Status of the Chandra Source Catalog Project

Over the past year, significant progress has been made in the development of the scientific algorithms, "Level 3" pipeline software, and associated infrastructure needed to generate the first release of the Chandra Source Catalog (CSC).

For those who are unfamiliar with the CSC project, the CSC will be the definitive catalog of all X-ray sources detected by the Chandra X-ray Observatory. The catalog will include fields of all Galactic latitudes, and sources from the entire detector field of view, although the first catalog release will be restricted to imaging data only (*i.e.*, no grating spectroscopy or ACIS continuous clocking data). For each detected X-ray source, the catalog will list the source position (accurate to ~1 arcsec), and a detailed set of source properties, which will ultimately include aperture and model fluxes in multiple bands to construct X-ray colors, source extent estimates, and spectral fits for bright sources. In addition to these traditional catalog elements, additional file-based data products will be included for each source individually from each observation in which a source is detected. These data products, which can be manipulated interactively by the user, include images, event lists, light curves, and spectra.

An estimate of the eventual size of the catalog can be obtained by projecting forwards from the sky coverage to date. Observations obtained during the first 6 years of the Chandra mission covered about 160 square degrees on the sky (including ~80 square degrees down to a flux level of 1.0×10^{-14} ergs cm^{-2} s^{-1}), with an estimated 150,000 detectable sources containing at least 10 counts. These numbers will continue to grow as the mission continues, with a 15 year prediction of ~400,000 sources distributed over ~400 square degrees, or ~1% of the sky.

Catalog Releases and User Interface

The catalog construction process will be carefully controlled, to ensure that each catalog release is a reliable, robust, and well characterized product to maximize the usefulness to the user community. A detailed statistical characterization of the source properties will accompany each release, including studies of astrometric and photometric accuracy, limiting sensitivity, completeness, and false source rates. Characterization requirements have been established, and a detailed characterization plan has been developed to meet these requirements.

User access to the catalog will be through a web-based browser interface in the first instance, with future support for a scripting language interface and virtual observatory workflows. High level requirements for the user interface have been developed, and detailed functional requirements are currently being evolved by studying a set of science "use cases."

Level 3 Pipelines, Software, and Algorithms

Prototype source detection and source properties Level 3 pipelines have been developed and tested. A test set of roughly 60 observations have been processed through the prototype pipelines. The results of these runs are providing valuable feedback aimed at improving the scientific integrity of the source analyses by identifying deficiencies in the current algorithms, and highlighting areas where pipeline operations could be improved. Scientific development of a new local background algorithm — one of the major missing components for the source detection pipeline that is essential to minimize the false source rate — is nearing completion, and is expected to be incorporated into the pipeline in the next few weeks. This algorithm combines a low-spatial frequency background component computed using a Poisson mean with a high spatial-frequency component that accurately identifies the ACIS readout streaks associated with bright sources in the field of view. A few weeks ago the CXC calibration team

provided a beta version of the SAOTrace ray-trace software that is capable of running directly on our Level 3 operations Beowulf cluster. This will alleviate a processing bottle-neck by allowing point spread function models to be generated directly in the source properties pipelines, taking advantage of the parallel processing capabilities of the Beowulf cluster, rather than processing serially on a Sun platform.

The first version of the science requirements for the remaining Level 3 pipeline were completed just before Christmas, and are presently being coded for testing. This "merge" pipeline matches each detected source in an observation with detections of the source in any other observations, and determines the "best estimates" of the source properties by combining data from all observations in which the source was detected. This process is complicated because of the spatially variable PSF (the source may be present at different off-axis angles in the observations), differing detector responses, and source time variability.

Hardware and Infrastructure

The algorithms required to build the CSC are very computationally intensive. To ensure that the catalog can be constructed in a reasonable time, a 15 node Linux Beowulf cluster has been installed for Level 3 pipeline operations, together with 3 Terabytes of scratch space. The automated processing infrastructure that runs the pipelines has been enhanced to manage pipeline processing on this cluster, and is being used to run the prototype pipelines. A recent upgrade to the third party Beowulf cluster software has addressed some early stability issues with the hardware, and the production platform now seems to be stable. Performance estimates based on the observed processing efficiency indicate that a production run of all the public observations to date will take of order 4 months using this cluster.

-- Ian Evans for the Chandra Source Catalog project team