Thoughts on Source Models

Ian N. Evans

Chandra Source Catalog Project Scientist

2019 May 08

Introduction

The *Chandra Source Catalog* (CSC) is the definitive catalog of serendipitous X-ray sources identified in publicly released imaging observations obtained by NASA's *Chandra* X-ray Observatory (CXO). The CSC is developed and published by the *Chandra* X-ray Center (CXC) and is supported by NASA contract NAS 8-3060 to the Smithsonian Astrophysical Observatory for operation of the CXC. Release 2.0 of the CSC, including properties for approximately 316,000 X-ray sources on the sky extracted from about 375,000 detections, will be released in the next few months.

One major aim of the CSC is to make available detailed estimates of the X-ray properties of astronomical sources detected by the CXO in a way that enables them to be immediately useful for scientific investigations by members of the (multi-wavelength) astronomical community who may be less familiar with the details of X-ray data and their reductions and analyses, while simultaneously maintaining the utility of the catalog for X-ray astronomy domain experts. Because of the relative complexity of X-ray data analyses, considerable thought has been expended over the years as to how to maximize the scientific utility of the catalog.

Detections vs. Sources

A key aspect of the CSC since its original inception is that we differentiate between *detections* and *sources*. *Detections* are what we observe on the detector, and they have associated detection properties (i.e., *measurements*) such as a position on the detector, photon event time of arrival (*Chandra*'s X-ray instruments detect individual photons), measured event pulse height, etc. *Sources* are our *best interpretation* of what the detections imply are the physical entities on the sky that emit the photons that we detect, and they have associated source properties (e.g., a position in world coordinates, a light-curve, an energy spectrum.)

Differentiating between detections and sources is not just relevant to the CSC. Indeed, at any wavelength what is detected may not represent what is present on the sky because of observational limitations that exist in any regime (e.g., instrumental PSF/beam/atmospheric seeing), so one should in all cases distinguish between the two concepts.

We do not imply that the source properties are associated with a *single* physical source of the photons on the sky for two reasons. First, unambiguously disentangling some properties of the source from properties of the intervening column may not be possible (e.g., we may conclude that a source has an energy spectrum that is well represented by an absorbed power-law, but the fraction of the absorbing column density that is intrinsic to the source, and the fraction that is due to the intervening column, may not be well determined). Second, additional measurements obtained at a later time may provide new data (such as higher resolution imaging) that show our interpretation of the detections to sources is incorrect.

The important point here is that there is a mapping from the detection properties to the source properties. That mapping may be relatively simple (e.g., world coordinate system on a linear image), or quite complex (e.g., measured pulse-height spectrum of an X-ray source, which requires a response matrix function [RMF], auxiliary response file [ARF], choice of spectral model, and forward fitting to estimate robustly).

An even more important point is that the mappings between detections and sources can be many-to-many.

For example, there may be multiple detections whose measured properties may be combined to provide estimates of source properties. In many cases those measurements may only provide estimates of the source properties at specific epochs since the source properties may change over time (e.g., position due to proper motion or parallax, flux or spectrum due to intrinsic temporal variability). Sometimes measurements will vary due to the measurement conditions rather than intrinsic source property changes (e.g., observed magnitude due to variable atmospheric extinction). Ideally, the latter will be corrected as part of the calibration process, but the measurement — and therefore the estimated source property — may have uncertainties with differing characteristics.

In the converse direction, a single detection may map to multiple sources, and so the detection properties may constrain source properties for multiple sources. This commonly happens for *Chandra*, where the point spread function (PSF) size increases by a factor of ~100 from the center of the field to the edge of the field (the PSF size is also energy-dependent). A low spatial resolution detection may provide a flux upper limit at a specific epoch for multiple sources resolved at higher spatial resolution but may not be useful to constrain other source properties such as position.

Temporal Variability

Since sources may be temporally variable (many compact sources observed in X-ray are temporally variable), measurement epoch is an important quantity as mentioned above. The CSC handles temporal variability by providing source properties (in addition to detection properties) for each measurement epoch. However, simply providing per-epoch properties is not sufficient for two reasons. First, in many cases the astronomer is not interested in source temporal variability, but rather wants some “canonical” estimate of the source properties. Second (and perhaps more important), source properties can often be better estimated by combining multiple measurements (i.e., higher S/N).

The CSC defines “best estimate” values for many time-dependent source properties that may provide “typical” or “most useful” values. The definition of what constitutes the “best estimate" is most appropriately determined by domain experts. As an example, the best estimate spectrum for a source that flares infrequently above the background level may be extracted from measurements obtained when the source is flaring. There may also be a meaningful method of grouping together multiple measurements obtained when the source is in the same state. For example, the CSC performs a multi-band Bayesian Blocks grouping of detection fluxes to identify groups of detections where the source is apparently in the same state (i.e., same fluxes and colors) and records source properties for each grouping. The "best estimates" for these source properties are those extracted from the grouping with the largest exposure time. The CSC also provides straight average values for certain properties (mostly aperture photometry) for comparison with other catalogs.

Including support for upper limits for certain source and detection properties (notably aperture photometry) can add significant value, especially for highly variable sources. The CSC routinely identifies *non*-detections, where the source location overlaps a given observation for which the source is not detected, and provides photometric upper limits based on the size of the local PSF.

As a general comment, aperture photometry properties need to have definitions of the detection (and, in many cases, detection background) regions, aperture fractions, and so on.

Uncertainties

Measurements have intrinsic uncertainties, and these detection uncertainties contribute to the uncertainty in the knowledge of the source properties. Their representations must be flexible to adequately provide a realistic knowledge of the uncertainties to the end user. While simple symmetric Gaussian errors are easily reported, uncertainties are often not Gaussian and, in many cases, Gaussian errors are not good representations of the real uncertainties. Asymmetric confidence intervals (possibly specified for multiple confidence levels) are more robust representations that may be appropriate in most cases. However, providing actual probability density functions (PDFs) for values being reported is a robust and relatively unambiguous representation. PDFs can also be used to define confidence intervals if a simpler representation is also required. The use of PDFs is particularly appropriate in the Poisson limit, and the CSC provides PDFs for aperture photometry properties and properties such as hardness ratios derived from aperture photometry properties.

Uncertainties along different axes may be strongly correlated (e.g., as is the case for imaging spectroscopy where the uncertainty in the source position parallel to the dispersion axis translates directly into the spectral coordinate uncertainty), and this information is essential to fully characterize the uncertainties for the end user.

Where the internal and external errors have significantly different scales, one should differentiate between the two where appropriate. As an example, if the internal astrometric uncertainty derived from a single image is smaller than the external astrometric uncertainty in a defined reference frame (such as ICRS), the relative source-to-source astrometry will be better defined than absolute source positions — and this is useful information for the end user.

Capturing Assumptions

Some source and detection properties are dependent on the assumptions used to derive them. For example, X-ray flux values computed using forward fitting are dependent on choice of spectral model (e.g., absorbed power law), so there has to be a way to link the derived properties to a description of the assumed model. Similarly, many measurements are dependent on calibration data so the same comment applies. While calibration data may seem to be quite straightforward to those used to typical optical imaging and low/medium resolution spectroscopy, they become much more critical when pushing the envelope; the sophisticated end user often wants to understand the statistics of the calibrations in order to trust the reported measurement uncertainties.

Measurement responses such as (n-D) PSFs are often not Gaussian and may not be well represented by a combination of analytic functions, so there should be a way to represent the responses (e.g., PSF's may vary significantly across the field, both in shape and dimension, as is the case for *Chandra*). Source properties may of necessity be extracted from (multiple) measurements with differing responses, so this applies on a per-property (or per-group of properties) basis.

Ensembles of sources (e.g., a catalog) and/or measurements may have specific statistical properties that should be captured in the source model.

In many cases linking the source model to actual data products may be the only reasonable way to address describing the more complex aspects of the source description. Model assumptions, calibrations, PDFs, light-curves, spectra, and so on may all be best represented as linked data products.

Conclusions

Although some of the thoughts laid out herein may seem overly complex, they are really reflecting the complexity of nature and our field of endeavor. We have had great success with the CSC user community by blending simpler representations of source and detection properties where feasible (e.g., detections with excellent S/N, users asking research questions that do not require the ultimate accuracy) with detailed and robust representations that can be used reliably when maximum fidelity and a clear understanding of uncertainties is required. Had we not included the latter, entire fields of research with the CSC and classes of users performing this research would not be possible.

I strongly urge the community to consider carefully optical and other wavebands, both the simple and complex, when constructing source models so that they can be of maximum use to scientists with varying degrees of expertise from newcomers to domain experts.