Astronomical Data Analysis Software and Systems XVII **P6.8**
ASP Conference Series, Vol. XXX, 2008
J. Lewis, R. Argyle, P. Bunclarck, D. Evans, and E. Gonzales-Solares, eds.

The Chandra Level 3 Master Source Pipeline: Automated Source Correlation from Heterogeneous Observations

Roger Hain, Ian Evans, Janet Evans, Kenny Glotfelty

Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA

Abstract. The Chandra X-ray Center's Level 3 source catalog seeks to automatically detect sources and compute their properties. Since Chandra is a pointed mission and not a sky survey, different sky regions are observed for a different number of times at varying orientations, resolutions, and other heterogeneous conditions. While this provides an opportunity to collect data from a potentially large number of observing passes, it also creates challenges in determining the best way to combine different detection results for the best characterization of the detected sources. The Chandra Level 3 Master Source Pipeline correlates data from multiple observations by updating existing cataloged source information with new data from the same sky region as they become available. We present real and simulated examples of different overlapping source detections processed in the current version of the Level 3 Master Source Pipeline. We explain how they are resolved into entries in the master source database, and examine the challenges of computing source properties for the same source detected multiple times. Future enhancements are also discussed.

1. Overview

The Chandra Source Catalog will contain targeted and serendipitous X-ray sources obtained from Chandra Level 3 (L3) processing of public observations (Evans, 2006). Automated pipeline processing will calibrate data, detect sources, calculate source properties, and combine individual detections into a single catalog of master sources. The L3 Catalog will be continuously updated, and fully characterized snapshots will be released as the latest version.

Automated source detection is currently performed on a single observation interval, or "obi". Sources that are detected together within the same observation interval are referred to as the set of "obi" sources. The master pipeline collects source property information from obi sources and (1) determines which obi sources from different obis are really observations of the same master source, and (2) determines how to combine source properties (position, flux, etc.) from individual obi sources into properties for a single master source.

2. Issues Combining Data from Multiple Observations

The same X-ray source may be detected by multiple Chandra observations with different apparent sizes and position errors. The same sources may appear as point-like when observed on-axis in some observations, and more dispered by

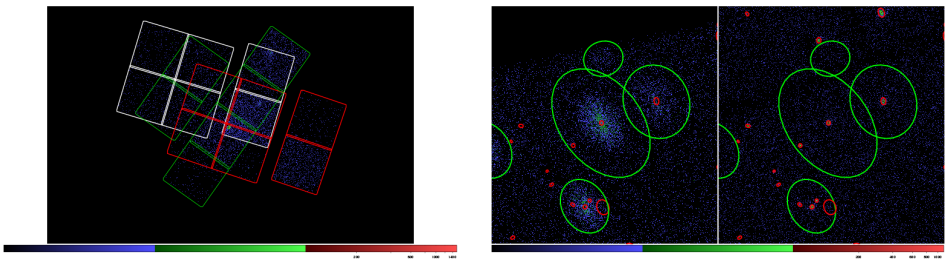

Figure 1. Left, Sky region coverage from M33 observation ids 786 (green), 1730 (red), 2023 (white). Right, two Rho Ophiucus observations. Left, obs id 635. Furthest right, obs id 637.

Chandra's non-linear point spread function (PSF) in other observations. Uncertainty in aspect reconstruction can also lead to offsets in source positions.

Figure 1 (left) shows the overlap of three different observations of the M33 region. The Chandra ACIS detector chips are outlined in different colors for each obi. This figure shows the potential for the same sources to be detected at different off-axis angles and PSFs from multiple observations.

Significant differences in source appearances between multiple observations are also shown in Figure 1 (right), where ellipses represent detections in the Rho Ophiucus region from two different observations. Events from observation id 635 are towards the left, and events from the same region as seen in observaion id 637 are shown on the rightmost part of the image. Source detections from 635 are shown on both images in green (larger ellipses), and detections from 637 are shown in red (smaller ellipses). The larger PSF size from 635 is one reason that data from multiple observations cannot simply be co-added prior to source detection. The various sizes and locations of the sources show a little of the complexity of determining how to map obi sources to master sources.

3. Updating the Database

In order to create instructions for the L3 database to update its set of master sources and their properties, the master pipeline generates a set of eight actions. The two primary actions are to MERGE an obi source with master source or to LINK an obi source to a master. In either case, the obi source is associated with the master source. In the case of a MERGE, the obi source can only merge with a single master source, and that master source's properties include the obi source's properties. For a LINK, the obi source is still associated with the master source, but its properties are not used in determining the master source's properties. Such a "confused" obi source can be linked to more than one master source.

For example, in Figure 1 (right), the rightmost larger green source will merge with the smaller, concentric red source to a single master source whose properties would be derived from both of these obi sources. However, the largest green source would not contribute its properties to any master sources, since it overlaps two (red) master sources and its properties cannot easily be assigned between them. A catalog user would know that the green obi source overlaps

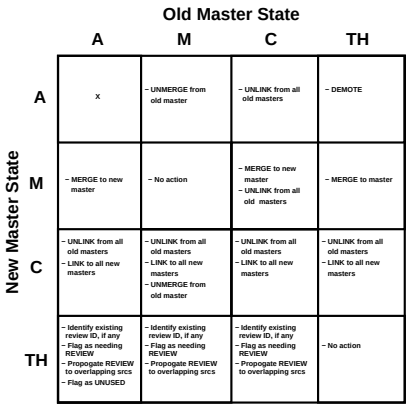

Figure 2. Database Actions. States: A-Absent; M-Merged; C-Confused; TH-Too Hard.

the two masters, but no properties from the green obi source would be used to derive the master source properties.

To determine what actions are necessary to update the database when a new obi is processed, the state of the master sources before and after the addition of the new obi sources is compared. In either the before or after state, obi sources will fall into the categories of merged; confused; too hard (no contribution to master sources) or be absent entirely. By comparing the old and new states of an obi source, a matrix of possible actions can be generated. (Figure 2).

Obis will not be processed through the master pipeline in any specific order, and it is a requirement that the final state of a master source be the same independent of the order in which individual obi sources are applied.

4. Transaction Lists

The database will have fixed releases with corresponding version numbers, but will also be available on a daily basis with the most recently processed obis. Since updates to the master sources can happen while the database is being accessed by users, some method is needed to prevent partial updates being made while source data are being retrieved. This is accomplished by implementing all the actions that may affect a master source at the same time in a single "transaction list".

Transaction lists are chosen so that any actions associated with a set of obi sources that are connected by overlaps are grouped together in a single list. This ensures that any action which may affect a master source is in a single file, and as long as the actions of that file are implemented together, no intermediate states will be found in the database.

An example of this is shown on the right side of Figure 3. Any actions which refer to the green sources, numbers 1-4, will all be grouped in the same transaction list since they are mutually overlapping. Actions affecting the red sources, numbers 5 and 6, are incorporated into a separate transaction list, since those sources do not overlap any of the green sources. This can also be

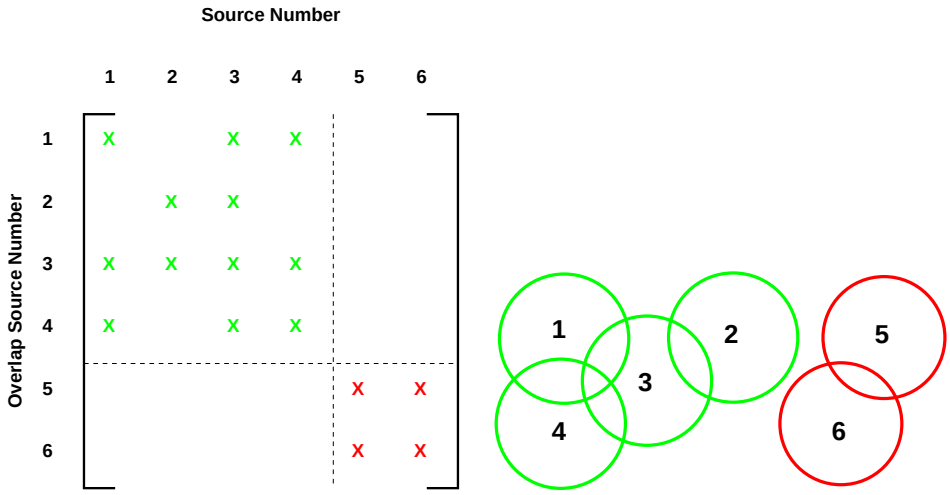

Figure 3. Transaction list grouping. Sample sources (right) and matrix representation (left).

represented by the symmetric block-diagonal matrix shown in on the left side of Figure 3. Having ordered the overlaps in this manner, each transaction list is formed from a block of the matrix.

5. Future Enhancements

Once obi sources are matched, their properties will be combined using methods which vary by property. Variability data for a master source may favor the obi source with the longest observation time, while position information may favor obi sources observed closest to the pointing axis, minimizing PSF dispersion. Trade-offs must be examined and implemented in the manner in which data are combined, as well. Some possible methods of joining data include using simple averages vs. error-weighted averages vs. weighting by time of observation.

Ultimately it is envisioned that the L3 pipeline will process individual obis together when master source properties are derived from multiple obi sources. This will allow combining data from multiple observations at a much earlier stage in processing compared to what will initially be done. This level of automation is expected at a later version of L3 processing software.

Acknowledgments. Support for development of the Level 3 Master Source Pipeline is provided by the National Aeronautics and Space Administration through the Chandra X-ray Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of the National Aeronautics and Space Administration under contract NAS8-03060.

References

Evans, I., et al. 2006, in ASP Conf. Ser. 351, ADASS XV, ed. C. Gabriel, C. Arviset, D. Ponz, & E. Solano (San Francisco: ASP), 65