Astronomical Data Analysis Software and Systems IV
ASP Conference Series, Vol. 77, 1995
R. A. Shaw, H. E. Payne, and J. J. E. Hayes, eds.

Source-searching in Photon-event Lists without Imaging

C. G. Page

X-ray Astronomy Group, University of Leicester, UK

Abstract. Many methods have been proposed for the detection of point sources in images. In re-processing data from the *ROSAT* extreme ultra-violet sky-survey we have made use of two different methods. One of these works directly on an event list, avoiding the losses which occur when events are binned into images.

1. Introduction

The first complete survey of the sky in the extreme ultra-violet (EUV) band was carried out by the Wide Field Camera (WFC) on *ROSAT*. The data, collected from 1990 August through 1991 January, were analyzed jointly by the five institutes in the UK *ROSAT* Consortium. The aim of our initial reduction (Page & Denby, 1992) was to obtain results without delay, so we did not expect to achieve the ultimate in sensitivity and completeness. The resulting *Bright Source Catalogue* (Pounds et. al. 1993) listed 383 EUV sources, a notable increase on the mere handful known previously.

The entire data-set of about 4×10^8 events has now been re-processed to get the best possible sensitivity, reliability, and completeness. Software improvements have been made in many areas, especially in determining the telescope aspect from the star-tracker data, in filtering out data with high background levels, and in searching for point sources.

2. Source Detection Methods

Textbooks on signal detection theory tell us that the matched filter is the optimum detector of a signal of known shape buried in stationary, Gaussian, additive noise. Even with our count-rates, clearly low enough to be in the Poissonian regime, matched filtering is still an attractive technique. Its simplest implementation involves binning the events into an image, convolving this with a unit replica of the point-spread function (PSF), and searching for peaks in the resulting array. This is relatively easy to compute, but there is one notable drawback: if the raw data have a Poissonian distribution there is in general no analytic function for the distribution of the convolution products. This makes it hard to set confidence limits for detections. A Gaussian approximation (e.g., Marshall 1994) is really only good enough to set a detection threshold for an initial screening pass.

This problem can be overcome by fitting the PSF to each peak in the raw image by least-squares and using a χ^2 test for significance. Unfortunately if, as often happens, some pixels have zero counts, determining the data variance presents a problem. Again there are ways around this (e.g., Kearns, Primini, & Alexander 1995), but the proper solution is to use maximum likelihood.

Maximum likelihood techniques are now widely used in astronomy: Cash (1979) pointed out the valuable properties of the C-statistic, which has a χ^2 distribution under almost all circumstances. This makes it easy to set a detection threshold, and to determine confidence limits in the fitted parameters. For an image where e_{ij} counts were expected and n_{ij} counts observed in pixel (i,j), the C-statistic to be minimized is:

$$C = -2\sum_{ij}(e_{ij} - n_{ij}\ln e_{ij})$$

The expected count in each pixel depends, of course, on the size and position of the source, the shape of the PSF, and on the background level.

Although this method produces good results, it is much slower. In order to process a large data volume the usual practice has been to scan the sky using a matched filter, then refine the search around each apparent peak and evaluate its significance using maximum likelihood. The PSS program, part of the Starlink ASTERIX package, uses just such a hybrid technique. PSS is widely used in the UK to analyze data from both the WFC and the XRT telescopes on *ROSAT*, and the author of the program, David Allan, has supplied a version to CEA Berkeley for handling *EUVE data.*

3. Source Searching without Forming an Image

When a telescope is scanned across the sky, as in a sky-survey, and it has a resolution which varies across the detector surface, some information is inevitably lost in forming an image. In such cases PSS uses an averaged point-spread function: for the WFC the resolution (FWHM) ranges from around 1 arc-minute at the center to 3 arc-minutes at the edge of the field of view, so the loss from using the mean PSF is small but not negligible.

The resolution range of the *ROSAT* XRT is rather larger. The solution proposed by Cruddace et al. (1988) for XRT survey analysis was to consider the detector surface divided into many concentric annuli, and form separate images from each one, before applying the maximum likelihood technique to the ensemble. For the WFC, however, such a division of the detector surface would not work as well because its off-axis PSF is far from circular.

A much better solution would be to avoid forming images at all and apply maximum likelihood to the raw event list. The C-statistic is actually simpler:

$$C = 2\left(E - \sum_{k}\ln e_k\right)$$

where e_k is the normalized probability density where each event was detected, and E is their sum, equal to the observed total count.

The main advantage of this is that each event can be processed using the PSF corresponding to its point of incidence on the detector. This should make better use of the higher spatial resolution at the center of the detector and avoid any loss of information from binning the events into an image.

A disadvantage is the number of computations required, given the need to cover the sky using a one-arcminute grid in two spectral bands. Close attention therefore had to be paid to efficiency, with a prime target being the function minimization, an inherently iterative technique. After examining the many methods listed by Press et al. (1992), it was found the downhill simplex method of Nelder & Mead (1965), said to be reliable but slow, was in fact among the fastest. This speed arises from the fact that it does not require the computation of derivatives of the function being minimized. An even greater performance gain resulted from using a source of strength S above zero but below the expected detection threshold in the first test at each grid point. Since the χ^2 surface of C against S is fairly smooth and roughly parabolic, the direction of the change in S after the first iteration shows whether there is any significant flux or not near that point. At the majority of points in the sky, therefore, only one iteration was required.

Even so it was still necessary to evaluate the kernel, and hence the PSF, over 10^{11} times. To make this efficient, several megabytes of memory were used to hold the PSF as a look-up table, sampled at suitable intervals in all four dimensions, and use quadrilinear interpolation. Another useful step was to evaluate the sum of the logarithms in the kernel as the logarithm of their product, with careful checks to avoid arithmetic overflow or underflow.

Rapid access to the event data was also essential. The raw data were stored on disk in regions about $2^\circ \times 2^\circ$ across with the events in time order. Each region was processed by reading all events into a table, with one column per attribute. When testing each grid point one needs to access all events in its vicinity: in our case those in a circle of radius 8 arc-minutes. The problem could have been solved by sorting the table into spatial order (as in an IRAFQPOE file) but sorting is inherently rather slow. Our solution was to turn the table into a set of linked-lists by adding two more arrays. The first array was 2-dimensional, like an image, but holding for each pixel just a pointer to the first event in the list which had a celestial position within that pixel. An extra column vector was then used to hold, for each event, a pointer to the next event located within the same pixel. The end of each pointer chain was set to zero. Thus all events in a small region could be accessed just by following, for each pixel, a short chain of pointers. Note that the event coordinates were left as floating-point numbers, they were rounded to integers only to determine membership of a particular pointer chain. The benefit of using this structure was that it could be created on a single pass through the table.

As a result of these changes, we were able to achieve a similar execution speed to that of the established PSS program. It is important to note, however, that in our data-set there are only a few events per pixel: your mileage may vary.

Determining whether this method was more or less sensitive that the standard PSS program turned out to be much more difficult. Since the differences were clearly small, the relative performance could be established only by carrying out extensive simulations. Simulating a suitably realistic sky with added sources takes long enough in the image domain, the generation of a sufficiently large set of photon-event lists with similar properties would have taken even longer.

We decided, therefore, to process the entire sky-survey data set using both the PSS program and an event-based maximum likelihood program in parallel, in the hope that the results from one would be clearly better than the other. The results were compared frequently during the processing, and this made it possible to identify and solve many problems which might otherwise have escaped notice. With these problems fixed, each program detected a small number of sources near the detection threshold which were missed by the other, but the results generally showed very good consistency, with very similar sensitivity. The two programs used somewhat different background estimation algorithms, and this may have masked any small difference in their inherent detection sensitivities.

The results, which have just been submitted for publication, are a substantial advance on our earlier work: there are now nearly 500 sources in total, with most of them detected in both filter bands.

4. Conclusions

Our main conclusion is not that one technique was clearly better than the other, but that we gained a lot from using the two of them in parallel. Some other projects, for example HIPPARCOS (Hog et al. 1992), have recognized the value of passing their data through two independent processing paths.

Acknowledgments. I acknowledge the contributions of many members of the UK *ROSAT* Consortium to the analysis of the WFC sky survey. The work was supported by the UK SERC/PPARC.

References

Cash, W., 1979 ApJ, 228, 939

Cruddace, R. G., Hasinger, G. R., & Schmitt, J. H. 1988, Astronomy from Large Databases, ESO Conference and Workshop Proceedings No. 28, p. 177

Hog, E., Kovalevsky, J., & Lindegren, L. 1992, ESA Bulletin, 43H

Kearns, K., Primini, F., & Alexander, D. 1995, this volume, p. **??**

Marshall, H. 1994, in Astronomical Data Analysis Software and Systems III, ASP Conf. Ser., Vol. 61, eds. D. R. Crabtree, R. J. Hanisch, & J. Barnes (San Francisco, ASP), p. 403marshallh

Nelder, J. A., & Mead, R. 1965 Comp. J., 7, 308

Page, C. G., & Denby, M. 1992, in Astronomical Data Analysis Software and Systems I, ASP Conf. Ser., Vol. 25, eds. D. M. Worrall, C. Biemesderfer, & J. Barnes (San Francisco, ASP), p. 351

Pounds, K. A., et al. 1993, MNRAS, 260, 77P

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes 2nd edition (New York, Cambridge University Press)