

# Using the Gregory-Loredo Algorithm for the Detection of Variability in the Chandra Source Catalog

Arnold H. Rots

Smithsonian Astrophysical Observatory, Cambridge, MA 02138, USA.  
email: arots@head.cfa.harvard.edu

**Abstract.** In searching for a reliable variability indicator for Chandra X-ray event data, we implemented the Gregory-Loredo (G–L) algorithm to assess variability and to generate light curves for the Chandra Source Catalog. A test was performed on 118 sources detected in a single observation, spanning the intensity range 5–24,000 photons over a total observing time of 102,000 sec. We conclude that the G–L algorithm is extremely robust, yielding both a reliable variability indicator and light curves with optimal resolution, while requiring a very modest amount of CPU time. The algorithm can also work on binned data, and is capable of handling data gaps as well as variations in effective area.

**Keywords.** Methods: statistical, catalogs, X-rays: general

---

## 1. Introduction

Determining whether a source of discrete (X-ray) photons exhibits variability is a non-trivial problem. We have tested the algorithm published by Gregory & Loredo (1992), modified for aperiodic variability, for use in the production of the Chandra Source Catalog (CSC; Evans *et al.* 2010). In brief,  $N$  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  $n_1, n_2, \dots, n_m$  occurring. Out of a total number of  $mN$  possible distributions, the multiplicity of this particular one is  $N!/(n_1!n_2!\dots n_m!)$ . The ratio of the latter to the former provides the probability that this distribution came about by chance. Hence, its inverse is a measure of the significance of the distribution. In that 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 Gregory & Loredo (1992) and Loredo (p. 87).

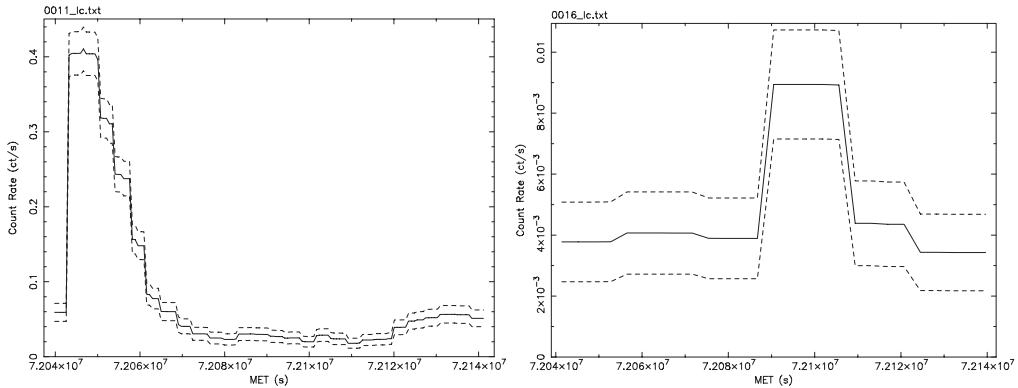
## 2. Evaluation and Usage

We tested the algorithm on all 118 sources found by the programme WAVDETECT in Chandra 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 programme was  $1.5\sigma$  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 were a few borderline cases on either side of the divide.

Our tests proved that the method works very well on event data, and is capable of dealing with data gaps. We added a capability of taking into account temporal variations in effective area. As a byproduct, a light curve with optimal resolution is delivered. That light curve is effectively the sum of the binnings weighed by their odds ratios, and as

such it represents the most optimal binning for the curve. The standard deviation  $\sigma$  is provided for each point of the light curve.

The odds ratio yields a probability for variability. The tests showed that there is an ambiguous range of probabilities:  $0.5 < P < 0.9$ , and in particular the range between 0.5 and 0.67 (above 0.9 all is variable, below 0.5 all is non-variable). For that range we have developed a secondary criterion based on the light curve, its average  $\sigma$  and the average count rate. We calculated the fractions  $f_3$  and  $f_5$  of the light curve that are within  $3\sigma$  and  $5\sigma$ , 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.



**Figure 1.** Two examples of light curves produced by our implementation, derived from sources with 8697 counts (left) and 14 counts (right). The dashed lines indicate the  $3\sigma$  levels.

### 3. Conclusion

G–L provides a robust algorithm for detecting temporal variability which is insensitive to the type and shape of the variability and which 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 programme appear to be near optimal for what we intended 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.

The programme is integrated into the production pipeline of the CSC. Fig. 1 provides two examples of light curves that are produced. Further information may be found at

<http://cxc.cfa.harvard.edu/csc/>

[http://cxc.cfa.harvard.edu/csc/why/gregory\\_loredo.html](http://cxc.cfa.harvard.edu/csc/why/gregory_loredo.html)

<http://hea-www.harvard.edu/~arots/GL/>

### Acknowledgement

This work has been supported by NASA under contract NAS 8-03060 to the Smithsonian Astrophysical Observatory for operation of the Chandra X-ray Center.

### References

- Rots, A. H., *et al.* 2010, *ApJS*, 189, 37  
 Gregory, P. C. & Loredo, T. J. 1992, *ApJ*, 398, 146