The Chandra Source Catalog: Statistical Characterization

Francis A. Primini[1], M. A. Nowak[2], J. C. Houck[2], J. E. Davis[2], K. J. Glotfelty[1], M. Karovska[1],
C. S. Anderson[1], N. R. Bonaventura[1], J. C. Chen[1], S. M. Doe[1], I. N. Evans[1], J. D. Evans[1], G. Fabbiano[1], E. Galle[1], D. G. Gibbs[1], J. D. Grier[1], R. Hain[1], D. M. Hall[1], P. N. Harbo[1], X. He[1], J. Lauer[1], M. L. McCollough[1], J. C. McDowell[1], J. B. Miller[1], A. W. Mitschang[1], D. L. Morgan[1], J. S. Nichols[1], D. A. Plummer[1], B. L. Refsdal[1], A. H. Rots[1], A. L. Siemiginowska[1], B. A. Sundheim[1], M. S. Tibbetts[1], D. W. Van Stone[1], S. L. Winkelman[1], P. Zografou[1]

[1]Smithsonian Astrophysical Observatory, [2]MIT Kavli Institute for Astrophysics and Space Research

SUMMARY

The Chandra Source Catalog (CSC) will ultimately contain more than ~250000 x-ray sources in a total area of ~1% of the entire sky, using data from ~10000 separate ACIS and HRC observations of a multitude of different types of x-ray sources (see Evans et al. Poster 472-01). In order to maximize the scientific benefit of such a large, heterogeneous dataset, careful characterization of the statistical properties of the catalog, i.e., completeness, sensitivity, false source rate, and accuracy of source properties, is required. Characterization efforts of other, large Chandra catalogs, such as the ChaMP Point Source Catalog (Kim et al. 2007, 2007ApJS..169..401K) or the 2 Mega-second Deep Field Surveys (Alexander et al. 2003, 2003AJ....126..539A), while informative, cannot serve this purpose, since the CSC analysis procedures are significantly different and the range of allowable data much less restrictive. We describe here the characterization process for the CSC. Our plans include both a comparison of real CSC results with those of other, deeper Chandra catalogs of the same targets as well as extensive simulations of blank-sky and point source datasets. We present preliminary results from our work to date. This work is supported by NASA contract NAS8-03060 (CXC).

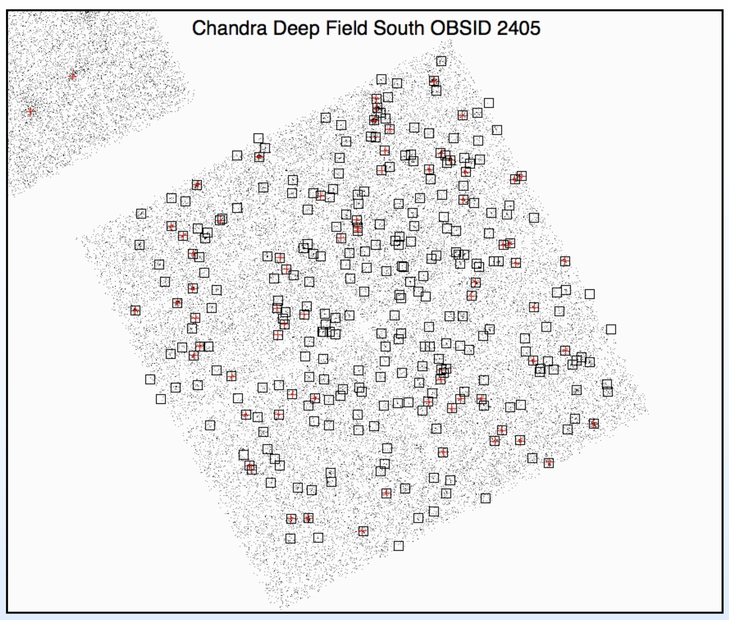
CONCLUSIONS

Preliminary analysis of both simulations and real datasets indicates that the statistical properties of the CSC may be characterized as follows:

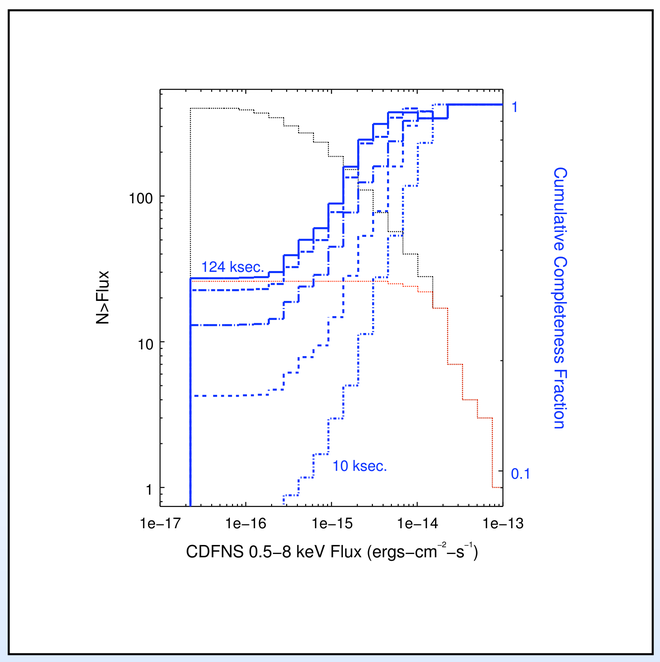
- False source rates of << ~1 per observation for exposures ≲ 50 ksec, and ≲1 per observation for exposures ≳ 120 ksec.;
- 50% completeness at broad-band fluxes of ~7×10^{-15} erg-cm^{-2}-s^{-1} for a ~10 ksec. observation, improving to ~9×10^{-16} erg-cm^{-2}-s^{-1} for ~125 ksec. observations;
- Positional uncertainties of ≲ 1″ (1σ) for sources with more than 10 counts, located within 3′ of the optical axis, ≲ 2″ for sources with more than 30 counts, within 10′, and ≲ 5″ for sources with more than 50 counts, within 15′;
- Flux accuracies of better than ~35% for sources with more than 30 counts, within 3′ of the optical axis, ~30% for sources with more than 50 counts, within 10′, and ≲ 5″ for sources with more than 100 counts, within 15′;

COMPARISONS WITH OTHER CATALOGS

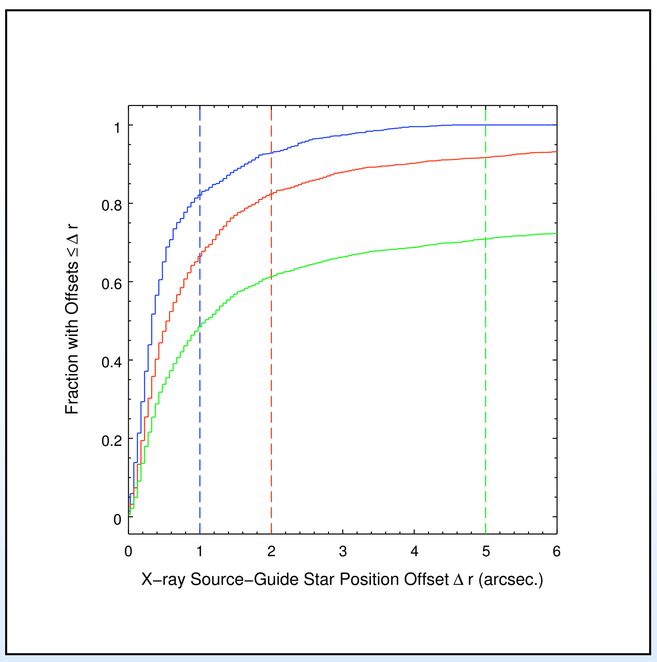
We compare the CSC to external catalogs based on Chandra data. Such catalogs provide an independent check on CSC source properties, using readily available data. However, they don't cover the full range of parameter space needed to fully characterize the CSC.



We compare CSC sources from individual observations to those detected in the Chandra Deep Fields (CDF) (Alexander et al. 2003, A.J., 126, 539). The CDFN and CDFS sources are derived from an analysis of stacked Chandra ACIS-I images, totaling ~1-2 Msec, and so can be considered complete at the exposures of individual observations. In the above 60 ksec observation, CDFS sources are indicated by black squares and CSC sources by red crosses. We consider CSC and CDFS sources to match if their positional offsets are less than 5 times the CSC position error. Since all CSC sources are associated with CDFS sources, there are no false sources in this observation (the CDFS catalog did not include the ACIS-S chips).



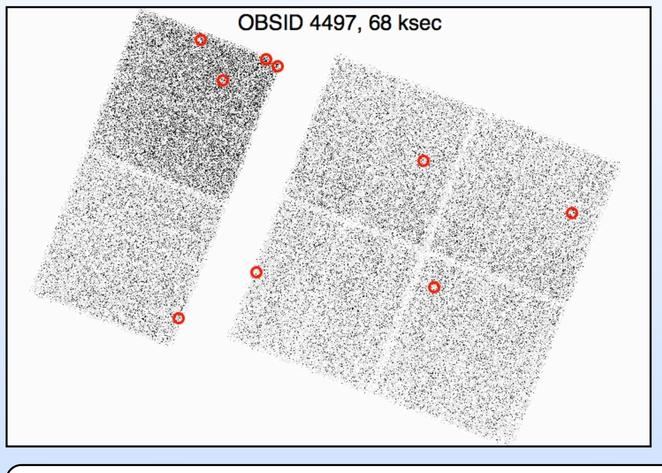
We examine 5 ACIS-I observations in the CDFN or CDFS catalogs, with exposures from ~10 - ~125 ksec. For each, we construct cumulative flux distributions for all CDF sources (black curve above), and for CDF sources associated with CSC sources (red curve). The ratio of the two distributions (blue curve) provides an estimate of the completeness fraction for the CSC. Flux distributions for a 10 ksec observation are shown. Completeness fractions for 10 and 124 ksec observations are labeled. Fractions for intermediate exposures of 30, 60, and 95 ksec are also plotted.



To estimate astrometric uncertainty, we compare the CSC to the Chandra Guide Star Catalog (http://cxc.harvard.edu/cgi-gen/cda/agasc/agascInterface.pl). The latter is guaranteed to include several stars in each CSC field, many of which may be x-ray sources. We consider CSC and AGASC sources to match if their positional offsets are less than 5 times the CSC position error. Here, we display the cumulative distributions of offsets for 3 different subsets of CSC sources: sources with > 10 net b-band counts and off-axis angle $\theta < 3'$ (blue), sources with > 30 net counts and $\theta < 10'$ (red), and sources with > 50 net counts and $\theta < 15'$ (green). Vertical dashed lines indicate 1σ catalog requirements for those sets (cxc.cfa.harvard.edu/csc/memos/files/Evans_Requirements.pdf, p.12).

SIMULATIONS

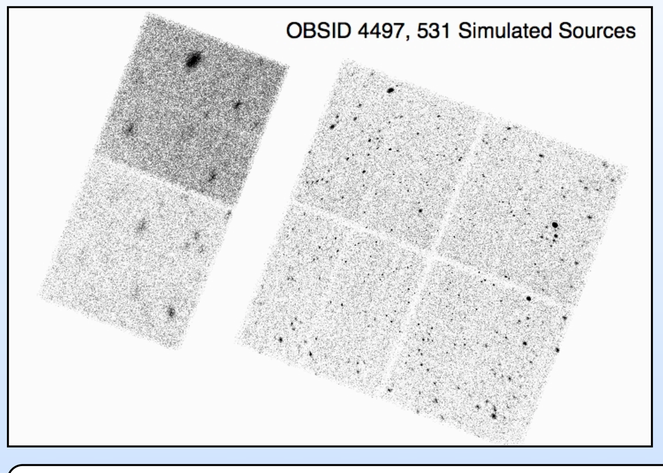
Simulated datasets can cover the full range of CSC source and field properties, but must be carefully designed and tested to ensure that they are both realistic and statistically independent. We simulate blank-sky datasets by sampling from ACIS background event lists in CALDB (http://cxc.harvard.edu/contrib/maxim/acisbg/COOKBOOK), normalized to match the metadata in selected ACIS-I and ACIS-S observations. Point sources are injected into blank-sky event lists using Marx, by sampling from a standard extragalactic logN-logS distribution and a uniform spatial distribution. Sources with both power-law and blackbody spectra are simulated. All simulations are then processed with the standard CSC source detection and properties pipelines. Additional characterization of source properties such as size or variability is done via simulations of input to the specific CSC analysis programs (see, e.g., Nowak et al., Poster 472-11).

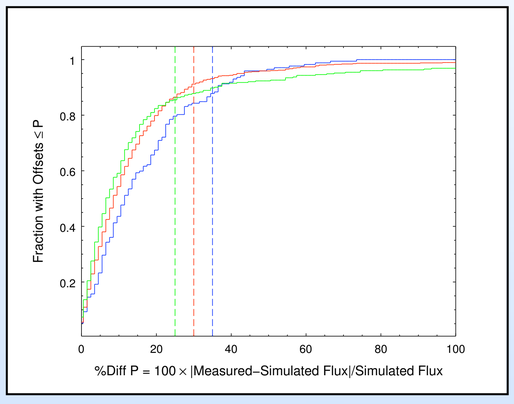


OBSID	CHIP SET†	TIME (ksec)	Sources (#Runs)	FSR
4497	ACIS-I	68	9 (50)	0.18
927	ACIS-I	125	64 (50)	1.28
4404	ACIS-S	30	5 (50)	0.1
7078	ACIS-S	51	18 (50)	0.36
4613	ACIS-S	118	88 (47)	1.87
4613‡	ACIS-S	118	8 (10)	0.8

† For the purposes of these simulations, the ACIS-I chip set consists of chips 0, 1, 2, 3, 5, & 6, and ACIS-S includes chips 2, 3, 5, 6, & 7. We are unable at present to simulate background in chip 8, due to its significant instrumental artifacts, and so exclude it from false source rate calculations.

‡ In our initial simulation runs, we normalized the background in each chip using the field background rates from Table 6.9 of the Chandra Proposer's Observatory Guide (see cxc.harvard.edu/proposer/POG_CYC10/html/chap6.html#tb:brates_i). Because of the more restrictive data screening in the CSC, these rates overestimate the actual CSC background rates. An additional run of this obsid, using background rates determined from source-free regions of the image yields a much smaller FSR.





The real CSC event list for OBSID 4497 is used to "seed" blank-sky simulations with metadata appropriate for that observation (exposure time, instrument configuration, aspect orientation and motion). Multiple blank-sky simulations are analyzed with the standard CSC pipeline to estimate false source rates. The image above shows all the false sources detected in 50 simulations of 4497, superimposed on a single blank-sky simulation. The table lists false source rates for several ACIS-I and ACIS-S simulation sets. Simulations of shorter exposures typically yield no false sources.

Flux accuracy is estimated by comparing simulated fluxes with those measured from CSC processing. In the image above, we show sources from a single simulation run for OBSID 4497. To the right, we display cumulative distributions of % difference from 10 simulation runs, containing ~2800 sources. Sources are divided into 3 different subsets: CSC sources with > 30 net b-band counts and off-axis angle $\theta < 3'$ (blue), sources with > 50 net counts and $\theta < 10'$ (red), and sources with > 100 net counts and $\theta < 15'$ (green). Vertical dashed lines indicate 1σ catalog requirements for those sets of sources (cxc.cfa.harvard.edu/csc/memos/files/Evans_Requirements.pdf, p. 20).

For more information about the *Chandra* Source Catalog, please refer to the public catalog web pages at http://cxc.cfa.harvard.edu/csc/