MIT Kavli Institute **Chandra X-Ray Center**

MEMORANDUM
July 10, 2006

To: file

From: David Huenemoerder

Subject: Adding Counts Spectra

Revision: 1

URL: TBD

File: ~dph/h3/CXC/TG/Scripts_devel/Add_pha/Doc/technote.tex

1 Adding PHA Counts Spectra

1.1 Why, or Why Not Combine PHA Counts Spectra?

Combining multiple observations, or even grating orders from a single observation has certain risks. To quote from the sage old advice found in the `ftools` help on `addspec`[1]:

> Users are WARNED that the addition of different PHA datasets is a dangerous exercise, and give [*sic*] rise to very misleading results from subsequent analysis. The two occasions where the addition of PHA datasets *may* be valid is [*sic*] A) to combine PHAs from different observations made by the same detector of a given source, B) to combine PHAs made at the same time by different but "identical" detectors. In case A), incorrect or misleading results will most likely be produced if the source exhibits variability (especially spectral variability), and could easily be produced if the response of the detector changes dramatically between the epochs of the observation. In case B), incorrect or misleading results could be produced if the detectors are not actually "identical", and their reponses [*sic*] differ significantly. In both cases, errors could also be introduced by the datasets having significantly different backgrounds (rates or spectra). Users are STRONGLY urged to consider these facts before using the task.

[1] ftp://heasarc.gsfc.nasa.gov/software/lheasoft/release/doc/fhelp/addspec.txt

The warning is even more apropos to *Chandra* , since the detectors are non-uniform, the PSF varies with off-axis angle, we have multiple-resolution instruments, response is time-variable, and rigorous CCD pileup modeling *requires* the integrity of individual observations.

Furthermore, we have analysis packages which support the dynamic assignment of data and responses, dynamic summing of data with proper folding of models through multiple responses, dynamic regridding, and efficient evaluation of models over multiple grids. It is far easier and more robust to manage the metadata of individual datasets during an interactive session than to correctly, consistently, and completely define merged equivalent quantities.

Nevertheless, there are situations when summed spectra are convenient, necessary, or inevitable. Consider the LETGS spectrum: it is a multi-order, multi-resolution sum of orders as delivered by the instrument. We can compute responses as if orders were distinct and combine them in analysis to model their sum. HETGS spectra are also, in a sense, summed: the dither moves some wavelengths between CCDs. We produce a binned spectrum across these transitions, and sum the ARFs to produce a commensurate response. For low signal-to-noise data, the statistical uncertainties may be far larger than any uncertainties incurred by summing counts from different responses (in which case summing is a convenience, not a necessity, since the sum could still be done dynamically during analysis).

It is easier to understand why one should *not* combine spectra than to justify combination: Wouldn't you rather have order-sorted LETGS spectra? If a source is variable, do you split the data into multiple spectra? If you have high- and low-resolution data, do you keep them separate? But given the subjective nature of analysis, dependence upon data quality, scientific goals, and collaborations using diverse analysis packages, we will pursue methods for producing combined data in file-based formats, with properly computed meta-data.

Users should be STRONGLY *urged to consider these facts* before summing spectra!

1.2 Mathematical Formalism

For some observed counts spectrum, i, extracted from some some source region of area Ω_s, we can express the detected counts, C, as a function of channel, h, in terms of the source model, S [photons $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$], effective area A ("ARF") [cm^2 counts phot^{-1}], and exposure, t [s]:

$$C_i(h) = S\,A_i\,t_i. \tag{1}$$

For simplicity, we have omitted the explicit spectral redistribution integrals; SA should be considered to be a shorthand for the operator, $\int dE\,R(E,h)A(E)S(E)$. The following is valid if the spectral redistribution functions are effectively identical for each of the summed observations, and the source spectrum is constant. We define the mean exposure time as

$$T = \frac{1}{N}\sum_{i=1}^{N} t_i, \tag{2}$$

and the exposure-weighted ARF as

$$A = \frac{1}{T} \sum_i t_i A_i. \tag{3}$$

Then the sum of counts can be written as:

$$C(h) = \sum_i C_i(h) = S\, A\, T. \tag{4}$$

The ARF-summing tool, `dmarfadd`, uses these definitions for T and A. Though `dmarfadd` was designed for use with the individual chip ARF's made by `mkgarf` for gratings, it is valid for summing multiple orders or observations with or without gratings, given the understanding that T is not the cumulative exposure, but that AT is the consistent product of exposure and effective area. Hence, we can simply sum the counts from multiple commensurate spectra, as long as we use `dmarfadd` to compute the corresponding ARF, and also require that the ARF's exposure take precedence during analysis.

1.3 Background Spectra

We can apply similar arguments for background spectra. For background counts spectrum, i, extracted from a background region of area Ω_b (detector or sky area, or for grating spectra, 1-dimensional extraction region width), we can express the detected counts, B, as a function of channel, h, in terms of the background model, b [counts area^{-1} s^{-1}], background scaling factor K ("`BACKSCAL`") [area], and exposure, t [s]:

$$B_i(h) = b\, K_i\, t_i. \tag{5}$$

Note the similarity in form to Equation 1 (though there is no implicit redistribution if we are only considering internal background). Here, K_i is analogous to A_i (we use K_i instead of Ω_b, the actual area, because `BACKSCAL` is usually expressed as a relative area by dividing by some reference area). Hence, we use the same definition of T, but in addition define the exposure-weighted `BACKSCAL` as

$$K = \frac{1}{T} \sum_i t_i K_i. \tag{6}$$

Then the sum of background counts can be written as:

$$B(h) = \sum_i B_i(h) = b\, K\, T. \tag{7}$$

We must now relate this background to the background counts, B_s, predicted in the source region, Ω_s, with corresponding K_{si}

$$B_{si}(h) = b\, K_{si}\, t_i. \tag{8}$$

with the exposure-weighted source-region `BACKSCAL`

$$K_s = \frac{1}{T} \sum_i t_i K_{si}. \tag{9}$$

These lead again to the same form,

$$B_s(h) = \sum_i B_{si}(h) = b\, K_s\, T. \tag{10}$$

Finally, since b is the same in Equations 7 and 10, and recognizing the T can be different for the source and background observations, we can trivially derive

$$B_s(h) = B(h)\, \frac{T_s}{T_b}\, \frac{K_s}{K_b}. \tag{11}$$

This form is identical to that used for analysis of a single source spectrum and single background spectrum, in which we scale the background spectrum by the ratio of exposures and backscales. This means that we can use the summed source and summed background counts spectra in conjunction with the mean exposure, the exposure-weighted backscale, and the exposure-weighted ARF to consistently and correctly model combined spectra in the same way we do individual spectra (under *a priori* assumption that combining the data is valid). We must be aware, however, that the important quantities are KT and AT, since the exposures are the mean values and the backscale and area are cumulative, an exposure-weighted sum. The analogous caveat applies here, that the "source" spectrum, b, is assumed to be constant.