

# Procedures for co-adding spectral files

A. Vikhlinin

## Abstract

This memos reviews the issues relevant for co-adding spectral files and related spectral calibration. We consider the case when two or more observations of (spectrally) the same region are to be co-added but the observations have different exposure or cover different area.

## 1 XSPEC spectral files and related calibration

XSPEC-type spectral files record the observed source spectrum as the number of photons (or counts rate) per channel, and also contain in the header pointers to the relevant calibration data. Here is an example of the header file (only relevant keywords are shown):

```

HDUCLAS2= 'TOTAL'      / Gross PHA Spectrum (source + bkgd)
HDUCLAS3= 'COUNT'     / PHA data stored as Counts (not count/s)
EXPOSURE=      15149.700195 / exposure (in seconds)
AREASCAL=      1.000000 / area scaling factor
BACKFILE= 'tprof_5289_reg6_sfi.bg'
BACKSCAL=      1.07365076257837
CORRFILE= 'tprof_5289_reg6_sfi.rdt'
CORRSCAL=      1
RESPFILE= 'tprof_5289_reg6_sfi.rmf' / response file
ANCRFILE= 'tprof_5289_reg6_sfi.arf' / ARF
POISSERR=      T / Poissonian errors to be assumed
STAT_ERR=      0 / no statistical error specified

```

1. Since HDUCLAS3='COUNT', the observed flux in each channel will be divided by EXPOSURE
2. **Background.** HDUCLAS2='TOTAL' signals that the spectrum needs to be background-subtracted. The background file is pointed to by the BACKFILE keyword; its structure is identical to the source spectrum.

The background is subtracted using the following procedure. Let

$$\begin{aligned}
 c_{s,i} &= \text{Observed flux [cnt] in channel } i \text{ in source file} \\
 t_s &= \text{Source exposure given by the EXPOSURE keywords} \\
 B_s &= \text{The value of the BACKSCAL in the source file,}
 \end{aligned}
 \tag{1}$$

and  $c_{b,i}$ ,  $t_b$ , and  $B_b$  are corresponding quantities from the background file. XSPEC calculates the source counts per channel as

$$c_i = c_{s,i} - \frac{t_s B_s}{t_b B_b} \times c_{b,i}.
 \tag{2}$$

If BACKSCAL = 1 in the background file, then BACKSCAL in the source file gives the factor by which the background spectrum needs to be multiplied. For example, if the ACIS count rate in the 10–12 keV band in your data is 4% higher than that in the blank-field dataset, you need BACKSCAL = 1.04 in the source spectrum.

3. **Correction file.** In addition to the background subtraction, a “correction file” pointed to by the CORRFILE keyword can be applied to the data. If  $C$  is the value of the CORRSCAL in the source file and its exposure is  $t_c$ , the source spectrum is

$$c_i = c_i - C \times \frac{t_s}{t_c} \times c_{c,i},
 \tag{3}$$

where  $c_{c,i}$  is the correction spectrum [cnt/chan].

4. AREASCAL — I do not use this keyword and always set it to one.
5. **Calculation of statistical uncertainties.** If POISSERR = T, XSPEC calculates statistical errorbars (after grouping) as

$$e_{s,i} = \sqrt{\max(c_{s,i}, 1)}, \quad (4)$$

(and identical relations for the uncertainties in the background and correction files), which are then propagated in equations 2–3.

If POISSERR = F, statistical uncertainties must be provided as the values of the STAT\_ERR column.

6. **Response calibration.** Finally, the effective area and spectral response files are pointed to by the ANCRFILE and RESPFILE keywords, respectively. These files tabulate the effective area of the telescope,  $A(E)$ , and redistribution function of the detector,  $R(i, E)$ . If the model spectrum is  $f(E)$ , the expected observed spectrum is

$$m_i = \int f(E) A(E) R(i, E) dE \quad (5)$$

## 2 Adding two spectra

### 2.1 Mathematics

When two spectra are to be added, the most optimal procedure from the statistical point of view is to add counts (not count rates),

$$c_{s,i} = c_{s,i}^{(1)} + c_{s,i}^{(2)} \quad (6)$$

**Adding exposures** If we add two observations of exactly the same region, clearly we should add exposure times,

$$t_s = t_s^{(1)} + t_s^{(2)}. \quad (7)$$

If we have two regions from the same observations, exposure obviously must be kept identical to either first or second spectrum,

$$t_s = t_s^{(1)}. \quad (8)$$

If we have different regions in different observations, it is unclear what to do with the exposure but it has little effect on the final spectrum except for the overall normalization. Below, we just assume that the exposure assigned to the co-added spectrum is  $t_s$ .

#### 2.1.1 Adding background and correction files

It is clear from eq. (6) that the background must be subtracted from co-added spectrum as

$$c_i = c_{s,i} - \frac{t_s^{(1)} B_s^{(1)}}{t_b^{(1)} B_b^{(1)}} \times c_{b,i}^{(1)} - \frac{t_s^{(2)} B_s^{(2)}}{t_b^{(2)} B_b^{(2)}} \times c_{b,i}^{(2)} \quad (9)$$

but we would like to produce a co-added background spectrum so that we can use the usual expression (eq. 2),

$$c_i = c_{s,i} - \frac{t_s B_s}{t_b B_b} \times c_{b,i}. \quad (10)$$

From (9) and (10) we have

$$\frac{t_s B_s}{t_b B_b} \times c_{b,i} = \frac{t_s^{(1)} B_s^{(1)}}{t_b^{(1)} B_b^{(1)}} \times c_{b,i}^{(1)} + \frac{t_s^{(2)} B_s^{(2)}}{t_b^{(2)} B_b^{(2)}} \times c_{b,i}^{(2)} \quad (11)$$

We have some freedom in assigning the EXPOSURE and BACKSCAL keywords in the coadded background spectrum. A natural choice seems to be to assign

$$B_b = B_s = 1 \quad \text{and} \quad t_b = 1. \quad (12)$$

With these definitions, the co-added background spectrum is

$$c_{b,i} = \frac{1}{t_s} \times \left[ \frac{t_s^{(1)} B_s^{(1)}}{t_b^{(1)} B_b^{(1)}} \times c_{b,i}^{(1)} + \frac{t_s^{(2)} B_s^{(2)}}{t_b^{(2)} B_b^{(2)}} \times c_{b,i}^{(2)} \right] \quad (13)$$

Identical considerations lead to the following procedure for adding correction files:

$$t_c = 1 \quad \text{and} \quad C = 1, \quad (14)$$

$$c_{c,i} = \frac{1}{t_s} \left[ C^{(1)} \times \frac{t_s^{(1)}}{t_c^{(1)}} \times c_{c,i}^{(1)} + C^{(2)} \times \frac{t_s^{(2)}}{t_c^{(2)}} \times c_{c,i}^{(2)} \right]. \quad (15)$$

**Error calculations** While procedure for co-adding the source spectra (eq. 6) obviously preserves Poisson errors, those for background or correction (eq. 13 and 15) do not. The best solution is to set `POISSERR = F` in the background and correction files and compute the `STAT_ERR` column as

$$e_i = \sqrt{c_i} \quad (16)$$

For some channels,  $e_i$  will be zero but this approach allows to essentially keep the Poisson error calculations when the channels are grouped. For the background or correction files  $e_i = 0$  usually will not lead to bad consequences because all  $e_i$  are anyway much smaller than the uncertainties in the source spectrum.

### 2.1.2 Adding response data

Even if the observed counts are simply added (eq. 6), the response data must be correctly weighted,

$$A(E) = w_1 A_1(E) + w_2 A_2(E), \quad (17)$$

$$R(i, E) = w_1 R_1(i, E) + w_2 R_2(i, E). \quad (18)$$

Strictly speaking, the weights  $w_1$  and  $w_2$  should be different for adding the effective area and redistribution files. The predicted co-added spectrum is

$$m_i = m_i^{(1)} + m_i^{(2)} = N_1 \int f(E) A_1(E) R_1(i, E) dE + N_2 \int f(E) A_2(E) R_2(i, E) dE, \quad (19)$$

where  $N_{1,2}$  are normalization factor reflecting exposure and area coverage of each observation. For the co-added response, we want

$$m_i = (N_1 + N_2) \times \int f(E) A(E) R(i, E) dE, \quad (20)$$

therefore

$$\int f(E) A(E) R(i, E) dE = \frac{1}{N_1 + N_2} \int f(E) (N_1 A_1(E) R_1(i, E) + N_2 A_2(E) R_2(i, E)) dE, \quad (21)$$

$$A(E) R(i, E) = \frac{N_1}{N_1 + N_2} A_1(E) R_1(i, E) + \frac{N_2}{N_1 + N_2} A_2(E) R_2(i, E). \quad (22)$$

When the redistribution functions are identical, we want  $R(i, E) = R_1(i, E) = R_2(i, E)$ , and therefore the weights for the area average are

$$w_1 = \frac{N_1}{N_1 + N_2}, \quad w_2 = \frac{N_2}{N_1 + N_2}. \quad (23)$$

If we use this “natural” choice of weights for both effective areas and redistribution functions, we have

$$A(E) R(i, E) = (w_1 A_1 + w_2 A_2)(w_1 R_1 + w_2 R_2) = w_1 A_1(w_1 R_1 + w_2 R_2) + w_2 A_2(w_1 R_1 + w_2 R_2) \quad (24)$$

instead of

$$A(E) R(i, E) = w_1 A_1 R_1 + w_2 A_2 R_2 \quad (25)$$

The error introduced vanishes in two cases,

1. Same redistribution functions,  $R_1 = R_2$ , because

$$w_1 A_1 (w_1 R_1 + w_2 R_2) + w_2 A_2 (w_1 R_1 + w_2 R_2) = w_1 A_1 R_1 + w_2 A_2 R_2 \quad (26)$$

2. Same effective area,  $A_1 = A_2$ , because

$$w_1 A_1 (w_1 R_1 + w_2 R_2) + w_2 A_2 (w_1 R_1 + w_2 R_2) = w_1 R_1 (w_1 A_1 + w_2 A_2) + w_2 R_2 (w_1 A_1 + w_2 A_2) \\ = w_1 A_1 R_1 + w_2 A_2 R_2 \quad (27)$$

Both extremes are valid when adding spectra from the single-type (either BI or FI) *Chandra* CCDs. If the spectral from different-type CCDs are added, they must be binned so that the response becomes diagonal for each type.

The above discussion shows that the natural choice for the weights in eq. (17) and (18) is

$$w_j = \frac{\text{area}_j \times \text{exposure}_j}{\sum_j \text{area}_j \times \text{exposure}_j} \quad (28)$$

The spectral files do not contain enough information to correctly assign the weights unless it can be assumed that the all areas are the same (in this case the weights are proportional to exposures). In the general case, the weights must be computed externally using, e.g., the exposure maps or/and the model distribution of the source surface brightness.

## 2.2 Software

All the necessary computations can be performed using FTOOLS tasks `mathpha`, `addrmf`, and `addarf`. Below, we provide the necessary procedures and discuss the issues.

### 2.2.1 Adding source spectra with `mathpha`

```
mathpha expr=spec1.pha+spec2.pha outfil=result.pha \
  units=COUNTS ncomments=0 \
  properr=no ERRMETH=POISS-0\
  areascal=NULL\
  exposure=CALC
```

When `exposure=CALC`, the exposures in the input spectra are added. If this is not what you want, it can be changed to e.g. `exposure=spec1.pha` (to copy the value from `spec1.pha`) or `exposure=10000` (to set it to 10 ksec).

### 2.2.2 Adding background and correction files

The weights in eq. (13) and (15) are best to calculate externally using the values of exposure and `*SCAL` keywords in the input file. In the following examples, we assume that the weights 0.01 and 0.003

```
mathpha expr="0.01*spec1.bg+0.003*spec2.bg" outfil=result.bg \
  units=RATE ncomments=0 \
  properr=yes errmeth=Gauss \
  areascal=NULL\
  exposure=1.0
```

This creates the correct `STAT_ERR` column and we do not want to restore the Poisson errors. Identical computation (with different weights) is applied to correction files.

### 2.2.3 Issues for `mathpha`

Sadly, `mathpha` allows division, `'/'`, as a valid operand. This means one cannot have input spectra in different directories. Either copy them or make soft links to the current directory.

### 2.2.4 Adding response files

The strategies for computing the weights for the spectral response files are discussed in §2.1.2. Let us assume that the correct weights are 0.75 and 0.25 (they should total to 1).

```
addrmf spec1.rmf,spec2.rmf 0.75,0.25 rmffile=result.rmf
addrmf spec1.arf,spec2.arf 0.75,0.25 result.arf
```

### 2.2.5 Updating headers

As a final step, we want to update the keywords in the resulting spectrum:

```
fparkey result.bg result.pha[spectrum] " BACKFILE
fparkey 1.0      result.pha[spectrum] " BACKSCAL
fparkey result.arf result.pha[spectrum] " ANCRFILE
fparkey result.rmf result.pha[spectrum] " RESPFILE
fparkey result.cor result.pha[spectrum] " CORRFILE
fparkey 1.0      result.pha[spectrum] " CORRSCAL
```

## 3 Adding more than two spectra

Mathematically, equations for adding the spectra, background, correction, and response files can be trivially generalized to the case of more than two input files. Software (mathpha, addrmf, addarf) also allows for several input files. However, if there are too many spectra to add, this can be done consecutively.

### 3.1 Adding many spectra “in turn”

The source and background/correction files can be easily added “in turn” using the procedure for two spectra outlined above:

```
spec.pha = spec1.pha + spec2.pha
spec.pha = spec.pha + spec3.pha
...
```

and identically for the background and correction files.

The procedure is only slightly more complicated for the response data. Let  $w_1, \dots, w_n$  are correct weights for adding response in  $n$  input spectra ( $\sum w_i = 1$ ). The ratio of first two responses in the result should be  $r_1 : r_2 = w_1 : w_2$ , therefore on the first step,

$$R_2 = \frac{w_1}{w_1 + w_2} r_1 + \frac{w_2}{w_1 + w_2} r_2 \quad (29)$$

On the second step,

$$R_3 = \frac{w_1 + w_2}{w_1 + w_2 + w_3} R_2 + \frac{w_3}{w_1 + w_2 + w_3} r_3 \quad (30)$$

and so on. This is the formal rule:

$$R_1 = r_1, \quad R_i = \frac{\sum_{j=1,i-1} w_j}{\sum_{j=1,i} w_j} R_{i-1} + \frac{w_i}{\sum_{j=1,i} w_j} r_i \quad (31)$$