Progress Report

Ian Evans
On behalf of the Chandra Source Catalog Project Team

Chandra Users' Committee Meeting
September 29, 2015

Highlights

- Started release 2 production run (*through* `stacker` *pipeline*) on 4/13/15

  Roughly 70% of observation stacks have completed processing as of 9/17/15
  - Total of 245,245 "TRUE" or "MARGINAL" detections using (somewhat conservative) detection thresholds used for preliminary detections list
    - Expect number of detections to increase once final calibration of thresholds is completed
- Published a (partial) preliminary list of detections on 8/3/15
  - Properties include positions (and errors), and proxies for aperture photometry
  - Content agreed to ahead of time by the CUC
  - Will continue to update as processing progresses
- post-`stacker` pipeline development in progress
  - post-`stacker` pipeline thread defined
  - Most complex algorithms have completed development or are currently under development, or (in one case) are ready to start development
  - Remaining algorithms that are ready to start development or waiting on spec are mostly evolutions of the release 1 versions
- Outreach
  - Poster at the AAS Special HEAD meeting in June/July
  - NASA Hyperwall presentations and catalog demos at IAU in August
  - Invited talk at ADASS 2015 in October

Usage Statistics

Current Release

- Catalog version: 1.1; Released: 2010 Aug 10
  - 106,586 master sources
  - 158,071 source detections
  - 5,110 observations with at least one detected source
- Subset of master source properties are available via HEASARC Browse, NED, and Vizier services
  - Usage statistics reported below do not include accesses via these services

Usage Statistics

Release 1.1	Reporting Period 2014 Oct 01 – 2015 Aug 31	
	Number	**% Non-CfA**
CSCview catalog browser initializations	145 /month	91%
CSCview catalog browser properties searches	206 /month	90%
Command-line (CLI) searches	997 /month*	73%
VO cone searches	7369 /month	94%
CSC Sky in Google Earth	513 visits/month	

* Excludes 30K searches (~ all non-CfA) from 2015 June

Software Release Summary and Production Status

Software Releases Completed

Release	Date	Summary
CAT 4.3.8	Apr 16	Bugfix patch
CAT 4.3.9	Apr 30	QA/corner cases
CAT 4.3.10	May 11	QA fixes
CAT 4.3.11	Jun 24	QA, MLE, `dmellipse`, repro upgrades
CAT 4.3.12	Jun 25	Addressed QA divide by zero error
CAT 4.3.13	Jul 07	Address stack library issues
CAT 4.3.14	Jul 09	PL/QA fixes for repros
CAT 4.3.15	Sep 04	Address long runs of `mkvtbkg` by creating smaller tiles

Release 2 Production Processing Statistics

- Through 2015 September 21, 18:20Z
  - Stacks submitted to processing: 5898
  - Completed processing/ingested into archive: 5359 (73.4% of stacks)
  - Currently processing: 106
  - Waiting for manual QA: 433[1]
  - Number of detections classified as TRUE: 166754
  - Number of detections classified as MARGINAL: 78491
  - Total number of TRUE or MARGINAL detections: 245245[2]

[1] Includes 175 forced manual QA with no other issues

[2] For comparison, the total number of detections in Rel. 1.1 was 158071

Preliminary Detections List

Content

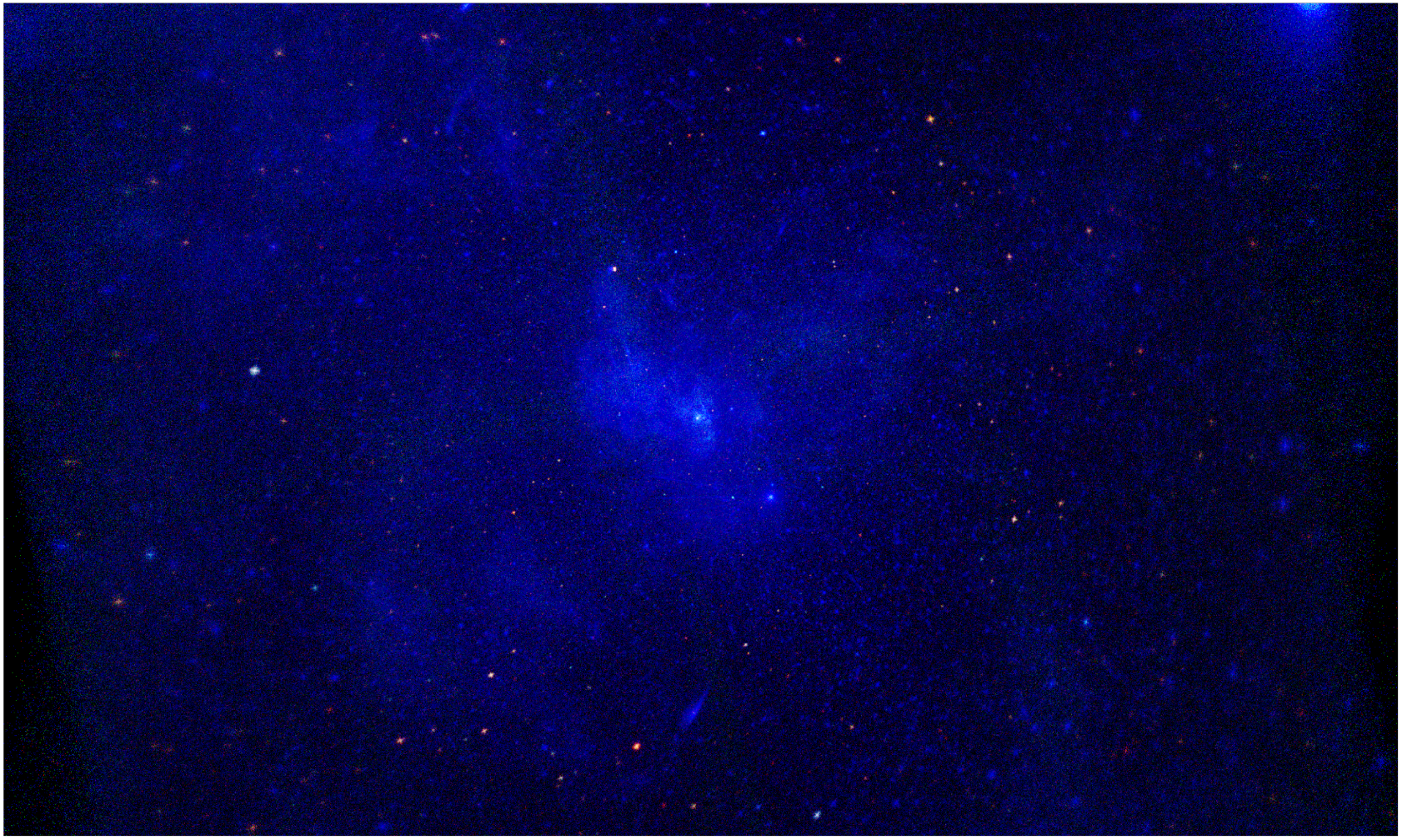
- A single FITS table containing a subset of information for all of the compact detections with preliminary classification TRUE or MARGINAL
- The table includes the following information:
  - Detection likelihood and preliminary classification
  - Fitted position and 95% internal-error error ellipse
  - Fitted detection amplitude and confidence interval in 4 bands for ACIS (broad, soft, medium, hard) and 1 band for HRC-I (wide)
    - Amplitude is a good proxy for aperture photometry for TRUE point sources
  - Classification of detection as point or extended
  - Fitted ellipse parameters for extended detections
- Documentation in the form of column descriptions and caveats is provided with the FITS data product

Preliminary Detections List (cont.)

Initial Release

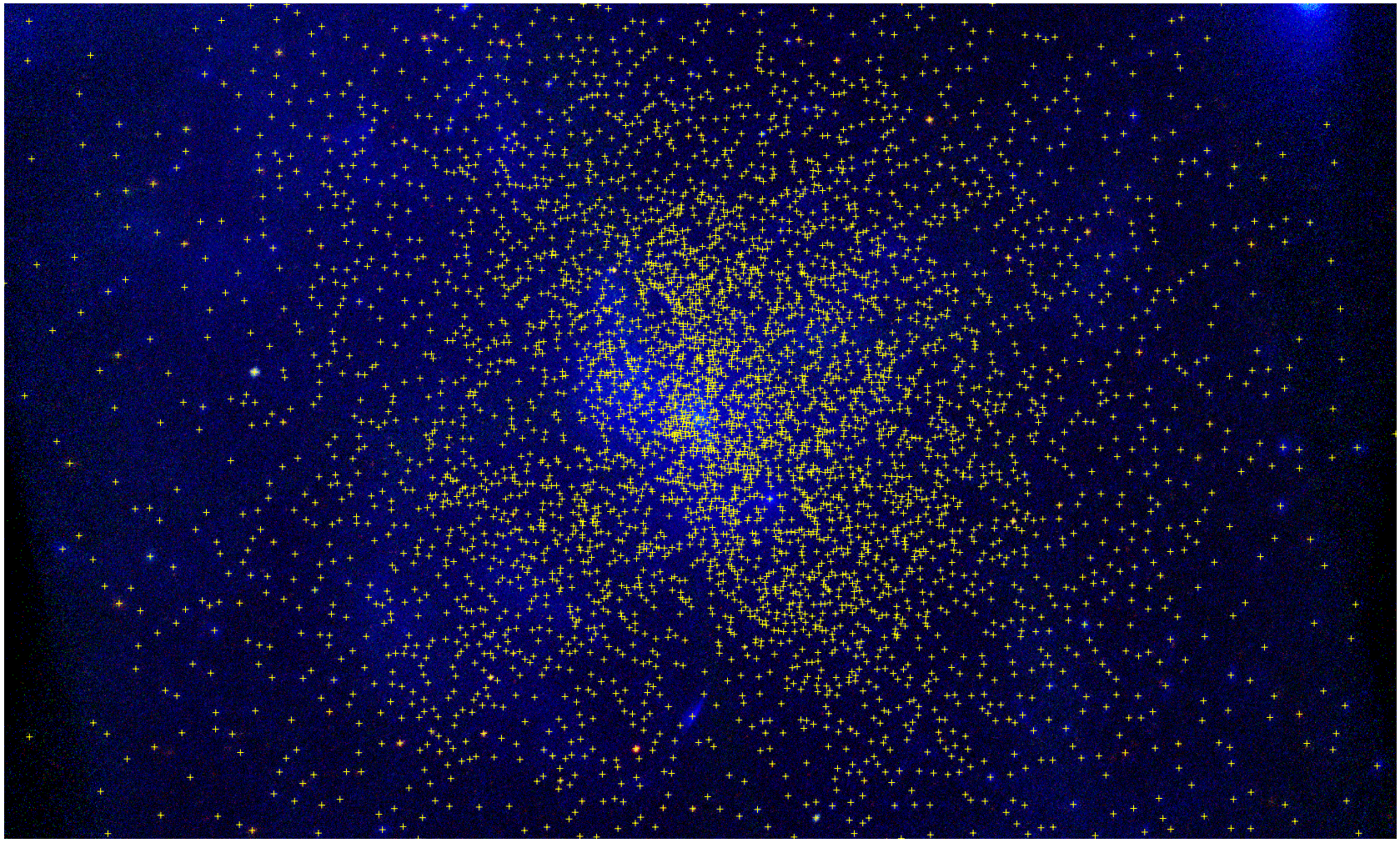
- The initial version of the preliminary detections list FITS table was released on 8/3/15 via the release 2 website http://cxc.cfa.harvard.edu/csc2/
- Included 241,074 detections
  - 163,869 identified as "TRUE" detections and 77,205 "MARGINAL" detections
    - Classifications are based on preliminary (pre-production) likelihood thresholds
  - 237,892 point source detections, 2,716 extended compact source detections, and 466 ambiguous
  - 186,238 detections with highest likelihood in the ACIS broad band, 18,895 in the ACIS hard band, 16,876 in the ACIS medium band, 17,200 in the ACIS soft band, 195 in the ACIS ultra-soft band, and 1,670 in the HRC wide band
- The availability of the preliminary detections list was announced through the standard CXC channels
  - Including the CSC website, *Chandra* Electronic Announcement, and Twitter
- The preliminary detections list will be updated periodically as processing proceeds
  - Additional announcements will be made in association with these updates

Release 2.0 Stack Example



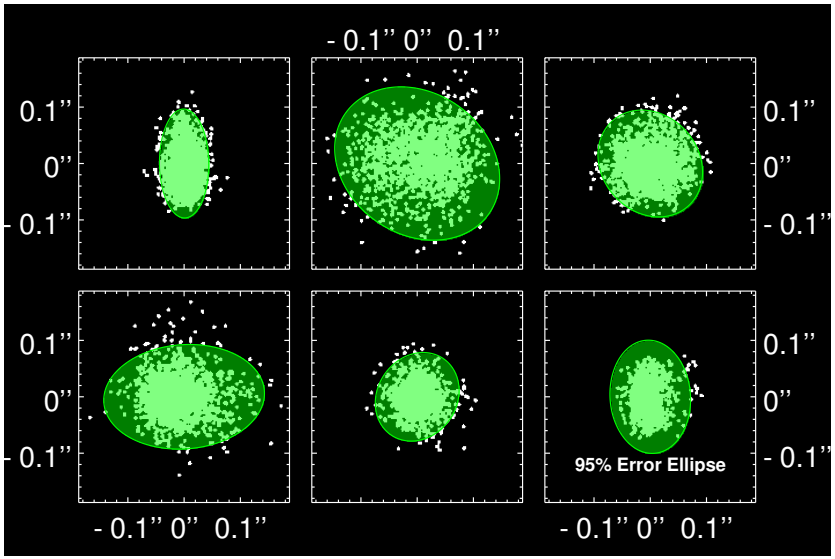
- Sgr A* — ACIS stack `acisfJ1745401m290028_001`
- 71 observations totaling ~2.2 Ms exposure
- Field is ~18 arcmin across
- Red = soft energy band (0.5–1.2 keV), green = medium energy band (1.2–2.0 keV), blue = hard energy band (2.0 –7.0 keV)

Release 2.0 Stack Example (cont.)



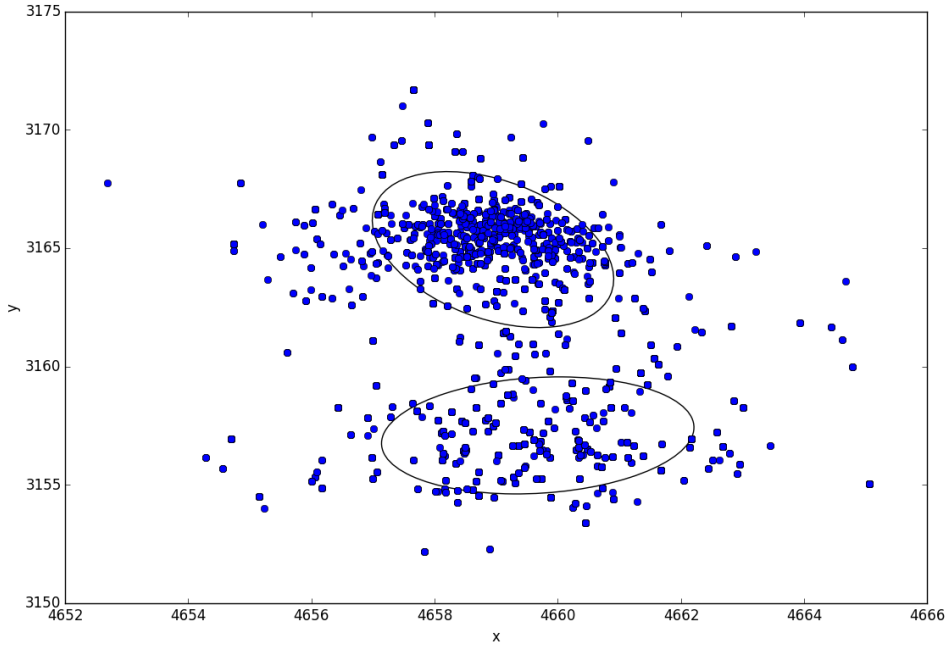
- Sgr A* — ACIS stack `acisfJ1745401m290028_001`
- 71 observations totaling ~2.2 Ms exposure
- Field is ~18 arcmin across
- Preliminary detections shown by yellow crosses
  - There are 4674 preliminary detections in this field; classification as TRUE, MARGINAL, or FALSE has not yet been completed

Source Position Uncertainties



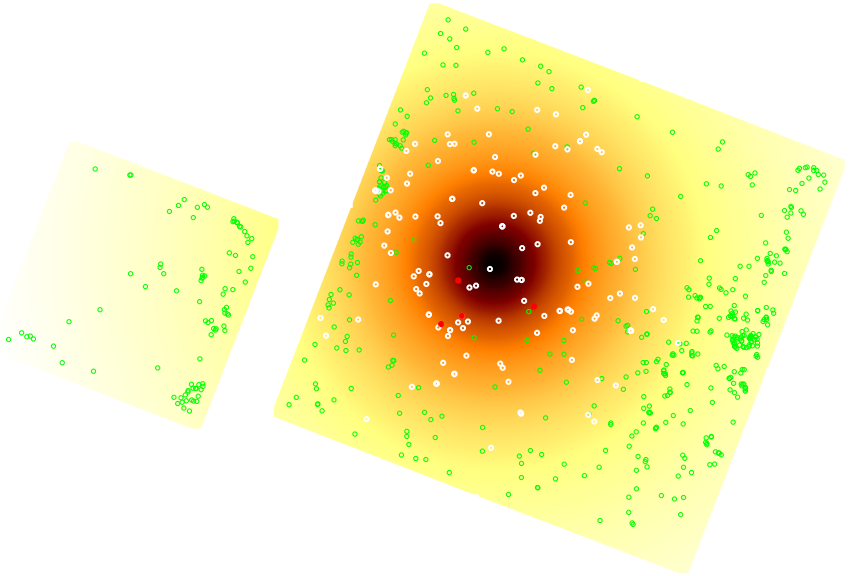
- *Left:* Reported position uncertainties for TRUE detections are computed from *Sherpa* pyBLoCXS MCMC position probability analyses for each source
- An *error ellipse* parameterization of the draws will be included in the catalog database
- The actual MCMC draws will be available as FITS files

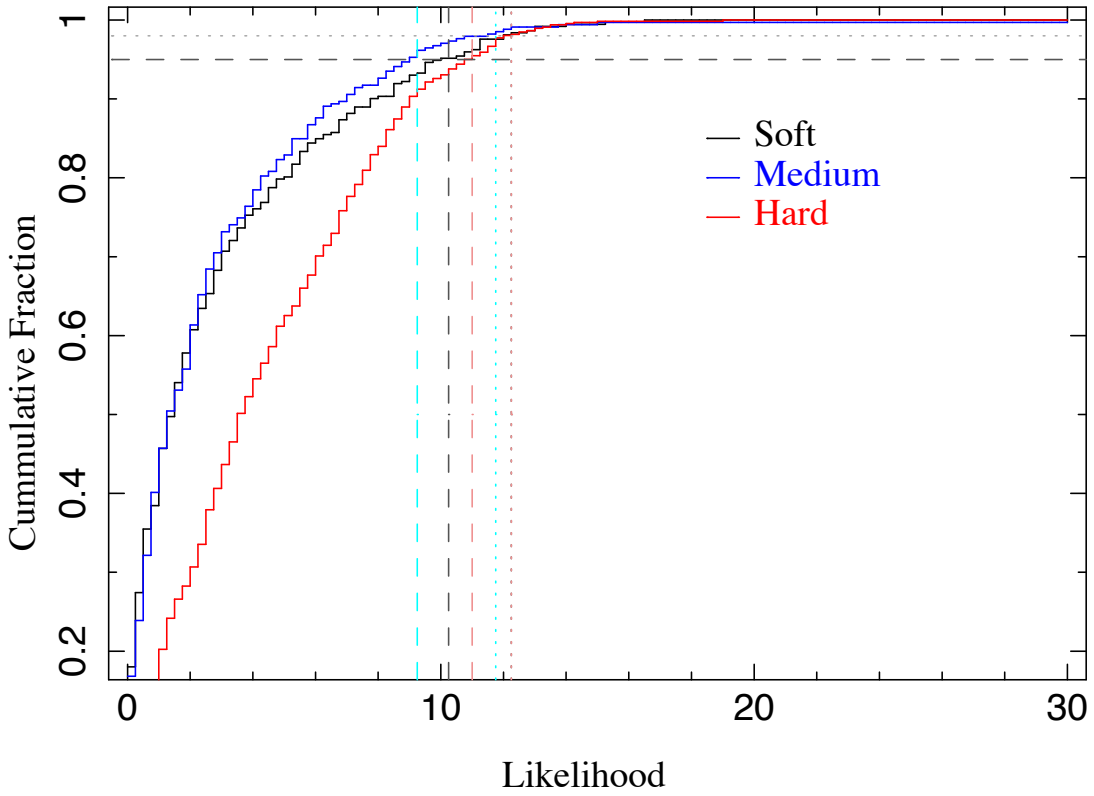
- *Right:* Multiple very close and overlapping sources not detected separately may be identified from the MCMC draws using Gaussian Mixture Models, with the appropriate number of components determined by AIC or BIC (Akaike or Bayes Information Criterion)
- In release 2.0, such detections will be flagged in the catalog
- In a future catalog release, feedback to the *Sherpa* maximum likelihood estimator may permit detailed multi-source fits to very close and overlapping detections



Detection Thresholds

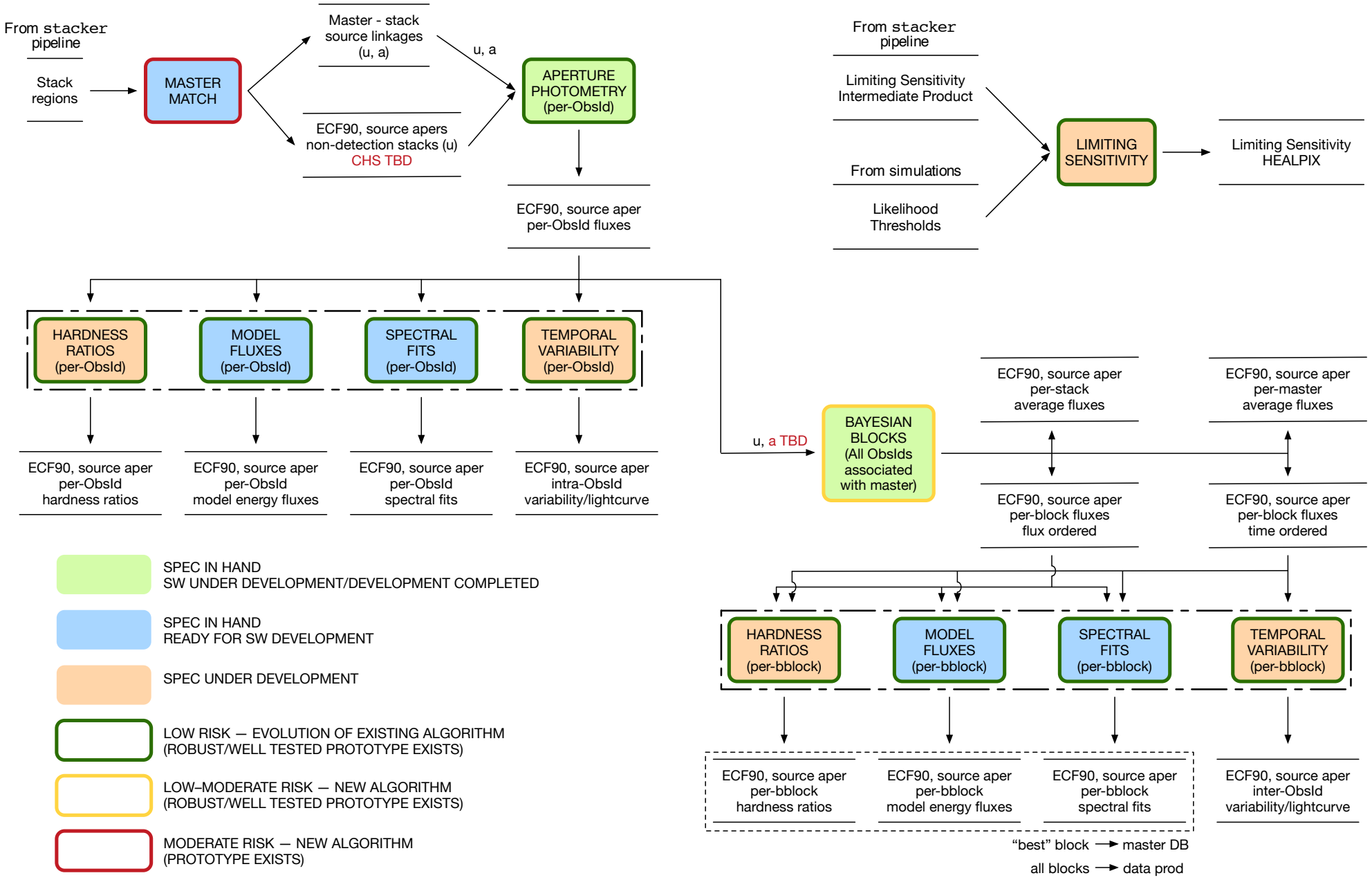
FALSE – MARGINAL and MARGINAL – TRUE detection thresholds are determined by simulations being run in parallel with catalog production



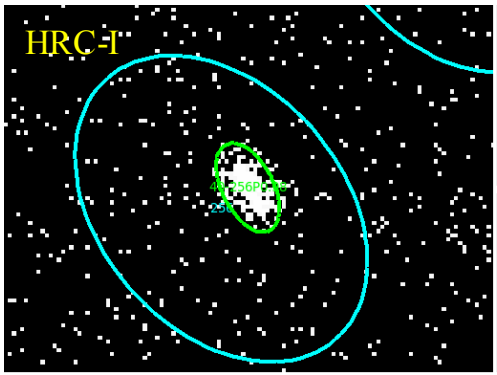


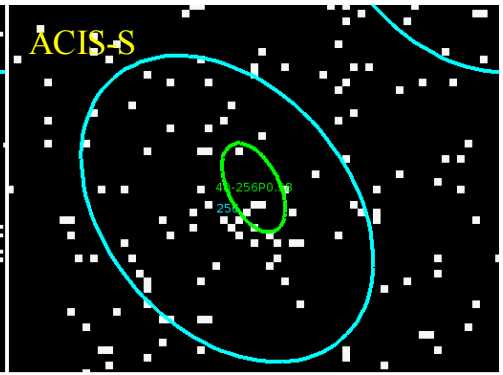
- *Left:* Example false source detections on blank ACIS field (overlayed on PSF map) — based on current simulation set (in progress), there are 241 total sources detected in 32 blank field realizations down to likelihood zero
- *Right:* Cumulative distribution of likelihoods in the soft, medium, and hard energy bands (summed over the field in this figure, although the actual distribution will likely depend on off-axis angle)
- Plan is to construct the MARGINAL – TRUE threshold based on Type I errors (e.g., 1% false source rate)
- Type II errors will also be considered when setting the FALSE – MARGINAL threshold
- Only MARGINAL and TRUE detections will be included in the catalog database
- However *all* detections (including those classified as FALSE) will be available to end users in FITS data products

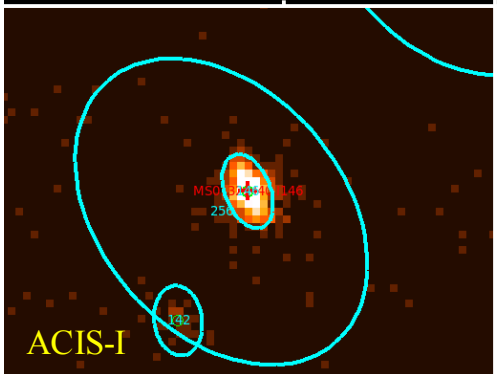
Post-stacker Pipeline Thread



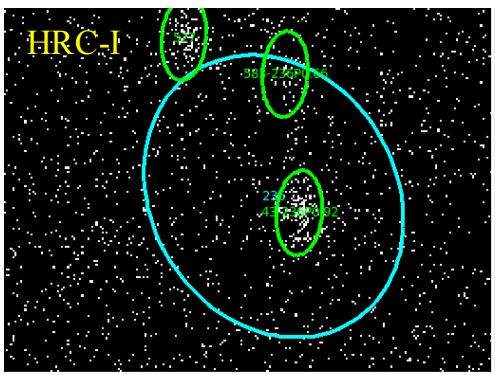
master_match

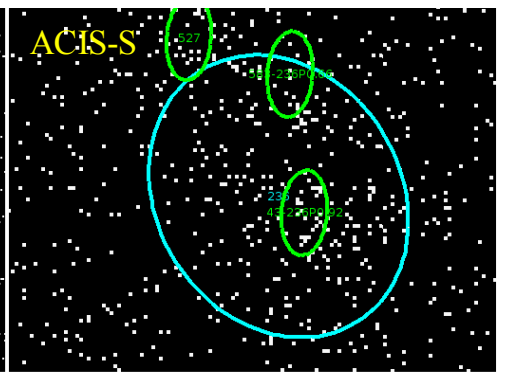


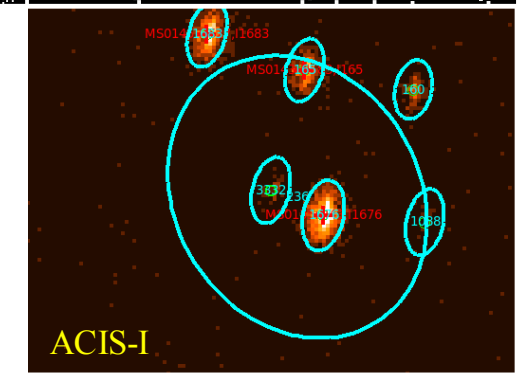




Region 1





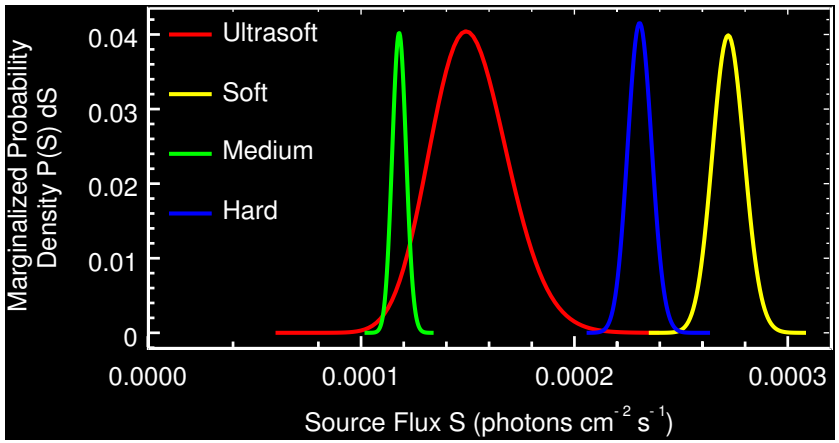


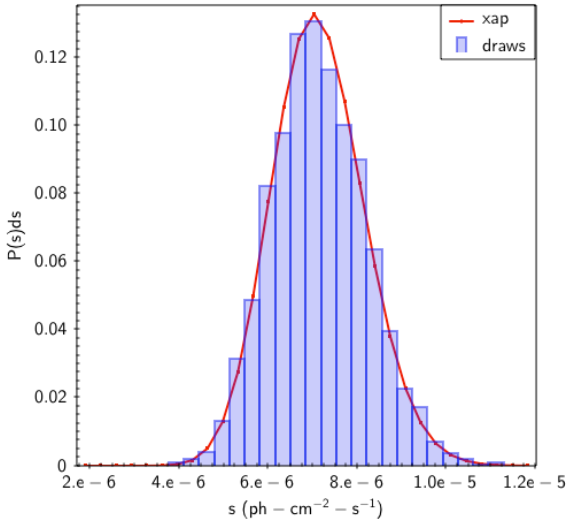
Region 2

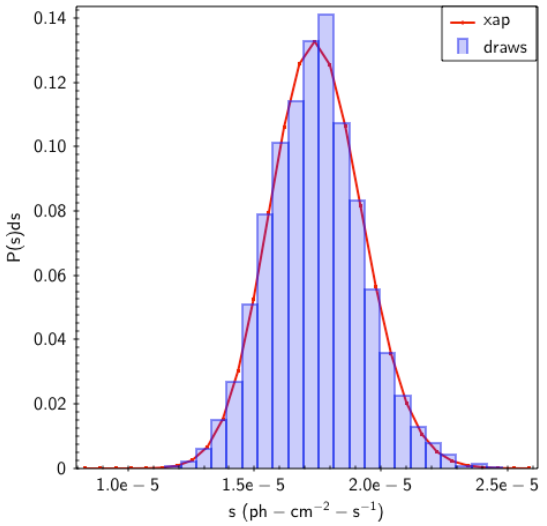
The matching algorithm is adapted from the Bayesian probabilistic cross-match formalism developed by Budavári & Szalay 2008 ApJ, **679**, 301

- In the upper panels, the HRC detections are shown in green, and the ACIS-S detections in cyan
- In the lower panels, both the ACIS-I and ACIS-S detections are shown in cyan
- In region 1, the on-axis HRC-I observation detects a single point source; the off-axis ACIS-S observation also appears to detect a single source (with a larger PSF), but there is a small position offset; the deeper on-axis ACIS-I observation identifies two point sources
  - There are two master sources — one corresponding to the bright (HRC-I/ACIS-I) detection and another corresponding to the fainter ACIS-I detection
  - The off-axis ACIS-S detection is ambiguously matched to the two resolved master sources
- In region 2, there are 6 master sources — 3 correspond to the matched HRC-I and ACIS-I detections, and 3 are detected only in the deeper ACIS-I observation
  - The off-axis ACIS-S detection is ambiguously matched to the overlapping master sources

Aperture Photometry



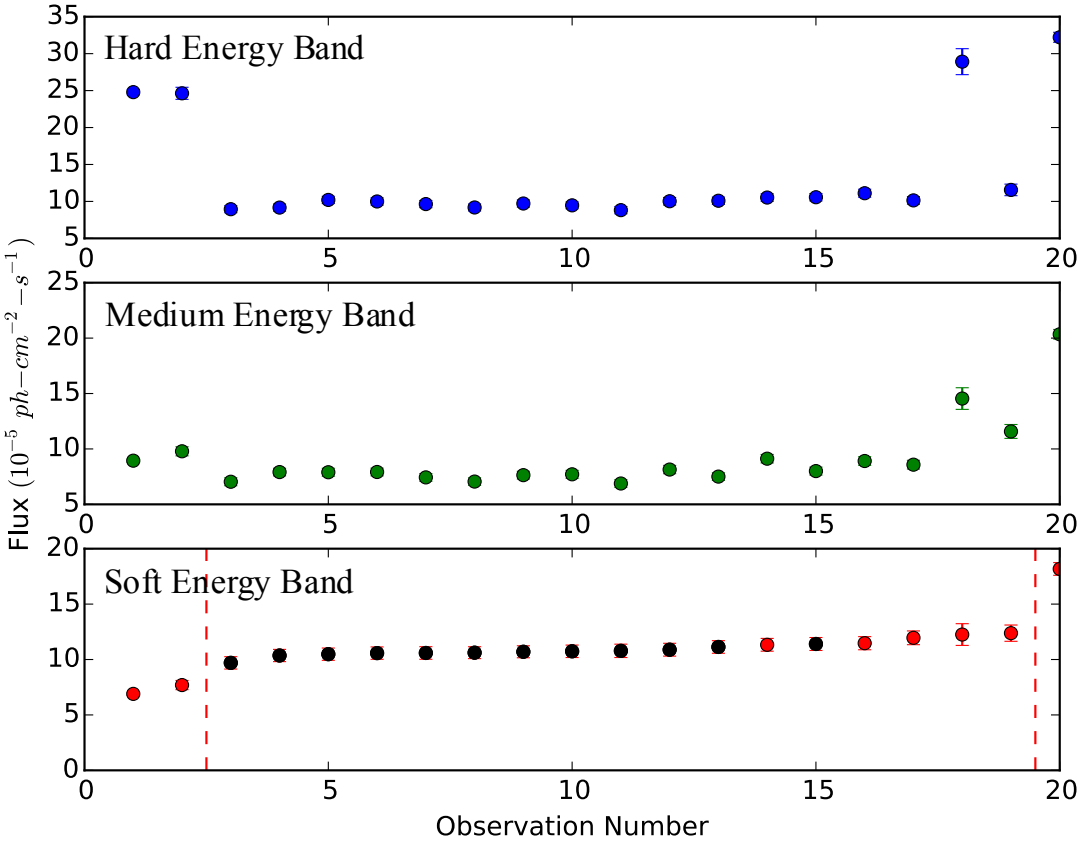




- *Left:* CSC 2.0 aperture photometry approach builds on current published work (Primini & Kashyap 2014 ApJ, **796**, 24) to determine probability density functions for net counts, rates, or fluxes, plus errors, from integral quantities (e.g., counts, average exposure, PSF fractions) in source and background apertures, without recourse to detailed PSF fitting
- Use *Sherpa* and pyBLoCXS MCMC method to characterize marginalized probability densities for source and background intensities (i.e. getting count rate + errors)
  - More tractable that direct numerical integration of joint posterior probability distribution (very compute/memory intensive)
  - More sources can be analyzed at the same time — critical in crowded fields such as Sgr A*
- *Center/Right:* Comparisons of exact compute-intensive method (red curve) and the new *Sherpa* pyBLoCXS MCMC method (purple histograms) for two sources
  - The red curve and the histograms agree very well, confirming the MCMC method can replace the exact method

Bayesian Blocks Example

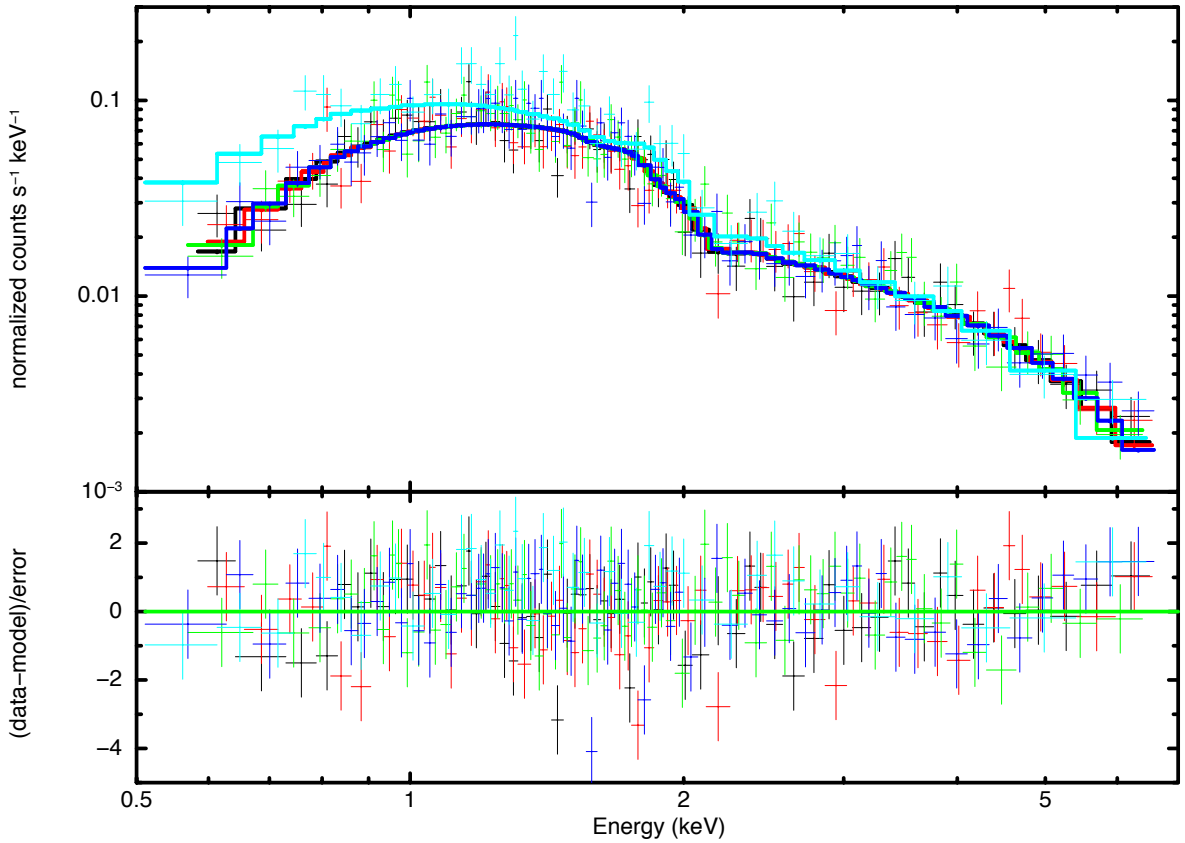
- *Right:* Example flux-ordered aperture photometry for 20 observations of a single source (based on soft energy band flux) — distinct blocks are split by dashed lines
- When combined with the data from the other bands, several multi-band blocks will be identified
- In this example, the "best" block (across all energy bands) consists of the observations shown in **black**
- Catalog aperture photometry values will be reported for this block; properties for other blocks will be accessible in FITS data products



- A Bayesian Blocks analysis identifies both *flux-ordered* and *time-ordered* blocks of observations with consistent aperture photometry across all energy bands simultaneously
- Aperture photometry, spectral, and temporal properties will be computed for each block to improve S/N
  - The assumption is that the source state is unchanged if the multi-band fluxes are the same
  - This will be true in many cases (if not, the source properties are also computed and recorded for each observation separately)
- Bayesian Blocks *include* aperture photometry upper limits from observations in which a source is not detected

Bayesian Block Spectral Fit

- *Right:* In this example, a 5-observation block was fitted by a single absorbed power-law model
- Each observation is shown indicated by a different color
- The best fit model is indicated by the solid lines — differences between the responses for the various observations is responsible for the different apparent shapes (in particular, for the observation shown in cyan the source was located on an ACIS BI CCD, whereas for the other observations the source was located on a FI CCD)



- Spectral fits are performed on each block identified in the flux-ordered Bayesian Blocks analysis, that has at least 150 net counts
  - Spectral fits include absorbed power-law, black body, bremsstrahlung, and APEC models
- In addition, model fluxes are computed using these models with predefined model parameters (other than normalization) for all TRUE sources irrespective of net counts
- Both spectral fits and model fluxes are also computed for each observation (that meets the net counts criteria) separately

Catalog Database Content

- Master Source Properties (X-ray sources on the sky)
  - Source name, position and position errors*, significance*, flags*, extent* (deconvolved), Bayesian block aperture photometry* (photon and energy fluxes, model fluxes), Bayesian block hardness ratios, Bayesian block spectral model fit properties, inter- and intra-observation temporal variability measures*, observation summary
- Stack Detection Properties (preliminary; X-ray detections)
  - Detection identification, position and position errors*, significance*, flags*, extent* (deconvolved), aperture photometry* (apertures, counts, count rates, photon and energy fluxes, model fluxes), stack identification, instrument information, processing information
- Per-Observation Detection Properties (X-ray detections)
  - Detection identification, position and position errors*, significance*, flags*, extent* (deconvolved), aperture photometry* (apertures, counts, count rates, photon and energy fluxes, model fluxes), hardness ratios, spectral model fit properties, intra-observation temporal variability measures*, observation identification, observation pointing, observation astrometry (aspect information), observation instrument configuration, processing information

* Multiple energy bands

- Note: numerical properties include associated lower and upper confidence limits

FITS Catalog Data Products

- Per-Observation Full Field Data Products
  - Event list, exposure corrected image*, background image*, exposure map*, adaptively smoothed exposure map*, aspect solution (incl. fine astrometry updates), aspect histogram, bad pixel map, field of view, pixel mask, extended source region polygons* (multiple contour levels)
- Stack Full Field Data Products
  - Event list, exposure corrected image*, background image*, exposure map*, field of view, limiting sensitivity*, merged source detection list
- Per-Observation Source Region Data Products
  - Region definitions, region event list, region image*, local PSF* (~50K counts), region exposure map*, PHA spectrum, ARF, RMF, light curve*, position error MCMC draws*, aperture photometry PDF*
- Stack Source Region Data Products
  - Region definitions, region event list, region image*, region exposure map*, position error MCMC draws*
- Master Source Data Products
  - Bayesian block aperture photometry PDFs*, Bayesian block spectral fits, Bayesian block model fluxes*, Bayesian block hardness ratios, Bayesian block temporal properties*, master light curve*

* Multiple energy bands

Release 2.0 Interfaces

- CSCview query interface
  - Enhanced to support release 2.0 content
  - Continued support for IVOA standards (including ADQL, SAMP, VOTable)
  - Limiting sensitivities will be reported directly by CSCview (4″ × 4″ HEALPIX)
  - Improved cross-match algorithm performance and input table capacity
  - Continued access to release 1.1 content (as a user selection)
  - Possible tighter integration with *Iris* (*Sherpa*/IVOA-standards based interactive SED and spectral fitting application) via SAMP
- Virtual Observatory (VO) interfaces
  - Support for Simple Cone Search (SCS), Simple Image Access Protocol (SIAP), and Table Access Protocol (TAP) data access standards
  - Same as release 1.1
- Command Line Interface (CLI)
  - Enhanced to support release 2.0 content, otherwise same as release 1.1
- *New* simple web interface
  - Interactive equivalent to the VO SCS, with results displayed in web browser
- CSC Sky
  - Same as release 1.1; will follow catalog release
- CSC/SDSS cross-match catalog
  - Same as release 1.1; will follow catalog release
  - Additional cross-match catalogs now possible more readily: `master_match` algorithm allows us to perform Bayesian catalog cross-matching *in house*

Outreach

Current Outreach

- An article describing release 2.0 was included in the 2015 *Chandra* Newsletter
- A poster describing release 2.0 was presented at the "High-Energy Large- and Medium-class Space Missions" workshop in June 2015 (Chicago)
- Release 2.0 was described in NASA "Hyperwall" presentations and demonstrations at the XXIX IAU General Assembly in August 2015 (Honolulu)

Planned Outreach

- An invited talk about release 2.0 will be presented at the ADASS XXV meeting in October 2015 (Sydney)
- Release 2.0 will be promoted at the AAS meeting in January 2016

Next Steps

Phase	Date	Summary
`stacker` phase II	Winter 2015	*New* — Reprocessing to correct processing/data issues that could not be tested prior to production start due to lack of time; update detection classification (TRUE/MARGINAL/FALSE) based on final likelihood thresholds

Update preliminary detections list based on final detection classifications

`master_match`	Spring 2016	Identify master sources; determine source ←→ detection linkages; determine source names Lien: develop `master_match` pipeline

`master_match` pipeline production run; update preliminary detections list with 2CXO source names

`source`	Summer 2016	Determine source properties; populate catalog databases; populate limiting sensitivity Lien: develop `source` pipeline

`source` pipeline production run; final QA review; catalog release