52-89

147932

p-3

N93-20795

EINSTEIN SLEW SURVEY: DATA ANALYSIS INNOVATIONS

M. ELVIS, D. PLUMMER, J. SCHACHTER AND G.FABBIANO
Center for Astrophysics, 60 Garden St. Cambridge MA 02138 USA

ABSTRACT Several new methods were needed in order to make the *Einstein* Slew X-ray Sky Survey. These may be useful for other projects.

SCALE OF THE SLEW SURVEY PROJECT

The *Einstein* X-ray observatory was not intended to make a survey of the sky. However in moving between targets the instruments were left on, so that exposure was accumulated on most of the sky. Eventually this led to the *Einstein* Slew Survey containing 818 bright X-ray sources, 40% of which were unknown (Including almost doubling the number of known BL Lac objects).

The idea of a sky survey made from *Einstein* data was alluring. However the project faced a number of problems: the software was on obsolete 16-bit machines; the data had never been looked at, was poorly documented, and was dispersed on ~2500 magnetic tapes; the star tracker data was unusable so there was no known aspect system; the gyro aspect data needed extrapolation over ~ 120° slews; the data set was fairly large, ~ 400 Mbytes; we needed an arcminute, all-sky exposure map ; and the 'sliding box' detect has $\sim 4 \times 10^7$ cells over the sky. No manager could commit large resources to such an uncertain a project. Fortunately technology made the Slew Survey feasible. Below we summarize the innovations which enabled the Slew Survey to be done.

EXPERIMENTAL APPROACH TO LARGE PROJECTS

By greatly lowering the investment in time and manpower needed to process a satellite data set modern computer hardware allows an 'experimental' approach to large data analysis projects. Instead of deriving optimum analysis methods from first principles the data can be processed with 'quick and dirty' algorithms, and reprocessed many times until an successful result is obtained. This resembles *rapid prototyping*. A small team (one full time programmer and one part-time scientist) were able to carry out the whole Slew project in less than two years.

The two keys are: (1)**Large on-line storage** (a 1-Gbyte disk in our case; optical WORMs for larger archives) has removed the manual overhead and time involved in loading tapes. To process the original 2000 magnetic tapes of *Einstein* data even once was a major undertaking. Now the data are all on WORMs, on-line; (2)**Large amounts of CPU** available without being noticed by other users of the system mean that a complete reprocessing becomes a small decision. We had ~100 times the CPU power available to the original *Einstein* project.

The processing time was thus reduced from months to just days. The whole Slew Survey was re-processed at least a dozen times to reach an acceptable solution.

This approach is only possible for a data product, such as a survey. Projects with a software deliverable cannot operate this way. Our software is not intended for public distribution, is almost undocumented, and is certainly unsupported. The quality assurance lies in the scientific paper describing the survey results, which includes many checks on quality.

PARALLEL PROCESSING ON A LAN

A local area network of workstations was used in a quite simple and primitive way to run up to 10 parallel streams of processing. A complete re-analysis of the survey could in this way be accomplished in *3 days instead of 3 to 4 weeks*. This, as noted above, changes the whole approach to the project.

The main problem in implementing this parallel system was the control list since occasionally two machines would attempt access simultaneously. Unix does not handle file locks well, so we resorted to 10 separate lists. Several off the shelf packages now exist for SUNs that manage parallel processing more flexibly. These should enable the whole SAO High Energy LAN of ~60 4-Mflop SPARC 1 machines to be combined. Depending on the application such a 'network supercomputer' may reach of order 200Mflop (roughly 2/3 of a Cray Y/MP).

PERCOLATION SOURCE DETECTION

Standard source detection algorithms in X-ray astronomy have been of the sliding-box type. They are thus 'sky-centered' *i.e.* every part of the survey area is examined and there are 150×10^6 box positions on the sky. To keep compute time reasonable we were forced to a different approach. In the Slew Survey the sky is sparsely populated with photons. Since there are 3×10^6 photons in the Slew Survey a photon centered-approach is some 50 times more efficient than a sky-centered approach.

We developed a method to find photons which were too closely associated with their neighbors for mere chance. This approach turned out to be identical to that used to locate groups of galaxies in the CfA redshift survey (Huchra and Geller 1984), and is a simple version of the class of 'percolation algorithms'.

'MINIMUM ACTION' IDENTIFICATIONS

Archive material

Two thirds of the Survey sources have been identified using archival resources (NED, SIMBAD, z-cat). Of the remaining sources 75% have plausible counterparts in one or more of the existing digitally accessible sky surveys: the IRAS Point Source and Faint Source Catalogs; the 5GHz 300ft Green Bank (87GB) radio survey; the HST Guide Star Catalog (GSC); the ROE/NRL and Minnesota catalogs of the optical sky surveys.

These resources now make it possible to produce spectral energy distributions for many X-ray sources in the Slew Survey without any new observing.

We can isolate the most likely counterparts. Source classes can be assigned with reasonable probability based on these distributions. In this way ~96% of the sources have some likely counterpart. The follow-up observing time needed on optical telescopes is thus minimized. Only a couple of dozen high Galactic latitude fields have no counterparts and need be observed with large telescopes.

Statistical Sieveing

Many of our identifications will be with objects bright enough to be in the HST GSC, but within our 2 arcmin radius error circles there are typically several GSC objects. A maximum likelihood method (de Ruiter *et al.* 1977, Prestage and Peacock 1983) is powerful for picking out the correct identifications. This calculates the likelihood that an optical candidate in the error circle is correct given its position in the error circle and the local background optical object density at that magnitude. This yields 57 unambiguous identifications with GSC objects that had no previous counterpart (Schachter *et al.* 1992). This method readily allows the inclusion of other factors, such as X-ray to optical ratios, or a stellar/non-stellar flag to separate galaxies from stars.

RAPID DISSEMINATION OF THE WHOLE DATA BASE

Naturally astronomers working on on a large project tend to milk all the interesting information from it before making it public. Instead The Slew Survey source catalog and the complete photon and timing data used to construct it were released on CD-ROM (Plummer *et al.* 1991) as soon as the survey was complete (Jan '91) and well before the paper describing the survey was submitted to the Ap.J. (Sept '91) or published (April '92). As software is developed it is placed on-line to assist a user's own Survey analysis. This will encourage others to work on the data and the Slew Survey will become more widely known, used, and referenced. Self-interest and public interest can converge!

Our aim is to get the Slew Survey 100% identified as soon as possible, so we can begin population studies (*e.g.* AGN evolution). The best way to do this is to get others to help. An on-line, regularly updated, source identification list is maintained at CfA (on *einline*, Karakashian *et al.* these proceedings). We will put any proposed identifications sent to us into this data base with proper credit to the discoverers, providing them with an incentive to get there first.

REFERENCES

de Ruiter, H. R., Willis, A. G., & Arp, H. C. 1977, A&A Supp., 28, 211.

Elvis M., Plummer D., Schachter J. and Fabbiano G., 1992, *ApJS*, in press.

Huchra J.P. and Geller M., 1984, *ApJ*, **257**, 423.

Karakashian T. *et al.* 1991, these proceedings.

Plummer D. *et al.* 1991, CD-ROM issued by SAO.

Prestage, R. M., & Peacock, J. A. 1983, MNRAS, 204, 355.

Schachter J. *et al.*, 1992, in preparation.