# Chandra Source Catalog (CAT2) Spectra and Spectral Fitting Specifications

*Michael L. McCollough, Aneta Siemiginowska, and Doug Burke*
*October 20, 2015*

We have divided the specifications for the spectra creation and spectral fitting into three stages in the pipeline. One that describes the creation of spectra, a second that will apply to overall spectral fitting, and then a third for specific considerations for the joint spectral fitting.

## I. Creation of Spectra and Related Files:

(1) General Spectral Files: For version 2 of the catalog (CAT2) we are proposing that we start with catalog version 1.1 (CAT1) of the spectral creation. This will include creating the source spectra, background spectra, RMF (for source region), and ARF files (for source region). For the ARF files we will include the correction for the finite aperture (i.e. arfcorr).

(2) Background regions used for spectral fits: We are currently proposing to use the CAT1 method for creating the background spectra for each source. We will evaluate this once the new pipeline starts producing spectra. (It will be important to note any differences in the background and source regions from the regions provided from source detection. This information will likely to be stored either in the spectral files or source region file.)

(3) Sources to Include: We plan to create the spectral files for all sources that are returned as TRUE and MARGINAL from MLE.

## II. Individual Spectral Fits:

(1) Basic Spectral Fitting: We are proposing that we start with the CAT1 version of the spectra fitting pipeline. This spectral fitting will be done on individual obsids with no merging (same as in CAT1). This should be written using Sherpa for the spectral fitting.

(2) Spectral Models: We have examined proposed spectral models and propose the following:

(a) We recommend that we should go with four spectral models: (1) Absorbed Power-Law (xsphabs*powlaw1d); (2) Absorbed Blackbody (xsphabs*xsbbody); (3) Absorbed Bremsstrahlung (xsphabs*xsbremss); and (4) Absorbed APEC (xsphabs*xsapec).

(b) For the APEC model there are four spectral parameters that can be varied (T, nH, z, abundance). We are proposing that the pipeline should allow four to be varied. But in the default fitting we will only allow T and nH to be varied. The parameters z and abundance should be frozen and defaults vaults will be provided. This approach is similar to what was done for the XMM spectral fit catalog.

(c) We will use the CAT1 method to set the initial value of the nH spectral fit parameter.

(d) We recommend we should use a two tiered fitting process (same as was done for CAT1). That is below a certain number of net counts (see below) we will use fixed values for the spectral parameters (to be provided) and do a fit by varying the amplitudes. Above or at that net number of counts we should do a full fit with the spectral parameters being allowed to vary.

(e) It is desirable to be able to add at a later time a third tier in which spectra with an even higher count rate could be fit with more complex models. But none of this should be considered a lien on the overall requirements. This last tier may be done as a separate user application outside of the pipeline.

(3) Spectral fitting constraints: We are proposing the following constraints on the individual spectral fits:

(a) Number of Counts per Spectrum: We are currently recommending the CAT1 values of 150 net counts in the 0.5-7.0 keV energy range (this value should be a parameter). After we have reviewed the new pipeline spectra we may have a new value for this parameter.

(b) Spectral Binning: We are currently recommending the CAT1 values of 16 counts per channel bin (this value should be a parameter). After we have reviewed the new pipeline spectra we may have a new value for this parameter.

(c) Spectral Boundaries: We are currently evaluating how we want to determine the spectral boundaries. A standard grouping command will be initial provided.

(d) Optimization Method: We recommend using the "Nelder-Mead algorithm (Simplex)". We plan on adjusting the normalization of the model prior to fitting to speed up the fitting process.

(e) Fit Statistic: We recommend using "chi2datavar".

(f) Flux Calculation: We recommend using "sample_flux" for determining the absorbed and unabsorbed fluxes. The number of iterations used in "sample_flux" will be a parameter which can be adjusted if necessary.

(g) Confidence Calculations: We recommend using "conf()" for determining the confidence levels (1 sigma).

(h) Fixed Spectral Parameters: For spectra that too few counts (see (a)) we will use fixed spectral parameters. For the Power-Law and Blackbody we will use the median parameter values of the previous good fits (reduced statistic $\leq 1.25$) from CAT1 fits for the spectra parameters (nH, gamma, kT). For the Bremsstrahlung and Absorbed APEC we will do fits to a sample of the CAT1 data. In all cases we will just fit the normalization of these models.

(4) Final Products (Individual Spectra): We are recommending have the same products as was provide (per model) in CAT1:

(a) The best fit values to the spectral parameters (for each model).

(b) The confidence levels to the spectral parameters (for each model).

(c) The amplitudes and fluxes (absorbed and unabsorbed) for the fits for the various spectra.

(d) The flux upper and lower limits (determined from the 1 sigma confidence levels from "sample_flux").

(e) The number of degrees of freedom (dof) for each model.

(f) The goodness of fit for the various models.

(5) Using the Cash statistic: The possibility of fitting the unbinned data using a Cash statistic is being considered. Determining the background spectra is likely to be an important issue for this approach. Progress has been made using this method with the development of the "wstat" which will be added Sherpa in the near future. We may still need to bin the data (at least 1 count per bin). But for the present we plan on going with the grouping and background subtraction approach that was used in CAT1. If this approach does work it will be easy to modify the proposed pipeline to use this method.

## III. Joint Spectral Fits

(1) Joint fitting: We recommend that we do a joint spectral fits to the data (instead of a combined spectral fit). We currently have a shell script that for a given "master ID" (MID) creates a short Sherpa script which can be used within Sherpa to perform the joint spectral fit. Copies of these of the shell script and examples of the Sherpa script and output are provided with these specifications.

(2) Spectral fitting constraints: We are proposing the following constraints on the joint spectral fits:

(a) Number of Counts in the Joint Spectrum: We are currently recommending that we use the same number of counts ( for a combination of all the spectra being fit) as for the individual case. This would be a values of 150 net counts in the 0.5-7.0 keV energy range (this value should be a parameter). After we have reviewed the new pipeline spectra we may have a new value for this parameter.

(b) Spectral Binning: We are currently recommending that each of the individual spectral used for joint fitting have a values of 16 counts per channel bin (this value should be a parameter). After we have reviewed the new pipeline spectra we may have a new value for this parameter.

(c) Spectral Boundaries: We are currently evaluating how we want to determine the spectral boundaries. A standard grouping command for each of the individual spectra will be initial provided.

(d) Optimization Method: Same as II.3.d.

(e) Fit Statistic: Same as II.3.e.

(f) Flux Calculation: We recommend using "sample_flux" for determining the absorbed and unabsorbed fluxes. The number of iterations used in "sample_flux" will be a parameter which can be adjusted if necessary. Under our current plan we will be calculating the flux for each of the Bayesian blocks provided (see below). For each block the normalizations for the spectra that are included in the block will be linked.

(g) Confidence Calculations: Same as II.3.g.

(h) Fixed Spectral Parameters: Same as II.3.h

(i) Spectral Parameters Linking: In all of the joint fits we will link all of the parameters including the normalizations.

(3) "datastack" Updates: A recommended change to the provided script is to update the script to use the various "datastack" functions that are available in Sherpa.

(4) Spectra to Include in Joint Fits: Due to the source variations the individual spectra used in the joint fit may need to be selected based on some source behavior. We have considered several methods of determining the spectral to be used (Hardness Ratios, Goodness of Fit, and Bayesian Blocks) Of these methods the one worked the best was the Bayesian Blocks. In this approach the flux ordered groups (by obsids/spectra) are provided from a Bayesian Blocks code (from Frank Primini) . For each of these blocks a joint spectral fit is made.

(5) Additional Considerations: Some additional matters that need to be considered for joint fits are the following:

(a) Tracking the spectra used in the joint fit: We need to track that spectra that are used for the joint fit. This will be tied to the results from the Bayesian Blocks analysis.

(b) Bayesian Blocks: All of the block spectral fit information will be keep in a separate data file (a new data product) for each of the MID for which there is a joint spectral fit.

(c) Best Bayesian Block: For each of the sources a single block (provided from the Bayesian Blocks analysis) will be chosen as being the most representative of the source and the spectral fit parameters will be put into the databases.

(6) Final Products: In addition to the things listed in (5) we are recommending to have the same products for the joint fits that are provide for the individual spectra fits:

(a) The best fit values to the spectral parameters (for each model).

(b) The confidence levels to the spectral parameters (for each model).

(c) The amplitudes and fluxes of the total fit for the various spectra.

(d) The flux upper and lower limits (determined from the 1 sigma confidence levels from "sample_flux").

(e) The number of degrees of freedom (dof) for each model.

(f) The goodness of fit for the various models.

(7) Using the Cash statistic: Same as II.5.