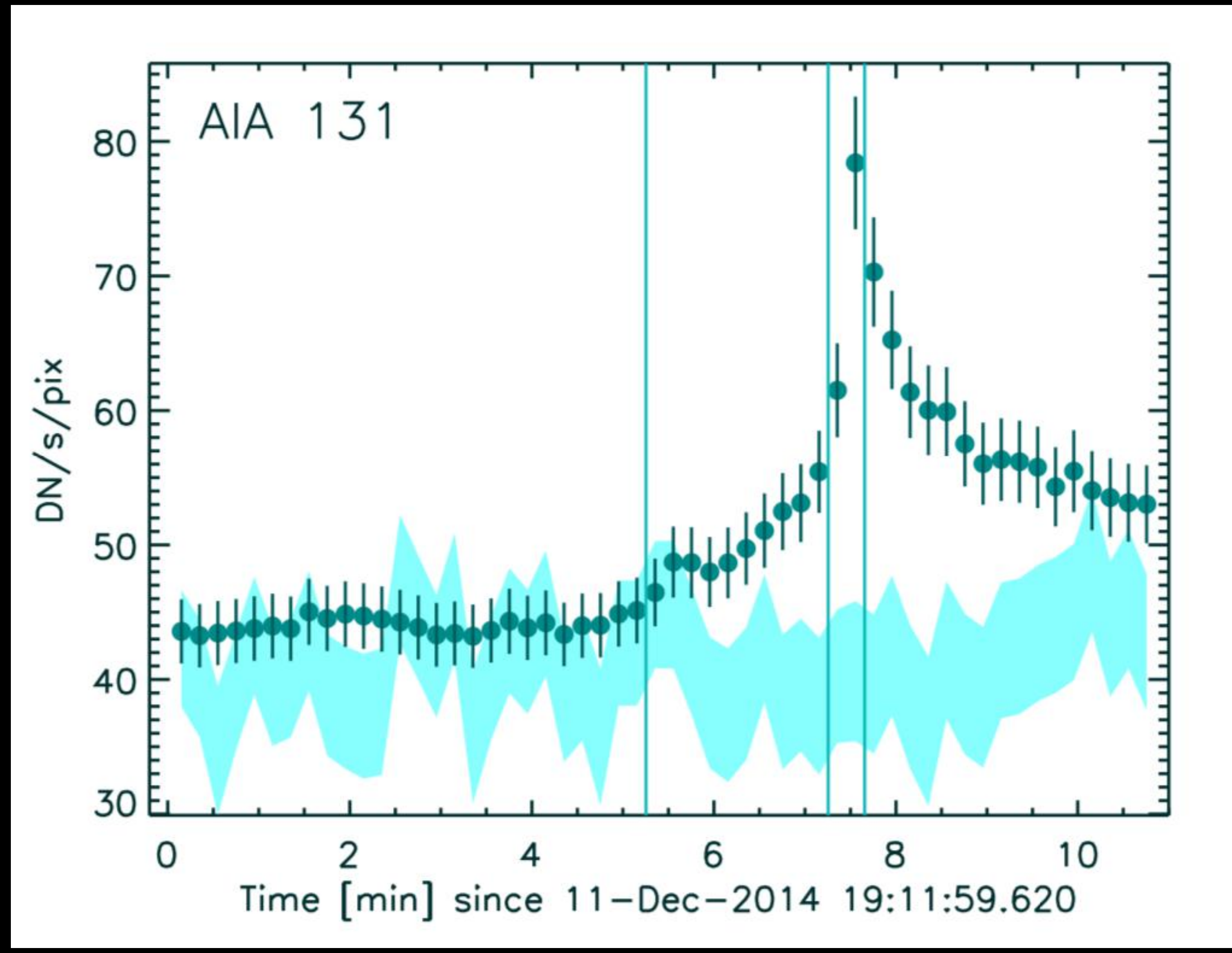
Emerging Multi-Domain Analysis

Full information: Image-Spectral-Time

Detecting flaring regions in the images of the Sun

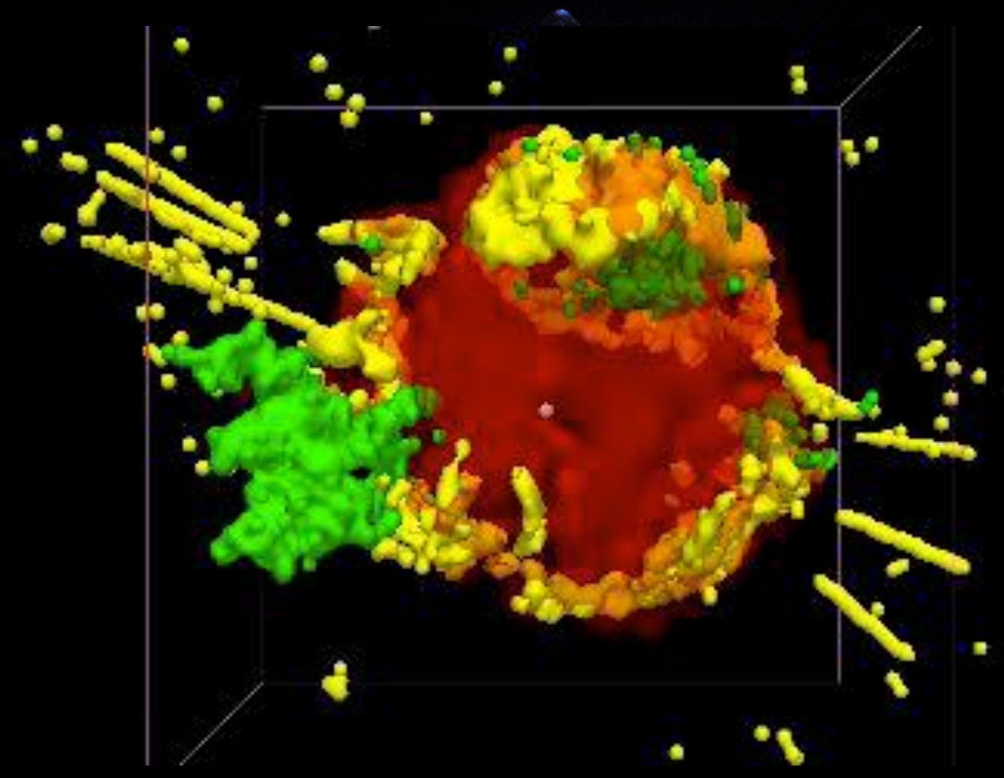


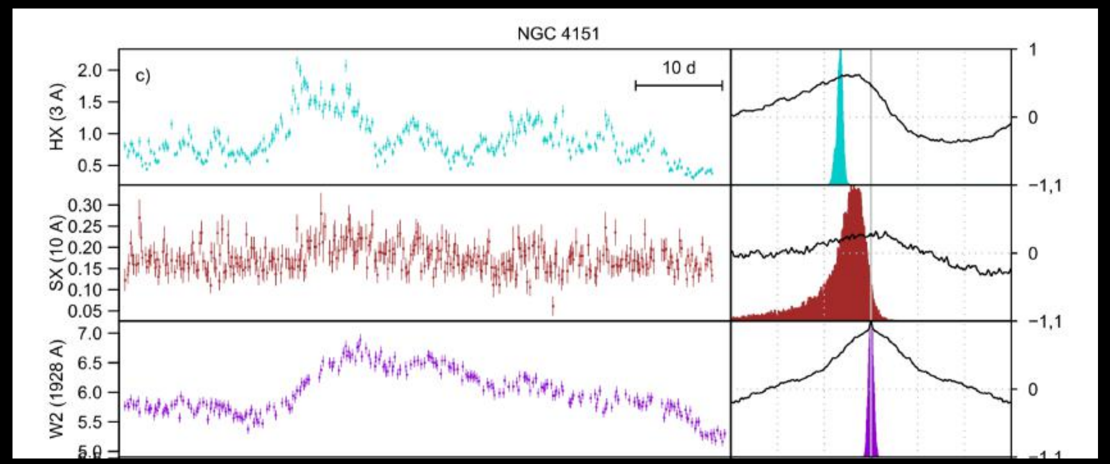
Change-points and Image Segmentation for Time Series Images - 4D-Automark

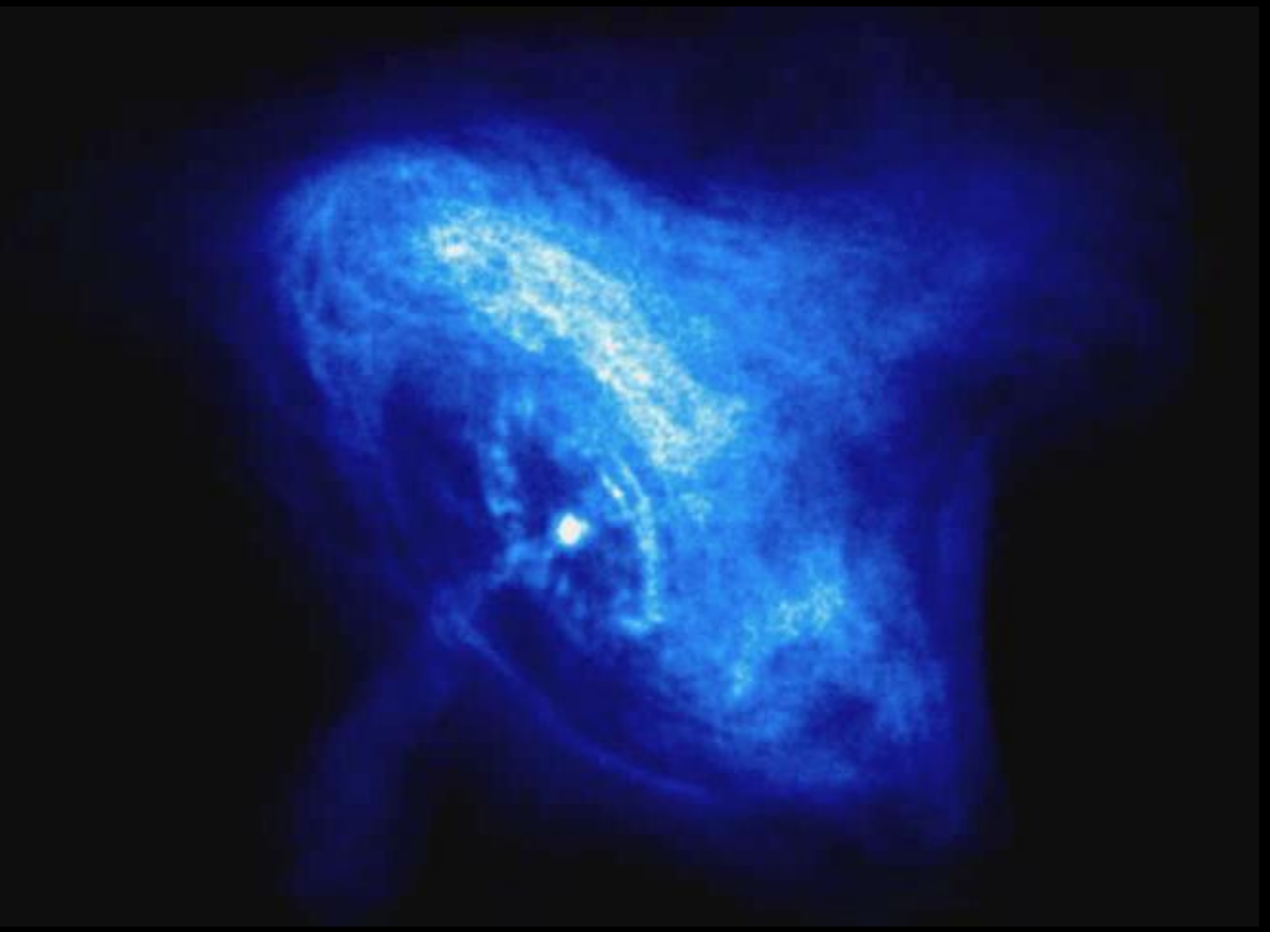


Xu et al 2021

Emerging Multi-Domain Analysis







Analysis	Description	Current Method	Challenges	Emerging
Spectral-Image $e(x_i, y_i, E_i)$	loss of time $\int e(x,y,t,E)\,dt$	source detection (VTP), spectral-image model, project, deproject in clusters, SNR	multi-spectra, averaging over image, overlapping sources, transients	(e)BASCS, BSS, Adaptive binning, ML
Spectral-Time $e(E_i, t_i)$	loss of location $\int e(x,y,t,E)\,dx\,dy$	multi-spectra, inter-band correlation	low counts spectra, non-even sampling, different apertures, multi-components	cross-spectrum, ABC, JAVELIN, Auto-Mark ML
Image-Time $e(x_i, y_i, t_i)$	loss of energy $\int e(x,y,t,E)\,dE$	image difference, source detection	spectral information, evolving boundaries, PSF, averaging	4D-automark spatial fitting, ML

Future Full Multi-Domain Analysis

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

Uncertainties: Data Collection

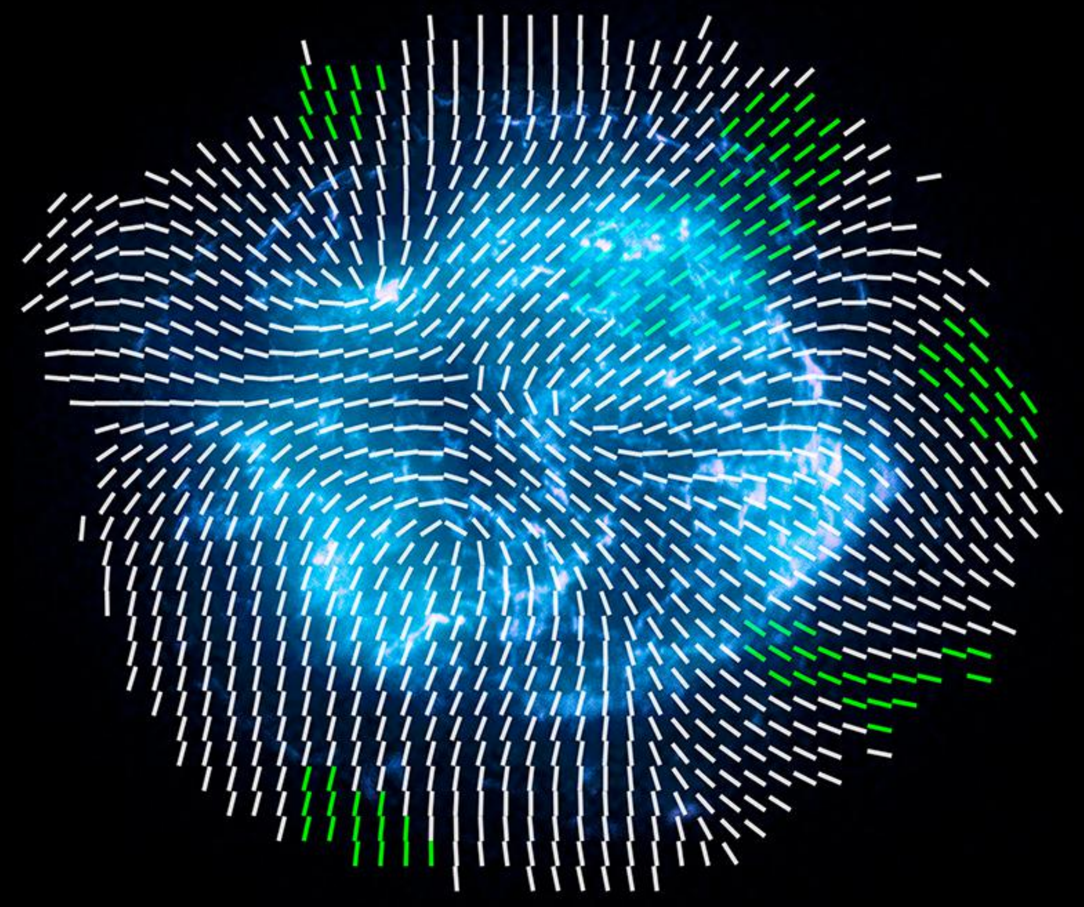
- X-rays are photon Events: sparsity, multi-dimensionality of the data
    - uncertainties in measurements: event location, energy and arrival time
        - calibration uncertainties (ARF, RMF, PSF)
        - instrumental effects (pileup, dead-time)
    - separating background and source events
    - overlapping sources in crowded fields due to point spread function (PSF) blurring
    - model image of computer generated PSF (uncertainties?)

Uncertainties: Science Inference

- Impact on scientific analysis and inference:
    - localization of photons, source position, identification of a source
    - Source intensity and flux
    - merged/combined data from multiple observations
    - X-ray structures:
        - detection of diffuse structures in images with Poisson background
        - define source boundaries
        - alignment between images in different bands: radio, optical, X-rays, volumes

Summary

- X-ray view **->** Universe is not calm



- Complex 4D X-ray data

  + new polarimetry data

- Computer generated models of physical processe characteristics (PSF, pileup etc.) often applied to the observed data

- Future methodology including emerging Machine Learning methods need to provide measurements of uncertainties impacting the scientific inference

- New methods have to be formatted for astronomers to be applied to their observations.

Reference

The Next Decade of Astroinformatics and Astrostatistics

Show affiliations Hide authors

Siemiginowska, Aneta ; Eadie, Gwendolyn ; Czekala, Ian ; Feigelson, Eric ; Ford, Eric B. ; Kashyap, Vinay ; Kuhn, Michael ; Loredo, Tom ; Ntampaka, Michelle ; Stevens, Abbie ; Avelino, Arturo ; Borne, Kirk ; Budavari, Tamas ; Burkhart, Blakesley ; Cisewski-Kehe, Jessi ; Civano, Francesca ; Chilingarian, Igor ; van Dyk, David A. ; Fabbiano, Giuseppina ; Finkbeiner, Douglas P. ; Foreman-Mackey, Daniel ; Freeman, Peter ; Fruscione, Antonella ; Goodman, Alyssa A. ; Graham, Matthew ; Guenther, Hans Moritz ; Hakkila, Jon ; Hernquist, Lars ; Huppenkothen, Daniela ; James, David J. ; Law, Casey ; Lazio, Joseph ; Lee, Thomas ; López-Morales, Mercedes ; Mahabal, Ashish A. ; Mandel, Kaisey ; Meng, Xiao-Li ; Moustakas, John ; Muna, Demitri ; Peek, J. E. G. ; Richards, Gordon ; Portillo, Stephen K. N. ; Scargle, Jeff ; de Souza, Rafael S. ; Speagle, Joshua S. ; Stassun, Keivan G. ; Stenning, David C. ; Taylor, Stephen R. ; Tremblay, Grant R. ; Trimble, Virginia ; Yanamandra-Fisher, Padma A. ; Young, C. Alex

Over the past century, major advances in astronomy and astrophysics have been driven by improvements in instrumentation. With the amassing of high quality data from new telescopes it is becoming clear that research in astrostatistics and astroinformatics will be necessary to develop new methodology needed in astronomy.

Publication:	Astro2020: Decadal Survey on Astronomy and Astrophysics, science white papers, no. 355; Bulletin of the American Astronomical Society, Vol. 51, Issue 3, id. 355 (2019)
Pub Date:	May 2019
arXiv:	arXiv:1903.06796
Bibcode:	2019BAAS...51c.355S
Keywords:	Astrophysics - Instrumentation and Methods for Astrophysics
E-Print Comments:	Submitted to the Astro2020 Decadal Survey call for science white papers

Astrostatistics News

About

Astrostatistics News (AN) is a newsletter designed to inform, promote, cultivate, and inspire the astrostatistics community.

The AN editors are Jessi Cisewski-Kehe (UW-Madison), David W. Hogg (NYU), Vinay L. Kashyap (CfA), and Aneta Siemiginowska (CfA). The AN was established in late 2022 with encouragement from the International Astrostatistics Association.

We anticipate 2 - 3 issues per year, with potential for more.

Subscribe to the Newsletter

https://www.astrostatisticsnews.com/

Astrostatistics News

Issue 1, December 2022
Issue Editors: Jessi Cisewski-Kehe, David W. Hogg, Vinay L. Kashyap, Aneta Siemiginowska

Introducing the Newsletter

Astrostatistics News (*AN*) is a newsletter designed to inform, promote, cultivate, and inspire the astrostatistics community. The AN editors are Jessi Cisewski-Kehe (UW-Madison), David W Hogg (NYU; Flatiron), Vinay Kashyap (CfA), and Aneta Siemiginowska (CfA). The *AN* was established in late 2022 with encouragement from the International Astrostatistics Association. We anticipate 2 - 3 issues per year, with potential for more.

Astrostatistics News Mission Statement

Astrostatistics News serves the astrostatistics community by highlighting and describing recent research developments in astrostatistics at an accessible level to the diverse backgrounds of its members, sharing interesting new algorithms, software, or data sets, promoting relevant events, and striving to inspire new researchers to join in the fun.

Subscribe to *Astrostatistics News*

To subscribe to *Astrostatistics News*, go to https://groups.google.com/g/astrostatistics-news and select the "Join group" button. You will need to be logged into your Google account to join the group.

Please forward this information to anyone who may be interested!

Thank you CHASC

https://hea-www.harvard.edu/astrostat/

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