Open Dartmouth: Faculty Open Access Articles

10-28-2013

Finding Rare AGN: XMM-Newton and Chandra Observations of SDSS Stripe 82

Stephanie M. LaMassa
Yale University

C. Megan Urry
Yale University

Nico Cappelluti
University of Maryland at Baltimore

Francesca Civano
Dartmouth College

Follow this and additional works at: https://digitalcommons.dartmouth.edu/facoa

Part of the External Galaxies Commons

Finding rare AGN: *XMM–Newton* and *Chandra* observations of SDSS Stripe 82

Stephanie M. LaMassa,[1][*] C. Megan Urry,[1] Nico Cappelluti,[2,3] Francesca Civano,[4,5]
Piero Ranalli,[2,6] Eilat Glikman,[1] Ezequiel Treister,[7] Gordon Richards,[8]
David Ballantyne,[9] Daniel Stern,[10] Andrea Comastri,[2] Carie Cardamone,[11]
Kevin Schawinski,[12] Hans Böhringer,[13] Gayoung Chon,[13] Stephen S. Murray,[4,14]
Paul Green[4] and Kirpal Nandra[13]

[1]*Yale Center for Astronomy and Astrophysics, Physics Department, Yale University, PO Box 208120, New Haven, CT 06520-8120, USA*
[2]*INAF – Osservatorio Astronomico di Bologna, Via Ranzani 1, I-40127 Bologna, Italy*
[3]*Center for Space Science and Technology, Physics Department, University of Maryland Baltimore College, 1000 Hilltop Circle, Baltimore, MD 21250, USA*
[4]*Physics and Astronomy Department, Dartmouth College, Wilder Lab, Hanover, NH 03755, USA*
[5]*Harvard–Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*
[6]*National Observatory of Athens, Institute of Astronomy, Astrophysics, Space Applications and Remote Sensing Metaxa & Pavlou St, Penteli GR-15236 Athens, Greece*
[7]*Universidad de Concepción, Casilla 160-c Concepción, Chile*
[8]*Department of Physics, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA*
[9]*Center for Relativistic Astrophysics, School of Physics, Georgia Institute of Technology, Atlanta 30332, GA, USA*
[10]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Mail Stop 169-221, Pasadena, CA 91109, USA*
[11]*The Harriet W. Sheridan Center for Teaching and Learning, Brown University, Box 1912, 96 Waterman Street, Providence, RI 02912, USA*
[12]*Institute for Astronomy, Department of Physics, ETH Zürich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*
[13]*Max-Planck-Institut Für Extraterrestriche Physik, D-85748 Garching, Germany*
[14]*Department of Physics and Astronomy, The Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA*

ABSTRACT

We have analysed the *XMM–Newton* and *Chandra* data overlapping $\sim$16.5 deg^2 of Sloan Digital Sky Survey Stripe 82, including $\sim$4.6 deg^2 of proprietary *XMM–Newton* data that we present here. In total, 3362 unique X-ray sources are detected at high significance. We derive the *XMM–Newton* number counts and compare them with our previously reported *Chandra* log$N$–log$S$ relations and other X-ray surveys. The Stripe 82 X-ray source lists have been matched to multiwavelength catalogues using a maximum likelihood estimator algorithm. We discovered the highest redshift ($z = 5.86$) quasar yet identified in an X-ray survey. We find 2.5 times more high-luminosity ($L_x \geq 10^{45}$ erg s^{-1}) AGN than the smaller area *Chandra* and *XMM–Newton* survey of COSMOS and 1.3 times as many identified by XBoötes. Comparing the high-luminosity active galactic nuclei (AGN) we have identified with those predicted by population synthesis models, our results suggest that this AGN population is a more important component of cosmic black hole growth than previously appreciated. Approximately a third of the X-ray sources not detected in the optical are identified in the infrared, making them candidates for the elusive population of obscured high-luminosity AGN in the early universe.

Key words: catalogues – surveys – galaxies: active – quasars: general – quasars: supermassive black holes – X-rays: galaxies.

1 INTRODUCTION

Supermassive black holes (SMBHs) that reside in galactic centres grow by accretion in a phase where they appear as active galactic nuclei (AGN). To understand AGN demography and evolution,

large samples over a range of redshifts and luminosities are necessary. Extragalactic surveys provide an ideal mechanism for locating large enough samples of growing black holes to study the ensemble statistically. Large-area surveys have been undertaken in the optical via, e.g. the Sloan Digital Sky Survey (SDSS; Ahn et al. 2012) and in the near-infrared (NIR) via the *Wide-Field Infrared Survey Explorer* (*WISE*; Wright et al. 2010) and the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al. 2007), locating over

[*] E-mail: stephanie.lamassa@yale.edu

100 000 AGN in the optical and millions of AGN candidates in the infrared.

However, optical selection is not ideal for studying high-luminosity, high-redshift AGN (quasars) that are heavily reddened or obscured. At redshifts greater than 0.5, diagnostic diagrams that use ratios of narrow emission lines to identify type 2 (obscured) AGN (e.g. Baldwin, Phillips & Terlevich 1981; Kewley et al. 2001; Kauffmann et al. 2003) become inefficient as Hα is shifted out of the optical. Such type 2 AGN can be found using alternate rest-frame optical diagnostics, using e.g. ratios of narrow emission lines versus $g - z$ colour (TBT; Trouille, Barger & Tremonti 2011) and versus stellar mass (MEx; Juneau et al. 2011), probing out to distances $z < 1.4$ and $z < 1$, respectively. Narrow rest-frame UV emission lines also allow identification of SMBH accretion at $z > 0.5$. Alternatively, obscured AGN candidates can be followed up with ground-based infrared spectroscopy to detect redshifted Hα and [N II]λ6584. However, the Kewley et al. (2001) and Kauffmann et al. (2003) boundaries between star-forming galaxies, composites and Sy2s are only calibrated at low redshifts. As galaxies beyond $z > 0.5$ have lower metallicities, updated diagnostics are necessary to unambiguously identify signatures of SMBH accretion (Kewley et al 2013a,b).

The reliability of infrared colour selection varies with the depth of the data, with *Spitzer* IRAC colour cuts (Lacy et al. 2004; Stern et al. 2005) and *WISE* colour cuts (Assef et al. 2013; Stern et al. 2012) being most applicable at shallow depths. At fainter fluxes, contamination from normal galaxies can become appreciable (Cardamone et al. 2008; Donley et al. 2012; Mendez et al. 2013). The revised IRAC colour selection from Donley et al. (2012) is more reliable for deeper data, yet at X-ray luminosities exceeding 10^{44} erg s^{-1}, 25 per cent (32 per cent) of the *XMM–Newton-* (*Chandra-*) selected AGN are not recovered with this mid-infrared (MIR) identification method.

X-rays provide an alternate way to search for AGN, complementing the optical and MIR identification techniques to provide a comprehensive view of black hole growth over cosmic time, because X-rays can pierce through large amounts of dust and gas. Their emission is visible out to cosmological distances as long as it is not attenuated by Compton-thick ($N_{\rm H} \geq 10^{24}$ cm^{-2}) obscuration. Normal star formation processes rarely exceed an X-ray luminosity above 10^{42} erg s^{-1} (e.g. Persic et al. 2004; Brandt & Hasinger 2005), whereas AGN luminosities extend to $\sim 10^{46}$ erg s^{-1}, making X-ray selection an efficient means for locating AGN at all redshifts. Indeed, X-ray surveys, such as the *Chandra* Deep Fields North (Alexander et al. 2003) and South (Giacconi et al. 2001; Xue et al. 2011), Extended *Chandra* Deep Field South (E-CDFS; Lehmer et al. 2005; Virani et al. 2006), *XMM–Newton* survey of the *Chandra* Deep Field South (Comastri et al. 2011; Ranalli et al. 2013), *XMM–Newton* and *Chandra* surveys of COSMOS (Cappelluti et al. 2007, 2009; Elvis et al. 2009; Brusa et al. 2010; Civano et al. 2012), XBoötes (Kenter et al. 2005; Murray et al. 2005), the *XMM–Newton* survey of the Lockman Hole (Brunner et al. 2008), *Chandra* observations of All-Wavelength Extended Groth Strip International Survey (Davis et al. 2007; Georgakakis et al. 2007), XDEEP2 (Goulding et al. 2012), the *XMM–Newton* Serendipitous Survey (Mateos et al. 2008) and the *Chandra* multiwavelength campaign (ChaMP; Kim et al. 2007), have identified thousands of AGN, contributing significantly to our knowledge of AGN demography and galaxy and SMBH co-evolution.

However, most of these X-ray surveys cover small (<1 deg^2) to moderate (3–5 deg^2) areas, sacrificing area for depth to uncover the faintest X-ray objects. The *XMM*-COSMOS (Cappelluti et al. 2007,

2009; Brusa et al. 2010), *Chandra* COSMOS (C-COSMOS Elvis et al. 2009; Civano et al. 2012) and ongoing *Chandra* COSMOS Legacy Project (PI: Civano) strikes a good balance of moderate area at moderate depth to populate a large portion of the $L_{\rm x} - z$ plane. But sources that are rare, like high-luminosity and/or high-redshift AGN, are under-represented in these small to moderate area X-ray samples as a larger volume of the Universe must be probed to locate them.

Considerable follow-up (optical/NIR imaging and spectroscopy) is needed to identify X-ray sources, and multiwavelength data are needed to classify these objects. Since spectroscopic campaigns and multiwavelength follow-up are time intensive, the output from wide area surveys, such as XBoötes (~ 9 deg^2; Kenter et al. 2005; Kochanek et al. 2012), ChaMP (~ 33 deg^2; Kim et al. 2007; Trichas et al. 2012) and *XMM*-LSS (~ 11 deg^2, the first part of the expanded *XMM*-XXL 50 deg^2 survey; Pierre et al. 2004; Chiappetti et al. 2013), has taken many years to achieve. The high-redshift X-ray-selected luminosity AGN population therefore remains poorly explored, prohibiting a comprehensive view of black hole growth.

To address this gap, we have begun a wide area X-ray survey in a region that already has a rich investment in multiwavelength data and a high level of optical spectroscopic completeness (> 800 objects deg^2): the SDSS Stripe 82 region, which spans 300 deg^2 along the celestial equator ($-60° < $ RA $< 60°$, $-1°.25 < $ Dec. $< 1°.25$). The current non-overlapping X-ray coverage in Stripe 82 from archival *Chandra* and archival and proprietary *XMM–Newton* observations is ~ 16.5 deg^2. The distribution of these pointings across Stripe 82 is shown in Fig. 1. As we are endeavouring to increase the survey area to ~ 100 deg^2, we dub the present survey 'Stripe 82X Pilot'. Here, we follow-up on the work presented in LaMassa et al. (2013) where we focused on just the *Chandra* overlap with Stripe 82, by adding in ~ 10.5 deg^2 of *XMM–Newton* observations, 4.6 deg^2 of which were obtained by us as part of an approved AO 10 proposal (PI: Urry), with the observations performed in 'mosaic' mode. We then match both catalogues to large optical (SDSS; Ahn et al. 2012), NIR (UKIDSS and *WISE*; Lawrence et al. 2007; Wright et al. 2010), ultraviolet (*GALEX*; Morrissey et al. 2007) and radio data sets (FIRST; Becker, White & Helfand 1995) in this region. Observations covering Stripe 82 with *Spitzer* (PI: Richards) and analysis of *Herschel* observation overlapping ~ 135 deg^2 of the region (Viero et al. 2013) are ongoing.

In Section 2, we discuss the reduction and analysis of the archival and proprietary mosaicked *XMM–Newton* data in Stripe 82. We use these data to calculate area–flux curves and in Section 3 present the $\log N - \log S$ relations, which we compare to the *Chandra* Stripe 82 number counts (LaMassa et al. 2013) and those from other X-ray
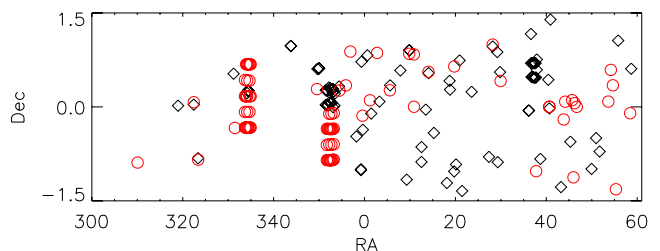

Figure 1. X-ray observations overlapping Stripe 82 used in this analysis, with *Chandra* observations shown as black diamonds and *XMM–Newton* pointings depicted as red circles. The dense *Chandra* pointings are part of the XDEEP2 survey (Goulding et al. 2012) while the dense *XMM–Newton* groupings represent the positions of the proprietary mosaicked observations we were awarded in AO 10.

surveys. We then describe in Section 4 the matching of the *XMM–Newton* and *Chandra* X-ray source lists with multiwavelength catalogues, producing multiwavelength source lists. In Section 5, we describe the general characteristics of the Stripe 82X sample so far. In particular, we highlight the interesting science gaps our data are primed to fill: uncovering the population of rare high-luminosity AGN at high redshift and identifying candidates for high-luminosity obscured AGN at $z > 1$. We have adopted a cosmology of $H_0 = 70$ km s^{-1} Mpc^1, $\Omega_M = 0.27$ and $\Lambda = 0.73$ throughout the paper.

2 *XMM–NEWTON* DATA REDUCTION

2.1 Archival observations

57 *XMM–Newton* EPIC non-calibration observations overlap Stripe 82. Of these, 24 were removed due to flaring, substantial pile-up and read-out streaks, small-window mode set-up or extended emission spanning the majority of the detector, all of which complicate serendipitous detections of point sources in the field. We were left with 33 archival observations well suited for our analysis, listed in Table 1 and shown as red circles in Fig. 1; for 3 of these we dropped the PN detector due to significant pile-up which did not affect the MOS detectors as seriously.

The raw observational data files were processed with *XMM–Newton* Standard Analysis System (SAS) version 11. SAS tasks *emchain* and *epchain* were run to generate MOS1 and MOS2 event files as well as PN and PN out-of-Time (OoT) event files. OoT events result from photons detected during CCD read-out, when photons are recorded at random position along the read-out column in the *y*-direction. The subsequent energy correction for these OoT events will then be incorrect. The fraction of OoT events is highest for the PN detector in full-frame mode, affecting ∼6.3 per cent of observing time. By generating simulated OoT event files, the PN images can be statistically corrected for this effect.

Good time intervals (GTIs) were applied to the data by searching for flaring in the high-energy background (10–12 keV for MOS, 12–14 keV for PN and PN OoT), removing intervals where the count rate was ≥3σ above the average. Low-energy flares were removed from this filtered event list by removing intervals where the count rate was ≥3σ above the average in the 0.3–10 keV range. In both the high-energy and low-energy cleaning, GTIs were extracted from single events (i.e. PATTERN = 0).

MOS images were extracted from all valid events (PATTERN 0–12), whereas the PN and PN OoT images were extracted from the single and double events only (PATTERN 0–4). To avoid emission line features from the detector background (i.e. Al Kα at 1.48 keV), the energy range 1.45–1.54 keV was excluded when extracting images from both the MOS and PN detectors. The PN background also has strong emission from Cu at ∼7.4 and ∼8.0 keV, so the 7.2–7.6 and 7.8–8.2 keV ranges were also excluded when extracting images from the PN detector. The PN OoT images were scaled by 0.063 to account for the loss of observing time due to photon detection during CCD read-out, and were then subtracted from the PN images. Finally, MOS and PN images were extracted in the standard 0.5–2, 2–10 and 0.5–10 keV ranges and were added among the detectors in each energy band.[1]

Exposure maps were generated using the SAS task *eexpmap* for each detector and energy range. Since vignetting, decrease in effec-

Table 1. Archival *XMM–Newton* observations in SDSS Stripe 82.

Obs. ID	RA	Dec.	Detectors	Exp time (ks)
0036540101	54.64	0.34	MOS1,MOS2,PN	21.77
0041170101	45.68	0.11	MOS1,MOS2,PN	50.04
0042341301[a]	354.44	0.26	MOS1,MOS2,PN	13.36
0056020301[a, b]	44.16	0.08	MOS1,MOS2,PN	23.37
0066950301[a]	349.54	0.28	MOS1,MOS2	11.45
0084230401[a]	28.20	0.99	PN	23.74
0090070201[a]	10.85	0.84	MOS1,MOS2,PN	20.53
0093030201[a]	322.43	0.07	MOS1,MOS2,PN	57.43
0101640201[a]	29.94	0.41	MOS1,MOS2,PN	10.54
0111180201[a]	310.06	−0.89	MOS1,MOS2,PN	16.31
0111200101[a, b, c]	40.65	0.00	MOS1,MOS2	38.39
0111200201[a, b, c]	40.65	0.00	MOS1,MOS2	37.99
0116710901[a]	54.20	0.59	MOS1,MOS2	7.64
0134920901	58.45	−0.10	MOS1,MOS2,PN	18.69
0142610101[a]	46.69	0.00	PN	65.89
0147580401	356.88	0.88	MOS1,MOS2,PN	15.12
0200430101	55.32	−1.32	MOS1,MOS2,PN	11.46
0200480401	37.76	−1.03	MOS1,MOS2,PN	16.07
0203160201[a, c]	46.22	0.06	MOS1,MOS2	15.08
0203690101	9.83	0.85	MOS1,MOS2,PN	47.31
0211280101[a]	355.89	0.34	MOS1,MOS2,PN	40.68
0303110401	14.07	0.56	MOS1,MOS2,PN	11.09
0303110801	359.55	−0.14	MOS1,MOS2,PN	9.63
0303562201	10.88	0.00	MOS1,MOS2,PN	6.57
0304801201	323.39	−0.84	MOS1,MOS2,PN	13.27
0305751001	1.20	0.11	MOS1,MOS2,PN	15.07
0307000701	45.97	−1.12	MOS1,MOS2,PN	15.84
0312190401[b]	43.82	−0.20	MOS1,MOS2,PN	11.63
0400570301	19.75	0.65	MOS1,MOS2,PN	25.94
0401180101	331.47	−0.34	MOS1,MOS2,PN	40.13
0402320201	53.64	0.09	MOS1,MOS2,PN	10.51
0403760301	2.76	0.86	MOS1,MOS2,PN	25.46
0407030101[a]	5.58	0.26	MOS1,MOS2,PN	27.15

[a]Detector mask manually updated to screen out regions of pile-up and extended emission.
[b]Overlapping observations that were run simultaneously through source detection software: 0056020301 and 0312190401 grouped together; 0111200101 and 0111200201 grouped together.
[c]PN detector removed from analysis due to significant pile-up.

tive area with off-axis distance, increases as a function of energy, we created spectrally weighted exposure maps, i.e. the mean energy at which the maps were calculated was found assuming a spectral model where, consistent with previous *XMM–Newton* surveys (e.g. Cappelluti et al. 2007), $\Gamma = 2.0$ in the soft band and $\Gamma = 1.7$ in the hard and full bands, since the spectral slope of the soft band in AGN tends to be steeper than the hard band. The same spectral model was used to derive energy conversion factors (ECFs) to transform count rates to physical flux units, where the ECF depends on the filter for the observation and was calculated via PIMMS[2] (see Table 2 for a summary). The exposure maps were added among the three detectors for each observation, normalized by these ECFs.[3]

Two regions in Stripe 82 had multiple X-ray observations (ObsIDs 0056020301, 0312190401 and 0111200101, 0111200201). In order to detect sources from these overlapping observations

---

[1] We note that in some observations, only one or two detectors had data. See Table 1.

[2] http://heasarc.nasa.gov/Tools/w3pimms.html
[3] In observations where only two detectors were active instead of three, the normalization was adjusted accordingly. No normalization was necessary for observations with only one detector.

Table 2. ECFs[a] for each detector and filter[b].

Band	PN Thin	PN Medium	PN Thick	MOS Thin	MOS Medium	MOS Thick
Soft (0.5–2 keV)	7.45	7.36	5.91	2.00	1.87	1.67
Hard (2–10 keV)	1.22	1.24	1.19	0.45	0.42	0.43
Full (0.5–10 keV)	3.26	3.25	2.75	0.97	0.91	0.85

[a] Energy conversion factors in units of counts $s^{-1}/10^{-11}$ erg cm^{-2} s^{-1}.
[b] Assuming a spectral model where $N_H = 3 \times 10^{20}$ cm^{-2} and $\Gamma = 2.0$ for the soft band and $\Gamma = 1.7$ for the hard and full bands. We note that for the PN detector, ECFs were adjusted to account for masking out energy ranges corresponding to background emission lines, as described in the text. For source detection, ECFs were summed among all detectors turned on during the observation.

simultaneously, the events files were mapped to a common set of WCS coordinates using SAS task *attcalc* to update the 'RA_NOM' and 'DEC_NOM' header keywords. The subsequent data products (e.g. images, exposure maps, background maps, detector masks) then share common coordinates. Before running the source detection in 'raster' mode (see Section 2.4), the header keywords 'EXP_ID' and 'INSTRUME' for these files were updated to common values.

2.2 Proprietary observations

We were awarded two *XMM–Newton* mosaicked pointings in AO 10 [PI: Urry, ObsIDs: 0673000101 ('Stripe 82 XMM field 1'), 0673002301 ('Stripe 82 XMM field 2')], covering ∼4.6 deg². With this observing strategy, each pointing has ∼4.56 ks of exposure time and is separated with 15 arcmin spacing. The exposure time in the regions with greatest overlap reaches a depth of ∼8 ks, after the effects of vignetting. The *XMM–Newton* mosaic procedure enables a relatively large region to be surveyed, in this case ∼2.5 deg² per mosaic, while minimizing overhead as after the first pointing, the EPIC offset tables do not need to be calculated (PN) and uploaded (MOS). Each mosaic was made up of 22 individual, overlapping pointings, for a total observing time of 240 ks between both mosaics.

We split the events files for the mosaicked observations into individual pseudo-exposures using the SAS task *emosaic_prep*. Each pseudo-exposure is then reduced in the same way as the archival pointings, producing cleaned events files, spectrally weighted exposure maps and appropriately modelled background maps (see below). As with overlapping archival observations, 'RA_NOM', 'DEC_NOM', 'EXP_ID' and 'INSTRUME' were updated to common values, but 'RA_PNT' and 'DEC_PNT' also had to be set manually to reflect the centre coordinates of each pointing for the point spread function (PSF) to be calculated correctly during source detection. One of the pointings from ObsID 0673002301 (pseudo-exposure field 22) was afflicted by flaring and consequently not used in the source detection. In total, approximately 4.6 deg² of Stripe 82 were covered in these observations. Soft-band images of these two mosaic fields are shown in Fig. 2 with the corresponding exposure maps in Fig. 3.

2.3 Background modelling

Following Cappelluti et al. (2007), we used the following algorithm to model the background. First, we created detection masks for each detector in each energy band for each observation and then ran the SAS task *eboxdetect* with a low detection probability (likemin = 4) to generate a preliminary list of detected sources.
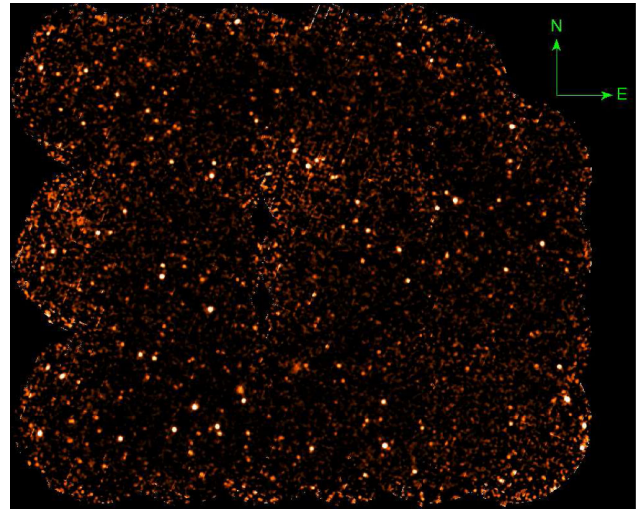
(a)
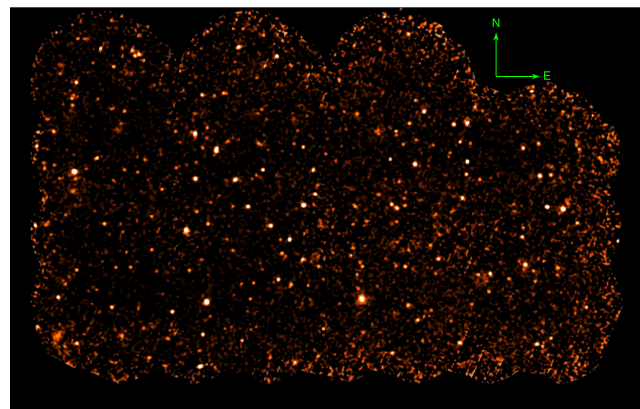
(b)

Figure 2. Mosaic 0.5–2 keV image of the (a) Stripe 82 XMM field 1 (ObsID: 0673000101, $333°47 < RA < 335°20, -0°58 < Dec. < 0°92$) and (b) Stripe 82 XMM field 2 (ObsID: 0673002301, $351°47 < RA < 353°45, -1°10 < Dec. < 0°15$). The images are background subtracted, corrected for the exposure map and smoothed with a Gaussian filter with a kernel of ∼12 arcmin (i.e. 3 pixels). The stretch of the colour map is 10^{-6}–3×10^{-6} counts s^{-1}.

The positions of these sources were then masked out when generating the background maps. Regions of significant extended emission (radius >1 arcmin), piled-up sources and read-out streaks were also masked out manually.

As noted by Cappelluti et al. (2007), the background has two components: unresolved X-ray emission which comprises the cosmic X-ray background (CXB) and local particle and detector background. The former background is subject to vignetting while the latter is not. The residual area (i.e. regions where no sources are detected) was split into two parts based on the median of the effective exposure. Regions above the median, with low vignetting, are dominated by the CXB whereas the detector background becomes more important below the median effective exposure. We set up templates to account for these two components of the background:

$$AM_{1,v} + BM_{1,unv} = C_1 \tag{1}$$

$$AM_{2,v} + BM_{2,unv} = C_2, \tag{2}$$

where $M_{1,v}$ and $M_{2,v}$ are the vignetted exposure maps for the areas above and below the median effective exposure time, respectively;
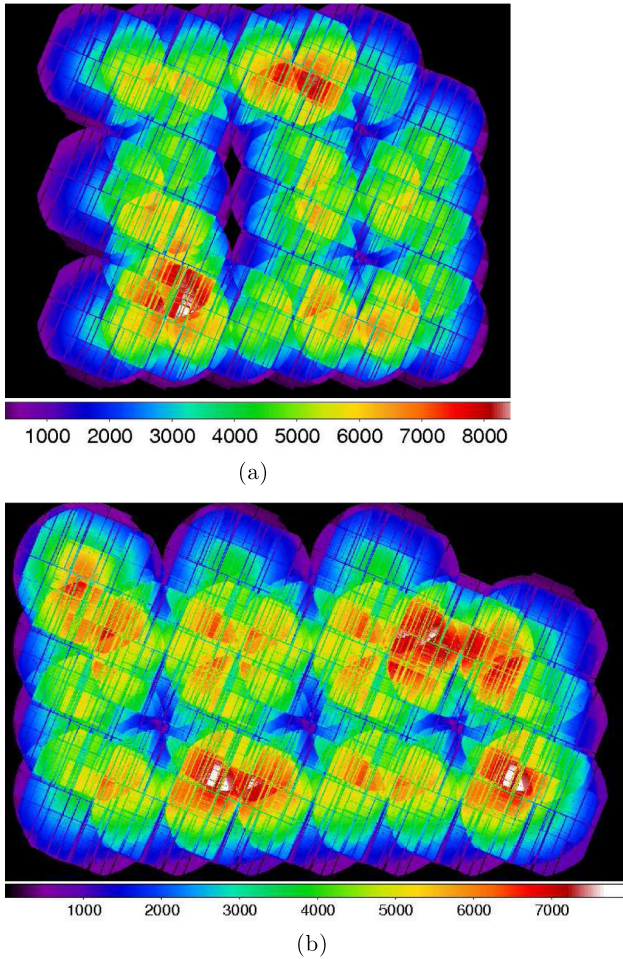
(a)

(b)

Figure 3. 0.5–2 keV exposure maps for the (a) Stripe 82 XMM field 1 and (b) Stripe 82 XMM field 2. Exposure time in ks, indicated by the colours shown.

$M_{1,unv}$ and $M_{2,unv}$ are the unvignetted template exposure maps, and C_1 and C_2 are the background counts. We solve this system of linear equations for the normalizations A and B. The vignetted and unvignetted exposure maps are normalized by A and B, respectively, and then added to obtain the background map for each detector and observation. The background maps among the multiple detectors were added, giving one background map per observation.

2.4 Source detection

We ran the source detection algorithm using the combined images, exposure maps and background maps generated as described above. We created detector masks on the combined images using the SAS task *emask*. For 15 observations, we manually updated these masks to screen out regions of extended emission and piled-up sources and read-out streaks, as noted in Section 2.3 (see Table 1). A preliminary list of sources was generated with the SAS task *eboxdetect*, which is a sliding box detection algorithm run in 'map' mode, where source counts are detected in a 5×5 pixel box with a low-probability threshold (likemin = 4). The source list generated by *eboxdetect* is used as an input for the SAS task *emldetect* which performs a maximum likelihood point PSF fit to the source count distribution, using a likelihood threshold (det_ml) of 6, where det_ml $= -\ln P_{random}$, with P_{random} being the Poisson probability that

a detection is due to random fluctuations. We ran *emldetect* with the option to fit extended sources, where the PSF is convolved with a β model profile. All extended sources (i.e. ext_flag >0 in the *emldetect* outputted source list) are omitted from further analysis in this paper.

For overlapping archival observations, *eboxdetect* and *emldetect* were run in 'raster' mode, i.e. these tasks were run on an input list of images, exposure maps, detector masks and background maps, which as noted above were remapped to a common WCS grid. The source detection algorithm was run separately for the soft, hard and broad-bands for the overlapping observations but simultaneously for the non-overlapping pointings; memory constraints precluded running *eboxdetect* and *emldetect* simultaneously for overlapping observations in multiple energy bands. The ECFs reported in Table 2 are summed among the detectors turned on for each observation and given as input in the source detection algorithm, converting count rates into physical flux units.

The 21-22 pointings for each mosaicked observation could not be fitted simultaneously for source detection due to computational memory constraints. Instead, each group of mosaicked pointings was split into sub-groups so that source detection was run on two adjacent 'rows' in RA to accommodate overlapping pointings. Other than the pointings on the Eastern and Western edges of the mosaic, each RA row was included in two source detection runs to account for overlap and ensure the deepest possible exposures. Similar to the overlapping archival observations, the source detection was run separately for the soft, hard and full bands. From the source lists, we then generated a list of individual sources and searched for the inevitable duplicate identifications of the same source, since portions of every field were in more than one source detection fitting run. Similar to the algorithm used for the *XMM–Newton* Serendipitous Source Catalogue to identify duplicates (Watson et al. 2009), if the distance between any two sources is less than d_{cutoff}, where $d_{cutoff} = \min(0.9 \times d_{nn,\,1}, 0.9 \times d_{nn,\,2}, 15$ arcsec, $3 \times \sqrt{\mathrm{ra\_dec\_err}_1^2 + \mathrm{sys\_err}^2} + \sqrt{\mathrm{ra\_dec\_err}_2^2 + \mathrm{sys\_err}^2})$, where d_{nn} is the distance between a source and its nearest neighbour in that pointing, ra_dec_err is the positional X-ray error returned by *emldetect*, and sys_err is the systematic positional error (taken to be 1 arcsec), we consider the sources to be the same. We then chose the source with the higher det_ml as the detection from which to derive the position, positional error, flux and flux error. We chose a maximum search radius of 15 arcsec based in part from the results of the simulations and matching the input simulated list to the detected source list, with this threshold maximizing identification of counterparts while minimizing spurious associations.

To merge the separate soft-, hard- and full-band source lists into one single source list for the archival overlapping and mosaicked observations, we identified duplicate sources using the method described above. The positions among (or between, for cases where a match was found in two rather than three bands) the bands were averaged and the positional errors were added in quadrature. In our final point source list, we remove extended objects (i.e. where ext > 0 as reported by *emldetect*) and only include the objects where det_ml ≥ 15 (5σ significance) in at least one of the energy bands, to reduce spurious identifications and assure our catalogue contains reliable X-ray detections (see Loaring et al. 2005; Mateos et al. 2008). As summarized in Table 3, we detected 2358 X-ray sources, of which 1607 were found in archival observations and 751 were discovered in our proprietary programme. Of this total number, 182 were detected only in the full band, 261 were identified solely in the soft band and 18 in just the hard band.

Table 3. Number of detected *XMM–Newton* sources[a].

Band	Archival	Proprietary	Total
Soft (0.5−2 keV)	1438	635	2073
Hard (2−10 keV)	432	175	607
Full (0.5−10 keV)	1411	668	2079
Total	1607	751	2358

[a]The numbers for the individual bands refer to the sources detected at del_ml ≥ 15 in that band, while the total band numbers indicate the sources detected at del_ml ≥ 15 in any given band.

2.5 Monte Carlo simulations: source detection reliability and survey coverage

To assess the source detection efficiency and the survey area as a function of limiting flux, we have performed detailed Monte Carlo simulations. First, we generated a list of random fluxes following a published $\log N - \log S$ distribution for each observation, using the fits to the *XMM–COSMOS* soft- and hard-bands number counts (Cappelluti et al. 2009) and the fit to the ChaMP full-band number counts (Kim et al. 2007). These simulated sources are placed in random positions across the detector. Using part of the simulator written for the *XMM–Newton* survey of the CDFS by Ranalli et al. (2013),[4] each input source list is convolved with the *XMM–Newton* PSF, generating simulated event lists for all detectors turned on during each observation. Similar to the procedure for the real data, images are extracted from these simulated events files and added among the detectors. The background map for each observation is added to the combined simulated image and then Poisson noise is added to the combined source image and background map to replicate real observations. The source detection on these simulated images is then executed in the same manner as the real data. We simulated 20 images per pointing, providing us with an adequate number of input and detected sources to gauge source detection reliability and assess survey sensitivity.

To estimate the fraction of spurious and confused sources, we compare the sources detected significantly from the simulations (det_ml ≥ 15) with the input source list. We consider a detected source within 15 arcsec of an input source as a match. Any detected object lacking an input counterpart is deemed spurious. The fraction of spurious sources is 0.49, 0.37 and 0.20 per cent in the soft, hard and full bands, respectively. Following the prescription of Cappelluti et al. (2007), a source is considered confused if $S_{out}/(S_{in} + 3\sigma_{out}) > 1.5$, where S_{out} and S_{in} are the output and input fluxes of the counterparts and σ_{out} is the error on the detected flux. We estimate our fraction of confused sources in the soft, hard and full bands as 0.34, 0.23 and 0.34 per cent, respectively.

From these simulations, we also accurately gauge our survey sensitivity by determining the distribution of fluxes for both input and significantly detected sources. The ratio of these distributions as a function of flux provides us with the area–flux curves shown in Fig. 4, where we show the area–flux curves separately for the *XMM–Newton* proprietary data (∼4.6 deg²), proprietary and archival *XMM–Newton* data (∼10.5 deg²), *XMM–Newton* and *Chandra* coverage (∼16.5 deg²) and C-COSMOS (∼0.9 deg²; Elvis et al. 2009) for comparison; we note that the fluxes in the *Chandra* hard (2–7 keV) and full (0.5–7 keV) bands were converted to 2–10 keV and 0.5–10 keV ranges using the assumed spectral
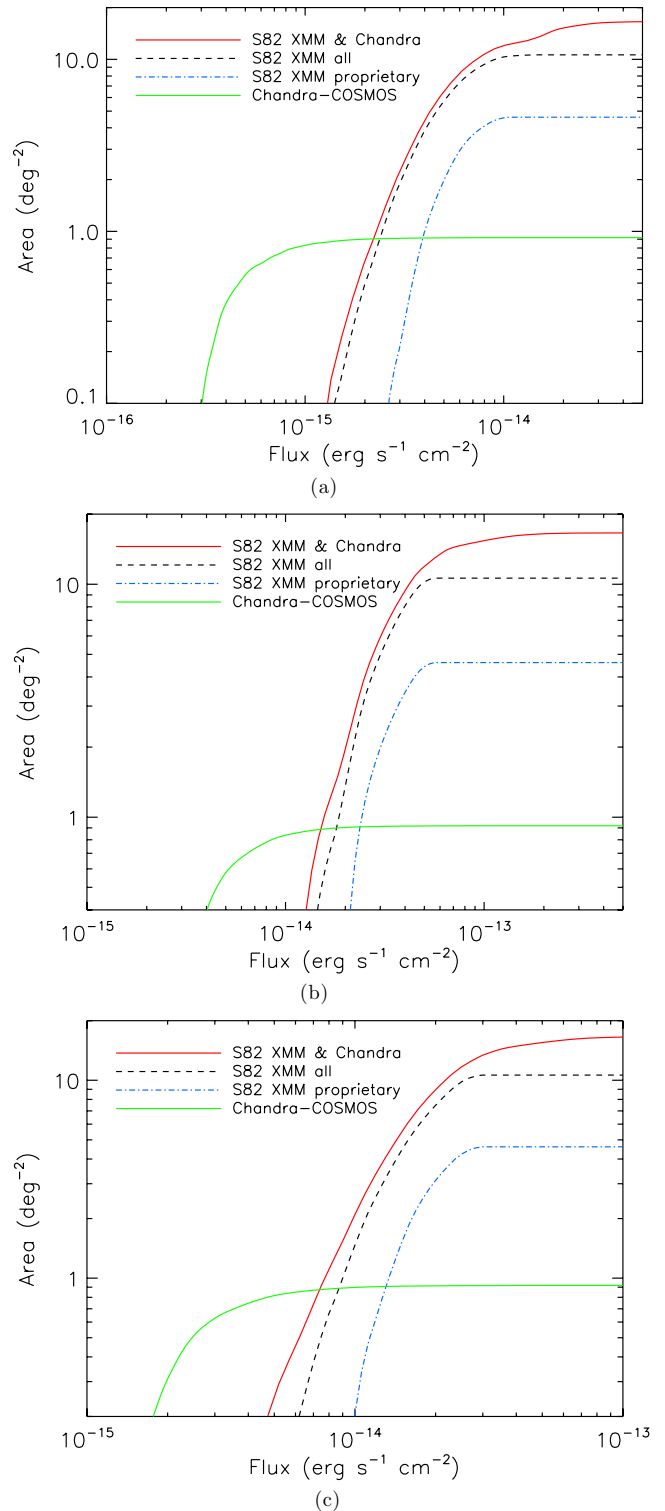
[4] https://github.com/piero-ranalli/cdfs-sim

Figure 4. Area–flux curves for the Stripe 82 X-ray coverage and C-COSMOS (Elvis et al. 2009) for comparison in the (a) soft, (b) hard and (c) full bands. From the *XMM–Newton* proprietary + archival area–flux curves, we produced the $\log N - \log S$ relationships in Fig. 5.

models of $\Gamma = 1.7$ for Stripe 82 and $\Gamma = 1.4$ for C-COSMOS. We reach down to approximate flux limits (at ∼0.1 deg² of coverage) of 1.4×10^{-15}, 1.2×10^{-14} and 5.6×10^{-15} erg s⁻¹ cm⁻² with half-survey area at 4.7×10^{-15}, 3.1×10^{-14} and 1.6×10^{-14} erg s⁻¹ cm⁻²

in the soft, hard and full bands, respectively. From these curves, we then generate the number counts below.

3 log N−log S

We present the number density of point sources as a function of flux, i.e. the log N−log S relation. In integral form, the cumulative source distribution is represented by

$$N(>S) = \sum_{i=1}^{N_s} \frac{1}{\Omega_i}, \qquad (3)$$

where $N(>S)$ is the number of sources with a flux greater than S and Ω_i is the limiting sky coverage associated with the *i*th source. The associated error is the variance:

$$\sigma^2 = \sum_{i=1}^{N_s} \left(\frac{1}{\Omega_i}\right)^2. \qquad (4)$$

To avoid biasing our log$N$−log$S$ relations by the inclusion of targeted sources, we removed the closest object located within 30 arcsec of the target RA and Dec., taken from RA_OBJ and Dec_OBJ in the FITS header. Of the 33 archival pointings, 18 had objects within 30 arcsec of the nominal target positions. Three of these were not detected at a significant level (i.e. det_ml $\geq$ 15) in any given band and were not in our final source list. Thus, only 15 sources were excluded when generating the log$N$−log$S$. Of the remaining 15 archival pointings, 12 had central regions masked out due to extended emission or pile-up (presumably from the targeted source) while the other 3 had no sources detected within 30 arcsec of the targeted position.

The number counts in the soft, hard and full bands are shown in Fig. 5. We have also overplotted the upper and lower bounds of the *Chandra* log$N$−log$S$ from Stripe 82 (S82 ACX) for comparison, where we have re-calculated the source fluxes and survey sensitivity from LaMassa et al. (2013) using the same spectral model as applied to the *XMM–Newton* data. We note that 12 *Chandra* non-cluster pointings used for generation of the log$N$−log$S$ presented in LaMassa et al. (2013) at least partially overlap the *XMM–Newton* observations, $\sim$1.2 deg^2. Since the hard and full bands are defined in S82 ACX up to 7 keV, the *Chandra* fluxes have been adjusted assuming a power-law model of $\Gamma = 1.7$ to convert to the energy ranges used in our *XMM–Newton* analysis (i.e. the *Chandra* fluxes have been multiplied by factors of 1.36 and 1.2 for the hard and full bands, respectively). The *XMM–Newton* and S82 ACX number counts are largely consistent, with slight discrepancies apparent at moderate fluxes in the hard band ($\sim$5 $\times$ 10^{-14} < $S_{2-10\,keV}$ < 2 $\times$ 10^{-13} erg cm^{-2} s^{-1}) and at the low *XMM–Newton* flux limit in the full band (<10^{-14} erg cm^{-2} s^{-1}). However, as noted in LaMassa et al. (2013), short exposure times in *Chandra* observations, which constitute the majority of Stripe 82 ACX, have an effect on the log$N$−log$S$ normalization in the hard band, making the offset between *XMM–Newton* and *Chandra* in this energy range unsurprising.

In Fig. 6, we compare the Stripe 82 ACX log$N$−log$S$ using the spectral model from LaMassa et al. (2013) and the one used here. In LaMassa et al. (2013), we adopted a spectral model used in *Chandra* surveys to which we compared our results while here we used a spectral model consistent with previous *XMM–Newton* surveys, such as *XMM*-COSMOS (Cappelluti et al. 2007). The difference in the hard-band number counts is slight with this change of assumed spectral model, but shifts the normalization to lower values in the soft and especially the full band where the median offset between the 1σ error bars in the discrepant ranges is $\sim$10 per cent.
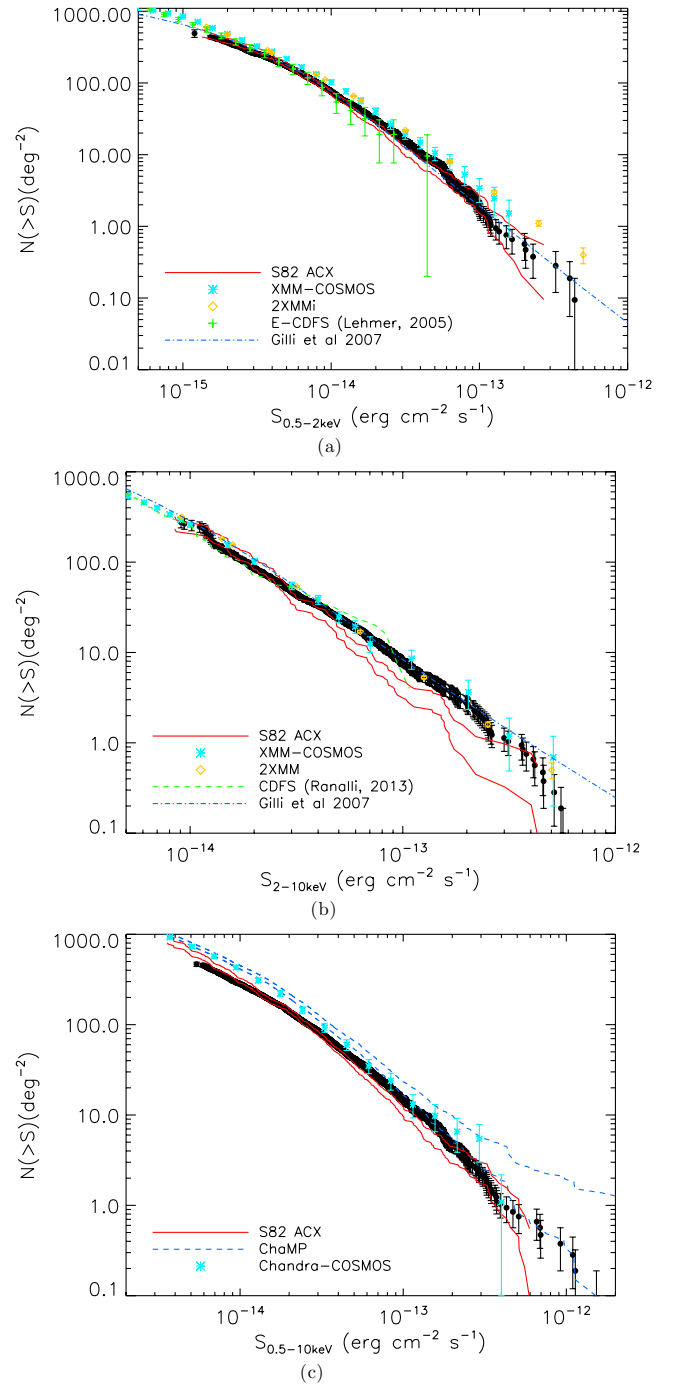

Figure 5. Stripe 82 *XMM–Newton* number counts (filled black circles) in the (a) soft, (b) hard and (c) full bands, with the 1σ confidence interval from Stripe 82 Archival *Chandra* (S82 ACX) overplotted in red. Comparison X-ray surveys are overplotted, ranging from small to moderate to wide area. There is general agreement; see text for discussion.

We also compare our log$N$−log$S$ relationships with those from previous X-ray surveys, spanning from wide (2XMMi, 132 deg^2; Mateos et al. 2008) to moderate (*XMM*-COSMOS, 2 deg^2; Cappelluti et al. 2007, 2009) to small areas (E-CDFS – 0.3 deg^2, *XMM*-CDFS – $\sim$0.25 deg^2; Lehmer et al. 2005; Ranalli et al. 2013). Where possible, we aim to compare our data with other *XMM–Newton* surveys. However, the *XMM–Newton* survey in the CDFS (Ranalli et al. 2013) only produced the log$N$−log$S$ in the hard band, so we use
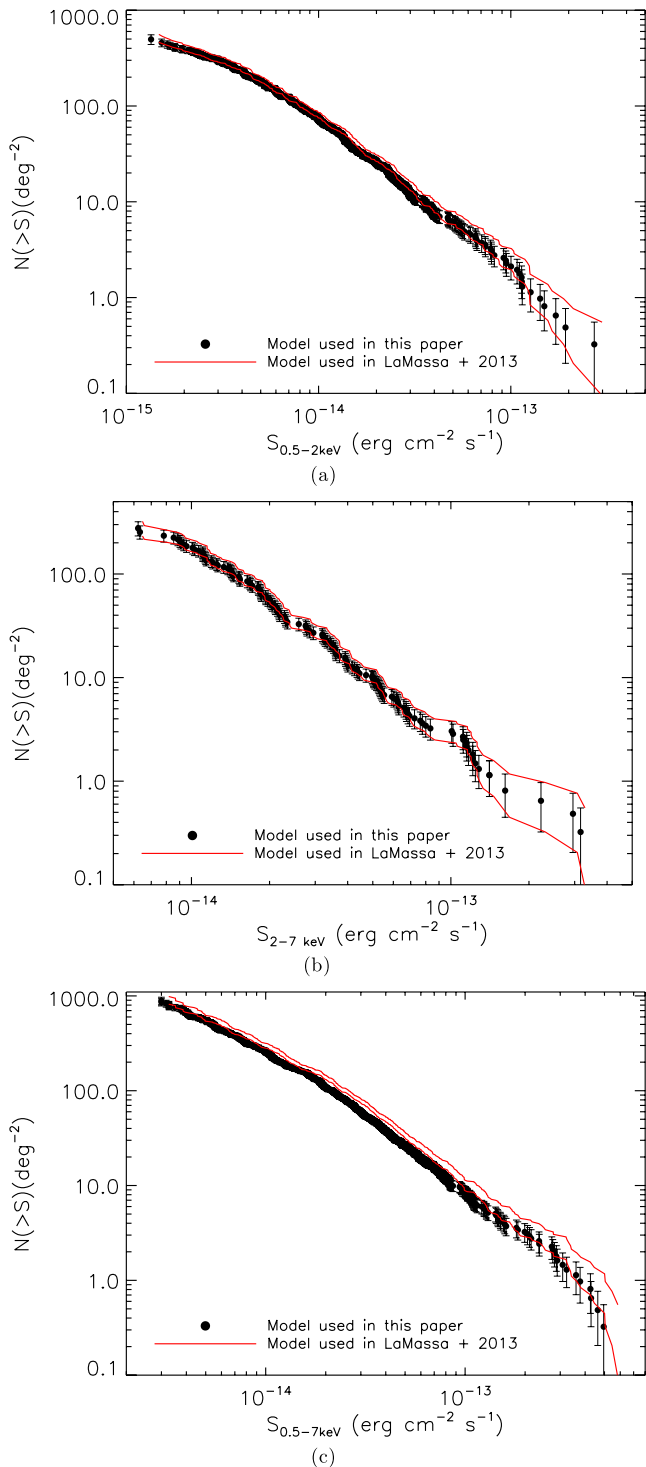
Figure 6. Comparison of the *Chandra* number counts in the (a) soft, (b) hard and (c) full bands using the assumed spectral model from LaMassa et al. (2013), $\Gamma = 1.4$ in all bands (1σ confidence interval shown by red lines) and the spectral model adopted for the *XMM–Newton* data, $\Gamma = 2$ in the soft band and $\Gamma = 1.7$ in the hard and full bands (black filled circles). The spectral model assumed for this paper shifts the number counts normalization to lower values with respect to the LaMassa et al. (2013) results, with the most significant offset (i.e. where the error ranges do not overlap) in the full band, with a median discrepancy of ~ 10 per cent.

the *Chandra* E-CDFS survey (Lehmer et al. 2005) for comparison in the soft band. No previous *XMM–Newton* survey has produced a full-band $\log N{-}\log S$, so we compare our Stripe 82 number counts with the small area C-COSMOS (0.9 deg^2; Elvis et al. 2009) and wide area ChaMP surveys (9.6 deg^2; Kim et al. 2007). As ChaMP defines the full band to be 0.5–8 keV, their fluxes were adjusted to match our 0.5–10 keV range using their adopted spectral model (i.e. multiplied by a factor of 1.18). We note that the spectral shapes over a broad-band are not well constrained, making the ECFs in this range approximate and comparisons with number counts using other model assumptions difficult to quantify; these comparisons are for illustrative purposes. The model predictions from Gilli, Comastri & Hasinger (2007) have also been overplotted in the soft and hard bands.

The Stripe 82 *XMM–Newton* number counts are consistent with previous *XMM–Newton* surveys in the hard band. The 2XMM $\log N{-}\log S$ from Mateos et al. (2008) is systematically higher than our data in the soft band. However, they note that their 0.5–2 keV number counts are higher than several other X-ray surveys, which they attribute to the inclusion of moderately extended sources in their catalogue. Similar to other surveys, we include only point sources, making this soft-band discrepancy with Mateos et al. (2008) not surprising. Stripe 82 *XMM–Newton* is fully consistent with E-CDFS (Lehmer et al. 2005) and the model predictions from Gilli et al. (2007) in this energy range. Though the normalization for the full-band Stripe 82 *XMM–Newton* $\log N{-}\log S$ seems low compared to ChaMP and C-COSMOS, this is likely due to differences in spectral models to convert from count rate to fluxes: ChaMP and C-COSMOS adopt a power-law model with $\Gamma = 1.4$, whereas we use $\Gamma = 1.7$. As shown in Fig. 6(c), the difference between these two spectral models shifts the full-band number counts in the right sense to account for the observed disagreement between Stripe 82 *XMM–Newton* and C-COSMOS and ChaMP. We also note that the ChaMP number counts seem to be somewhat higher than other *Chandra* surveys (LaMassa et al. 2013) while C-COSMOS shows better agreement with our calculations.

As we show below, these X-ray objects do preferentially sample the high-luminosity AGN population and include candidates for interesting rare objects: reddened quasars and high-luminosity AGN at high redshift. In a future paper, we will quantify the evolution of these sources by generating the quasar luminosity function, beginning with the $\log N{-}\log S$ relations presented here.

4 MULTIWAVELENGTH SOURCE MATCHING VIA MAXIMUM LIKELIHOOD ESTIMATOR

The Stripe 82 X-ray source lists represent the *XMM–Newton* objects found above and the *Chandra* sources detected at $\geq 4.5\sigma$ level from all pointings overlapping the Stripe 82 area. In LaMassa et al. (2013), we presented only those observations that did not target galaxy clusters, covering an area of ~ 6.2 deg^2, garnering 709 objects. Inclusion of the previously omitted *Chandra* pointings adds an additional 1.2 deg^2 to produce a total of 1146 X-ray sources. About 1.5 deg^2 of the full 7.4 deg^2 of *Chandra* coverage in Stripe 82 overlaps the *XMM–Newton* pointings. Using the method described above to find duplicate observations of the same X-ray object, we cross-matched the *XMM–Newton* and *Chandra* source lists, finding 3362 unique objects over ~ 16.5 deg^2 of non-overlapping area.

To assign multiwavelength counterparts to the Stripe 82 X-ray sources, we employed a maximum likelihood estimator (MLE) algorithm which takes into account the distance between potential matches and the brightness of the ancillary counterpart (Sutherland

& Saunders 1992). The ancillary source at the closest distance to the X-ray object, as found using the nearest-neighbour method, may not be the true match, but may instead be a spurious association due to random chance. As there are many more faint than bright objects, an association between a bright source and an X-ray target is more likely to represent a true counterpart than a match to a faint source. The MLE technique codifies this statistically, assigning reliability values to each potential match and has been successfully implemented in multiwavelength catalogue matching in previous X-ray surveys (e.g. Brusa et al. 2005, 2007, 2010; Cardamone et al. 2008; Luo et al. 2010; Civano et al. 2012).

All the objects within a search radius (r_{search}) around each X-ray target are assigned a likelihood ratio (LR), which is the probability that the correct counterpart is found within r_{search} divided by the probability of finding an unassociated object by chance:

$$LR = \frac{q(m)f(r)}{n(m)}, \tag{5}$$

where $q(m)$ is the expected normalized magnitude distribution of ancillary counterparts, $f(r)$ is the probability distribution of the positional errors (which is assumed to be a two-dimensional Gaussian, where σ is derived by adding the X-ray and ancillary positional errors in quadrature) and $n(m)$ is the magnitude distribution of background sources. For the positional *Chandra* uncertainty, we added the major and minor axes of the 95 per cent confidence level error ellipse, err_ellipse_r0 and err_ellipse_r1, in quadrature, while *XMM–Newton* positional errors are from the *emldetect* source detection script added in quadrature to a 1 arcsec systematic error.[5] As noted below, for ancillary catalogues where a positional error is not quoted, we adopted a uniform, survey-dependent, positional uncertainty. Since *Chandra* has higher resolution and a smaller on-axis PSF than *XMM–Newton*, we chose different radii to search for ancillary counterparts for each catalogue. For *Chandra* objects, $r_{search} =$ 5 arcsec (Civano et al. 2012) while for *XMM–Newton* sources, r_{search} = 7 arcsec (Brusa et al. 2010); the positional errors for 88 per cent of *Chandra* and 99.9 per cent of *XMM–Newton* sources are below the adopted search radii.

To determine the background distribution $n(m)$, we isolate the ancillary sources within an annulus around each X-ray source, with inner and outer radii of 7 and 30 arcsec for *Chandra* and 10 and 45 arcsec for *XMM–Newton*. The inner radius is chosen to avoid the inclusion of real counterparts and the outer radius is picked to ensure a large number of sources to estimate the background while minimizing overlap with other X-ray sources. Within these annular regions, there were 53 *Chandra* pairs and 7 *Chandra* triples (~11 per cent of the sample) and 49 *XMM–Newton* pairs and 3 *XMM–Newton* triples (~4 per cent of the sample), i.e. only a small fraction of the background histogram has duplicate objects.

We then calculate $q'(m)$ by first finding the magnitude distribution of ancillary objects within r_{search} of each X-ray source and dividing by the area to obtain the source density magnitude distribution. Similarly, we divide $n(m)$ by the search area, and take the difference between the former and latter which gives us the expected source density magnitude distribution. Finally, multiplying this distribution

Table 4. Number of X-ray sources detected in ancillary databases.

Catalogue	*Chandra*	*XMM–Newton*	Total[a]
X-ray	1146	2358	3362
SDSS	676	1283	1892
WISE	595	1324	1855
UKIDSS	543	1266	1754
GALEX	164	301	447
FIRST	42	82	119
Spec-zs	306	497	759

[a]Duplicate sources between the *Chandra* and *XMM–Newton* catalogues removed.

by the search area gives us $q'(m)$. We then normalize $q'(m)$ to Q, the ratio of the number of X-ray sources with counterparts found within r_{search} to the total number of X-ray sources, producing $q(m)$ (see Civano et al. 2012).

From LR, we then calculate a reliability value for each source:

$$R = \frac{LR}{\Sigma_i(LR)_i + (1 - Q)}, \tag{6}$$

where the sum over LR is for each possible counterpart found within r_{search} around an individual X-ray source. We use R to discriminate between true counterparts and spurious associations. Since R depends on the source density and magnitude distribution of the ancillary sources, the critical R value (R_{crit}) we adopt to accept a match as 'real' differs among catalogues and strikes a fine balance between missing true counterparts and adding contamination from chance proximity to an unrelated source. To calibrate R_{crit}, we shifted the positions of the X-ray sources by random amounts, with offsets ranging from ~21 to ~35 arcsec, and re-ran the matching code. Any matches found should be due to random chance. We then plotted the distribution of reliability values for these spurious associations to estimate the contamination above R_{crit}; full details regarding the estimate of false matches are given in Appendix A. We impose a lower limit on R_{crit} of 0.5, even in the cases where the reliability values for the shifted X-ray positions are consistent with zero. If there were multiple counterparts per X-ray source, or multiple X-ray sources per counterpart, the match with the highest reliability was favoured.

In the online catalogues (available at CDS and searchable with VizieR; Ochsenbein, Bauer & Marcout 2000), we list the X-ray sources, fluxes and matches to the ancillary multiwavelength catalogues, including the non-aperture-matched photometry. Duplicate observations of the same X-ray object between the *Chandra* and *XMM–Newton* source lists are marked in the online tables. Objects not included in the $\log N - \log S$ relations, i.e. targets of observations and for *Chandra* objects, all sources identified in observations targeting galaxy clusters, are also noted. If the X-ray flux is not detected at a significant level in any individual band ($<4.5\sigma$ for *Chandra* and det_ml < 15 for *XMM–Newton*), the flux is listed as null in the online catalogues. A high-level summary of the number of sources matched to each optical, NIR and ultraviolet catalogue is reported in Table 4, with the magnitude/flux density distributions for these counterparts shown in Fig. 7. Appendix B details the columns for the online versions of the catalogues.

4.1 Sloan Digital Sky Survey

Due to the high density of sources in SDSS, as well as sub-arcsecond astrometry precision, we matched the X-ray sources separately to
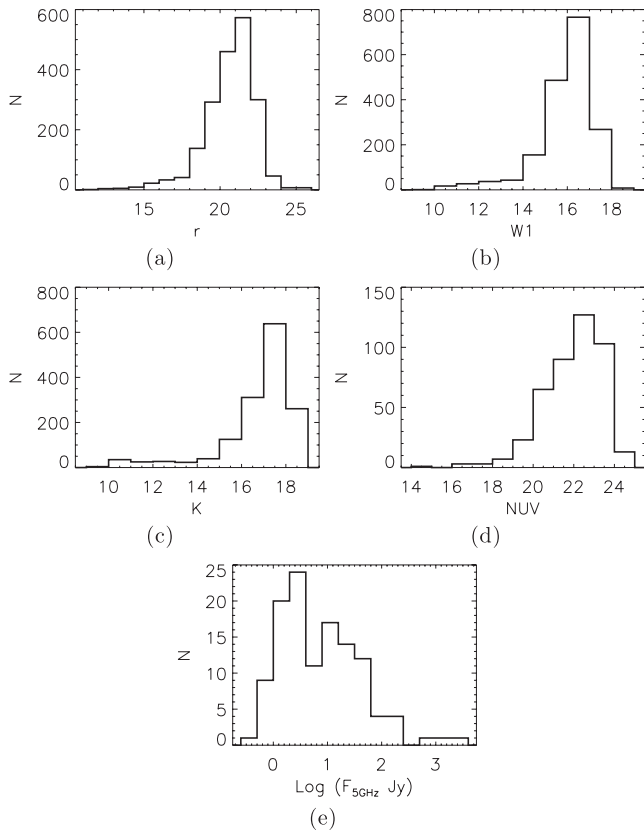
Figure 7. Magnitude/flux density distributions of X-ray counterparts from (a) SDSS, (b) *WISE* (c) UKIDSS (d) *GALEX* and (e) FIRST.

the *u*, *g*, *r*, *i* and *z* bands, using single-epoch photometry from Data Release 9 (DR9; Ahn et al. 2012). A uniform 0.1 arcsec error was assumed for all SDSS positions (Rots & Budavári 2011).

After vetting each individual band source list to include only objects exceeding R_{crit}, we combined these source lists into a matched SDSS/*Chandra* catalogue and an SDSS/*XMM–Newton* catalogue. We visually inspected the cases where multiple SDSS objects (from separate band matchings) were paired to one X-ray source and chose the most likely counterpart by selecting the one with the greatest number of matches and/or the brightest object. We also imposed quality control cuts to assure that the broad-band SEDs and derived photometric redshifts we will generate in a future paper (after careful aperture matching) are robust. We therefore require the SDSS objects to not be saturated[6] or blended[7] and to have the photometry well measured.[8] After this vetting, every remaining SDSS match was visually inspected to remove objects contaminated by optical artefacts from e.g. diffraction spikes, or proximity to a close object that was not caught in the pipeline flagging.

We identified 748 and 1444 SDSS counterparts to *Chandra* and *XMM–Newton* sources, corresponding to 65 and 61 per cent of the sample, respectively, that exceeded R_{crit}. However, 72 and 161 of these were rejected due to failing the quality control checks and visual inspection described above (marked as 'yes' in the 'SDSS_rej'

flag in the online catalogues, leaving 676 and 1283 reliable matches to *Chandra* and *XMM–Newton* sources, or 59 and 54 per cent of the X-ray sources). In a follow-up paper in which we will generate the broad-band SEDs, we will use co-added data of the 50–60 epochs of Stripe 82 scans to search for counterparts for the remaining ∼35 per cent of the X-ray sources (see Jiang et al. 2009; McGreer et al. 2013 for studies of *z* > 5 QSOs using co-added SDSS Stripe 82 data).

4.1.1 Spectroscopy

We searched spectroscopic data bases to find redshifts corresponding to our matched X-ray/SDSS catalogues, using SDSS DR9, 2SLAQ (Croom et al. 2009), WiggleZ (Drinkwater et al. 2010) and DEEP2 (Newman et al. 2012). This yielded spectroscopic redshifts for 306 *Chandra* sources (∼27 per cent of the sample): 286 from SDSS DR9; 10 from 2SLAQ; 3 from WiggleZ and 7 from DEEP2. For the *XMM–Newton* sources, 497 optical counterparts had spectroscopic redshifts (∼21 per cent of the sample): 468 from SDSS DR9, 20 from 2SLAQ, 4 from WiggleZ and 5 from DEEP2. We manually checked the spectra for the 25 SDSS sources where warning flags were set or for any object with *z* > 5: three spectra were re-fitted to give more reliable redshifts; 11 were discarded due to poor spectra that could not be reliably fitted, and we confirmed the redshifts for the remaining 11 objects. In Table 4, the number of reported redshifts do not include the 11 that we discarded. 28 *XMM–Newton* sources had spectroscopic redshifts but unreliable photometry; we retain the redshift, but not the photometric, information for these objects. In the online catalogues, we indicate the data base from which the spectroscopic redshifts were found, with *z*-source of 0, 1, 2, 3 and 4 referring to SDSS, 2SLAQ, WiggleZ, DEEP2 and SDSS spectra refitted/verified by us, respectively.

4.2 *WISE*

For a source to be included in the *WISE* All Sky Source Catalog (Wright et al. 2010; Cutri et al. 2012), an SNR > 5 detection was required for one of the four photometric bands, *W*1, *W*2, *W*3 or *W*4, corresponding to wavelengths 3.4, 4.6, 12, and 22 μm, with resolution 6.1, 6.4, 6.5 and 12.0 arcsec. The X-ray sources were matched to the *W*1 band since this band has the greatest number of non-null values, including both detections and upper limits. In the full Stripe 82 area, no *WISE* sources had null W1 detections, so we do not miss any potential *WISE* counterparts by matching to only the *W*1 band. The matching was performed on all *W*1 values, regardless of whether the magnitude corresponded to a detection or an upper limit (i.e. where the *W*1 SNR is below 2). The RA and Dec. errors were added in quadrature to provide an estimate of the *WISE* astrometric error.

If any bands suffered from saturation,[9] spurious detections associated with artefacts (i.e. diffraction spikes, persistence from a short-term latent image, scattered halo light from a nearby bright source or optical ghost image from nearby bright source), contamination from artefacts or moon level contamination,[10] we

---

[6] (NOT SATUR) OR (SATUR AND (NOT SATUR_CENTER))

[7] (NOT BLENDED) OR (NOT NODEBLEND)

[8] (NOT BRIGHT) AND (NOT DEBLEND_TOO_MANY_PEAKS) AND (NOT PEAKCENTER) AND (NOT NOTCHECKED) AND (NOT NO-PROFILE)

[9] We consider the band to be affected by saturation if the fraction of saturated pixels exceeded 0.05, i.e. we could not rule out saturation at the 2σ level.

[10] We consider moon_lev ≥5 as contaminated, where moon_lev is the number of frames affected by scattered moonlight normalized by the total number of frames in the exposure multiplied by 10, and spans from 0 ≤ moon_lev ≤ 9.

consider the magnitude in that band unreliable. If every band did not pass these quality control tests, then the source is not included in our final tally since we will not use the *WISE* data for generating the SEDs. For extended sources (where ext_flag > 0), the magnitudes measured from the profile-fitting photometry (i.e. *wnmpro*, where *n* goes from 1–4) are unreliable. For these objects, we therefore focus on the magnitudes and quality flags associated with the elliptical apertures, *wngmag*, where *n* goes from 1–4. Again, if all bands have null elliptical magnitudes and/or non-zero quality control flags, the source is not included in our catalogue. The extended sources have the WISE_ext flag set to 'yes' in the online catalogues.

When matching the *Chandra* catalogue to *WISE*, we imposed an R_{crit} of 0.75 and found 595 counterparts that passed the photometry quality control checks, or 52 per cent of the *Chandra* sample. Eight of these were extended. Photometry of 30 sources was compromised, 20 of which were extended. Our R_{crit} threshold for the *XMM–Newton* source list was 0.9, with 1324 counterparts identified with acceptable photometry (56 per cent of the sample), of which 8 were extended. 65 sources did not pass the quality control checks, of which 40 were extended. The X-ray sources with *WISE* counterparts removed for not passing the quality control checks are marked as 'yes' in the WISE_rej field in the online catalogues.

4.3 UKIDSS

We searched for the UKIDSS Large Area Survey (LAS) Data Release 8 (Hewett et al. 2006; Casali et al. 2007; Lawrence et al. 2007; Warren et al. 2007) for NIR counterparts to the Stripe 82 X-ray sources; details regarding maintenance of the UKIDSS science archive are described by Hambly et al. (2008). We used the LAS *YJHK* Source table, which contains only fields that have coverage in every filter and merges the data from multiple detections of the same object. Only primary objects were selected[11] so that we worked with a clean input list with no duplicate NIR sources. We a priori removed objects flagged as noise, i.e. those sources with mergedClass set to zero and PNoise ≤ 0.05; that is, we only retained objects that are consistent with real detections (not noise) at greater than the 2σ level for our candidate source list and background histograms. The UKIDSS positional uncertainties are set to NULL in the catalogue. Dye et al. (2006) quote that the internal accuracy can be ~100 mas in each coordinate and the external accuracy is ~80 mas in each coordinate. Adding the 180 mas uncertainty in quadrature for each coordinate gives a positional error of ~0.25 arcsec which we apply uniformly to all UKIDSS sources.

The X-ray source catalogues were matched separately to each UKIDSS band: *Y* (0.97–1.07 μm), *J* (1.17–1.33 μm), *H* (1.49–1.78 μm) and *K* (2.03–2.37 μm). The output matches were culled to include only sources exceeding R_{crit} and these individual band lists were then combined. Based on our test of shifting the X-ray positions by random amounts, we chose the values R_{crit} = 0.85, 0.75, 0.8 and 0.75 for the *Y, J, H* and *K* bands, respectively, for the *Chandra* matches; we used R_{crit} = 0.5, 0.6, 0.7, and 0.5 for the *Y, J, H* and *K* bands, respectively, in the *XMM–Newton* source matching.

When merging the individual UKIDSS matches with the X-ray source lists, more than one UKIDSS counterpart was matched to 1 *Chandra* object and to 45 *XMM–Newton* objects. We inspected these

cases by eye and generally chose the brightest potential candidate in the most number of bands as the preferred match. In cases where the brightnesses were similar, we favoured the candidate with the greatest number of matches among the UKIDSS bands. We note that a handful of these multiple potential candidates were duplicate observations of a bright star or a bright star and associated diffraction spike. We found 543 UKIDSS counterparts to the 1146 *Chandra* sources (47 per cent) and 1266 UKIDSS counterparts to the 2358 *XMM–Newton* objects (54 per cent). None of these IR sources was affected by saturation, i.e. there were no instances where 'MergedClass' was set to −9 and 'PSaturated' (probability of saturation) was 0 for all objects.

4.4 *GALEX*

The *GALEX* catalogue comprises sources detected over several surveys, including Deep, Medium and All Sky Imaging Surveys (DIS, MIS and AIS, respectively) as well as a Guest Investigator program. For a trade-off between depth and coverage, and to cleanly remove duplicate observations of the same source, we extracted objects from the MIS survey only. Since the survey has overlapping tiles (see Morrissey et al. 2007) multiple observations of the same source can appear in the catalogue. To choose the best candidate list, we queried the MIS data base from Galex Release 7 for primary sources, i.e. those that are inside the pre-defined position ('SkyGrid') cell within the field (see Budavári et al. 2009). We further require that each primary is within 5 arcmin of the field centre. Following the prescription of Bianchi et al. (2011), we considered objects within 2.5 arcsec as possible duplicates: if they are part of the same observation, i.e. had the same 'photoextractid,' they are considered unique sources but if they are from different observations, the data corresponding to the longest exposure were used. We note that in many cases, sources with the same 'photextractid' but different 'objids' (which identifies unique sources) were actually unmerged far-ultraviolet (FUV) and near-ultraviolet (NUV) detections of the same source, where one observation had either an FUV non-detection while there was an NUV detection or vice versa. However, since we matched the X-ray source lists separately to the NUV and FUV catalogues, such duplicates do not affect the results of our analysis.

The *Chandra* and *XMM–Newton* source lists were matched to this cleaned *GALEX* catalogue using $R_{crit} = 0.5$ for each band. We used the individual source positional errors reported in the *GALEX* data base, rather than applying a systematic positional error to all sources. Matching the NUV and FUV detections separately, rather than focusing on the *GALEX* sources with detections in both bands, has the advantage that we locate ultraviolet counterparts that are detected in one band and not the other. We then merged the results of the individual band matching, locating *GALEX* counterparts for 164 *Chandra* and 301 *XMM–Newton* objects, corresponding to 14 and 13 per cent of each parent sample, respectively.

4.5 FIRST

Due to the low space density of both radio and X-ray sources, we matched our X-ray source lists to the FIRST (Becker et al. 1995; White et al. 1997) catalogue using a simple nearest-neighbour approach rather than MLE: the closest radio object within a search radius of 5 arcsec for *Chandra* sources and within 7 arcsec for *XMM–Newton* objects was chosen as the true counterpart. We used the FIRST catalogue released in 2012, which contains all sources detected between 1993 and 2011, with a detection limit of 0.75 mJy

---

[11] The 'priOrSec' flag is set to zero if there are no duplicate observations of the same source or to the best 'frameSetId' for duplicated observations. The SQL syntax to isolate primary observations is then '(priOrSec = 0 OR priORSec = frameSetId)'.

over part of Stripe 82 ($319°6 < $ RA $ < 49°5$, $-1° < $ Dec. $< 1°$), and 1 mJy detection limit for the rest of the region (Becker et al. 2012). We identified radio counterparts for 42 *Chandra* sources (4 per cent of the sample) and 82 *XMM–Newton* objects (3 per cent of the sample). From shifting the X-ray positions by random amounts, we expect spurious associations for one *Chandra* source within 5 arcsec and four *XMM–Newton* objects within 7 arcsec. Two *Chandra* sources had two potential radio counterparts within r_{search}, but these X-ray sources were within the search radius of each other, so these duplicate potential matches are expected. Within the 7 arcsec *XMM–Newton* search radius, two potential counterparts were found for four X-ray sources. In all of these cases, the nearest neighbour was also the brightest radio object.

5 DISCUSSION

Here, we use the results of the catalogue matching to discuss general characteristics of the X-ray sources in Stripe 82, highlighting the science areas our survey is uniquely poised to investigate.

5.1 Probing the high X-ray luminosity regime of black hole growth

We calculated full-band X-ray luminosities for the sources with spectroscopic redshifts. After removing duplicate matches between the *Chandra* and *XMM–Newton* source catalogues and isolating the objects with luminosities exceeding 10^{42} erg s^{-1}, the X-ray luminosity above which there are few or no starburst-dominated X-ray sources (e.g. Persic et al. 2004; Brandt & Hasinger 2005), we confirm that 645 of the 759 Stripe 82 X-ray sources with optical spectra are AGN; the remaining sources have X-ray luminosities consistent with star-forming galaxies or low-luminosity AGN or are stars. Below, we compare the X-ray luminosity distribution with other X-ray surveys and with model predictions and comment on the interesting sources we have discovered.

5.1.1 Comparison with other X-ray surveys

The comparison X-ray surveys plotted in Fig. 8 span from deep, small area (the GOODS and MUSYC survey of E-CDFS and CDF-S, $\sim$0.3 deg^2; Giavalisco et al. 2004; Treister et al. 2004; Cardamone et al. 2010), to moderate area and moderate depth (*XMM*- and *Chandra*-COSMOS, $\sim$2.1 deg^2; Cappelluti et al. 2009; Brusa et al. 2010; Civano et al. 2012), to wide area and shallow depth (XBoötes, $\sim$9 deg^2; Kenter et al. 2005; Kochanek et al. 2012). Again, we focus on only the X-ray sources with spectroscopic redshifts, with a completeness of $\sim$28 per cent, $\sim$45 per cent and $\sim$44 per cent for E-CDFS + CDF-S, COSMOS and XBoötes, respectively. For reference, the spectroscopic completeness of Stripe 82X, prior to any dedicated follow-up observations, is currently $\sim$23 per cent, though the smaller X-ray surveys peer deeper, garnering many more faint optical counterparts where spectroscopic follow-up opportunities are limited. We report observed, full-band luminosities, which is 0.5–10 keV for the *XMM–Newton* surveys (obtained for *XMM*-COSMOS by summing the individual soft- and hard-band fluxes while Civano et al. (2012) provides full-band fluxes for C-COSMOS objects), 0.5–7 keV for the Stripe 82 *Chandra* sources and XBoötes, and 0.5–8 keV for E-CDFS + CDF-S.

As Fig. 8 illustrates, survey area determines the AGN population sampled. Small area surveys (e.g. E-CDFS + CDF-S) identify faint objects but leave the high-luminosity objects sparsely sampled. Moving to wider areas expands the parameter space to higher
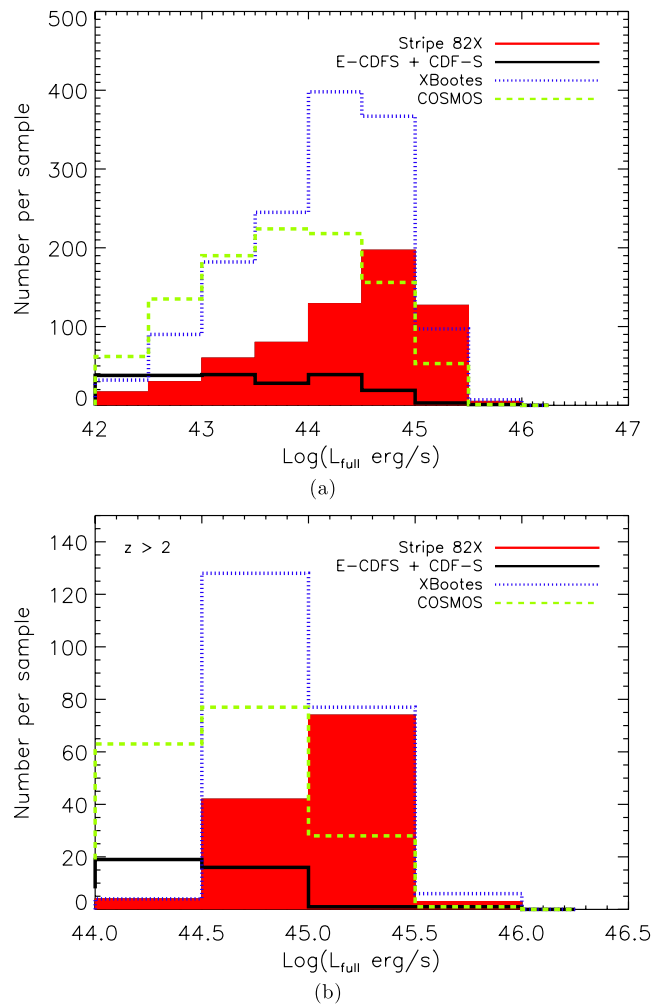

Figure 8. X-ray luminosity distribution for Stripe 82X sources (red filled) compared to other X-ray selected samples for (a) all sources with spectroscopic redshifts and (b) objects with spectroscopic redshifts greater than 2. Wider area surveys are necessary to sample appreciable numbers of high-redshift and high-luminosity AGN; small area surveys (e.g. E-CDFS + CDF-S, black; Giavalisco et al. 2004; Treister et al. 2004; Cardamone et al. 2010) lack sufficient volume to detect rare objects. Though Stripe 82X has a lower level of spectroscopic completeness than the other surveys, and *XMM*- and C-COSMOS (green dashed; Cappelluti et al. 2009; Brusa et al. 2010; Civano et al. 2012) and XBoötes (blue dot–dashed; Kenter et al. 2005; Kochanek et al. 2012) have benefited from dedicated multiyear follow-up campaigns, we immediately identify more $L_x > 10^{45}$ erg s^{-1} AGN at all redshifts and comparable numbers at $z > 2$.

luminosities since these objects are rare and more volume must be probed in order to locate them. This becomes very apparent at $z > 2$ (Fig. 8b).

Wider area surveys, such as COSMOS and XBoötes have higher levels of spectroscopic completeness than Stripe 82X due to dedicated multiyear spectroscopic campaigns. However, prior to any follow-up, we have identified $\sim$2.5 times more high-luminosity AGN ($L_{0.5-10\,keV} > 3 \times 10^{44}$ erg s^{-1}) than *XMM*- and *Chandra*-COSMOS at all redshifts. Compared to XBoötes, Stripe 82X pilot has $\sim$30 per cent more $L_{0.5-10\,keV} > 10^{45}$ erg s^{-1} AGN when considering all redshifts, and finds almost as many in the young universe. Though the current spectroscopic completeness of Stripe 82X pilot is comparatively lower, more area is covered, enabling identification of more high-luminosity AGN.

However, comparing the source density of the brightest objects among these two wider area surveys with Stripe 82X pilot indicates that there are still more high-luminosity AGN left to find. The space density of $L_{0.5-10\,\mathrm{keV}} > 10^{45}$ erg s^{-1} AGN found in COSMOS is 26 deg^{-2} (i.e. 54 AGN in 2.1 deg^2) and in XBoötes is 12 deg^{-2} (104 in 9 deg^2), while currently Stripe 82X has a space density of 8 deg^{-2} (132 AGN in 16.5 deg^2). We therefore anticipate that with additional spectroscopic completeness, the source density of high-luminosity AGN will subsequently increase. Of the $L_{0.5-10\,\mathrm{keV}} > 10^{45}$ erg s^{-1} AGN already identified in Stripe 82 that have optical classifications, one is a narrow-line AGN (optically classified as a 'galaxy') while the remaining have broad lines.

5.1.2 Comparison with model predictions

With the larger data set presented here, we expand on the work of LaMassa et al. (2013) and compare the luminosity distribution of X-ray AGN we immediately identify with X-ray background population synthesis predictions of Treister, Urry & Virani (2009), Gilli et al. (2007) and Ballantyne et al. (2011). We input the observed area–flux curves for *Chandra* and *XMM–Newton* into the Treister et al. (2009) simulator,[12] and convolved the predicted $\log N - \log S$ distributions from Gilli et al. (2007)[13] and Ballantyne et al. (2011) with our observed area–flux curves; we note that since the Gilli et al. (2007) predictions only allow the hard-band flux to be defined from 2 to 10 keV, we corrected the output fluxes in the *Chandra* band to our 2–7 keV range, using our assumed spectral model where $\Gamma = 1.7$. As Gilli et al. (2007) do not provide model predictions in the full band, we compare our observed full-band numbers to the models from Treister et al. (2009) and Ballantyne et al. (2011). The predicted luminosity bins represent intrinsic, rest-frame luminosities, while the Stripe 82X data are observed luminosities. We removed any *Chandra* pointings that overlapped *XMM–Newton* pointings since the latter has more effective area. Since we did detect a handful of *Chandra* sources in these removed pointings that were not identified by *XMM–Newton*, the histograms presented in Figs 9 and 10 are a sub-set of the total data.

The models from Gilli et al. (2007) predict more AGN at all redshifts and more high-luminosity ($>10^{45}$ erg s^{-1}) AGN at $z > 2$ than the Treister et al. (2009) models while the Ballantyne et al. (2011) model predicts more $L > 10^{44}$ erg s^{-1} AGN than the Treister et al. (2009) model. As Ballantyne et al. (2011) produces predictions based on three different input luminosity functions (Ueda et al. 2003; La Franca et al. 2005; Aird et al. 2010), which show a significant range in expected AGN numbers within most luminosity bins, discrepancies among models can be attributed to differences in X-ray luminosity functions (XLFs). Gilli et al. (2007) uses the XLF from Hasinger, Miyaji & Schmidt (2005) in the 0.5–2 keV band, estimating the contribution of moderately obscured AGN ($10^{21} < N_H < 10^{24}$ cm^{-2}) in the hard band by calculating the difference between the Ueda et al. (2003) and La Franca et al. (2005) XLFs and the Hasinger et al. (2005) XLF (after converting the latter to the hard band). The predictions from Treister et al. (2009) are calibrated on the hard-band XLF from Ueda et al. (2003).

Due to our limited spectroscopic completeness (see Section 4.1.1), we have identified fewer AGN than predicted given
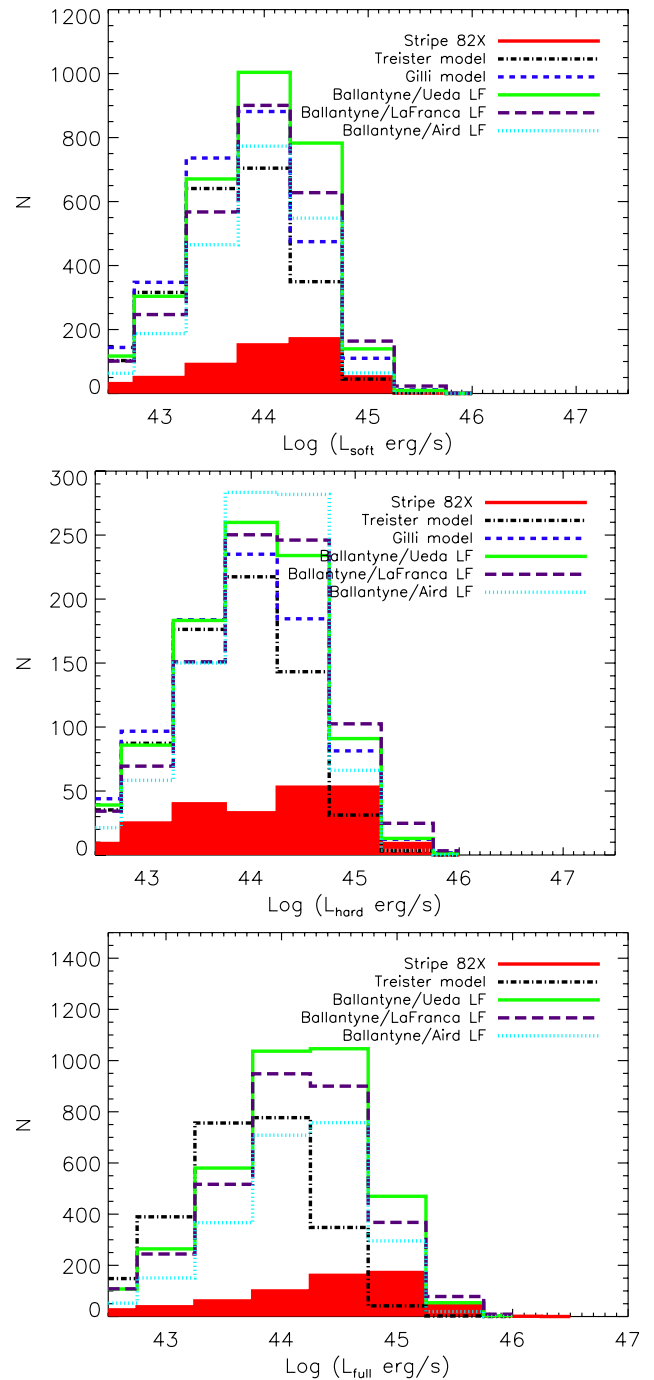

Figure 9. Luminosity distribution for Stripe 82X sources (red filled) for all sources with spectroscopic redshift compared to population synthesis models from Treister et al. (2009, dash–dotted black line), Gilli et al. (2007, dashed blue line) and Ballantyne et al. (2011, assuming different luminosity functions noted in the legend) in the (top) soft, (middle) hard and (bottom) full energy bands; models from Gilli et al. (2007) over the full X-ray band are not available. At high luminosities, we have already identified more AGN than predicted by the Treister et al. (2009) model and almost as many as predicted by the Gilli et al. (2007) model and Ballantyne et al. (2011) model, depending on the assumed luminosity function.

the constraints from our data. However, these 'missing' objects are predominantly at low to moderate luminosities ($<10^{45}$ erg s^{-1}). When considering objects at all spectroscopic redshifts, we found more high-luminosity AGN than predicted by the Treister et al.

---

[12] Model predictions from the work of Treister et al. (2009) for a range of input values are publicly available at http://agn.astroudec.cl/j_agn/main.html
[13] http://www.bo.astro.it/~gilli/counts.html, where we use their assumed spectral model of $\Gamma = 1.9$ to convert the hard-band luminosity bins into soft-band luminosity bins required by their code.
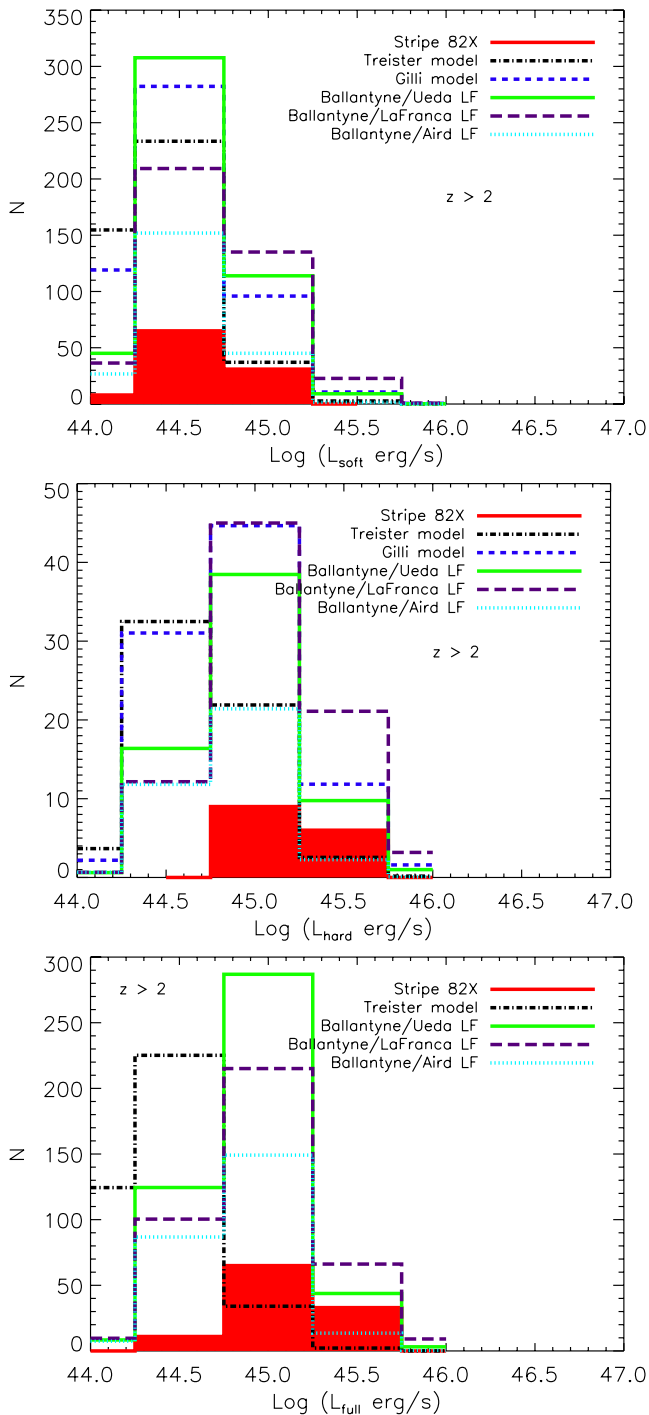
Figure 10. Same as Fig. 9 but for spectroscopic redshifts greater than 2. We identified more high-luminosity AGN than predicted by Treister et al. (2009) and Ballantyne et al. (2011) with the Aird et al. (2010) luminosity function as input in the hard and full bands, suggesting that the Stripe 82X large area survey will provide important constraints to black hole growth in the high-luminosity, high-redshift regime.

(2009) model, most of those predicted by the Gilli et al. (2007) model and a significant fraction to most, depending on the luminosity function, of those predicted by the Ballantyne et al. (2011) model. The same result applies to objects at $z > 2$ in the hard and full bands [though we do find more high-luminosity objects than the Ballantyne et al. (2011) predictions based on the Aird et al.

(2010) models in these bands], while in the soft band, the Treister et al. (2009) and Ballantyne et al. (2011) models predict slightly more high-luminosity objects than we have yet discovered. Currently, the discrepancies between our observations and the Treister et al. (2009) model are within $\sim 2\sigma$ assuming Poisson uncertainties. Given the lower space density of these objects compared to surveys with higher spectroscopic completeness (i.e. Section 5.1.1), it seems clear that more high-luminosity AGN will be confirmed. Even a small increase in the high-luminosity population would surpass the predictions of Gilli et al. (2007) and the more conservative numbers of Ballantyne et al. (2011). We also expect that our luminosity distribution is systematically lower than the predictions as the latter use intrinsic, rather than observed, luminosities as input. The systematic effect would shift the Stripe 82X sources into higher luminosity bins if corrected for absorption, making our comparison at high luminosities conservative.

Finding a greater number of high X-ray luminosity AGN relative to the model predictions is consistent with what was reported earlier by LaMassa et al. (2013), namely, that population synthesis models need to be refined to properly account for the high luminosity AGN regime. As this unexplored population is more numerous than predicted, quantifying its impact on AGN demography and evolution is critical for fully understanding black hole growth. Increased spectroscopic completeness will inform us as to the significance of the offset.

5.2 SMBH growth at high redshift

Though $z > 5$ quasars identified by SDSS have been followed up with dedicated *Chandra* observations (e.g. Brandt et al. 2002; Shemmer et al. 2006; Vignali et al. 2005), not many have been found in X-ray surveys. Thus far, only six have been confirmed spectroscopically: $z = 5.19$ in CDF-N (Barger et al. 2003), $z = 5.4$ from the *Chandra* Large Area Synoptic X-ray Survey (Steffen et al. 2004), $z = 5.3$ and $z = 5.07$ from C-COSMOS (Civano et al. 2011), and two from ChaMP, with the most distant object having a redshift of 5.41 (Trichas et al. 2012). None have been located in the 4 Ms of CDF-S, demonstrating that area trumps depth for locating these very high redshift sources. In this pilot survey of Stripe 82X, we have discovered only one object beyond a redshift of 5, but it is the most distant X-ray-selected quasar from an X-ray survey to date, at $z = 5.86$ with $L_{0.5-10\,\text{keV}} = 4.4 \times 10^{45}$ erg s^{-1} (*Chandra* source, MSID = 165442). The SDSS spectrum of this source is shown in Fig. 11, revealing broad Lyα (i.e. this source is classified as a broad-line AGN).

Such objects are expected to be quite rare. For instance, model predictions from Treister et al. (2009) estimate that only three AGN at $z > 5$ with $L_{0.5-10\,\text{keV}} > 10^{45}$ erg s^{-1} exist in this survey area, given the observed full-band area–flux curves. Similarly, using the soft-band area–flux curves, the models from Gilli et al. (2007) predict five AGN at $z > 5$ with $L_{0.5-2\,\text{keV}} > 10^{45}$ erg s^{-1}, but less than one when applying an exponential decline to the high-z luminosity function. These types of objects will be below the flux limit of *eRosita* (Merloni et al. 2012; Kolodzig et al. 2012), making the Stripe 82X survey important for constraining black hole formation models.

5.3 Obscured AGN beyond the local universe

5.3.1 WISE *AGN candidates*

In Fig. 12, we plot the *WISE W1 − W2* colour as a function of *W1* for the 1713 Stripe 82 X-ray sources with significant
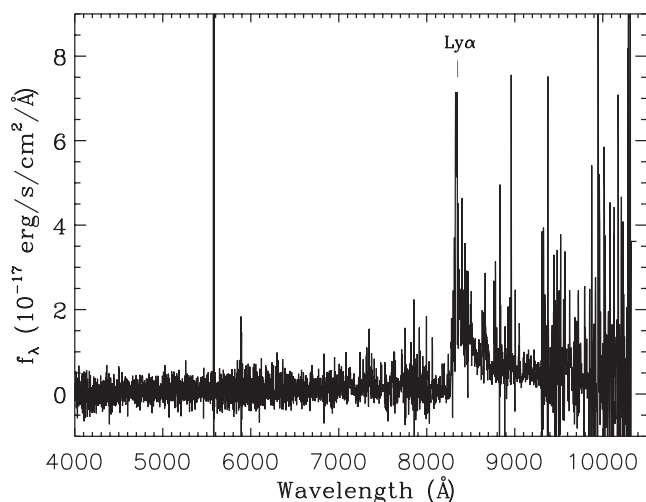
Figure 11. SDSS spectrum of the highest redshift quasar yet discovered in an X-ray survey at $z = 5.86$ with an X-ray luminosity of 4.4×10^{45} erg s^{-1}. The Lyα transition is marked. This source was discovered in the archival *Chandra* data (MSID = 165442).

detections (SNR ≥ 2) in both bands on top of the contours for all *WISE* sources with significant $W1$ and $W2$ colours in the full 300 deg^2 Stripe 82 area. The colour cut of $W1 - W2 \geq 0.8$ used to identify *WISE* AGN candidates (Stern et al. 2012; Assef et al. 2013) is overplotted, with 904 of our X-ray/*WISE* objects falling within this region, or 53 per cent of the total. This contrasts with the results of Stern et al. (2012) who find that in the COSMOS field, a majority of X-ray sources with *WISE* counterparts have blue colours, i.e. $W1 - W2 < 0.8$; for instance, only 91 out the 244 *XMM–Newton/WISE* sources in COSMOS (38 per cent) have *WISE* AGN candidate colours. A higher fraction of the Stripe 82 X-ray sources have infrared colours consistent with obscured AGN.
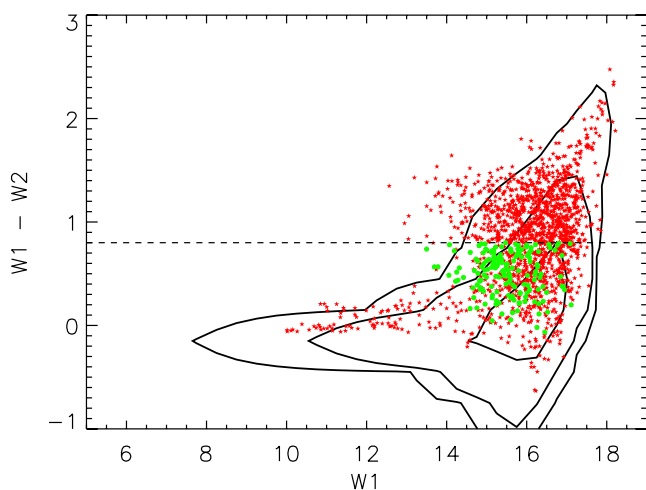
Figure 12. *WISE* colour $W1 - W2$ as a function of $W1$ with the contours indicating the density (10^3, 10^4, 10^5, 10^6 objects per contour) of all *WISE* objects in 300 deg^2 if Stripe 82, with our X-ray objects overplotted as red stars and the $W1 - W2 \geq 0.8$ AGN candidate colour cut (e.g. Assef et al. 2013; Stern et al. 2012) marked by the dashed line. About half of our X-ray objects have redder colours, while 2/3 of the objects with spectra that are X-ray identified as AGN also exceed this boundary. The 166 spectroscopically identified X-ray AGN ($L_x > 10^{42}$ erg s^{-1}) with bluer colours are shown by the green circles.

509 of the 1713 Stripe 82 X-ray sources with significant $W1$ and $W2$ detections have spectroscopic redshifts and X-ray luminosities indicative of AGN activity ($L_x > 10^{42}$ erg s^{-1} Persic et al. 2004; Brandt & Hasinger 2005). Of these, 165, or 32 per cent, have *WISE* colour $W1 - W2 < 0.8$ (green circles in Fig. 12). These results indicate that two-thirds of our spectroscopically confirmed AGN are obscured (red *WISE* colours) and that identifying AGN candidates based on a simple colour cut can miss up to a third of bluer AGN that can be recognized via other selection mechanisms, e.g. the optical and X-ray. As pointed out by Stern et al. (2012), this result reinforces the complementarity of MIR and X-ray selection in providing comprehensive views of SMBH growth.

5.3.2 Optically normal galaxies

At $z > 0.5$, the BPT diagnostic line ratio diagrams used to discriminate between type 2 AGN and star-forming galaxies (e.g. Baldwin et al. 1981; Kewley et al. 2001; Kauffmann et al. 2003) become challenging in optical surveys as Hα is shifted out of the rest-frame bandpass. Candidates for obscured AGN would then have to be identified via alternative optical diagnostics, such as ratios of narrow emission lines versus stellar mass (MEx; Juneau et al. 2011) or versus rest-frame $g - z$ colour (TBT; Trouille et al. 2011). Follow-up of type 2 AGN candidates with ground-based NIR spectroscopy to observe the traditional BPT line diagnostics is also possible (Kewley et al. 2013a,b), but is inefficient for studying a statistical ensemble of this population. Conversely, calculating an object's X-ray luminosity provides a more efficient identification mechanism. In our survey, we have identified 22 X-ray AGN at $z > 0.5$ with luminosities exceeding 10^{43} erg s^{-1} that were classified as galaxies in SDSS, 2SLAQ or DEEP2 based on their optical spectra. One of these objects is extremely bright as noted above, $L_x = 10^{45}$ erg s^{-1}, and is an example of the kind of highly luminous obscured AGN our survey is designed to uncover. Currently these sources only represent 3 per cent of our AGN sample, but we expect that more of these objects will be discovered during our spectroscopic follow-up campaign.

5.3.3 Optical dropouts

We identified 748 and 1444 optical counterparts to the *Chandra* and *XMM–Newton* sources, respectively, though 72 and 161 are discarded due to poor photometry. How many of the $\sim$400 *Chandra* and $\sim$900 *XMM–Newton* X-ray objects lacking SDSS counterparts ($r > 23$) do we find in the infrared? Most of these optical dropouts are either reddened by large amounts of dust or live at high redshift, so that the rest-frame optical light is shifted to redder wavelengths.

To answer this question, we look at two classes of optical dropouts: the X-ray sources with optical counterparts below $R_{\rm crit}$, including objects where no SDSS counterparts are found within the search radius, and the sub-set of X-ray sources without any optical counterpart within $r_{\rm search}$. In the former case, a true counterpart can be misclassified as a random association, especially if it is faint. The latter number then gives us a lower limit on the number of infrared bright optical dropout X-ray sources. Comparison of the flux limits for SDSS, *WISE* and UKIDSS to the type 1 quasar SED (i.e. broad-line AGN) from Elvis et al. (1994) demonstrate that SDSS is deeper than the *WISE* or UKIDSS observations, making the detection of IR sources that are SDSS dropouts a significant finding. We summarize these results in Table 5, detailing the number of

Table 5. Number of optical dropouts detected in X-rays.

Catalogue	*Chandra*	*XMM–Newton*	Total[a]
No SDSS counterpart within r_{search} or above R_{crit}			
X-ray	398	914	1312
IR[b]	112	371	472
WISE	95	313	401
UKIDSS	43	149	189
No SDSS counterpart within r_{search}			
X-ray	317	486	781
IR[2]	88	161	240
WISE	73	124	192
UKIDSS	37	82	116

[a]After removing duplicate sources between the *Chandra* and *XMM–Newton* catalogues.
[b]Detected in *WISE* or UKIDSS.

optical dropouts found in the IR generally and the numbers identified in the *WISE* and UKIDSS catalogues specifically. We note that the greater percentage of optical dropouts that have no counterpart within the search radius for the *Chandra* catalogue compared to *XMM–Newton* can be understood by the larger search radius used for the latter catalogue.

Over 30 per cent of the optical dropouts are detected in the infrared, making them candidates for the elusive population of obscured high-luminosity AGN at high redshift. We plot the *WISE* colours of the 151 dropouts (∼12 per cent of optical dropouts) that have significant *W*1, *W*2 and *W*3 detections (i.e. SNR > 2 in each band) in Fig. 13 for the optical dropout X-ray sources. The *WISE* colours are overlaid on the diagram from Wright et al. (2010), where the coloured loci represent different classes of astronomical objects. A majority of the optical dropouts detected in X-rays have *WISE* colours consistent with active galaxies, with nearly half having infrared colours akin to quasars. These are prime candidates for high-luminosity type 2 AGN or highly reddened quasars and will be followed up by us with NIRSPEC on Keck and ISAAC on ESO's VLT. For the remaining 840 optical dropouts without infrared associations (25 per cent of the X-ray sample), deeper optical and infrared imaging is necessary to identify the multiwavelength counterparts to the X-ray sources.

6 CONCLUSIONS

We have reduced and analysed the ∼10.5 deg^2 of *XMM–Newton* data overlapping SDSS Stripe 82, including ∼4.6 deg^2 of proprietary data awarded to us in AO 10. From these observations, we detected 2358 unique X-ray sources at high significance, with 2073, 607 and 2079 in the soft (0.5–2 keV), hard (2–10 keV), and full (0.5–10 keV) bands, respectively. The $\log N - \log S$ relations show general agreement with previous surveys in these bands, given the effect that choice of spectral model affects the normalization in the full band. Using an MLE algorithm (Sutherland & Saunders 1992; Brusa et al. 2005, 2007, 2010; Cardamone et al. 2008; Luo et al. 2010; Civano et al. 2012), we identified multiwavelength counterparts to Stripe 82 X-ray sources, finding:

(i) 1892 optical matches from SDSS, of which 759 have spectroscopic redshifts; 1855 *WISE* counterparts; 1754 UKIDSS matches; 447 ultraviolet counterparts from *GALEX*; and 119 radio sources from FIRST (using nearest-neighbour matching rather than MLE due to low source densities).
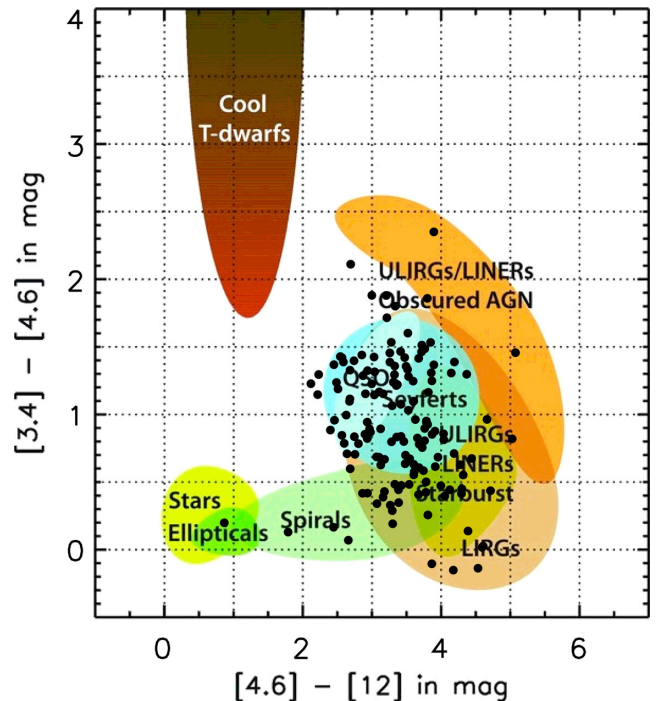

Figure 13. *WISE* colour–colour diagram for X-ray sources (filled circles) detected significantly in the *W*1, *W*2 and *W*3 bands that have no optical counterpart within the search radius or where optical sources are found within r_{search} but are below R_{crit} and are therefore not likely candidates for the true optical counterpart to the X-ray source. The coloured loci represent the classes of objects with these *WISE* colours, defined by Wright et al. (2010). Most of the optical dropouts are consistent with active galaxies.

(ii) Focusing on the sub-set of sources with spectroscopic redshifts, Stripe 82X harbours more high-luminosity ($L_x \geq 10^{45}$ erg s^{-1}) AGN than E-CDFS and CDFS (∼0.3 deg^2; Cardamone et al. 2010), *XMM*- and *Chandra*-COSMOS (∼2.1 deg^2; Cappelluti et al. 2009; Brusa et al. 2010; Civano et al. 2012) and even the larger XBoötes survey (∼9 deg^2; Kenter et al. 2005; Kochanek et al. 2012). Though these other surveys benefited from years of spectroscopic follow-up, Stripe 82X covers a wider area and thereby already uncovers more rare objects (high-luminosity AGN at all redshifts and in the early universe, at $z > 2$). These numbers will increase with the spectroscopic follow-up we are currently undertaking.

(iii) We have compared the luminosity distribution of X-ray sources with spectroscopic redshifts with the population synthesis model predictions from Treister et al. (2009), Gilli et al. (2007) and Ballantyne et al. (2011) taking into account the observational constraints of our observed area–flux curves in the soft, hard and full X-ray bands. As we showed in LaMassa et al. (2013) using a sub-set of these data and the full-band (Treister et al. 2009) model predictions given our area–flux curves, we discovered more high-luminosity (>10^{45} erg s^{-1}) AGN than predicted by Treister et al. (2009). Though the Gilli et al. (2007) and Ballantyne et al. (2011) models predict more AGN, we have found most of those predicted at high luminosity (depending on the luminosity function for the Ballantyne et al. 2011 model), and this number will continue to increase with our spectroscopic follow-up. Refinement to models is clearly indicated to better account for this important regime of black hole growth. As these rare, high-luminosity AGN are more numerous than previously predicted, understanding their census,

evolution and connection to the host galaxy becomes an important piece in completing the puzzle of cosmic black hole growth.

(iv) We have found the most distant, spectroscopically confirmed X-ray selected quasar in an X-ray survey to date, at $z = 5.86$.

(v) About a third of the X-ray sources that are optical dropouts are identified in the infrared, making them candidates for reddened quasars and/or high-luminosity type 2 AGN at high redshift. Most of those with significant detections in the *W*1, *W*2 and *W*3 *WISE* bands have colours consistent with active galaxies, with more than half of them having quasar colours. We have a Keck-NIRSPEC campaign and were awarded ESO VLT ISAAC DDT time to follow-up these objects.

The Stripe 82X survey provides an important pathfinder mission to *eRosita*, scheduled to be launched in 2014, which will survey the entire sky in 0.5–10 keV X-rays, though with a poorer resolution than *Chandra* and *XMM–Newton* ($\sim$25 arcsec) and with an 0.5–2 keV flux limit that is five times higher than our proprietary *XMM–Newton* mosaicked observations (Merloni et al. 2012). *eRosita* expects to uncover millions of AGN, of which a few tens will be $z > 6$ QSOs. An efficient method will then need to be devised to isolate the very high-redshift population: results of Stripe 82X, with the wealth of multiwavelength data, will help to inform robust identification techniques applicable to *eRosita*.

In other luminosity ranges, X-ray selection has uncovered a different, if overlapping, population of AGN compared to optical selection. While the jury is still out on how this impacts black hole growth at high luminosity, it is clear that the answer requires large samples selected at X-ray energies, so that the optical- and X-ray samples can be compared.

REFERENCES

Ahn C. P. et al., 2012, ApJS, 203, 21
Aird J. et al., 2010, MNRAS, 401, 2531
Alexander D. M. et al., 2003, AJ, 126, 539
Assef R. J. et al., 2013, ApJ, 772, 26
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Ballantyne D. R., Draper A. R., Madsen K. K., Rigby J. R., Treister E., 2011, ApJ, 736, 56
Barger A. J., Cowie L. L., Capak P., Alexander D. M., Bauer F. E., Brandt W. N., Garmire G. P., Hornschemeier A. E., 2003, ApJ, 584, L61
Becker R. H., White R. L., Helfand D. J., 1995, ApJ, 450, 559
Becker R. H., Helfand D. J., White R. L., Gregg M. D., Laurent-Muehleisen S. A., 2012, VizieR Online Data Catalog, 8090, 0
Bianchi L., Efremova B., Herald J., Girardi L., Zabot A., Marigo P., Martin C., 2011, MNRAS, 411, 2770
Brandt W. N., Hasinger G., 2005, ARA&A, 43, 827
Brandt W. N. et al., 2002, ApJ, 569, L5
Brunner H., Cappelluti N., Hasinger G., Barcons X., Fabian A. C., Mainieri V., Szokoly G., 2008, A&A, 479, 283
Brusa M. et al., 2005, A&A, 432, 69
Brusa M. et al., 2007, ApJS, 172, 353
Brusa M. et al., 2010, ApJ, 716, 348
Budavári T. et al., 2009, ApJ, 694, 1281
Cappelluti N. et al., 2007, ApJS, 172, 341
Cappelluti N. et al., 2009, A&A, 497, 635
Cardamone C. N. et al., 2008, ApJ, 680, 130
Cardamone C. N. et al., 2010, ApJS, 189, 270
Casali M. et al., 2007, A&A, 467, 777
Chiappetti L. et al., 2013, MNRAS, 429, 1652
Civano F. et al., 2011, ApJ, 741, 91
Civano F. et al., 2012, ApJS, 201, 30
Comastri A. et al., 2011, A&A, 526, L9
Croom S. M. et al., 2009, MNRAS, 392, 19
Cutri R. M. et al., 2012, Explanatory Supplement to the WISE All-Sky Data Release Products. p. 1, available at: http://wise2.ipac.caltech.edu/docs/release/allsky/expsup/index.html
Davis M. et al., 2007, ApJ, 660, L1
Donley J. L. et al., 2012, ApJ, 748, 142
Drinkwater M. J. et al., 2010, MNRAS, 401, 1429
Dye S. et al., 2006, MNRAS, 372, 1227
Elvis M. et al., 1994, ApJS, 95, 1
Elvis M. et al., 2009, ApJS, 184, 158
Evans I. N. et al., 2010, ApJS, 189, 37
Georgakakis A. et al., 2007, ApJ, 660, L15
Giacconi R. et al., 2001, ApJ, 551, 624
Giavalisco M. et al., 2004, ApJ, 600, L93
Gilli R., Comastri A., Hasinger G., 2007, A&A, 463, 79
Goulding A. D. et al., 2012, ApJS, 202, 6
Hambly N. C. et al., 2008, MNRAS, 384, 637
Hasinger G., Miyaji T., Schmidt M., 2005, A&A, 441, 417
Hewett P. C., Warren S. J., Leggett S. K., Hodgkin S. T., 2006, MNRAS, 367, 454
Jiang L. et al., 2009, AJ, 138, 305
Juneau S., Dickinson M., Alexander D. M., Salim S., 2011, ApJ, 736, 104
Kauffmann G. et al., 2003, MNRAS, 346, 1055
Kenter A. et al., 2005, ApJS, 161, 9
Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, ApJ, 556, 121
Kewley L. J., Dopita M. A., Leitherer C., Davé R., Yuan T., Allen M., Groves B., Sutherland R., 2013a, ApJ, 774, 100
Kewley L. J., Maier C., Yabe K., Ohta K., Akiyama M., Dopita M. A., Yuan T., 2013b, ApJ, 774, 10
Kim M., Wilkes B. J., Kim D.-W., Green P. J., Barkhouse W. A., Lee M. G., Silverman J. D., Tananbaum H. D., 2007, ApJ, 659, 29
Kochanek C. S. et al., 2012, ApJS, 200, 8
Kolodzig A., Gilfanov M., Sunyaev R., Sazonov S., Brusa M., 2013, A&A, 558, 89

La Franca F. et al., 2005, ApJ, 635, 864

Lacy M. et al., 2004, ApJS, 154, 166

LaMassa S. M. et al., 2013, MNRAS, 432, 1351

Lawrence A. et al., 2007, MNRAS, 379, 1599

Lehmer B. D. et al., 2005, ApJS, 161, 21

Loaring N. S. et al., 2005, MNRAS, 362, 1371

Luo B. et al., 2010, ApJS, 187, 560

McGreer I. D. et al., 2013, ApJ, 768, 105

Mateos S. et al., 2008, A&A, 492, 51

Mendez A. J. et al., 2013, ApJ, 770, 40

Merloni A. et al., 2012, preprint (arXiv:1209.3114)

Morrissey P. et al., 2007, ApJS, 173, 682

Murray S. S. et al., 2005, ApJS, 161, 1

Newman J. A. et al., 2013, ApJS, 208, 5

Ochsenbein F., Bauer P., Marcout J., 2000, A&AS, 143, 23

Persic M., Rephaeli Y., Braito V., Cappi M., Della Ceca R., Franceschini A.,
     Gruber D. E., 2004, A&A, 419, 849

Pierre M. et al., 2004, J. Cosmol. Astropart. Phys., 9, 11

Ranalli P. et al., 2013, A&A, 555, A42

Rots A. H., Budavári T., 2011, ApJS, 192, 8

Shemmer O. et al., 2006, ApJ, 644, 86

Steffen A. T., Barger A. J., Capak P., Cowie L. L., Mushotzky R. F., Yang
     Y., 2004, AJ, 128, 1483

Stern D. et al., 2005, ApJ, 631, 163

Stern D. et al., 2012, ApJ, 753, 30

Sutherland W., Saunders W., 1992, MNRAS, 259, 413

Treister E. et al., 2004, ApJ, 616, 123

Treister E., Urry C. M., Virani S., 2009, ApJ, 696, 110

Trichas M. et al., 2012, ApJS, 200, 17

Trouille L., Barger A. J., Tremonti C., 2011, ApJ, 742, 46

Ueda Y., Akiyama M., Ohta K., Miyaji T., 2003, ApJ, 598, 886

Viero M. P. et al., 2013, ApJ, 772, 77

Vignali C., Brandt W. N., Schneider D. P., Kaspi S., 2005, AJ, 129,
     2519

Virani S. N., Treister E., Urry C. M., Gawiser E., 2006, AJ, 131, 2373

Warren S. J. et al., 2007, preprint (arXiv:astro-ph/0703037)

Watson M. G. et al., 2009, A&A, 493, 339

White R. L., Becker R. H., Helfand D. J., Gregg M. D., 1997, ApJ, 475,
     479

Wright E. L. et al., 2010, AJ, 140, 1868

Xue Y. Q. et al., 2011, ApJS, 195, 10

APPENDIX A: RELIABILITY THRESHOLDS FOR COUNTERPART SELECTION

As mentioned in the main text, our goal is to optimize selection of multiwavelength counterparts to the Stripe 82 X-ray sources by maximizing the number of true associations while minimizing contamination from chance coincidences. We inspected the distribution of source 'reliabilities' calculated via MLE and picked a critical threshold (R_{crit}) above which we expect a vast majority of the ancillary objects represent true counterparts. By shifting the X-ray positions by random amounts and running the MLE code, the distribution of resulting reliabilities provides an empirical estimate of the contamination in our matched catalogues.

In Figs A1–A7, we compare the reliability distribution of each wavelength band to which we matched (solid black histogram) with the reliability distribution after shifting the X-ray positions (by ∼21 to ∼35 arcsec) overplotted (blue histogram). The dotted line indicates the specific R_{crit} value we used for that band. In the captions, we note the number of spurious associations expected, i.e. ancillary counterparts matched to random positions on the sky, above R_{crit}. We stress that contamination percentages that can be calculated from this test are not exact, but are instead meant to provide an empirical method for calibrating the reliabilities on a band-by-band basis.
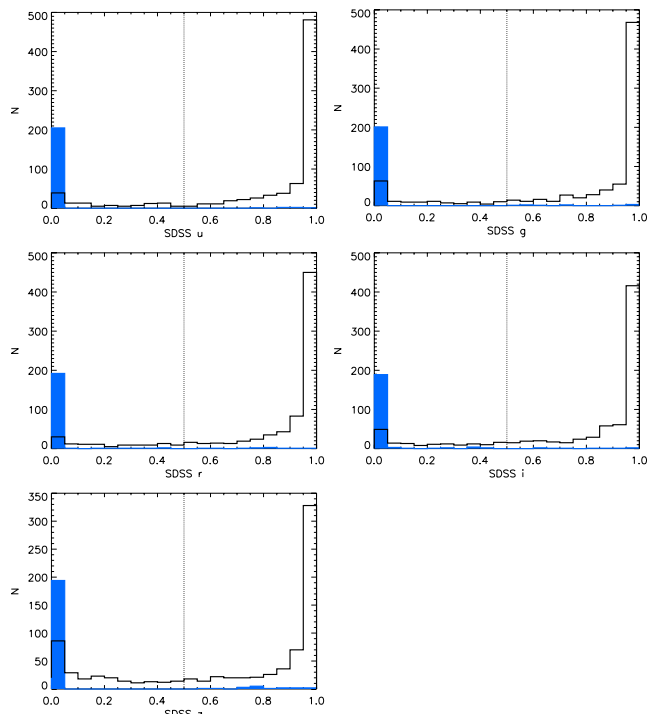

Figure A1. Reliability distributions for each SDSS band matched to *Chandra* sources. The number of spurious associations above R_{crit} is predicted to be 6, 10, 11, 9 and 17 in the *u*, *g*, *r*, *i* and *z* bands, respectively. The dot–dashed line indicates the adopted reliability threshold for claiming a counterpart.
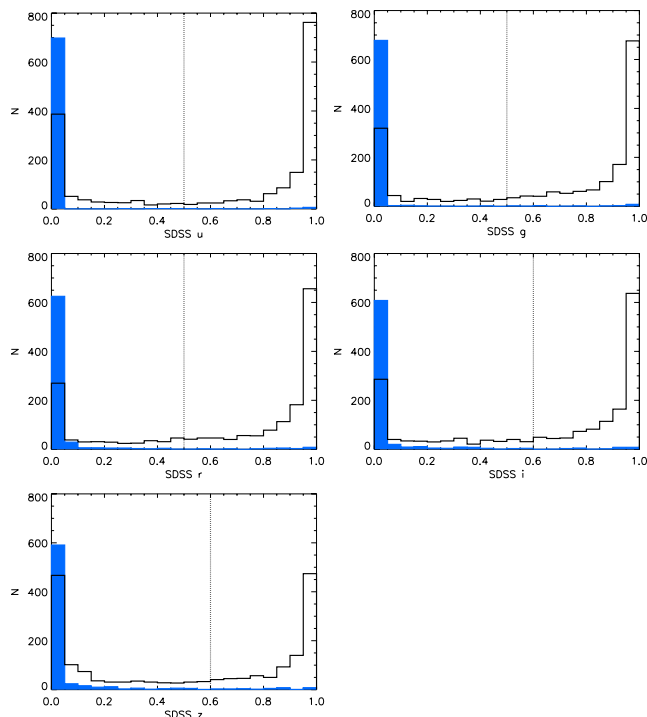

Figure A2. Reliability distributions for each SDSS band matched to *XMM–Newton* sources. The number of spurious associations above R_{crit} is predicted to be 12, 22, 28, 28 and 34 in the *u*, *g*, *r*, *i* and *z* bands, respectively. The dot–dashed line indicates the adopted reliability threshold for claiming a counterpart.
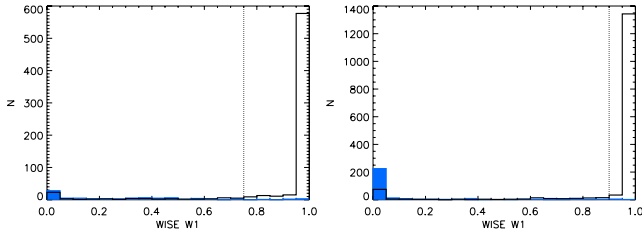
Figure A3. Reliability distributions for *WISE* band *W*1 matched to (left) *Chandra* sources and to (right) *XMM–Newton* objects. The number of spurious associations above R_{crit} is predicted to be 6 for the *Chandra* source list and 7 for the *XMM–Newton* catalogue. The dot–dashed line indicates the adopted reliability threshold for claiming a counterpart.
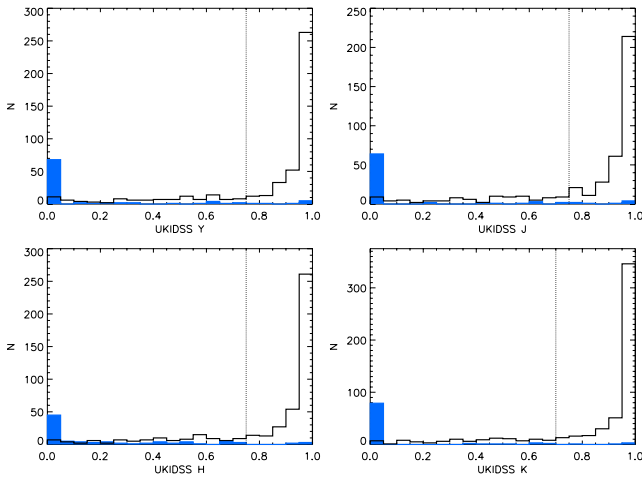

Figure A4. Reliability distributions for each UKIDSS band matched to *Chandra* sources. The number of spurious associations above R_{crit} is predicted at 8, 8, 5 and 5 in the *Y*, *J*, *H* and *K* bands, respectively. The dot–dashed line indicates the adopted reliability threshold for claiming a counterpart.

Figure A5. Reliability distributions for each UKIDSS band matched to *XMM–Newton* sources. The number of spurious associations above R_{crit} is predicted to be 21, 30, 21 and 27 in the *Y*, *J*, *H* and *K* bands, respectively. The dot–dashed line indicates the adopted reliability threshold for claiming a counterpart.
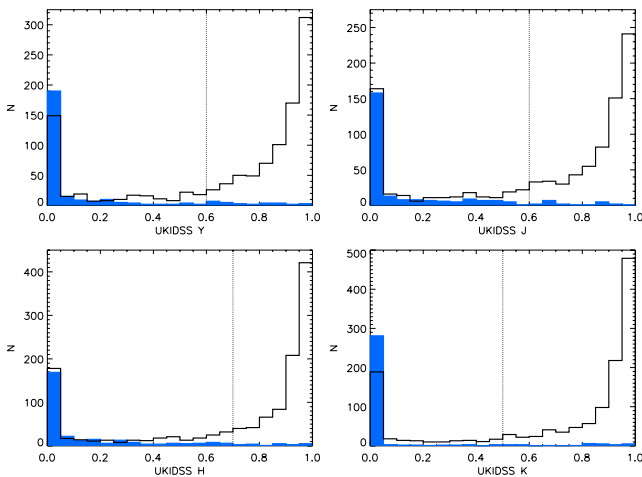
As in all multiwavelength surveys, a handful of true counterparts may be missed, falling below R_{crit}, while several random coincident matches may be promoted as real matches. However, our empirical tests indicate that this effect is at the few per cent level at most.
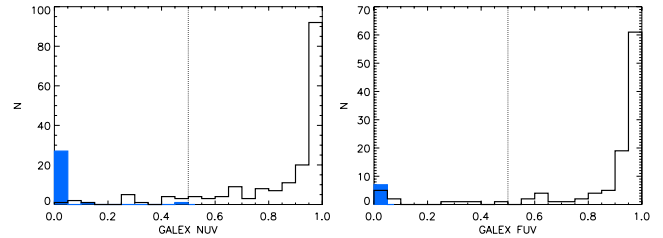

Figure A6. Reliability distributions for each *GALEX* band matched to *Chandra* sources. When shifting the X-ray positions by random amounts and re-running the MLE code, we find no spurious associations above R_{crit} in either band. The dot–dashed line indicates the adopted reliability threshold for claiming a counterpart.
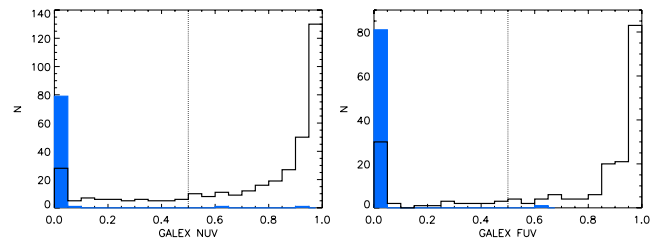

Figure A7. Reliability distributions for each *GALEX* band matched to *XMM–Newton* sources. We find two spurious counterparts above R_{crit} in the NUV band and one in the FUV band. The dot–dashed line indicates the adopted reliability threshold for claiming a counterpart.

APPENDIX B: COLUMN DESCRIPTIONS FOR ONLINE VERSIONS OF THE CATALOGUES

Non-significant X-ray fluxes have zero values in the online catalogues. When reporting the ancillary multiwavelength data, numeric values of -999 and null strings indicate that a reliable counterpart was not identified for that X-ray source.

B1 *Chandra*

(1) MSID: *Chandra* Source Catalog identification number (Evans et al. 2010).

(2) ObsID: *Chandra* observation identification number.

(3) RA: *Chandra* RA (J2000).

(4) Dec: *Chandra* Dec. (J2000).

(5) RADec_err: *Chandra* positional error (arcsec).

(6) Dist_nn: Distance to nearest *Chandra* source (arcsec).

(7) Soft_Flux: 0.5–2 keV flux (10^{-14} erg cm^{-2} s^{-1}). Set to 0 if flux is not significant at $>4.5\sigma$ level.

(8) Soft_flux_error_high: higher bound on 0.5–2 keV flux (10^{-14} erg cm^{-2} s^{-1}). If flux is 0, this is the flux upper limit.

(9) Soft_flux_error_low: lower bound on 0.5–2 keV flux (10^{-14} erg cm^{-2} s^{-1}).

(10) Hard_flux: 2–7 keV flux (10^{-14} erg cm^{-2} s^{-1}). Set to 0 if flux is not significant at $>4.5\sigma$ level.

(11) Hard_flux_error_high: higher bound on 2–7 keV flux (10^{-14} erg cm^{-2} s^{-1}). If flux is 0, this is the flux upper limit.

(12) Hard_flux_error_low: lower bound on 2–7 keV flux (10^{-14} erg cm^{-2} s^{-1}).

(13) Full_flux: 0.5–7 keV flux (10^{-14} erg cm^{-2} s^{-1}). Set to 0 if flux is not significant at $>4.5\sigma$ level.

(14) Full_flux_error_high: higher bound on 0.5–7 keV flux (10^{-14} erg cm^{-2} s^{-1}). If flux is 0, this is the flux upper limit.

(15) Full_flux_error_low: lower bound on 0.5–7 keV flux (10^{-14} erg cm^{-2} s^{-1}).

(16) Lum_soft: log 0.5–2 keV luminosity (erg s^{-1}).

(17) Lum_hard: log 2–7 keV luminosity (erg s^{-1}).

(18) Lum_full: log 0.5–7 keV luminosity (erg s^{-1}).

(19) In_XMM: set to 'yes' if X-ray source is in the *XMM–Newton* Stripe 82 catalogue.

(20) Removed_LogN_LogS: set to 'yes' if X-ray source was not part of the log$N$−log$S$ relation published in (LaMassa et al. 2013).

(21) SDSS_Rej: set to 'yes' if SDSS counterpart is found but rejected due to poor photometry.

(22) SDSS_Objid: SDSS object identification number.

(23) SDSS_RA: SDSS RA (J2000).

(24) SDSS_Dec: SDSS Dec. (J2000).

(25) SDSS_Rel: MLE reliability of SDSS match to X-ray source.

(26) SDSS_Dist: distance between X-ray and SDSS source (arcsec).

(27) u_mag: SDSS *u* mag.

(28) u_err: SDSS *u* mag error.

(29) g_mag: SDSS *g* mag.

(30) g_err: SDSS *g* mag error.

(31) r_mag: SDSS *r* mag.

(32) r_err: SDSS *r* mag error.

(33) i_mag: SDSS i mag.

(34) i_err: SDSS *i* mag error.

(35) z_mag: SDSS *z* mag.

(36) z_err: SDSS *z* mag error.

(37) Specobjid: SDSS spectroscopic object identification number.

(38) Class: optical spectroscopic class (if available).

(39) Redshift: spectroscopic redshift.

(40) z_src: source of spectroscopic redshift; 0 – SDSS, 1 – 2SLAQ, 2 – WiggleZ, 3 – DEEP2, 4 – SDSS spectra refitted/verified by us.

(41) WISE_Name: *WISE* name.

(42) WISE_RA: *WISE* RA (J2000).

(43) WISE_Dec: *WISE* Dec. (J2000).

(44) WISE_sigra: *WISE* RA error (arcsec).

(45) WISE_sigdec: *WISE* Dec. error (arcsec).

(46) WISE_Rel: MLE reliability of *WISE* match to X-ray source.

(47) WISE_Dist: distance between X-ray and *WISE* source (arcsec).

(48) W1: *WISE W*1 mag. All *WISE* magnitudes are from profile-fitting photometry, unless the WISE_ext flag is set to 'yes,' in which case the magnitudes are associated with elliptical apertures.

(49) W1sig: *WISE W*1 error.

(50) W1SNR: *WISE W*1 SNR. Any *WISE* magnitudes with SNR <2 are upper limits.

(51) W2: *WISE W*2 mag.

(52) W2sig: *WISE W*2 error.

(53) W2SNR: *WISE W*2 SNR.

(54) W3: *WISE W*3 mag.

(55) W3sig: *WISE W*3 error.

(56) W3SNR: *WISE W*3 SNR.

(57) W4: *WISE W*4 mag.

(58) W4sig: *WISE W*4 error.

(59) W4SNR: *WISE W*4 SNR.

(60) WISE_ext: set to 'yes' if *WISE* source is extended.

(61) WISE_rej: set to 'yes' if *WISE* counterpart is found but rejected for poor photometry.

(62) UKIDSS_ID: UKIDSS ID.

(63) UKIDSS_RA: UKIDSS RA (J2000).

(64) UKIDSS_Dec: UKIDSS Dec. (J2000).

(65) UKIDSS_Rel: MLE reliability of UKIDSS match to X-ray source.

(66) UKIDSS_Dist: distance between X-ray and UKIDSS source (arcsec).

(67) Ymag: UKIDSS *Y* mag.

(68) Ysig: UKIDSS *Y* error.

(69) Hmag: UKIDSS *H* mag.

(70) Hsig: UKIDSS *H* error.

(71) Jmag: UKIDSS *J* mag.

(72) Jsig: UKIDSS *J* error.

(73) Kmag: UKIDSS *K* mag.

(74) Ksig: UKIDSS *K* error.

(75) UKIDSS_Rej: UKIDSS counterpart found but rejected for poor photometry.

(76) GALEX_Objid: *GALEX* object identification number.

(77) GALEX_RA: *GALEX* RA (J2000).

(78) GALEX_Dec: *GALEX* Dec. (J2000).

(79) NUV_poserr: *GALEX* NUV positional error (arcsec).

(80) FUV_poserr: *GALEX* FUV positional error (arcsec).

(81) GALEX_Rel: MLE reliability of *GALEX* match to X-ray source.

(82) GALEX_Dist: distance between X-ray and *GALEX* source (arcsec).

(83) NUV_mag: *GALEX* NUV mag.

(84) NUV_magerr: *GALEX* NUV error.

(85) FUV_mag: *GALEX* FUV mag.

(86) FUV_magerr: *GALEX* FUV error.

(87) FIRST Name: IAU name of FIRST counterpart.

(88) FIRST_RA: FIRST RA (J2000).

(89) FIRST_Dec: FIRST Dec. (J2000).

(90) FIRST_Dist: distance between X-ray and FIRST source (arcsec).

(91) FIRST_Flux: FIRST 5 GHz flux density (Jy).

(92) FIRST_err: FIRST 5 GZ flux density error (Jy).

B2 *XMM–Newton*

(1) Rec_no: unique record number assigned to each *XMM–Newton* source.

(2) ObsID: *XMM–Newton* observation identification number.

(3) RA: *XMM–Newton* RA (J2000).

(4) Dec: *XMM–Newton* Dec (J2000).

(5) RADec_Err: *XMM–Newton* positional error (arcsec).

(6) Dist_nn: distance to nearest *XMM–Newton* source (arcsec).

(7) Soft_flux: 0.5–2 keV flux (10^{-14} erg cm^{-2} s^{-1}). Flux is 0 if det_ml < 15 in the soft band.

(8) Soft_flux_err: error in 0.5–2 keV flux (10^{-14} erg cm^{-2} s^{-1}).

(9) Hard_flux: 2–10 keV flux (10^{-14} erg cm^{-2} s^{-1}). Flux is 0 if det_ml < 15 in the hard band.

(10) Hard_flux_err: error in 2–10 keV flux (10^{-14} erg cm^{-2} s^{-1}).

(11) Full_flux: 0.5–10 keV flux (10^{-14} erg cm^{-2} s^{-1}). Flux is 0 if det_ml < 15 in the full band.

(12) Full_flux_err: error in 0.5–10 keV flux (10^{-14} erg cm^{-2} s^{-1}).

(13) Lum_soft: log 0.5–2k keV luminosity (erg s^{-1}).

(14) Lum_hard: log 2–10 keV luminosity (erg s^{-1}).

(15) Lum_full: log 0.5–10 keV luminosity (erg s^{-1}).

(16) In_Chandra: set to 'yes' if source is found in the *Chandra* catalogue.

(17) Removed_LogN_LogS: set to 'yes' if source is removed from log$N$−log$S$ calculation presented in the main text.

(18) SDSS_Rej: set to 'yes' if SDSS counterpart is found but rejected due to poor photometry.

(19) SDSS_Objid: SDSS object identification number.

(20) SDSS_RA: SDSS RA (J2000).

(21) SDSS_Dec: SDSS Dec. (J2000).

(22) SDSS_Rel: MLE reliability of SDSS match to X-ray source.

(23) SDSS_Dist: distance between X-ray and SDSS source (arcsec).

(24) u_mag: SDSS *u* mag.

(25) u_err: SDSS *u* mag error.

(26) g_mag: SDSS *g* mag.

(27) g_err: SDSS *g* mag error.

(28) r_mag: SDSS *r* mag.

(29) r_err: SDSS *r* mag error.

(30) i_mag: SDSS *i* mag.

(31) i_err: SDSS *i* mag error.

(32) z_mag: SDSS *z* mag.

(33) z_err: SDSS *z* mag error.

(34) Specobjid: SDSS spectroscopic object identification number.

(35) Class: optical spectroscopic class (if available).

(36) Redshift: spectroscopic redshift.

(37) z_src: source of spectroscopic redshift; 0 – SDSS, 1 – 2SLAQ, 2 – WiggleZ, 3 – DEEP2, 4 – SDSS spectra re-fit/verified by us.

(38) WISE_Name: *WISE* name.

(39) WISE_RA: *WISE* RA (J2000).

(40) WISE_Dec: *WISE* Dec. (J2000).

(41) WISE_sigra: *WISE* RA error (arcsec).

(42) WISE_sigdec: *WISE* Dec. error (arcsec).

(43) WISE_Rel: MLE reliability of *WISE* match to X-ray source.

(44) WISE_Dist: distance between X-ray and *WISE* source (arcsec).

(45) W1: *WISE* *W*1 mag. All *WISE* magnitudes are from profile-fitting photometry, unless the WISE_ext flag is set to 'yes,' in which case the magnitudes are associated with elliptical apertures.

(46) W1sig: *WISE* *W*1 error.

(47) W1SNR: *WISE* *W*1 SNR. Any *WISE* magnitudes with SNR <2 are upper limits.

(48) W2: *WISE* *W*2 mag.

(49) W2sig: *WISE* *W*2 error.

(50) W2SNR: *WISE* *W*2 SNR.

(51) W3: *WISE* *W*3 mag.

(52) W3sig: *WISE* *W*3 error.

(53) W3SNR: *WISE* *W*3 SNR.

(54) W4: *WISE* *W*4 mag.

(55) W4sig: *WISE* *W*4 error.

(56) W4SNR: *WISE* *W*4 SNR.

(57) WISE_ext: set to 'yes' if *WISE* source is extended.

(58) WISE_rej: set to 'yes' if *WISE* counterpart is found but rejected for poor photometry.

(59) UKIDSS_ID: UKIDSS ID.

(60) UKIDSS_RA: UKIDSS RA (J2000).

(61) UKIDSS_Dec: UKIDSS Dec. (J2000).

(62) UKIDSS_Rel: MLE reliability of UKIDSS match to X-ray source.

(63) UKIDSS_Dist: distance between X-ray and UKIDSS source (arcsec).

(64) Ymag: UKIDSS *Y* mag.

(65) Ysig: UKIDSS *Y* error.

(66) Hmag: UKIDSS *H* mag.

(67) Hsig: UKIDSS *H* error.

(68) Jmag: UKIDSS *J* mag.

(69) Jsig: UKIDSS *J* error.

(70) Kmag: UKIDSS *K* mag.

(71) Ksig: UKIDSS *K* error.

(72) UKIDSS_Rej: UKIDSS counterpart found but rejected for poor photometry.

(73) GALEX_Objid: *GALEX* object identification number.

(74) GALEX_RA: *GALEX* RA (J2000).

(75) GALEX_Dec: *GALEX* Dec. (J2000).

(76) NUV_poserr: *GALEX* NUV positional error (arcsec).

(77) FUV_poserr: *GALEX* FUV positional error (arcsec).

(78) GALEX_Rel: MLE reliability of *GALEX* match to X-ray source.

(79) GALEX_Dist: distance between X-ray and *GALEX* source (arcsec).

(80) NUV_mag: *GALEX* NUV mag.

(81) NUV_magerr: *GALEX* NUV error.

(82) FUV_mag: *GALEX* FUV mag.

(83) FUV_magerr: *GALEX* FUV error.

(84) FIRST Name: IAU name of FIRST counterpart.

(85) FIRST_RA: FIRST RA (J2000).

(86) FIRST_Dec: FIRST Dec. (J2000).

(87) FIRST_Dist: distance between X-ray and FIRST source (arcsec).

(88) FIRST_Flux: FIRST 5 GHz flux density (Jy).

(89) FIRST_err: FIRST 5 GZ flux density error (Jy).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Catalogues listing the X-ray sources, fluxes and matches to the ancillary multiwavelength catalogues, including the non-aperture-matched photometry. (http://mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/stt1837/-/DC1).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

This paper has been typeset from a TEX/LATEX file prepared by the author.