
# Progress Report

**Ian Evans**
**On behalf of the Chandra Source Catalog Project Team**

**Chandra Users' Committee Meeting**
**May 01, 2015**

# Production Status



## Started Production — Monday, April 13

- Ran 1 stack end-to-end in production with full team checkout
- Initially injected 1500 stacks into production queue
  - Data archived and (temporarily) saved on disk for ongoing team checkout
  - Checking out manual QA process (will transition to DSOps team for remainder of production)
  - Initially forcing most large (> 8 ObsId) stacks to manual QA for review

## Production Processing Statistics

- Through 2015 April 30, 03:50Z

| | |
|---|---|
| Stacks approved for processing: | 3114 |
| Held from processing: | 42[1] |
| Ready and awaiting kick off: | 574 |
| Total kicked off to date: | 2498 |
| Completed processing/ingested into archive: | 2209 (30.2% of stacks) |
| Currently processing: | 19 |
| Waiting for manual QA: | 268[2] |
| Paused due to error: | 2 |
| Number of detections classified as TRUE: | 73876 |
| Number of detections classified as MARGINAL: | 33663 |
| Total number of TRUE or MARGINAL detections: | 107539[3] |

[1]All HRC-I currently being held from processing
[2]Includes 77 forced manual QA with no other issues
[3]For comparison, the total number of detections in Rel. 1 was 158071

# Chandra Source Catalog Highlights



## Highlights

- Developed/refined pipelines through `stacker` pipeline
  - Refined algorithms based on science testing and data review
    - Finalized several "bad detection" filters and parameters
    - Finalized selection of point/extended source models and parameters
    - New algorithm to determine whether an "extended" source is really a high density of point sources
  - Resolved poor position error determinations
    - Now use "*get_draws*" Bayesian algorithm in *Sherpa*
    - Studied results to characterize; tuned parameters for point sources
    - Worked "extended" sources case (more free parameters; <2% of sources)
  - Identified and resolved bugs/unexpected issues as necessary
  - Conducted pipeline/parameter walkthru; settled production parameter settings
- Science Testing
  - Continued to support extensive science testing through `stacker` pipeline
  - Generating and running simulations to set key parameters for production
- Infrastructure
  - Review/feedback of automated and manual QA system
  - Tested archive ingest and replication to establish runtimes
  - Tested system with Linux test archive in preparation for June migration
  - Prepared system for Phase II run of any anomalous processing
  - Established a system for Reprocessing
  - Baselined HW and disk space for production

## Software Releases Completed

| Release | Date | Summary |
|---|---|---|
| **Science data test and feedback (Oct 21)** | | |
| CAT 4.2.6 | Nov 05 | Address SAOTrace issue; 0 counts on ccd (corner case); fix to MLE for off center points |
| CAT 4.2.4.1 | Dec 08 | JBOSS server fix for VO services (Rel. 1) |
| CAT 4.2.7 | Dec 20 | Fine tune file formats, QA, parameter defaults |
| CAT 4.2.8 | Jan 16 | CSC Web Services upgrade: CSCview, CSC VO, CSC web services upgrade (Rel. 1) |
| CAT 4.2.9 | Feb 12 | Pixel Mask corrections in MLE |
| **Pipeline/Parameter Walkthru/Review (Feb 24)** | | |
| CAT 4.3 | Mar 05 | Full rebuild of CAT system; Feedback from PL/param review |
| CAT 4.3.1 | Mar 13 | Complete parameter feedback; QA, CalDB default set to match SDP (v4.6.7) |

# CSC Software Release Summary (cont.)



## Software Releases Completed

| Release | Date | Summary |
|---|---|---|
| CAT 4.3.2 | Mar 19 | `stacker` PL review feedback; MLE file/data fixes |
| CAT 4.3.3 | Mar 24 | Select `wavdetect`/`mkvtbkg` alg. for source position; Rotation angle algorithm update; `mrgsrc3` draws errors stack/obi info fix; fine astrometry QA check |

**CXC science review of CSC data — (Mar 25 – Apr 02)**

| | | |
|---|---|---|
| CAT 4.3.4 | Mar 31 | QA feedback from internal review; param update on `wavdetect`/`mkvtbkg` tools; MLE 90 degree ext src rot angle fix; data prod header review/feedback |
| CAT 4.3.5 | Apr 02 | Science feedback; parameter setting tweaks |
| CAT 4.3.6 | Apr 07 | Phase II prep update; MLE random seed update |
| CAT 4.3.7 | Apr 09 | Final parameter setting science feedback; QA |

**Ready for CSC Rel. 2.0 production run through `stacker` pipeline**

# Testing



## Software Release Testing

- Regression test of 6 stacks run after each software release verified by catalog software team
- Followed by randomly selected set of 400 stacks from draft input list evaluated by both catalog science and software teams
  - Science team focus on scientific validation of updates in release, general scientific validation, and identification of anomalies or incorrect behavior
- Added "stress" test of 20 stacks as necessary
  - Verifies specific issues but has a long run time

## Science Team Testing Focus Since October

- Completed analysis of HRC-I simulations of point and extended sources to set *preliminary* HRC-I likelihood thresholds (similar to previously reported effort for ACIS)
- Compared point and extended compact source fits and evaluated thresholds for selecting point vs. extended source fit
- Compared properties of isolated source fits with those of sources in dense groups (single-source vs. multi-source bundles)
- Test compact source fit confidence intervals determined from MCMC analysis (Sherpa `get_draws`)
  - Source position errors, point and extended source amplitude errors, and extended source parameter errors

# Testing (cont.)



## "External" CXC Science Team Testing

- Recommended by CUC and supported by CXC Director's Office
- Added support by two additional SDS members to support Sherpa `get_draws` confidence interval analysis
  - Team members will provide additional support through final catalog release
- Added support by 8 additional testers from various CXC groups for the "CXC science review of CSC data" test (Mar 25 – Apr 02)
  - Goals were:
    - Provide scientific validation of the quantities in the merged source detection (`mrgsrc3`) data files
    - Look for anomalies in individual source fits
    - Look for statistical anomalies and trends
  - No substantive science issues identified as part of this test, although "lost" detections because of hardware network issues identified (subsequently added check in software to flag this case)
  - Two of the testers also reviewed data from the final pre-production software release test run, with no significant issues identified

# Input Data Selection

## Production Data Preparation

- Pre-filters define acceptable input ObsIds
  - ACIS (timed readmode, faint, vfaint, faint with bias, or graded datamodes) or HRC-I imaging observations
  - Moving target observations excluded
  - SIM Z offset between -20.0 and +20.0 mm; defocus between -0.1 and +0.2 mm
  - Processed through DS 8.4.2 (Repro 4) or later
  - No unresolved errors during observation or standard data processing
- 9578 ACIS ObsIds and 805 HRC-I ObsIds pass the pre-filters
  - These combine to form 6957 ACIS stacks and 345 HRC-I stacks, using a 60 arcsec match metric with complete linkage
    - The two largest ACIS stacks are centered on CDFS (81 ObsId stack totaling 5808 ks) and on Sgr A* (71 ObsId stack totaling 2460 ks)
    - The two largest HRC-I stacks are centered on M31 (61 ObsId stack totaling 1112 ks) and on Vega (41 ObsId stack totaling 136 ks)
- Additional filtering applied in `precaldet` PL
  - E.g., enhanced background flare filtering
  - Eliminates poor quality data but not observations (unless the entire observation is compromised)

# Next Steps

## Production Next Steps

- Continue injecting stacks until processing midterm recess
- Processing midterm **recess** — Late May (~3–4 weeks downtime)
  - *Pause* CSC Rel. 2 production
  - Prepare early-results detection list FITS table
  - Install DS 10.4 — CXCDS operations release (Jun 03)
    - Includes migration of all CXC database and replication servers (including CSC servers) to Linux
    - Upgrade L3 production network hardware
- Processing re-start — Late June
  - Release preliminary detection list FITS tables periodically
  - Estimate completion of processing thru `stacker` PL ~early Fall
    - Includes time for planned final calibration of TRUE/MARGINAL/FALSE detection likelihood thresholds and possible rerun of a subset of data (subset of extended detections, bright/dim detection bundles)

# Next Steps (cont.)



## Software Working/Planned Next Steps

- **CSC Rel. 2.0 production run `master_match` and `source` pipelines**
  - Lien: develop `master_match` and `source` PLs; limiting sensitivity population
- **CSC Rel. 2.0 production run `master` pipeline, final QA review, and release**
  - Lien: `master` PL; populate databases

## Pipeline Definitions

| | |
|---|---|
| `master_match`: | Matches sources across multiple stacks; computes best estimate source positions and assigns source names |
| `source`: | Computes source photometric, spectral, and temporal properties from each observation and stack in which a source is detected |
| `master`: | Produces a canonical set of databased properties for a source |

# Preliminary Detections List



## Content

- A single FITS table containing a subset of information for all of the compact detections with preliminary classification TRUE or MARGINAL
- The table will include the following information:
  - Detection likelihood and preliminary classification
  - Fitted position and 95% internal-error error ellipse
  - Fitted detection amplitude and confidence interval in 4 bands for ACIS (broad, soft, medium, hard) and 1 band for HRC-I (wide)
    - Amplitude is a good proxy for aperture photometry for TRUE point sources in the ACIS b band
  - Classification of detection as point or extended
  - Fitted ellipse parameters for extended detections

## Timeline

- The early-results detections list FITS table will be constructed during the processing midterm recess
  - FITS table and associated documentation will be made available through the CSC website and advertised via a *Chandra* Electronic Announcement
- Updates to the preliminary detections list will be distributed and announced using the same framework

# Preliminary Detections List File Definition

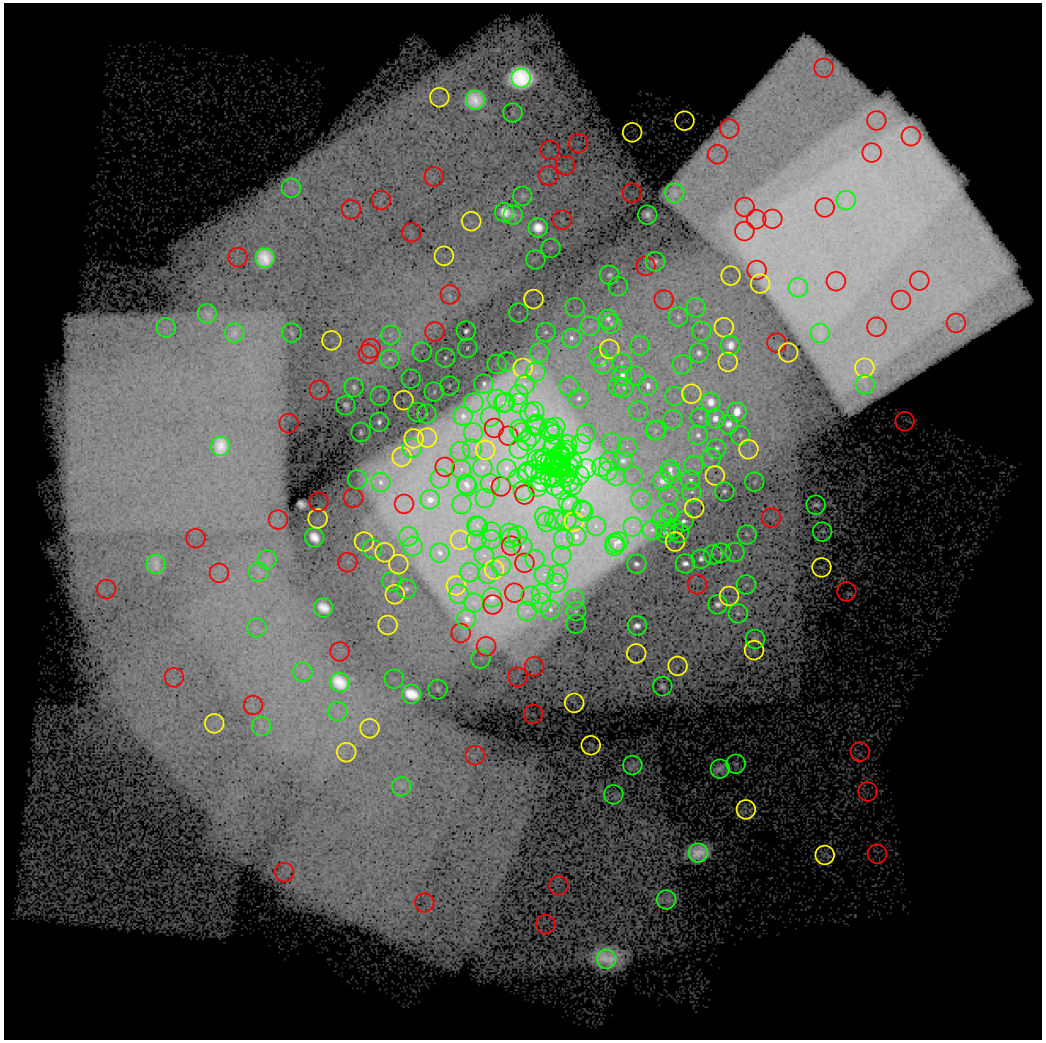


## DETLIST HDU: PRELIMINARY DETECTIONS LIST

| TTYPE | Comment Field (for TTYPE) | TUNIT | Example |
|---|---|---|---|
| detect_id | Stack_id.component | | 'acisfJ0123456p012345_001.0001' |
| likelihood | Highest source log likelihood | | 100.0 |
| src_quality | Preliminary source quality (TRUE/MARGINAL) | | 'TRUE' |
| eband | Energy band of maximum likelihood value | | 'b' |
| theta | Highest likelihood src model mean off-axis ang | arcmin | 0.0 |
| streak_src_flag | Is source on a streak? | | 'F' |
| ra | Highest likelihood source model position RA | deg | 20.94000 |
| dec | Highest likelihood source model position Dec | deg | 1.39583 |
| err_ellipse_ra | 95% confidence error ellipse position RA | deg | 20.94000 |
| err_ellipse_dec | 95% confidence error ellipse position Dec | deg | 1.39583 |
| err_ellipse_r0 | 95% confidence error ellipse semi-major axis | arcsec | 0.35 |
| err_ellipse_r1 | 95% confidence error ellipse semi-minor axis | arcsec | 0.30 |
| err_ellipse_ang | 95% confidence error ellipse smaj axis rotang | deg | 60.0 |
| src_rdata_b | Raw data counts in ECF90 aperture ACIS b band | count | 100 |
| src_ampl_b | Best fit point src model amplitude ACIS b band | photon/cm**2/s | 1.00E-08 |
| src_ampl_lolim_b | 68% confidence lower limit on src_ampl_b | photon/cm**2/s | 9.00E-09 |
| src_ampl_hilim_b | 68% confidence upper limit on src_ampl_b | photon/cm**2/s | 1.10E-08 |
| *Above 4 quantities repeated for ACIS s, m, h bands and HRC w band* | | | |
| eband_ext | Energy band of extended source model properties | | 's' |
| extsrc_class | Extended source classification | | 'POSSIBLE' |
| ext_rdata | Raw data counts in source region aperture | count | 40 |
| ext_ampl | Best fit ext src model amplitude | photon/cm**2/s | 2.50E-09 |
| ext_ampl_lolim | 68% confidence lower limit on ext_ampl | photon/cm**2/s | 1.50E-09 |
| ext_ampl_hilim | 68% confidence upper limit on ext_ampl | photon/cm**2/s | 3.50E-09 |
| ext_smaj | Best fit extended source model semi-major axis | arcsec | 5.0 |
| ext_smin | Best fit extended source model semi-minor axis | arcsec | 3.5 |
| ext_rotang | Best fit extended source model rotation angle | deg | 30.0 |
| ext_smaj_lolim | 68% confidence lower limit on ext_smaj | arcsec | 4.5 |
| ext_smaj_hilim | 68% confidence upper limit on ext_smaj | arcsec | 5.5 |
| ext_smin_lolim | 68% confidence lower limit on ext_smin | arcsec | 3.0 |
| ext_smin_hilim | 68% confidence upper limit on ext_smin | arcsec | 4.0 |
| ext_rotang_lolim | 68% confidence lower limit on ext_rotang | deg | 10.0 |
| ext_rotang_hilim | 68% confidence upper limit on ext_rotang | deg | 50.0 |

# Sample Production Run Datasets





M31 stack acisfJ0042432p411544_001 (21 ObsIds, 299 ks) with preliminary TRUE detections identified in green and preliminary MARGINAL detections identified in yellow. Preliminary FALSE detections are shown in red in this figure. They are retained to allow a planned final calibration of the thresholds based on ongoing simulation runs, and will always be available to end users in the `mrgsrc3`) data files.

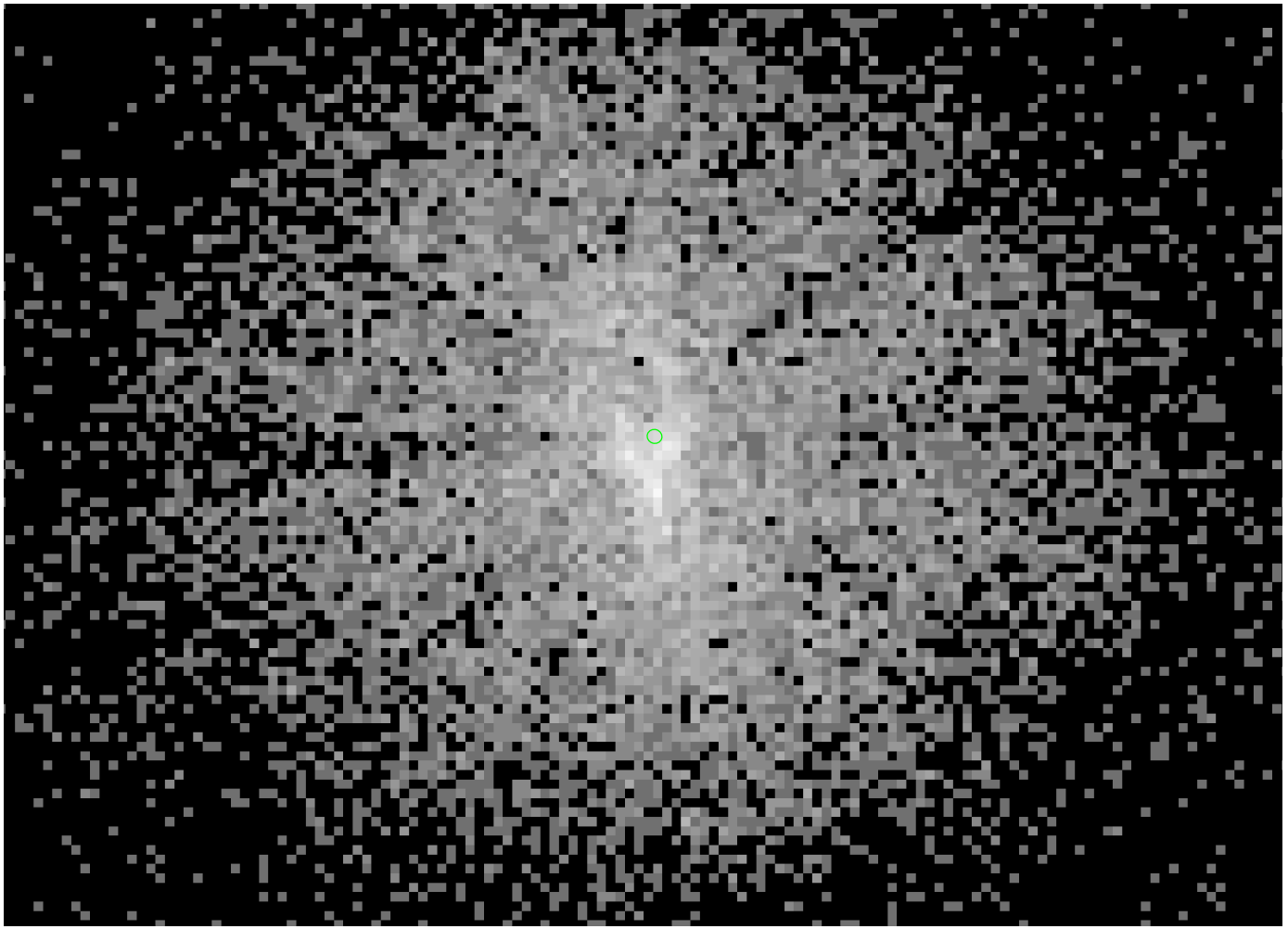
# Sample Production Run Datasets (cont.)



The center of the M31 stack acisfJ0042432p411544_001 with (~95%) absolute position error ellipses for detections shown in green. The error ellipses shown are computed by summing the fit error ellipses computed by Sherpa `get_draws` in quadrature with the systematic error component (0.16 arcsec radial) computed from Rel. 1 of the catalog. The exact systematic error component for Rel. 2 of the catalog will be recomputed after the completion of `master_match`.

# Sample Production Run Datasets (cont.)

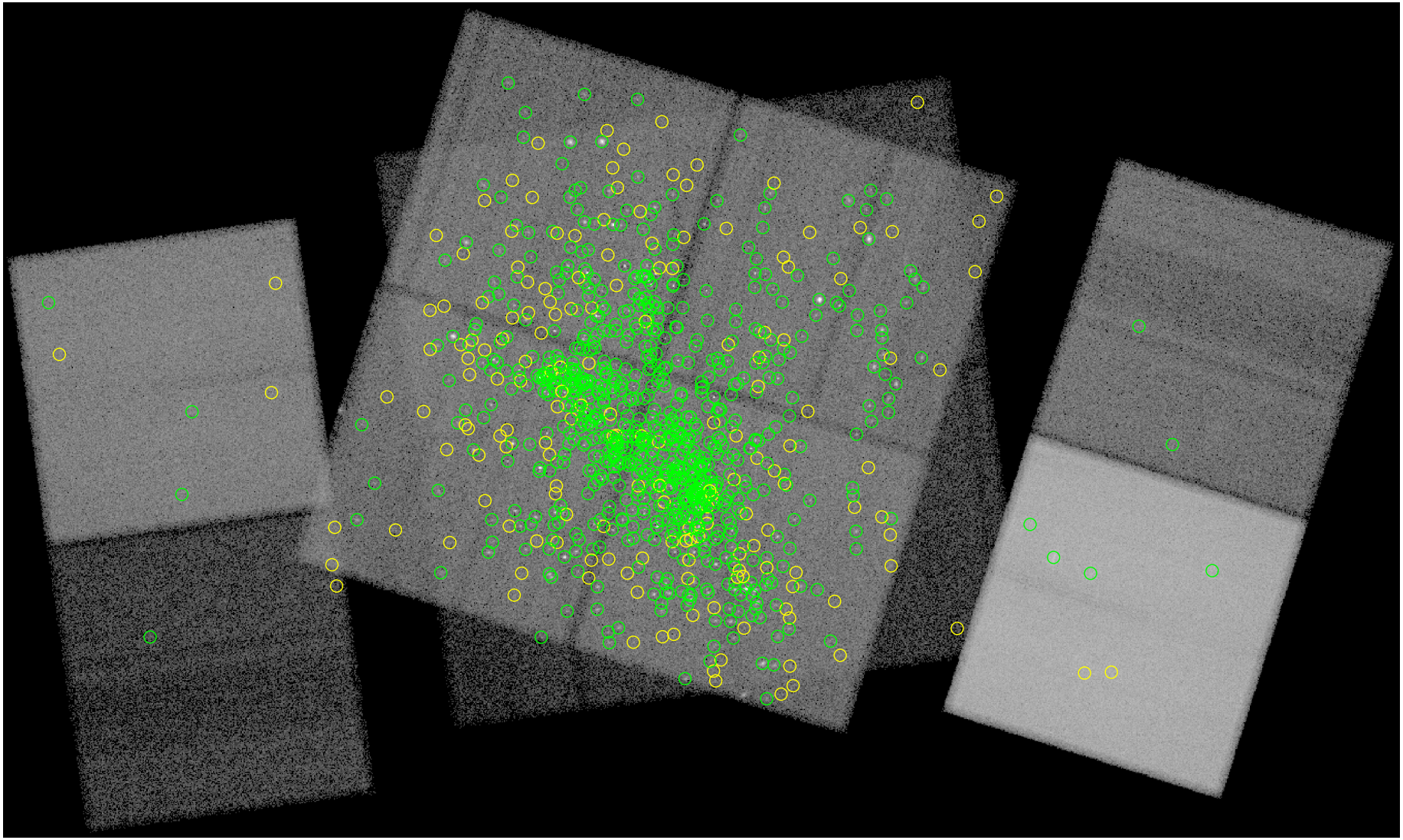




Even far off axis in the M31 stack acisfJ0042432p411544_001, for bright sources the computed position error ellipses can be quite small because the PSF fitting provides more robust results than the simple centroid determination used in Rel. 1 when the PSF is asymmetric. Here the source is 15.3 arcmin off-axis and the error ellipse semi-major axis is ~0.39 arcsec. Systematic errors in the plate scale calibration may dominate the true uncertainties at large off-axis angles.

# Sample Production Run Datasets (cont.)

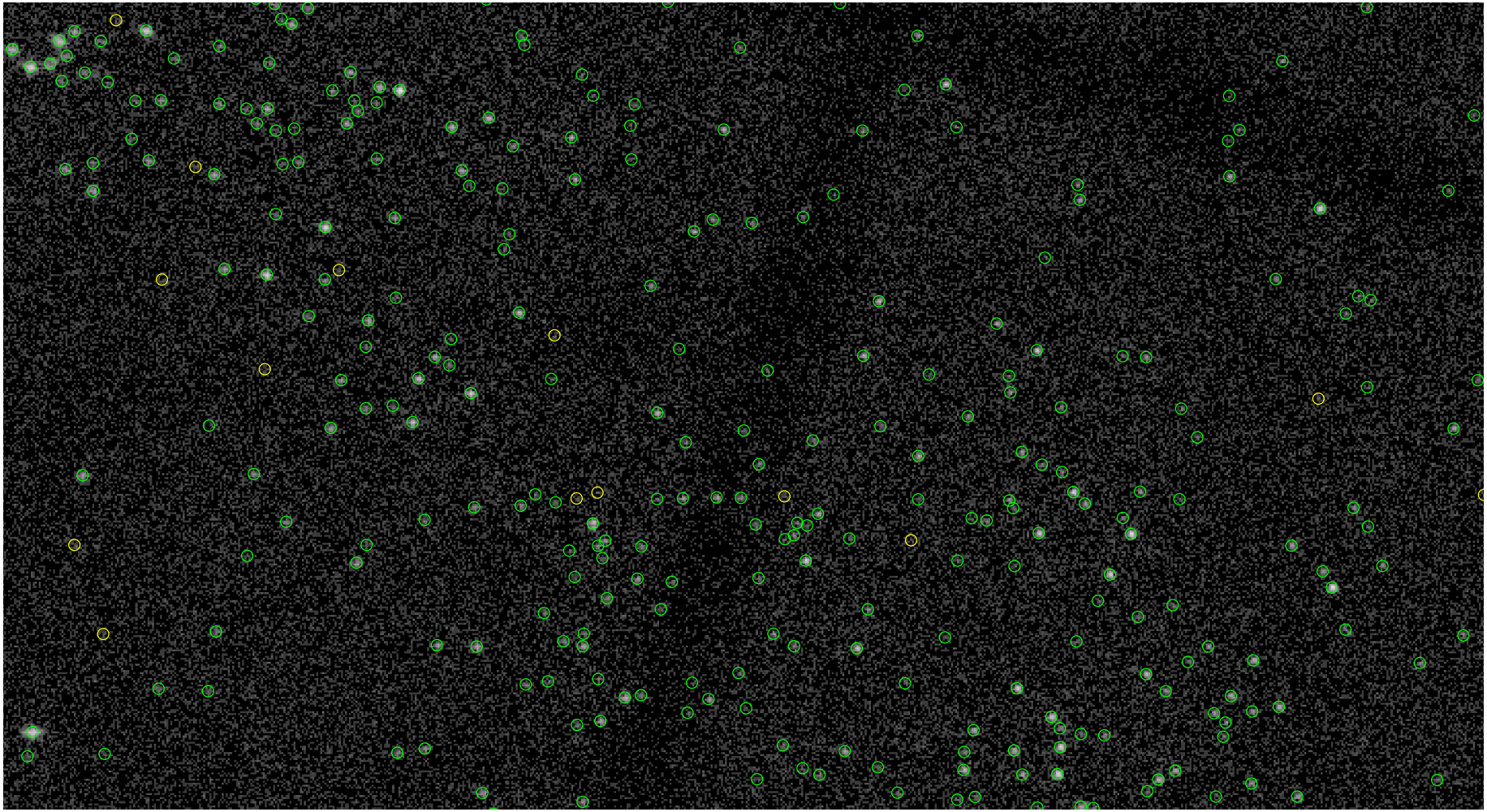




NGC1893 stack acisfJ0522499p332815_001 (5 ObsIds, 457 ks) with preliminary TRUE and MARGINAL detections identified.

# Sample Production Run Datasets (cont.)

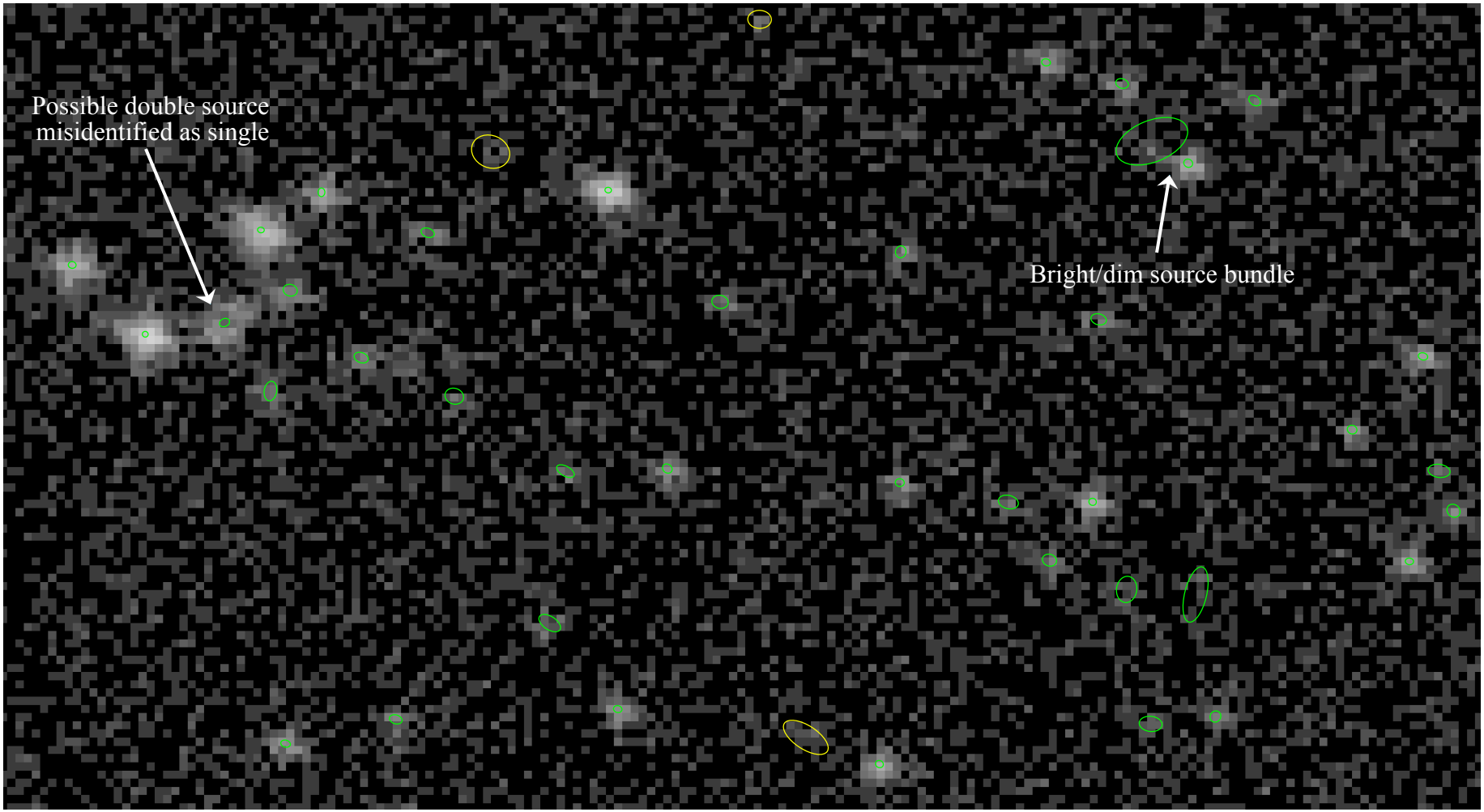




The center of the NGC1893 stack acisfJ0522499p332815_001 with preliminary TRUE and MARGINAL detections identified.

# Sample Production Run Datasets (cont.)

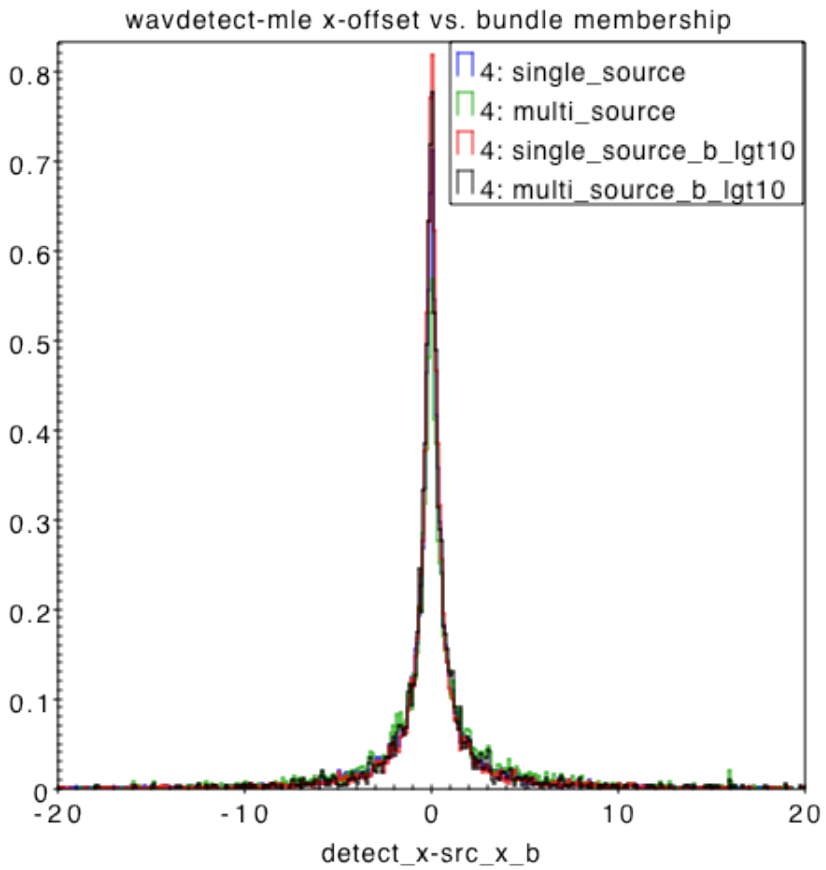


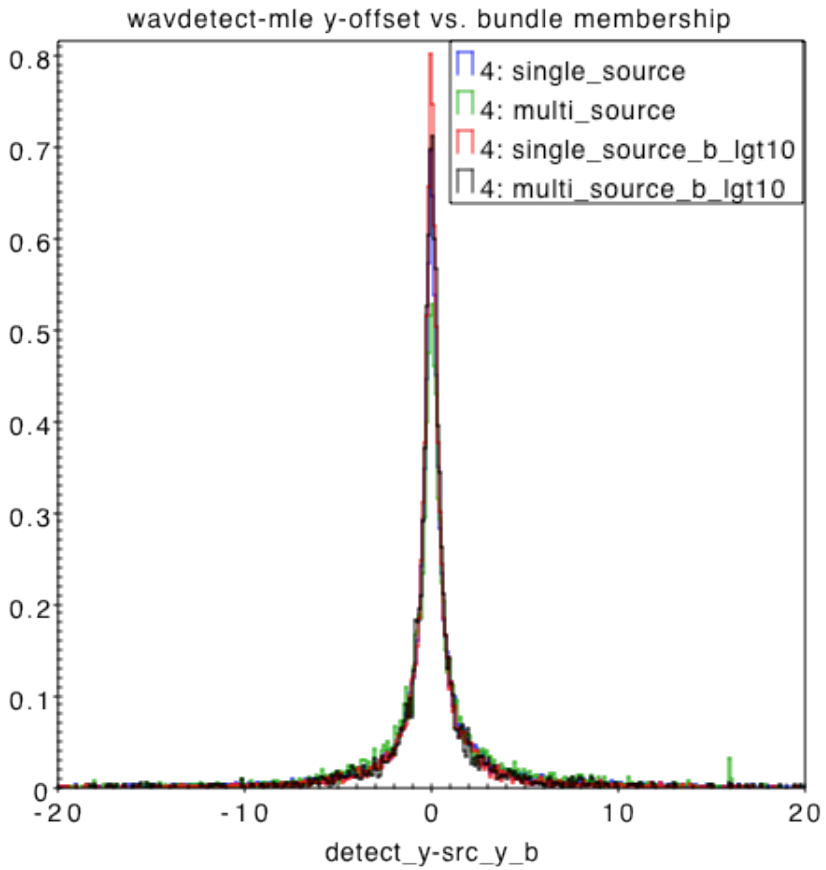



The center of NGC1893 stack acisfJ0522499p332815_001 with (~95%) absolute position error ellipses shown. Fainter sources tend to have larger error ellipses, as expected. Double sources misidentified as single at the source detection stage will not be fitted correctly by MLE in Rel. 2 (nor were they in Rel. 1). Error ellipses for faint sources near bright sources (bright/dim bundles) are sometimes incorrect, and are candidates to be rerun before the final catalog release.

# Preliminary Detection Properties



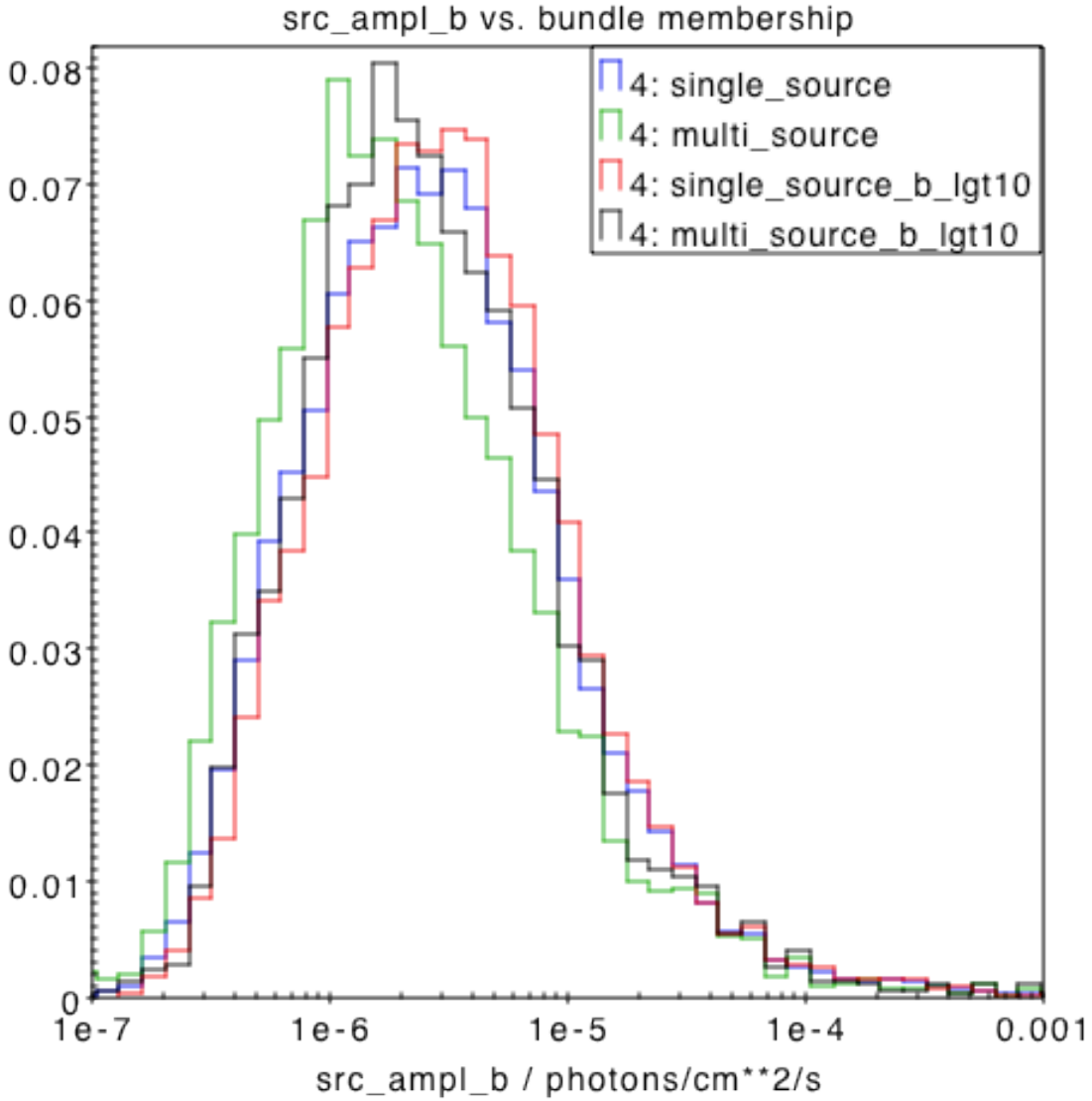



Comparison of position offsets between initial `wavdetect` and final fitted positions shows no major difference between isolated detections and those detected in multi-source bundles, indicating that any systematic errors in fitted positions of detections in multi-source bundles are small compared to random errors.

# Preliminary Detection Properties (cont.)

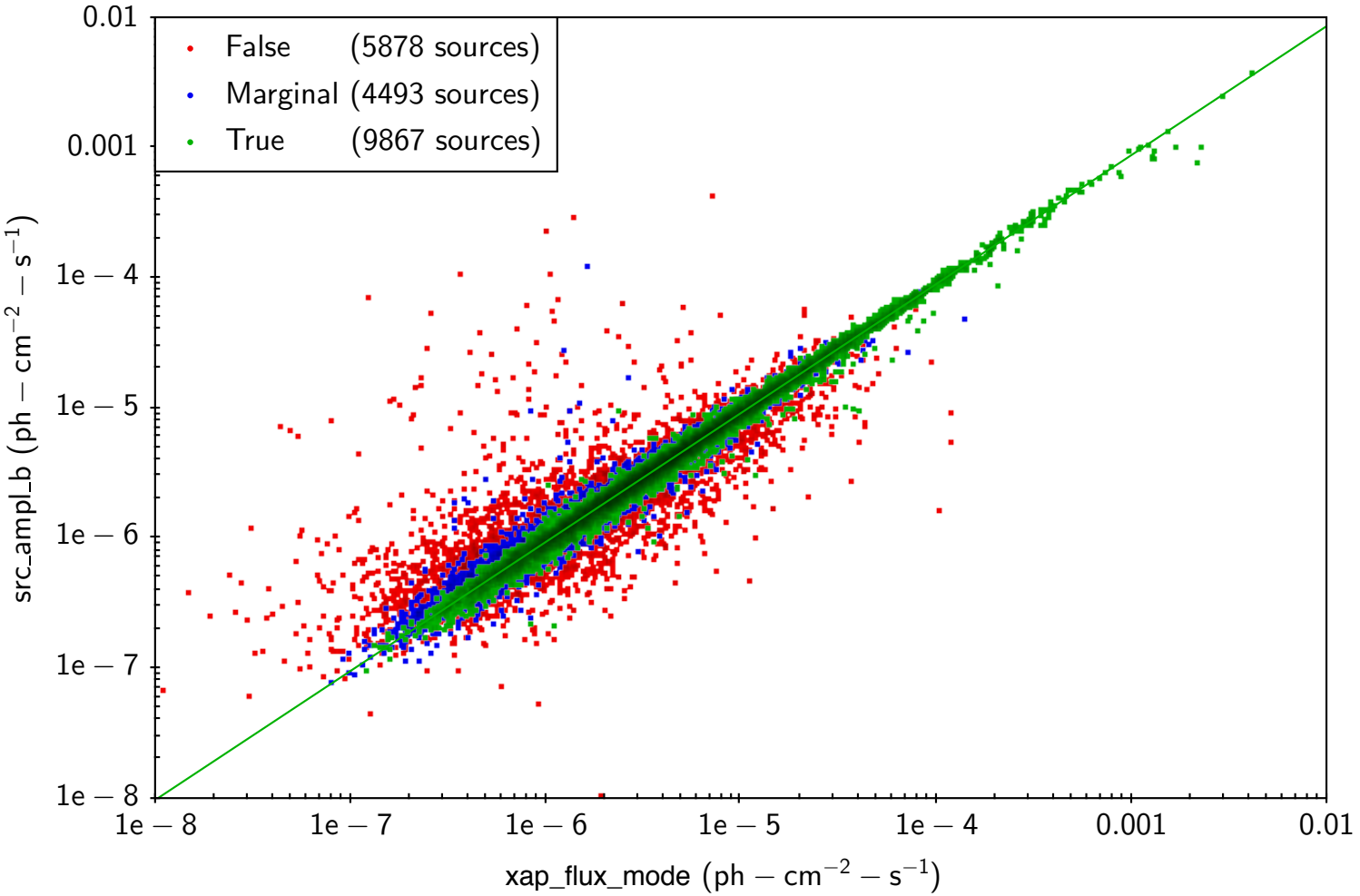





Similarly, the distributions of fitted source amplitude show no major difference between isolated sources and those detected in multi-source bundles.

# Preliminary Photometric Properties

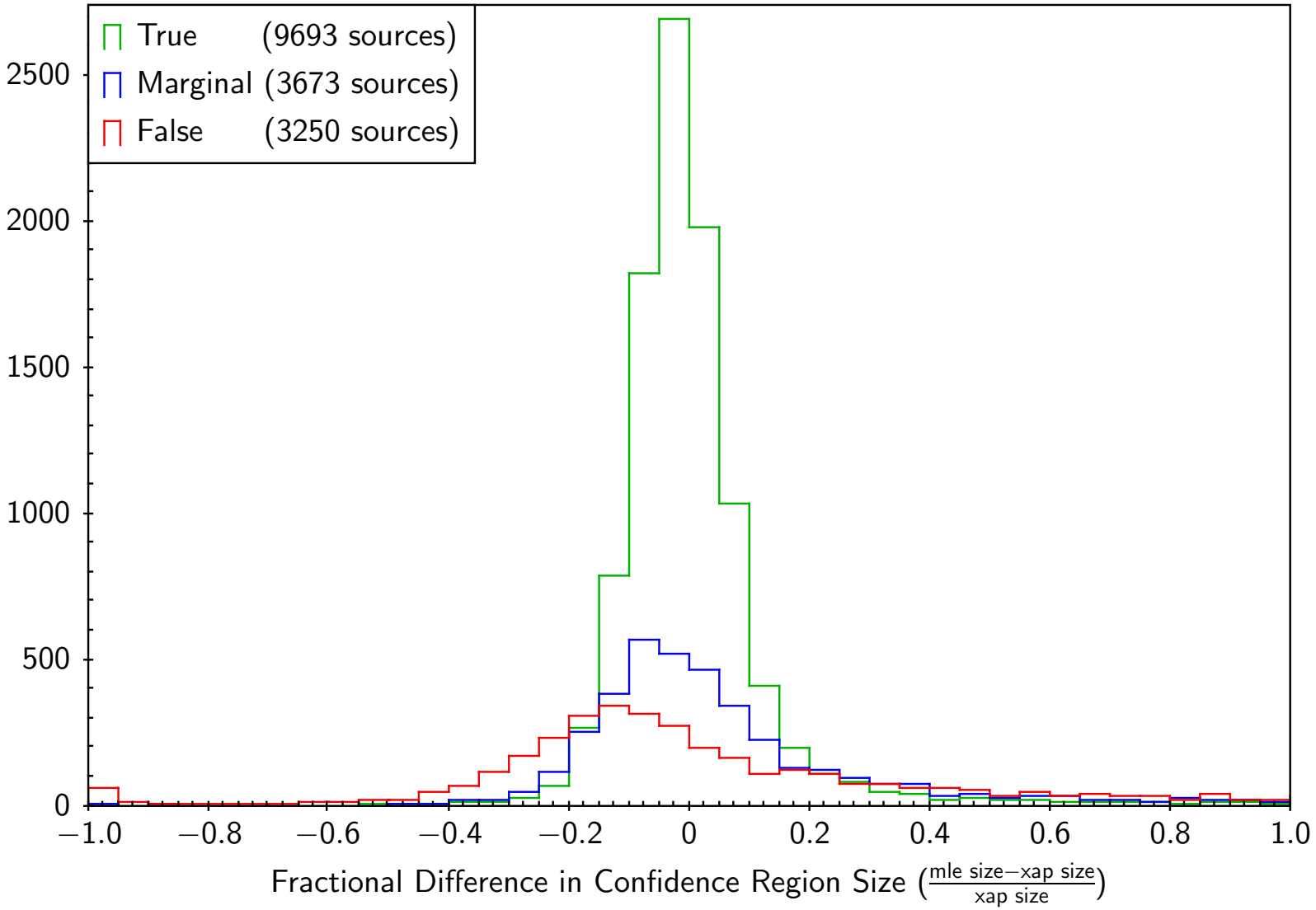





Correlation between ACIS b band fitted amplitude and the corresponding modal photon flux using the full Bayesian X-ray aperture photometry algorithm that will be used in the CSC Rel. 2 source properties pipeline. Analysis includes ~20,000 detections in single-source bundles from single observations. The correlation demonstrates that the fitted amplitude is a good proxy for the flux for detections with preliminary classifications of TRUE. The straight line is an unweighted fit to log(flux) with a slope of 0.991 and a correlation coefficient of 0.995.

# Preliminary Photometric Properties (cont.)

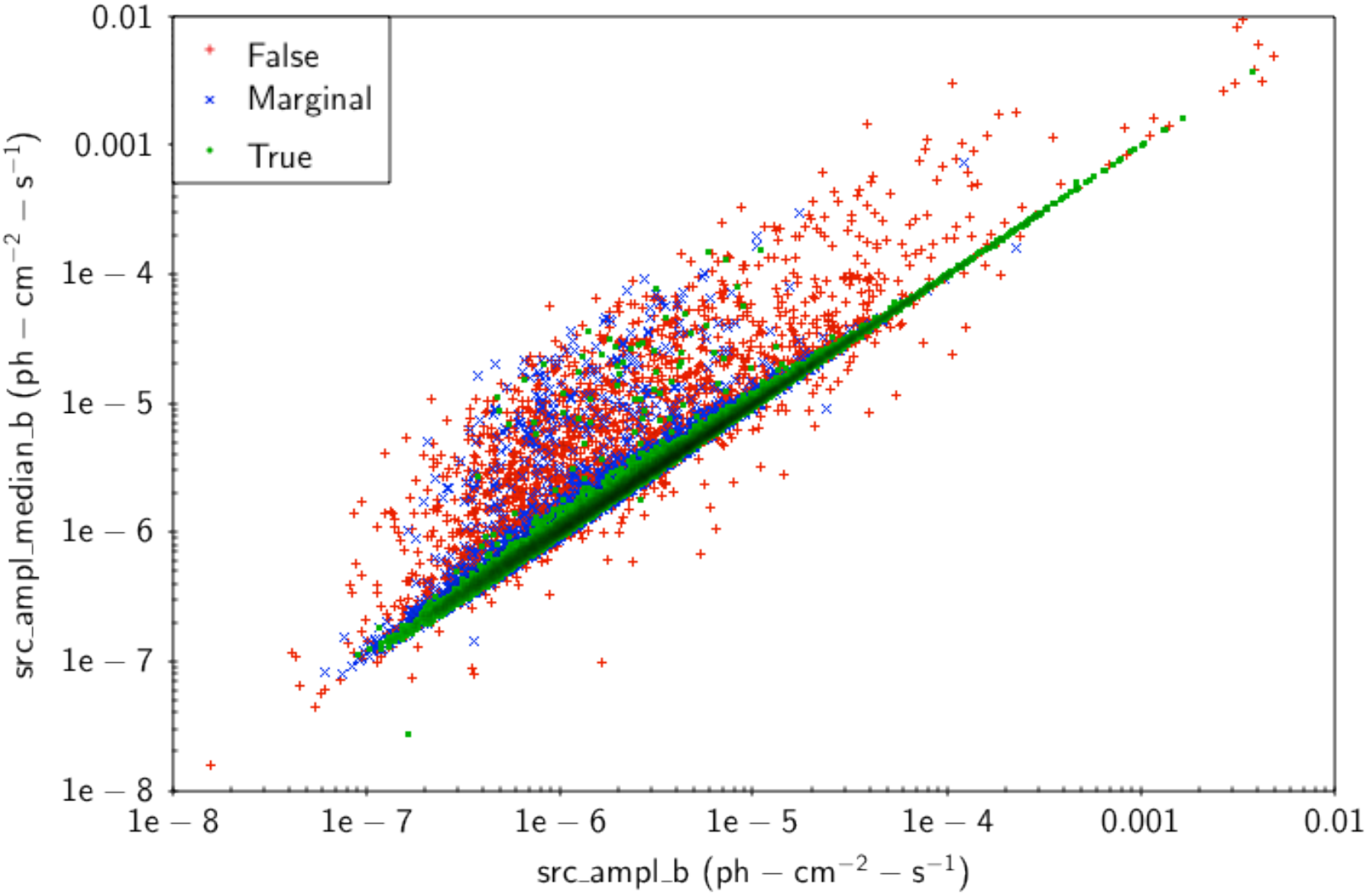





Comparison of the ACIS b band fitted amplitude confidence interval (upper confidence limit − lower confidence limit) and the corresponding photon flux confidence interval using the full Bayesian X-ray aperture photometry algorithm.

# Preliminary Photometric Properties (cont.)

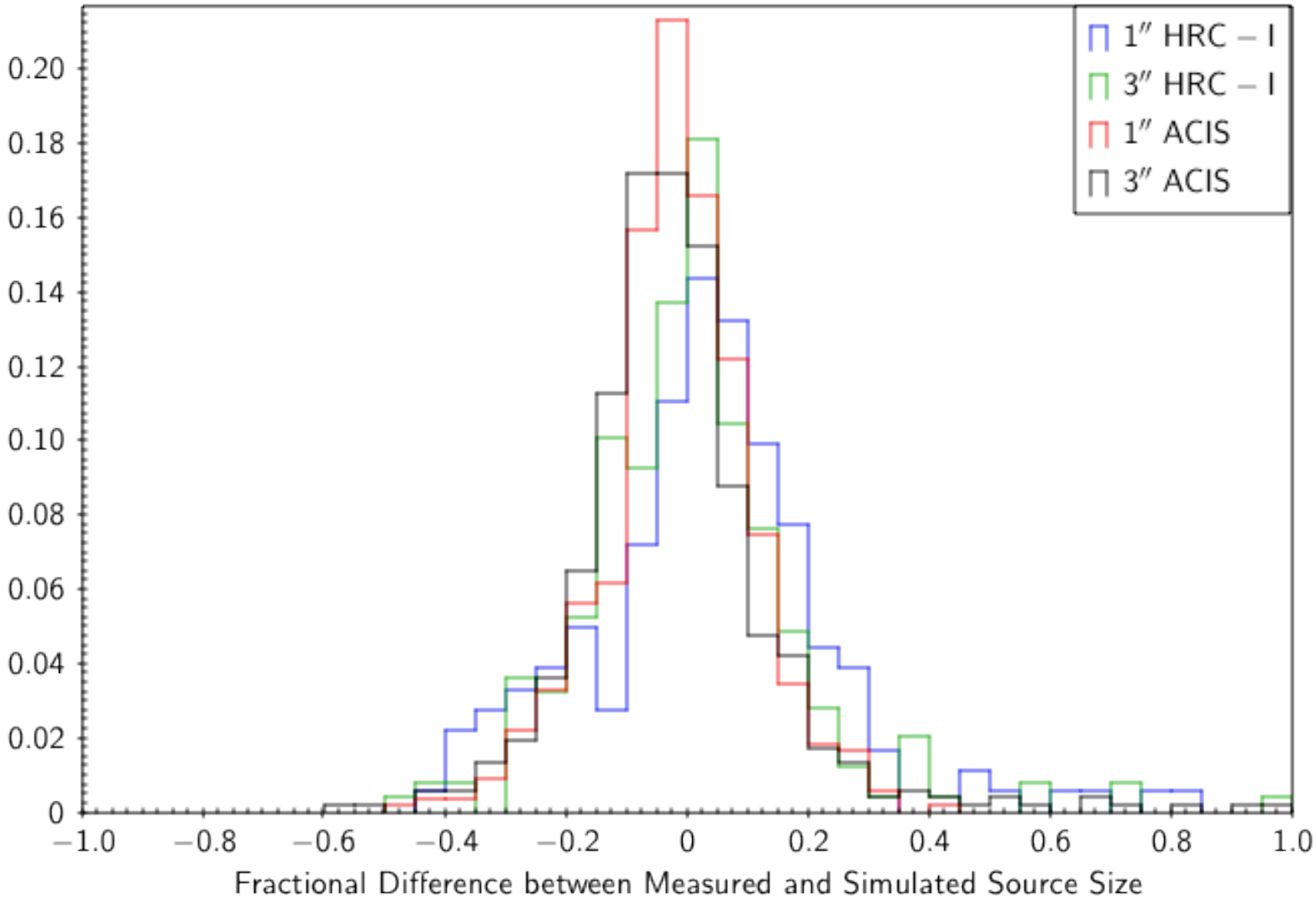





Comparison of the median of the sampled posterior distribution in detection amplitude (determined using `get_draws`) with the best-fit value for point sources. There is good agreement for detections with preliminary classifications of TRUE. Detections below this threshold show larger discrepancies.

# Preliminary Source Extent Properties

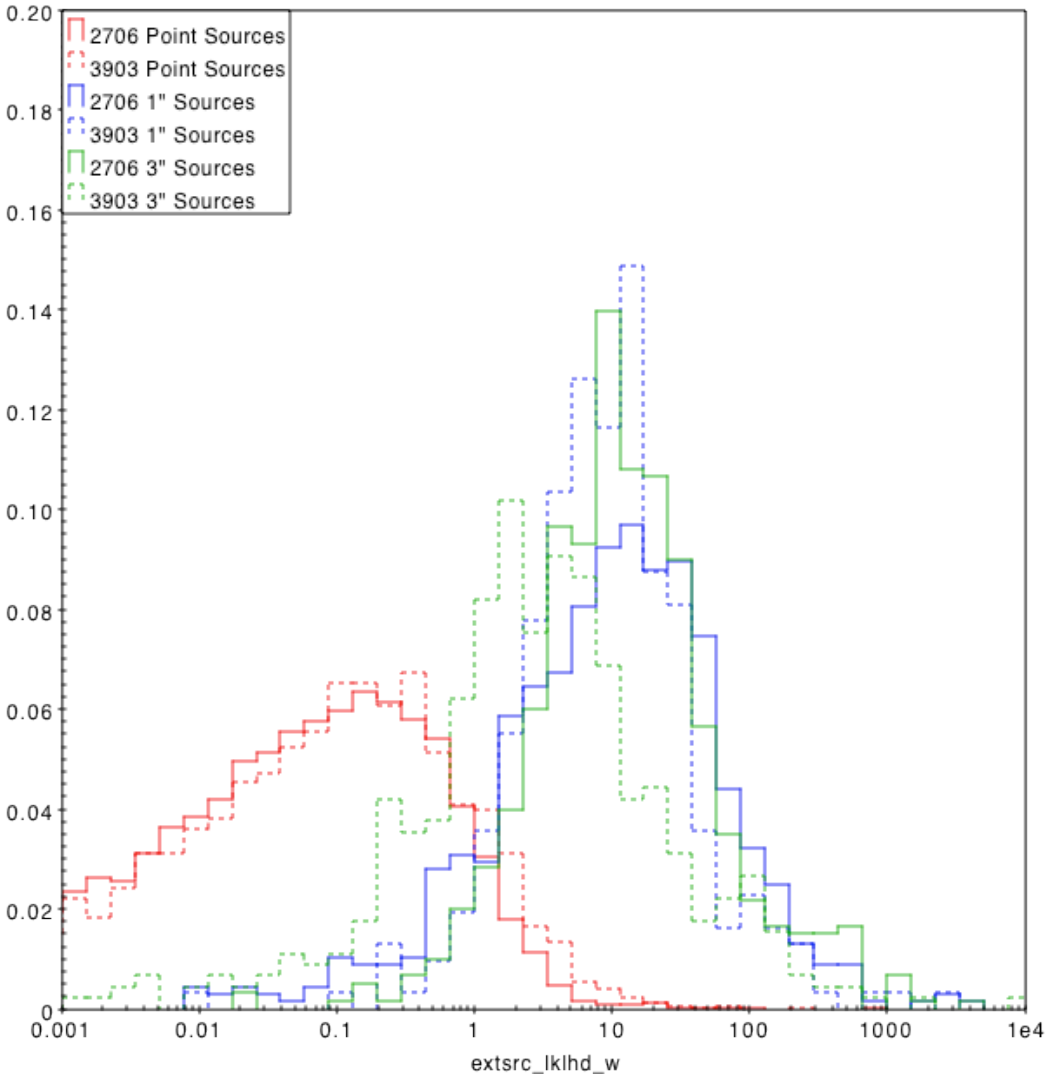





Comparison of the fitted sizes of non-point compact detections with actual simulated sizes. For both ACIS and HRC-I the agreement is good to ±20% in general. Compact extended sources are modeled using a rotated elliptical Gaussian convolved with the local PSF.

# Preliminary Source Extent Properties (cont.)






Comparison of the Simulations of 7 ks and 40 ks HRC-I observations indicates that a threshold of ~10 for extended source likelihood will yield a low probability of incorrectly classifying a point source as extended (same as for ACIS), but the detection efficiency for true extended sources is somewhat lower than for ACIS observations.

## Current and Planned CSC Release 2 Outreach

- Advertised at the "15 Years of Science with *Chandra*" symposium and the January 2015 AAS meeting (Boston)
- Planning a poster at the upcoming "High-Energy Large- and Medium-class Space Missions" workshop in June 2015 (Chicago)
- Planning a CSC Rel. 2 presentation on the NASA "Hyperwall" at the XXIX IAU General Assembly in August 2015 (Honolulu)
- Presenting an invited talk at ADASS XXV in October 2015 (Sydney)
- An article describing CSC Rel. 2 will be included in the upcoming 2015 *Chandra* Newsletter
- Availability of the preliminary detections list will be announced through the standard CXC channels (including the CSC website, *Chandra* Electronic Announcement, and Twitter [*@chandraxray*]), and will be submitted for inclusion in the electronic AAS News Digest
  - Additional announcements about Rel. 2 of the catalog will also be announced through these forums