Specifications for Chandra Source Catalog Release 2: Variability Flags and Hardness Ratios

Michael Nowak, MIT-CXC, April 13, 2016, Version 1.0

1 Overview

The variability and hardness ratios in most instances will be derived from similar starting points[1], namely the probability distributions for the source photometric fluxes, specifically the distributions derived for the ECF90 aperture regions. For this reason, we describe the variability and hardness ratio[2] specifications together.

In what follows, we will describe specifications for variability tests in different energy bands, and in the case of hardness ratios different combinations of energy bands. Each of these tests will be applied to different groupings of observations. As regards the grouping of observations, there are three major categories that need to be distinguished: 1) properties related to source regions from individual observations, 2) properties related to groupings of individual observations derived from one or more detect stacks, i.e., Master Sources and Master Averages, and 3) properties related to other subdivisions of the data, e.g., individual stacks and subdivisions by Bayesian Blocks analyses (the latter itself possibly being performed on either a per-stack or a per-Master Source basis).

The outline of this specification is therefore as follows. In §2 we discuss derived properties for the source detection regions as applied to *individual* observations. §3 describes derived variability and color properties for multiple observations of the same source region, i.e., Master Sources and Master Averages. In this section, I also briefly discuss the possibility of creating regions for "undetected sources". §4 then discusses how flux, variability, and hardness ratio calculations should be generalized to other subdivisions of the data, such as Bayesian Blocks decompositions. Within each section, we first describe variability, and then hardness ratios.

2 Individual Observations

Both the variability and hardness ratios for individual observations are to be calculated using the same ECF90 source regions and events used for the aperture photometry calculations outlined in Frank Primini's memo (Sept. 5, 2013; with further mathematical background outlined in Primini & Kashyap 2014, ApJ, 796, 24). These regions are obtained from successful source detections as derived from the stacked individual observations. These regions will be used to create hardness ratios for individual observations that comprise a stack, as well as to assess inter-observation variability for the stack-derived sources, even if that given observation contributes no discernible signal to the stack detection.

For every successful source detection in a given stack — regardless of whether or not there are discernible events within a given energy band or within an individual observation at that location — both "source" and "background" regions will be used to generate aperture photometry results (even if only upper limits) for all energy bands and individual observations that comprise that stack.

These aperture photometry regions will be used also for variability and hardness ratio calculations in the individual observations that comprise the stack. We will, however, only perform variability and hardness ratio calculations for sources for which we will provide photometry results. That is, if MLE flags a source

[1]The major exception being the variability tests performed on *events from individual observations*, which adhere to the methods used in the first catalog release.

[2]For purposes of this memo, "hardness ratio" is synonymous with "color", and refer to the same concept, namely the difference between aperture fluxes in two bands, ratioed to the sum of these same two aperture fluxes.

as false and we therefore do not generate photometry products, we will not generate variability and hardness ratio results for that source.

2.1 Variability

Variability studies for regions in individual observations comprise the only set of properties described in this memo that *do not* rely upon the *probability distributions* derived for photometry calculations; however, they do rely upon the *source regions* defined for the aperture photometry calculations. Variability studies, just as for Catalog Release 1.1, will rely upon calculations with counts and regions, with further knowledge of the region area vs. time as ascertained from the `dither_region` tool. As for Release 1.1, corrections will be made for the region area as a function of time (e.g., dithering over a chip edge or bad pixel), but no corrections will be made for response variations (e.g., dithering between chip nodes).

A lightcurve of counts vs. time will be extracted for each region in each energy band, and likewise the `dither_region` tool will be run for each of these regions. Identical to the the Catalog Release 1.1, the counts vs. time lightcurve will be coupled with the area fraction vs. time curve and run through the existing Kuiper, Kolmogorov-Smirnov (K-S), and Gregory-Loredo (`glvary`) tools. That is, the normalized cumulative distribution function for the counts vs. time will be compared to the normalized cumulative distribution function for the dither region area vs. time for the Kuiper and K-S tests, while both lightcurves will be input to the `glvary` tool.

The following properties, identical to Catalog Release 1.1, will be calculated and stored:

- `ks_prob_[u/s/m/h/b/w]` - The Kolmogorov-Smirnov test probability for each energy band.
- `kp_prob_[u/s/m/h/b/w]` - The Kuiper test probability for each energy band.
- `var_prob_[u/s/m/h/b/w]` - The Gregory-Loredo test probability for each energy band.

As for Catalog Release 1.1, the time averaged lightcurve in each energy band that is calculated by the `glvary` tool will be stored, and further the following properties of this lightcurve (also calculated and output by the `glvary` tool) will be retained and stored:

- `var_mean_[u/s/m/h/b/w]` - The time averaged Gregory-Loredo lightcurve mean count rate in each energy band.
- `var_sigma_[u/s/m/h/b/w]` - The standard deviation of the time averaged Gregory-Loredo lightcurve in each energy band.
- `var_max_[u/s/m/h/b/w]` - The minimum count rate in the time averaged Gregory-Loredo lightcurve.
- `var_min_[u/s/m/h/b/w]` - The maximum count rate in the time averaged Gregory-Loredo lightcurve.

The following property derived from the Gregory-Loredo analyses will also be calculated and stored identically to Catalog Release 1.1:

- `var_index_[u/s/m/h/b/w]` - An index in the range 0–10 that combines the odds ratio calculated by the `glvary` tool with the fraction of the `glvary` lightcurve within its 3σ and 5σ bounds.

The calculated variability index can be found in the header of the probability file output by the `glvary` tool.

There are two additional variability properties from Catalog Release 1.1 which we should retain for the next catalog release; however, I cannot find *precise definitions* for them on either our catalog web pages or within the catalog description paper (Evans et al., 2010, ApJ, 189, 37). **Subject to review of their precise definitions, and determination of whether we want to modify these definitions,** we will create and store the following values.

- dither_warning_flag - A Boolean set to TRUE if the peak of the power spectrum of the count rate lightcurve occurs at the dither frequency, or at a beat frequency of the dither frequency, of the observation. Not stated in our published definitions are: which lightcurve we use to calculate the power spectrum (I am presuming it is the b or w band glvary lightcurve), how precisely the peak frequency must match a dither frequency, and which possible beat frequencies are considered.
- var_code - A binary code that signifies which bands (u/s/m/h/b/w) contain significant variability. Not stated in our published definitions is which combination of the three variability tests, and their respective significance levels, we use to define "significant variability".

These are both useful values to retain; however, we should verify their current definitions.

2.2 Hardness Ratios

Hardness ratios in Catalog Release 1.1 all have the form hard_xy, where we have chosen the convention that x is always the *higher* energy band. This is a perfectly reasonable convention, so we retain it for Release 2. The Catalog Release 1.1 definition of hardness ratio, however, follows (flux_apper_x - flux_apper_y)/flux_apper_b, where b is the broad energy band. For mathematical simplicity (see the definitions below), we are changing this definition so that the ratio is defined as (x-y)/(x+y). To be more precise, let us define the ECF90 aperture flux in each energy band, x and y, as F_x and F_y, respectively, and their sum as F_{xy}. For purposes of aperture photometry estimates, each of these values are random variables with associated probability distributions, $P_x(F_x)$, $P_y(F_y)$, and $P_{xy}(F_{xy})$. The calculations of $P_x(F_x)$ and $P_y(F_y)$ follow the specifications of Frank Primini's memo. In what follows we will not need to explicitly calculate $P_{xy}(F_{xy})$.

The hardness ratio, H_{xy} will be *defined* by the relationships $F_x \equiv (1 + H_{xy})F_{xy}/2$ and $F_y \equiv (1 - H_{xy})F_{xy}/2$, even though we are not directly providing estimates of F_{xy} in the catalog. This allows us to write the two-dimensional probability function:

$$P_{x,y}(F_x, F_y)\, dF_x dF_y \;=\; P_x(F_x)P_y(F_y)\frac{F_{xy}}{2}\, dH_{xy}dF_{xy} \;\; . \qquad (1)$$

Note that $P_{xy}(F_{xy}) \neq P_{x,y}(F_x, F_y)$. $P_{xy}(F_{xy})$ is a one-dimensional probability distribution for the summed flux in the two energy bands, while $P_{x,y}(F_x, F_y)$ is a two-dimensional probability distribution for the fluxes in the individual energy bands. Furthermore, we are going to assume that the probability distributions $P_x(F_x)$ and $P_y(F_y)$ are independent of one another. This allows us to set $P_{x,y}(F_x, F_y) = P_x(F_x)P_y(F_y)$. Although this restricts us to non-informative priors on the hardness ratio, one could still use informative priors on each of the individual flux bands. Additionally, the photometry calculations used to create the flux probability distributions already assume independent energy bands.

The joint distribution of the fluxes is now used to calculate the hardness ratio distribution. Specifically, the probability distribution for the hardness ratio becomes:

$$\begin{aligned} P_{H_{xy}}(H_{xy})\, dH_{xy} &= \int_{F_{xy}=0}^{\infty} P_x(F_x)P_y(F_y)\frac{F_{xy}}{2}\, dH_{xy}dF_{xy} \\ &= \int_{F_{xy}=0}^{\infty} P_x\left(\frac{(1+H_{xy})F_{xy}}{2}\right) P_y\left(\frac{(1-H_{xy})F_{xy}}{2}\right)\frac{F_{xy}}{2}\, dH_{xy}dF_{xy} \;\; . \qquad (2) \end{aligned}$$

For the Release 2 catalog, we will report the value of H_{xy}^{max}, which we define as the hardness ratio that maximizes $P_{H_{xy}}$. Although it is unlikely to occur in practice, calculation of this maximum value should allow for cases where we have multiple estimates of H_{xy}^{max} (e.g., possibly due to employing a discrete approximation to our estimate of $P_{H_{xy}}$). In such cases we will report the midpoint of the range of values

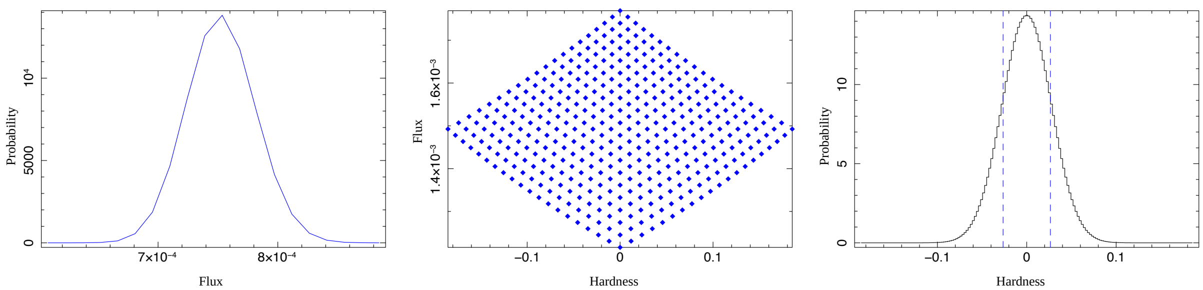

Figure 1: Left: A flux probability distribution (provided to me by Frank), with 20 bins. For illustration, we will assume that both flux channels have the same distribution. Middle: The discrete points at which the flux probability distributions are stored mapped to their associated discrete points in the summed flux/hardness space. (Note that this particular mapping is *more symmetric* than will be typical, since I have used the exact same flux probability distribution for each band.) Right: The resulting hardness probability distribution estimated at 161 grid points (i.e., $\approx$ eight times the number of bins in a single flux probability bin), with a simple estimate of its 68% confidence limits.

of $\{H_{xy}^{max}\}$. The error bar ranges will be defined by an amplitude cut on the probability distribution $P_{H_{xy}}$. Specifically, we will determine a value P^{cut} such that

$$\int_{P_{H_{xy}} \geq P^{cut}} P_{H_{xy}}(H_{xy}) \, dH_{xy} \approx 0.68 \ . \tag{3}$$

The values of H_{xy} can run from -1 to 1. The lower limit on H_{xy}, $\equiv H_{xy}^{low}$, is the lowest value ≥ -1 that is part of the integration boundaries, and the upper limit on H_{xy}, $\equiv H_{xy}^{up}$, is the highest value ≤ 1 that is part of the integration boundaries. Note that this definition does not lead to "symmetric" error bars since in general we expect that

$$\int_{H_{xy} < H_{xy}^{low}} P_{H_{xy}}(H_{xy}) \, dH_{xy} \neq \int_{H_{xy} > H_{xy}^{up}} P_{H_{xy}}(H_{xy}) \, dH_{xy} \ . \tag{4}$$

If one band has especially low counts compared to the other, we can expect the values of these two integrals to be very different from one another. (That is, it will be likely that in such a case one of the error bounds will be either -1 or 1.)

The individual probability components $P_x(F_x)$ and $P_y(F_y)$ are those calculated for aperture photometry as described in Frank Primini's memo. These are being stored as discrete values in a FITS file. It is important to note that the current storage scheme is using *probability densities*, and that the calculations above are also using *probability densities*. As regards performing the actual calculation of the probability density for the hardness ratio, and its subsequent integration over flux, *I am not requiring that a specific scheme be followed*. Assuming that we store the probability densities as vectors of length N, then we one can create an $N \times N$ grid for the two-dimensional probability distributions. With such a grid, one could: 1) directly numerically integrate out (i.e., marginalize over) the summed flux, F_{xy}, on this grid, 2) fit a two-dimensional Gaussian to the values on the grid and then integrate over one of the dimensions of the fit, or 3) apply a two-dimensional spline fit to the values on the grid and then integrate out the summed flux variable. Here I present an example of the latter procedure, as one suggestion for calculating the hardness ratio probability distribution.

It should be noted, however, that this grid will *not* have evenly spaced values of H_{xy} or F_{xy}. Furthermore, being finite, the grid will be truncated around the edges (see Fig. 1), so in general we should expect that even integrating the full two-dimensional probability distribution over the entire grid will result in a value not equal to unity. It is hoped (but needs to be verified) that the flux distributions will fall off sufficiently rapidly near the integration edges so that the integral nearly evaluates to unity; however, we expect that a small renormalization of the resulting probability distribution will need to be applied. Note also that the contours of constant flux are lines with slope $= -1$. There might be ways of taking advantage of this fact in integrating over this variable when turning the two-dimensional probability distribution into a one-dimensional hardness distribution.

In Fig. 1, I show a simple example of creating a hardness distribution from two (in this case, identical) flux probability distributions. The only assumption made here was that the input probability distributions fell-off sufficiently rapidly in their tails that round off errors (and truncation of the 2D grid; middle Fig. 1) in the numerical integration were insignificant. The hardness ratio distribution was calculated via a very simple integration over summed flux in the two dimensional summed flux/hardness probability density, as determined from a bilinear interpolation of $P_x(F_X)P(F_y)F_{xy}/2$ mapped to the F_x vs. F_y plane[3]

The 68% confidence limits were determined by reverse sorting (high to low value) the resulting hardness probability distribution, multiplied by the hardness bin width, and then performing a cumulative sum. The value of the probability distribution was determined for that last point where the cumulative sum was ≤ 0.68. The hardness limits were then taken as the lowest and highest hardness bin boundaries with a probability distribution greater than or equal to this value. In these calculations, the summed flux grid ran from the summed lowest bins to the summed highest bins of the input flux probability ranges, while the hardness bin ran similarly over the extremes that could be calculated from the input flux probability distributions. The number of grid points used for the summed flux and the hardness axes were taken to be four times (plus one) of the sum of the number of input grid points in each of the flux probability distributions. This might be oversampling the hardness probability distribution somewhat, but was done to create a smooth curve. Again, this procedure is an admittedly crude estimate, but may suffice for our purposes. (One can find two dimensional interpolation algorithms in `SciPy`, e.g., `http://docs.scipy.org/doc/scipy/reference/tutorial/interpolate.html`; the simple scheme here ran in about 0.07 s on a single core on my 2.8 GHz i7 Apple laptop.)

The `S-lang` code used to create the above probabilities is reproduced here below. (Any scheme that will work with greater precision and/or speed is an acceptable substitute!)

```
variable ix, iy, px, py;

variable hard_pdf = "acisf00390_000N004_r0001b_mpdfs.fits[1]";
variable soft_pdf = "acisf00390_000N004_r0001b_mpdfs.fits[1]";

% Intensity, Probability Density
(ix,px) = fits_read_col(hard_pdf,"inten","margpdf");
(iy,py) = fits_read_col(soft_pdf,"inten","margpdf");

variable lx = length(ix), ly = length(iy), nh=4*(lx+ly)+1, nf=@nh;

variable min_ix=min(ix), max_ix=max(ix), min_iy=min(iy), max_iy=max(iy);

% Extremes of color & flux (assuming rectangular ix/iy grid)
variable hlo=(min_ix-max_iy)/(min_ix+max_iy),
```

[3]A bilinear interpolation was used to create an evenly and slightly over-sampled grid over which to numerically integrate the probability distribution. *This is not a strict requirement.* The numerical schemes I chose were used for simplicity and transparency, *not speed or accuracy*. For example, there are 2D interpolation schemes that are suitable for creating non-evenly spaced grids; however, I found many of those schemes to be “black boxes”. The only requirement is that we must be able to create the 2D summed flux/hardness probability distribution, and then numerically marginalize over the summed flux variable. The schemes presented here appear to work, and can be considered a baseline against which we can compare other schemes.

```
         hhi=(max_ix-min_iy)/(max_ix+min_iy),
         flo=min_ix+min_iy, fhi=min_ix+max_iy;

% Flux/Probability density pairs at grid boundaries for interpolation
define flx_prob_bounds(fx,fy)
{
   variable idx_lo, idx_hi, xlo, xhi, ylo, yhi, pxlo, pxhi, pylo, pyhi;

   idx_lo = wherelast(ix <= fx);
   idx_hi = wherefirst(ix >= fx);

   if(idx_lo==NULL) {
      xlo=fx; pxlo=0.; xhi=ix[idx_hi]; pxhi=px[idx_hi];}
   else if(idx_hi==NULL) {
      xhi=fx; pxhi=0.; xlo=ix[idx_lo]; pxlo=px[idx_lo];}
   else if (idx_lo==idx_hi) {
      if (idx_lo) { idx_lo--; } else { idx_hi++; }
      xlo=ix[idx_lo]; pxlo=px[idx_lo]; xhi=ix[idx_hi]; pxhi=px[idx_hi];}
   else {
      xlo=ix[idx_lo]; pxlo=px[idx_lo]; xhi=ix[idx_hi]; pxhi=px[idx_hi];}

   idx_lo = wherelast(iy <= fy);
   idx_hi = wherefirst(iy >= fy);

   if(idx_lo==NULL) {
      ylo=fy; pylo=0.; yhi=iy[idx_hi]; pyhi=py[idx_hi];}
   else if(idx_hi==NULL) {
      yhi=fy; pyhi=0.; ylo=iy[idx_lo]; pylo=py[idx_lo];}
   else if (idx_lo==idx_hi) {
      if (idx_lo) { idx_lo--; } else { idx_hi++; }
      ylo=iy[idx_lo]; pylo=py[idx_lo]; yhi=iy[idx_hi]; pyhi=px[idx_hi];}
   else {
      ylo=iy[idx_lo]; pylo=py[idx_lo]; yhi=iy[idx_hi]; pyhi=py[idx_hi];}

   % These define the 4 corners bounding input flux pair
   return xlo, pxlo, xhi, pxhi, ylo, pylo, yhi, pyhi;
}

% Bilinear interpolation of summed flux/hardness probability density
% for a given input pair of fluxes in individual bands
define bilinear_flux_hardness(fx,fy)
{
   variable xlo, pxlo, xhi, pxhi, ylo, pylo, yhi, pyhi;
   (xlo, pxlo, xhi, pxhi, ylo, pylo, yhi, pyhi) =
      flx_prob_bounds(fx,fy);

   variable fq11, fq12, fq21, fq22;
   fq11 = pxlo*pylo*(xlo+ylo)/2;
```

```
   fq12 = pxlo*pyhi*(xlo+yhi)/2;
   fq21 = pxhi*pylo*(xhi+ylo)/2;
   fq22 = pxhi*pyhi*(xhi+yhi)/2;

   variable p2d = (fq11*(xhi-fx)*(yhi-fy)+fq21*(fx-xlo)*(yhi-fy)+
     fq12*(xhi-fx)*(fy-ylo)+fq22*(fx-xlo)*(fy-ylo))/(xhi-xlo)/(yhi-ylo);

   return p2d;
}

% Hardness & Flux Grids, Hardness Probability
variable i, j, h_lo, h_hi, f_lo, f_hi, df, dh, h, ft,
         ph=Double_Type[nh];

(h_lo,h_hi) = linear_grid(hlo,hhi,nh);
(f_lo,f_hi) = linear_grid(flo,fhi,nf);

ft = (f_lo+f_hi)/2.; df=(f_hi[0]-f_lo[0]); dh=(h_hi[0]-h_lo[0]);

_for i (0,nh-1,1){
   h=(h_lo[i]+h_hi[i])/2.;
   ph[i]=sum(df*array_map(Double_Type, &bilinear_flux_hardness,
                          (1+h)*ft/2, (1-h)*ft/2               )); }

% Normalized hardness probability distribution
ph = ph/sum(ph*dh);

variable imax = where(ph==max(ph));
% Reported hardness ratio
variable h_best = (h_lo[imax[0]]+h_hi[imax[-1]])/2.;

variable phint = ph*dh;
variable ias = reverse(array_sort(phint));
variable phint_sum = cumsum(ph[ias])*dh;
variable ipsum = wherelast(phint_sum < 0.68);
variable ph_cut = ph[ias[ipsum]];
variable ilo = wherefirst(ph >= ph_cut);
variable ihi = wherelast(ph >= ph_cut);

% Hardness Limits
variable h_lolim = h_lo[ilo], h_hilim=h_hi[ihi];
```

Based upon the above definitions, the following properties (with names identical to Catalog Release 1.1, although definitions slightly different than the earlier release) will be calculated and stored *for each observation comprising a given stack*:

- `hard_[ms/hs/hm]` - The hardness value, as defined above (i.e., the peak of the probability distribution, or the midpoint between the extremes of multiple identical peaks) for the softer band subtracted from the harder band, relative to their sum. The hardness ratios for the medium to soft, hard to soft, and hard to medium bands only will be calculated; we are not calculating any hardness ratios involving the ultra-soft (u), broad (b), or wide (w) energy bands.

- hard_[ms/hs/hm]_lolim - The lower limits of the above hardness ratios, as defined in this memo. Taken together with hard_[ms/hs/hm]_hilim, this encompasses the 68% confidence limit. These are *not symmetric error bars*.

- hard_[ms/hs/hm]_hilim - The upper limits of the above hardness ratios, as defined in this memo. Taken together with hard_[ms/hs/hm]_lolim, this encompasses the 68% confidence limit. These are *not symmetric error bars*.

It is important to note that the above quantities, being based upon probability distributions, can and should be calculated for observations within the stack that lack any discernible signal in one or more energy bands. So long as a photometry probability distribution can be calculated for all energy bands, each of the above colors will be able to be calculated. (Some care might need to be taken with the numerical integration near the $F_x = F_y = 0$ boundary.)

The following property is new to Catalog Release 2.

- hard_[ms/hs/hm]_prob - The probability distributions for the medium to soft, hard to soft, and hard to medium energy bands (stored as hardness vs. probability density, with suggested column names color and margPDF, the latter column name is currently consistent with the flux probability density file provided to me by Frank Primini). It should be stored in a similar manner (and perhaps the same file) as the aperture flux probability distributions. In terms of its gridding, it should have on the order of a few times the sum of the number of grid points in the input flux distributions, with the hardness ratio grid running over the extreme values that can be calculated from the extreme values in the input flux probability distributions, i.e., $(min(F_x) - max(F_y))/(min(F_x) + max(F_y))$ to $(max(F_x) - min(F_y))/(max(F_x) + min(F_y))$.

3 Multiple Observations (Master Sources and Master Averages)

Once we have gone beyond the level of individual observations and individual energy bands and look at properties related to groups of observations (Master Sources and Master Averages), *newly calculated*[4] variability and hardness ratios both will be determined from the probability distributions derived from the aperture photometry calculations.

For cases of grouped observations (whether we are considering Master Source, Master Average, or Bayesian Block groupings of observations), an important difference remains between how the variability and hardness ratio calculations use aperture flux probability distributions. Variability estimates still make use of the aperture flux probability distributions for *individual observations*. Hardness ratio estimates, however, will make use of the aperture flux probability distributions promoted to the Master Source or Master Average level that have been created from the *combination* of individual observations. The philosophy will be that to the extent that a set of flux probability distributions are calculated at the Master Source or Master Average level, corresponding Master Source and Master Average hardness ratio estimates will be created from these flux distributions.

As for Catalog Release 1.1, we also will include information about color variability, which *does* utilize the hardness ratio probability densities calculated for *individual* observations. This is described in §3.3 below.

[4]As in Catalog Release 1.1, certain already calculated values of variability probability will be promoted to the Master Source/Master Average level.

3.1 Regions for "Undetected Sources"

There are potential sets of "source regions" that do not currently fit into any of our paradigms of Master Source, Master Averages, or Bayesian Blocks calculations. These regions might represent fairly rare cases, and furthermore they would require additional calculations that by definition won't have been performed on the first pass through the pipeline. We can consider whether the properties from these regions should be: 1) ignored[5], 2) incorporated into the definitions of the Master Average (as opposed to Master Source, which likely will be comprised only of the high S/N observations of a source), or 3) taken as part of another class alongside Master Source and Master Average. As such, the following should be considered a preliminary outline for considering these regions, and should be at a low priority for implementation in Release 2 of the Catalog.

The specific case considered here is that of "undetected" sources which would require us to create source regions outside of the normal stack detection pipeline in order to generate additional photometry, variability, and hardness ratio products. An example of such an "undetected" source is as follows. Imagine that there are two or more stacks of observations that have overlapping regions. If a given location with a successful detection in one or more stacks does not have a successful detection in other overlapping stacks, those stacks with non-detections will never have had regions defined and photometry calculated at that location. Therefore one would have to *post facto* generate ECF90 regions for each observation and energy band in the stack[s] lacking detections. These regions would then be used to relaunch the pipeline, and generate photometry, variability, and color products.

There are four major points to address, in concert with the photometry specifications, as regards regions for "undetected" sources:

- *The source location to be used for generating photometry upper limits in stacks without source detections should be the best-fit Master Source locations.* For the case of only two overlapping stacks (one with a detection, the other without), this source location is that from from the single stack with a detection. If there are two or more stacks with detections, any overlapping stacks will take the best-fit sky positions calculated for the Master Source, with the Master Source location being determined only using stacks with positive source detections.

- *For cases of ambiguous cross-matches among stack detections, where we already will be generating multiple Master Sources, one should generate multiple different corresponding "undetected source" regions for any overlapping stacks without a corresponding source.* An obvious case where this would be necessary is where two different on-axis sources in one stack are ambiguously matched to the same off-axis source in a second stack, while neither has a match to a third stack that encompasses the same locations. In this case, the "non-detection regions" would be different from one another, with these "non-detection" regions possibly substantially overlapping each other. Upper limits must still be calculated for each of these "non-detection" regions. This rule is essentially the same as above: each Master Source location should be used to generate regions for photometry limit calculations in any overlapping stacks with non-detections.

- *The region size to be used for an energy band and observation lacking an independent source detection is the ECF90 region suitable for a point source, given the point spread function, at the Master Source location.* That is, for sake of uniformity across the catalog — and the fact that the source extent itself might be subject to a large degree of uncertainty, be variable across observations, or be different in each energy band — one should always use "point source" appropriate ECF90 region sizes and shapes to calculate variability and color properties.

[5] My understanding of what we did in Release 1.1 as regards Master Sources was that we only matched *positive* detections. That is, we did not take a positive detection from one observation and compare it to a non-detection at the same source position in another observation. Thus, we have some precedent from the first catalog for not considering "undetected sources".

- *For purposes of hardness ratio and variability properties that involve more than a single stack, "undetected source" photometry upper limits will only be required for overlapping stacks.* That is, once we go beyond consideration of hardness ratios and variability within a stack, we should not compare to any *individual observation*, unless that individual observation is itself a "stack".

Again, to the best of my knowledge, this is not a complication that we considered or allowed for in Release 1.1 of the catalog. It will require further processing after the pipeline has been run through to completion once. Therefore, inclusion of such "undetected source regions" is a decision that we can postpone for a later date.

3.2 Variability

We should first note that the variability test described here should be considered distinct from the Bayesian Blocks decompositions, discussed below. Although the Bayesian Blocks decompositions are in themselves variability tests, where any successful decomposition is an indication of source variability, the variability analysis discussed below is a more broad and general analysis than the Bayesian Blocks analysis.

Bayesian Blocks is simultaneously answering two questions: does a source indicate variability *above a given fixed threshold*, and if so, what is the best subdivision of time intervals into blocks of statistically uniform flux? For the variability tests described here, we are answering a single, highly general, question: to what degree are the individual observations inconsistent with a source of constant flux? From this point of view, Bayesian Blocks is answering a restrictive question, akin to the Gregory-Loredo test used for individual observations, whereas the test described below is answering a general question, akin to the Kolmogorov-Smirnov and Kuiper tests applied to the individual observations. Thus, it is likely that we will find a number of sources without any Bayesian Blocks decompositions that will still have significant variability in the long-term lightcurve as determined by the procedures described below.

The test begins with the probability distributions, ${P_x}^i({F_x}^i)$, for the aperture flux in energy band x, ${F_x}^i$, for the i^{th} observation out of the N observations that make up the Master Source or Master Average under consideration. *Note that this includes probability distributions for individual observations and energy bands without discernible signal at the Master Source or Master Average location.* We expect that a given Master Source or Master Average can be comprised of multiple independent stacks of observations, with each stack itself comprised of multiple observations. In what follows, the index i does not make any distinction among separate stacks and represents *regions at the same location from all individual observations, including those with only upper limits to the aperture flux, that comprise the Master Source or Master Average*.

First, we calculate the N means and N standard deviations for each of these distributions, as labeled by i for each individual observation:

$$\overline{F_x}^i = \int {F_x}^i {P_x}^i({F_x}^i) \, d{F_x}^i \ \ , \ \ {\sigma_x}^{i2} = \int (\overline{F_x}^i - {F_x}^i)^2 {P_x}^i({F_x}^i) \, d{F_x}^i \ \ . \tag{5}$$

The N values $\overline{F_x}^i$ are useful to retain internally as a convenience for the calculation described below; however, the N values of ${\sigma_x}^{i2}$ *must be retained internally for calculation of a Master Source parameter value that we will report:* (`var_inter_sigma_[u/s/m/h/b/w]`).

We now perform two fits with the aperture flux probability distributions. For the first fit, we find the N fitted rates, ${F_x}^{i max}$, that maximize the log of the likelihood function:

$$\log L^{i max} = \sum_i \log {P_x}^i({F_x}^{i max}) \ \ . \tag{6}$$

This is equivalent to minimizing the negative log of the likelihood. These values are just the maximum likelihood fluxes that already have been calculated for aperture photometry (and may even be stored in the header of the probability distribution FITS file). The values from the files can be reused here. (Furthermore, these values of ${F_x}^{imax}$ are reused below, so having them previously calculated and stored is worthwhile.)

The second fit is the *single* rate, F_x^{max} that maximizes the log of the likelihood function.

$$\log L_2^{max} = \sum_i \log {P_x}^i (F_x^{max}) \ . \tag{7}$$

Again, the maximization can be trivially recast as a minimization.

The values of F_x^{max} and ${F_x}^{imax}$ must be retained internally to calculate `var_inter_sigma_x`. Specifically, for each band and each observation calculate

$$|F_x^{max} - {F_x}^{imax}|/{\sigma_x}^i \ . \tag{8}$$

The *maximum* value of this quantity (searching over the individual observations i) in *each* energy band then becomes `var_inter_sigma_x`.

As an important technical note, in my own tests with aperture flux probability densities provided to me by Frank Primini, I have performed the above minimizations with Akima interpolations *of the logarithms of the tabulated probability densities*. (Again, a number of interpolation routines are available in `SciPy`.) This is as opposed to splining the probability function itself, and then taking the log of the splined values. The former procedure yielded better behavior in the wings of the probability distribution. The minimizations were performed using a `simplex` routine available in `ISIS`. Under `ISIS`, the form of the minimization took:

```
max_log_L = -simplex_new.optimize( params, param_min, param_max,
                                   &function, {function_input} );
```

where `max_log_L` is the log likelihood for the best fit, `params` holds the best fit count rate(s), `function` evaluates the negative of the log likelihood functions, and {`function_input`} passes information to the log-likelihood function, e.g., the spline fit to the log of the likelihood function. The above can be replicated in `Sherpa` with minimal difficulty. χ^2 minimization is just a special case of the above, with `max_log_L` $\rightarrow -\chi^2$, and {`function_input`} passing the data as well as information about the fit model. The negative log likelihood function just needs to be properly written so as to be passed properly to the `Sherpa` minimization routines.

Once the best-fit log-likelihoods are calculated, the `ISIS` routine passes back the quantity:

$$D \equiv 2 \left(\log(L_1^{max}) - \log(L_2^{max}) \right) \ , \tag{9}$$

and the number of degrees of freedom, i.e., $N - 1$. *This is just the basic likelihood ratio test.*

Since $L_1^{max} \geq L_2^{max}$, $D \geq 0$. Furthermore, we presume that in the absence of any true inter-observation variability, D will be distributed as χ^2 with $N-1$ degrees of freedom. The routine therefore also passes back a "false alarm" probability, p, which is given by one minus the cumulative distribution for the χ^2 statistic:

$$p \equiv 1 - F(\chi_N^2 = D; N - 1) \ . \tag{10}$$

If p is small, then the inter-observation variability is more likely to be real. The function $F(D; N - 1)$ can be found in a number of numerical libraries, including `scipy.special`. It can also be written in terms of Gamma functions and incomplete Gamma functions.

The following is an outline of how I programmed the inter-observation variability test in `ISIS`/`S-lang`. A version based upon `Sherpa` could follow a very similar procedure.

1. Read the N aperture flux probability distributions for a given energy band and Master Source/Master Average location. These N probability distributions will not only represent source detections, but often will include probability distributions for non-detections.

2. Calculate the mean and variance for each distribution, including the non-detections. My tests did this by integrating a spline-fit to determine each quantity. The means and variances are used to determine reasonable starting values for maximizing the log-likelihood functions, and the variance is later used to calculate `var_inter_sigma_[u/s/m/h/w]_[ms/ma]`. *These individual means and variances can be calculated and stored in the FITS file headers during the aperture photometry probability density calculation steps.*

3. Create a fit function for the negative log-likelihood functions. My tests utilized spline fits to the logs of the probability distributions. One fit function should take a single parameter: a common fitted rate for all N distributions. The other fit function should take N parameters: an individual rate for each of the N distributions.

4. Minimize $-\log(L_2)$ with a single rate, and minimize $-\log(L_1)$ with N rates. As a good "starting guess" for the former, use the error weighted mean:

$$\left(\sum_i^N \frac{\overline{F_x}^i}{{\sigma_x}^{i2}} \middle/ \left(\sum_i^N \frac{1}{{\sigma_x}^{i2}}\right)\right) , \tag{11}$$

   where $\overline{F_x}^i$ and ${\sigma_x}^{i2}$ are the means and standard deviations calculated for the individual distributions (Eq. 5). As a good starting guess for the latter, use the N individual means calculated from the distributions (also Eq. 5). The parameter minimum and maximum values here are also taken from the tabulated data (i.e., we do not search for values outside of the tabulated range).

5. Calculate the standard likelihood ratio statistic, $D = 2(\log(L_1^{max}) - \log(L_2^{max}))$, the degrees of freedom, $N - 1$, and convert these values into a false variability probability value, p, using the χ^2 cumulative distribution function (Eq. 10).

6. Use Eq. 8 in each band to calculate `var_inter_sigma_[u/s/m/h/b/w]_[ms/ma]`.

7. Return and store the results.

8. Perform the above steps separately for the observations comprising a Master Source and the observations comprising a Master Average.

Note: Currently we are planning to define several different groupings of individual observations. A cohort is the set of observations upon which the detect algorithm is run, with the stack being the further subset of those observations from which source properties (e.g., aperture flux) are derived for a given source region. A grouping of detections matched by position from one or more stacks will form a Master Source (for a curated set of detections, which may use a variety of criteria to determine the "best" detections) or a Master Average (for considering all observations from stacks where a successful detection has been achieved at that location). In the above, and in what follows, I am suggesting that we save variability and hardness ratio properties for the Master Source and Master Average classes, which I have designated by appending `_ms` and `_ma` to the familiar Release 1.1 Master Source property variable names. *This naming convention is merely a placeholder for this memo, and can be changed to conform to any naming scheme, e.g., being used for the aperture photometry products. Alternatively we can retain the naming scheme used for Master Sources in Release 1.1 (i.e., drop the suffixes `_ms` or `_ma`), so long as the named property is clearly identified its associated grouping class in the catalog release.*

Given the above caveat as regards the source property naming convention, the following properties, with similar names but new definitions as compared to Catalog Release 1.1, will be stored as part of the Master Source and Master Average properties:

- `var_inter_index_[u/s/m/h/b/w]_[ms/ma]` - This retains the same definition as given in Table 7 of Evans et al. (2010; ApJS, 189, 37), except that in lieu of the inter-observation variability reduced χ^2_ν as calculated in Release 1.1, we will use $D/(N-1)$ from the probability test described above (Eq. 9).

- `var_inter_prob_[u/s/m/h/b/w]_[ms/ma]` - The p value in each energy band as defined by Eq. 10.

- `var_inter_sigma_[u/s/m/h/b/w]_[ms/ma]` - The largest difference between mean flux and an individual band flux, relative to the flux standard deviation in that individual observation, in each energy band (Eq. 8).

The following properties, identical to Catalog Release 1.1, will be promoted from their values calculated at the individual observation step, and stored as part of the Master Source and Master Average properties:

- `ks_intra_prob_[u/s/m/h/b/w]_[ms/ma]` - The highest value of the Kolmogorov-Smirnov test probability (in each energy band) from the individual observations comprising the Master Source or Master Average.

- `kp_intra_prob_[u/s/m/h/b/w]_[ms/ma]` - The highest value of the Kuiper test probability (in each energy band) from the individual observations comprising the Master Source or Master Average.

- `var_intra_index_[u/s/m/h/b/w]_[ms/ma]` - The highest value of `var_index` (in each energy band) from the individual observations comprising the Master Source or Master Average.

- `var_intra_prob_[u/s/m/h/b/w]_[ms/ma]` - The highest value of the Gregory-Loredo test probability (in each energy band) from the individual observations comprising the Master Source or Master Average.

Logic must be included to account for the fact that these quantities will not have been calculated for observations and energy bands where there were no source counts. Note: I am suggesting that we drop inclusion of `var_intra_sigma_[u/s/m/h/b/w]`, the scaled count rate variability in counts s^{-1}, since it is a quantity without a lot of context at the Master Source level. Both `var_mean` and `var_sigma` will be accessible via the associated individual observations.

Finally, similarly to Catalog Release 1.1 we shall include the Master Source and Master Average properties:

- `var_flag_[ms/ma]` - A Boolean set to `FALSE` if all values of `var_inter_index` and `var_intra_index` from all individual observations and energy bands comprising the Master Source or Master Average lie below given values, and is otherwise set to `TRUE`. *These values (which might be distinct for `var_inter_index` and `var_intra_index` are still TBD.*

3.3 Hardness Ratios

At the Master Source or Master Average levels, calculation of the hardness ratios are straightforward *presuming that we have a probability densities defined for the Master Source or Master Average flux in each energy band.* Exactly which probability distributions will be promoted to the Master Source and Master Average levels are TBD. However, for whatever choices are made, the Master Source and Master Average hardness ratios then *follow exactly the same procedures as outlined in Eqs. 1–4*, but utilizing the appropriate aperture flux probability distributions for the Master Source or Master Average. The definitions of the hardness ratio and its lower and upper limits are exactly analogous to the individual observation cases.

The following Master Source and Master Average properties, with names based upon those from Catalog Release 1.1 but definitions following the individual observation procedures outlined in Eq. 1–4 and §2.2, will be calculated and stored:

- hard_[ms/hs/hm]_[ms/ma] - The hardness value for the softer band subtracted from the harder band, relative to their sum. The hardness ratios for the medium to soft, hard to soft, and hard to medium bands only will be calculated.

- hard_[ms/hs/hm]_lolim_[ms/ma] - The lower limits of the above hardness ratios. Taken together with hard_[ms/hs/hm]_hilim_[ms/ma], this encompasses the 68% confidence limit. These are *not symmetric error bars*.

- hard_[ms/hs/hm]_hilim_[ms/ma] - The upper limits of the above hardness ratios. Taken together with hard_[ms/hs/hm]_lolim_[ms/ma], this encompasses the 68% confidence limit. These are *not symmetric error bars*.

The following property is new to Catalog Release 2.

- hard_[ms/hs/hm]_prob_[ms/ma] - The probability distributions for the medium to soft, hard to soft, and hard to medium energy bands, stored as hardness vs. probability density, in a similar manner as for the hardness ratio probability density for the individual observations. (This probably could be stored along with the Master Source and Master Average aperture flux probability densities.)

The next step is to determine whether or not the Master Source or Master Average show color variability. This follows the procedures of §3.2, specifically Eqs. 5–10, except with ${P_{H_{xy}}}^i({H_{xy}}^i)$ substituting for ${P_x}^i({F_x}^i)$; $\overline{H_{xy}}^i$ and ${\sigma_{xy}}^i$ substituting for $\overline{F_x}^i$ and ${\sigma_x}^i$, and H_{xy}^{max} and ${H_{xy}}^{imax}$ substituting for F_x^{max} and ${F_x}^{imax}$. That is, we determine the mean and variances of the individual hardness ratio probability distributions (calculated following the outline of §2.2), and then determine the best fit mean hardness ratio and the best fit individual hardness ratios using these same probability distributions. We then use these best fit hardness ratios to calculate $\log$ likelihoods, $D/(N-1)$, and a p value for the proposition that the hardness ratios remain unchanged for the individual observations that comprise the Master Source.

This suggests two new sets of Master Source and Master Average properties related to the hardness ratios for Release 2 of the Catalog:

- var_inter_hard_[ms/hs/hm]_prob_[ms/ma] - The p value as defined by Eq. 10 using hardness ratio probability densities instead of aperture flux probability densities.

- var_inter_hard_[ms/hs/hm]_sigma_[ms/ma] - The largest difference between mean hardness ratio and an individual observation hardness ratio, relative to the standard deviation of hardness ratio in that individual observation (i.e., the hardness ratio equivalent of Eq. 8).

From these we will create a hardness ratio variability flag for Master Sources and Master Averages, exactly analogous to the same named property from Release 1.1 of the Catalog:

- var_inter_hard_flag_[ms/ma] - A Boolean set to FALSE if var_inter_hard_[ms/hs/hm]_prob_[ms/ma], for all values, lie below a given threshold (*the value of this threshold is to be determined*), and is otherwise set to TRUE.

4 Bayesian Blocks, and Other Groupings

We have discussed a number of other groupings of the individual objects. These have included: determining source probabilities on a per-stack basis (this removes any ambiguity in the Master Match process, by limiting the individual observations to those for which a successful fit has been performed to the joint data), a Bayesian Blocks analysis for flux-ordered observations in a single stack, a Bayesian Blocks analysis for flux-ordered observations for a single Master Average[6], and a Bayesian Blocks analysis for time-ordered observations for a single Master Average.

The constituents of individual stacks are well-defined and are not dependent upon subsequent analysis (in contrast to the Bayesian Blocks decompositions). Thus, it is straightforward to apply all the methods of §3, *in an identical manner as for the master sources and master averages*, to determine variability and hardness ratio properties for a "stack source". There should be no recoding necessary. Below I list the properties to record that are analogous to those calculated for the Master Source and Master Averages, and designate them with the suffix `_stk` (again, merely a notational convenience for purposes of this memo).

As regards the Bayesian Blocks analyses, the fact that these already are in of themselves variability analyses, we merely shall set flags to indicate blocks have been found in a given band. However, there is no particular variability analysis that we are performing that becomes more sensitive when considering a combined block. (We can, however, also indicate whether or not the block is composed of individual observations that are themselves variable, as we do for Master Sources.) However, the blocks can have hardness ratio values determined, following the exact procedures in §3.3, using the aperture photometry probability densities generated for each block.

4.1 Variability

For the stack-sources, we should record: `var_inter_index_x_stk`, `var_inter_prob_x_stk`, `var_inter_sigma_x_stk`, `ks_intra_prob_x_stk`, `kp_intra_prob_x_stk`, `var_intra_index_x_stk`, `var_intra_prob_x_stk`, `var_flag_stk`.

For the Bayesian Blocks analyses, we should set Boolean flags to `TRUE` for any Blocks analysis in a given energy band that indicates a subdivision. Specifically, we can set the following flags (names subject to modification):

- `blx_flux_stk_flag_[s/m/h/w]` - Results of the flux-ordered Bayesian Blocks analysis, performed on a stack.
- `blx_flux_ma_flag_[s/m/h/w]` - Results of the flux-ordered Bayesian Blocks analysis, performed on the Master Average.
- `blx_lc_ma_flag_[s/m/h/w]` - Results of the time-ordered Bayesian Blocks analysis, performed on the Master Average.

Note that the current plan is to perform Bayesian Blocks analyses in the `s`, `m`, `h`, and `w` bands.

4.2 Hardness Ratios

For the stack-sources, we should record: `hard_[ms/hs/hm]_stk`, `hard_[ms/hs/hm]_lolim_stk`, `hard_[ms/hs/hm]_hilim_stk`, `hard_[ms/hs/hm]_prob_stk`, `var_inter_hard_[ms/hs/hm]_prob_stk`, `var_inter_hard_[ms/hs/hm]_sigma_stk`, `var_inter_hard_flag_stk`.

[6] We originally discussed this possibility before developing a concept of Master Source vs. Master Average. Master Average covers all stacks where a source is detected, and it is likely that the Master Source definition will be in part derived from a Bayesian Blocks analysis of the Master Average. Therefore, I am presuming that at the Master level, the Bayesian Blocks analysis will be applied to the Master Average.

For the Bayesian Blocks analyses, for any blocks for which we record aperture photometry values, we should also record: hard_[ms/hs/hm]_#, hard_[ms/hs/hm]_lolim_#, hard_[ms/hs/hm]_hilim_#, and hard_[ms/hs/hm]_prob_#, where these quantities are calculated and stored for *each block of the Bayesian Blocks decomposition* (hence the designation of _# at the end, to indicate that there are multiple values for each Bayesian Block group). The Bayesian Blocks analyses themselves will ensure that the var_inter_hard properties are not required in these cases.

5 Suggested Priorities

§6 gives a rough overview as to what concepts are essentially the same from Catalog Release 1.1, and hence require less detailed development, and which are newer. Most of the variability quantities for single observations are simple reiterations of prior work, albeit with new regions. The next most important set of quantities to tackle are the hardness ratio quantities for single observations. We know what the individual observation regions and aperture flux probabilities are going to be, and the mechanics of the hardness ratio calculations in these cases directly translates over to the cases with more complex groupings (stack, Master Source, Master Average, and the Bayesian Blocks groupings). The next most well-defined grouping is the individual stack, followed by the Master Average (all overlapping stacks with positive source detections associated with each other via the match process). The Bayesian Blocks and Master Source properties are the groupings whose final definitions still need to be agreed upon in conjunction with the aperture photometry specifications. Finally, I am currently not aware of plans that utilize regions along the lines of those described in §3.1. Such regions must be created at any rate must be created as a separate post-processing step, thus should be relegated to the lowest priority for further development.

6 Summary

In the tables below, I summarize all the variability and hardness ratio properties discussed in this memo. I further distinguish whether they are to be applied to individual observations (o), an individual stack (stk), the group of observations comprising the Master Source (mstr src), the group of observations comprising the Master Average (mstr avg), or one of the Bayesian Blocks decompositions: flux ordered observations in a stack (blx [flx/stk]), flux ordered observations in a Master Average (blx [flx/mstr]), or time ordered observations in a Master Average (blx [lc/mstr]). As described above, for the most part variability quantities are calculated from *individual* observations, even when calculating a "group quantity", while the hardness ratio quantities are calculated from the relevant *group* photometry properties.

variability	obs	stk	mstr src	mstr avg	blx [flx/stk]	blx [flx/mstr]	blx [lc/mstr]
`ks_prob_x`	o						
`kp_prob_x`	o						
`var_prob_x`	o						
`var_mean_x`	o						
`var_min_x`	o						
`var_max_x`	o						
`var_index_x`	o						
`dither_warning_flag`	o						
`var_code`	o						
`var_inter_index_x_y`		1	1	1			
`var_inter_prob_x_y`		1	1	1			
`var_inter_sigma_x_y`		1	1	1			
`ks_intra_prob_x_y`		o	o	o			
`kp_intra_prob_x_y`		o	o	o			
`var_intra_index_x_y`		o	o	o			
`var_intra_prob_x_y`		o	o	o			
`var_flag_y`		1	1	1			
`var_inter_hard_x_prob_y`[1]		2	2	2			
`var_inter_hard_x_sigma_y`[1]		2	2	2			
`var_inter_hard_flag_y`[1]		2	2	2			
`blx_flux_stk_flag_x`					2		
`blx_flux_ma_flag_x`						2	
`blx_lc_ma_flag_x`							2

`_x` refers to possible energy bands (`u/s/m/h/b/w` – see text for relevant bands for each quantity)
`_y` refers to possible type of grouping of individual observations, `_stk` (stack), `_ms` (Master Source), or `_ma` (Master Average)
[1] Both a variability and a hardness ratio property.
o – quantity essentially identical to that from the original catalog, Release 1.1. (The quantity may be being applied in new circumstances, e.g., a stack source.)
1 – quantity similar to one from Catalog Release 1.1; however, either based upon different input quantities, or new computational definitions as outlined in the memo. (The quantity may be being applied in new circumstances, e.g., a stack source.)
2 – quantity/concept new to Catalog Release 2, as outlined in this memo.

hardness	obs	stk	mstr src	mstr avg	blx [flx/stk]	blx [flx/mstr]	blx [lc/mstr]
hard_z	1						
hard_z_lolim	1						
hard_z_hilim	1						
hard_z_prob	2						
hard_z_y		1	1	1			
hard_z_lolim_y		1	1	1			
hard_z_hilim_y		1	1	1			
hard_z_prob_y		2	2	2			
var_inter_hard_x_prob_y[1]		2	2	2			
var_inter_hard_x_sigma_y[1]		2	2	2			
var_inter_hard_flag_y[1]		2	2	2			
hard_z_#[2]							
hard_z_lolim_#[2]					2	2	2
hard_z_hilim_#[2]					2	2	2
hard_z_prob_#[2]					2	2	2

_x refers to possible energy bands (u/s/m/h/b/w – see text for relevant bands for each quantity)
_y refers to possible type of grouping of individual observations, _stk (stack), _ms (Master Source), or _ma (Master Average)
_z refers to possible hardness ratio comparison bands, _ms (medium to soft), _hs (hard to soft), or _hm (hard to medium)
[1] Both a variability and a hardness ratio property.
[2] Calculated and stored for *each* block (numbered by #) in the resulting Bayesian Blocks decomposition.
o – quantity essentially identical to that from the original catalog, Release 1.1. (The quantity may be being applied in new circumstances, e.g., a stack source.)
1 – quantity similar to one from Catalog Release 1.1; however, either based upon different input quantities, or new computational definitions as outlined in the memo. (The quantity may be being applied in new circumstances, e.g., a stack source.)
2 – quantity/concept new to Catalog Release 2, as outlined in this memo.