Progress Report

Ian Evans
On behalf of the Chandra Source Catalog Project Team

Chandra Users' Committee Meeting
October 10, 2012

Current Catalog Release

Summary

- Current catalog version: 1.1; Released: 2010 Aug 10
  - 106,586 master sources
    - Includes 104,628 ACIS-only, 1,034 HRC-only, 924 both ACIS and HRC
  - 158,071 source detections
    - Includes 152,296 ACIS, 5,775 HRC
  - 5,110 observations with at least one detected source

Usage Statistics

Release 1.1	2011 Sep 01 – 2012 Feb 28		2012 Mar 01 – 2012 Aug 31	
	Number	% Non-CfA	Number	% Non-CfA
CSCview catalog browser initializations	266 /month	85%	300 /month	90%
VO cone searches	28K /month	99%	36K /month	98%
CSC Sky in Google Earth	2093 visits/ month	~100%	2116 visits/ month	~100%

Release 2 Summary

Expected Number of Distinct Sources on the Sky	~ 280,000
Expected Number of Source Detections	~ 380,000
Expected Number of Observations	~ 7,000
Expected Limiting Net Source Counts (point source, on-axis)	~ 5
Instrument Data Included	ACIS and HRC-I Imaging
Source Detection Runs On	Stacks of observations with the same instrument, and pointings within 60 (TBR) arcsec
Sources Types Included	Point and compact sources in regions without extended emission Regions of extended emission delineated by convex hulls Convex hull thresholds TBD Point sources within extended emission convex hulls Detection sensitivity may be reduced in these regions

Science Highlights Since Last CUC Meeting

- Working algorithm development and prototypes for release 2.0
  - Supported release 2 pipeline design
  - Developed/developing algorithms and tool prototypes for limiting sensitivity, background determination/extended source detection, and aperture photometry
  - Investigating source detection performance
    - Concluded `wavdetect` does not perform satisfactorily in areas with non-uniform background or extended emission
    - Evaluating "hybrid" approach to detect candidate sources
      - » Use `wavdetect` to identify candidates in regions without extended emission
      - » Use `mkvtbkg` to identify convex hull surrounding extended emission
      - » Use `mkvtbkg` to identify point source candidates superimposed on extended emission
    - Use `mle` to evaluate whether candidate sources are real or not
  - Investigating criteria for merging variable sources
- Assessing test pipeline output and providing feedback

Software Highlights Since Last CUC Meeting

- Continuing development of release 2 pipelines and tools
  - Defined release 2 pipeline design requirements
    - Multi-step architecture runs processing in distinct phases
  - Completed initial implementation through "stacker" pipeline
  - Working later-stage tools as specs become available
- Supporting science testing using actual and simulated data
- Defined requirements for new production hardware
  - Installed cluster "*sylvester*" at Hampshire Street
    - 288 processing cores, 1536 GB memory, 32.4 TB local disk
    - Being used extensively for pipeline testing
  - Release 1 production cluster "*morris*" moved to CDP for development testing

Catalog-Related Releases

CAT 4.0	18 May	Beta 1 (internal); Pipelines (PL) thru Stacker
CAT 4.1	10 Aug	Beta 2 (internal); add HRC; add archive thru Stacker
CAT 4.2	~ Oct	Beta 3 (internal); PL refinements; add QA; add LimSens
CAT 4.3	~ Dec	PL refinements; add Master Match PL + archive; add QA

New Algorithm and Tool Development

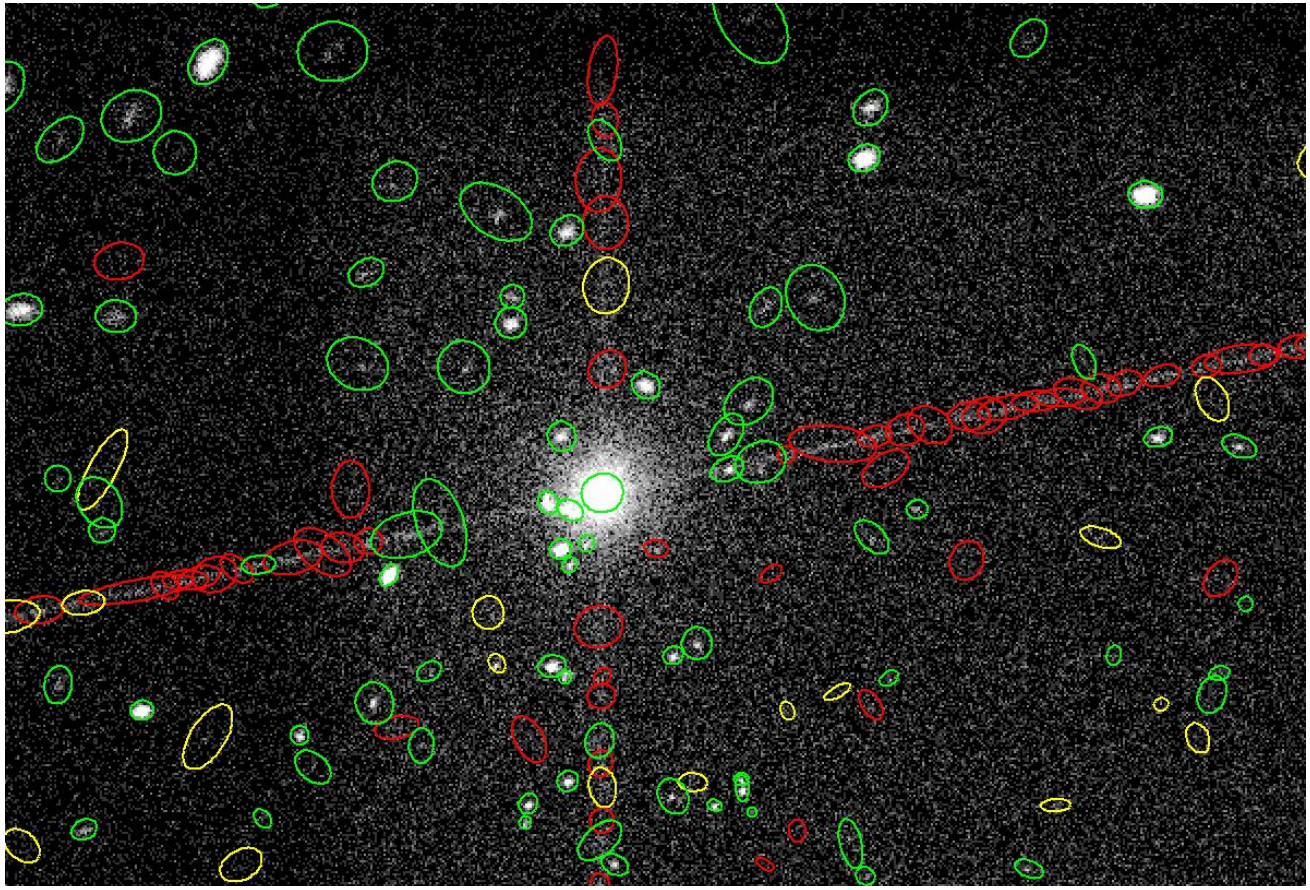
- Several new algorithms are being developed and implemented as new tools by the catalog team
  - `mle` — Maximum likelihood estimator tool fits local PSF convolved with elliptical Gaussian to candidate point and compact sources to compute source likelihood that can be used to evaluate whether candidate sources are real
    - `mle` does a simultaneous fit across multiple observations that include the same candidate source
      - » "Bad" regions (e.g., readout streaks) can be rejected on an ObsId-by-ObsId basis
  - `mkvtbkg` — Voronoi tesselation background tool creates robust observation background by identifying and excluding "bright" sources (point, compact, and extended)
    - Background is being investigated for use with `wavdetect` and `mle`
    - As a side-effect, `mkvtbkg` generates polygons surrounding extended emission that can be used to construct convex hulls to identify extended sources
      - » Point sources within a convex hull can be identified by using a different threshold

Release 2 Algorithms and Tools (cont.)

New Algorithm and Tool Development (cont.)

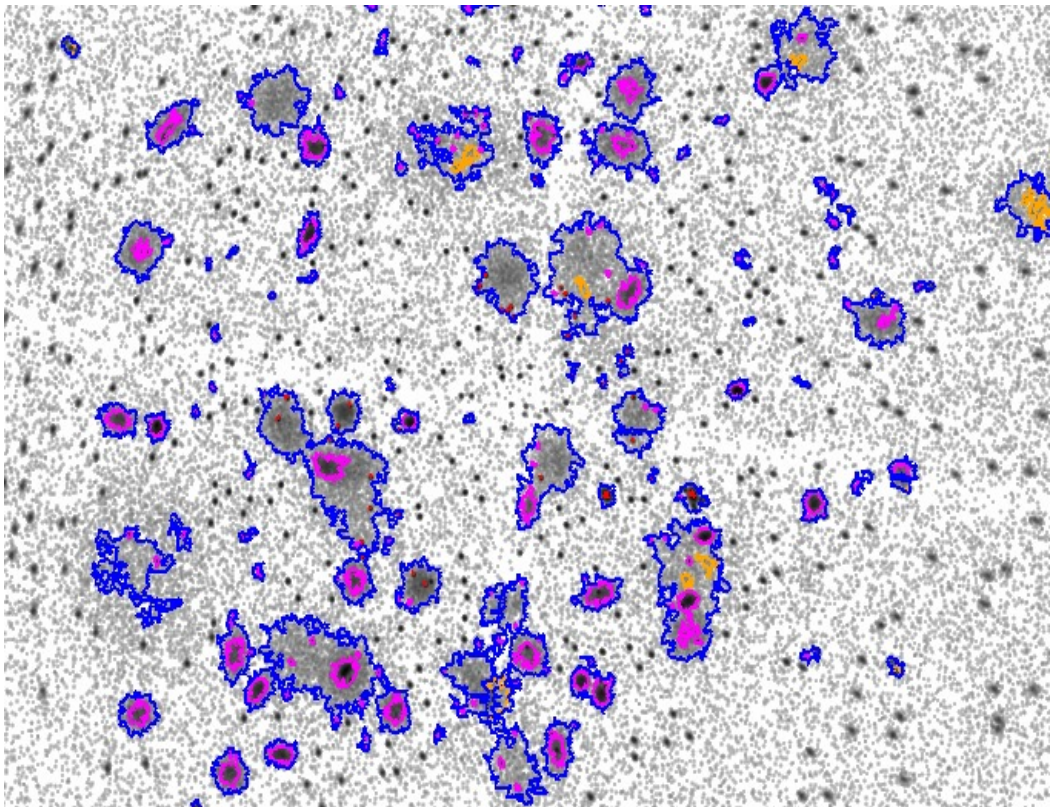
- `limsens` — Computes point source limiting sensitivity for each location on the sky based on defined likelihood threshold(s)
  - Similar to that used for the 2XMM catalog (Carrera et al. 2007, A&A, 469, 27)
  - Will be computed on a grid of 4″ × 4″ HEALPIX pixels
- New aperture photometry algorithm — determines intensities of multiple sources with overlapping source regions simultaneously, and uses informative Bayesian priors to combine data from multiple observations
  - Determines source intensities for convex hull sources
  - Determines upper limits for non-detections (e.g., from other observation stacks covering the same region of sky)
  - Computed intensity probability distributions will be used to determine hardness ratios and inter-observation temporal variability

Point and Compact Source Detection



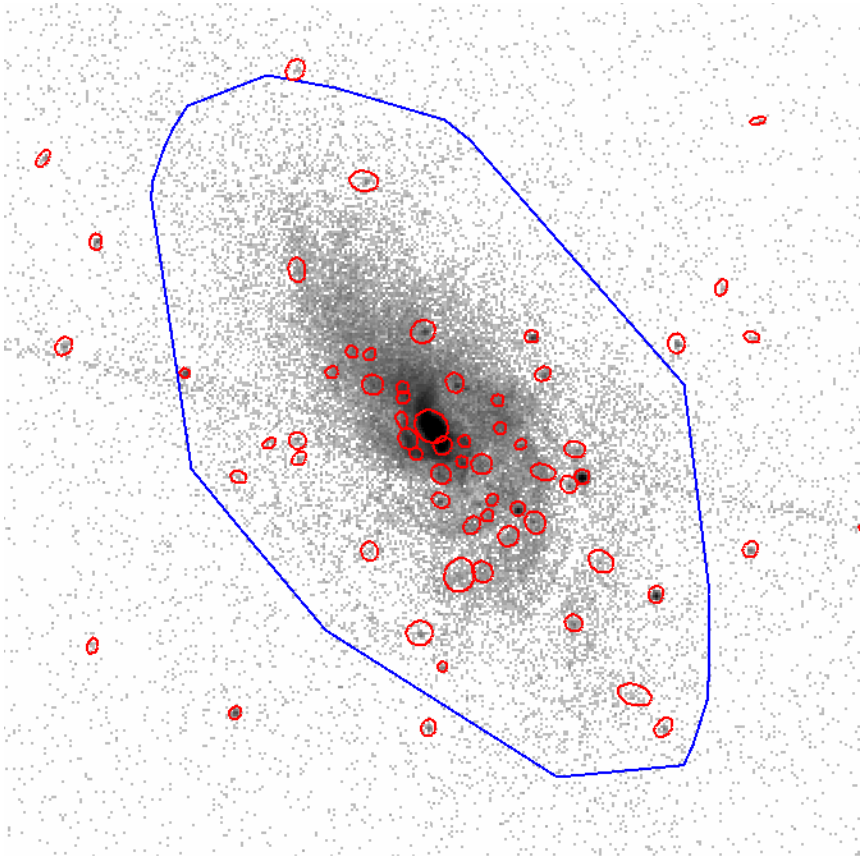
- Preliminary source regions identified by wavdetect on a stack of 3 ObsIds (735, 9122, & 12301)
- Sources with MLE likelihood ≥ 9 are shown in green; possible sources with 5 ≤ likelihood < 9 are shown in yellow; non-sources with likelihood < 5 are shown in red
- Detections that occur only within a masked-out region surrounding each ObsId's readout streak are rejected automatically

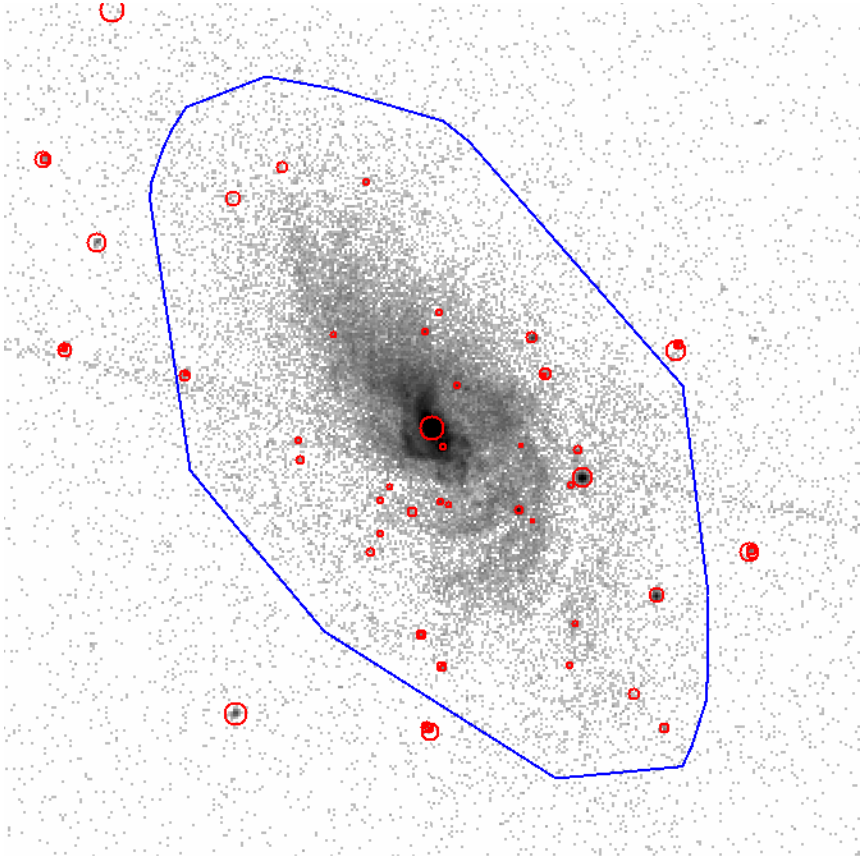
Extended Source Detection



- mkvtbkg uses Voronoi tesselation of the event distribution to creates robust observation backgrounds
  - As a side-effect, mkvtbkg generates polygons surrounding extended emission
    - » Select Voronoi polygons with area smaller than some threshold
    - » Merge adjacent polygons to construct extended source polygons
- In this example, extended source polygons for three thresholds are shown in blue, orange, or magenta
- Resulting polygons are complex and are impacted by inter-chip gaps and readout streaks
  - Unwanted polygon vertices (e.g., near readout streak) are eliminated and a convex hull is constructed around the remaining vertices

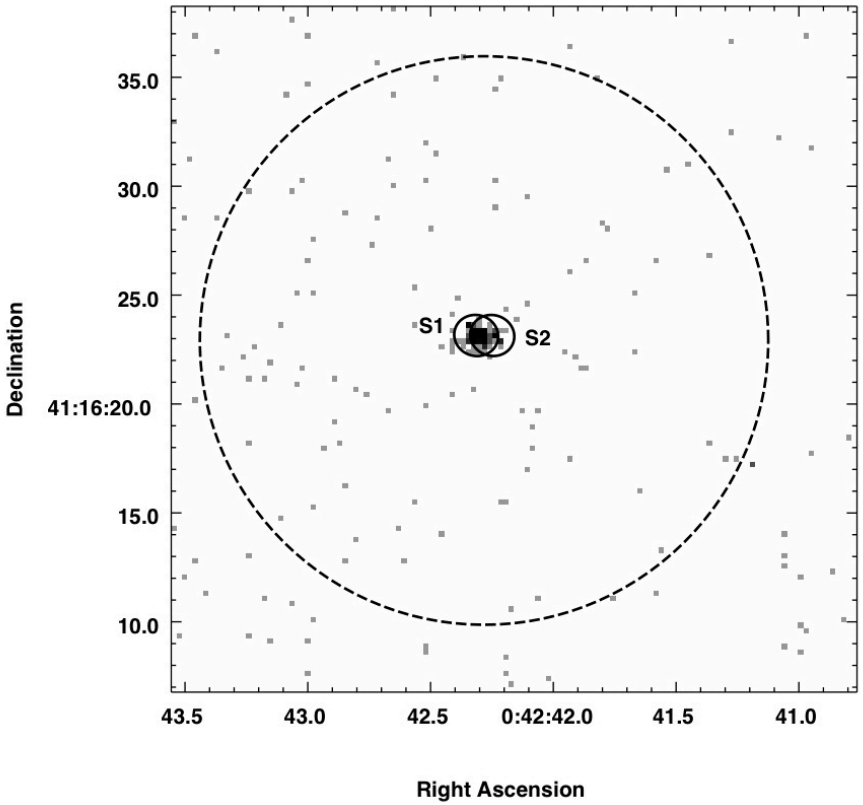
Extended Source Detection (cont.)



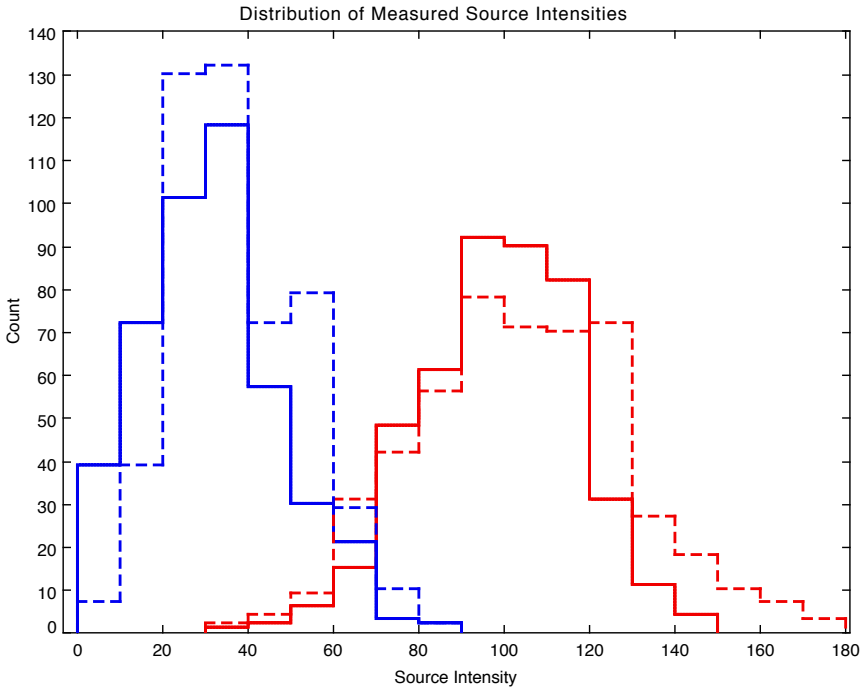


- Convex hull generated by mkvtbkg shown in blue
- (Left) Red sources from wavdetect include detections on the extended emission that are unrelated to physical sources
- (Right) Red sources from mkvtbkg are determined by appropriately filtering Voronoi polygons brighter than some threshold and more accurately identify point sources superimposed on the extended emission
  - The detection sensitivity for these sources may be lower than for point and compact sources in the field

Aperture Photometry

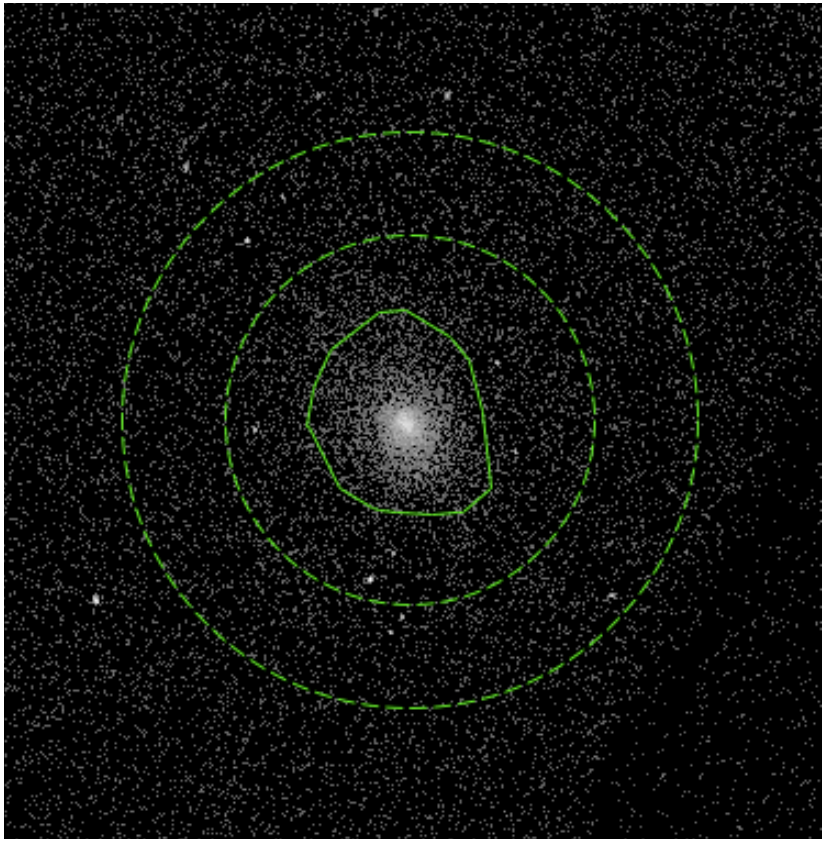


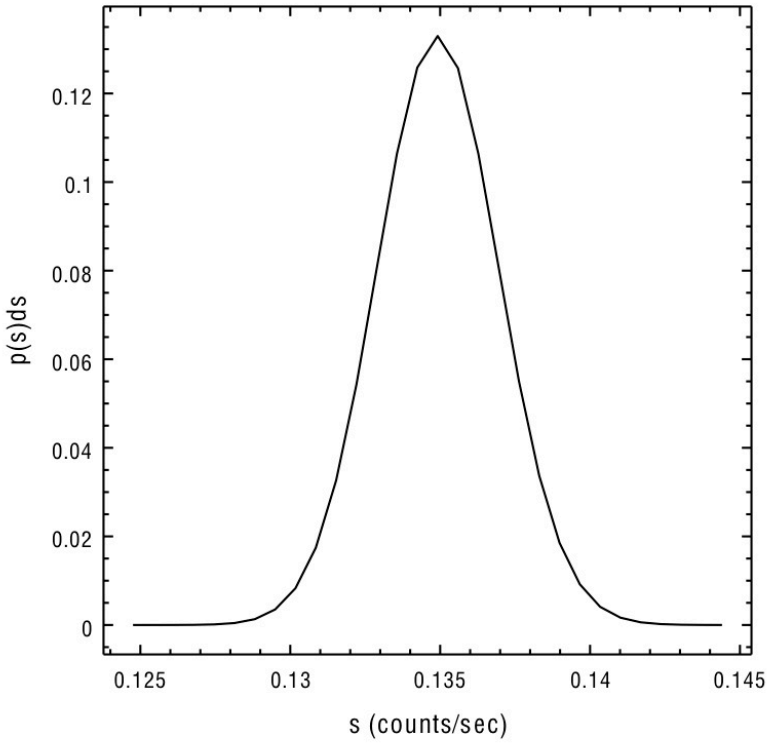
- Simulation of two point sources separated by 0.75″ at 0.5′ off-axis, with true intensities of 100 and 30 counts
- The 90% PSF ECF region is shown for each source



- Distribution of source count intensities derived using the new (solid) and release 1.1 (dashed) aperture photometry algorithms (500 simulations)
- New algorithm reproduces true source counts more accurately

Extended Source Photometry

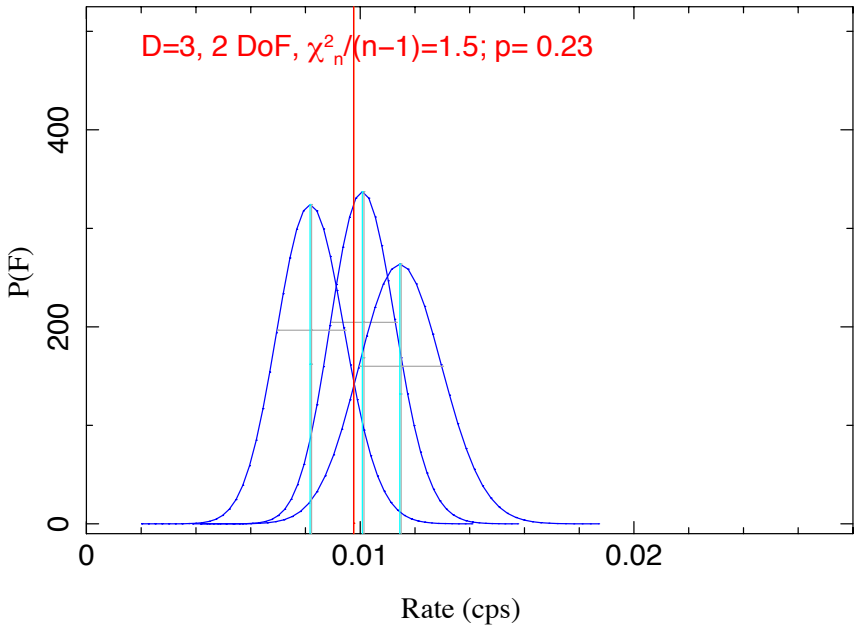


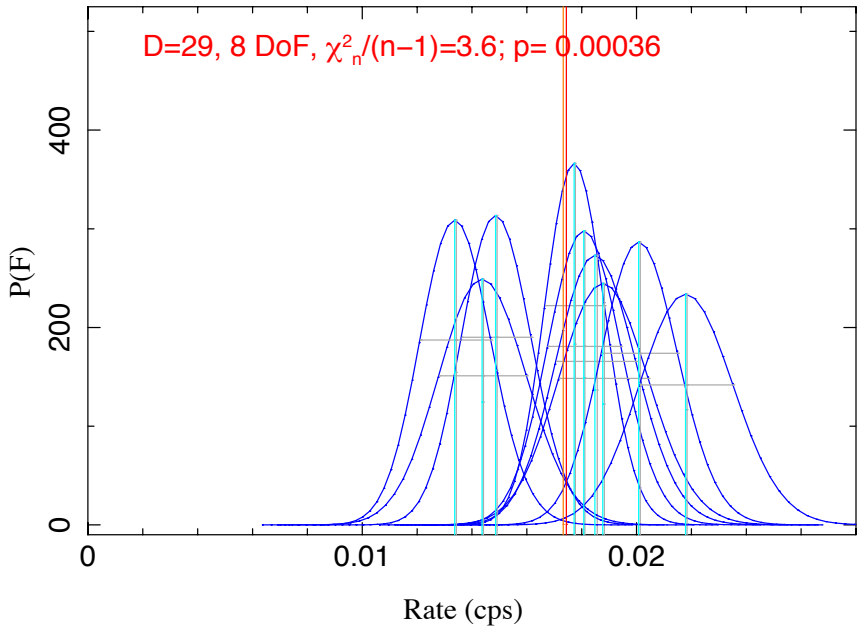


- Extended source region derived from convex hull surrounding source on ObsId 800
- Background region circular annulus shown dashed

- Posterior probability distribution for net count rate in convex hull source
- Ultimately, task is to determine extended source intensity after correcting for any detected embedded point sources

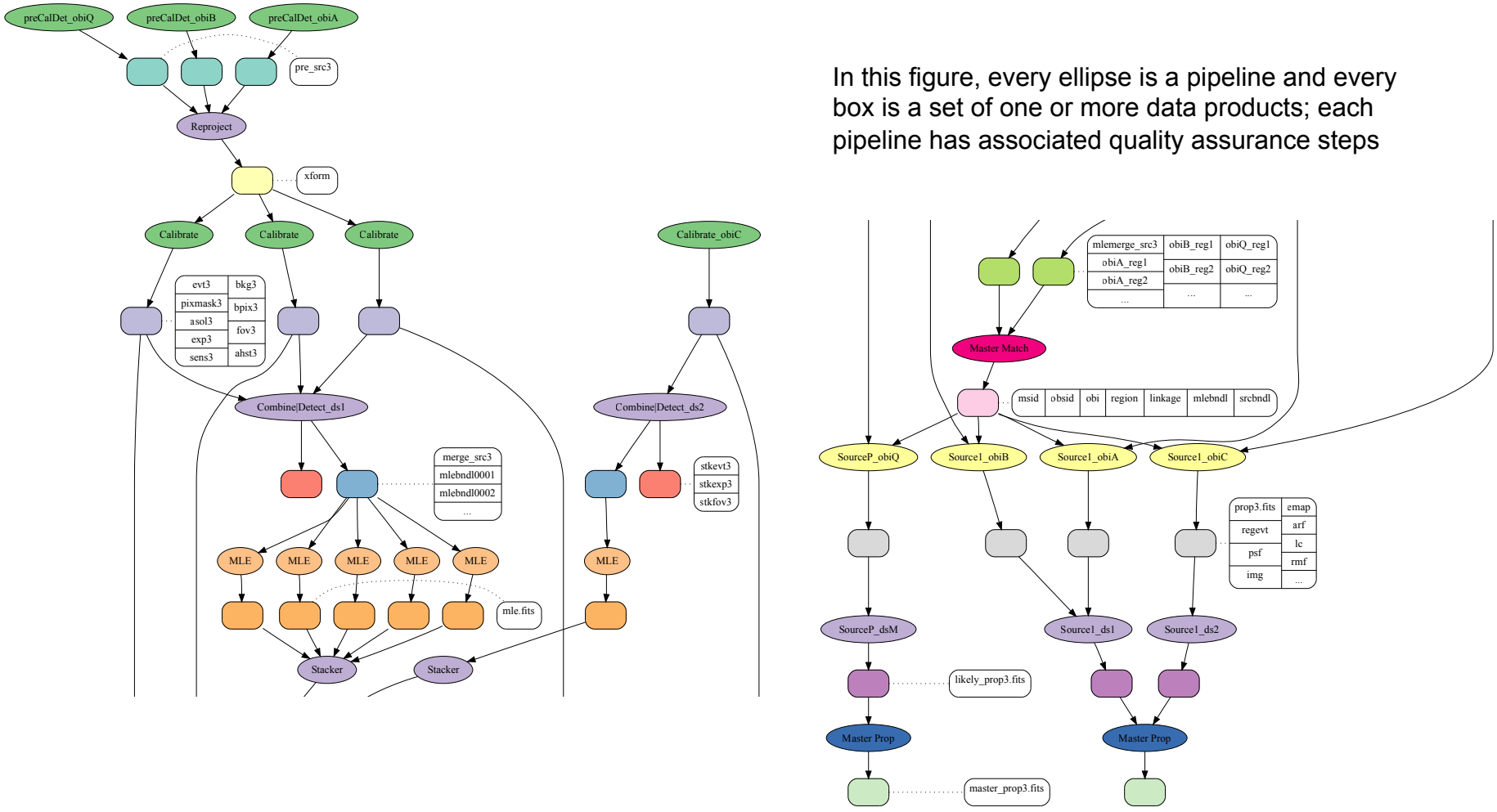
Temporal Variability





- Investigating more intelligent (than rel. 1) approaches for handling temporally variable sources when computing source properties from multiple observations (e.g., when fitting model spectra)
  - — (e.g.) calculate the difference, D, in log likelihood based upon fitting with one rate (red line) vs. N rates (light blue lines), and determine probability of that difference
  - — (Left) Source intensity distributions derived from aperture photometry for 3 observations of a source — there is a 23% chance that a non-variable source would appear to vary that much — reject this as being variable and composite these 3 observations
  - — (Right) Source intensity distributions derived from aperture photometry for 9 observations of a source — the variability is statistically significant
- Investigating ways to reasonably group observations into non-variable composite subsets (e.g., using a Bayesian Blocks analysis)

Release 2 Pipeline



In this figure, every ellipse is a pipeline and every box is a set of one or more data products; each pipeline has associated quality assurance steps

For release 2, the pipeline configuration is significantly more complex than the release 1 configuration because of the requirement to combine stacks of close observations prior to source detection

The current CAT release implements all pipelines (not necessarily completely) through “stacker” (bottom left)

Operations plan

- Targeting catalog release 2 in late 2013
- Take advantage of multi-step architecture to process in distinct phases while subsequent pipeline steps are being developed / tested
  - Each phase needs completed science review and sign-off prior to start
- Phase 1: run for record through "stacker" pipeline after CAT 4.3
  - Creates all per-ObsId-stack source detections
  - Dependent on completion of Repro IV (expected early 2013)
    - Avoids operational complexity by eliminating need to run aspect pipeline
- Phase 2: run "master match" for record after CAT 4.4
  - Merges source detections from multiple overlapping ObsId-stacks
- Phase 3: run source pipelines after CAT 4.5
  - Generates source properties for each detected source
- Phase 4: perform final QA