Astronomical Data Analysis Software and Systems XII
ASP Conference Series, Vol. 295, 2003
H. E. Payne, R. I. Jedrzejewski, and R. N. Hook, eds.

The XMM-Newton Serendipitous Sky Survey

M. G. Watson

Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK on behalf of the XMM-Newton *Survey Science Centre (SSC)*

Abstract. The main properties of the *XMM-Newton* serendipitous sky survey are outlined. The *XMM-Newton* Survey Science Centre's role in the survey is described with emphasis on its follow-up and identification programme and the production of the *XMM-Newton* catalogue.

1. Introduction

Serendipitous X-ray sky surveys have been pursued with most X-ray astronomy satellites since the *Einstein* Observatory. The resultant serendipitous source catalogues have made a significant contribution to our knowledge of the X-ray sky and our understanding of the nature of the various Galactic and extragalactic source populations. The *XMM-Newton* Observatory (Jansen et al. 2001) provides unrivaled capabilities for serendipitous X-ray surveys by virtue of the large field of view of the X-ray telescopes with the EPIC X-ray cameras[1] (Turner et al. 2001; Strüder et al. 2001), and the high throughput afforded by the heavily nested telescope modules. This capability ensures that each *XMM-Newton* observation provides significant numbers of previously unknown serendipitous X-ray sources in addition to data on the original target (Watson et al. 2001).

A follow-up and identification programme for the *XMM-Newton* sources and the compilation of a high quality serendipitous source catalogue from the *XMM-Newton* EPIC observations are major responsibilities of the *XMM-Newton* Survey Science Centre (SSC; Watson et al. 2001). This paper describes these two aspects of the SSC's activities, both of which are ongoing projects.

2. The Role of the XMM-Newton Survey Science Centre (SSC)

The SSC is an international collaboration involving a consortium of ten institutions in the UK, France, Germany, Italy and Spain. The SSC has several major roles in the project: the development of the scientific processing and analysis software for *XMM-Newton*, the routine 'pipeline' processing of all the observations, a follow-up and identification programme for the *XMM-Newton*

[1] There are three EPIC cameras on-board *XMM-Newton*, one in the focal plane of each of the co-aligned X-ray telescope modules: two EPIC MOS cameras and one EPIC pn camera. In normal operations observations are made with all three cameras simultaneously.

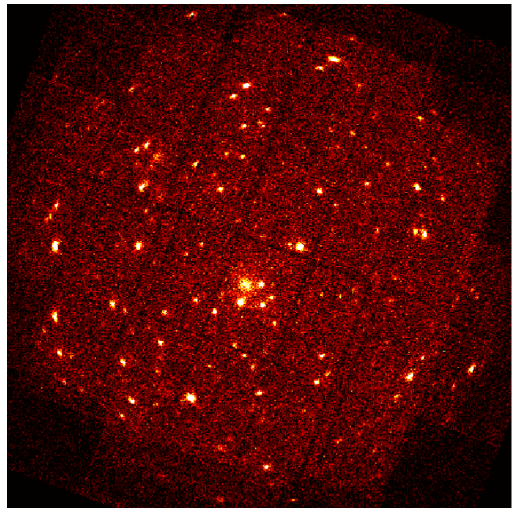

Figure 1. Example *XMM-Newton* X-ray image. Data shown are the combined images from all three EPIC cameras in the 0.5–4.5 keV band. The target of the observation is a relatively faint quasar.

serendipitous survey and the compilation of the *XMM-Newton* Serendipitous Source Catalogue. Here I briefly outline the SSC's activities in the first two areas, the others being covered in later sections.

Working closely with ESA's Science Operations Centre (SOC) staff, the SSC has played a large role in the development of the scientific analysis software for the *XMM-Newton* project: the 'SAS.' SAS modules are used in a fixed configuration for the routine processing of the *XMM-Newton* data, and in an interactive configuration to carry out custom analysis of the data. The SSC also carries out the routine 'pipeline' processing of all the *XMM-Newton* observations from each of the three science instrument packages. The aim of this processing is to provide a set of data products which will be of immediate value for the *XMM-Newton* observer as well as for the science archive. The *XMM-Newton* data products include calibrated, 'cleaned' event lists which provide the starting point for most interactive analysis of the data as well as a number of secondary high-level products such as sky images, source lists, cross-correlations with archival catalogues, source spectra and time series. The data products also constitute the starting point for the compilation of the serendipitous catalogue (see section 5.).

3. Serendipitous Science with XMM-Newton

The high throughput, large field of view and good imaging capabilities of *XMM-Newton* mean that it detects significant numbers of serendipitous X-ray sources in each pointing in addition to the main target of the observation, as is illustrated in Figure 1. Typical observations yield ~ 50 serendipitous sources per field. As *XMM-Newton* makes of the order 700 observations per year, covering ~ 150 sq.deg. of the sky, the number of serendipitous sources is thus growing at a rate of ~ 50000 sources per year, i.e., the annual rate is comparable in size to the complete *ROSAT* All Sky Survey, but reaches fluxes 2–3 orders of magnitude fainter. *XMM-Newton* data thus provides a deep, large area sky survey which represents a major resource for a wide range of programmes (see section 7.). The extended energy range of *XMM-Newton*, compared with previous imaging X-ray missions such as *ROSAT* and the *Einstein* Observatory, means that *XMM-Newton* detects significant numbers of obscured and hard-spectrum objects (e.g., obscured AGNs) which are absent from earlier studies.

In comparison with *Chandra*, typical *XMM-Newton* observations cover a significantly larger sky area but do not go as deep due to the superior imaging provided by the *Chandra* mirrors. *XMM-Newton* has a very significant advantage at photon energies above 4–5 keV: at low energies the ratio of *XMM-Newton*/*Chandra* effective areas is 3–4 but this increases to ~ 6 at 5 keV and > 10 at 7 keV.

4. The 'XID' Follow-up Programme

4.1. Need for Follow-up Studies

In order to exploit the full potential of the *XMM-Newton* serendipitous survey in the context of a wide range of scientific programmes, the key initial step is the 'identification' of the X-ray sources, i.e., a knowledge of the likely classification into different object types. For *XMM-Newton* serendipitous sources, the X-ray observations themselves will provide the basic parameters of each object: the celestial position, X-ray flux and colours for all sources and information on the X-ray spectrum, spatial extent and temporal variability for the brighter objects detected. This information alone will, in some cases, be sufficient to provide a clear indication of the type of object, but for the vast majority of sources, additional information will be required before a confident classification of the object can be made. Some of this information can come from existing astronomical catalogues, or from existing or planned large-scale optical, IR and radio surveys (e.g., SDSS, 2MASS, Denis, FIRST/NVSS), but full exploitation of the *XMM-Newton* serendipitous survey requires a substantial programme of new observations, primarily in the optical and IR using ground-based facilities. The SSC's XID follow-up programme is designed to meet these aims. The main new observational elements of the XID programme, the 'Core Programme' and the 'Imaging Programme' are outlined below.

4.2. The XID Core Programme

The aim of the Core Programme is to obtain the identifications for a well-defined sample of X-ray sources drawn from selected *XMM-Newton* fields, primarily using optical/IR imaging and spectroscopy. Imaging is required both to locate potential candidates accurately and reveal their morphology, whilst the optical spectroscopy provides the diagnostics needed both for object classification and for determining basic object parameters such as redshift and spectral slope. The principal objective is to obtain a completely identified sample which can be used to characterise the *XMM-Newton* source population overall sufficiently well that we can use the basic X-ray and optical parameters to assign a 'statistical' identification for a large fraction of **all** the sources in the *XMM-Newton* serendipitous source catalogue.

Table 1. XID Core Programme samples.

Sample	Flux range[a]	Sky dens.[b]	# EPIC[c]	r′ mag.[d]
FAINT	$\geq 10^{-15}$	2200	5–10	23–25
MEDIUM	$\geq 10^{-14}$	340	30–50	21–23
BRIGHT	$\geq 10^{-13}$	10	1000	17–21
GALACTIC	$\geq 5\ 10^{-15}$	~ 300	40	wide range

[a] X-ray flux in erg cm^{-2} s^{-1} in the 0.5–4.5 keV band

[b]source density (deg^{-2})

[c]no.of EPIC fields required

[d]typical r' magnitude range of counterparts

The Core Programme has two main parts: one for the high and one for the low galactic latitude sky. The high galactic latitude part consists of three samples, each containing $\approx$ 1000 X-ray sources in three broad flux ranges as summarised in Table 1. The size of the subsamples is dictated by the need to identify enough objects to reveal minority populations. Studying sources at a range of X-ray fluxes is necessary because we already know that the importance of different source populations changes with X-ray flux level.

Substantial progress has been made on the XID Core Programme over the last two years, in particular in the medium, bright and Galactic samples (Barcons et al. 2002a,b; Della Ceca et al. 2002; Motch et al. 2002). A large fraction of the ground-based observations needed to support this programme have come from the AXIS programme which gained a substantial international time allocation on the Canary Islands telescopes (Barcons et al. 2001). For example a significant number of identifications have already been made in 19 medium sample fields included in the XID programme. Of the 239 medium sample sources in these fields around 2/3 of these have already been identified (see Figure 2); the remaining sources are typically too faint for 4m-class telescopes and will be pursued with larger facilities over the next year. When complete, the XID medium sample will bridge the gap between the deepest surveys aimed at studying the most distant Universe, and shallower surveys which aim to characterise the local and high luminosity X-ray source populations.

4.3. The XID Imaging Programme

The XID imaging programme aims to obtain optical/IR photometry and colours for a large number of *XMM-Newton* fields. The rationale is that a combination of X-ray flux & X-ray colours (from the *XMM-Newton* data) and optical magnitude

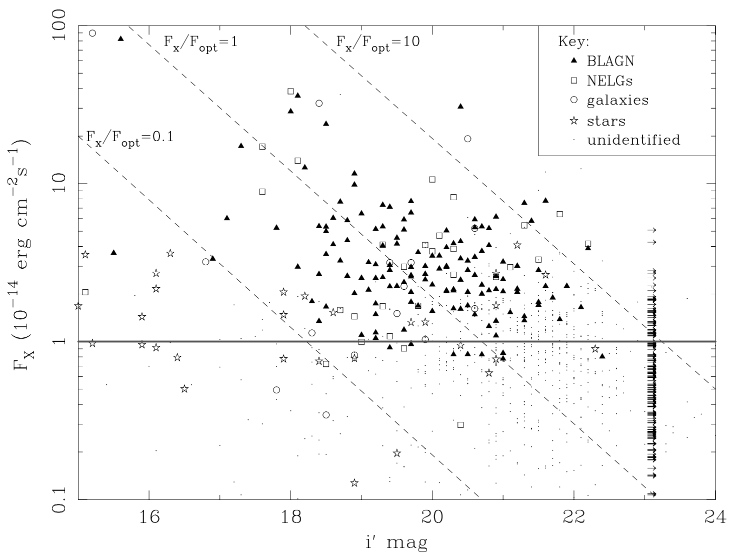

Figure 2. Distribution of X-ray and optical fluxes for serendipitous sources identified in the XID programme. The symbols indicate object classifications as given in the key. The nominal limit for the XID **medium** sample is indicated by the horizontal line.

and optical colours (from new ground-based observations) will provide the key parameters which make possible an accurate 'statistical' identification of the *XMM-Newton* sources. This will be possible using the results from the Core Programme to provide the 'calibration' of the basic parameters for different object types.

To date deep multi-colour optical imaging has been obtained for more than 150 *XMM-Newton* fields using 2m-class telescopes in both hemispheres. The prospects are good to expand this to 250–300 fields over the next two years.

5. The XMM-Newton Catalogue

5.1. Catalogue fields

The first installment of the catalogue (Watson et al. 2002) will include around 700 *XMM-Newton* observations ('fields') selected purely on the public availability of the fields at the planned release date (end 2002) and the observation being made in one of the appropriate EPIC imaging modes. No other selection of fields, e.g., in terms of sky location, exposure time etc., has been made. Figure 3 shows the sky distribution of the catalogue fields. The overall sky distribution is reasonably uniform, although there are some biases such as the paucity of fields in the Cygnus region due to *XMM-Newton* visibility constraints. The average exposure time per catalogue field is $\sim$ 20 ksec for the EPIC MOS cameras and $\sim$ 15 ksec for the EPIC pn camera.

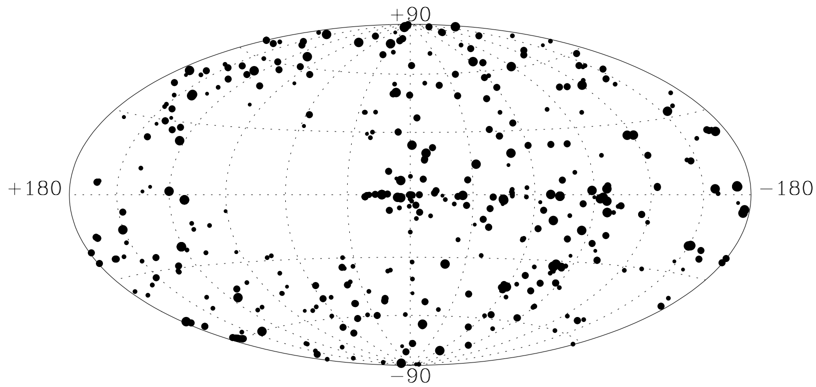

Figure 3. Sky distribution of the *XMM-Newton* fields in the catalogue shown in Galactic coordinates. Size of the symbols reflects the total number of serendipitous sources detected: the average number is ~ 50 per field.

5.2. Data Processing

Data processing for the production of the *XMM-Newton* catalogue is based closely on the standard SSC pipeline (see section 2.) used to process each *XMM-Newton* dataset for distribution to the observer, and population of the *XMM-Newton* Science Archive. Catalogue processing uses a fixed software and calibration data configuration in order to guarantee uniformity. The main processing stages for each *XMM-Newton* observation are:

- production of calibrated events from the ODF science frames;
- generation of the appropriate low-background time intervals using a threshold optimized for point source detection;
- generation of multi-energy-band X-ray images and exposure maps from the calibrated events;
- a four-stage source detection and parameterization procedure:
    - generation of a preliminary source list using a sliding-box technique and local background estimation;
    - generation of a background map from 2-D spline fits to the images with bright sources excised;
    - generation of a refined source list again using a sliding-box technique but employing the background map; and
    - maximum-likelihood fitting and parameterization of sources in the refined list.
- merging of the three camera-level source lists into an EPIC-level source list with merging on the basis of positional coincidence alone; and
- cross-correlation of the source list with a variety of archival catalogues and other resources using the CDS facilities in Strasbourg.

The source search approach utilized involves simultaneous fitting of five energy band images for each EPIC camera, thus producing a source list for each camera which contains source and detection information in each energy band (as well as the total band). The camera lists combined in the penultimate stage described above thus produce a merged source list which forms the reference source list for that observation.

5.3. Quality Control

Although the source detection algorithm described above is now mature, typically producing reliable source lists from most *XMM-Newton* observations, the approach is not perfect and is known to have problems in producing reliable results in a number of (rare) circumstances. Each *XMM-Newton* field included in the catalogue is therefore visually screened to locate such defects. Where problems are noted the sources affected are 'flagged' and the flag values transferred to the source lists. In rare cases ($< 10\%$ of the total) the entire field has significant problems, e.g., very high background or very high surface brightness diffuse sources, which mean that it is of marginal value for detecting serendipitous sources. Such fields will be excluded from the final catalogue. Apart from these rare cases, the median fraction of sources flagged as being spurious amounts to only 4% overall, reflecting the maturity of the source detection approach.

6. Catalogue Properties

6.1. Source Numbers

The working catalogue contains a total of $\sim$ 78000 source detections in any EPIC camera (i.e., in one or more cameras) and a total of $\sim$ 11000 sources detected in all three EPIC cameras. These numbers refer to a broad-band (0.2–12 keV) detection above a likelihood of 10, corresponding to $\approx 4\sigma$. At this significance the *a priori* probability of spurious detections is low, corresponding to < 1 spurious source per field, although simulations are underway to verify the calibration of the likelihood parameterisation. The total sky area covered for detections in any camera is $\sim$ 130 sq.deg., whilst for detection in all three cameras the area is $\sim$ 90 sq.deg.

For the released version of the catalogue these numbers will be reduced by the fraction flagged as spurious (and the small number of fields totally excluded). We also anticipate setting a somewhat higher likelihood threshold once we have completed our investigation of the reliability of detections as a function of likelihood, being pursued by simulations.

6.2. Source Count Distribution

Figure 4 shows the $\log N - \log S$ distribution for all EPIC pn sources in the working catalogue. The distribution is **not** corrected for sky coverage, i.e., how the actual sky area covered varies with X-ray flux. Comparing the uncorrected $\log N - \log S$ with the expected source counts demonstrates that the catalogue is essentially complete down to an X-ray flux $f_X \approx 4 \times 10^{-14}$ erg cm^{-2} s^{-1} (0.2–12 keV) (equivalent to $f_X \approx 2 \times 10^{-14}$ erg cm^{-2} s^{-1} in the 0.5–2 keV band and $f_X \approx 8 \times 10^{-15}$ erg cm^{-2} s^{-1} in the 2–10 keV band). This limit is in line with expectations given the exposure time distribution of catalogue fields. Around 30% of the catalogue sky area is covered to $f_X \approx 10^{-14}$ erg cm^{-2} s^{-1} (0.2–12 keV). Work to establish the definitive coverage corrections is underway and will

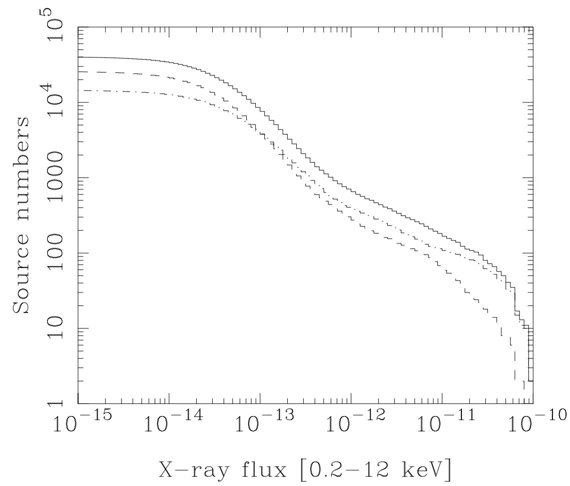

Figure 4. Uncorrected $\log N - \log S$ distribution for all EPIC pn detections in the working catalogue. The solid curve is for all sources, the dashed curve is for high latitude sources ($|b| > 20°$) and the dashed-dot curve for low latitude sources ($|b| < 20°$).

be part of the ancillary information included with the released version of the catalogue.

6.3. Astrometry

For each *XMM-Newton* field, the catalogue processing (section 5.2.) attempts to correct the astrometric reference frame using cross-correlation of the *XMM-Newton* source list with the USNO A2.0 catalogue (Monet et al. 1998). The technique employed involves finding the maximum likelihood in a grid of trial astrometric shifts and rotations with a likelihood function depending on the angular separation between each potential *XMM-Newton*–USNO object match. If an acceptable solution is found ($>$ 70% of fields) the resultant astrometric correction is applied to the *XMM-Newton* source list for that field. (The cases where an acceptable solution is not found are primarily fields with low numbers of X-ray sources and/or fields with high optical object density).

The results of applying this technique to the catalogue fields can be employed to quantify the initial accuracy of the astrometry of each *XMM-Newton* field (i.e., before correction). Fits to the distribution of shifts in RA, Dec and field rotation imply that the intrinsic accuracy of the *XMM-Newton* field astrometry (as determined solely from the in-orbit attitude solution) can be characterized by a Gaussian with $\sigma \approx 1.5$ arcsec. *After* correction using this technique, the residual field astrometric errors are of the order 0.5–1 arcsec, close to the nominal 1 arcsec astrometric accuracy of the USNO catalogue itself.

As the typical *statistical* error-circle for a faint *XMM-Newton* source has $\sigma_{\rm stat} \approx 1 - -2$ arcsec, the size of the field systematic component determined justifies the ~ 5 arcsec positional accuracy which has been assumed to date as the effective $\sim 90\%$ confidence radius of uncorrected positions, e.g., for the identification of *XMM-Newton* source counterparts (see section 4.).

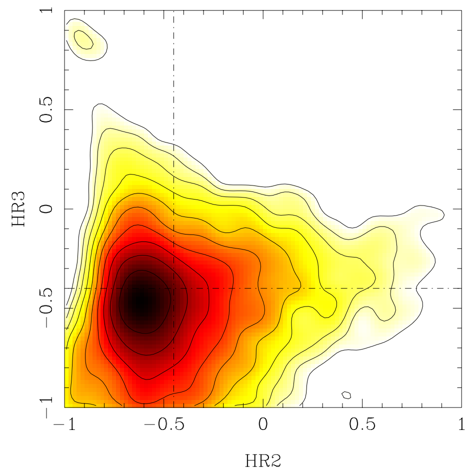

Figure 5. X-ray colour-colour plot showing the distribution of HR2-HR3 values for EPIC pn catalogue sources ($|b| > 20°$ only). The gray-scale and contours indicate the logarithmic density of points down to $\sim 1\%$ of the peak. The dashed-dot lines indicate the approximate HR2, HR3 values for the X-ray background. HR2 is the ratio (B-A)/(A+B) and HR3 is the ratio (C-B)/(B+C) where A= 0.5–2 keV count rate, B=2–4.5 keV count rate and C=4.5–7.5 keV count rate.

7. Scientific Potential of the Catalogue

The *XMM-Newton* catalogue represents a significant resource that can be used for a variety of astrophysical projects. Although deep Chandra and *XMM-Newton* pencil-beam surveys (e.g., Mushotzky et al. 2000; Hasinger et al. 2001; Giacconi et al. 2001; Brandt et al. 2001) have probed the faintest parts of the extragalactic source population, the *XMM-Newton* catalogue also can make a major contribution. The *XMM-Newton* catalogue reaches modest depths ($f_X \approx 10^{-14}$ erg cm^{-2} s^{-1}) at high coverage (tens of square degrees). As this flux limit is where the bulk of the objects that contribute to the X-ray background lie (due to the fact the $\log N - \log S$ distribution breaks to a flatter slope at around this flux), the large samples of medium-deep flux sources that the *XMM-Newton* catalogue provides will thus be a significant resource for X-ray background studies.

The *XMM-Newton* catalogue also provides a rich resource for generating well-defined samples for specific studies, utilizing the fact that X-ray selection is a highly efficient (arguably the most efficient) way of selecting certain types of object, notably AGN, clusters of galaxies, interacting compact binaries and very active stellar coronae. AGN samples have obvious value in evolution studies and cluster samples can provide, potentially, key measurements of cosmological parameters. Selecting 'clean' samples requires knowledge of the likely parameter ranges of different types of object: some of this is already known but further 'calibration' of this concept is one of the key aims of the SSC's XID programme (section 4.; Watson et al. 2001; Barcons et al. 2002a,b). The inclusion of matches with archival catalogues for each *XMM-Newton* catalogue source also provides a valuable starting point for investigation of well-defined samples.

To illustrate some of the potential described above, Figure 5 shows the overall distribution of X-ray colours in the catalogue. This figure provides a glimpse of the power of the *XMM-Newton* catalogue for providing interesting samples. The spectrum of the X-ray background corresponds to HR2 ≈ -0.45, HR3 ≈ -0.4, but evidently the bulk of the catalogue sources have spectra softer than this. Thus merely by extracting a subset of *XMM-Newton* sources with X-ray colours harder than these limits, one automatically selects those objects that must be an important constituent of the background. Optical and near-IR observing programmes to investigate the nature of the hard source samples constructed in this manner are already underway.

As well as the potential for building up large samples, the brighter sources in the *XMM-Newton* catalogue also provide the prospect of obtaining high quality X-ray spectra and time series data (and morphology). Amongst the catalogue sources more than 15% have more than 200 EPIC pn counts, enough for a reasonable spectral characterization and crude variability indications, whilst 3% have more than 1000 counts, sufficient for a very good X-ray spectral measurement or variability analysis.

Acknowledgments. The SSC's activities relating to the *XMM-Newton* serendipitous sky survey are a consortium effort involving all the SSC institutions. I gratefully acknowledge the efforts of many colleagues which continue to contribute to the success of these projects.

References

Barcons, X. et al. 2001, New Century of X-ray Astronomy, eds. H. Inoue & H. Kunieda (San Francisco: ASP), 160

Barcons, X. et al. 2002a, A&A, 382, 522

Barcons, X. et al. 2002b, Astron Nachr., in press

Brandt, W. N. et al. 2001, AJ, 122, 1

Della Ceca, R. et al. 2002, New visions of the X-ray Universe in the XMM-Newton and Chandra era, in press, astro-ph/0202150

Giacconi, R. et al. 2001, ApJ, 551, 624

Hasinger, G. et al. 2001, A&A, 365, L45

Jansen, F. et al. 2001, A&A, 365, L1

Monet, D. et al. 1998, The USNO-A2.0 Catalogue, (U.S. Naval Observatory, Washington DC; VizieR Online Data Catalog, 1252)

Motch, C. et al. 2002, New visions of the X-ray Universe in the XMM-Newton and Chandra era, in press, astro-ph/0203025

Strüder, L. et al. 2001, A&A, 365, L18

Turner, M. J. L. et al. 2001, A&A, 365, L110

Watson, M. G. et al. 2001, A&A, 365, L51

Watson, M. G. et al. 2002, Astron Nachr., in press