



CHANDRA
SOURCE CATALOG

Progress Report

Ian Evans

On behalf of the Chandra Source Catalog Project Team

Chandra Users' Committee Meeting

April 27, 2010

Summary

- Production of catalog release 1.1 is approaching completion
 - Includes public HRC-I imaging observations, and several minor updates to release 1.0 data, but otherwise retains the same general characteristics as rel. 1.0
 - Expect to release in next few weeks
- Catalog interface updates
 - Updated version of CSCview with several new and enhanced capabilities will be released with version 1.1 of the catalog
 - Released two updates to CSC Sky in Google Earth catalog visualizer
 - Released CSC 1.0 – SDSS DR7 cross-match database interface
 - Released CSC 1.0 limiting sensitivity map interface
- Statistical characterization of release 1.1 is progressing
- Release 2.0 planned enhancements have been identified, and prototyping/testing of some critical path items has started

Science Highlights Since Last CUC Meeting

- Completed extensive testing and verification of CSC 1.1 production system releases
- Currently performing manual quality assurance reviews of release 1.1 bulk processing
- Continued working on statistical characterization of HRC processing
- Updated public web site with latest user documentation and threads
 - <http://cxc.cfa.harvard.edu/csc/>
- Documents and publications
 - 2 catalog-related posters presented at the 11th HEAD meeting
 - Worked extensive revisions to catalog description paper in response to referee's comments
 - Working catalog statistical characterization and aperture photometry algorithm papers

Software Highlights Since Last CUC Meeting

- Completed several catalog processing system test releases, leading to catalog release 1.1 production system release
- Released several new and updates user interfaces
- Software team managed production processing of catalog release 1.1
- User Interface Releases

CSC Sky	1.0	27 Oct	Added color-coded FoVs for all Chandra observations
Xmatch	1.0	29 Oct	CSC 1.0-SDSS DR7 cross-match interface release
LimSens	1.0	09 Dec	Web-accessable CSC limiting sensitivity service release
CSC Sky	1.1	21 Dec	Added catalog source markers and pop-up data balloons
- Catalog Production System Releases

CAT	3.2.1	04 Dec	HRC science test (vs. Req. Doc. 1.1) feedback updates
CAT	3.2.2	18 Dec	ACIS data migration updates
CAT	3.2.3	28 Jan	New algorithm for HRC position error ellipse
CAT	3.2.4	19 Feb	HRC RMF fix/CSC 1.1 production system release
- Upcoming Releases

CAT	3.2.5	May	CSCview 1.1 release; support for CSC 1.1
CSC	1.1	May	Chandra Source Catalog release 1.1

CSC Release 1.0

- Released version 1.0 on March 4, 2009 and version 1.0.1 on July 21, 2009
- 94,676 master sources
 - 60 tabulated source properties (287 data columns inc. conf. limits/energy bands)
- 135,914 source detections
 - 120 tabulated properties (599 data columns)
- 3,912 observations
- 4,998,528 (FITS and JPEG) data files, 4,377,277 MB

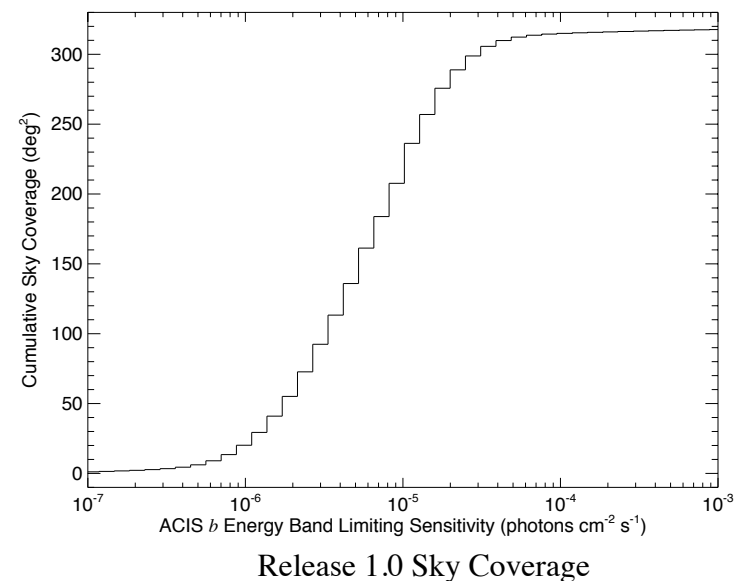
Usage Statistics Since Release

- Interfaces intended primarily for professional use
- ~7K queries via CSCview (~65% non-CfA)
- ~17K queries via VO cone search (~75%)
- ~350K queries* via scripting interface (~95%)
- ~75K data file downloads (~55%)

*Includes ~266K from a single external user

CSC Sky in Google Earth Statistics

- Mix of professional and non-professional users
- ~140K page hits/month
- ~1700 user sites/month
- ~11GB/month data volume
- Significant variation from month to month



New CIAO Tools

- 14 new CIAO user tools were developed from CSC 1.0 processing tools

<i>Tool Name</i>	<i>CIAO Version</i>	<i>Description</i>
aprates	4.1	Calculate source aperture photometry properties
dmellipse	4.1	Calculate best fit ellipse for specified encircled fraction
eff2evt	4.1	Calculate energy flux from event energies
lim_sens	4.1	Create a limiting sensitivity map
mkpsfmap	4.1	Look up PSF size for each pixel in an image
acis_streak_map	4.1.2	Create a high spatial frequency background map
dither_region	4.1.2	Calculate region on detector covered by a sky region
evalpos	4.1.2	Get image values at specified world coordinates
glvary	4.1.2	Search for variability using the Gregory-Loredo algorithm
pileup_map	4.1.2	Create image that gives indication of pileup
modelflux	4.1.2	Calculate spectral model energy flux
srcentent	4.1.2	Compute source extent
create_bkg_map	4.2	Create a background map from event data
dmimgpm	4.2	Create a low spatial frequency background map

Release 1.1 Contents and Enhancements

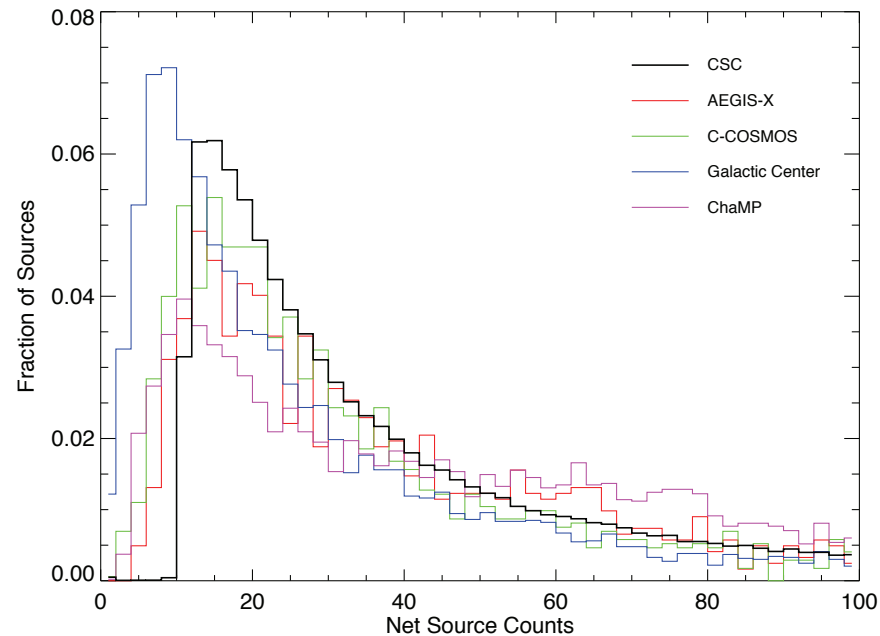
- Adds 621 public HRC-I imaging ObsIds since the start of the mission
 - Like release 1.0, observations of highly extended sources are not included
- Adds 641 ACIS imaging ObsIds that became public since release 1.0
 - The same photometric system is used for release 1.1 as release 1.0 — since this is an incremental release (*i.e.*, rel. 1.0 data are not in general reprocessed), recent updates to effective areas and contamination model are *not* included
- Includes updates to source properties to address release 1.0 deficiencies
 - Tabulated source position uncertainties updated to include 0.16 arcsec absolute position uncertainty deduced from comparison of CSC 1.0 and SDSS DR7
 - Chip coordinates for all source detections have been recomputed to include mean aspect solution $dy/dz/d\theta$ offsets (typically updates chip coordinates by a few pixels; sky coordinates are not affected)
 - Variability code corrected for 14,133 source detections
 - Dither warning flag corrected for 11,921 source detections
- CSC 1.0 ObsIds reprocessed
 - 161 ACIS ObsIds that had $> 3\%$ photometric error due to an exposure map bug
 - 73 ACIS ObsIds that did not have final time-dependent gain calibrations available prior to CSC 1.0 release
 - 8 ACIS ObsIds that were reprocessed for miscellaneous reasons

Processing Status

- Release 1.1 production is almost complete
 - Started processing on April 5
 - All observations have completed bulk processing, except for some requiring manual quality assurance review (~120 as of 4/20)
 - Expect QA will trigger reprocessing of order a few dozen observations
- Planning to release CSC 1.1 in May
- Release 1.1 adds
 - ~31K new source detections (~80% ACIS/20% HRC)
 - ~16K new master sources
 - ~7K prior master sources are updated with data from new observations
 - These number may change slightly following quality assurance

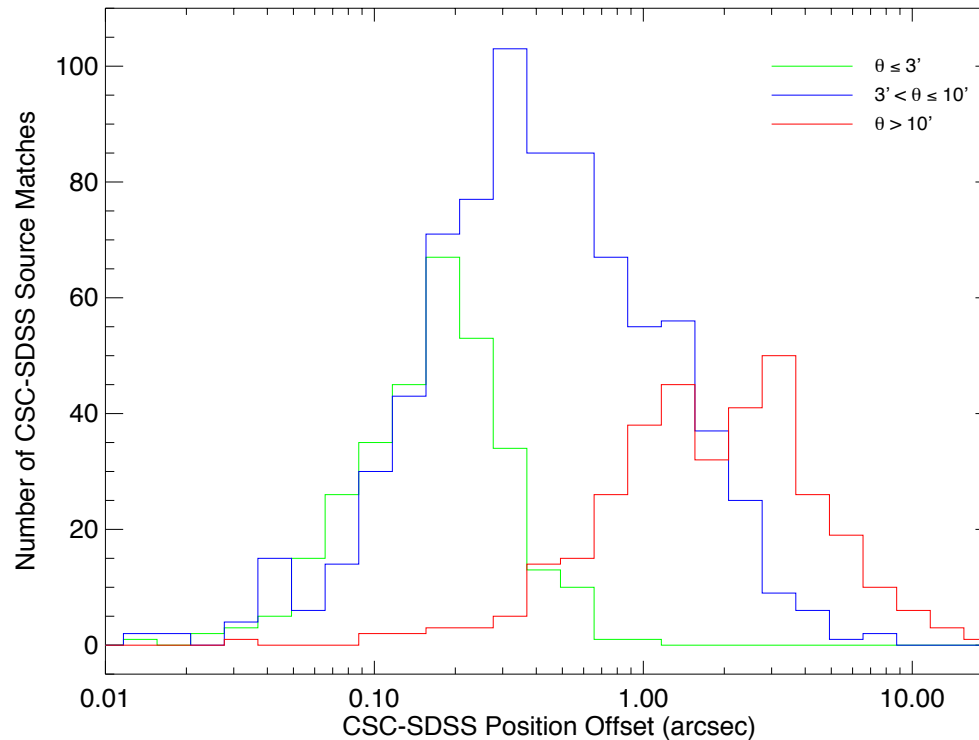
Catalog Statistical Characterization

- Statistical properties of release 1 are available on the catalog web site
 - <http://cxc.cfa.harvard.edu/csc/char.html>
- A paper describing the statistical properties of release 1.0 is in preparation (Primini et al.)
- A number of additional characterization studies of release 1.0 have been performed in response to some questions raised by the catalog description paper referee
 - *e.g.*, magnitude of `wavdetect` centroiding errors, fraction of sources with underestimated energy flux errors when using the ARF-correction method (<1%), fraction of sources affected by spectral fit bias due to lack of energy-dependent aperture corrections (~2.5%), ...
 - Many of these studies will be detailed further in the characterization paper
- Characterization simulations for release 1.1 are continuing
 - Some studies have led to revisions of production pipeline algorithms
 - HRC point source simulations are in progress



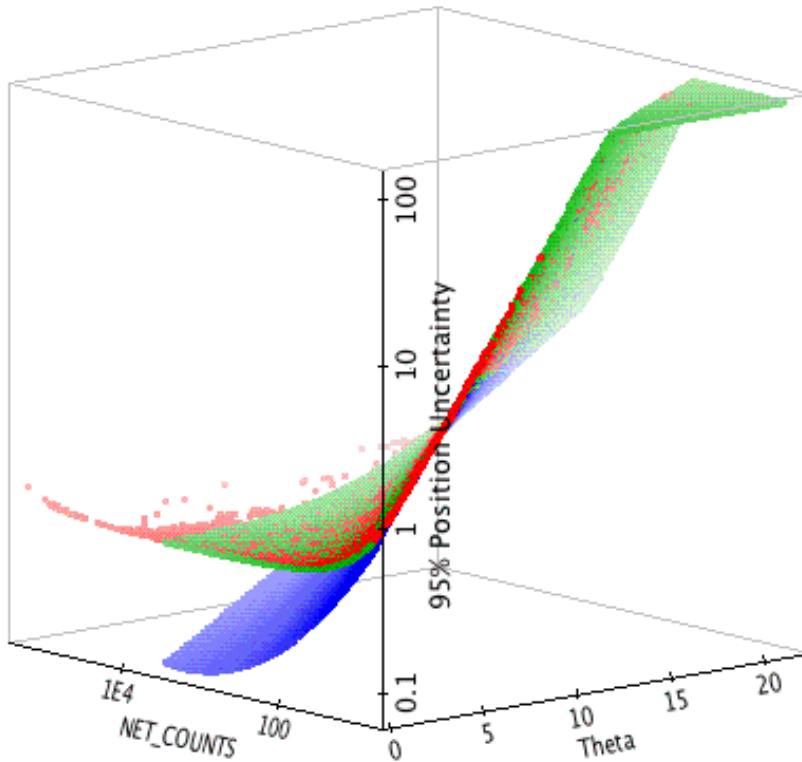
CSC Net Source Count Threshold

- CSC lower threshold is determined by a 3σ flux significance threshold
- Distribution of net counts for CSC sources detected in the ACIS broad (0.5–7.0keV) band vs. four other Chandra catalogs constructed from ACIS observations with comparable exposures shows that while these other catalogs do include sources with fewer net counts than the CSC, the additions are in general not large
 - ~5%, ~9%, ~9%, and ~24%, for AEGIS-X, the Galactic Center catalog, C-COSMOS, and ChaMP, respectively
 - We attribute the larger percentage in ChaMP to the restricted fields-of-view and the careful manual screening of each source detection used when constructing that catalog



CSC Absolute Astrometric Uncertainty

- Distribution of absolute positional offsets for 1,488 CSC sources that are identified with QSOs included in the SDSS QSO Catalog
 - The mean absolute offset is ~ 0.22 arcsec for CSC sources with off-axis angles within 3 arcmin of the aimpoint, ~ 0.62 arcsec for sources within 10 arcmin, and ~ 1.2 arcsec for all matching sources



- CAT3.2.4 HRC Source Position Uncertainties
- Pos_Err Model
- Pos_Err Model with Absolute Astrometric Error Added

- The figure verifies that production system HRC position errors are computed correctly using the new fit

HRC Source Position Errors

- As a result of release 1.1 production system testing and HRC data characterization, a new fit to the HRC position errors was derived that improves on the ChaMP model that is appropriate for ACIS
- The new fit was added to the release 1.1 production pipeline for HRC data

CSCview 1.1

- New version will be released with CSC 1.1
- Includes numerous enhancements
 - Updated standard queries include HRC observations
 - Metadata display window on Query and Results pages provides brief descriptions of source properties (including units) and file-based data products
 - Improved user interface simplifies changes to queries, and simplifies data product selection and retrieval
 - Query results display is now a single page, and the results can be resorted by clicking on column name headers
 - Source preview allows user to display “quick-look” JPEG images of selected source region and associated local PSF images, as well as full-field images
 - Cone searches now display the distance of the source from the user-specified search position
 - Save to file supports output in tab-separated-value (TSV) and VOTable formats
 - TSV is an RDB format with enhanced header metadata commonly used by Vizier and other astronomical applications
 - IVOA Simple Application Messaging Protocol (SAMP) interface allows CSCview to communicate with other SAMP-aware applications such as ds9, TOPCAT, Aladin, ...

CSCview GUI Enhancements

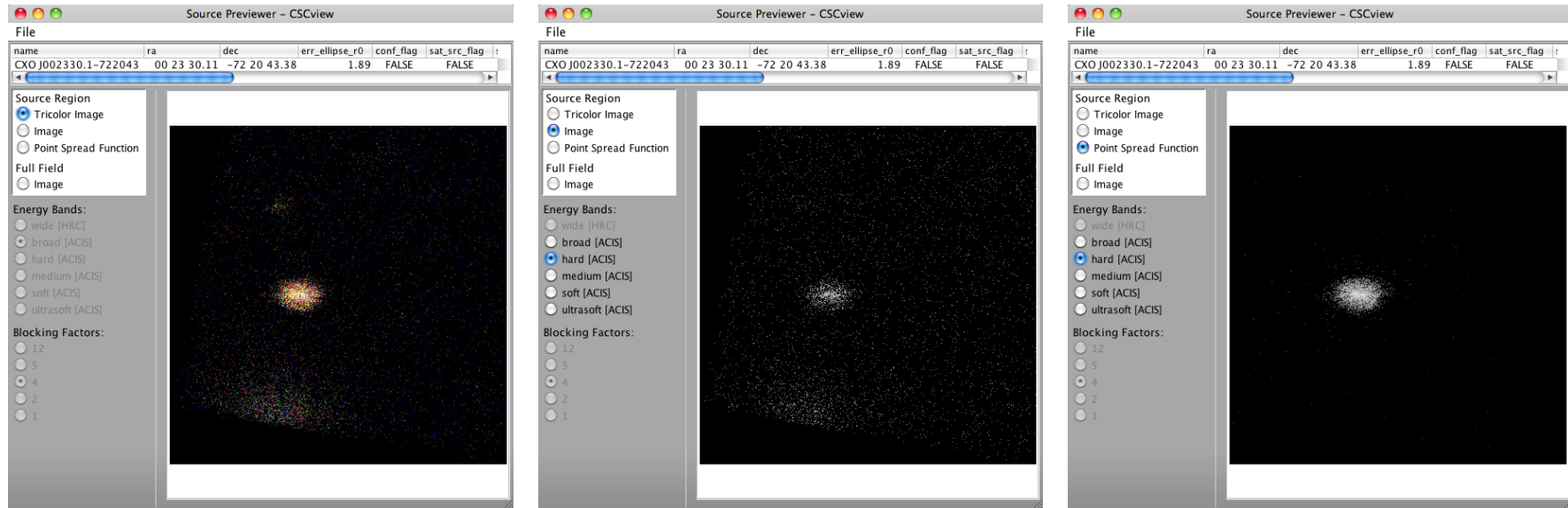
New image viewer provides quick-look source, PSF, and full field images

Revised interface simplifies selection of specific data products and energy bands for user's source selection

The screenshot displays the Chandra Source Catalog search interface. The main window shows a table of search results with columns for name, ra, dec, err_ellipse_r0, significance, flux_aper_b, flux_aper_w, conf_flag, sat_src_flag, and flux_aper_lolim_t. A sidebar on the left allows for filtering data products, including Source Region (Event List, Image, Spectrum, ARF, RMF, Exposure Map, PSF, Lightcurve, Region), Full Field (Event List, Image, Background Image, Exposure Map, Sensitivity Map, Aspect Histogram, Bad Pixel File, Field of View), and Energy Bands (wide [HRC], broad [ACIS], hard [ACIS], medium [ACIS], soft [ACIS], and ultrasoft [ACIS]). A metadata display window is open at the bottom, showing a table of properties for a selected source.

Table	Name	Datatype	Units	Description
Master Sources	name	varchar		Source name in the format 'CXO Jhhmmss.s +/- ddmms'
Master Sources	ra	double	deg	Source position, ICRS right ascension
Master Sources	dec	double	deg	Source position, ICRS declination
Master Sources	err_ellipse_r0	double	arcsec	Major radius of the 95% confidence level error ellipse
Master Sources	conf_flag	boolean		Source regions overlap (source is confused)
Master Sources	sat_src_flag	boolean		Highest source flux significance across all observations
Master Sources	significance	double		Aperture-corrected net energy flux inferred from the source region aperture, calculated by counting X-ray events; ACIS broad energy band
Master Sources	flux_aper_b	double	erg/s*cm^2	Aperture-corrected net energy flux inferred from the source region aperture, calculated by counting X-ray events (68% lower confidence limit); ACIS broad energy
Master Sources	flux_aper_lolim_b	double	erg/s*cm^2	

New metadata display provides brief descriptions of tabulated properties (including units), standard queries, and file-based data products



Source Preview

- Pop-up source preview window allows user to quickly visualize source regions as tricolor images (*left*), band-specific images (*center*), and associated PSF images (*right*), as well as full-field images
- Could be expanded for other kinds of previews (*e.g.*, plot source spectra, lightcurves, ...)

Release 2.0 Goals

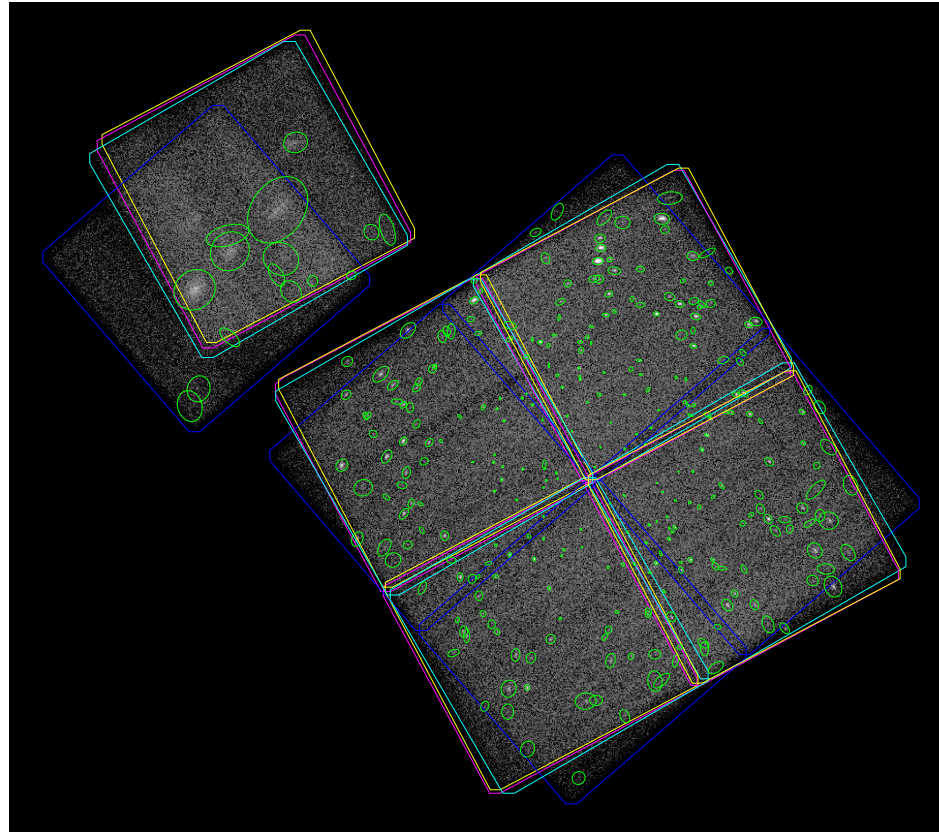
- Co-add multiple observations of the same field prior to source detection
 - Restricted to observations with pointings within ~ 30 arcsec, so PSF isn't "too different" for a single source between observations
 - Restricted to observations with "similar" instrument configurations
- Detect sources to a fainter threshold (~ 5 net counts instead of ~ 10)
- Improve handling of moderately extended sources ($\lesssim 60$ arcsec), especially in the cores of galaxies
- Include numerous algorithm enhancements identified over past ~ 2 years
 - Most of these are incremental improvements to existing algorithms, or require application of existing tools within the pipelines
 - *Examples include*
 - » Simultaneous solution for aperture photometry properties for sources with overlapping source regions
 - » Use of spectrally-weighted exposure maps and PSFs, and energy-dependent aperture corrections in spectral fits (vs. release 1 approach of using effective monochromatic energy approximation)
 - » Using non-detections in observations to set upper limits for sources detected in other observations
- Do all this without significantly affecting source detection efficiency, completeness, and false source rates!

Development Process

- Developed a draft list of proposed enhancements and “areas of concern”
 - Currently working on top-level and infrastructure requirements flow-down
- Identified a set of critical tests to be completed in the short term
 - Started developing prototypes for critical software
- Next steps:
 - Complete detailed processing thread design
 - Assess impact of new requirements on archive, pipelines, and processing infrastructure
 - Begin developing detailed algorithm requirements
- Like release 1.0, we plan to post relevant documents on the public website

Detect/Verify Approach

- Source detection is performed using `wavdetect`, with a lower detection significance threshold than release 1.0, on co-added stacks of observations
 - Set significance threshold to robustly detect sources down to faint limit
 - Except for ACIS readout streak sources, most false source detections in release 1.0 are associated with chip edges or BI–FI chip boundaries
- Compute the probability of a source being real using new MLE code being developed in Sherpa
 - Probability is then used to decide whether detection is real, false, or uncertain
 - Further processing based on probability: for example, could choose to treat to record limited information about questionable sources, while still providing complete set of source properties and data products for real sources
 - Resulting Sherpa package could be adapted for release in CIAO, and would provide a simple way for users to assess source probabilities, especially for split observations

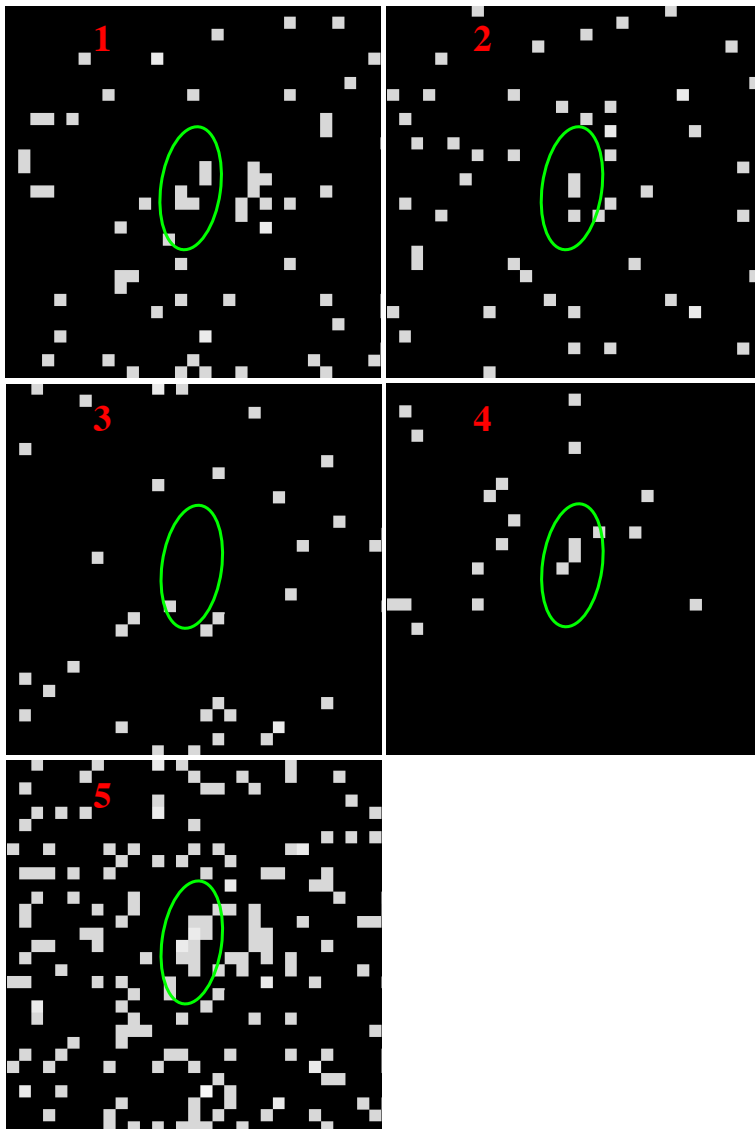


Example

- Four co-added CDFS observations [ObsIds 2312 (132 ks), 2313 (132 ks), 2405 (60 ks), 2406 (30 ks)]
- Filtered source detections from co-added observation are shown in green
 - In this test run, there is a hard limit of 6 net counts; there are eight 6–7 count sources, ten 7–8 count sources, eleven 8–9 counts, and several hundred brighter sources

Sherpa MLE Implementation

- Compute the probability of a source being real using new MLE code being developed in Sherpa
 - The algorithm performs a simultaneous best-fit of the local PSF model at the location of the source on the detector for each of the observations
 - Observations where the source location is compromised (source position is on a chip edge, BI–FI boundary, ...) are excluded from the MLE computation
 - Final algorithm will convolve the PSF model with a simple extended source model, and extract best-fit extended source model parameters
 - Must run the MLE on various combinations of the observations in the stack to maximize source probability for variable sources
- Similar approach has been used by 2XMM and C-COSMOS catalogs
 - XMMSAS `emldetect` code is used as a basis for the Sherpa implementation
- Software implementation is written in Python
 - Significantly simplifies development process by leveraging the investment made in developing the Sherpa infrastructure in Python



Work in progress

- Images 1–4 are the same source region extracted from distinct CDFS observations (~354 ks total)
- Image 5 is a co-added stack of the 4 observations
- Co-added source is ~8.5 net counts (but we have reliably detected and identified ~6.5 net count sources)
- MLE source probability for observations 1–4 is 1.0
- MLE also identifies source as real including only observations 2–4

```

Method = moncar
Statistic = cstat
Initial fit statistic = 192.927
Final fit statistic = 57.1544 at function evaluation 9140
Data points = 160
Degrees of freedom = 155
Probability [Q-value] = 1
Reduced statistic = 0.368738
Change in statistic = 135.772
  src.fwhm 2.55884e-05
  src.xpos 53.1507
  src.ypos -27.8437
  src.ampl 2.71444
  bkg.c0 1.11947e-09
cstat_no_source = 1470.68031045
cstat_point_source = 57.1544400059
Delta-cstat-stat = 1413.52587044
Probability of true source detection: 1.0
65.857641 secs
  
```

Short Term Plans

- Catalog Releases
 - Release 1.1
- Public Interfaces
 - CSCview GUI
 - Release 1.1 updates previously discussed
 - Support for cross-matching with user supplied catalogs
 - CSC Sky in Google Earth
 - Next release will include display of HRC observations and be updated for release 1.1
 - Catalog limiting sensitivity service
 - Next release will include HRC observations and be updated for release 1.1
 - CSC-SDSS cross-match
 - Update for release 1.1
 - Enhanced command-line interface
 - Simplified access to catalog file-based data products via URL
 - External Interfaces
 - Access to a subset of catalog master source properties through Vizier
 - Access to a subset of catalog master source properties through NED
 - » Expect to complete both after release 1.1

Longer Term Plans

- Catalog releases
 - Catalog release 2
 - Co-add multiple observations of the same field prior to source detection to achieve fainter limiting sensitivities
 - Improve background modeling
 - Improve aperture photometry in crowded fields
 - Improve extended source handling
 - Better spectral weighting approach and handling of aperture corrections
 - Numerous other algorithmic improvements
 - Future releases
 - Simultaneous source detection across overlapping observations with different detectors and pointings (and thus very different local PSFs)
 - Detection and classification of very extended sources