The Chandra Source Catalog 2.0: Data Processing Pipelines

Joseph B. Miller[1], Christopher Allen[1], Jamie A. Budynkiewicz[1], Judy C. Chen[1], Danny G. Gibbs II[1], Charles Paxson[1], Craig S. Anderson[1], Douglas Burke[1], Francesca Civano[1], Raffaele D'Abrusco[1], Stephen M. Doe[2], Ian N. Evans[1], Janet D. Evans[1], Giuseppina Fabbiano[1], Kenny J. Glotfelty[1], Dale E. Graessle[1], John D. Grier[1], Roger M. Hain[1], Diane M. Hall[3], Peter N. Harbo[1], John C. Houck[1], Jennifer Lauer[1], Omar Laurino[1], Nicholas Lee[1], J. Rafael Martinez-Galarza[1], Michael L. McCollough[1], Jonathan C. McDowell[1], Warren McLaughlin[1], Douglas L. Morgan[1], Amy E. Mossman[1], Dan T. Nguyen[1], Joy S. Nichols[1], Michael A. Nowak[4], David A. Plummer[1], Francis A. Primini[1], Arnold H. Rots[1], Aneta Siemiginowska[1], Beth A. Sundheim[1], Michael S. Tibbetts[1], David W. Van Stone[1], Panagoula Zografou[1]

[1]Smithsonian Astrophysical Observatory [2] formerly Smithsonian Astrophysical Observatory [3]Northrop Grumman Mission Systems
[4]MIT Kavli Institute for Astrophysics and Space Research

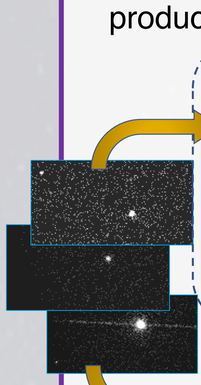
Processing Overview

With the construction of the second Chandra Source Catalog (CSC2.0), came new requirements using new techniques to define a software system that analyzes deeper into the X-ray sky as seen by Chandra. A software system capable of processing ~10,000 observations and nearly 320,000 point and compact X-ray sources was developed, built, and run in a partnership between scientists and software developers.

- 20 data processing pipelines organized into 3 distinct phases: detection, master matching and source properties
  - Detection phase - adjusts for shifts in fine astrometry, combines observations, detects sources, and assesses the likelihood that sources are real
  - Master Match phase - detections across stacks of observations are analyzed for coverage of the same source to produce a master source
  - Source Property phase - each source is characterized by over 350 properties including photometry, spectroscopy, and variability, at the observation, stack, and master levels across 6 energy bands
- 2 additional pipelines - one to compute limiting sensitivity and another to produce source properties on large extended sources characterized by convex hulls
- Quality Assurance (QA) steps ensure accuracy of data products by handling corner cases through manual intervention where the software tools may get confused
- Processing pipelines are designed to run in parallel on a high performance computing cluster (HPCC) – see poster #238.04
  - Processing is done on the smallest unit for a given task - pipelines may be run on an observation, a stack of observations, an ensemble (or group of stacks), or even a single source.
  - Distribution is managed by the Catalog Automatic Processing system, which optimizes the processing time

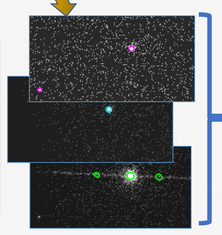
Detecting Sources

Prior to processing, observations (OBI) within one arcmin of each other, are grouped into stacks. While the collection of pipelines operates on a stack, each pipeline may operate on a component of the stack (e.g an observation in the stack or even a group of detections). The results of this phase include source detections, stacked products and per-obi products.



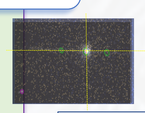
Multi-OBI Stacks ONLY

PreCal and Detect
- Calibrate each OBI using the same CALDB and software
- Run source detection to find bright sources in the OBI



Fine Astrometry
- Match bright source between OBIs
- Compute offsets between each OBI and the OBI with the most sources
- Using a weighted average, find a common tangent point for all OBIs

QA: Cases that couldn't find a reliable match



xforms

Maximum Likelihood Estimator (MLE)

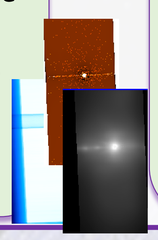
1st Pass
- Use MLE for initial fit, correcting for detect outliers
- Re-center source at fitted position

2nd Pass
- Refine position using MLE
- Compute likelihood
- Calculate errors using Markov Chain Monte Carlo

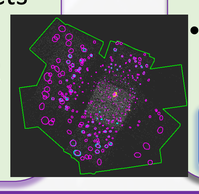
Rebundle and Rerun

QA: Adjust the fit for sources with poor fits

Calibrate
- Calibrate each OBI using the same CALDB and software
- Transform multi-OBI stacks to the same tangent plane
- Make per OBI products - images, backgrounds, etc.



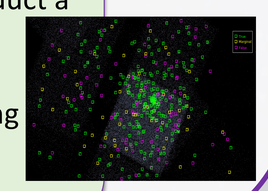
Combine and Detect
- Create stacked products by combining per-OBI products
- Detect sources on the stacked event files using 2 methods – wavdetect and Voronoi Tesselation



Validate Sources
- Chose best detection between 2 detection algorithms
- Organize detections into bundles of overlapping sources

QA: Reviewer can manipulate detections in difficult fields

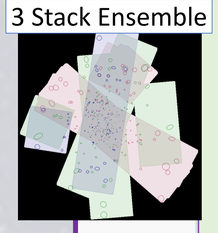
Stack Final Detection List
- Create the mrgsrc product a combination of all processing steps
- Qualify detections using likelihoods into True Marginal or False



Master Match

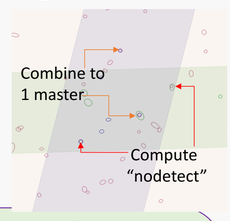
After all detections are found, the true and marginal detections within groups of overlapping stacks (ensembles) are compared against each other to determine the final list of sources in the field.

3 Stack Ensemble



Find Master Sources
- Identify overlapping detections with significant overlap as the same source
- Ignore overlapping sources from the same stack
- Sources that overlap multiple sources have an ambiguous association

QA: Cases when too many overlaps cannot clearly identify the actual sources



Find Missed Detections
- Find areas in valid parts of the observation where no detection was found
- "Nodetects" are used as upper limits in photometry

Source Properties

Once all the sources have been identified, each source is characterized at the OBI, stack and master levels. The catalog identifies on the order of 350 properties for each source. When expanded out for each band and when including errors, there are on the order of 1600 properties computed.

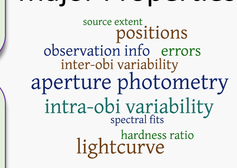
Per-OBI Properties
- Compute properties for each OBI the detect source is valid.

Bayesian Blocks
- For all detections associated with the source, create groups of similar OBIs with based on flux

Stack Properties
- Compute properties averages across all OBIs in the detect stack

Master Properties
- For each Bayesian group, compute properties averages across all OBIs in the group

Major Properties



CHANDRA SOURCE CATALOG

This work has been supported by NASA under contract NAS 8-03060 to the Smithsonian Astrophysical Observatory for operation of the Chandra X-ray Center.