Effectiveness of the Gregory-Loredo Algorithm for Detecting Temporal Variability in Chandra Data

Arnold Rots, CXC/SAO

2005-09-29

Abstract

We describe application of the Gregory-Loredo algorithm for detecting temporal variability in Chandra data. We have performed a test on 118 sources spanning the intensity range of 5 to 24000 photons over 102000 s.

We conclude that the G-L algorithm, when combined with a secondary criterion, is extremely robust, yielding a reliable variability indicator as well as a light curve with optimal resolution.

Introduction

This note describes using the Gregory-Loredo algorithm (1992, ApJ **398**, 146) to detect temporal variability in sources identified in the L3 pipeline (intra-ObI only), based on the event files.

Briefly, N events are binned in histograms of m bins, where m runs from 2 to m_{max}. The algorithm is based on the likelihood of the observed distribution $n_1, n_2, \ldots, n_m$ occurring. Out of a total number of m^N possible distributions the multiplicity of this particular one is $N!/(n_1! \,.\, n_2! \ldots n_m!)$. The ratio of the latter to the former provides the probability that this distribution came about by chance. Hence the inverse is a measure of the significance of the distribution. In this way we calculate an odds ratio for m bins versus a flat light curve. The odds are summed over all values of m to determine the odds that the source is time-variable. For more details, see the paper.

The method works very well on event data and is capable to deal with data gaps. We have added the capability to take into account temporal variations in effective area. As a byproduct, it delivers a light curve with optimal resolution.

Although the algorithm was developed for detecting periodic signals, it is a perfectly suitable method for detecting plain variability by forcing the period to the length of the observation.

Implementation

We have implemented the G-L algorithm as a standard C program, operating on simple ASCII files for ease of experimentation.

Input data consist of a list of event times and, optionally, good time intervals with, optionally, normalized effective area (i.e., 1.0 for full exposure).

Two output files are created: odds ratios as a function of m and a light curve file which includes ±3σ curves.

Usage:

[-i]	input file with event times (infile) [stdin]
[-om]	results file: probabilities as a function of m (outfile) [stdout]
[-ntrng]	maximum number of (good) time intervals allowed (integer, >= 0) [200]

[-n]	maximum number of events to be accepted (integer, >= 0) [70000]
[-tb]	Start of time range (-1 if not used) (double) [-1]
[-te]	End of time range (-1 if not used) (double) [-1]
[-rfrac]	fraction of events to be included in subsample (double, 0.000 to 1.000) [1]
[-rseed]	seed for random subsample selection (integer) [1]
[-olc]	resulting output file with light curve (outfile) [stdout]
[-log]	If yes, prints running log on standard out (boolean) [false]
[-mmin]	the minimum number of model bins to use (integer, 2 to 3000) [2]
[-mmax]	the maximum number of model bins (integer, 2 to 3000) [see below]
[-U]	display this message

If m_{max} is not explicitly specified, the algorithm is run twice. The first time all values of m are used, up to the minimum of 3000 and $(t_e - t_b) / 50$; i.e., variability is considered for all time scales down to 50 s which is about 15 times the most common ACIS frame time. The sum of odds $S(m) = \text{sum}\ (O(i),\ i = m_{min} .. m) / (m - m_{min} + 1)$ is calculated as a function of m and its maximum is determined. Then the algorithm is run again with m_{max} set to the highest value of m for which $S(m) > \max(S) / \sqrt{e}$.

The light curve that is generated by the program essentially consists of the binnings weighed by their odds ratios and represents the most optimal binning for the curve. The standard deviation σ is provided for each point of the light curve.

The program provides information on the total odds ratio O (or, rather, its $^{10}\log$), the corresponding probability of a variable signal, the m value with the maximum odds ratio and the odds-weighted first moment of m, as well as the characteristic time scales represented by these two values.

There is an ambiguous range of probabilities: $0.5 < P < 0.9$, and in particular the range between ½ and ⅔ (above 0.9 all is variable, below 0.5 all is non-variable). For this range we have developed a secondary criterion, based on the light curve, its average σ, and the average count rate. We calculate the fractions f_3 and f_5 of the light curve that are within 3σ and 5σ, respectively, of the average count rate. If $f_3 > 0.997$ AND $f_5 = 1.0$ for cases in the ambiguous range, the source is deemed to be non-variable.

Finally, the program assigns a *variability index*:

Variability Index	**Condition**	**Comment**
0	$P <= ½$	Definitely not variable
1	$½ < P < ⅔$ AND $f_3 > 0.997$ AND $f_5 = 1.0$	Not considered variable
2	$⅔ <= P < 0.9$ AND $f_3 > 0.997$ AND $f_5 = 1.0$	Probably not variable
3	$0.5 <= P < 0.6$	May be variable
4	$0.6 <= P < ⅔$	Likely to be variable
5	$⅔ <= P < 0.9$	Considered variable
6	$0.9 <= P$ AND $O < 2.0$	Definitely variable
7	$2.0 <= O < 4.0$	Definitely variable
8	$4.0 <= O < 10.0$	Definitely variable
9	$10.0 <= O < 30.0$	Definitely variable
10	$30.0 <= O$	Definitely variable

The code is structured such that all I/O is executed though three function calls. It is a simple matter to replace these three functions, e.g., to change to output in FITS format.

Test Results

The program was run on all 118 sources found by *wavdetect* in ObsId 635. The total time span of the observation was 102 ks and the sources varied between 5 and 24000 counts. The average time to run the program was 1.5 s per source. 71 sources were found to be variable with an odds ratio > 1.0 (probability > 0.5). Visual inspection of the light curves of all 118 sources found 54 that are variable, though there are a few borderline cases on either side of the divide.

Examples of output files and a summary of the test results are presented in the Appendix.

Analysis

The following table summarizes the number of variable sources detected, the false, and the missed detections, as a function of odds ratio and probability range.

Odds ratio range	**Probability range**	**Good detections**	**False detections**	**Missed with secondary criterion**	**False with secondary criterion**
1.0 – 2.0	½ – ⅔	2	7	1	0
2.0 – 9.0	⅔ – 0.9	5	10	2	0
> 9.0	> 0.9	47	0	0	0

Using G-L just by itself is problematic in the probability range 0.5 – 0.9, considering the required trade-off between missed detections and false detections. We solved this problem by designing the secondary criterion that is based on the fraction of the light curve within 3σ and 5σ. The final result is that three borderline variable sources are missed (with emphasis on *borderline*), but there are no spurious detections. However, any user who is concerned about missing potential candidates should be encouraged to inspect all sources with a variability index greater than 0.

The G-L algorithm, as expected, is pleasantly insensitive to the shape of the light curve, something that is a known problem with the current implementation of the K-S test. It is also not over-interpreting the data in low count rate sources, requiring a statistically significant deviation from a flat distribution before yielding an odds ratio greater than one.

The light curves (and the ± 3σ curves), in providing precisely the desired resolution, are of the kind that we would want to include in the L3 product package.

Attached are ten figures, highlighting the different types of cases. Dashed lines represent the ± 3σ curves. Figs. 1 – 7 provide a typical cross-section of the variable sources.

Fig. 8 shows one of the cases that cannot very well be handled by simple statistics: it fails all criteria, but there appears to be a definite trend; obviously, the judgment as to what is "definite" is subjective and we do not consider the source variable. To some extent, the example in Fig. 5 is in that category, too.

Finally, the test has brought to light the issue of time variable exposure (or effective area). It appears in a number of sources with characteristic times that are harmonics of one of the two dither periods (707 and 1000 s). Figs. 9 and 10 show that this can be properly taken care of by providing the program with normalized effective area as a function of time.

Conclusion

We conclude that G-L provides a robust algorithm for detecting temporal variability that is insensitive to the type and shape of variability and that takes properly into account the uncertainties in the count rate, requiring a statistically significant departure from a flat count rate for it to declare variability. The light curves provided by the program appear to be near-optimal for what we intend to present to users.

The addition of the secondary criterion results in a reliable test, though careful users may want to inspect the light curves of all sources with a non-zero variability index.

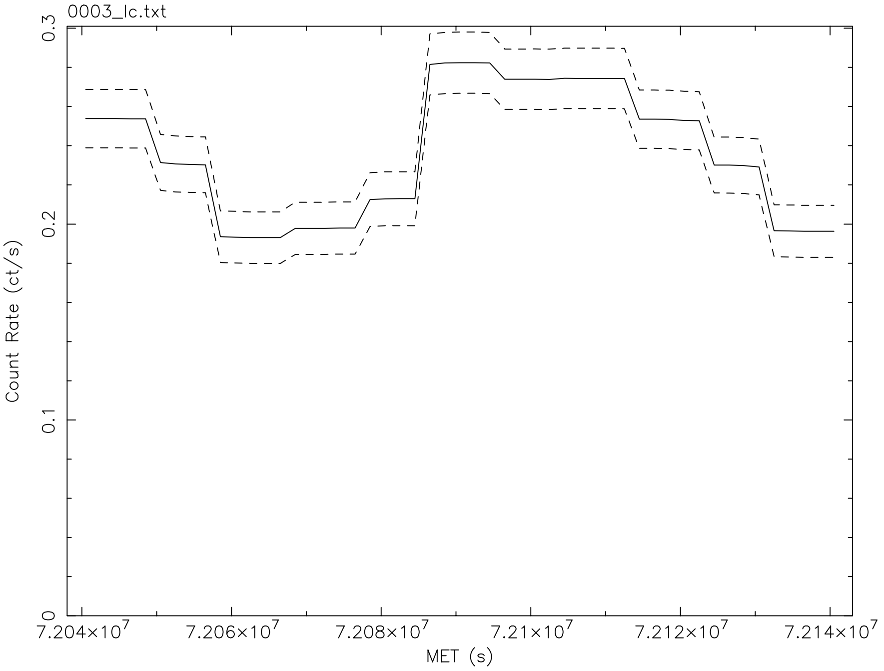

Fig. 1. Source 3: 24093 counts. Even though the SNR is high, the resolution of the light curve is fairly low since higher resolution is not warranted by its shape.

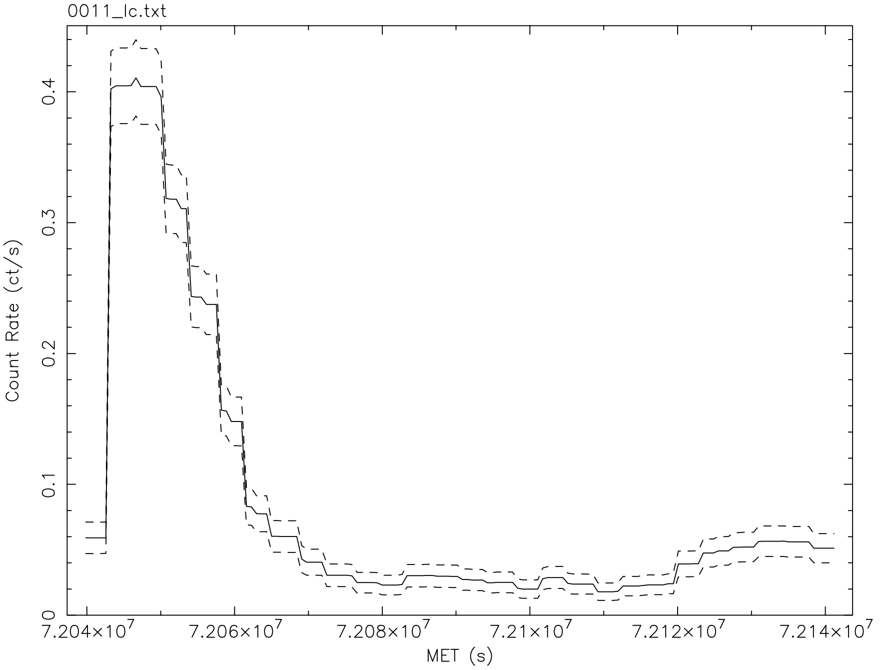

Fig. 2. Source 11: 8697 counts. The timescale of the changes in this source are very much shorter than in source 3; hence the resolution is higher.

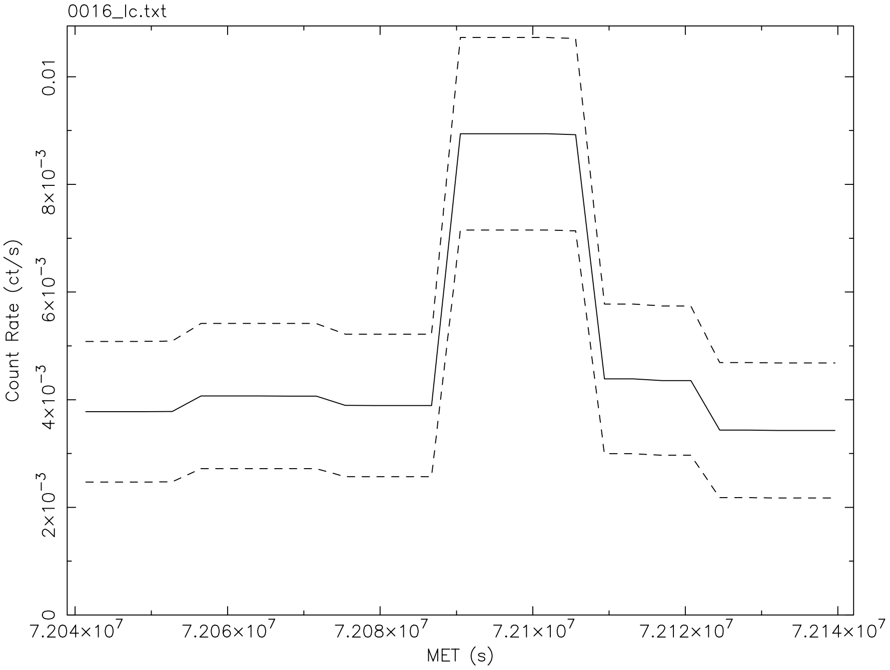

Fig. 3. Source 16: 484 counts. The odds ratio was high (^{10}log(odds)=8.6).

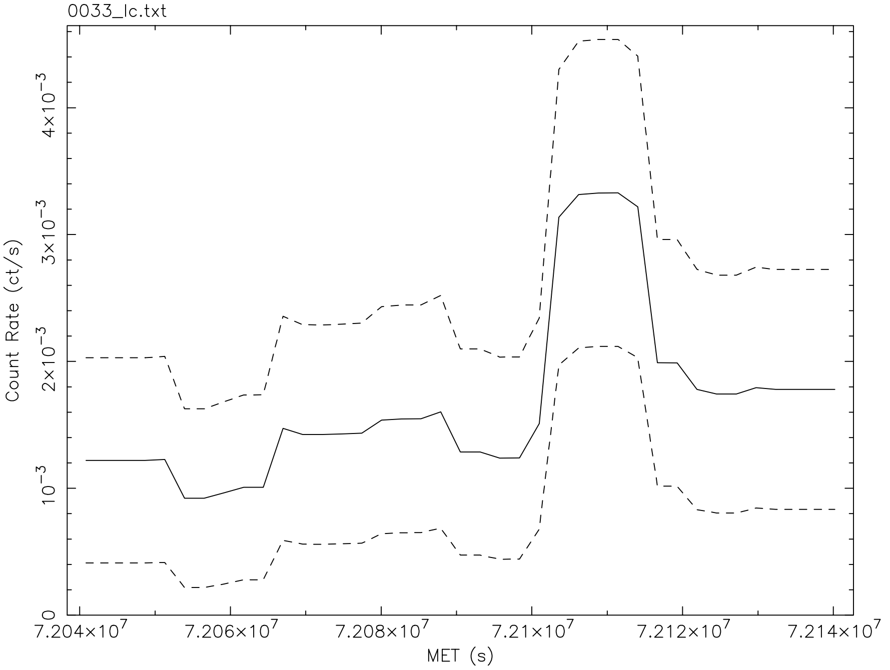

Fig. 4. Source 33: 171 counts.

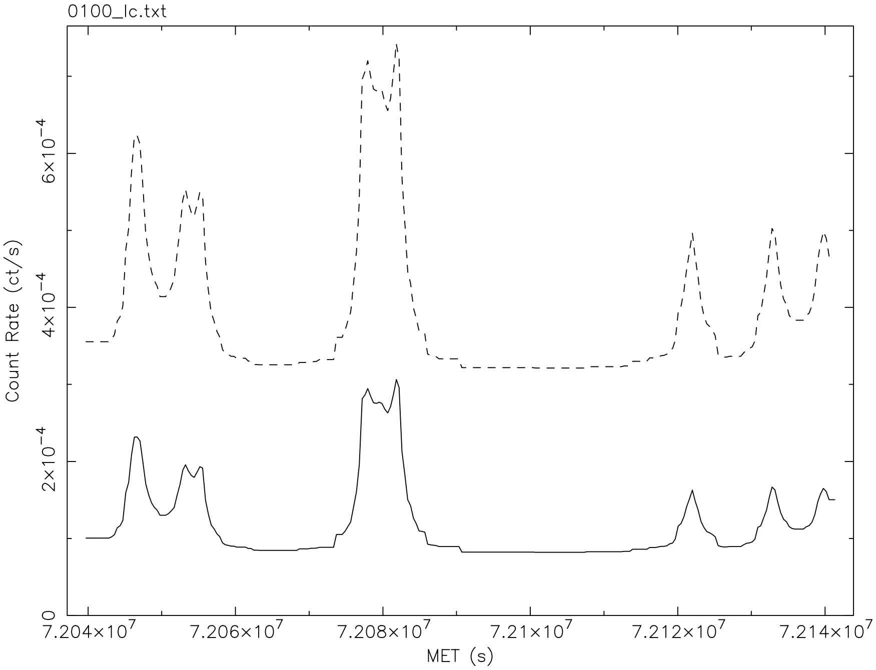

Fig. 5. Source 100: 12 counts. The odds ratio is only 2.2 and on the basis of the 3σ fraction it should be rejected, but it certainly looks variable, although it is a borderline case.

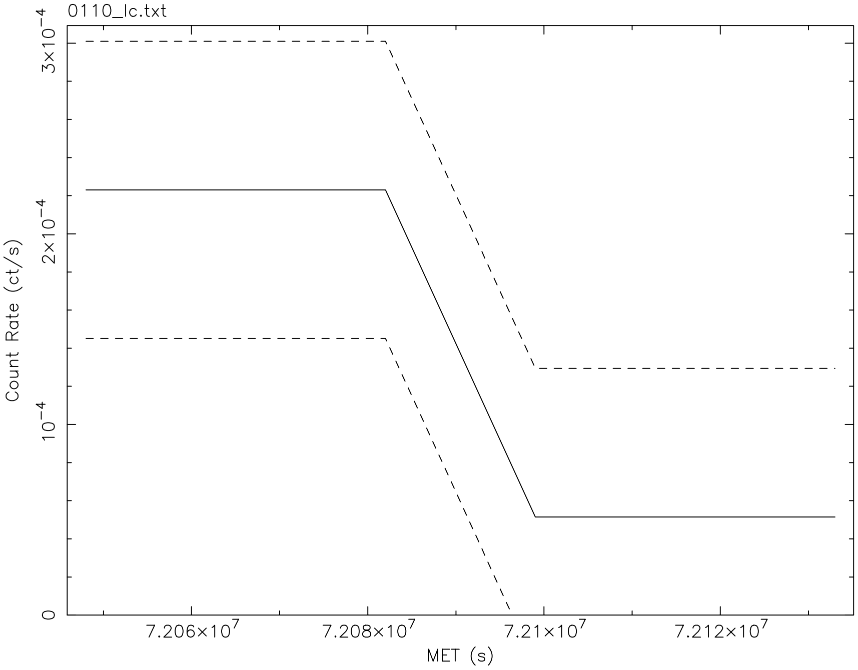

Fig. 6. Source 110: 14 counts. A low count rate, but variable, nevertheless.

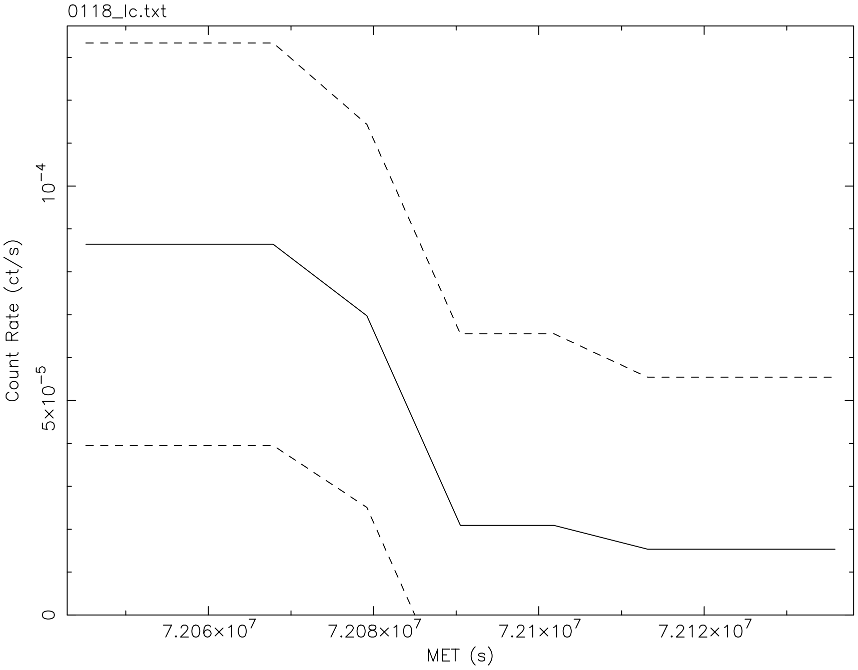

Fig. 7. Source 118: 5 counts. With only 5 counts I would not stake my reputation on this one. f_3 backs that up: probably not variable.

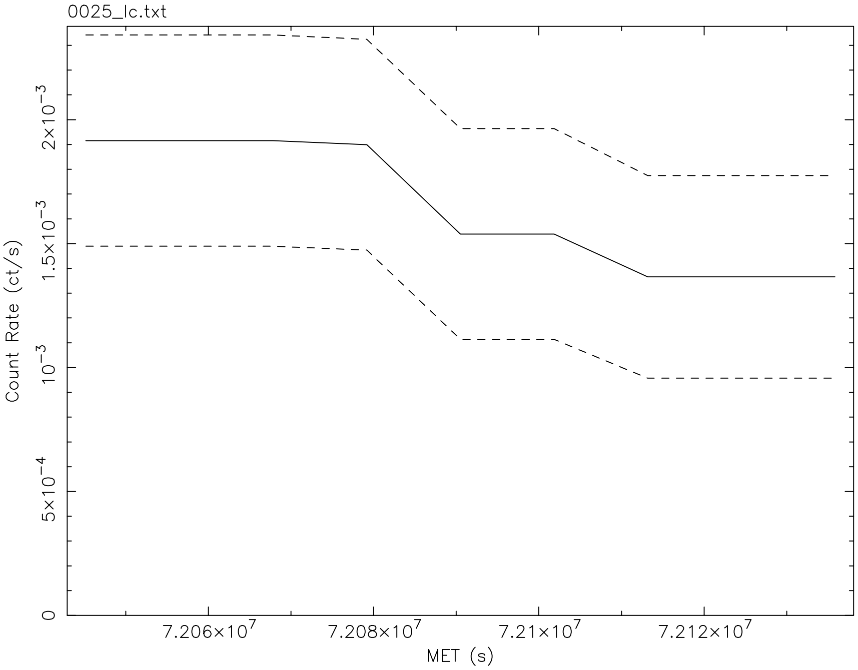

Fig. 8. Source 25: 170 counts. This is an example of a borderline case where there is no statistically significant change while there is yet an unmistakable trend. I still would not consider this source variable.

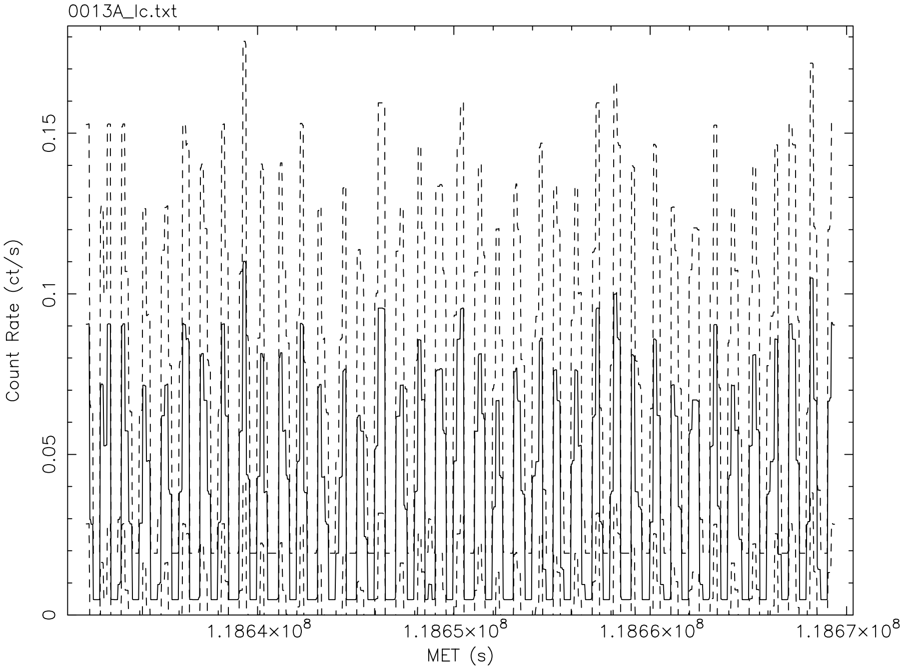

Fig. 9. Source 13 of ObsId 1575 (1376 counts), showing clearly the 707 s dither period.

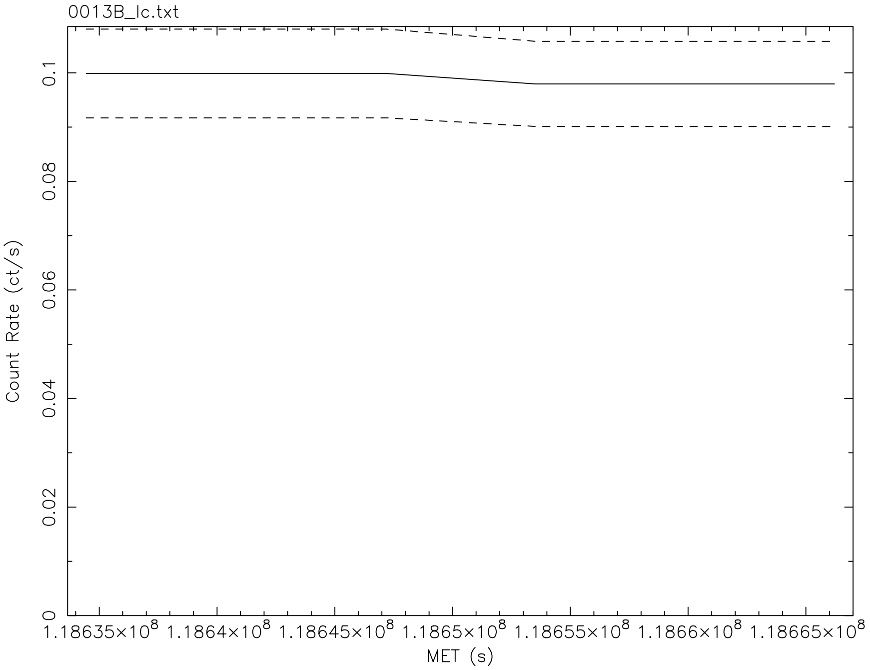

Fig. 10. Source 13 of ObsId 1575 taking into account the normalized effective area as a function of time.

Appendix

Example of output file with odds ratios as a function of *m*.

```
#    >>>  G L V A R Y  <<<
#
# Data file:  0001_
# Time range: 72039530.664659 to 72141499.614917
# Time zero point: 0.000000
#
# Run with 8674 SINGLE events, total integration time 101968.950258 sec
# Total time span covered: 101968.950258 sec
# 10Log ( Odds for variable signal ):           39.753
# Probability of a variable signal  :            1.000
# mmin = 2, mmax = 14
# Fraction of light curve within 3 sigma of average rate: 0.333333, 5 sigma:
0.880952
# Variability index: 10
#

# First moment over m, characteristic time:       9.00      11331.7
# m with maximum odds, characteristic time:       9         11329.9
#
# m           Probability          Sum(Odds)
#
  2           0.000000000000       0.000000
  3           0.000000000358       0.000000
  4           0.000000000000       0.000000
  5           0.000000000000       0.000000
  6           0.001739450095       0.004523
  7           0.000000000000       0.003769
  8           0.000000000000       0.003230
  9           0.997008295938       1.622965
 10           0.000000000000       1.442636
 11           0.000000000000       1.298372
 12           0.001252253609       1.181818
 13           0.000000000000       1.083333
 14           0.000000000000       1.000000
```

Example of light curve output file.

```
#    >>>  G L V A R Y  <<<
#
# Data file:  0001_
# Time range: 72039530.664659 to 72141499.614917
# Time zero point: 0.000000
#
# Run with 8674 SINGLE events, total integration time 101968.950258 sec
# Total time span covered: 101968.950258 sec
# 10Log ( Odds for variable signal ):           39.753
# Probability of a variable signal  :            1.000
# mmin = 2, mmax = 14
# Fraction of light curve within 3 sigma of average rate: 0.333333, 5 sigma:
0.880952
# Variability index: 10
#

#    Time            <F>            sigma        <F>-3*sigma     <F>+3*sigma
#
 72040744.58     0.074508293     0.002434692     0.067204217     0.081812369
 72043172.42     0.074508293     0.002434692     0.067204217     0.081812369
 72045600.25     0.074508293     0.002434692     0.067204217     0.081812369
 72048028.09     0.074502112     0.002434583     0.067198362     0.081805863
 72050455.92     0.074502112     0.002434583     0.067198362     0.081805863
 72052883.75     0.073710951     0.002423006     0.066441932     0.080979970
 72055311.59     0.073710951     0.002423006     0.066441932     0.080979970
 72057739.42     0.073710337     0.002423000     0.066441338     0.080979336
 72060167.26     0.073710337     0.002423000     0.066441338     0.080979336
 72062595.09     0.075820101     0.002453685     0.068459046     0.083181156
 72065022.93     0.075825399     0.002453778     0.068464064     0.083186734
 72067450.76     0.075825399     0.002453778     0.068464064     0.083186734
 72069878.60     0.075825399     0.002453778     0.068464064     0.083186734
 72072306.43     0.075825399     0.002453778     0.068464064     0.083186734
 72074734.27     0.082169759     0.002542570     0.074542048     0.089797470
 72077162.10     0.082169759     0.002542570     0.074542048     0.089797470
 72079589.93     0.082169759     0.002542570     0.074542048     0.089797470
 72082017.77     0.082176529     0.002542683     0.074548480     0.089804577
 72084445.60     0.082176529     0.002542683     0.074548480     0.089804577
 72086873.44     0.081033740     0.002527049     0.073452594     0.088614886
 72089301.27     0.081033740     0.002527049     0.073452594     0.088614886
 72091729.11     0.081020337     0.002526882     0.073439691     0.088600982
 72094156.94     0.081020337     0.002526882     0.073439691     0.088600982
 72096584.78     0.073987792     0.002427082     0.066706545     0.081269039
 72099012.61     0.073977932     0.002426912     0.066697197     0.081258667
 72101440.45     0.073977932     0.002426912     0.066697197     0.081258667
 72103868.28     0.073977932     0.002426912     0.066697197     0.081258667
 72106296.11     0.073977932     0.002426912     0.066697197     0.081258667
 72108723.95     0.123074792     0.003017199     0.114023193     0.132126390
 72111151.78     0.123074792     0.003017199     0.114023193     0.132126390
 72113579.62     0.123074792     0.003017199     0.114023193     0.132126390
 72116007.45     0.123040355     0.003016742     0.113990131     0.132090580
 72118435.29     0.123040355     0.003016742     0.113990131     0.132090580
 72120863.12     0.093064133     0.002684086     0.085011876     0.101116390
 72123290.96     0.093064133     0.002684086     0.085011876     0.101116390
```

```
72125718.79     0.093011353     0.002683583     0.084960604     0.101062103
72128146.63     0.093011353     0.002683583     0.084960604     0.101062103
72130574.46     0.088265162     0.002623430     0.080394871     0.096135453
72133002.29     0.088260601     0.002623359     0.080390524     0.096130679
72135430.13     0.088260601     0.002623359     0.080390524     0.096130679
72137857.96     0.088260601     0.002623359     0.080390524     0.096130679
72140285.80     0.088260601     0.002623359     0.080390524     0.096130679
```

Summary of Test Results

```
Source  Counts   Var  log(Odds)  Probab-     f3        f5
                 Inx              ility
===========================================================

 0029     257     0     -1.318    0.046    1.0000    1.0000
 0030     700     0     -1.314    0.046    1.0000    1.0000
 0088     251     0     -1.165    0.064    1.0000    1.0000
 0086     250     0     -1.071    0.078    1.0000    1.0000
 0012    1156     0     -1.042    0.083    1.0000    1.0000
 0064     374     0     -0.958    0.099    1.0000    1.0000
 0040     116     0     -0.937    0.104    1.0000    1.0000
 0027    2710     0     -0.769    0.146    0.8000    1.0000
 0077      89     0     -0.689    0.170    1.0000    1.0000
 0045      72     0     -0.669    0.176    1.0000    1.0000
 0068     157     0     -0.653    0.182    1.0000    1.0000
 0076      69     0     -0.401    0.284    1.0000    1.0000
 0063      55     0     -0.314    0.327    1.0000    1.0000
 0025     170     0     -0.313    0.327    1.0000    1.0000
 0079      56     0     -0.303    0.333    1.0000    1.0000
 0073      33     0     -0.274    0.347    1.0000    1.0000
 0039      76     0     -0.265    0.352    1.0000    1.0000
 0089      35     0     -0.259    0.355    1.0000    1.0000
 0083     106     0     -0.245    0.363    1.0000    1.0000
 0099      37     0     -0.233    0.369    1.0000    1.0000
 0057      34     0     -0.221    0.376    1.0000    1.0000
 0046      94     0     -0.177    0.399    1.0000    1.0000
 0069      30     0     -0.175    0.400    1.0000    1.0000
 0075     100     0     -0.155    0.412    1.0000    1.0000
 0106      58     0     -0.155    0.412    1.0000    1.0000
 0061      31     0     -0.150    0.415    1.0000    1.0000
 0091      48     0     -0.147    0.416    1.0000    1.0000
 0047     160     0     -0.137    0.422    1.0000    1.0000
 0087      36     0     -0.136    0.423    1.0000    1.0000
 0090      20     0     -0.123    0.430    1.0000    1.0000
 0071      28     0     -0.104    0.440    1.0000    1.0000
 0067      18     0     -0.102    0.442    1.0000    1.0000
 0092      18     0     -0.099    0.443    1.0000    1.0000
 0094      38     0     -0.098    0.444    1.0000    1.0000
 0111      17     0     -0.091    0.448    1.0000    1.0000
 0085      20     0     -0.079    0.455    1.0000    1.0000
 0097      14     0     -0.068    0.461    1.0000    1.0000
 0113      14     0     -0.061    0.465    1.0000    1.0000
 0093      13     0     -0.060    0.466    1.0000    1.0000
 0105      12     0     -0.053    0.470    1.0000    1.0000
 0070      49     0     -0.037    0.478    1.0000    1.0000
 0098      24     0     -0.031    0.482    1.0000    1.0000
 0050      55     0     -0.018    0.490    1.0000    1.0000
 0065     184     0     -0.018    0.490    0.9997    1.0000
 0082      34     0     -0.018    0.489    1.0000    1.0000
 0038     124     0     -0.013    0.492    1.0000    1.0000
 0114       7     0     -0.009    0.495    1.0000    1.0000
 0108       9     1      0.012    0.507    1.0000    1.0000
 0112       9     1      0.030    0.518    1.0000    1.0000
 0109      12     1      0.033    0.519    1.0000    1.0000
 0101      46     1      0.036    0.521    1.0000    1.0000
 0117      17     1      0.145    0.583    1.0000    1.0000
 0095      25     1      0.153    0.587    1.0000    1.0000
 0107      39     1      0.161    0.592    1.0000    1.0000
 0034     104     1      0.295    0.664    1.0000    1.0000
 0103      28     2      0.334    0.683    1.0000    1.0000
 0100      12     2      0.349    0.691    1.0000    1.0000
```

0115	21	2	0.371	0.701	1.0000	1.0000
0044	162	2	0.385	0.708	1.0000	1.0000
0102	33	2	0.475	0.749	1.0000	1.0000
0036	314	2	0.486	0.754	0.9998	1.0000
0081	222	2	0.504	0.762	0.9995	1.0000
0104	82	2	0.509	0.763	1.0000	1.0000
0118	5	2	0.582	0.793	1.0000	1.0000
0072	131	2	0.624	0.808	0.9993	1.0000
0066	63	2	0.646	0.816	1.0000	1.0000
0048	136	2	0.739	0.846	1.0000	1.0000
0051	167	3	0.071	0.541	0.8095	1.0000
0018	377	5	0.651	0.817	0.6667	1.0000
0084	16	5	0.849	0.876	0.9838	1.0000
0024	542	5	0.855	0.877	0.8000	0.9111
0031	287	6	1.006	0.910	0.3333	1.0000
0110	14	6	1.079	0.923	0.0000	1.0000
0059	64	6	1.082	0.924	0.7778	1.0000
0096	93	6	1.220	0.943	0.9995	1.0000
0033	171	6	1.243	0.946	0.8718	0.8974
0042	222	6	1.265	0.948	0.9997	1.0000
0058	530	6	1.308	0.953	0.7917	1.0000
0060	58	6	1.468	0.967	0.9136	0.9753
0074	47	6	1.787	0.984	0.9663	1.0000
0049	67	7	2.250	0.994	1.0000	1.0000
0037	100	7	2.269	0.995	0.9103	0.9359
0080	135	7	2.919	0.999	0.8333	0.8611
0028	336	7	2.972	0.999	0.8333	0.8333
0043	149	7	3.081	0.999	0.9352	0.9769
0035	129	7	3.111	0.999	0.8730	0.9524
0041	2040	7	3.154	0.999	0.7222	0.7222
0062	38	7	3.164	0.999	0.8718	0.8718
0116	14	7	3.434	1.000	0.9989	0.9993
0052	98	7	3.498	1.000	0.0000	1.0000
0032	291	8	4.030	1.000	0.4815	1.0000
0023	220	8	4.096	1.000	0.9682	1.0000
0056	68	8	4.968	1.000	0.9444	0.9921
0055	69	8	5.490	1.000	0.6190	0.8095
0021	659	8	6.178	1.000	0.3333	0.3333
0053	57	8	6.292	1.000	0.9815	0.9815
0015	4847	8	8.525	1.000	0.0000	0.6667
0016	484	8	8.626	1.000	0.8148	0.8148
0054	138	9	11.108	1.000	0.9799	0.9856
0019	3247	9	21.207	1.000	0.5000	0.6667
0014	1323	9	29.815	1.000	0.7424	0.9091
0026	199	10	30.290	1.000	0.9603	0.9683
0001	8674	10	39.896	1.000	0.3333	0.8810
0078	328	10	50.032	1.000	0.9000	0.9333
0007	1115	10	58.955	1.000	0.5333	0.8000
0020	1025	10	72.159	1.000	0.7667	0.9167
0003	24093	10	77.424	1.000	0.1569	0.4314
0006	2831	10	79.677	1.000	0.3333	0.6889
0004	6775	10	93.660	1.000	0.1282	0.4872
0022	651	10	117.251	1.000	0.1961	0.2941
0008	4445	10	142.491	1.000	0.8756	0.9813
0009	1106	10	181.623	1.000	0.2444	0.4444
0010	1102	10	231.745	1.000	0.8989	0.9579
0017	785	10	367.387	1.000	0.1149	0.1724
0013	1401	10	523.414	1.000	0.1560	0.2766
0005	2992	10	627.994	1.000	0.0575	0.0575
0011	8697	10	2158.204	1.000	0.0333	0.0333
0002	17450	10	5767.780	1.000	0.0000	0.0172