JSM 2021 Aug 8: 220142: Statistical Answers to astrophysical questions

Low hanging fruit

Common analysis problems in astronomy with[out] principled solutions

Vinay Kashyap
Center for Astrophysics | Harvard & Smithsonian

& CHASC Astrostatistics Collaboration

Low hanging fruit

There are a lot of analysis challenges in astro data that seem very simple

but have [mostly] not been dealt with rigorously

These are ripe for statisticians to step in and guide future analyses.

I will present a smorgasbord of such challenges, along with pointers to how our CHASC astrostatistics collaboration has been dealing with them

A Menu

1. Fitting datasets of different sizes and different qualities
2. Uncertainties on PSF encircled-energy contours
3. Systematic uncertainties in point spread and spectral redistribution
4. Calibration of effective areas of multiple telescopes
5. ~~Setting bounds on the range of a power-law distribution~~
6. ~~Changepoints in a 2/3/4D datacube~~
7. ~~Segmentation of a photon events list~~
8. ~~Boundary of a blob~~

1. Fitting datasets of different sizes and different qualities

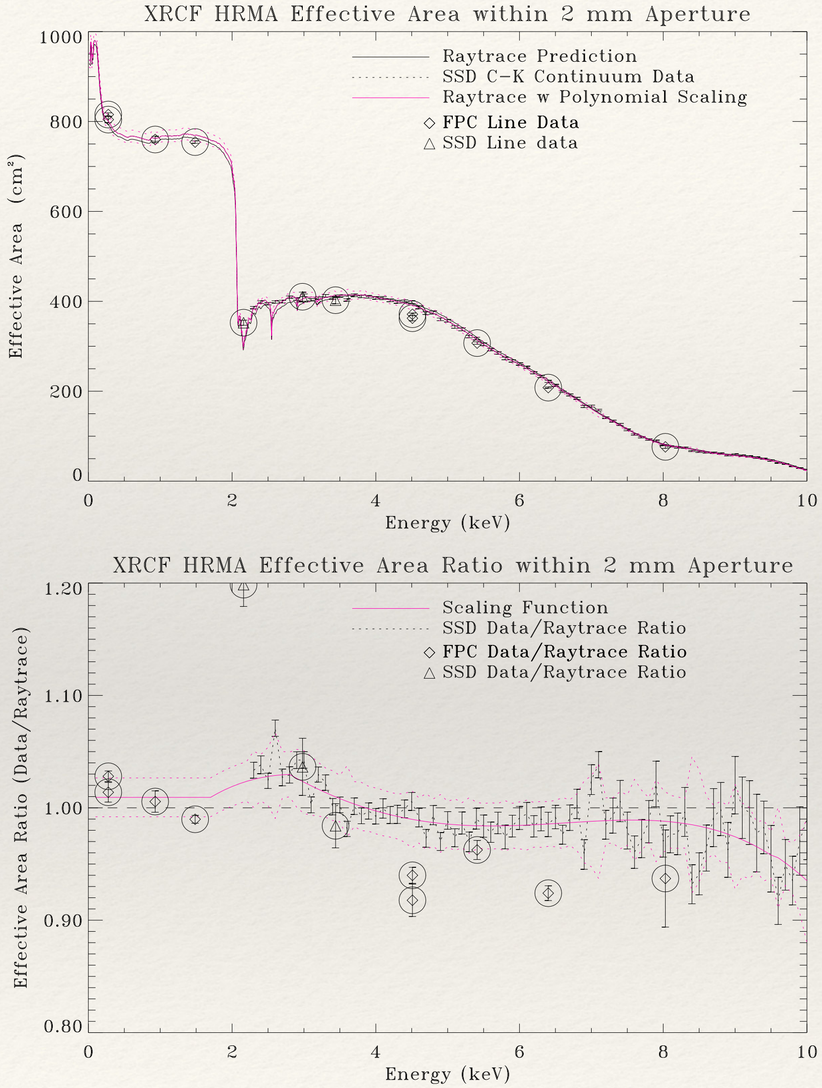
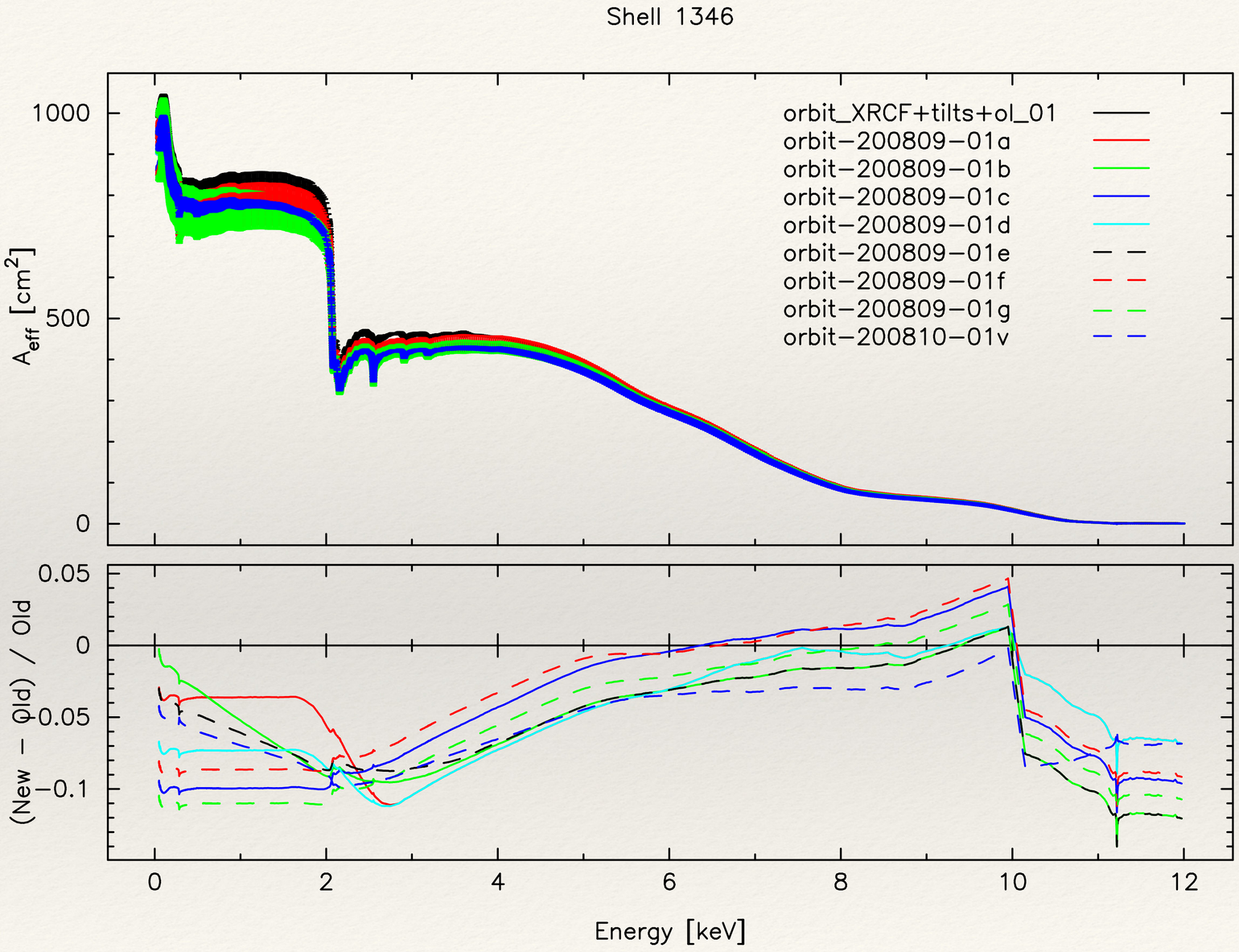


Figure 9: Calibration data vs. raytrace prediction. Top panel shows the XRCF HRMA effective area within 2 mm aperture. Bottom panel shows the effective area ratio of data/raytrace.



Zhao, CXC Memo 2008

1. Fitting datasets of different sizes and different qualities

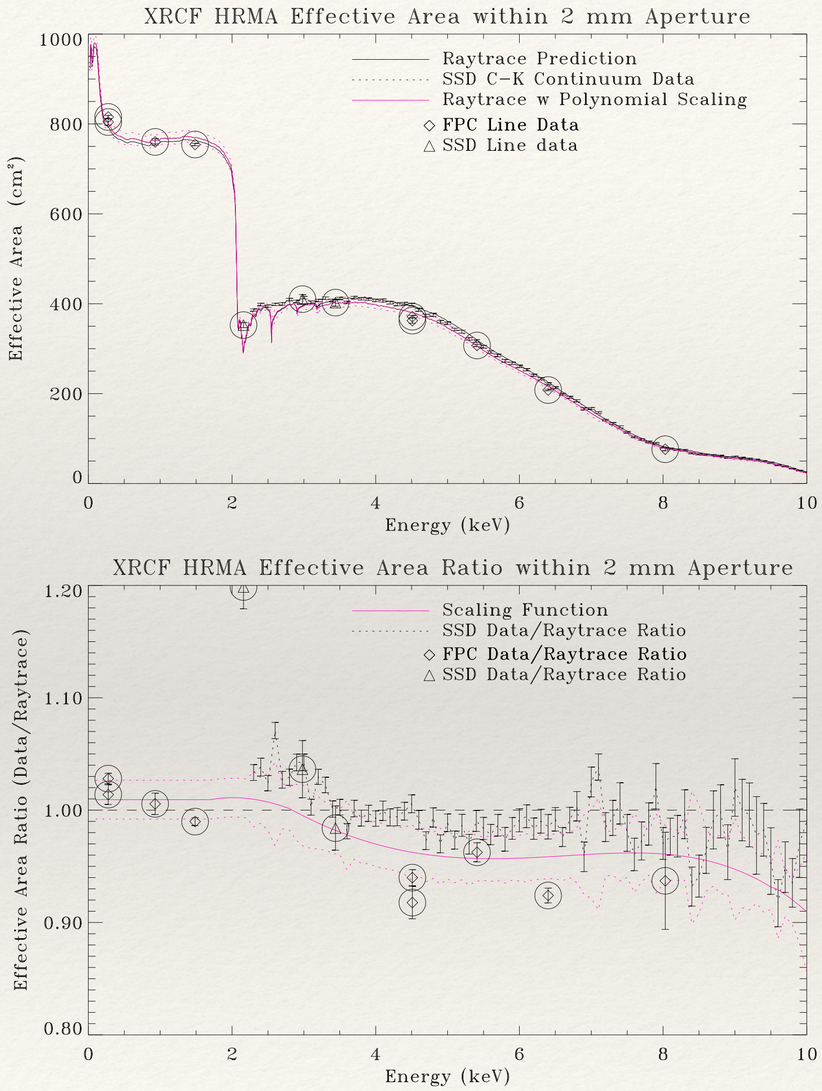
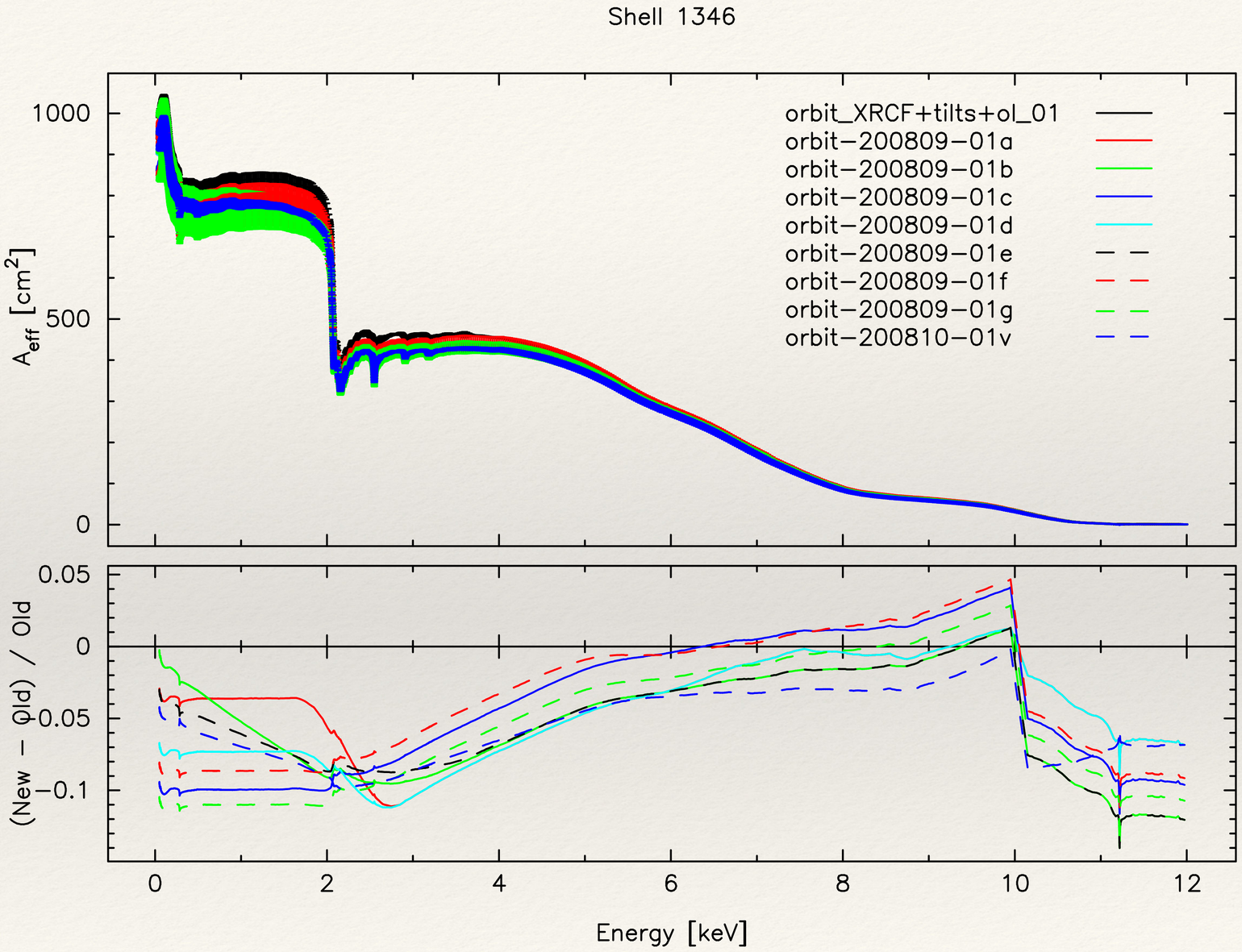


Figure 10: Calibration data vs. raytrace prediction. Top panel shows the XRCF HRMA effective area within 2 mm aperture. Bottom panel shows the effective area ratio of data/raytrace and pulled down by giving 50% the weight to the FPC data.



Zhao, CXC Memo 2008

1. Fitting datasets of different sizes and different qualities

(a) flat FPC+4th order poly SSD <2 keV, FPC 1/2 >2 keV

(b) 2nd order poly FPC <2 keV, 1/2 FPC+flat SSD >2 keV

(c) No FPC, extrapolate SSD to <2 keV

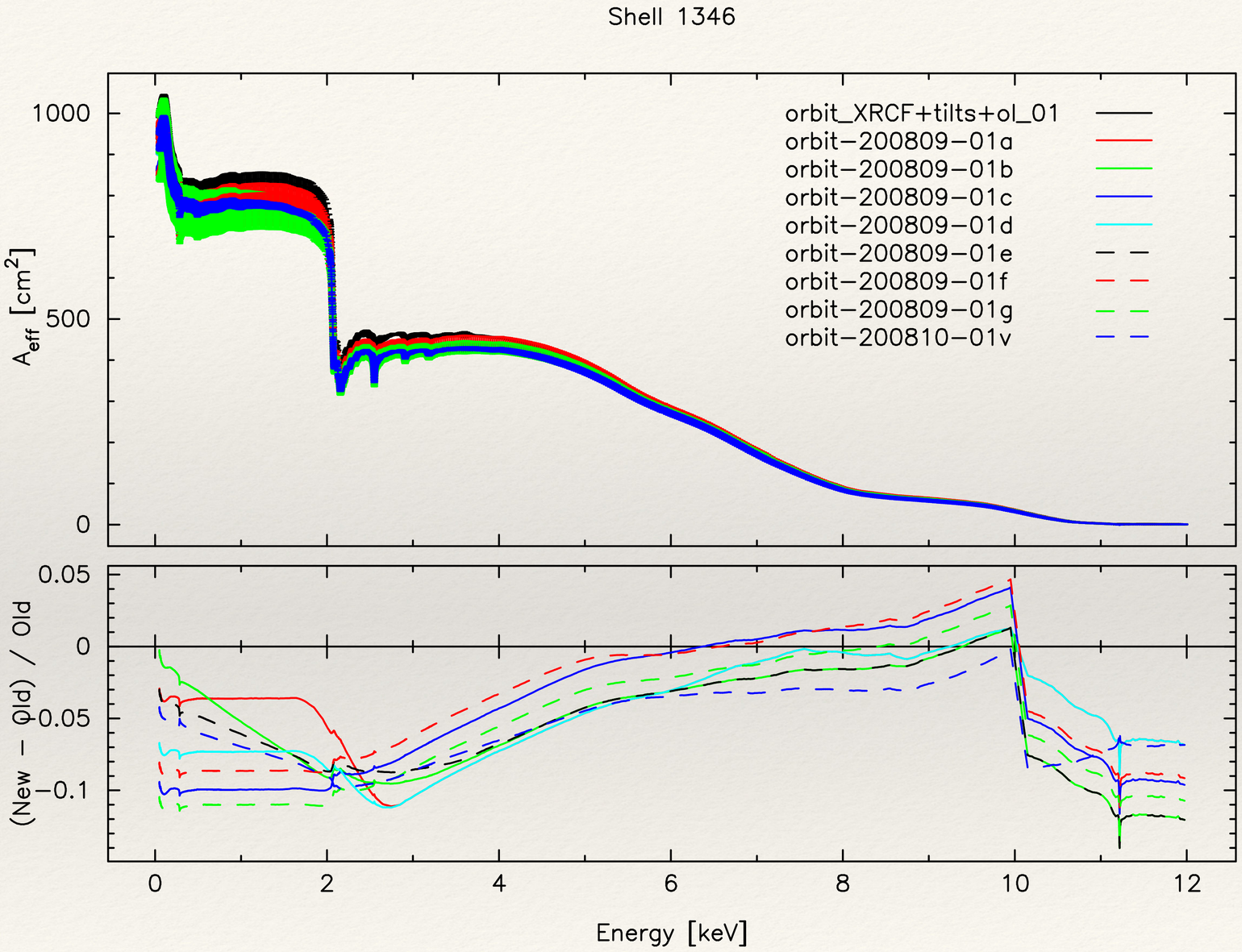
(d) As (a), but ignore FPC C-K

(e) As (b), but ignore FPC C-K

(f) Average SSD, average FPC, average the two averages

(g) As (f), but ignore FPC C-K

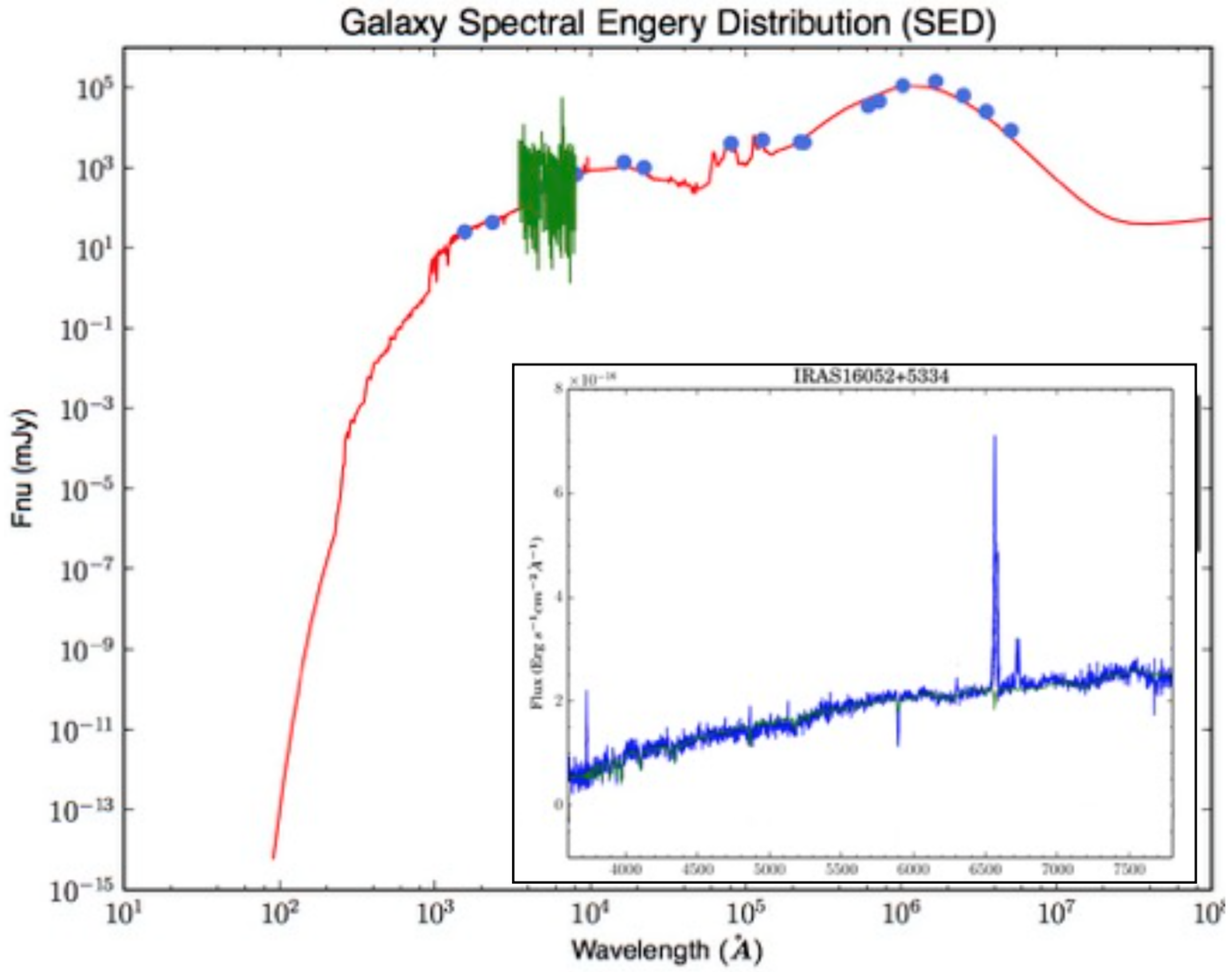
(v) straight line fits to FPC and SSD, then average the two straight lines

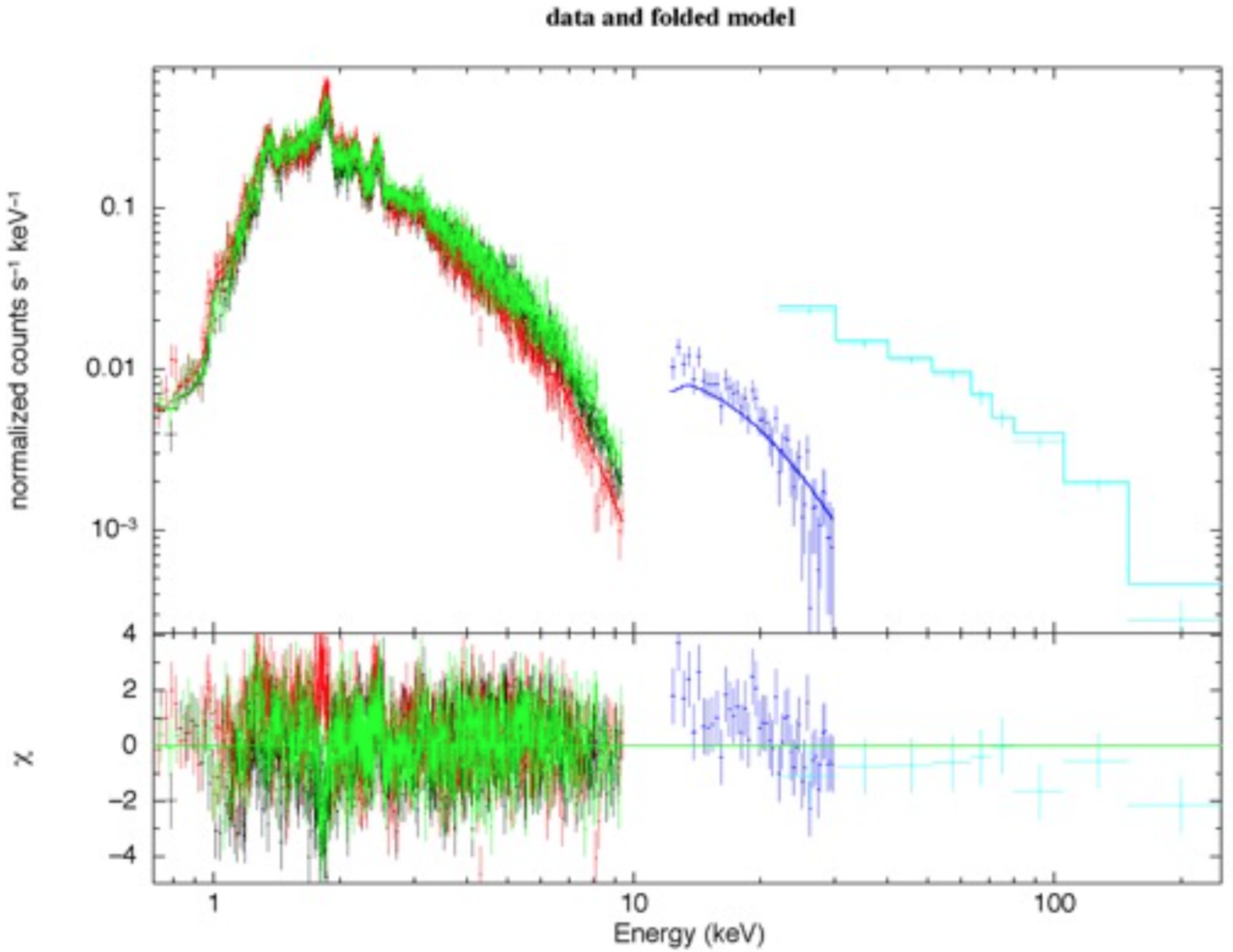


Zhao, CXC Memo 2008

1. Fitting datasets of different sizes and different qualities

When the quality of the information is not commensurate with the size of the data



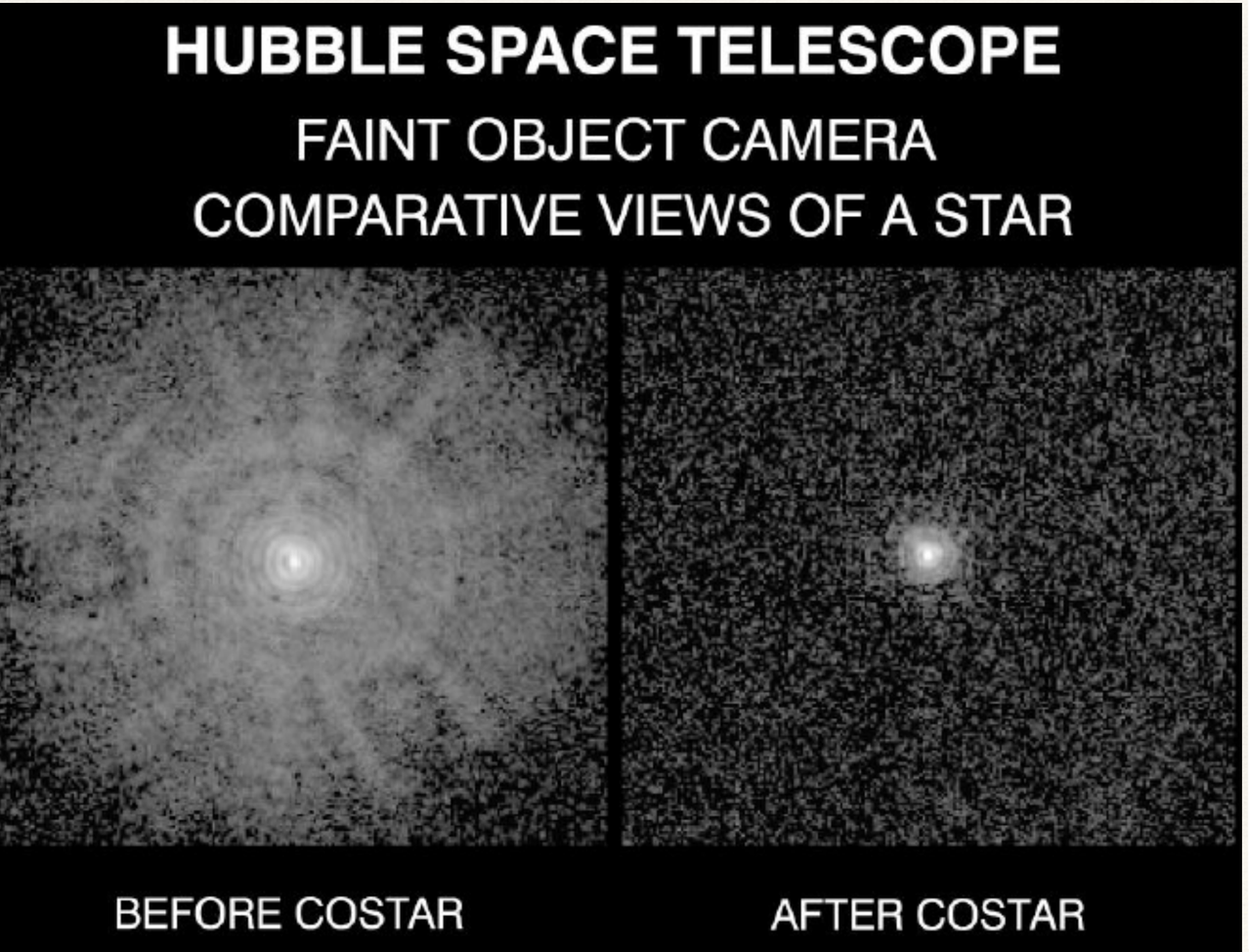


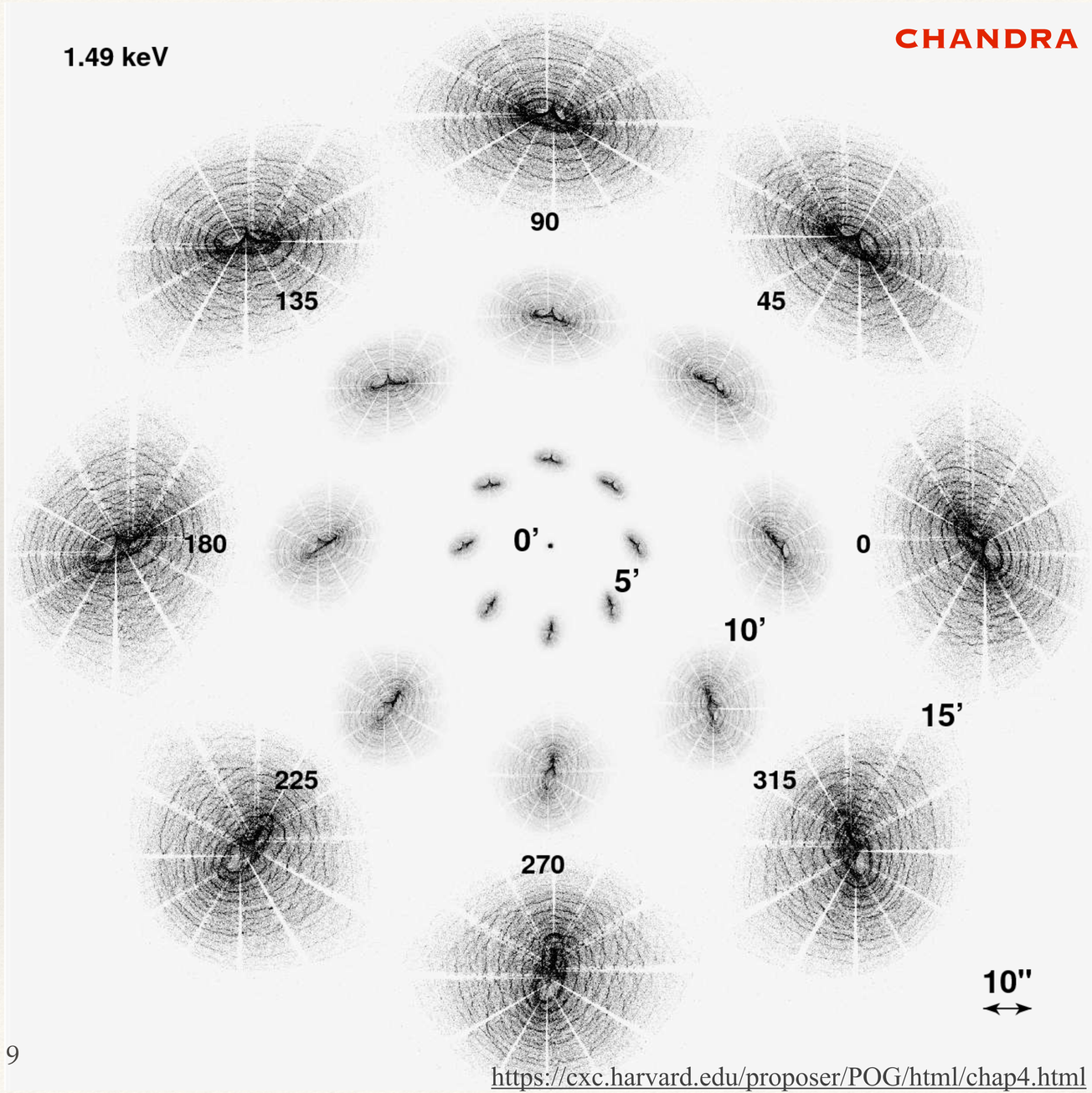
How to…

… weight datasets of different sizes and qualities in a joint analysis?

2. Uncertainties on encircled-energy contours of PSFs

PSF : Point Spread Function
aka resolution, defines image sharpness



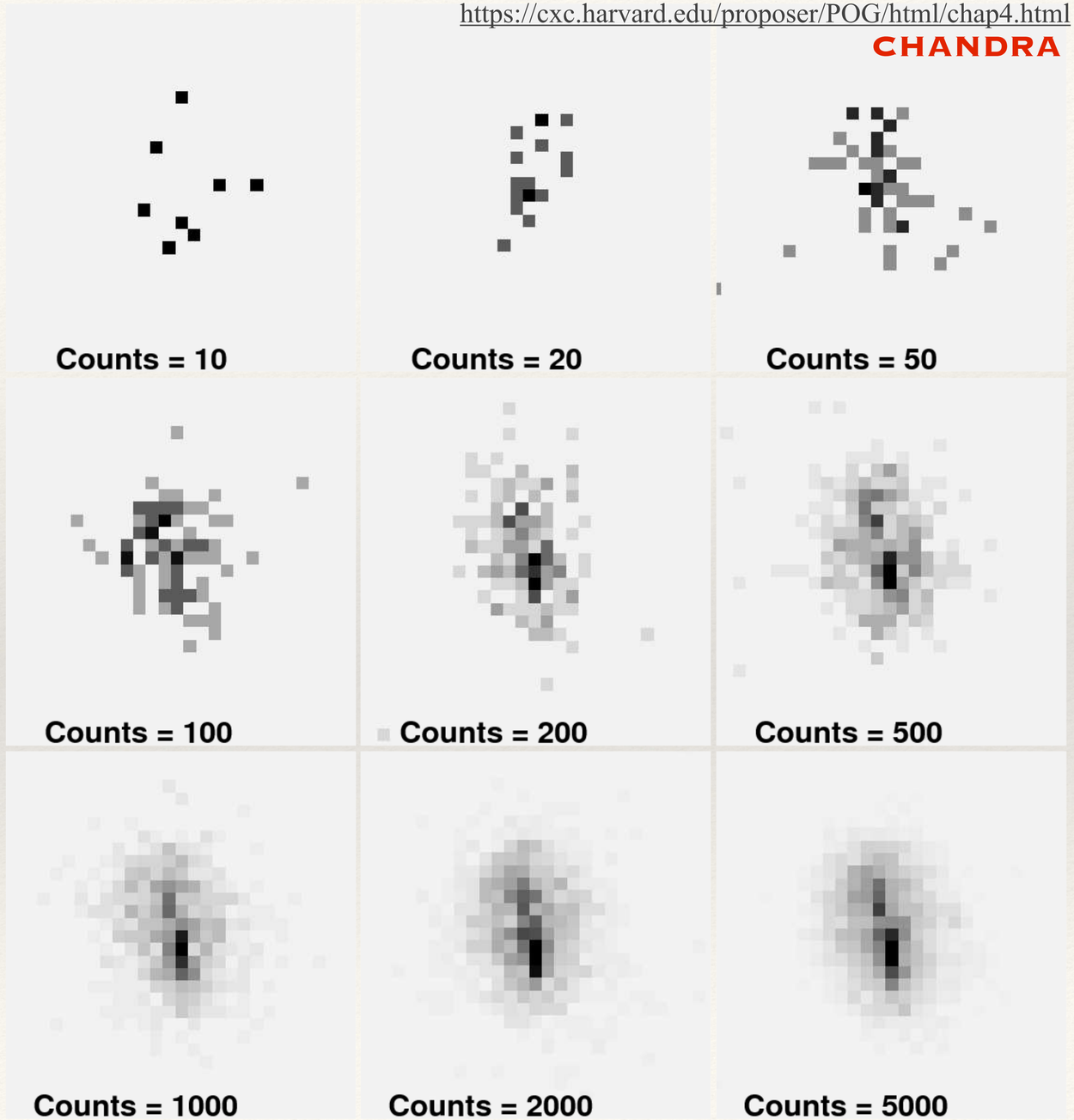


https://cxc.harvard.edu/proposer/POG/html/chap4.html

2. Uncertainties on encircled-energy contours of PSFs

https://cxc.harvard.edu/proposer/POG/html/chap4.html

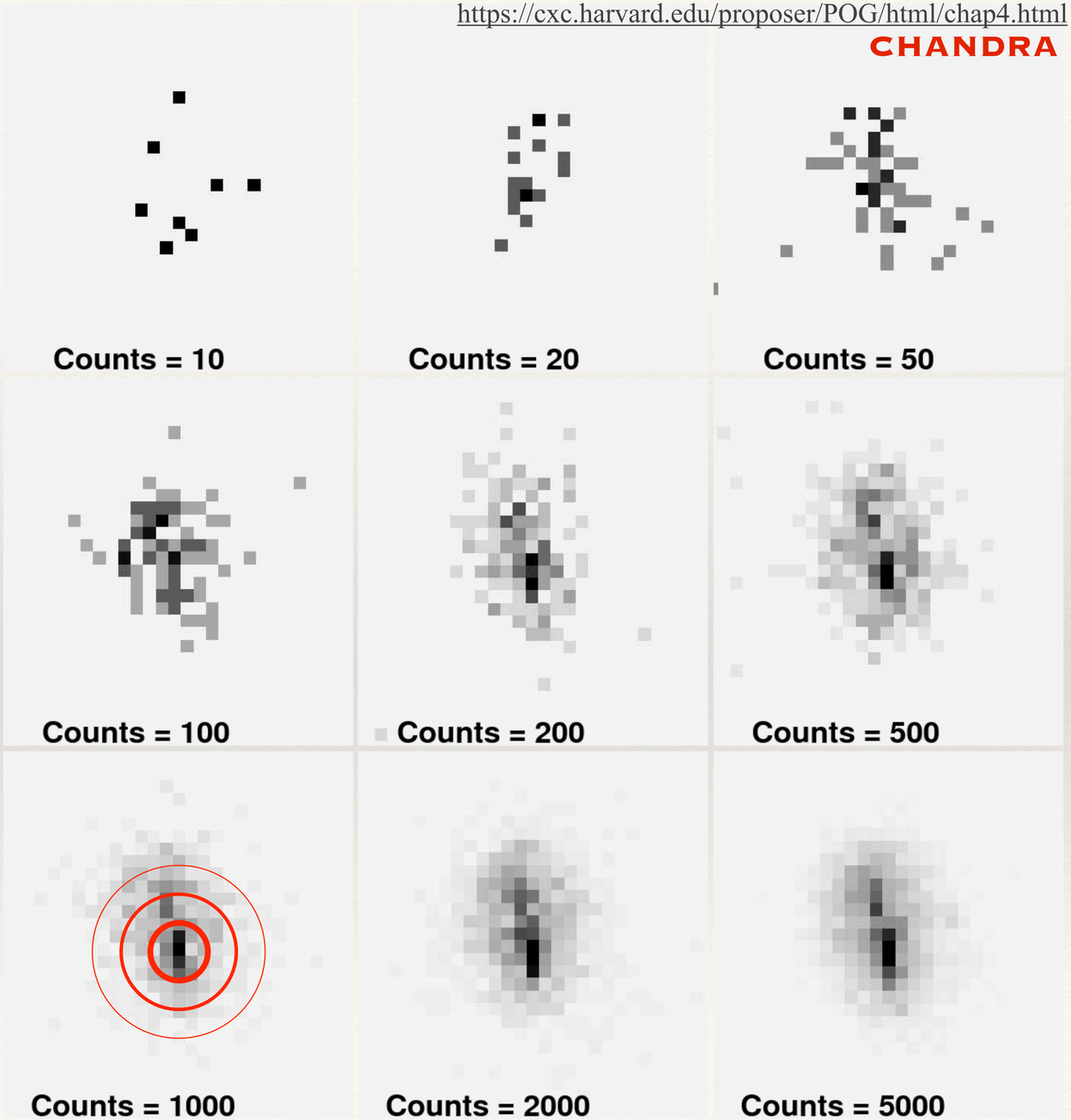
How to
— define an unambiguous enclosed-energy (EE) iso contour?
(multiple peaks, asymmetrical contours, energy dependence, Poisson counts fluctuations)
— require that contour of smaller EE should be wholly contained within contour of larger EE?
— define uncertainty intervals on the EE areas?
(Binomial distribution, maybe, for radially symmetric PSFs)
— deal with [sparse] [Poisson] background?



2. Uncertainties on encircled-energy contours of PSFs

https://cxc.harvard.edu/proposer/POG/html/chap4.html

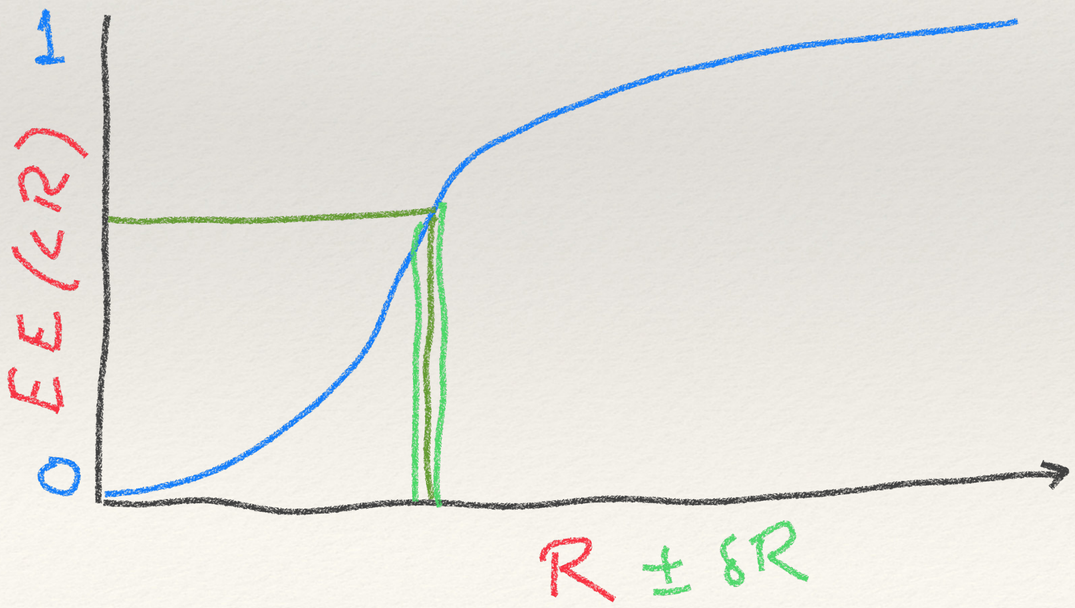
How to
— define an unambiguous enclosed-energy (EE) iso contour?
(multiple peaks, asymmetrical contours, energy dependence, Poisson counts fluctuations)
— require that contour of smaller EE should be wholly contained within contour of larger EE?
— define uncertainty intervals on the EE areas?
(Binomial distribution, maybe, for radially symmetric PSFs)
— deal with [sparse] [Poisson] background?

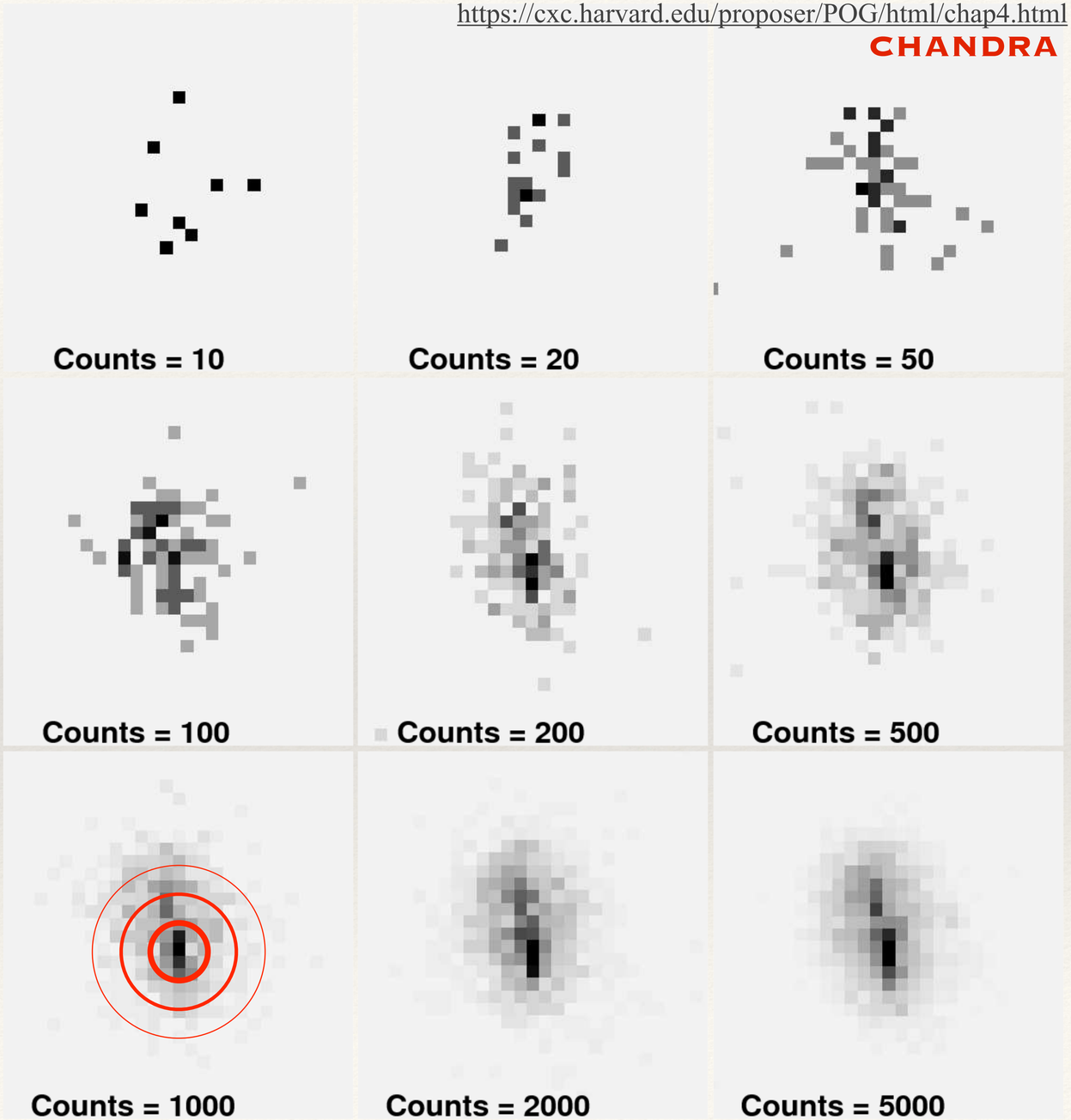


2. Uncertainties on encircled-energy contours of PSFs

https://cxc.harvard.edu/proposer/POG/html/chap4.html

How to
— define an unambiguous enclosed-energy (EE) iso contour?
(multiple peaks, asymmetrical contours, energy dependence, Poisson counts fluctuations)
— require that contour of smaller EE should be wholly contained within contour of larger EE?
— define uncertainty intervals on the EE areas?
(Binomial distribution, maybe, for radially symmetric PSFs)
— deal with [sparse] [Poisson] background?





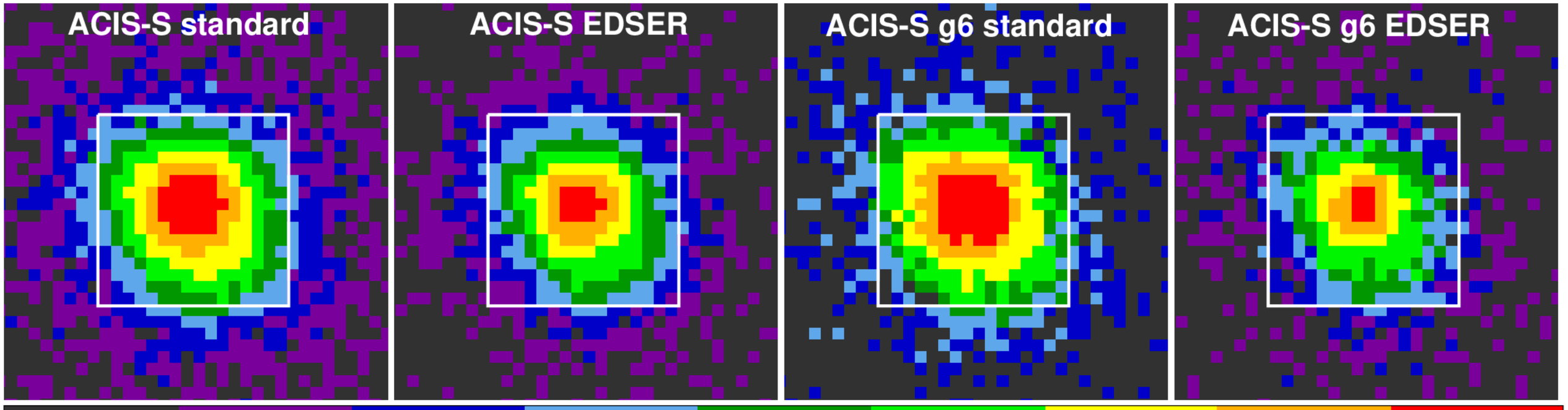
How to…

… weight datasets of different sizes and qualities in a joint analysis?

… characterize shapes, sizes, and uncertainties of complex PSFs?

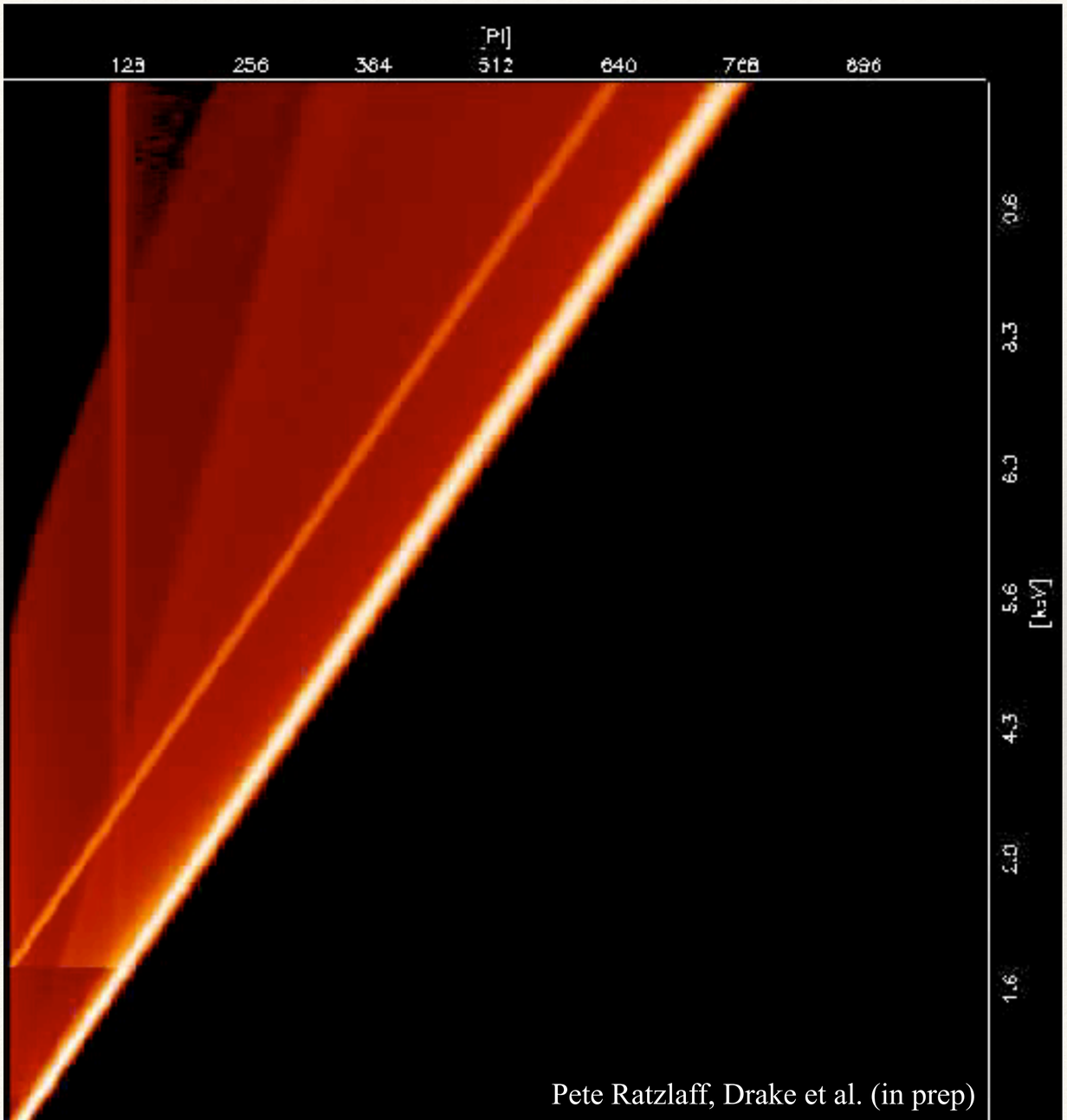
3. Systematic uncertainties in point spread and spectral redistribution

Ping Zhao et al. (in prep)



Variations in observed Chandra ACIS Point Spread Function
showing variations due to
- event quality (where within the pixel does the photon land)
- processing (different corrections for event quality), and
- Poisson fluctuations in the wings

3. Systematic uncertainties in point spread and spectral redistribution



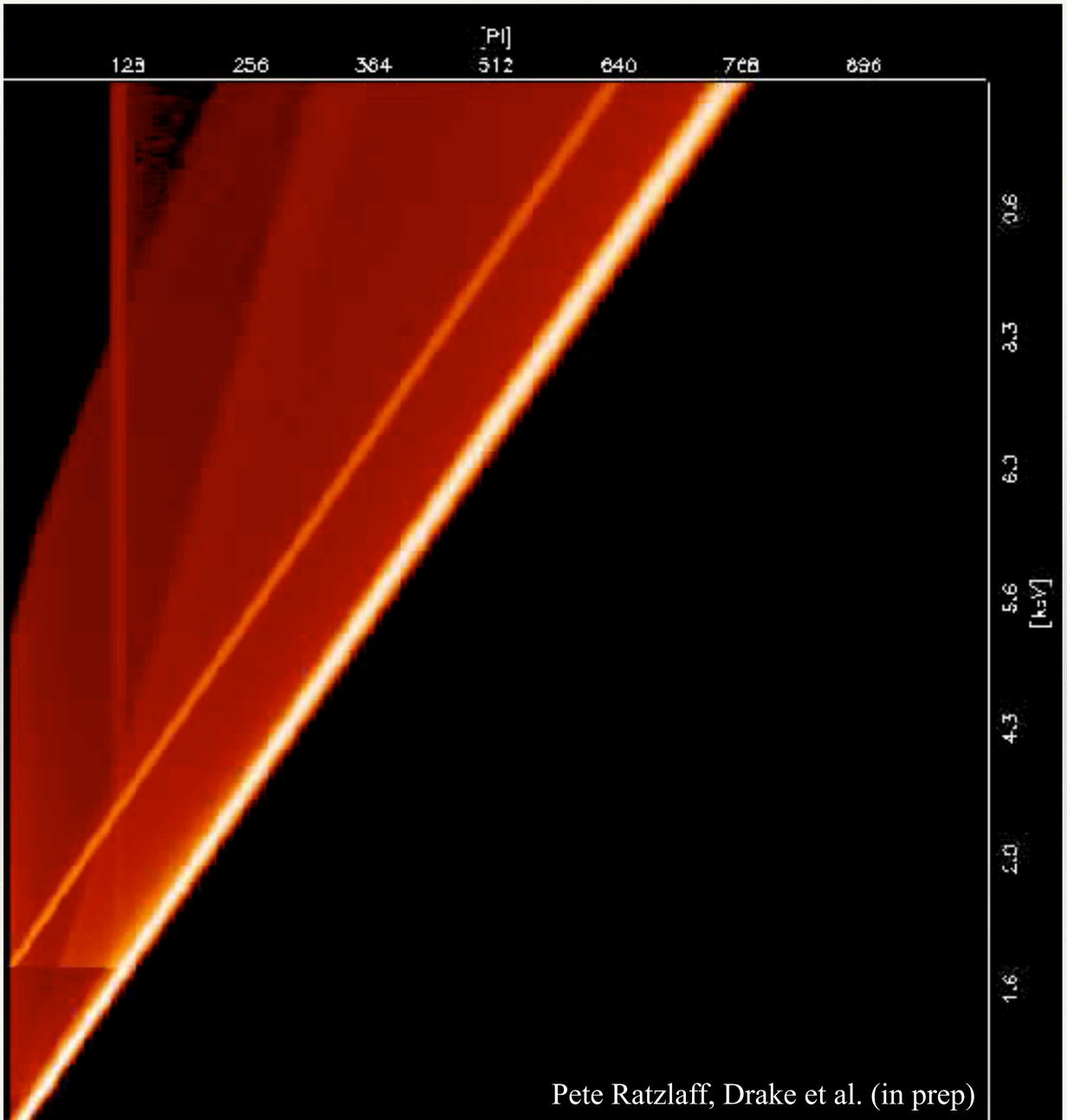
Pete Ratzlaff, Drake et al. (in prep)

Modeling the Unknown Knowns

Incoming photons are "scrambled" by a probabilistic instrument response. The instrument response itself is known only imperfectly.

Consequences: misidentification of atomic emission lines, bias in estimation of source properties like plasma temperature.

3. Systematic uncertainties in point spread and spectral redistribution



Pete Ratzlaff, Drake et al. (in prep)

Modeling the Unknown Knowns

Incoming photons are "scrambled" by a probabilistic instrument response. The instrument response itself is known only imperfectly.

Consequences: misidentification of atomic emission lines, bias in estimation of source properties like plasma temperature.

How to…

… weight datasets of different sizes and qualities in a joint analysis?

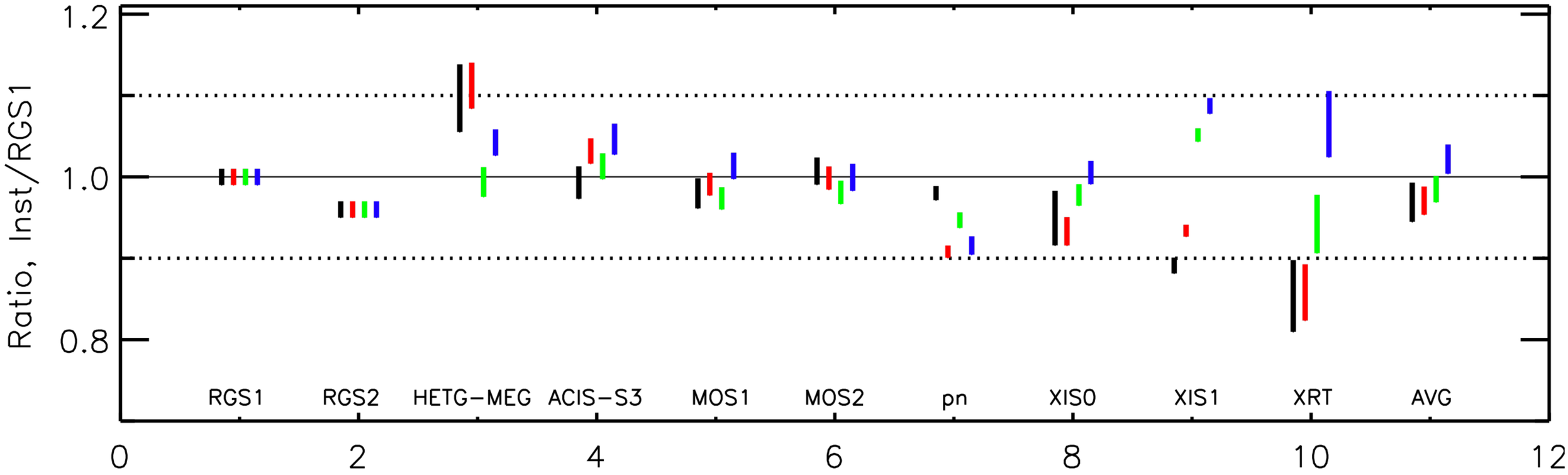
… characterize shapes, sizes, and uncertainties of complex PSFs?

… understand uncertainties in PSF and RMFs?

Effective Area
photon collecting geometric area, reduced by efficiency of observing a photon of given energy

4. Calibration of effective areas of multiple telescopes

OVII black **OVIII red** **NeIX green** **NeX blue** ***DePasquale(SAO)***



5

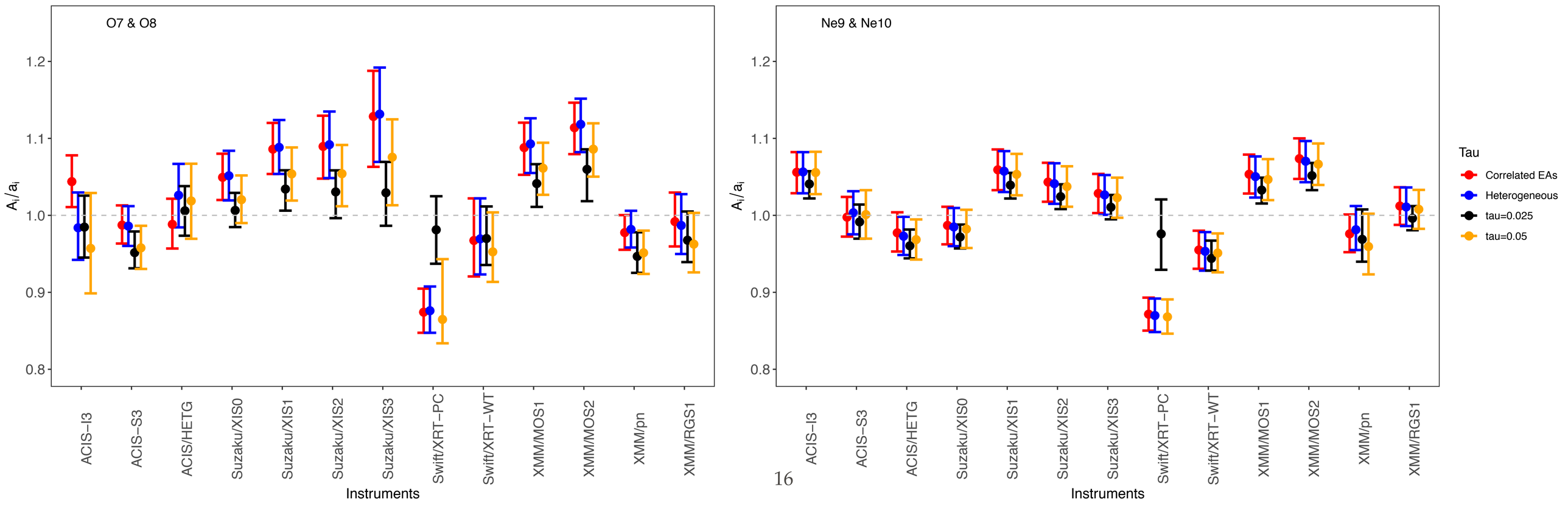
Paul Plucinsky ***IACHEC April 2011***

https://iachec.org/wp-content/presentations/2011/Plucinsky.pdf

4. Calibration of effective areas of multiple telescopes

- Chen, Y., et al., 2018, *Calibration Concordance for Astronomical Instruments via Multiplicative Shrinkage*, Journal of the American Statistical Association, DOI: 10.1080/01621459.2018.1528978

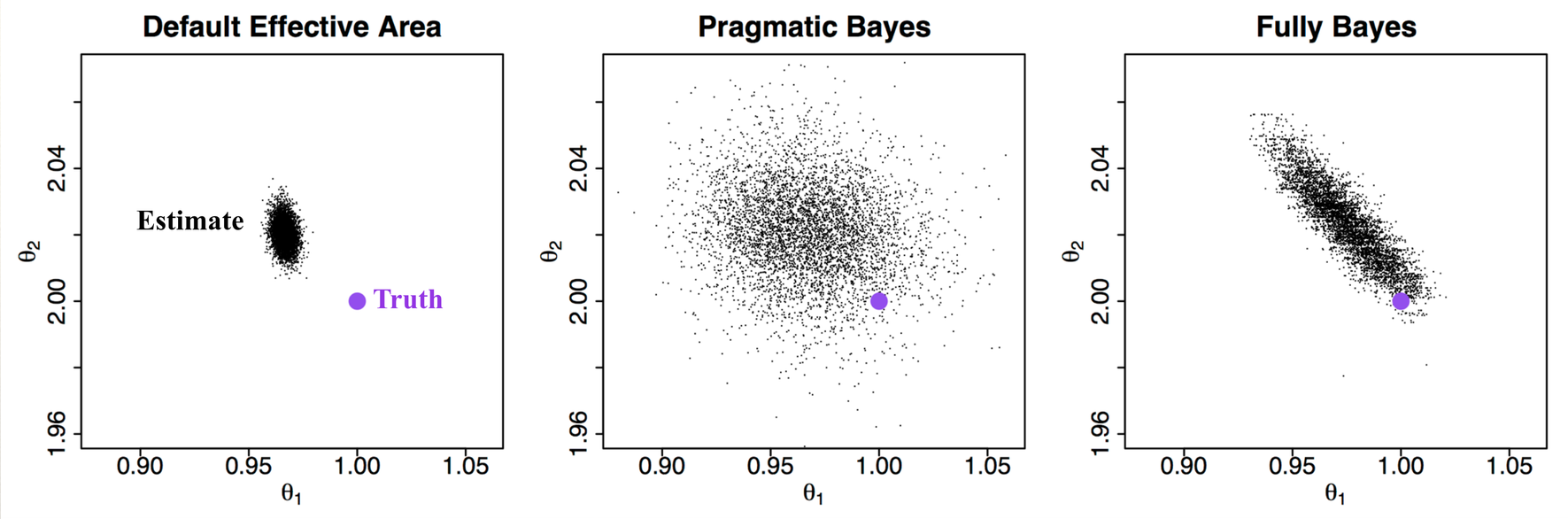
- Marshall, H., et al., 2021, *Concordance: In-flight Calibration of X-ray Telescopes without Absolute References*, Astrophysical Journal, in review



4. Calibration of effective areas of multiple telescopes

Pragmatic Bayes and Full Bayes

- Lee, H., et al., 2011, *Accounting for Calibration Uncertainties in X-ray Analysis: Effective Areas in Spectral Fitting*, Astrophysical Journal, 731, 126
- Xu, J., et al., 2014, *A Fully Bayesian Method for Jointly Fitting Instrumental Calibration and X-ray Spectral Models*, Astrophysical Journal, 794, 97
- Yu, X., et al., 2018, *Incorporating Uncertainties in Atomic Data into the Analysis of Solar and Stellar Observations: A Case Study in Fe XIII*, Astrophysical Journal, 866, 146
- Yu, X., et al., 2021, *Effect of Systematic Uncertainties on Density and Temperature Estimates in Coronae of Capella*, in preparation

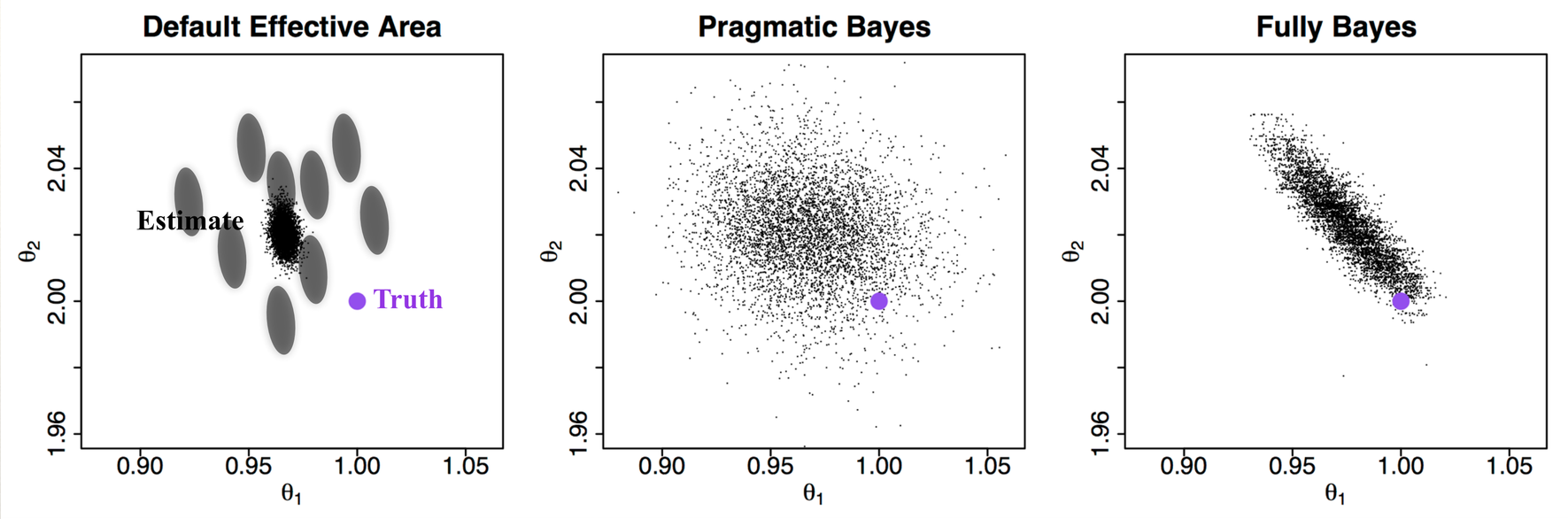


$p(\theta|D,\varepsilon^{(def)})$

4. Calibration of effective areas of multiple telescopes

Pragmatic Bayes and Full Bayes

- Lee, H., et al., 2011, *Accounting for Calibration Uncertainties in X-ray Analysis: Effective Areas in Spectral Fitting*, Astrophysical Journal, 731, 126
- Xu, J., et al., 2014, *A Fully Bayesian Method for Jointly Fitting Instrumental Calibration and X-ray Spectral Models*, Astrophysical Journal, 794, 97
- Yu, X., et al., 2018, *Incorporating Uncertainties in Atomic Data into the Analysis of Solar and Stellar Observations: A Case Study in Fe XIII*, Astrophysical Journal, 866, 146
- Yu, X., et al., 2021, *Effect of Systematic Uncertainties on Density and Temperature Estimates in Coronae of Capella*, in preparation



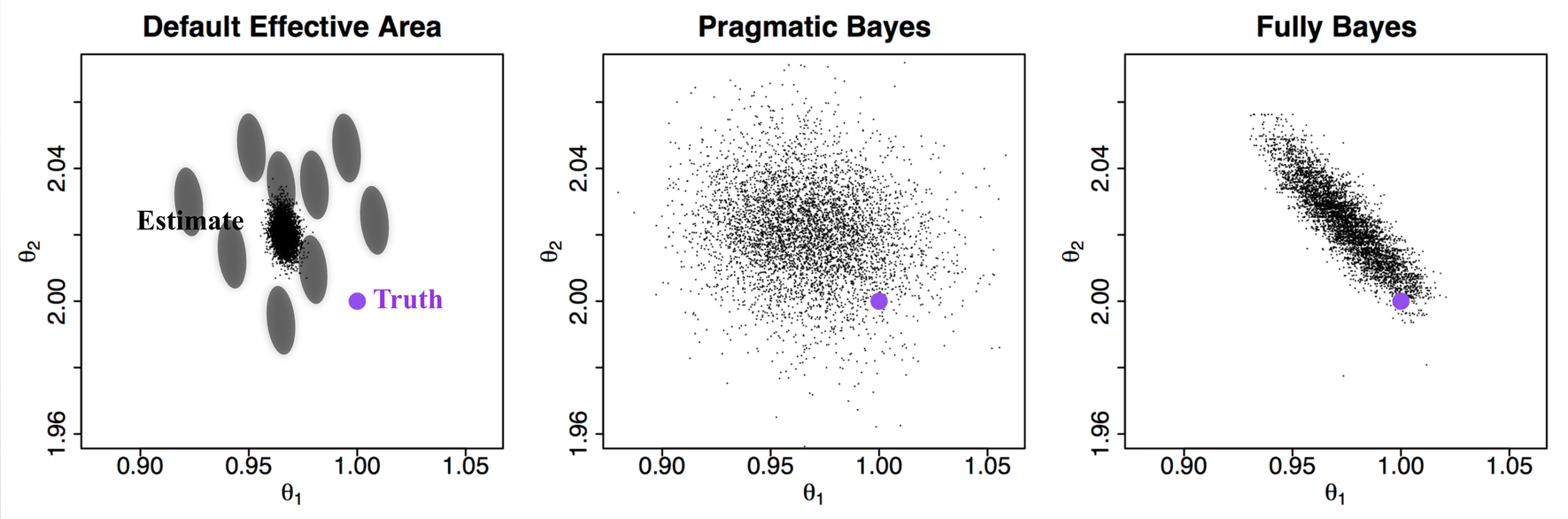
$p(\theta|D,\varepsilon^{(def)})$

$p(\theta|D,\varepsilon^{(m)})$

4. Calibration of effective areas of multiple telescopes

Pragmatic Bayes and Full Bayes

- Lee, H., et al., 2011, *Accounting for Calibration Uncertainties in X-ray Analysis: Effective Areas in Spectral Fitting*, Astrophysical Journal, 731, 126
- Xu, J., et al., 2014, *A Fully Bayesian Method for Jointly Fitting Instrumental Calibration and X-ray Spectral Models*, Astrophysical Journal, 794, 97
- Yu, X., et al., 2018, *Incorporating Uncertainties in Atomic Data into the Analysis of Solar and Stellar Observations: A Case Study in Fe XIII*, Astrophysical Journal, 866, 146
- Yu, X., et al., 2021, *Effect of Systematic Uncertainties on Density and Temperature Estimates in Coronae of Capella*, in preparation



$p(\theta|D,\varepsilon^{(def)})$

$p(\theta|D,\varepsilon^{(m)})$

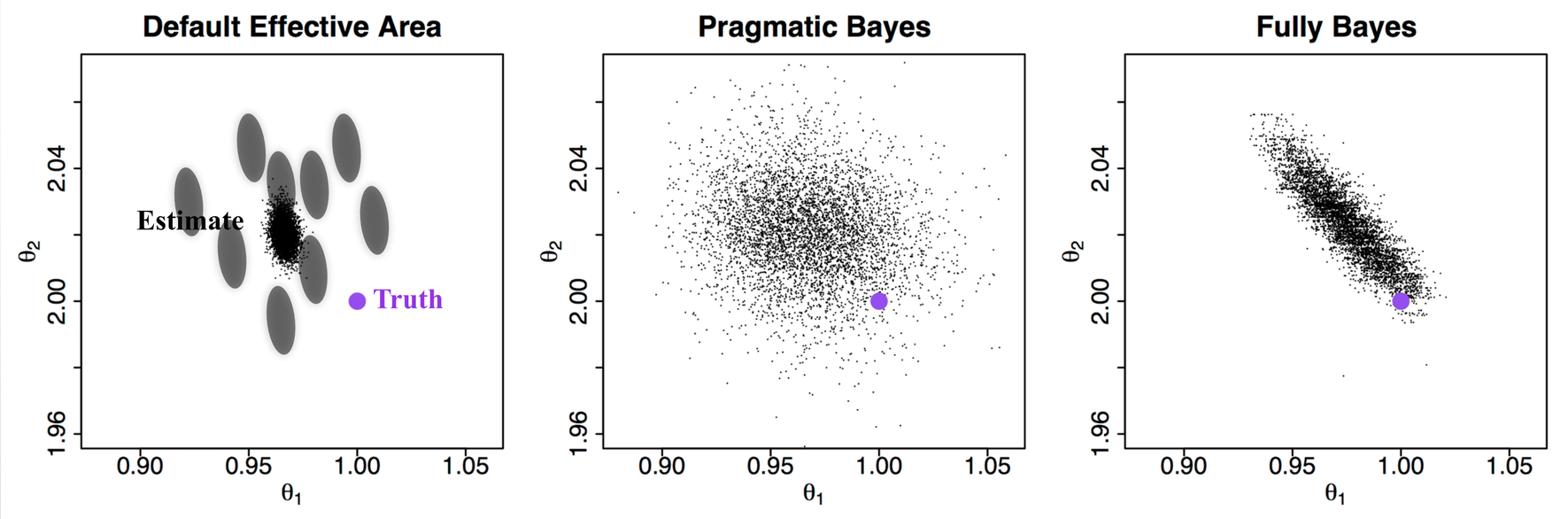
$p(\varepsilon,\theta|D)$

$\rightarrow p(\theta|D,\varepsilon) \cdot p(\varepsilon)$

4. Calibration of effective areas of multiple telescopes

Pragmatic Bayes and Full Bayes

- Lee, H., et al., 2011, *Accounting for Calibration Uncertainties in X-ray Analysis: Effective Areas in Spectral Fitting*, Astrophysical Journal, 731, 126
- Xu, J., et al., 2014, *A Fully Bayesian Method for Jointly Fitting Instrumental Calibration and X-ray Spectral Models*, Astrophysical Journal, 794, 97
- Yu, X., et al., 2018, *Incorporating Uncertainties in Atomic Data into the Analysis of Solar and Stellar Observations: A Case Study in Fe XIII*, Astrophysical Journal, 866, 146
- Yu, X., et al., 2021, *Effect of Systematic Uncertainties on Density and Temperature Estimates in Coronae of Capella*, in preparation



$p(\theta|D,\varepsilon^{(def)})$

$p(\theta|D,\varepsilon^{(m)})$

$p(\varepsilon,\theta|D)$

$\rightarrow p(\theta|D,\varepsilon) \cdot p(\varepsilon)$

$p(\varepsilon,\theta|D)$

$\rightarrow p(\theta|D,\varepsilon) \cdot p(\varepsilon|D)$

How to…

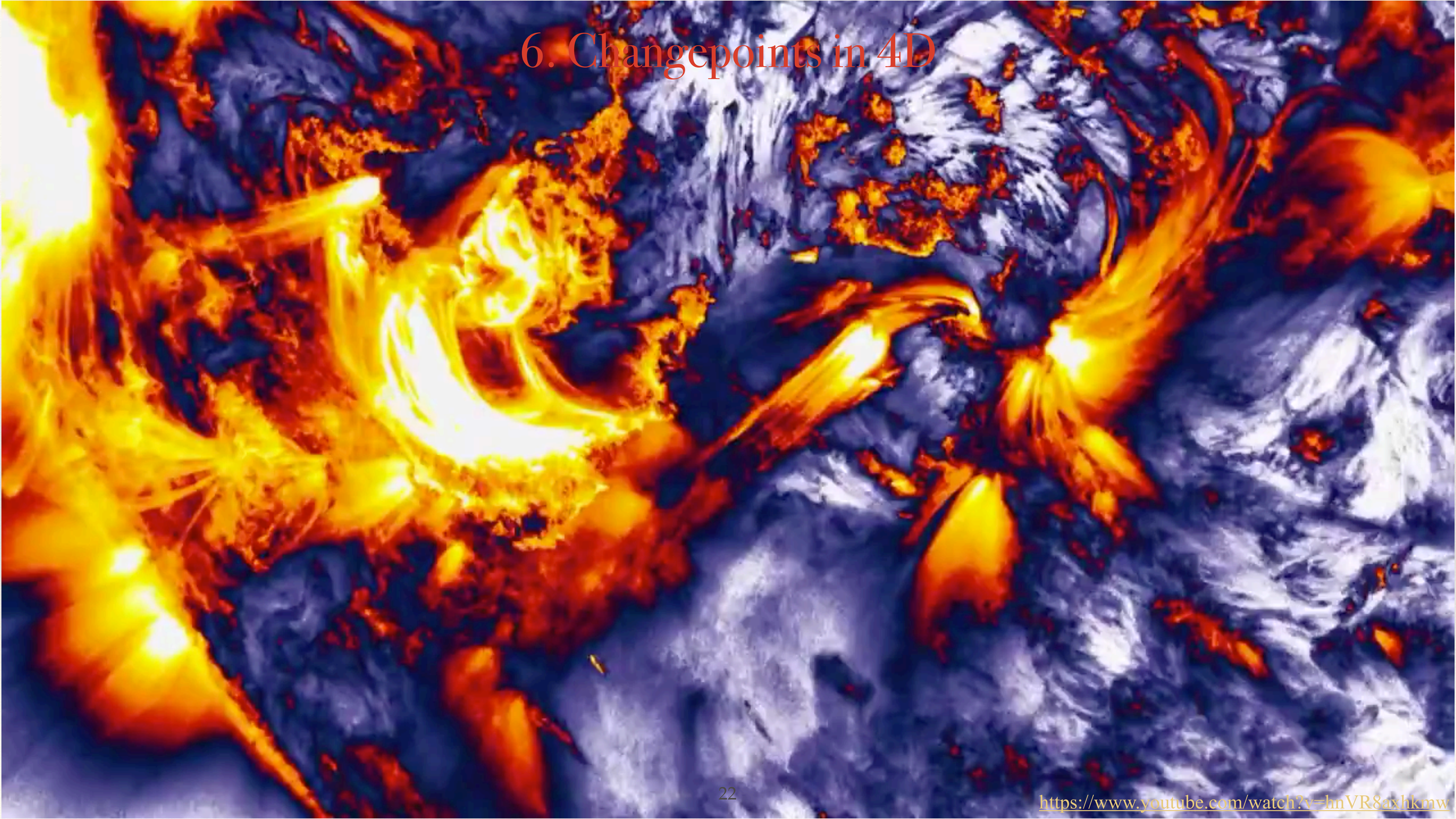
… weight datasets of different sizes and qualities in a joint analysis?

… characterize shapes, sizes, and uncertainties of complex PSFs?

… understand uncertainties in PSF and RMFs?

… incorporate systematics and correct for them?

6. Changepoints in 4D



https://www.youtube.com/watch?v=hnVR8axhkmw

How to…

… weight datasets of different sizes and qualities in a joint analysis?

… characterize shapes, sizes, and uncertainties of complex PSFs?

… understand uncertainties in PSF and RMFs?

… incorporate systematics and correct for them?

… model only the known knowns while [not] ignoring the known unknowns?

… find changepoints in a 2/3/4D datacube?

… robustly segment sparse data?

… locate the boundary of an extended source?