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

MEMORANDUM
October 7, 2010

To:	Jonathan McDowell, SDS Group Leader
From:	Glenn E. Allen, SDS
Subject:	The afterglow detection efficiency of acis_run_hotpix
Revision:	1.1
URL:	http://space.mit.edu/CXC/docs/docs.html#aft_det_eff
File:	/nfs/cxc/h2/gea/sds/docs/memos/afterglow_detection_efficiency_1.1.tex

1 Introduction

A cosmic-ray "afterglow" is produced when a large amount of charge is deposited on a CCD by a cosmic ray. Most of the charge is clocked off of the CCD in a single frame. However, a small amount can be captured in charge traps, which release the charge relatively slowly. As a result, a sequence of events can appear in a single detector pixel over a few frames[1] as the trapped charge is released. A selected sample of afterglows is plotted in Figure 1. Notice that the events in an afterglow need not occur in consecutive frames. There can be gaps of a few frames between events. Also note that while the summed pulse heights of the events typically decrease from frame to frame, the pulse heights can increase toward the end of an afterglow when the pulse heights are relatively low.

2 Afterglow Detection Algorithms

To date, two algorithms have been used by the CXC to identify cosmic-ray afterglows. The first algorithm was implemented in the CIAO tool acis_detect_afterglow and used for pipeline processing from the summer of 2000 to the fall of 2004. This algorithm searches for occasions when events are detected in two or more consecutive frames on the same CCD pixel. While the events are flagged as potential cosmic-ray afterglows and excluded from Level 2 event-data files, the corresponding pixels are not included in the observation-specific bad-pixel file. This algorithm finds many afterglow events, but at the expense of discarding X-ray events associated with real astrophysical sources. The fraction of the source events that are discarded depends on the brightness and variability of the source.

In an attempt to minimize the loss of source events, another algorithm[2] was developed and implemented in the CIAO tool acis_run_hotpix.[3] This second algorithm searches for detector pixels that have an unusually large number of events. Suspicious pixels are added to the observation-specific bad-pixel file only if the neighboring pixels do not have a significant excess of events. This condition helps insure that events associated with dithered sources are not discarded. Events associated with afterglows are flagged and excluded from Level 2 event-data files. The newer algorithm has been used for pipeline processing (and reprocessing) since

[1] For the data sets listed in Table 1, none of the identified afterglows has a duration longer than 17 frames (53 s).
[2] For a detailed description of the algorithm, see http://space.mit.edu/CXC/docs/docs.html#hotpix.
[3] Actually, acis_run_hotpix is a wrapper around the tools acis_find_hotpix, acis_classify_hotpix and acis_build_badpix.

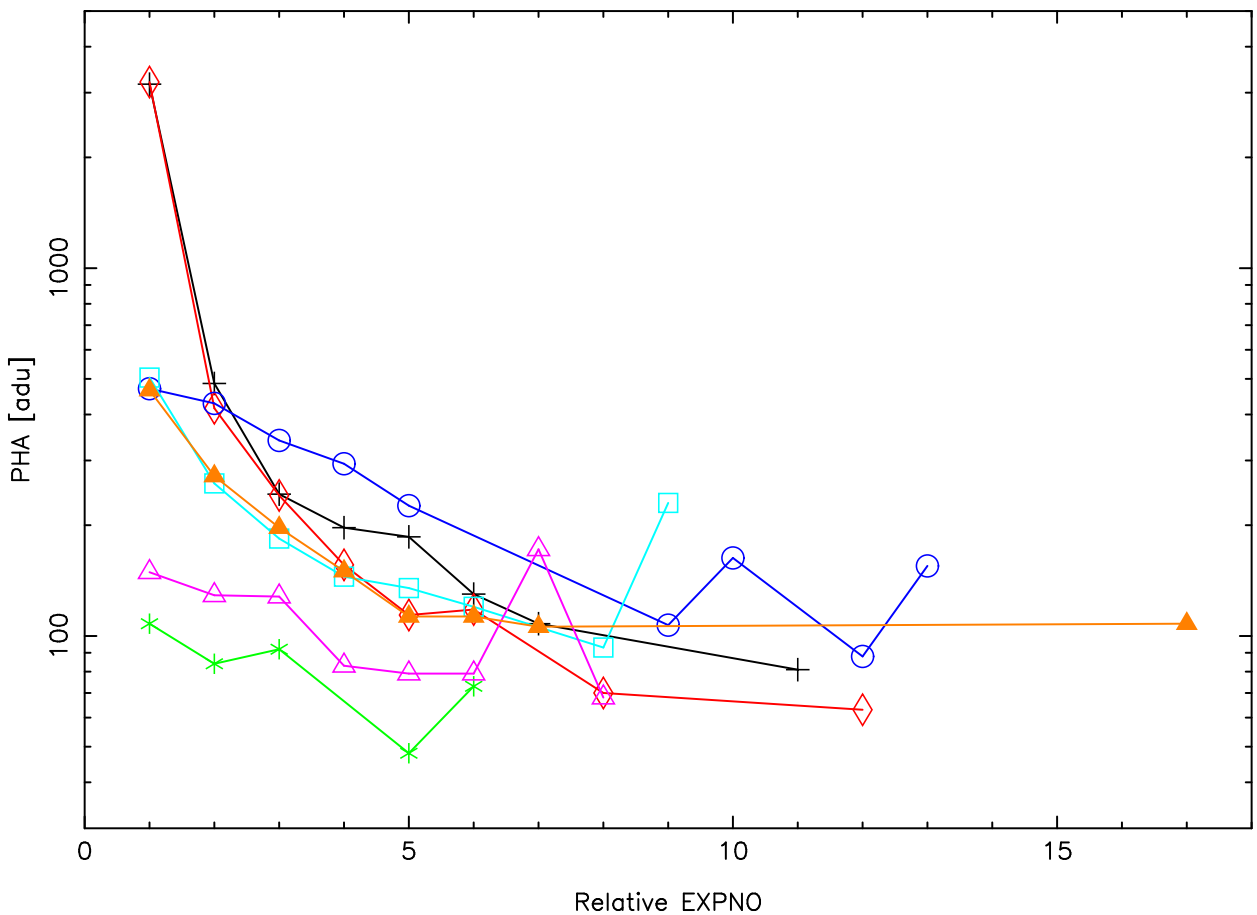

Figure 1: The pulse height v. relative frame number for a selected sample of afterglows. Note that events need not occur in consecutive frames. There can be gaps of a few frames without events in the middle of an afterglow. While the summed pulse height typically decreases from frame to frame, it can increase toward the end of an afterglow when the pulse height is relatively low.

Table 1: ACIS Data Used

No.	OBS_ID	DATE-OBS	DETNAM	DATAMODE	No. events (Level 1)	No. events (Level 2)	EXPOSURE (s)
1	5841	2005-03-14	ACIS-01236	VFAINT	1,393,605	118,614	44,457
2	5842	2005-03-16	ACIS-01236	VFAINT	1,441,036	123,932	46,428
3	5843	2005-03-19	ACIS-01236	VFAINT	1,366,983	117,729	44,463
4	5844	2005-03-21	ACIS-01236	VFAINT	1,397,942	120,062	45,852
5	7999	2006-11-25	ACIS-012367	VFAINT	1,373,989	207,567	29,660
6	8000	2007-05-26	ACIS-012367	VFAINT	2,239,625	337,188	46,668
7	8001	2007-04-02	ACIS-012367	VFAINT	2,370,265	358,654	50,092
8	8002	2006-12-19	ACIS-012367	VFAINT	1,222,501	188,204	29,651
9	8003	2007-04-02	ACIS-012367	VFAINT	2,332,759	425,946	47,043
10	8004	2006-11-27	ACIS-012367	VFAINT	746,425	109,178	15,734
Total					15,885,130	2,107,074	400,048

the fall of 2004. While it is relatively gentle on astrophysical sources, it does let some afterglows "slip through the cracks." Users have reported that some of the afterglows that are missed are subsequently identified as potential sources by the CIAO tool wavdetect.

This memo describes an analysis of the efficiency with which acis_run_hotpix identifies afterglows and summarizes the features of the afterglows that are missed by acis_run_hotpix and found by wavdetect. This study has highlighted the need for an improved afterglow-detection algorithm. A specification for such an algorithm will be presented elsewhere.

3 Test Data

The data used to test the efficiency of acis_run_hotpix are listed in Table 1. These data, which are from some of the "COSMOS" and "EGS" observations, were obtained in the spring of 2005, late 2006 and the spring of 2007. The observations were performed using the ACIS-I array and one or two of the ACIS-S CCDs. The very-faint data mode was used for all of the observations. Collectively, the ten OBS_IDs include 2.11 million events (Level 2) and 400.0 ks of observing time.

4 Analysis

Each data set in Table 1 was reprocessed in the same manner using the following sequence of steps.

1. The CIAO tool acis_run_hotpix was used to create a new observation-specific bad-pixel file. Pixels identified as having cosmic-ray afterglows are included in the output file.

2. The CIAO tool acis_process_events was used to reprocess the Level 1 event data with the new bad-pixel file. Events associated with the afterglows identified by acis_run_hotpix have STATUS bit 16 set to one.

3. The CIAO tool dmcopy was used to remove "bad" events, including those that are identified as being part of an afterglow.

4. The CIAO tool wavdetect was used to search for potential astrophysical sources.

5. A set of S-Lang functions was used to examine each potential wavdetect "source." These functions

   a. identify the Level 1 events within a two pixel (0.984 arcsec) radius of a source location,

   b. histogram the chip coordinates of the selected events,

   c. identify the pixel with the largest number of events as the pixel on which an afterglow may have occurred,

Table 2: Numbers of Potential Afterglows and Sources

OBS_ID	wavdetect $N_{\rm tot}^{\rm wav}$	wavdetect $N_{\rm aft}^{\rm wav}$	acis_run_hotpix $N_{\rm aft}^{\rm arh}$
5841	66	12	166
5842	61	10	179
5843	59	11	172
5844	60	10	166
7999	60	13	146
8000	60	13	160
8001	83	18	175
8002	66	16	144
8003	66	11	176
8004	41	7	91
Total	622	121	1575

d. determine if there is a set of two or more events on the pixel for which the time between consecutive events does not exceed ten frames (i.e. the minimum afterglow criteria),

e. identify the frame numbers (i.e. EXPNO values) associated with the beginning and end of the potential afterglow,

f. count the number of events in the afterglow,

g. estimate the average number of background events per pixel associated with the afterglow,

h. calculate the number of trials performed (i.e. the number of pixels searched for afterglows) and

i. calculate the pre- and post-trials significances of the afterglow.

For each data set, Table 2 includes the total number of potential astrophysical sources and afterglows identified by wavdetect ($N_{\rm tot}^{\rm wav}$). The column $N_{\rm aft}^{\rm wav}$ represents the subset of this number that satisfy the minimum afterglow criterion described in step 5d. Therefore, $N_{\rm tot}^{\rm wav} - N_{\rm aft}^{\rm wav}$ is an estimate of the number of potential astrophysical sources identified by wavdetect. The column $N_{\rm aft}^{\rm arh}$ is the number of afterglows that were identified by acis_run_hotpix. These afterglows were removed from the data before it was processed with wavdetect.

5 Discussion

An inspection of Table 2 suggests that

- about 7.1% (i.e. $121/(121+1575)$) of the afterglows are missed by acis_run_hotpix and
- about 19% (i.e. $121/622$) of the potential sources identified by wavdetect are afterglows.

These percentages should be regarded as rough estimates because they are based on five assumptions that may or may not be valid.

1. One assumption is that all 1,575 afterglows identified by acis_run_hotpix are legitimate afterglows. A visual inspection of the pulse-height and flight-grade values v. frame number for these afterglows suggests that this assumption is essentially valid.

2. Another assumption is that all of the afterglows that are missed by acis_run_hotpix are identified by wavdetect. This assumption was not tested because an improved version of the afterglow-detection algorithm has not been coded.

3. A third assumption is that the 121 wavdetect "sources" identified as afterglows are legitimate afterglows. A visual inspection of the pulse-height and flight-grade values v. frame number for each of these afterglows suggests that a few of the afterglows with small numbers of events may not be afterglows. Otherwise the afterglows appear to be genuine.

4. A fourth assumption is that all real sources with a significant number of events were identified by wavdetect. This assumption was not tested because it is beyond the scope of the present study. The present study is focused on tests of the tool acis_run_hotpix.

5. The last assumption is that none of the 501 (i.e. 622−121) non-afterglow sources identified by wavdetect is a false detection. Again, this assumption was not tested because it is beyond the scope of the present study.

In addition to these five assumptions, the results may be sensitive to the cosmic- and x-ray background flux spectra and the clocking mode of the ACIS CCDs.

Despite these cautionary comments, it is clear that the tool acis_run_hotpix does not identify all real afterglows. The primary reason that some of the afterglows are missed is that the post-trials probabilities (P_1) of the afterglows are too large to be considered significant. Since the default probability threshold is 0.001,[4] this means that the afterglows that are missed by acis_run_hotpix have $P_1 > 10^{-3}$. Here,

$$P_1 = 1 - (1 - P_0)^{N_\mathrm{t}}\,, \tag{1}$$

where N_t is the number of trials (i.e. the number of pixels searched for afterglows) and

$$P_0 = 1 - \left[\left(\sum_{i=0}^{N_\mathrm{a}-1} \frac{N_\mathrm{b}^i}{i!}\right) + \frac{1}{2}\frac{N_\mathrm{b}^{N_\mathrm{a}}}{N_\mathrm{a}!}\right] e^{-N_\mathrm{b}} \tag{2}$$

is the pre-trials, Poisson probability of obtaining at least N_a events in an afterglow for a mean number of background events per pixel N_b. For example, consider the afterglow in Figure 2 that is marked with a red $\times$ surrounded by a green square. There are nine events in this afterglow (i.e. $N_\mathrm{a} = 9$). The expected number of events due to the background $N_\mathrm{b} = 0.396$. Using equation 2, one can show that the pre-trials probability $P_0 = 2.39 \times 10^{-10}$. Since 5,068,030 pixels were searched for evidence of afterglows, the post-trials probability of this afterglow is only $P_1 = 1.21 \times 10^{-3}$ (eqn. 1). Therefore, the afterglow is not significant enough to be identified by acis_run_hotpix. In terms of the afterglows plotted in Figure 2, the afterglow lies above the blue line. This line delineates the boundary between $P_1 < 10^{-3}$ (the lower, right-hand side) and $P_1 > 10^{-3}$ (the upper, left-hand side). Note that all but two of the potential afterglows identified by wavdetect lie above this boundary. That is, the afterglows missed by acis_run_hotpix (and found by wavdetect) have too few events (given N_b) for $P_1 < 10^{-3}$.

It is curious that two of the afterglows identified by wavdetect lie below the blue line in Figure 2. One of these has $N_\mathrm{a} = 11$ and $N_\mathrm{b} = 0.561$. The corresponding pre- and post-trials probabilities are $P_0 = 1.36 \times 10^{-11}$ and (for $N_\mathrm{t} = 6{,}102{,}143$) $P_1 = 8.29 \times 10^{-5}$, respectively. Therefore, this afterglow is significant. While the estimate of the number of background events per pixel (0.561) is high compared to the mean number of "background" events per pixel for the entire node (0.207), the difference is not large enough for the afterglow events to be considered part of an astrophysical source (see sec. 1.5.14 of http://space.mit.edu/CXC/docs/docs.html#hotpix). Therefore the afterglow should have been identified by acis_run_hotpix. The second of these two afterglows has $N_\mathrm{a} = 11$, $N_\mathrm{b} = 0.208$, $P_0 = 3.38 \times 10^{-16}$, $N_\mathrm{t} = 6{,}102{,}129$ and $P_1 = 2.06 \times 10^{-9}$. Since this afterglow is significant (and not associated with an astrophysical source), it too should have been identified by acis_run_hotpix. It is not clear why these afterglows were missed.

In other words, the number of events in the afterglows that are missed is too small for the detection algorithm to identify the afterglows as significant. As shown in Figure 3, acis_run_hotpix finds the vast majority of the afterglows that contain eight events or more. However, the efficiency with which afterglows are identified by the tool drops quickly as the number of events in an afterglow declines. As shown in Figure 2, the probability of identifying an afterglow also depends on the mean number of background events. Below the blue line, $P_1 < 10^{-3}$. In this region, all but two afterglows were identified by acis_run_hotpix. I don't know why two afterglows below the blue line were missed by acis_run_hotpix. According to the algorithm described in the spec, these two afterglows with $N_\mathrm{a} = 111$ should have been identified. Also note that some of the afterglows identified by acis_run_hotpix lie above the blue line. The reason that most of these afterglows lie above the line is that the algorithm used by acis_run_hotpix to calculate N_a is flawed. The value of N_a includes all of the events on the pixel, whether they are part of the afterglow or not.

[4] The probability threshold is adjustable using the parameter probthresh.

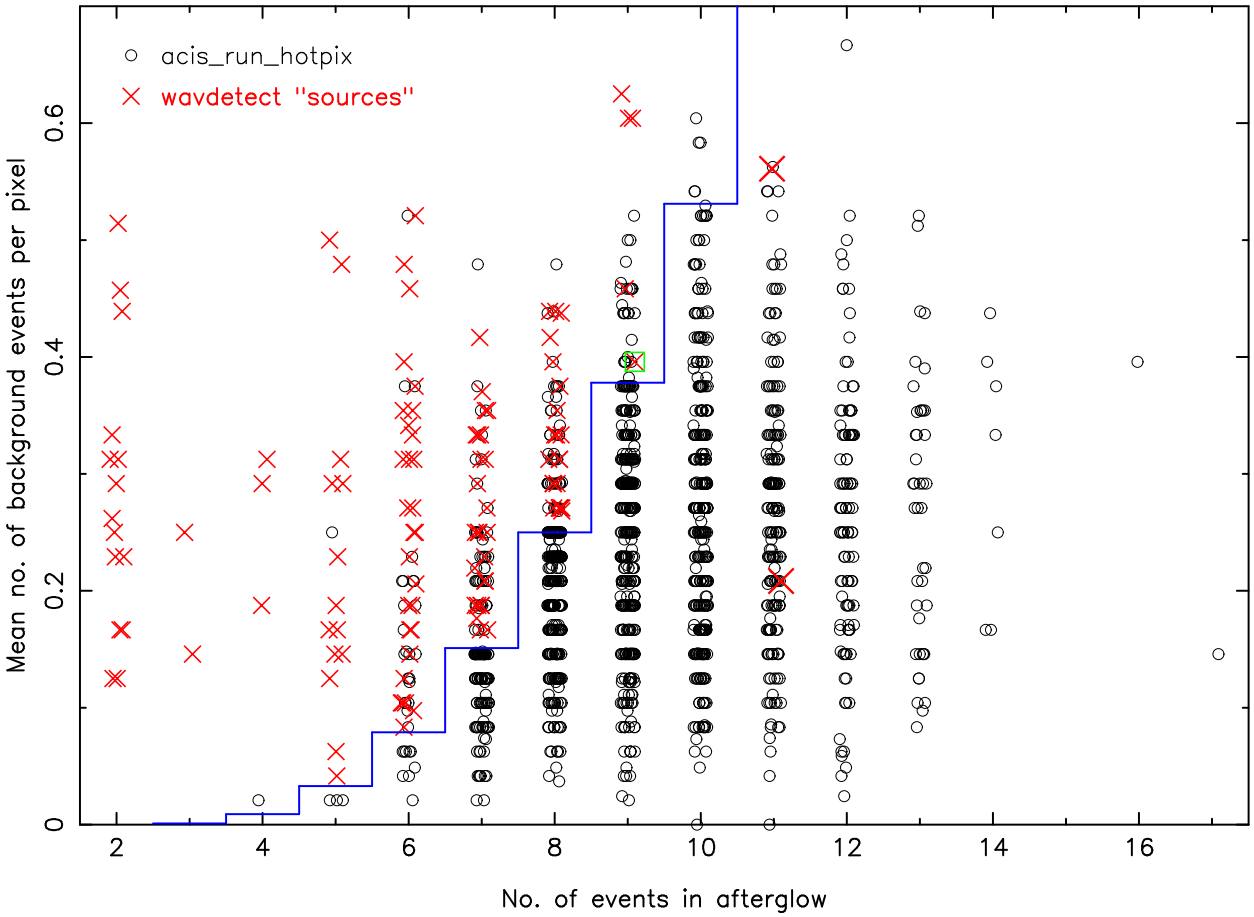

Figure 2: A plot of the number of events in a potential afterglow ($N_{\rm a}$) v. an estimate of the mean number of background events per pixel ($N_{\rm b}$). A small random number has been added to the number of events in the afterglow for display purposes only. The blue line is the boundary between the lower, right-hand region where a potential afterglow is statistically significant (after the trials penalty) and the upper, left-hand region where a potential afterglow is not statistically significant. The exact location of this boundary depends on the number of trials (i.e. the number of pixels searched) and the search algorithm used. The boundary shown here is appropriate for the current two-dimensional afterglow search algorithm in acis_run_hotpix assuming that 6,230,112 pixels are searched for afterglows. Notice that two of the afterglows identified by wavdetect lie below the blue line and a small fraction of the afterglows identified by acis_run_hotpix lie above the blue line. These afterglows are discussed in the text.

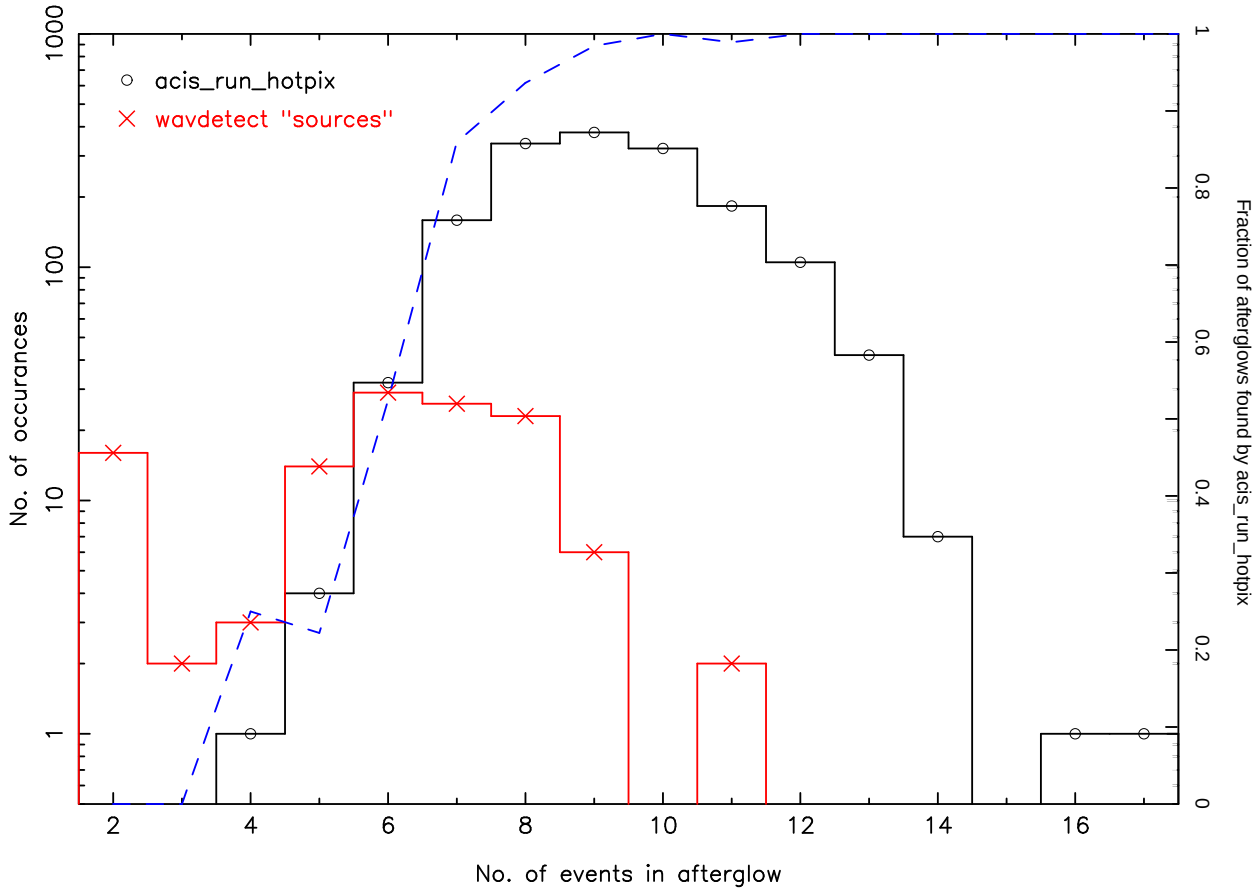

Figure 3: The black (○) and red (×) histograms depict the distributions of the number of potential afterglows identified by the tools acis_run_hotpix and wavdetect, respectively, v. the number of events in the afterglow. These histograms use the vertical scale on the left-hand side of the plot. For the ten data sets examined, no afterglow has a duration longer than 17 frames (53 s). The dashed blue curve is an estimate of the fraction of the potential afterglows that is identified by acis_run_hotpix (i.e. is the black histogram divided by the sum of the two histograms). This curve uses the vertical scale on the right-hand side of the plot. As depicted, acis_run_hotpix identifies more than 90% of the afterglows that contain at least eight events.

6 Recommendations

The tool acis_run_hotpix fails to identify some afterglows with a moderate number of events that wavdetect identifies as potential sources. This problem can be minimized by

- increasing the default value (10^{-3}) of the acis_run_hotpix parameter probthresh,
- reducing the default value (10^{-6}) of the parameter sigthresh, or
- enhancing the algorithm used to search for potential afterglows.

Of these three possibilities, I recommend the later. The algorithm in acis_run_hotpix computes the significance of an afterglow using the total number of events on the background pixels integrated over the entire observation. Since the durations of afterglows are relatively short, the significance of the afterglow could be computed using the number of background events integrated over a short time interval. In this fashion, it should be possible to identify most of the afterglows that are now missed by acis_run_hotpix and found by wavdetect.