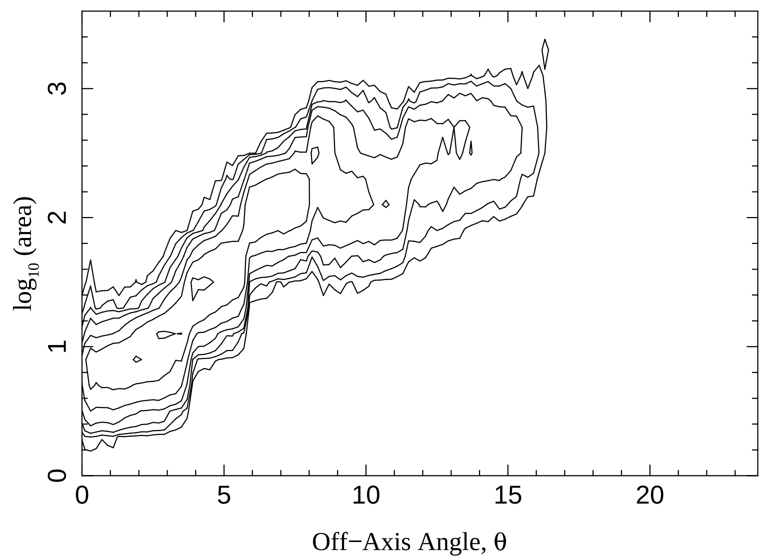

Figure 1: For the detected sources in the current CSC, the source extraction area as a function of off-axis angle. The contours are density of detections in the x-y plane, in factor of two increments (with the highest contour being close to the peak density).

Background Filtering for Release 2 of the Chandra Source Catalog

Michael Nowak, MIT-CXC, May 1, 2012

In trying to determine what background flare level to remove, I am going to harken back to my memo of June 25, 2007. That memo relied on simple Poisson statistics, and an estimate of the signal-to-noise ratio in the Gaussian limit. Even though we are going with MLE for release 2 of the catalog, for simplicity I will continue to use the Gaussian approximation. It should still be qualitatively correct, and quantitatively not horrible. The goal here will be, given the detections from the first catalog, to determine what would happen to signal-to-noise ratios if good time intervals were reduced to remove times of higher background.

The bottom line result is that, barring extremely large background flares, almost *any* removal of time decreases the S/N of the *majority* of *currently* detected sources. This is for the simple reason that most sources are detected within $\approx 9'$ of the optical axis, and have relatively small source extraction areas (i.e., mean radii $\lesssim 6$ pixels), coupled with the fact that *Chandra* has intrinsically low background. Thus, the S/N ratio in most sources is simply the square root of the detected counts. Extremely large flares are required to have the background term become significant.

This is partly illustrated in Fig. 1, where I show the density of sources as a function of off-axis viewing angle and the logarithm of the source extraction region area. This is further iterated in Fig. 2a, where I show the existing $(S/N)^2$ as a function of off-axis angle. (Here and throughout this work, all figures will be using the b-band results from the catalog.)

We can quantify how the signal-to-noise changes if we include times of flaring. We have for non-flare times:

$$\left(\frac{S}{N}\right)^2 = \frac{\mathcal{S}^2 T^2}{\mathcal{S}T + 2\mathcal{R}\mathcal{B}T} \qquad (1)$$

where T is the observation time, $\mathcal{S}$ is the background subtracted source rate, $\mathcal{B}$ is the background rate, and $\mathcal{R}$ is the square of the ratio of the source region area to the background region area. Now imagine an additional time, T_F, characterized by an *additional* background rate, $\mathcal{B}_F$. If we include this time in the

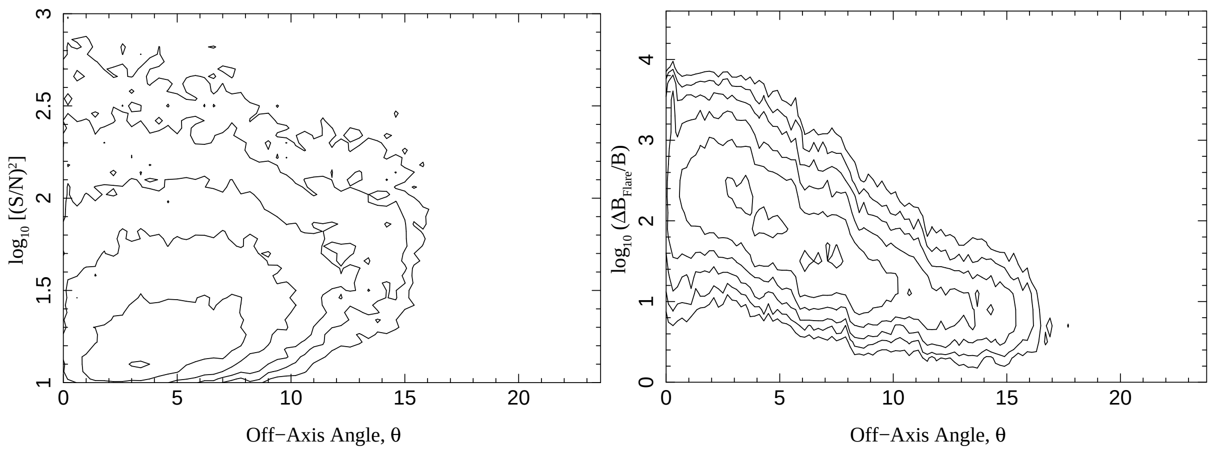

Figure 2: Left: The $\log_{10}$ of the squared signal-to-noise as a function of off-axis angle in the current CSC. Right: For the detected sources in the current CSC, the ratio of *additional* background flare level required such that removal of that flare results in an increased signal-to-noise. The contours are density of detections in the x-y plane, in factor of two increments (with the highest contour being close to the peak density).

source detection, we then have:

$$\left(\frac{S}{N}\right)^2_F = \frac{\mathcal{S}^2(T+T_F)^2}{\mathcal{S}(T+T_F)+2\mathcal{R}(\mathcal{B}(T+T_F)+\mathcal{B}_F T_F)} \quad . \tag{2}$$

Comparing the two, we can show that signal-to-noise *decreases* if

$$\frac{\mathcal{B}_F}{\mathcal{B}} > \left(\frac{T+T_F}{T}\right)\left(1+\frac{\mathcal{S}}{2\mathcal{R}\mathcal{B}}\right) \quad . \tag{3}$$

Thus, the above (in the limit that $T_F \rightarrow 0$, as assumed in all the figures shown here) becomes the minimum flare amplitude for which it becomes beneficial to remove the flare time. This is plotted in Fig. 2b for the existing source catalog. Again, most currently detected sources would require very large background flares in order for removing the flare time to lead to an increased signal-to-noise.

Assuming one does excise the flare times, one can also ask what is the fractional change in $(S/N)^2$. This is given by:

$$\Delta\left(\frac{S}{N}\right)^2 / \left(\frac{S}{N}\right)^2 = 1 - \frac{T+T_F}{T}\left(1+\frac{2\mathcal{R}\mathcal{B}_F}{\mathcal{S}+2\mathcal{R}\mathcal{B}}\left(\frac{T_F}{T+T_F}\right)\right)^{-1} \quad . \tag{4}$$

We can check this against the current catalog by comparing the current $(S/N)^2$ to what it would be, say, for a flare that were an additional 10% of the observing time with an *additional* $\mathcal{B}_F$ of $14 \times \mathcal{B}$. This is shown in Fig. 3. Most sources would have a *decreased* $(S/N)^2$, but only by $\approx 10\%$ (i.e., the fractional duration of the flare). However, there are some sources whose $(S/N)^2$ would improve, and they can potentially improve by wider margin than 10%. This would tend to be the lower $(S/N)^2$ sources.

Fig. 4 shows the fraction of existing catalog sources whose signal-to-noise would improved if background flares of a given fractional increase in amplitude were excised from the data. In general, it takes fairly large flares before the majority of the *existing* catalog could be improved. We do see that there is a distinct bend in the curve for $\mathcal{B}_F \gtrsim 10$. Also shown in Fig. 4 is the fraction of improved source detections as a function of $(S/N)^2$ for different assumed background flare levels. Note that these curves drop below the nominal detection threshold for the catalog, since here we are considering only the b-band. These are

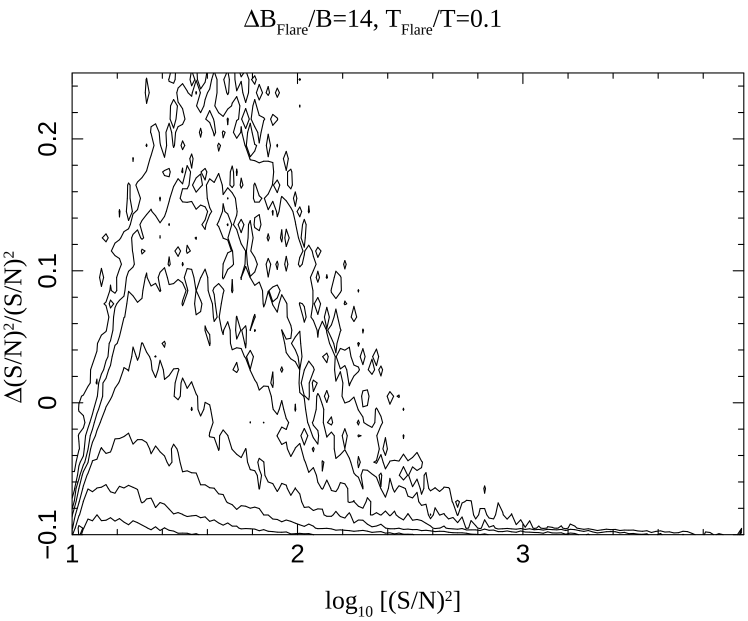

Figure 3: For the detected sources in the current CSC, the fractional change in the square of their signal-to-noise ratio, as a function of the $\log_{10}$ of their squared signal-to-noise, if one removes a factor of 10 (additional flux) flare that has a duration of 10% of the non-flare time. The contours are density of detections in the x-y plane, in factor of two increments (starting close to the peak density).

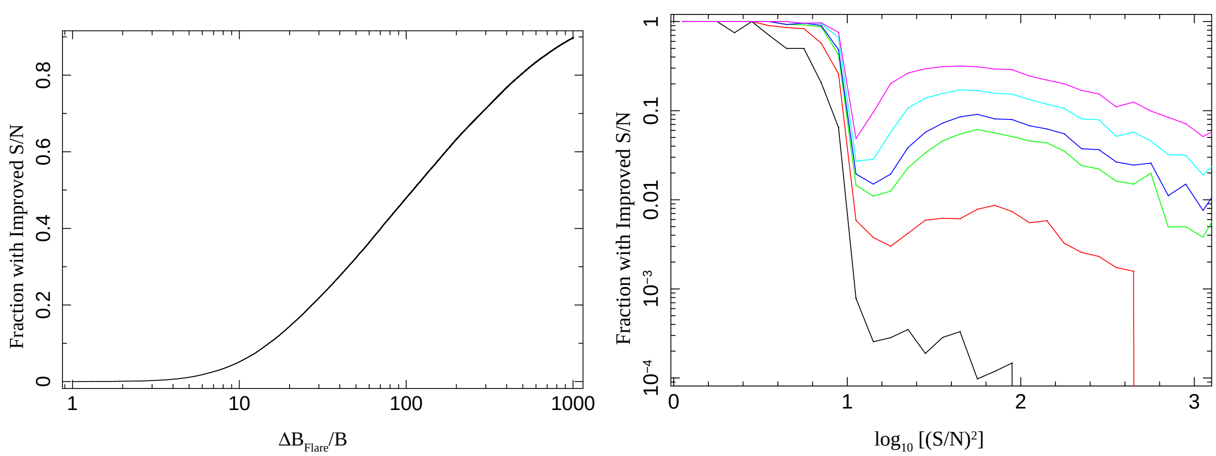

Figure 4: Left: The fraction of CSC sources whose signal-to-noise ratio would improve if a background flare with a given change in level, relative to the quiescent background, is removed. Right: The fraction of sources, as a function of the $\log_{10}$ of their squared signal-to-noise ration, whose signal-to-noise would improve if background flares of a given amplitude (measured as an *addition* relative to the quiescent background level) were removed. Lines are for relative additions of 2, 4, 8, 10, 16, and 32 times the quiescent level.

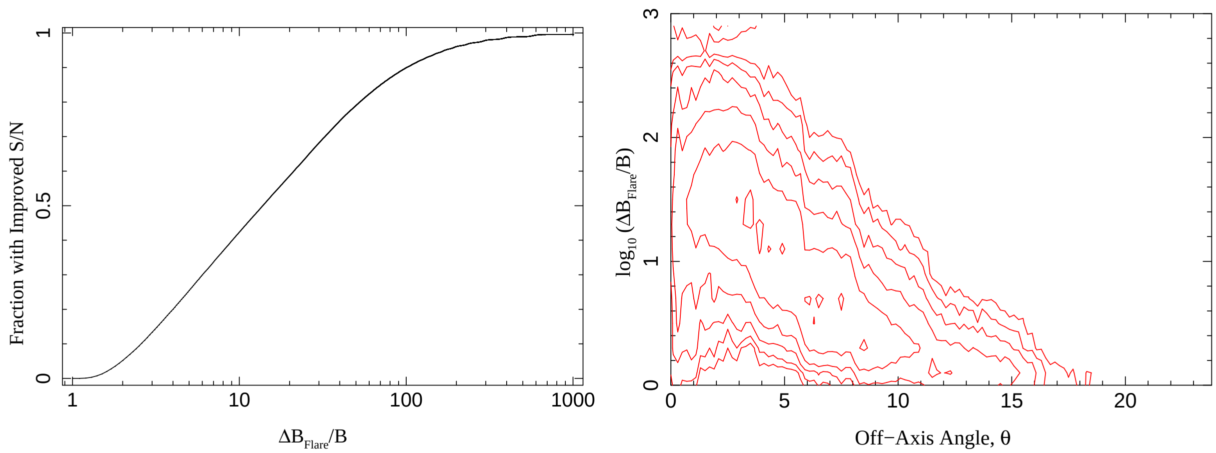

Figure 5: Assuming a catalog of faint sources with the same locations, areas, and background rates as the current CSC, but presuming a $(S/N) = 2$. Left: The fraction of these fake, faint sources whose signal-to-noise ratio would improve if a background flare with a given change in level, relative to the quiescent background, is removed. Right: For these faint, fake sources, the minimal fractional change in background level that is required, as a function of off-axis angle, in order for the flare removal to improve the signal-to-noise.

obviously sources whose detections rely on other bands. Note also that in general one finds, unsurprisingly, a much larger fraction of these faint sources with an improved $(S/N)^2$ when removing background flares.

Part of the driving force behind the next catalog is to go after even fainter sources, so let us consider a hypothetical set of sources with $(S/N)^2 = 4$. (I realize that this perhaps fainter than we plan to go – perhaps we are more realistically planning on going down to only $(S/N)^2 = 6$, i.e., 6 count sources for the most part.) To do this, I presumed a catalog with the same background count rates and source sizes and distributions as the current catalog, but simply replaced all the $(S/N)^2$ values by 2. (Throughout, the $(S/N)^2$ is calculated via equation 1, which *usually*, but not always, reduces down to $(S/N)^2 \approx$ the number of source counts.) Fig. 5 shows the fraction of these sources whose $(S/N)^2$ is improved by removing flares of a given amplitude. More than half of these fake sources have improved signal-to-noise if we remove flares with amplitudes $\mathcal{B}_F \gtrsim 14$. As shown in Fig. 5b, we still might be slightly decreasing $(S/N)^2$ for sources within $\lesssim 5'$ of the optical axis, but overall we will be improving the signal-to-noise. Again, it is always a trade-off between the benefits to larger size/fainter objects vs. slight decrease in signal-to-noise to smaller size/brighter objects.

To give some perspective, in the current catalog, only $\approx 1.4\%$ of the *sources* (as opposed to the observations, which is a different measure) have any significant amount of flare time removed under our current, somewhat conservative, criteria. The removed time ranges from $\approx$ 5–85% of the observation time, with a (detected source) average of $\approx$ 17% of the time removed[1] Flare time removal has not been a major factor in the catalog to date. (However, we have had a fairly quiet Sun for much of the time period covered by the existing catalog.)

This then begs the question of how many additional sources do we expect to gain in the next release of the catalog? As a rough cut at this, in Fig. 6 I plot the number of sources in the current catalog as a function of detected $\log_{10}[(S/N)^2]$. There is a rollover near our catalog inclusion criteria of $(S/N)^2 \approx 10$. The tail of this distribution[2] is reasonably well fit by $5 \times 10^6 \, \exp(-2.2 \, \log_{10}[(S/N)^2])$. Extrapolating that back to

[1]I estimated this by taking the ratio of `livetime` to `gti_elapse` time for detected sources whose dither warning flag was `FALSE`, and only considered sources where this ratio was $< 94.5\%$.

[2]Whereas it is true that the existing "$\log N$-$\log S$" curve for the *Chandra* source catalog is comprised of a disparate set of populations and kinds of sources, each with their own distributions, and there are complex selection criteria involved, I am pre-

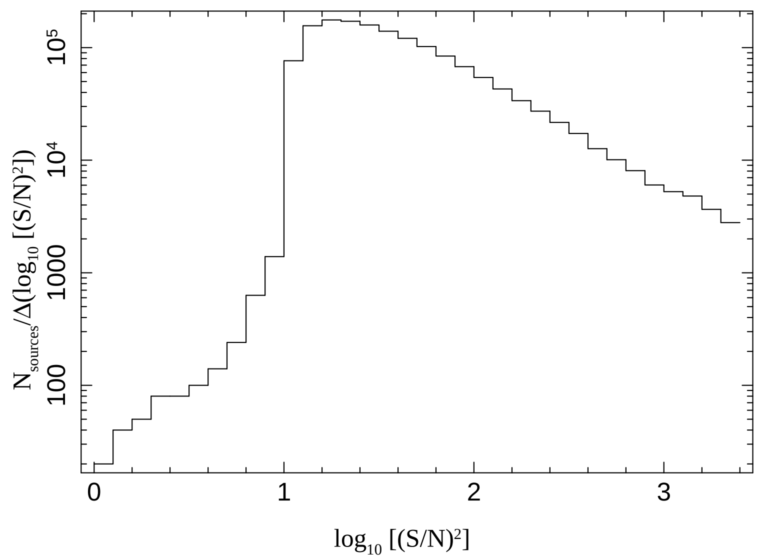

Figure 6: Assuming a catalog of faint sources with the same locations, areas, and background rates as the current CSC, but presuming a $(S/N) = 2$. Left: The fraction of these fake, faint sources whose signal-to-noise ratio would improve if a background flare with a given change in level, relative to the quiescent background, is removed. Right: For these faint, fake sources, the minimal fractional change in background level that is required, as a function of off-axis angle, in order for the flare removal to improve the signal-to-noise.

$(S/N)^2 = 4$ implies that there are 600k (!!!) sources total to be found. However, statistical fluctuations mean that we will have a similar rollover as to the one we see here, which would instead imply that there are a total of ≈ 350k sources. If instead we cutoff at $(S/N)^2 \approx 6$, and again accounting for a similar rollover, we might expect a total of ≈ 250k sources.

This means that whatever criterion we choose predominantly will affect the ≈ 100k–200k sources yet to be discovered. It is their (low) signal-to-noise that is of the greatest concern, and the results of Fig. 5 take precedence. Again, even for these low signal-to-noise ratios, most sources are fairly tolerant to large background flares. Somewhere around a value of $\mathcal{B}_F \approx 14$ is where more than half the faintest sources have improved signal-to-noise after background flare removal, and it is near the point where we begin to see a rise in the fraction of current catalog sources that will have improved signal-to-noise. (Again, however, *most* current sources would take a slight hit in $(S/N)^2$ for flare removal.) For the faint sources, this 50% threshold value for the background flare rises slightly if our catalog cutoff is something closer to $(S/N)^2 = 6$.

This leads me to the following suggestions:

- As for the current catalog, we *should not* use an iterative scheme, i.e., identify sources, estimate background, filter background, re-identify sources, re-estimate background, filter again, …. A single pass on the background should suffice.
- We could be less aggressive in background flare filtering. Filtering times where the *total* background is $\gtrsim 15\times$ the quiescent background would lead to an improvement in signal-to-noise for the majority of yet to be discovered faint sources.
- Given the fairly limited amount of time filtered in the current catalog, we can instead choose to remain

suming that so long as there aren't radical changes in the way users perform observations with *Chandra*, this is a decent *empirical* distribution that provides a reasonable estimate of the average population to be found. It is essentially the philosophy of "bootstrap" statistical tests — when your real distribution is too complex to model properly, the fairest approximations are actual measurements.

fairly conservative in our filter criteria. This wouldn't have a huge hit on the catalog exposure, and we would avoid any "law of unintended consequences". E.g., all of the above presumes that the background is *well-characterized*. Does characterizing the background become more difficult as we admit longer periods of larger background flares? That issue is not considered here.

- However, the same argument that says so little of our source-summed exposure time is affected by flares to begin with, we can take a risk at being looser with the flare criteria, and again not adversely affect the total statistics of the catalog.

My own inclination is to allow more time into the catalog by only filtering large flares, $\geq 15\times$ the quiescent level. I am generally in favor of letting the user downstream adopt a more stringent set of criteria should they desire.