The Chandra Source Catalog 2.0: Early Cross-matches

Arnold H. Rots[1], Tamás Budavári[5], Dan T. Nguyen[1], Douglas Burke[1], Roger M. Hain[1], Francesca Civano[1], Christopher Allen[1], Joseph B. Miller[1], Craig S. Anderson[1], Jamie A. Budynkiewicz[1], Judy C. Chen[1], Raffaele D'Abrusco[1], Stephen M. Doe[2], Ian N. Evans[1], Janet D. Evans[1], Giuseppina Fabbiano[1], Danny G. Gibbs II[1], Kenny J. Glotfelty[1], Dale E. Graessle[1], John D. Grier[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], Joy S. Nichols[1], Michael A. Nowak[4], Charles Paxson[1], David A. Plummer[1], Francis A. Primini[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 [5]The Johns Hopkins University

Cross-matching the Chandra Source Catalog (CSC) with other catalogs presents considerable challenges, since the Point Spread Function (PSF) of the Chandra X-ray Observatory varies significantly over the field of view. For the second release of the CSC (CSC2) we have been developing a cross-match tool that is based on the Bayesian algorithms by Budavari, Heinis, and Szalay (ApJ 679, 301 and 705, 739), making use of the error ellipses for the derived positions of the sources.

However, calculating match probabilities only on the basis of error ellipses breaks down when the PSFs are significantly different. Not only can bonafide matches easily be missed, but the scene is also muddied by ambiguous multiple matches. These are issues that are not commonly addressed in cross-match tools. We have applied a satisfactory modification to the algorithm that, although not perfect, ameliorates the problems for the vast majority of such cases.

Here we present some early cross-matches of the CSC2 catalog with obvious candidate catalogs—SDSS and Wise—and report on the determination of the absolute astrometric error of the CSC2 .

Cross-matching

- Cross-matching historically started out as a visual exercise
- Subsequently, cross-matching catalogs was based on a fixed radial proximity match (say, within 1")
- This becomes problematic, as it does not work well for catalogs that are:
  - taken from very different parts of the electromagnetic spectrum— detected sources may represent physically different objects which may or may not be visible at other frequencies/energies/wavelengths
  - derived from observations with significantly different PSFs— a source in a lower-resolution catalog may represent a blend of multiple sources from a higher-resolution one
- The reliability of the matches can be significantly improved by calculating rigorous match probabilities based on the detected positions and the detailed uncertainties therein
- The approach we have chosen is to use the algorithms developed by Budavári & Szalay: for *n* catalogs, calculate a Bayes Factor for each *n*-tuple of sources based on their error ellipses; assign a flat prior probability to all tuples; calculate a posterior probability based on the prior and Bayes Factor; iterate by using a new the prior based on the sum of probabilities; for algorithmic details, see the white box in the lower right corner of this poster
- A self-consistency argument sets the acceptance threshold based on the final sum of probabilities
- Although the examples given here are catalog pair matches, the algorithm can handle matching any number of catalogs simultaneously; this is done efficiently by identifying sets of connected components based on pairwise match probabilities, using a graph theory library

Resolution

Accidents can happen when the spatial resolutions of the catalogs differ significantly.

- When a high-resolution catalog has two or more sources fairly closely together, these may be blended into a single source in a low resolution catalog
  - The error ellipse of the large blended source may well be small compared to the separations of the high resolution sources, resulting in incomplete or no matches, yielding match probabilities that are plainly wrong
  - A similar situation can arise when the two PSFs are similar in size and shape, but have a significant eccentricity and very different position angles
- We have solved this by calculating Bayes factors for each source tuple based on the error ellipses and based on the PSF ellipses, and using the larger of the two for that source tuple

Ambiguous Matches

The resolution problem raises an important issue that has not been commonly recognized in cross-match operations: with widely varying PSF sizes from different catalogs, matches — even those with a high probability — are not necessarily 1-to-1 anymore; they may be one-to-one, one-to-many, or many-to-one.

- As long as a source has a high probability match with no more than one source per (other) catalog, it can be unambiguously identified as representing the same physical object
- As soon as there are matches with more than one source in a single catalog, the source becomes ambiguously identified — unless those other sources turn out to be ambiguous
  - That last case refers to high resolution sources being matched with multiple low resolution sources
  - To disentangle this, we keep track of pairwise potentially ambiguous matches, defined as matches with sources that have a much smaller PSF

CSC2 and SDSS DR13

This cross-match involved 1574 distinct contiguous areas covered by the CSC2, matching with all SDSS "good stars" within the bounding box of each CSC2 area

	CSC2	SDSS DR13
Sample coverage area (sq deg)	197	402
Total number of sources/stars	87,276	2,609,153
Unambiguous matches	17,705	17,705
Ambiguous matches	1,061	14

CSC2 and AllWISE SC

This cross-match involved 4352 distinct contiguous areas covered by the CSC2, matching with all sources from the AllWISE catalog that are within the bounding box of each CSC2 area

	CSC2	WISE
Sample coverage area (sq deg)	568	1147
Total number of sources/stars	330,758	22,633,344
Unambiguous matches	132,519	132,519
Ambiguous matches	5,146	1,863

Astrometric Uncertainty

The systemic astrometric uncertainty (1-σ) of the CSC2 was determined to be 0.29" on the basis of a statistical analysis of the separations of matched pair components from the cross-match against the SDSS DR13 stars. The same technique was used for CSC1.

Details

Bayes Factors: $$B = 2^{n-1}\frac{\sqrt{|\bar{\bar{\boldsymbol{K}}}|}}{\prod_{i=1}^{n}\sqrt{|\bar{\bar{\boldsymbol{C}}}_i|}}\exp\left\{\frac{1}{2}\left(\overline{\boldsymbol{y}}^T\cdot\bar{\bar{\boldsymbol{K}}}^{-1}\cdot\overline{\boldsymbol{y}}-\sum_{i=1}^{n}\overline{\boldsymbol{x}_i}^T\cdot\bar{\bar{\boldsymbol{C}}}_i^{-1}\cdot\overline{\boldsymbol{x}_i}\right)\right\}$$

Where $\boldsymbol{C}_i$ are the covariance matrices of the sources, $\boldsymbol{x}_i$ their positions, $\boldsymbol{K}$ the inverse of the sum of the inverse covariance matrices, and $\boldsymbol{y}$ the center position.

First Prior: $P_0(0)=\frac{\min(N_i;\, i=1..n)}{\prod_{i=1}^{n}N_i}$ Posterior for tuple j: $P_j(k)=\left(1+\frac{1-P_0(k)}{B_j\cdot P_0(k)}\right)^{-1}$

Iterate on subsequent Priors: $P_0(k+1)=\frac{\sum_j P_j(k)}{\prod_{i=1}^{n}N_i}$ (N_i: number of sources in catalog i)

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.