

# The X-SERVS Survey: New *XMM-Newton* Source Catalog for the XMM-LSS field

C.-T. J. Chen<sup>1,2,\*</sup>, W. N. Brandt<sup>1,2,3</sup>, B. Luo<sup>4</sup>, and Friends

<sup>1</sup>*Department of Astronomy & Astrophysics, 525 Davey Lab, The Pennsylvania State University, University Park, PA 16802, USA*

<sup>2</sup>*Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA*

<sup>3</sup>*Department of Physics, The Pennsylvania State University, University Park, PA 16802, USA*

<sup>4</sup>*School of Astronomy and Space Science, Nanjing University, Nanjing 210093, China*

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

We present source catalogs from the XMM-Large Scale Structure survey region (XMM-LSS), one of the XMM-Spitzer Extragalactic Representative Volume Survey (X-SERVS) fields. The survey is comprised of 1.1 Ms new *XMM-Newton* AO-15 data and 1.6 Ms archival data, totaling 2.7 Ms flare-filtered exposure over a 5.3 deg<sup>2</sup> area. We provide X-ray catalogs of 5218 sources detected in the soft (0.5–2 keV), hard (2–10 keV), and full (0.5–10 keV) bands. Sources are first detected using *XMM-Newton* SAS task EWAVELET with THRESHOLD  $\geq 4$  and further verified by the SAS task EMLDETECT with DET\_ML  $\geq 6$ , which is equivalent to a false-detection probability of  $P = 0.0025$ . A total of 2375 new X-ray sources are detected compared to the XMM-XXL-North survey in the same area. Our flux limits and distributions are comparable to those of the XMM-COSMOS survey. The median fluxes in the soft, hard, and full bands (in erg cm<sup>-2</sup> s<sup>-1</sup>) are  $2.65 \times 10^{-15}$ ,  $1.48 \times 10^{-14}$ , and  $8.78 \times 10^{-15}$ , respectively. We identify multiwavelength counterparts for 99.5% of the X-ray sources, of which 93% of them have matching likelihood ratios satisfying our reliability thresholds. The reliabilities of these high likelihood ratio counterparts are further confirmed to be  $\approx 97\%$  reliable using the available *Chandra* sources over  $\approx 5\%$  of XMM-LSS. Results of multiwavelength identifications are also included in the source catalog, along with basic optical-to-infrared photometry as well as spectroscopic redshifts from the publicly available surveys.

**Key words:** keyword1 – keyword2 – keyword3

## 1 INTRODUCTION

PLACE HOLDER

## 2 XMM-NEWTON OBSERVATIONS IN THE XMM-LSS REGION AND DATA REDUCTION

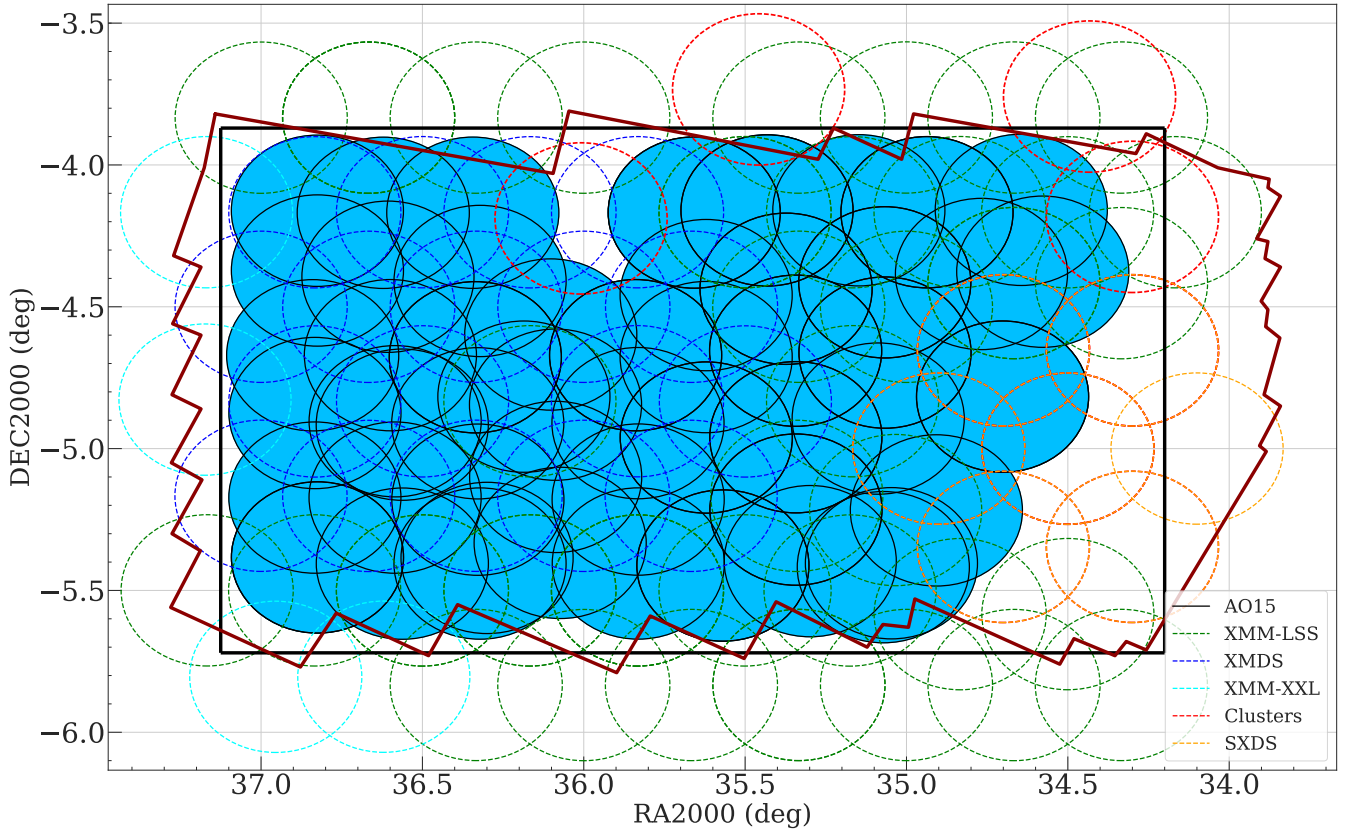
### 2.1 XMM-Newton data in the XMM-LSS region

The XMM-LSS field has been targeted by a number of *XMM-Newton* surveys of different sensitivities. The original XMM-LSS survey was a  $\approx 11$  deg<sup>2</sup> field typically covered by observations of  $\approx 10$  ks exposure time per pointing, each separated by 20′ (Pacaud et al. 2006; Pierre et al. 2016). Within the 11 deg<sup>2</sup> field,  $\approx 4$  deg<sup>2</sup> were observed by the

*XMM-Newton* Medium Deep Survey (XMDS, 20–25 ks exposure depth, Chiappetti et al. 2005). In addition, the Subaru *XMM-Newton* Deep Survey (SXDS, Ueda et al. 2008), adjacent to the XMDS field, covers a 1.14 deg<sup>2</sup> area and reaches  $\approx 50$  ks exposure per pointing (Ueda et al. 2008). Moreover, the XMM-LSS field recently became a part of the 25 deg<sup>2</sup> XMM-XXL-North field (Pierre et al. 2016), which has similar *XMM-Newton* coverage as the original XMM-LSS survey.

In addition to the X-ray data, XMM-LSS is among the extragalactic fields with extensive multiwavelength coverage. In particular, the central  $\approx 5$  deg<sup>2</sup> area of the XMM-XXL-North field (i.e., the combination of XMDS and SXDS fields) was selected to be one of the *Spitzer* Extragalactic Representative Volume Survey fields. This sky region is covered uniformly by multiple photometric and spectroscopic surveys (see Sec. 5 for more details), and has also been selected to be one of the deep drilling fields of upcoming Dark Energy Survey and LSST. Compared to the multiwavelength

\* E-mail: ctchen@psu.edu



**Figure 1.** Locations of the *XMM-Newton* observations used in this work. The AO-15 observations are marked as the blue-filled circles with solid boundaries. The archival observations are marked as dashed circles. Circles with green, orange, blue, and cyan colors are for XMM-LSS, SXDS, XMDS, and XMM-XXL observations, respectively. We also plot the RA/DEC range of our catalog selection area as the solid rectangle, and the *Spitzer* SERVS footprint is marked as the red polygon. Note that our AO-15 observations do not cover the entirety of the source-extraction region, because the existing data from SXDS (bottom-right corner, orange circles) and from deep X-ray cluster follow-up observations (top-middle and top-right, the red circles) are already at the desired depth.

**Table 1.** The *XMM-Newton* data used to create the source catalog include 155 pointings with a total of 2.7 Ms of background-screened exposure time, of which 1.1 Ms is from the new AO-15 observations<sup>a</sup>. Columns from left to right: target field, *XMM-Newton* revolution, *XMM-Newton* ObsID, observation starting time, Right Ascension and Declination (J2000), cleaned exposure time for PN, MOS1, and MOS2. This table is available in its entirety online.

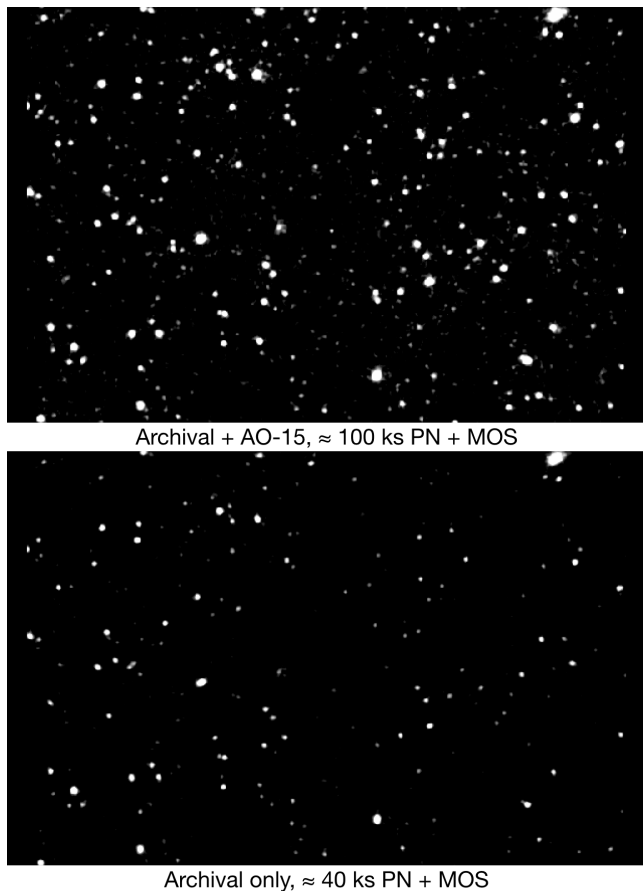
<sup>a</sup>: MOS only. For PN the total background-screened time is 2.3 Ms, of which 0.9 Ms is from the new AO-15 observations.

Field	Revolution	ObsID	Date (UT)	R.A.	Decl.	GTI (PN) (ks)	GTI (MOS1) (ks)	GTI (MOS2) (ks)
AO-15	3054	0780450101	2016-08-13T01:34:06	35.81072	-5.15989	20.91	23.61	23.61
XMM-LSS	1205	0404965101	2006-07-09T08:08:08	35.80953	-5.48532	3.44	10.36	9.91
XMDS	287	0111110401	2001-07-03T14:01:54	35.97582	-5.15253	21.40	27.20	27.40
SXDS	118	0112370101	2000-07-31T21:57:54	34.47819	-4.98115	39.13	42.70	42.83
XMM-XXL-N	2137	0677580101	2011-08-10T01:53:35	38.00217	-4.49993	4.94	5.93	5.52
XLSSJ022404.0-041328	0928	0210490101	2005-01-01T19:08:30	36.03267	-4.20230	80.28	87.98	87.98

data, archival *XMM-Newton* observations covering this sky region span a wide range of exposure time. In order to advance studies of accreting SMBHs and their environments, deep X-ray observations with similar areal coverage are integrally required in addition to the rich multiwavelength data in this field. To this end, we obtained *XMM-Newton* AO-15 observations taken between July 2016 to February 2017 with a total of 1.3 Ms exposure time. Our AO-15 data include 67 *XMM-Newton* observations. In addition to the new data, we made use of all the overlapping archival *XMM-Newton* observations to create a deep *XMM-Newton* catalog contiguous

ously covering most of the SERVS data in the XMM-LSS region. After excluding observations that were completely lost due to flaring background (see §2.2), the archival data used here include 51 observations culled from the 10 ks XMM-LSS survey, eighteen 20–25 ks observations from XMDS, 4 mosaic-mode observations<sup>1</sup> obtained as part of the XMM-XXL survey (Pierre et al. 2016), 4 archival *XMM-Newton* observations targeting galaxy clusters identified in the XMM-

<sup>1</sup> Each mosaic-mode observation is comprised of  $\approx$  ten 10 ks exposures.



**Figure 2.** *Top* – Background-subtracted, smoothed image at the 0.5–10 keV band for a  $0.8 \times 0.6$  deg<sup>2</sup> region centered at RA=35.58°, DEC=−4.965°. This image is created using both archival data and the new AO-15 data. *Bottom* – Same as the top image, but only the archival data are included. The typical vignetting-corrected exposure times are shown in the bottom of both panels. The exposure time of the full survey region is shown in Fig. 3.

XXL-North and XMM-LSS surveys ( $\approx 30 - 100$  ks), and the ten 50 ks observations from SXDS. We present the details of each observation in Table 1, and show the positions of each XMM observations used in this work in Fig 1.

Our AO-15 observations were separated into two epochs to minimize the effects of background flaring. We first observed the XMM-LSS sky-region in the SERVS footprint with  $\approx 1$  Ms *XMM-Newton* exposure time during July–August 2016. We then screened these first observations for flaring backgrounds (§2.2) and re-observed the background contaminated sky-regions using the remaining 0.3 Ms. Notably, we also observed the SXDS region in which one of the SXDS observations carried out in 2002 was severely affected by background flares. In this work, we present an X-ray source catalog obtained from a  $5.3$  deg<sup>2</sup> sky-region with  $34.2^\circ \leq \text{RA} \leq 37.125^\circ$  and  $-5.72^\circ \leq \text{Dec} \leq -3.87^\circ$ . The sky-region is primarily selected by the footprint of our AO-15 observations, with additional SXDS data within the SERVS footprint in the corner. A total of  $\approx 3$  Ms of *XMM-Newton* observations is used for generating the X-ray source catalog. In Fig. 2, we show the background-subtracted, 0.5–10 keV image (see §3 for the details of data analysis) from a  $\approx 0.5$

deg<sup>2</sup> region in XMM-LSS generated using the combined AO-15 and archival data. An image generated using only the archival data is also shown for comparison.

## 2.2 Data preparation and background flare filtering

We use the *XMM-Newton* Science Analysis System (SAS) 16.1.0 and HEASOFT 6.21 for our data analysis. The *XMM-Newton* Observation Data Files (ODFs) were processed with the SAS tasks EPICPROC (EPPROC and EMPROC for PN and MOS, respectively) to create MOS1, MOS2, PN, and PN out-of-time (OOT) event files for each ObsID. For observations taken in mosaic mode or with unexpected interruptions due to strong background flares, we use the SAS task EMOSAIC\_PREP to separate the event files into individual pseudo-exposures and assign pseudo exposure IDs.<sup>2</sup> For the mosaic mode observations, we also determine the RA/DEC of each pseudo-exposure using the AHFRA and AHFDEC values in the attitude files created using the SAS task ATTHKGEN.

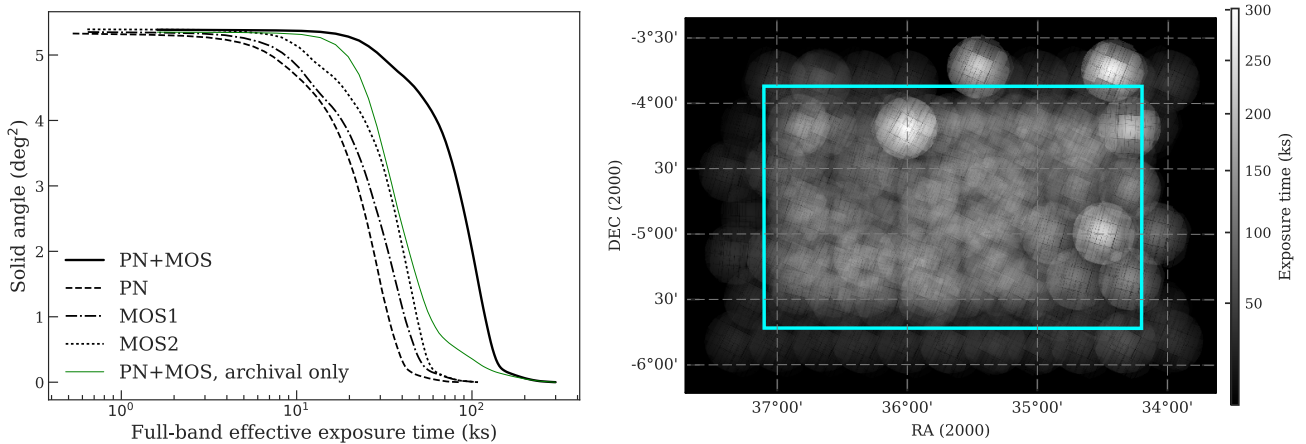
For each event file, we create single event light curves in time bins of 100 s for high (10–12 keV) and low (0.3–10 keV) energies using EVSELECT to search for time intervals without significant background flares (the “good time intervals”, GTIs). We first remove time intervals with 10–12 keV count rates exceeding  $3\sigma$  above the mean, and then repeat the  $3\sigma$  clipping procedure for the low-energy light curves. Since the background flares usually manifest as a high-count-rate tail in addition to the Gaussian-shape count-rate histogram, adopting the  $3\sigma$  clipping rule can effectively remove the high-count-rate tail while retaining useful scientific data. For a small number of event files with intense background flares, we instead filter the event files using the nominal count rate thresholds suggested by the *XMM-Newton* Science Operations Centre.<sup>3</sup> We exclude 12 pointings with GTI < 2 ks from our analysis. A total of 2.7 Ms of exposure remains after flare filtering, including 1.1 Ms from AO 15 and 1.6 Ms from the archival data. After screening for background flares, we further exclude events in energy ranges that overlap with the instrumental background lines (Al K $\alpha$  lines at 1.45–1.54 keV for MOS and PN; Cu lines at 7.2–7.6 keV and 7.8–8.2 keV for PN).

From the background-screened, instrumental-line-removed event files, we extract images with a  $4''$  pixel size using EVSELECT<sup>4</sup> in the following bands: 0.5–2 (soft) keV and 2–10 keV (hard), and 0.5–10 keV (full) bands. For each image, we generate exposure maps with and without

<sup>2</sup> For mosaic-mode observations, multiple exposures are stored in a single event file for each EPIC camera after ODF processing. EMOSAIC\_PREP separates the single event file into multiple pseudo event files with the same ObsID but different EXP\_ID and EXPIDSTR, which are informative FITS file keywords required by the SAS tasks. For observations with unexpected interruptions, the exposure is separated into multiple event files with the same ObsID and EXP\_ID by EPICPROC, but with different EXPIDSTR values. EMOSAIC\_PREP assigns a separate EXP\_ID for each event file, which is necessary for SAS source-detection tasks.

<sup>3</sup> <https://www.cosmos.esa.int/web/xmm-newton/sas-thread-epic-filterbackground>

<sup>4</sup> The  $4''$  pixel size is set by the parameter IMAGEBINSIZE= 80 in the SAS task EVSELECT.



**Figure 3.** *Left* – The black solid line shows the cumulative survey solid angle as a function of full-band effective (i.e., vignettted) exposure for observations used in this work. Distributions for individual instruments are also shown as the dashed line (PN), dash-dotted line (MOS1), and dotted line (MOS2). For comparison, the cumulative survey solid angle for the archival *XMM-Newton* data is also shown as the thin green line. *Right* – Full-band survey effective exposure map (PN + MOS). The 5.3 deg<sup>2</sup> survey region from which the X-ray source catalog is constructed is marked as the cyan rectangular box. Except for several regions with deep *XMM-Newton* follow-up observations of galaxy clusters, the *XMM-Newton* coverage in our survey region is generally homogeneous.

vignetting-corrections using the SAS task EEXPMAP. We set USEFASTPIXELIZATION=0 and ATTREBIN=0.5 in order to obtain more accurate exposure maps. We note that the exposure maps without vignetting-corrections are only used for generating background maps in §3. Detector masks were also generated using the SAS task EMASK. We show the distribution of exposure values across the *XMM-LSS* field and the PN+MOS exposure map in Fig. 3. The median PN exposure time of the central  $\approx 4.5$  deg<sup>2</sup> region covered by SERVS is 48.5 ks. For the full 5.3 deg<sup>2</sup> survey region, the median PN exposure time is 45.8 ks.

### 3 THE MAIN X-RAY SOURCE CATALOG

#### 3.1 First-pass source detection and astrometric correction

The astrometric accuracy of *XMM-Newton* observations can be affected by the pointing uncertainties of *XMM-Newton*. This uncertainty is usually smaller than a few arcsec, but can be as large as  $\approx 10''$  (e.g., Cappelluti et al. 2007; Watson et al. 2008; Rosen et al. 2016). To achieve better astrometric accuracy and to minimize any systematic offsets between different *XMM-Newton* observations, we run an initial pass of source detection for each observation and then use the first-pass source list to register the *XMM-Newton* observations onto a common WCS frame. The first-pass source detection methods are outlined below:

- (i) For the exposure taken by each instrument for each observation, we generate a temporary source list using the SAS task EWAVELET with a low likelihood threshold (THRESHOLD=4). EWAVELET is a wavelet-based algorithm that runs on the count-rate image generated using the image and exposure map extracted as described in §2.2.
- (ii) We use the temporary source list as an input to generate background images using the SAS task ESPLINEMAP with METHOD=MODEL. The METHOD=MODEL option of ES-

PLINEMAP fits the source-excised image with two templates: the vignettted exposure map, and the un-vignettted exposure map. The former represents the cosmic X-ray background with an astrophysical origin, while the latter represents the intrinsic instrumental noise. ESPLINEMAP then finds the best-fit linear combination of the two templates and generates the background map. The details of this method are described in Cappelluti et al. (2007). The background maps are used for the PSF-fitting based source detection task described in Step (iv).

- (iii) We run EWAVELET again for each observation. This time the source list is generated by running EWAVELET on the exposure map and image coadded across the PN, MOS1, and MOS2 exposures (when available) with the default likelihood threshold (THRESHOLD=5).
- (iv) For each EWAVELET source list, we use the SAS task EMLDETECT to re-assess the detection likelihood and determine the best-fit X-ray positions. EMLDETECT is a PSF-fitting tool, which performs maximum-likelihood fits to the input source considering the *XMM-Newton* PSF, exposure values, and background levels of the input source on each image. EMLDETECT also convolves the PSF with a  $\beta$ -model brightness profile for clusters and uses the result to determine if the input source is extended. We use a stringent likelihood threshold (LIKMIN= 10.8) to ensure that astrometric corrections are calculated based on real detections.
- (v) For the mosaic-mode observations, the multiple pointings under the same ObsID were already registered on the same WCS frame. Therefore, we do not correct the astrometry for each pseudo-exposure but only consider the astrometric offsets on an ObsID-by-ObsID basis. The source lists for the mosaic-mode observations were generated using the SAS task EMOSAIC-PROC, which is a mosaic-mode wrapper for procedures similar to (i)-(iv) described above.

For steps (iv) and (v), the source searching was conducted simultaneously on the images of the three EPIC-cameras as the astrometric offsets between PN, MOS1, and



MOS2 are negligible. For each ObsID, we cross-correlate the high-confidence EMLDETECT list of point sources (with the EMLDETECT flag EXT=0) with the optical source catalog culled from the Subaru Hyper Suprime Cam Survey Public Data Release 1 (HSC-PDR1; Aihara et al. 2017), which is an ultra-deep optical photometric catalog with sub-arcsec angular resolution. The astrometry of HSC-PDR1 is calibrated to the PAN-STARRS1  $3\pi$  survey and has a  $< 0.05''$  astrometric uncertainty. More details of the HSC-PDR1 catalog can be found in Aihara et al. (2017), and it is also briefly discussed in §5. For astrometric corrections, we limit the optical catalog to HSC sources with  $i = 18 - 23$  to minimize possible spurious matches and matches to bright stars that might have proper motions or parallaxes. The offset between each ObsID and the HSC catalog is calculated based on a maximum likelihood algorithm similar to the SAS task EPOSCORR. The major difference between our approach and EPOSCORR is that we use an iterative optimization approach compared to the grid-searching algorithm adopted by EPOSCORR. During each iteration, we cross-correlate the optical catalog with the X-ray catalog using a  $3''$  search radius and exclude all duplicated matches (typically only  $< 5\%$  of the X-ray sources have more than one optical counterpart in the bright HSC-PDR1 catalog). We then calculate the required corrections that maximize the cross-correlation likelihood. After each iteration, we apply the best-fit astrometric offsets onto the source list and next repeat the catalog cross-correlation steps and re-calculate the required additional corrections for the source list that already has updated coordinates from the previous iteration. The required corrections usually converges after 1–2 iterations. For the purpose of frame-correction, we adopt the X-ray positional uncertainties calculated based on the PSF-fitting likelihood ratios provided by EMLDETECT ( $\sigma_{eml}$  hereafter). The positional uncertainty information is necessary because the required corrections should be based more on X-ray sources with better positions within each observation. To avoid over-weighting sources with extremely small  $\sigma_{eml}$ , we also include an irreducible systematic uncertainty when finding the best-fit values for frame-correction.<sup>5</sup> The median number of X-ray sources in an ObsID with only one HSC counterpart within  $3''$  is 32.

The required frame-correction offsets calculated using our approach are less than  $3''$  in both RA and DEC and are generally consistent with the results calculated using EPOSCORR, with a median difference of  $0.1''$ . For several obsIDs the difference between our offsets and the EPOSCORR offsets are non-negligible ( $> 0.5''$ ). We visually inspect these ObsIDs and find that our approach does improve the alignments between the optical and corrected X-ray images. The event files and the attitude file for each ObsID are then projected onto the WCS frame of the HSC catalog by updating the relevant keywords using a modified version of ALIGN\_EVT (Ranalli et al. 2013). Since the sky coordinates for the event files of the mosaic-mode pseudo-pointings are derived based on the reference point centered at the nominal RA and DEC positions of the mosaic-mode ObsIDs, we also recalculate the sky coordinates for these event files with the SAS task ATTCALC using the true pointing positions as the reference

point, which is necessary for using regular SAS tasks for mosaic-mode pseudo-exposures.

### 3.2 Second-pass source detection

We re-create images, exposure maps, detector masks, background maps using the frame-corrected event files and attitude files. We then run source-detection tasks for the second time considering all XMM-Newton observations listed in Table 1. Similar to the approach used for the XMM-H-ATLAS survey (Ranalli et al. 2015), we divide the XMM-LSS field into a grid when running the second-pass source detection. We then use a custom-built wrapper of relevant SAS tasks to carry out the second-pass source detection, which is similar to the GRIDDETECT<sup>6</sup> tool built for the XMM-H-ATLAS survey (Ranalli et al. 2015).

The cell sizes are determined by the number of EWAVELET sources and the number of ObsIDs encompassed because the number of images that can be processed by a single EMLDETECT thread is limited. For each cell in the grid, we co-add the images and exposure maps for all observations with footprint inside the cell and run EWAVELET with a low detection threshold<sup>7</sup> on the co-added image and exposure map. For each cell, we only keep EWAVELET sources within the RA/DEC range of the cell (with a  $1'$  padding on each side of the cell). We then use the EWAVELET list as an input for EMLDETECT to assess the detection likelihood. Instead of running on the co-added image, EMLDETECT takes the image, exposure map, background map, and detector mask of each input observation into account. The EMLDETECT point-source list of the full XMM-LSS region is constructed from the union of the sources from all cells after removing duplicates due to the “padding”. We search for sources in three different bands: 0.5–2 keV (soft), 2–10 keV (hard), and 0.5–10 keV (full). We consider sources with EMLDETECT likelihood larger than  $DET\_ML = 6.0$  to be reliably detected. This corresponds to a false-detection probability of  $2.5 \times 10^{-3}$ . A total of 5218 sources satisfy this criterion in at least one of the three bands (see §3.4). For readers interested in an even more reliable catalog, we suggest using the  $DET\_ML = 10.8$  detection threshold (equivalent to a false-detection probability of  $2 \times 10^{-5}$ ) for the final X-ray catalog. A total of 4895 sources have  $DET\_ML \geq 10.8$  in at least one band. The detection thresholds are determined by extensive simulations that are discussed in §4. We show the spatial distribution of the 5218 detected sources in Fig. 4.

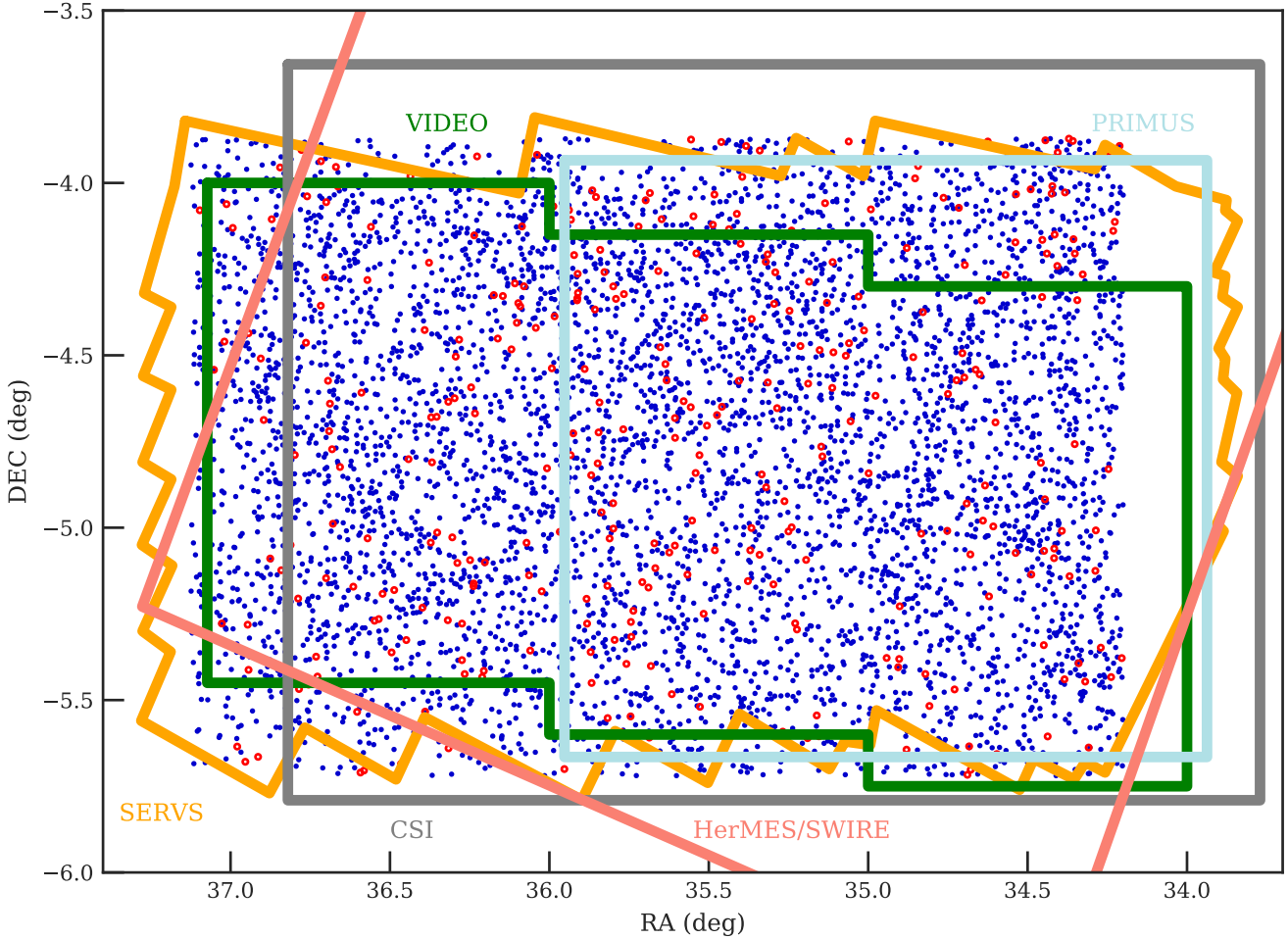
### 3.3 Astrometric accuracy

We investigate the positional accuracy of the XMM-Newton sources by comparing the second-pass X-ray catalog with the HSC-PDR1 catalog. Similar to the frame-correction procedures described in §3.1, we search for unique optical counterparts around the X-ray positions using a  $3''$  search radius. For the 5136 X-ray sources detected in the full-band during the second-pass source-searching process, a total of 2085 X-ray sources are found to have only one  $i = 18 - 23$  HSC counterpart within  $3''$ . We use the separations between the

<sup>5</sup> We assume the systematic uncertainties to be  $0.5''$  as suggested by Watson et al. (2008).

<sup>6</sup> <https://github.com/piero-ranalli/griddetect>.

<sup>7</sup> THRESHOLD=4.



**Figure 4.** Spatial distribution of the 5218 sources detected in this work. We have identified reliable multiwavelength counterparts (see Sec. 5.1 and Sec. 5.2 for details) for 93% of the *XMM-Newton* sources (blue dots), while the remaining 7% of sources are marked as open red circles. Some of the excellent multiwavelength coverage of the XMM-LSS field is also shown as labeled (see §5 for details)

optical and X-ray positions of this subsample as a means to determine empirical X-ray positional uncertainties, which is a commonly adopted practice in deep X-ray surveys (e.g. Luo et al. 2010; Xue et al. 2011; Luo et al. 2017).

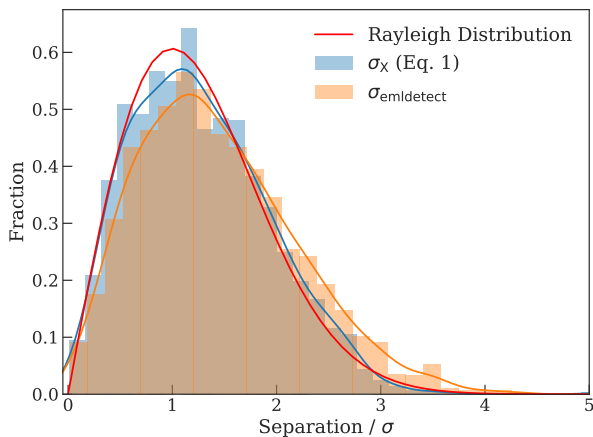
The X-ray positional accuracy is determined by how well the PSF-centroid location can be measured, which usually depends on the number of counts of the detected source and the PSF size of the instrument (i.e., off-axis angle). For the vast majority of the X-ray sources presented in this work, the detected photons come from at least three different observations, hence the dynamical range of effective off-axis angle for each sources detected on the coadded image is relatively small, and thus the X-ray positional uncertainty is mostly dependent on the number of counts available for detected sources. Using the angular separations between the 2085 X-ray sources and their unique optical counterparts, we derive an empirical relation between the number of X-ray counts,  $C$ ,<sup>8</sup> and the 68% positional uncertainty radius ( $r_{68\%}$ ) for the full-band-detected X-ray sources,

$\log_{10} r_{68\%} = -0.31 \times \log_{10} C + 0.85$ . The parameters are chosen such that 68% of the sources have positional offsets smaller than the empirical relation.

For this work, we define the X-ray positional uncertainty,  $\sigma_x$ , to be the same as the uncertainties in RA and DEC where  $\sigma_{RA} = \sigma_{DEC} = \sigma_x$ . Under this definition,  $\sigma_x$  is  $r_{68\%}$  divided by a factor of 1.515 (e.g., Pineau et al. 2017). Because the separations in both RA and DEC behave as a univariate normal distribution with  $\sigma_{RA}$  and  $\sigma_{DEC}$ , respectively,<sup>9</sup> the angular separation should therefore follow the joint probability distribution function of the uncertainties in the RA and DEC directions. Since we assume  $\sigma_{RA} = \sigma_{DEC}$ , the angular separation between an optical source and an X-ray source should follow the univariate Rayleigh distribution with the scaling parameter  $\sigma_x$ , where  $\sigma_x = \sigma_{RA} = \sigma_{DEC}$  (see §4 of Pineau et al. 2017, for details). Since  $\sigma_x$  is not the standard deviation of a 2D normal distribution, the 68% uncertainty of the angular separation is represented as  $r_{68\%} = 1.515\sigma_x$ . The factor 1.515 is determined by integrating the Rayleigh

<sup>8</sup> An upper limit of 2000 is set on  $C$  because the improvement of positional accuracy is not significant for larger source counts (e.g., Luo et al. 2017).

<sup>9</sup> Here we consider the positional uncertainties of the HSC-PDR1 catalog to be negligible compared to the *XMM-Newton* positional uncertainties.



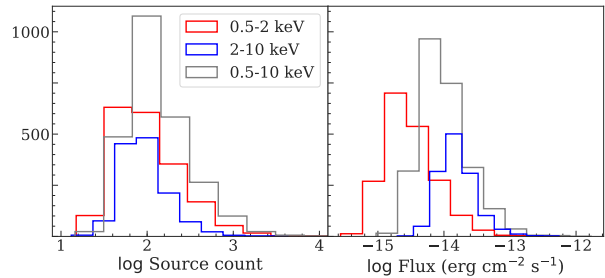
**Figure 5.** Histograms of the normalized positional offsets, a dimensionless quantity, and comparison with the expected Rayleigh distribution, the solid red curve. The distribution for the positional offsets normalized by the empirically derived positional uncertainty,  $\sigma_x$ , for the sources matched to bright optical counterparts is marked as the blue histogram along with the kernel density estimation. The distributions of the positional offsets normalized by the EMLDETECT-derived positional uncertainties,  $\sigma_{\text{eml}}$ , are also shown as the orange histogram.

distribution until the cumulative probability reaches 0.68. For reference, 90%, 95%, and 99.73% uncertainties correspond to  $2.146\sigma_x$ ,  $2.448\sigma_x$ ,  $3.439\sigma_x$ , respectively.

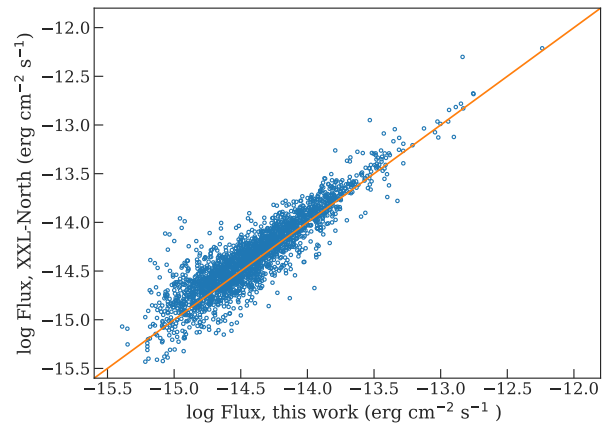
For each energy band, we repeat the same process to find the best-fit relation for  $\sigma_x$  using the following equation:

$$\log_{10} \sigma_x = \alpha \times \log_{10} C + \beta. \quad (1)$$

Due to the large PSF size of *XMM-Newton*, it is possible that the positional uncertainties were underestimated when deriving Eq. 1 using only the optical sources within the  $3''$  search radius. Therefore, we adopt an iterative process. For each iteration, we use the derived  $\sigma_x$  to find reliable matches using the likelihood-ratio matching method described in §5.1. We then re-derive Eq. 1 using the reliable matches, and the updated astrometric uncertainties are used for running likelihood-ratio matching again. This is a stable process, as the parameters converge after 2–3 iterations. The average positional uncertainties ( $\sigma_x$ ) for our soft-band, hard-band, and full-band X-ray catalogs are  $1.35''$ ,  $1.37''$ , and  $1.31''$ , respectively. In Fig. 5, we show a comparison of the normalized separation ( $\text{Separation}/\sigma$ ) between the full-band X-ray sources and their bright optical counterparts with  $\sigma$  derived using Eq. 1,  $\sigma_x$ , and the  $\sigma$  calculated by EMLDETECT,  $\sigma_{\text{eml}}$ . The agreement between the Rayleigh distribution and the  $\text{Separation}/\sigma_x$  distribution of our sample demonstrates that our empirically derived  $\sigma_x$  are reliable indicators of the true positional uncertainties. As for  $\sigma_{\text{eml}}$ , previous studies have identified that an irreducible systematic uncertainty should be added to  $\sigma_{\text{eml}}$  for the normalized separation to follow a Rayleigh distribution (e.g., Watson et al. 2008), but the nature of this systematic uncertainty remains unclear. We use  $\sigma_x$  as the positional uncertainties of our X-ray catalog, but  $\sigma_{\text{eml}}$  is also included in the final catalog for completeness.



**Figure 6.** Left – Source count distributions for the sources detected in the soft (red), hard (blue), and full (gray) bands. Right – Flux distributions of the sources detected in the three bands. Colors are the same as the left panel.



**Figure 7.** Comparison of the soft-band X-ray fluxes of our X-ray sources and those of the *XMM-Newton* counterparts identified in the XXL-North source catalogs (Liu et al. 2016) within a  $10''$  radius.

### 3.4 The main X-ray source catalog

We detect 5136, 3878, 2655 point sources with  $\text{DET\_ML} \geq 6.0$  in the 0.5–2 keV, 2–10 keV, and 0.5–10 keV bands, respectively. We report the details of the main X-ray source catalog in Table A of Appendix A. The extended sources (identified by the  $\text{EXT} > 0$  flag of EMLDETECT) are not included as the properties of the extended X-ray emission are beyond the scope of this work.<sup>10</sup> We associate catalogs from the three energy bands using a similar approach to that adopted by the *XMM-Newton* Serendipitous Source Catalogue. We consider two sources from different catalogs to be the same if their angular separation is smaller than any of the following quantities: (1)  $10''$  (2) distance to the nearest-neighbor in each catalog (3) quadratic sum of the 99.73% positional uncertainties from both bands. The final source catalog is the union of the sources detected in the three energy bands. We check for potential duplicated sources by visually inspecting all sources with distance to the nearest-neighbor ( $\text{DIST\_NN}$ ) less than  $10''$ , and only one set of sources is found to be duplicated, resulting in a total of 5218 unique sources. There are

<sup>10</sup> There are 68, 11, and 77 sources identified as  $\text{EXT} > 0$  by EMLDETECT in the 0.5–2 keV, 2–10 keV, and 0.5–10 keV bands, respectively.

2928 sources with more than 100 PN+MOS counts in full-band, and 126 sources with more than 1000 X-ray counts. A unique X-ray source ID is assigned to each of the 5218 sources at this stage.

We also derive the count-rate to flux energy conversion factors (ECFs) assuming a power-law spectrum with photon index  $\Gamma = 1.7$ , which is typical for X-ray AGNs (e.g., [Ranalli et al. 2015](#)), and Galactic absorption,  $N_H = 3.57 \times 10^{20} \text{ cm}^{-2}$ .<sup>11</sup> The conversion factors for PN and MOS are (7.06, 1.99), (1.24, 0.44), and (3.17, 0.96) counts  $\text{s}^{-1}/10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1}$ , in the 0.5–2 keV, 2–10 keV, and 0.5–10 keV bands, respectively. For each source detected by EMLDETECT, the flux from each EPIC camera is calculated separately using the corresponding ECF. The final flux of the source is the error-weighted mean of the fluxes from the three EPIC cameras, when available. The median fluxes for the soft, hard, and full bands are  $2.7 \times 10^{-15}$ ,  $1.5 \times 10^{-14}$ , and  $8.8 \times 10^{-15} \text{ erg cm}^{-2} \text{ s}^{-1}$ , respectively. We show the source count and flux distributions of the sources detected in the three energy bands in Fig. 6.

For sources that are detected in less than three bands, we calculate the source-count upper limits using the mosaicked background map of the band in which the source is not detected. The mosaicked background map of each band is generated by summing the background maps from all individual observations. According to the Poisson probability set by the EMLDETECT detection likelihood threshold ( $P_{\text{Random}}$ , the probability of the detected source being a random Poisson fluctuation due the background), we can calculate the minimum required total counts ( $m$  in the following equation) to exceed the expected number of background counts,  $B$ , using the regularized upper incomplete  $\Gamma$  function (which is equivalent to Eq. 2 of [Civano et al. \(2016\)](#) if  $m \in \mathbb{N}$ ) :

$$P_{\text{Random}} = \frac{1}{\Gamma(m)} \int_B^\infty t^{m-1} e^{-t} dt \quad (2)$$

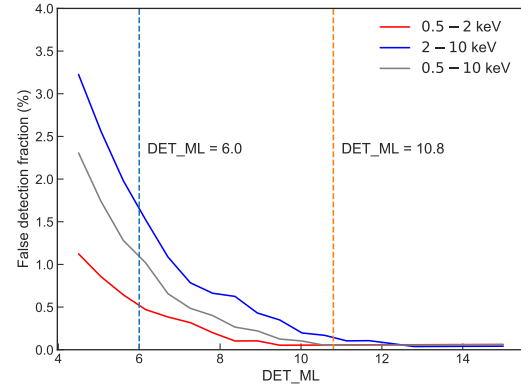
Here  $\Gamma(m)$  is the  $\Gamma$  function. The upper limits are set at  $P_{\text{Random}} = 2.5 \times 10^{-3}$  (consistent with  $\text{DET\_ML} = 6$ ). For each non-detected source in each band, we determine the background counts by summing up the background map within the circle with 70% encircled energy fraction (EEF). We then calculate  $m$  using the SCIPY function `SCIPY.SPECIAL.GAMMAINCCINV`.<sup>12</sup> Since  $m$  is the required total counts to exceed random background fluctuations at the given probability, the flux upper limit is calculated based on the following equation similar to [Cappelluti et al. \(2009\)](#); [Civano et al. \(2016\)](#):

$$S = \frac{m - B}{\exp \times \text{EEF} \times \text{ECF}} \quad (3)$$

Here  $\text{EEF} = 0.7$ , and  $\exp$  is the median exposure time within the 70% EEF circle. The upper limits of the source count and source count rate are also corrected for PSF loss. The flux upper limits are calculated as the exposure-time-weighted mean of the three EPIC detectors.

<sup>11</sup> Derived using the `COLDEN` task included in the CIAO software package. The Galactic column density is calculated at the center of the source-detection region at  $\text{RA} = 35.6625^\circ$ ,  $\text{DEC} = -4.795^\circ$ .

<sup>12</sup> This is the inverse function of Eq. 2.



**Figure 8.** The fraction of spurious sources detected at different  $\text{DET\_ML}$  based on simulations. The detection thresholds relevant to our catalog are marked as the dashed lines.

For each source detected in either the soft or the hard band (or both), we calculate its hardness ratios (HR) defined as  $(H - S)/(H + S)$ , where  $H$  and  $S$  are the source counts weighted by the effective exposure times in the hard and the soft bands, respectively. Note that the source counts here are the default output of EMLDETECT, which is the sum of the counts from all three EPIC detectors. We report this value in our catalog for direct comparison with previous studies. The uncertainties on HR are calculated based on the count uncertainties from the output of EMLDETECT using the error propagation method described in §1.7.3 of Lyons (1991).

As a comparison, we also match our X-ray catalog with the XXL-North catalog ([Liu et al. 2016](#)) by searching for counterparts within a  $10''$  radius. A total of 2843 X-ray sources from XXL-North are found to have a counterpart within the  $10''$  radius in our X-ray catalog. For these matched sources, we show the comparison between the soft-band X-ray fluxes reported in the XXL-North catalog and those in our catalog in Fig. 7. As expected, the majority of the archival sources detected in our catalog have consistent soft-band fluxes. Since the SXDS observations were also used for constructing the XXL-North ([Liu et al. 2016](#)) catalog, the 2843 sources matched to the XXL-North catalogs are considered as the same sources detected in ([Liu et al. 2016](#)), and we include the IDs from the original catalog for these matched objects in our source catalog (Table A). In our source extraction region, 208 sources from the original ([Liu et al. 2016](#)) catalog do not have a counterpart in our point-source catalog. Of these 208 sources, 50 of them can be matched to the extended sources excluded from our point source catalog with a  $1'$  matching radius. Another 110 of these 208 sources can be matched to the sources with  $\text{DET\_ML} < 6.0$  that were excluded from our final catalog described in §3.2. The remaining sources comprise  $\approx 1.6\%$  of the XXL-North catalog in our source extraction region. Visual inspections suggest that the vast majority of these sources might be spurious detections, but we cannot rule out the possibility that some sources are missed in our catalog due to their X-ray variabilities (e.g., [Yang et al. 2016](#)). The properties of sources that exhibit strong X-ray variabilities will be presented in a separate work.



**Table 2.** Sensitivity curves. Column 1: Soft-band flux. Column 2: Soft-band survey solid angle (DET\_ML = 6). Column 3: Soft-band survey solid angle (DET\_ML = 10.8). Columns 4–6: Similar to Columns 1–3 but for the hard band Columns 6–9: Similar to Columns 1–3 but for the full band. This table is available in its entirety online.

$\log S_{0.5-2\text{keV}}$ (erg cm <sup>-2</sup> s <sup>-1</sup> )	$\Omega_{0.5-2\text{keV}}^6$ (deg <sup>2</sup> )	$\Omega_{0.5-2\text{keV}}^{10.8}$ (deg <sup>2</sup> )	$\log S_{2-10\text{keV}}$ (erg cm <sup>-2</sup> s <sup>-1</sup> )	$\Omega_{2-10\text{keV}}^6$ (deg <sup>2</sup> )	$\Omega_{2-10\text{keV}}^{10.8}$ (deg <sup>2</sup> )	$\log S_{0.5-10\text{keV}}$ (erg cm <sup>-2</sup> s <sup>-1</sup> )	$\Omega_{0.5-10\text{keV}}^6$ (deg <sup>2</sup> )	$\Omega_{0.5-10\text{keV}}^{10.8}$ (deg <sup>2</sup> )
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
-15.40	1.855	0.044	-14.53	0.826	0.038	-14.81	0.782	0.038
-15.38	2.129	0.055	-14.51	1.040	0.046	-14.80	0.999	0.046
-15.37	2.391	0.069	-14.50	1.277	0.054	-14.78	1.243	0.055
-15.35	2.633	0.082	-14.48	1.524	0.063	-14.77	1.491	0.065
-15.34	2.856	0.098	-14.47	1.790	0.074	-14.75	1.763	0.078

## 4 SURVEY SENSITIVITY AND $\log N - \log S$

### 4.1 Monte Carlo simulation

To assess our survey sensitivity, we perform Monte Carlo simulations of mock X-ray observations. For each simulation, we generate a list of mock X-ray sources by sampling from the  $\log N - \log S$  relations reported in the XMM-COSMOS survey (Cappelluti et al. 2009, for the 0.5–2 keV and 2–10 keV bands) and the Chandra Multiwavelength Project survey (ChaMP, Kim et al. 2007, for the 0.5–10 keV band). The maximum flux of the mock X-ray catalogs is set at  $10^{-11}$  cm<sup>-2</sup> s<sup>-1</sup>. The minimum flux of the mock X-ray sources at each energy band is set as 0.5 dex lower than the minimum detected flux (e.g. LaMassa et al. 2016). We randomly place the mock X-ray sources in the RA/DEC range covered by the XMM-Newton observations used in this work. We then use a modified version of the simulator written for the XMM-Newton survey of the CDF-S (Ranalli et al. 2013), CDFS-SIM,<sup>13</sup> to create mock event files. CDFS-SIM converts X-ray fluxes to PN and MOS count rates with the same model used for deriving the ECFs, and it then randomly places X-ray events around the source location according to the count rates, the XMM-Newton PSFs at the given off-axis angle, and the real exposure maps. We extract images from the simulated event files using the same methods described in §3. For each observation, the simulated image is then combined with a simulated background, which is created by re-sampling the original background map according to Poisson distributions to create simulated images that mimic the real observations. For each energy band, a total of 20 simulations are created.

We run the same two-stage source-detection procedures described in §3 on the simulated data products. For each simulation, we match the detected sources to the input sources within a  $10''$  search radius by minimizing the quantity  $R^2$ :

$$R^2 = \left(\frac{\Delta\text{RA}}{\sigma_{\text{RA}}}\right)^2 + \left(\frac{\Delta\text{DEC}}{\sigma_{\text{DEC}}}\right)^2 + \left(\frac{\Delta\text{RATE}}{\sigma_{\text{RATE}}}\right)^2 \quad (4)$$

Here  $\Delta\text{RA}$  and  $\Delta\text{DEC}$  are the difference between the simulated RA/DEC positions and the RA/DEC positions obtained by running source detection on the simulated images.  $\Delta\text{RATE}$  is the difference between the simulated count rates and the detected count rates.  $\sigma_{\text{RA}}$ ,  $\sigma_{\text{DEC}}$ , and  $\sigma_{\text{RATE}}$  are the uncertainties of RA, DEC and count rates of the detected

sources. Minimizing  $R^2$  takes into account the flux and positional differences between the input catalog and the sources detected in the simulated images (e.g., Cappelluti et al. 2007; Ranalli et al. 2015). For detected sources without any input sources within the  $10''$  radius, we consider them to be spurious detections. In Fig. 3.4, we show the spurious fraction as a function of DET\_ML for the soft, hard, and full bands. For our catalog, we consider sources with DET\_ML > 6 to be reliably detected. At this threshold, the spurious fractions are 0.73%, 2.01%, and 1.68% for the soft, hard, and full bands, respectively. As for the threshold at DET\_ML = 10.8, the spurious fractions become 0.08%, 0.34%, and 0.10%, respectively. The low spurious fraction in the soft-band is likely due to its lower background levels. For the full X-ray source catalog of 5218 sources, these spurious fractions translate to  $\approx 41$  spurious detections for DET\_ML > 6, and 12 spurious detections for DET\_ML > 10.8.

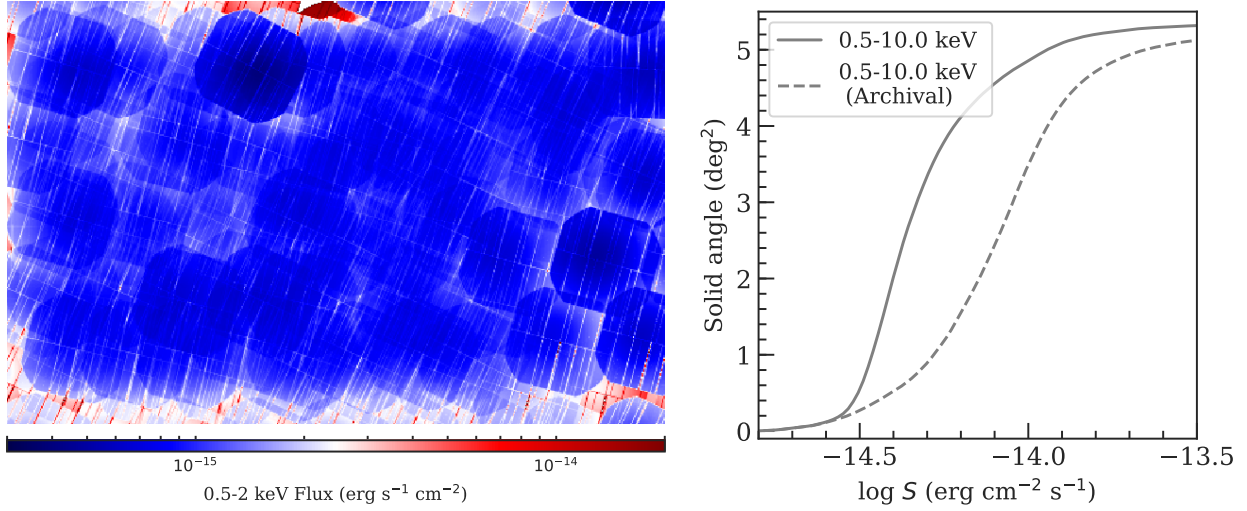
### 4.2 Survey sensitivity, sky coverage, and $\log N - \log S$

We create sensitivity maps of our survey region in different bands using the background maps and exposure maps generated as described in §2.2. The mosaicked background and exposure maps are binned to  $5 \times 5$  pixels ( $20'' \times 20''$ ). For each pixel on the binned, mosaicked background map, the minimum required source counts to exceed the random background fluctuations are calculated using Eq. 2. The sensitivity is then calculated using Eq. 3 with the corresponding EEF and ECF values. We show the soft-band sensitivity map in Fig. 9-left. We also generated a soft-band sensitivity map using only the archival data. To visualize the improvement upon the archival data, we compare the full-band sky coverage obtained from all available XMM-Newton data in our survey region with the sky coverage obtained using only the archival data. We show the comparison in Fig. 9-right, demonstrating the improved survey homogeneity with the new XMM-Newton observations. The sky coverage corresponding to the DET\_ML  $\geq 6.0$  catalog in our survey region in the soft, hard, and full-bands are shown in Fig. 10. The sensitivities curves for both the DET\_ML  $\geq 6.0$  and DET\_ML  $\geq 10.8$  catalogs are presented in Table 2.

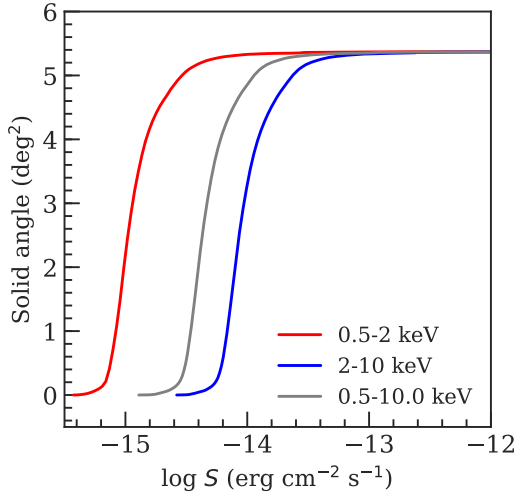
We calculate the  $\log N - \log S$  relations of our survey using the sky coverage curves described above and the following equation:

$$N(> S) = \sum_{i=1}^{N_s} \frac{1}{\Omega_i} \quad (\text{deg}^{-2}). \quad (5)$$

<sup>13</sup> <https://github.com/piero-ranalli/cdfs-sim>



**Figure 9.** *Left* – Soft-band sensitivity map of the source-extraction region (the same as the cyan box shown in Fig. 3). *Right* – Comparison of the full-band sky coverages between this work (solid line) and the archival *XMM-Newton* observations (dashed line), demonstrating the more homogeneous sensitivity across the wide field enabled by the new data.



**Figure 10.** Sky coverage in the soft, hard, and full bands of our X-ray survey in XMM-LSS.

Here  $N(> S)$  represents the total number of detected sources with fluxes larger than  $S$ , and  $\Omega_i$  is the sky coverage associated with the flux of the  $i$ th source. The  $\log N - \log S$  relations of our survey are shown in Fig. 11, along with a selection of the relations from the literature. The flux differences caused by different choices of power-law indices and/or the slight difference in energy range have been corrected assuming the  $\Gamma = 1.7$  power-law spectrum used in this work. Considering factors such as different spectral models and/or methods of generating survey sensitivity curves, our  $\log N - \log S$  relations are consistent with the relations reported in the literature.

## 5 MULTIWAVELENGTH DATA IN THE XMM-LSS FIELD

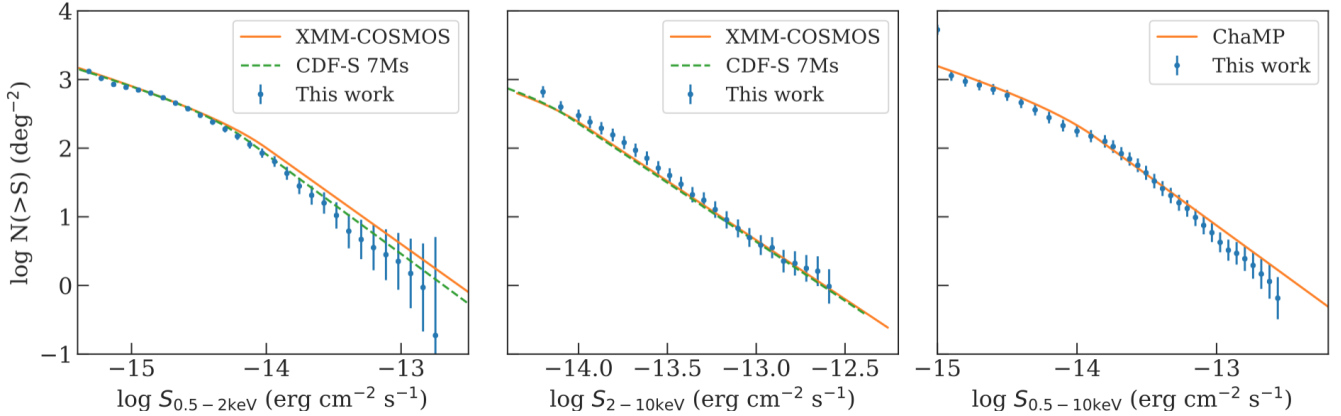
The XMM-LSS region is one of the most extensively observed extragalactic fields. The publicly available multiwavelength observations in the XMM-LSS region utilized in this work are the *Spitzer* Extragalactic Representative Volume Survey (Mauduit et al. 2012), the *Spitzer* SWIRE survey (Lonsdale et al. 2003), the VISTA Deep Extragalactic Observations (VIDEO) survey (Jarvis et al. 2013), the CFHTLS-wide survey (Hudelot et al. 2012), and the Subaru HSC Survey Public Data Release 1 (HSC-PDR1, Aihara et al. 2017).

We focus on finding the correct counterparts for our X-ray sources in four deep optical-to-near-IR (OIR) catalogs, SERVS, VIDEO, CFHTLS, and HSC-PDR1. SERVS is a post-cryogenic *Spitzer* IRAC survey in the near-IR 3.6 and 4.5  $\mu\text{m}$  bands with  $\approx 2\mu\text{Jy}$  survey sensitivity limits and  $\approx 5\text{ deg}^2$  solid angle coverage in the XMM-LSS region. We make use of the highly reliable two-band SERVS catalog built using SEXTRACTOR, obtained from the *Spitzer* Data Fusion catalog,<sup>14</sup> which has  $\approx 4 \times 10^5$  sources. The *Spitzer* Data Fusion catalog already integrated data from the *Spitzer* Wide-area Infrared Extragalactic survey (SWIRE, Lonsdale et al. 2003), which include photometry in all four IRAC bands and the photometry in MIPS 24, 70, and 160  $\mu\text{m}$ . 87% of the X-ray sources have at least one SERVS counterpart within their 99.73% positional-uncertainty radius ( $r_{99\%}$  hereafter), which is calculated based on the quadratic sum of the 99.73% X-ray positional uncertainties and the corresponding OIR positional uncertainties.

VIDEO is a deep survey in the near-infrared  $Z$ ,  $Y$ ,  $J$ ,  $H$  and  $K_s$  bands with  $\approx 80\%$  completeness at  $K_s < 23.8$  and a total of  $\approx 5.7 \times 10^5$  sources. 84% of the X-ray sources have at least one VIDEO counterpart within  $r_{99\%}$ .

The CFHTLS-W1 survey covers the entirety of our X-ray data, with an 80% completeness limit of  $i' = 24.8$ .

<sup>14</sup> <http://www.mattia Vaccari.net/df/>.

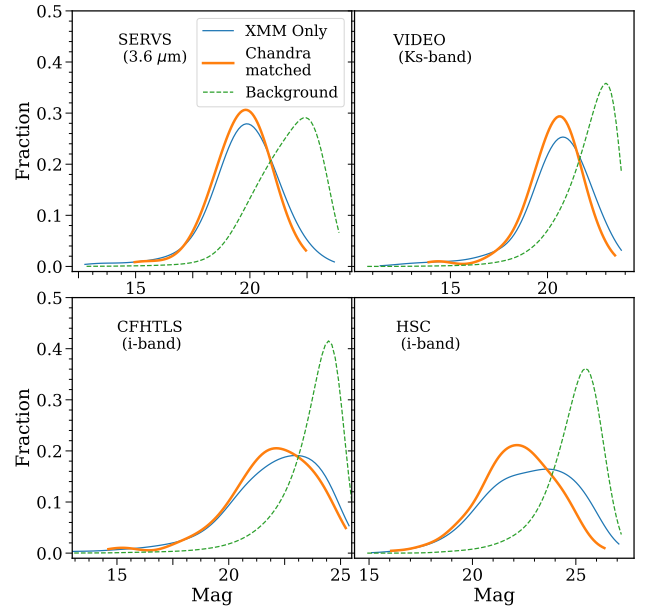


**Figure 11.** The  $\log N - \log S$  relations for our catalog in the soft-band (left), hard-band (middle), and full-band (right). The error bars represent the  $1\sigma$  uncertainties obtained using Eq. 4 of Cappelluti et al. (2009). For comparison, we also show some  $\log N - \log S$  relations from the literature with appropriate energy range and power-law index corrections. The  $\log N - \log S$  relations of our survey are generally consistent with previous studies.

We select the CFHTLS sources in the RA/DEC ranges marginally larger ( $1'$ ) than our source detection region. We limit the CFHTLS sources to those with  $SNR > 5$  in the  $i'$ -band. The total number of sources in the  $i'$ -band selected catalog is  $\approx 8.1 \times 10^5$ . 96% of the X-ray sources in our catalog have at least one CFHTLS counterpart within  $r_{99\%}$ .

The XMM-LSS field is entirely encompassed by the 108  $\text{deg}^2$  HSC-PDR1-wide survey. Inside the XMM-LSS field, HSC-PDR1 also has “ultra-deep” ( $\approx 1.77 \text{ deg}^2$ ) and “deep” ( $\approx 5 \text{ deg}^2$ ) surveys, which overlap with the SXDS and XMDS regions, respectively. The limiting magnitudes in the  $i$ -band for the ultra-deep, deep, and wide HSC-PDR1 surveys are 27.2, 26.5, and 26.4, respectively. We select the  $i$ -band detected HSC-PDR1 sources in the RA/DEC ranges slightly larger than our source detection region. The HSC-PDR1 sources used for our source-matching are the union of the ultra-deep, deep, and wide surveys.<sup>15</sup> The HSC catalogs from the three different depth layers were merged according to their OBJECT\_ID. To avoid duplications, sources from the HSC wide survey within  $0.5''$  of any sources from either the deep or ultra-deep survey are discarded. Similarly, sources from the deep survey with distance less than  $0.5''$  from any ultra-deep sources are also discarded. The total number of HSC-PDR1 sources in our source detection region is  $\approx 3.2 \times 10^6$ , and  $\approx 96\%$  of the X-ray sources in our main catalog have at least one HSC-PDR1 counterpart within  $r_{99\%}$ .

Although CFHTLS is not as deep as HSC-PDR1 at  $g$ ,  $r$ ,  $i$ ,  $z$  bands, it has the complementary  $u^*$ -band photometry. Including photometry from both optical surveys also ensures that we will not miss an optical counterpart due to bad photometry caused by artifacts such as satellite tracks in either surveys.



**Figure 12.** Kernel density estimations of the magnitude distributions (solid lines) for the expected counterparts in SERVS (top-left), VIDEO (top-right), CFHTLS (bottom-left), and HSC-PDR1 (bottom-right). We show the distributions obtained using the full XMM-Newton catalog ( $q(m)_{\text{XMM-Newton}}$ ), and the distributions obtained using the Chandra sources in the XMM-LSS field ( $q(m)_{\text{Chandra}}$ ). The magnitude distributions of the background, unrelated sources are also shown in each panel as the dashed curves. This figure demonstrates that  $q(m)_{\text{Chandra}}$  significantly improves upon the background-dominated  $q(m)_{\text{XMM-Newton}}$  for the deep OIR catalogs in the bottom panels.

<sup>15</sup> We select sources with the DETECT\_IS\_PRIMARY and IDETECTED\_NOTJUNK flags set as TRUE, and CENTROID\_SDSS\_FLAGS set as FALSE. According to the HSC-PDR1 example script for selecting “clean objects”, we also exclude the HSC sources with FLAGS\_PIXEL\_EDGE, FLAGS\_PIXEL\_SATURATED\_CENTER, FLAGS\_PIXEL\_CR\_CENTER, FLAGS\_PIXEL\_BAD flags in the  $i$ -band to avoid unreliable  $i$ -band sources.

### 5.1 The maximum likelihood estimation method

To match reliably the X-ray sources to the OIR catalogs with much higher source densities, we employ the maximum likelihood estimation method (MLE hereafter) similar to previous deep X-ray surveys, (e.g., Brusa et al. 2007; Luo et al. 2010, 2017). The likelihood ratio ( $LR$  hereafter) is defined as

the ratio between the probability that the source is the correct counterpart, and the probability that the source is an unrelated background object (Sutherland & Saunders 1992):

$$LR = \frac{q(m)f(r)}{n(m)}. \quad (6)$$

Here  $q(m)$  is the magnitude distribution of the expected counterparts in each OIR catalog,  $f(r)$  is the probability distribution function of the angular separation between X-ray and OIR sources, and  $n(m)$  is the magnitude distribution of the background sources in each OIR catalog.

We calculate the background source magnitude distributions using OIR sources between  $10''$  and  $50''$  from any sources in our X-ray catalog.

As discussed in §3.3, the probability distribution function of the angular separation should follow the Rayleigh distribution:

$$f(r) = \frac{r}{\sigma_x^2} \exp\left(-\frac{r^2}{2\sigma_x^2}\right). \quad (7)$$

Note that Eq. 7 is different from the two-dimensional Gaussian distribution function that maximizes at  $r = 0$ , and thus the  $LR$  values calculated in this work are not directly comparable to previous works that adopted a Gaussian  $f(r)$ .

Due to the larger positional uncertainties of *XMM-Newton* and the high source densities of the OIR catalogs, deriving an accurate magnitude distribution of the expected counterparts,  $q(m)$ , using *XMM-Newton* data is challenging. Therefore, we obtain  $q(m)$  for our X-ray sources by first matching our *XMM-Newton* catalog to the *Chandra* Source Catalog 2.0 (CSC 2.0; Evans et al. 2010) to take advantage from the higher angular resolution and positional accuracy of *Chandra*. We derive the positional uncertainties of the *Chandra* sources in our survey region using the same empirical approach described in Xue et al. (2011) by selecting CSC sources in the RA/DEC range of our catalog, and matching them onto HSC-PDR1 using a  $1.5''$  radius. We select CSC sources that are uniquely matched to our X-ray catalogs within the 95% uncertainties (*Chandra* and *XMM-Newton* positional uncertainties are added in quadrature). A total of 223 sources in our *XMM-Newton* catalog are matched to a unique *Chandra* source in CSC. We match these *Chandra* sources to the four OIR catalogs using Eq. 6, with  $q(m)$  derived using the iterative approach described in Luo et al. (2010). The  $q(m)$  derived from the CSC sources,  $q(m)_{Chandra}$ , is then used as the expected magnitude distribution for OIR counterparts of our *XMM-Newton* sources. We note that the X-ray flux distributions in the soft, hard, and full bands of the *Chandra*-matched subsample are very similar to those of our entire *XMM-Newton* catalog, and therefore  $q(m)_{Chandra}$  should be consistent with the intrinsic magnitude distributions of the real OIR counterparts of our full X-ray catalog. The counterpart matching processes are run on four different OIR catalogs: SERVS, VIDEO, CFHTLS, and HSC-PDR1. For illustration, we show the magnitude distributions of the background sources and the distributions of the expected counterparts derived using CSC sources in Fig. 12.

For comparison, we also obtain  $q(m)$  for the full *XMM-Newton* catalog without using the *Chandra* positions,  $q(m)_{XMM-Newton}$ . We again use the Luo et al. (2010) iterative method, but with a  $3''$  initial search radius.

$q(m)_{XMM-Newton}$  is also plotted on Fig. 12. It is evident that for ultra-deep OIR catalogs such as HSC-PDR1 and CFHTLS,  $q(m)_{XMM-Newton}$  is skewed toward the faint background sources compared to the *Chandra*-matched subsample. For the other catalogs, we find no qualitative difference between  $q(m)_{Chandra}$  and  $q(m)_{XMM-Newton}$ , but we still use  $q(m)_{Chandra}$  for consistency.

We next compute the  $LR$  values for all OIR sources within a  $10''$  radius of X-ray sources using Eq. 6. For each OIR catalog, we choose the  $LR$  thresholds ( $LR_{th}$ ) such that the reliability and completeness parameters are maximized (see Eq. 5 of Luo et al. 2010 for details). Counterparts with  $LR > LR_{th}$  are considered to be reliably matched. The results are reported in Table 2. For each OIR catalog, we show the number of all X-ray sources with at least one OIR counterpart within the 99.73% positional uncertainties of the X-ray sources,  $N_{All}$ , and the number of X-ray sources with at least one reliably matched source with  $LR > LR_{th}$ ,  $N_{Reliable}$ . For an X-ray source with only one OIR counterpart within the search radius but where its matching  $LR$  is smaller than  $LR_{th}$ , we consider the counterpart to be an acceptable match if the separation from the OIR position to the X-ray position is less than  $r_{99\%}$ . The total number of such cases for each OIR catalog is listed as  $N_{Acceptable}$  in Table 2.

Motivated by the spurious matching rates (see §5.2 for the cross-matching reliability analysis), we first select a “primary” counterpart for each X-ray source from, in priority order of, SERVS, VIDEO, CFHTLS, and HSC-PDR1. After selecting the primary OIR counterpart, we then associate different OIR catalogs with each other using a simple nearest-neighbor algorithm. Thanks to the much smaller positional uncertainties of the OIR catalogs, we adopt a homogeneous search radius of  $1''$  for the OIR catalog associations, which is the approach used by the *Spitzer* Data Fusion database (Vaccari 2016).

As a result, 4854 ( $\approx 93\%$ ) X-ray sources have at least one reliable match. Of these sources, 3898 are matched to SERVS, 371 are matched to VIDEO, 382 are from CFHTLS, and 174 are from HSC. There are also 959 sources with multiple counterparts having  $LR > LR_{th}$  in various OIR catalogs. For these sources, we select a “secondary” counterpart based on the following priority order: (1) 62 second-best matches from SERVS (2) 29 second-best matches from VIDEO (3) 458 best matches from CFHTLS (4) 255 best matches from HSC (5) 96 second-best matches from CFHTLS (6) 100 second-best matches from HSC. Finally, there are 22 X-ray sources with three reliable counterparts; all of the tertiary counterparts are from CFHTLS (6) and HSC (16), which are also reported in our final catalog.

For the remaining X-ray sources without any of the counterparts with  $LR > LR_{th}$ , 61 ( $\approx 1\%$ ) of them are considered to have “acceptable” matches, where 48, 7, and 6 of them are matched to SERVS, VIDEO, and CFHTLS, respectively. For the remaining sources, the vast majority of them still have at least one OIR counterpart within the relevant 99.73% positional uncertainty circle. We note that the vast majority of these sources still have  $DET\_ML \geq 10.8$  in at least one band. We find only 24 sources are completely isolated. Visual inspection of these sources shows that most of the 24 sources coincide with a bright star, thus making the pipeline OIR photometry unreliable. We plot the positional



**Table 3.** MLE counterpart matching results. Column 1: Catalog name. Column 2: Survey magnitude limit for each catalog in AB. Column 3: Survey area. Column 4: Positional uncertainty for each OIR catalog. Column 5:  $LR$  threshold. Column 6: Total number of X-ray sources with at least one counterpart within the 99.73% positional uncertainty radius in each catalog. Column 7: Average number of OIR sources within the 99.73% positional uncertainty radius of the X-ray sources. Column 8: Total number of X-ray sources with at least one counterpart with  $LR > LR_{th}$ . Column 9: Total number of X-ray sources without any  $LR > LR_{th}$  counterparts but having only one OIR source within the 99.73% positional uncertainty radius. Columns 10–12: See §5.2 for details. Column 10: The fraction of X-ray sources in the “associated population” based on the results of Monte Carlo simulations. Column 11: False matching rates determined using Monte Carlo simulations. Column 12: False matching rates determined based on comparing the matching results of the 223 *XMM-Newton* sources that also have archival *Chandra* detections, i.e., the fraction of these 223 X-ray sources to have different *XMM-Newton* matching results than the *Chandra* results. For the summary row, Columns 10–12 are calculated as the weighed sum (based on the number of primary counterparts from each catalog) of the results from all four OIR catalogs.

Catalog	Limiting Magnitude	Area deg <sup>2</sup>	$\sigma$	$LR_{th}$	$N_{All}$	$\overline{N}_{99\%}$	$N_{Reliable}$	$N_{Acceptable}$	$f_{AP}$	False Rate (Simulation)	False Rate ( <i>Chandra</i> )
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
SERVS	$3.6\mu m < 23.1$	5.0	$0.5''$	0.34	4530	1.46	3898	259	96.8%	4.2%	2.7%
VIDEO	$K_s < 23.8$	4.5	$0.3''$	0.24	4285	1.88	3804	90	86.3%	8.0%	5.6%
CFHTLS-wide	$i < 24.8$	5.4	$0.2''$	0.27	5022	2.37	4138	129	75.6%	15.6%	9.2%
HSC-PDR1	$i < 26.5$	5.4	$0.1''$	0.13	5002	4.39	4192	52	78.6%	18.4%	12.7%
Summary	N/A	N/A	N/A	N/A	5194	N/A	4854	61	93.1%	5.8%	2.9%

offsets between the X-ray sources and the reliably matched sources in Fig. 13.

## 5.2 Counterpart identification reliability

We assess the reliability of the MLE matching results using the Monte Carlo simulation approach described in Broos et al. (2007) and Xue et al. (2011). Compared to the simple estimation based on matching OIR catalogs to a random X-ray catalog, the Broos et al. (2007) method usually provides a much more realistic assessment of the matching reliability. As described in Broos et al. (2007) and Broos et al. (2011), we consider our X-ray sources consist of two different intrinsic populations, the “associated population” and the “isolated population”. The associated population is comprised of X-ray sources that do have a real counterpart in the corresponding OIR catalog, and the X-ray sources that should not have any OIR counterparts belong to the isolated population. For the associated population, counterpart-matching procedures can produce three different outcomes: (1) an X-ray source is matched to its correct counterpart (correct match, or CM), (2) an X-ray source is matched to an incorrect counterpart (incorrect match, or IM), and (3) no counterparts were recovered (false negative, or FN). For the isolated population, there are two possible matching results: (1) no counterparts are found (true negative, or TN), and (2) an OIR source is identified as a counterpart (false positive, or FP).

In order to estimate the fractions of X-ray sources in both populations for our catalog, we simulate each population separately. The details of the simulation setup can be found in the appendix of Broos et al. (2007) and §5 of Broos et al. (2011). We briefly summarize the simulation procedures as the following. (1) For the “associated population”, we remove all OIR sources considered to be a match in §5.1, then randomly shuffle the OIR catalogs. We then create OIR counterparts for each X-ray source in our catalog based on the X-ray and OIR positional uncertainties, and the expected magnitude distributions derived in §5.1. (2) For the “isolated population”, we create mock X-ray sources

that are at least  $20''$  away from any real X-ray sources. A total of 100 simulations are carried out for each population, and we run the MLE matching procedures on each simulation as described in §5.1. The simulations of the isolated populations usually produce a much higher spurious fraction (i.e., the number of false-positives divided by the size of the X-ray catalog). For the SERVS, VIDEO, CFHTLS, and HSC-PDR1 catalogs, the median spurious fractions of the isolated populations are 19%, 24%, 30%, and 40%, respectively. For the associated populations, the spurious fractions (defined as  $N_{IM}/(N_{IM} + N_{CM})$ ) for SERVS, VIDEO, CFHTLS, and HSC-PDR1 are 3%, 5%, 7%, and 9%, respectively.

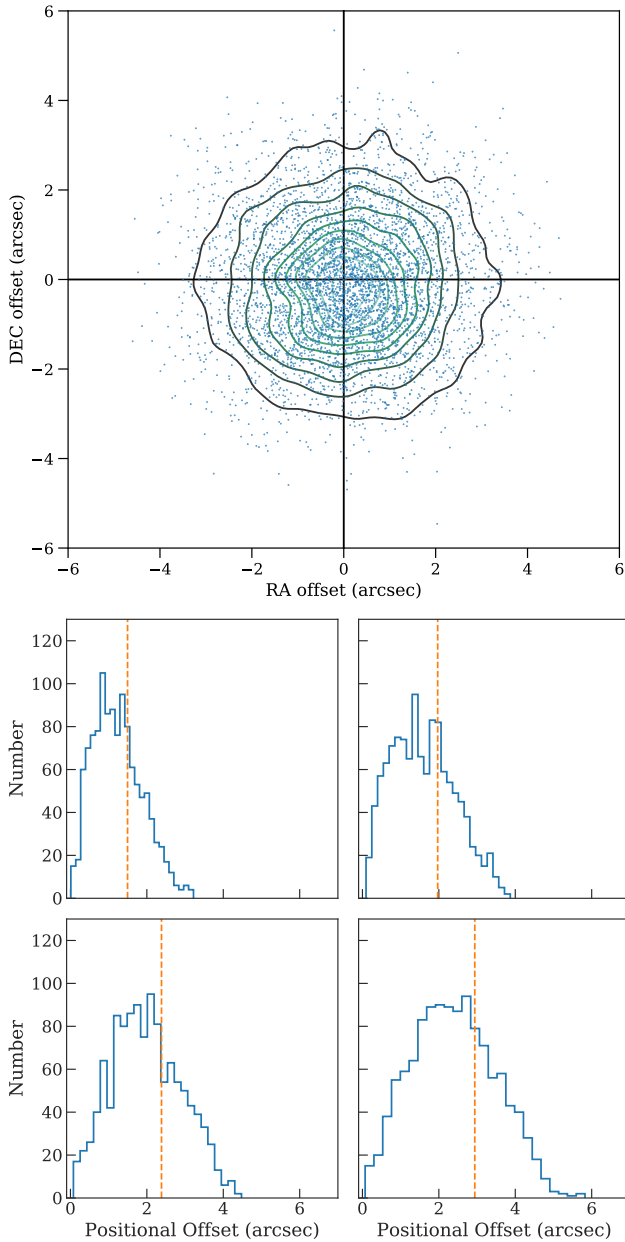
For the MLE matching results of the real data, X-ray sources that were not reliably matched to any counterparts (with a total number of  $N_{negative}$ ) should contain a mixture of the FNs of the associated population and the TNs of the isolated population. Therefore, we can use the median FN and TN from simulations to estimate the fraction of X-ray sources in the associated population ( $f_{AP}$ ):

$$N_{negative} = N_{FN} \times f_{AP} + N_{TN} \times (1 - f_{AP}). \quad (8)$$

With  $f_{AP}$ , we can estimate the expected number of X-ray sources to have a spurious match as the weighted sum of the numbers of IM and FP. The false matching rate,  $f_{False}$  should therefore be:

$$f_{False} = (N_{IM} \times f_{AP} + N_{FP} \times (1 - f_{AP})) / (N_{positive}). \quad (9)$$

Here we consider  $N_{positive}$  as the combination of both the “reliable” and “acceptable” matches reported in Table 3. We carry out simulations for each OIR catalog. The values of  $f_{False}$  and  $f_{AP}$  for each OIR catalog are also reported in Table 3. Due to the high  $f_{AP}$  values, the false matching rates of our matching results are mostly determined by the spurious fractions of the associated populations, which are much lower than those of the isolated populations. Notably, adopting the *Chandra*-matched counterpart magnitude density,  $q(m)_{Chandra}$ , does reduce the false-matching rates compared to those derived using  $q(m)_{XMM-Newton}$ . For the SERVS and VIDEO catalogs, the improvements are marginal ( $< 0.5\%$ ),



**Figure 13.** *Top:* Distribution of the positional offsets in the RA vs. DEC plane for the 4854 reliably matched sources. Note that the median positional offsets are  $< 0.1''$  in both the RA and DEC directions. *Bottom:* Histograms of positional offsets for the 4854 reliably matched sources, divided into four bins based on their positional uncertainties. In each panel, we also mark the 68% positional offset value as the dashed line.

while the improvements for CFHTLS and HSC-PDR1 are more significant ( $\approx 2\%$  and  $6\%$ , respectively.)

We further scrutinize the MLE matching reliabilities by making use of the 223 CSC sources and their multiwavelength matching results described in §5.1. We assess the reliability of the matching results of these *Chandra* sources using the aforementioned Monte Carlo method, and find that the false-match fractions are 0.9%, 1.4%, 2.8%, and 3.3%, for SERVS, VIDEO, CFHTLS, and HSC-PDR1, re-

spectively. For each catalog, we also directly compare the reliable matches obtained with *XMM-Newton* and *Chandra* positions for these 223 sources, and find that 97%, 94%, 91%, and 87% of the *Chandra* matching results and the *XMM-Newton* results are the same for the SERVS, VIDEO, CFHTLS, and HSC catalogs, respectively. The low disagreement rate between the matching results obtained using *Chandra* positions and *XMM-Newton* positions are consistent with the false-matching rates calculated based on the Monte Carlo simulation considering the small number of *Chandra*-detected subsample.

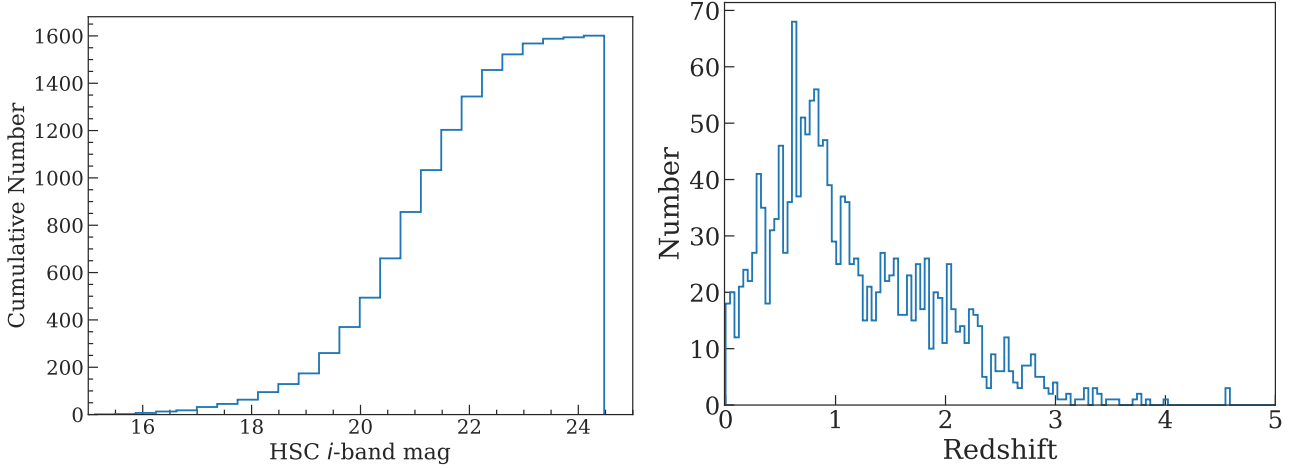
Similar to the full *XMM-Newton* catalog, we also select “primary” counterparts for the *Chandra* sources using the same priority orders. 85%, 10%, 1%, and 4% of the *Chandra* sources have their “primary” counterparts from SERVS, VIDEO, CFHTLS, and HSC-PDR1, respectively. When comparing the primary counterparts of these *Chandra* sources and the primary counterparts of the corresponding *XMM-Newton* sources,  $\approx 97\%$  of them are identical, suggesting that the matching results of the *XMM-Newton* catalog are highly reliable.

### 5.3 Redshifts

The XMM-LSS region is covered by a number of spectroscopic redshift surveys that target galaxies with various optical magnitude constraints: the PRISM Multi-Object Survey (PRISM; Coil et al. 2010), the VIMOS Public Extragalactic Redshift Survey (VIPERS; Garilli et al. 2014), the VIMOS VLT Deep Survey (VVDS; Le Fevre et al. 2013). As part of the SDSS-BOSS program, 3042 X-ray sources found in the XXL-N field ( $25 \text{ deg}^2$ ) with  $r < 22.5$  were all followed-up by SDSS (Menzel et al. 2016). Also, there are three other redshift surveys in the XMM-LSS region that target near-IR selected galaxies, including the spectroscopic follow-up observations of the UKIDSS Ultra-Deep Survey (UDSz; Bradshaw et al. 2013; McLure et al. 2013), the 3D-HST Survey (Skelton et al. 2014; Momcheva et al. 2016) in the UDS region, and the Carnegie-Spitzer-IMACS Redshift Survey (CSI; Kelson et al. 2014).

We adopt the same nearest-neighbor matching criterion with a  $1''$  matching radius to associate these redshifts to each OIR catalog. The redshift for each X-ray source is determined by the coordinates of its primary OIR counterpart. In cases where redshifts from different catalogs do not agree with each other, we choose redshifts using the following order (ranked by spectral resolution): SDSS, VVDS, VIPERS, UDSz, PRIMUS (reliable), CSI (reliable), 3D-HST, PRIMUS (acceptable), CSI (acceptable). Of the 5218 sources in our main X-ray source catalog, 1762 of them have spectroscopic redshifts ranging from  $0.002 < z < 4.57$ . We list the properties of each redshift catalog in Table 4. The cumulative histogram of the *i*-band magnitudes of the sources with redshifts are shown in Fig 14-left. We also show the redshift histogram in Fig 14-right in bins of  $z = 0.04$ . Note that there are several redshift spikes indicative of X-ray large-scale structures (e.g., Luo et al. 2017).

We note that high-quality photometric redshifts for AGNs are not yet available for the our survey region, but are already being constructed for the general galaxy populations. In particular, in a  $\approx 1 \text{ deg}^2$  area within the XMM-LSS region, Nyland et al. (2017, N17 hereafter) has presented a



**Figure 14.** *Left* – Cumulative distribution of the HSC  $i$ -band magnitudes for the X-ray sources with spectroscopic redshifts. *Right* – Distribution of the redshifts in bins of  $\Delta z = 0.04$  for the 1762 X-ray sources with spectroscopic redshift measurements from the literature. The redshift spikes are likely associated with X-ray large-scale structures (e.g., Luo et al. 2017).

**Table 4.** Redshift catalogs used in this work. Column 1: Redshift survey name. Column 2: Survey instrument. Column 3: Survey sensitivity. Column 4: Targeting fields. Column 5: Survey area. Column 6: Total number of redshifts matched the main X-ray catalog. Column 7: Total number of redshifts assigned to the X-ray sources in the main catalog.

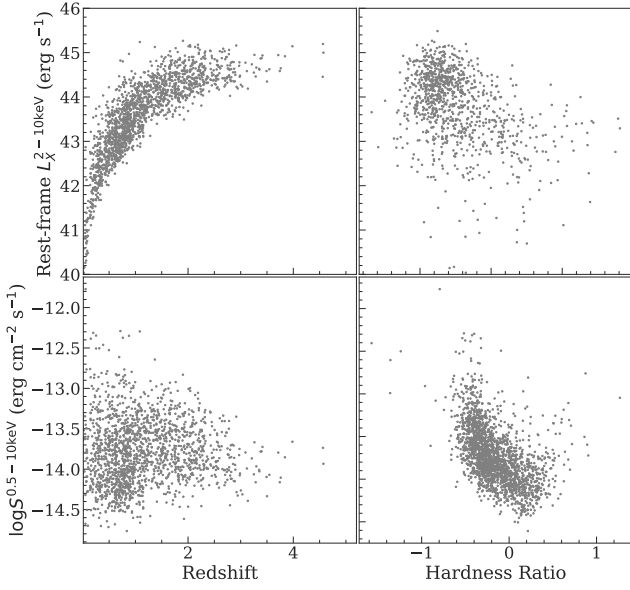
Catalog (1)	Instrument (2)	Survey sensitivity (3)	Targeting fields (4)	Area (5)	$N_{\text{matched}}$ (6)	$N_{\text{assigned}}$ (7)
3D-HST	WFCS G141 Grism	$JH_{\text{IR}} \lesssim 24$	UDS	191.2 arcmin <sup>2</sup>	13	5
CSI	IMACS (Uniform-Dispersion Prism)	$[3.6\mu\text{m}]_{\text{AB}} \lesssim 21$	XMM-LSS	6.9 deg <sup>2</sup>	300	57
PRIMUS	IMACS (Low-Dispersion Prism)	$i \lesssim 23.5$	XMM-LSS	2.9 deg <sup>2</sup>	763	364
SDSS	BOSS	$r \lesssim 22.5$	XXL-North	25 deg <sup>2</sup>	1104	1104
UDSz	VIMOS/FORS2	$K < 23$	UDS	0.5 deg <sup>2</sup>	24	17
VVDS	VIMOS	$17.5 \lesssim i \lesssim 24.5$	XMDS+SXDS	3 deg <sup>2</sup>	87	43
VIPERS	VIMOS	$i \lesssim 22.5$	XMM-LSS	7.8 deg <sup>2</sup>	350	172

“forced-photometry” catalog, which is ideal for photometric redshift measurements. We make use of the 12-band photometry from  $u'$  to IRAC  $4.5\mu\text{m}$  (from CFHTLS, VIDEO, and SERVS) in N17 to derive photometric redshifts for the X-ray sources in this region using the methods described in Yang et al. (2014). We match the N17 to the coordinates of the primary counterparts of our source catalog using a  $1''$  matching radius, and matched 902 X-ray sources. We exclude 177 X-ray sources that are classified as a broad-line AGN according to their optical spectra. Since the flux uncertainties in N17 do not account for uncertainties in PSF modeling, we adopt an additional 3% flux errors, which is typical of PSF modeling uncertainties (e.g., §5.3 of Yang et al. 2014). Following the recipe of Yang et al. (2014), we measure the photometric redshifts using the SED-fitting code EAZY (Brammer et al. 2008) using the default galaxy template and settings. As described in §5.6 of Yang et al. (2014), we perform iterative procedures to adjust photometric zero points. The zero-point corrections are  $\lesssim 0.1$  mag. For each source, EAZY calculates a parameter  $Q_z$  (see Eq. 8 of Brammer et al. 2008) to indicate photometric redshift quality. We consider photometric redshifts with  $Q_z < 1$  as reliable (see §6.3 of Yang et al. 2014). There are 91 sources with  $Q_z < 1$  and reliable spectroscopic redshifts, which can be used to assess the quality of photometric redshift measurements. We find a normalized

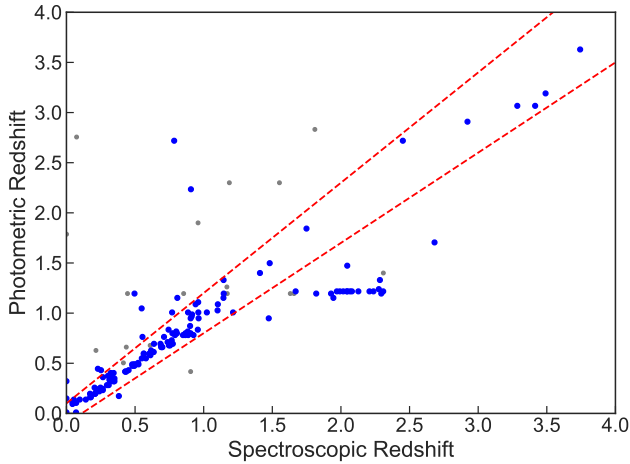
median absolute deviation (NMAD) of  $\sigma_{\text{NMAD}} = 0.024$ , with an outlier (defined as  $|\Delta z|/(1 + z_{\text{spec}}) > 0.15$ ) fraction of 7.7%, which is comparable to the photometric redshift reliability reported in Yang et al. (2014). We compare the photometric and spectroscopic redshifts for these 91 sources in Fig 16. In the 1 deg<sup>2</sup> area covered by N17, we obtain 349 reliable photometric redshifts for four sources that do not have spectroscopic redshift measurements, increasing the fraction of sources with redshifts from  $\approx 32\%$  to  $\approx 70\%$ . We note that a similar forced-photometry catalog for our full survey region including the deep HSC-PDR1 photometry is currently being constructed (Pforr et al., in preparation), we expect to achieve at least 70% photometric redshift completeness for our full X-ray catalog when the full forced-photometry catalog becomes available.

#### 5.4 Source properties and classification

In this section we briefly discuss some of the properties of the 4854 sources with reliable or acceptable counterparts. For the X-ray sources with secure redshifts, we calculate their rest-frame 2–10 keV luminosity assuming a  $\Gamma = 1.7$  power-law spectrum and corrected for the Galactic absorption. We show the  $L_X - z$  distribution of our sample in Fig. 15, along with the  $L_X$  vs. HR, HR vs. full-band flux, and full-band



**Figure 15.** Properties of the 1762 X-ray sources with spectroscopic redshift measurements, including (1) the  $L_{2-10 \text{ keV}} - z$  distribution (top-left), (2)  $L_{2-10 \text{ keV}}$  vs. hardness ratio (top-right), (3)  $0.5-10 \text{ keV}$  flux vs. redshift (bottom-left), (4)  $0.5-10 \text{ keV}$  flux vs. hardness ratio (bottom-right).



**Figure 16.** Spectroscopic and photometric redshifts for the 109 sources that are not classified as a broad-line source according to their original spectroscopic flags. Of the 109 sources, 91 sources with high quality photometric redshifts ( $Q_z < 1$ ) are marked as the blue circles.

flux vs. redshift distributions. In Fig. 17, we also compare the luminosity, flux, and redshift distributions of our sample to those from archival X-ray surveys, including XMM-COSMOS, COSMOS-Legacy, and Stripe 82X. The comparisons in the middle and right panels are limited to sources with available spectroscopic redshifts in the Stripe-82 and XMM-LSS regions. The right panel of Fig. 17 demonstrates that our catalog occupies a valuable part of parameter space among X-ray surveys by more than doubling the source counts of the XMM-COSMOS survey, which will enable a

wide range of science that was previously limited by either survey sensitivity or cosmic variance.

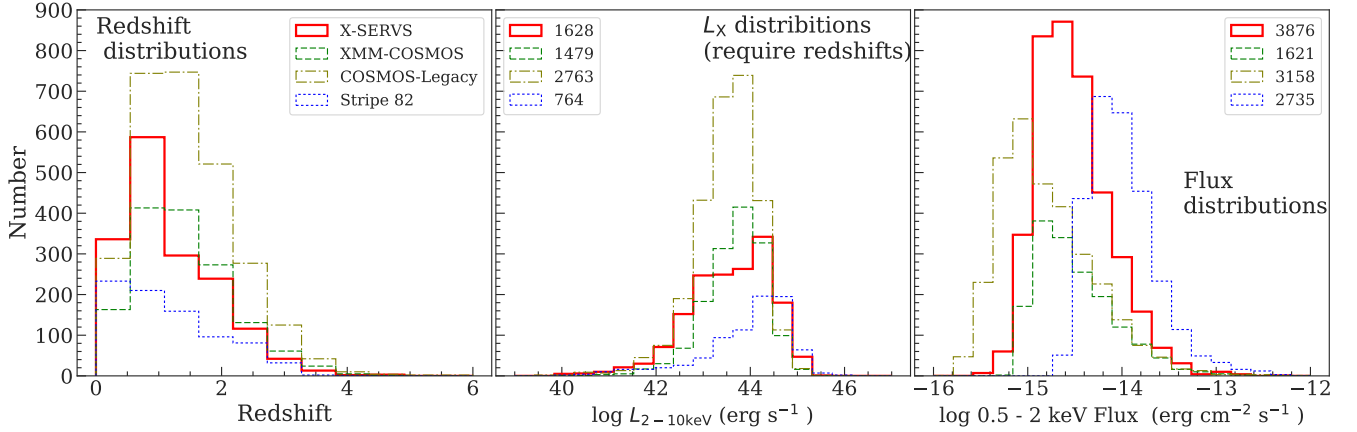
We also include basic source classifications in our catalog. For sources with spectroscopic observations, we directly make use of the spectroscopic classifications when available. A total of 831 sources are classified as AGNs based on the broad-line spectroscopic flags specified in SDSS, VIPERS, or VVDS catalogs. For the other sources, we use the criteria described in Luo et al. (2017) to select AGNs, including: (1) A X-ray luminosity threshold where we regard sources with rest-frame  $L_{2-10 \text{ keV}} > 3 \times 10^{42} \text{ erg s}^{-1}$  as an AGN. A total of 1504 sources satisfy this criterion. (2) X-ray bright sources with X-ray-to-optical and X-ray-to-near-IR flux ratios larger than  $\log f_x/f_R > -1$  and  $\log f_x/f_{Ks} > -1.2$ , respectively. There are 3409 sources with  $\log f_x/f_R > -1$  and 4071 sources with  $\log f_x/f_{Ks} > -1.2$ . We show the flux-ratio distributions in Fig. 18. The total number of sources classified as AGNs is 4661, or  $\approx 89\%$  of the total sample. For the other  $\approx 12\%$  of sources, some of them might be powered by star formation processes in galaxies given their relatively weak X-ray to OIR ratios. However, we expect the majority of these sources will be identified as X-ray AGNs when redshift information become available. Since the median X-ray flux of the 12% unclassified sources will exceed the X-ray AGN luminosity threshold at  $z \geq 0.3$ , below which most of the galaxies would have already been included in existing wide-field redshift surveys in XMM-LSS.

## 6 SUMMARY

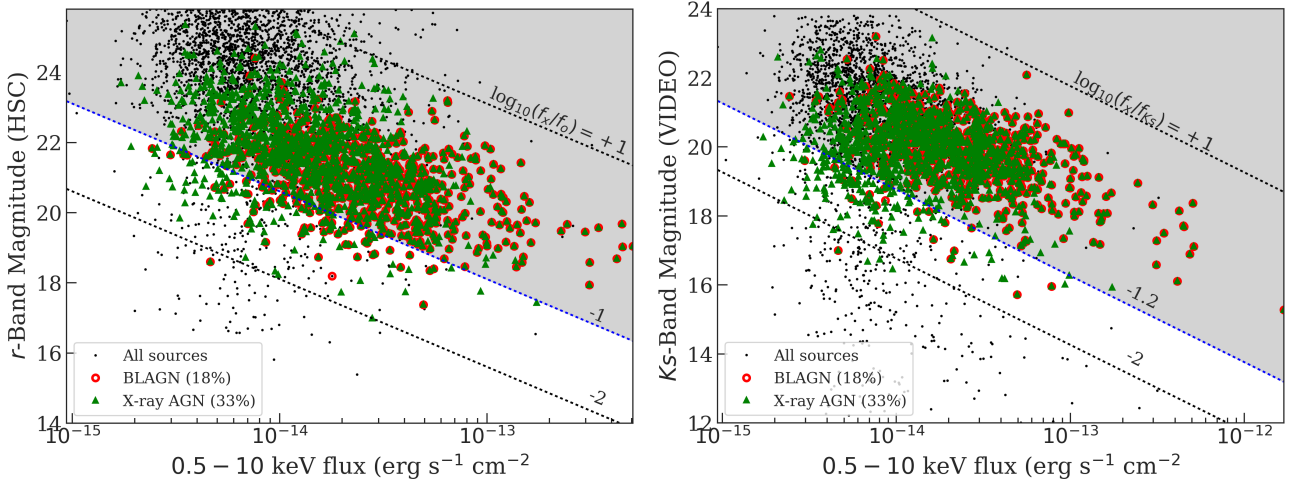
In this work, we present a new X-ray source catalog in the XMM-LSS region constructed using both new AO-15 and archival *XMM-Newton* data. The main results are the following:

1. Our X-ray catalog is constructed based on data in a  $\approx 5.3 \text{ deg}^2$  rectangular region centered at RA= 35.58°, DEC=−4.965°. A total of 155 pointings from 149 different *XMM-Newton* ObsIDs are used, with a total of  $\approx 2.7 \text{ Ms}$  background-filtered exposure time (1.1 Ms from AO-15). The median value of the cleaned PN exposure time is 46 ks for the full  $5.3 \text{ deg}^2$  field, reaching the desired homogeneity and survey depth.
2. The main X-ray source catalog is generated using EWAVELET and EMLDETECT. We list all 5218 sources with EMLDETECT DET\_ML > 6 in the soft-band (0.5–2 keV), hard-band (2–10 keV), or the full-band (0.5–10 keV). Of the 5218 sources, we find 2843 of them to be the same X-ray sources identified in previous X-ray surveys in our survey area (the XMM-XXL-North surveys; Liu et al. 2016), and 2375 are newly discovered X-ray sources.
3. The median fluxes in  $\text{erg cm}^{-2} \text{ s}^{-1}$  for the three X-ray bands are  $2.7 \times 10^{-15}$  (0.5–2 keV),  $1.5 \times 10^{-14}$  (2–10 keV),  $8.8 \times 10^{-15}$  (0.5–10 keV). There are 2928 sources with more than 100 X-ray counts in full-band (PN + MOS), and 126 sources with more than 1000 X-ray counts.
4. Monte Carlo simulation suggests that the number of spurious sources should be  $\approx 41$  with a DET\_ML= 6.0 threshold,  $\approx 99.2\%$  reliability. If we raise the detection threshold to DET\_ML= 10.8, or 99.8% reliability, and the expected number of spurious sources would be 12.





**Figure 17.** A comparison between this work (solid red line), XMM-COSMOS (green dashed line), COSMOS-Legacy (brown dash-dotted line), and Stripe 82-X (blue dotted line) surveys. Distributions shown in panels from left to right are: redshift,  $\log L_{2-10 \text{ keV}}$ , and 0.5–2 keV flux, respectively. For the left and middle panels, the histograms are for the same set of sources with redshift measurements, with source numbers marked in the middle panel. Right panel shows the distribution of soft-band fluxes for the soft-band detected sources in each catalog, no redshift information is required. The numbers of the soft-band sources are also shown in the right panel.



**Figure 18.** *Right* – Distributions of the full-band (0.5–10 keV) X-ray flux and the optical flux in the  $r$ -band. *Left* – Full-band X-ray flux versus the near-IR flux in the  $Ks$ -band. In both plots, the shaded regions mark the “AGN” regime as defined by the  $\log_{10} f_x/f_o > -1$  (right) or the  $\log_{10} f_x/f_o > -1.2$  (left) thresholds as described in §4.4 of [Xue et al. \(2011\)](#). For sources with spectroscopic redshift measurements, we also make those with  $L_X > 3 \times 10^{42} \text{ erg s}^{-1}$  as the green triangles. The 831 sources with optical spectra consistent with a broad-line AGNs are also marked as open red circles.

5. The absolute astrometry of the *XMM-Newton* catalog is registered to the WCS frame of the Subaru HSC-PDR1 survey. The positional uncertainties for the X-ray sources are determined based on an empirical relation between the X-ray-to-optical positional offsets and the X-ray source counts. Our empirical positional uncertainties are well-characterized by the Rayleigh distribution. The median positional uncertainties in the soft, hard, and full-bands are: 1.31″, 1.35″, and 1.37″, respectively.
6. We search for OIR counterparts in the SERVS, VIDEO, CFHTLS, and HSC-PDR1 surveys, and we find that 99.5% (5196/5218) of the X-ray sources have at least one OIR counterpart within the 99.73% positional uncertainties. We also find that  $\approx 93\%$  (4854/5218) of the X-ray sources have at least one reliable OIR counterpart. We also collect

1762 secure redshifts from SDSS, VIPERS, VVDS, UDSz, PRIMUS, CSI, and 3D-HST.

7. We test the matching results using a subsample of 223 X-ray sources with a reliable *Chandra* counterpart from CSC 2.0. We find that  $\approx 97\%$  of the matching results from *XMM-Newton* and *Chandra* are identical, suggesting our multi-wavelength matching results are highly reliable.
8. We classify 4584 X-ray sources as AGNs. The classification is based on their optical spectra from SDSS, VIPERS, or VVDS (831); X-ray luminosity larger than  $3 \times 10^{42} \text{ cm}^{-2} \text{ s}^{-1}$  (1504); and large X-ray-to-optical and X-ray-to-NIR flux ratios (4479).

The X-ray source catalog presented in this work is the first  $> 2 \text{ deg}^2$  X-ray survey with sensitivity comparable to

that of COSMOS. The 5.3 deg<sup>2</sup> wide-area and 46 ks depth survey will enable a wide range of studies. For instance, the large AGN sample and the excellent multiwavelength coverage will provide the means to exploring the behaviors of AGNs in the multidimensional space of galaxy parameters. The wide-area of this survey will also enable studies of AGN triggering mechanisms as a function of environments. In the near future, the combination of AGN samples from this work, COSMOS, and additional X-ray surveys in other X-SERVS fields will sample the full range of cosmic large-scale structures, greatly alleviating the cosmic variance uncertainties seen in previous COSMOS results and therefore advancing our understandings of coevolution of SMBHs and their host galaxies.

## ACKNOWLEDGMENTS

**To be completed** We acknowledge the support of NASA grant NNX17AF07G (CTC, WNB). This work is based on observations taken by the 3D-HST Treasury Program (GO 12177 and 12328) with the NASA/ESA HST, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555.

Funding for PRIMUS is provided by NSF (AST-0607701, AST-0908246, AST-0908442, AST-0908354) and NASA (Spitzer-1356708, 08-ADP08-0019, NNX09AC95G).

HSC-PDR1: NAOJ / HSC Collaboration.

VIPERS: This paper uses data from the VIMOS Public Extragalactic Redshift Survey (VIPERS). VIPERS has been performed using the ESO Very Large Telescope, under the "Large Programme" 182.A-0886. The participating institutions and funding agencies are listed at <http://vipers.inaf.it>. VVDS: This research uses data from the VIMOS VLT Deep Survey, obtained from the VVDS database operated by Césam, Laboratoire d'Astrophysique de Marseille, France.

SDSS :

CSI :

## APPENDIX A: MAIN CATALOG DESCRIPTION

Here we describe the columns of the main X-ray source catalog, Table A. Throughout the table, we mark null values as -99.

### X-ray properties

Columns 1–100 give the X-ray properties of our sources. Columns for the soft-band results are marked with the "SB\_" prefix. Columns for the hard and full band results are marked with the HB\_ and FB\_ prefixes, respectively.

- (1) Column 1: The unique source ID (XID) assigned to each X-ray source.
- (2) Columns 2–3: RA and DEC in degrees of the X-ray source. The positions are determined based on EMLDETECT. Based on availability, we use the positions from, in priority order of, full-band, soft-band, and hard-band as the primary position of the X-ray source. Band-specific positions are listed in Columns 8–13.
- (3) Column 4: X-ray positional uncertainty ( $\sigma_x$ ) in arcsec based on the empirical relation between source counts and positional offsets to the HSC-PDR1 catalog. Note that this is

not the  $\sigma$  of a 2D-Gaussian distribution but rather the scaling parameter of the univariate Rayleigh distribution (see §3.3 and Pineau et al. 2017 for details). The positional uncertainties are based on those of the full-band. For sources without a full-band detection, the soft or hard-band positional uncertainties are listed. See §3.3 for details.

- (4) Columns 5–6: 68% and 99.73% X-ray positional uncertainties in arcsec based on the Rayleigh distribution, see §3.3 for details.
- (5) Column 7: Positional uncertainties calculated by EMLDETECT,  $\sigma_{\text{eml}}$ . Similar to  $\sigma_x$ , we list the full-band values when possible and list soft or hard-band  $\sigma_{\text{eml}}$  for sources not detected in the full band.
- (6) Columns 8–13: RA and DEC in degrees of the source in the soft, hard, and full bands, respectively.
- (7) Columns 14–16: The source detection threshold in each band, DET\_ML, which is computed using EMLDETECT.
- (8) Columns 17–19: Total (PN + MOS1 + MOS2) exposure time in seconds in each band.
- (9) Columns 20–28: PN, MOS1, and MOS2 exposure time in seconds in each band.
- (10) Columns 29–31: Total background map values (PN + MOS1 + MOS2) in counts per pixel in each band.
- (11) Columns 32–40: PN, MOS1, and MOS2 background map values in counts per pixel in each band.
- (12) Columns 41–43: Total (PN + MOS1 + MOS2) net counts in each band.
- (13) Columns 44–52: PN, MOS1, and MOS2 net counts in each band.
- (14) Columns 53–64: Uncertainties of total, PN, MOS1, and MOS2 net counts in each band.
- (15) Columns 65–76: Total, PN, MOS1, and MOS2 net count rates in each band, in count s<sup>-1</sup>.
- (16) Columns 77–88: Uncertainties of total, PN, MOS1, and MOS2 net count rates in each band, in count s<sup>-1</sup>.
- (17) Columns 89–94: Flux and flux uncertainty in different bands, in erg cm<sup>-2</sup> s<sup>-1</sup>. The conversion factors between count rates and fluxes are derived assuming a power-law spectrum with a  $\Gamma = 1.7$  photon index and Galactic absorption column density for each EPIC detector, See §3.4 for details. The flux and uncertainties reported here are the error-weighted average of all EPIC detectors
- (18) Columns 95–97: Hardness ratio, defined as (H-S)/(H+S), where H is the total (PN + MOS1 + MOS2) net counts divided by total exposure time in the hard band and S is the total net counts divided by total exposure time in the soft band. The uncertainties of HRs are calculated based on the count uncertainties using the error propagation method described in §1.7.3 of Lyons (1991). Sources detected only in the full-band are set to -99 in all three columns. We note that the on of the CCD on MOS1 was affected by a micrometeorite impact, therefore H and S are sometimes calculated based on only results from two cameras with non-zero exposure time.
- (19) Columns 98: Rest-frame, observed 2–10 keV X-ray luminosity (only corrected for Galactic absorption) computed as in §5.4.
- (20) Columns 99: CSC 2.0 source name of the nearest *Chandra* source in CSC.
- (21) Columns 100: XXL-North catalog source name of the nearest *XMM-Newton* source in Liu et al. (2016).

### Multiwavelength matching results

Columns 101–112 show the multiwavelength matching results based on the MLE method described in §5.1. In these columns, the 99.73% positional uncertainty radius represents the quadratic sum of the positional uncertainties of each X-ray source and the corresponding OIR catalog (see Table 3).

- (1) Columns 101–104: Number of sources from each OIR catalog within the 99.73% positional uncertainty radius of each X-ray source. The average values of these columns are listed in Column (7) of Table 3.
- (2) Columns 105–108: Number of sources from each OIR catalog that satisfies  $LR \geq LR_{th}$ .
- (3) Column 109: Flag set to 1 if the X-ray source has at least one reliable counterpart with  $LR > LR_{threshold}$  from any of the four OIR catalogs. See §5.1 for details.
- (4) Column 110: Flag set to 1 if the X-ray source has no reliable counterparts, but has at least one acceptable counterpart from any of the four OIR catalogs. See §5.1 for details.
- (5) Column 111: Flag set to 1 if the X-ray source has at least one counterpart from any of the four OIR catalogs, but none of the OIR counterparts satisfy the reliable or acceptable criteria described in §5.1.
- (6) Column 112: Flag set to 1 if the X-ray source has no counterpart from any of the four OIR catalogs within the 99.73% positional uncertainty radius.

### Multiwavelength properties

Columns 113–184 give the multiwavelength properties from each OIR catalog for the primary counterparts matched to X-ray sources using the MLE methods. Properties from SERVS, SWIRE, VIDEO, CFHTLS, and HSC-PDR1 are marked with prefixes SERVS\_, SWIRE\_, VIDEO\_, CFHT\_, and HSC\_, respectively.

- (1) Column 113: Catalog from which the primary counterpart is selected. The primary counterpart is chosen in priority order of: SERVS, VIDEO, CFHTLS, and HSC-PDR1, which is based on the matching reliability of each OIR catalog. See §5.2 for details.
- (2) Column 114–116: RA and DEC in degrees of the primary counterpart and its separation in arcsec from the X-ray source.
- (3) Column 117: The matching likelihood ratio of the primary counterpart.
- (4) Columns 118–129: RA, DEC, and Object ID of the primary counterpart culled from the original OIR catalogs.
- (5) Columns 130–133: SERVS 1.9" aperture photometry and the associated uncertainties in the 3.6 $\mu$ m and 4.5 $\mu$ m bands.
- (6) Columns 134–141: SWIRE 1.9" aperture photometry and the associated uncertainties in the 3.6 $\mu$ m, 4.5 $\mu$ m, 5.8 $\mu$ m, and 8.0 $\mu$ m.
- (7) Columns 142–143: SWIRE 1.9" aperture photometry and the associated uncertainties in the 24 $\mu$ m band.
- (8) Columns 144–153: VIDEO PSF photometry and uncertainties in AB magnitude in the  $Z$ ,  $Y$ ,  $J$ ,  $H$ , and  $K_s$  bands.
- (9) Columns 154–163: CFHTLS PSF photometry and uncertainties in AB magnitude in the  $u$ ,  $g$ ,  $r$ ,  $i$ ,  $z$  bands.
- (10) Columns 164–173: HSC cmodel photometry and uncertainties in AB magnitude in the  $g$ ,  $r$ ,  $i$ ,  $z$ ,  $y$  bands.

- (11) Columns: 174–176: RA, DEC, and Object ID from the original redshift catalogs for the primary counterparts.
- (12) Column 177: Spectroscopic redshift adopted for the X-ray source. The redshifts are chosen based on the spectral resolution of the observations. See §5.3 for details.
- (13) Column 178: The catalog from which the redshift is culled from.
- (14) Column 179: Original redshift flag from one of the redshift catalogs. For SDSS, see <http://www.sdss.org/dr14/algorithms/bitmasks/#ZWARNING> for the definition of flags. For VVDS, see §3.4 of Le Fevre et al. (2013) for the definition of flags. For VIPERS, see §4.3 of Garilli et al. (2014) for the definition of flags. For PRIMUS, see <http://primus.ucsd.edu/version1.html#ztags> for the definition of flags. For CSI, see §4.6 of Kelson et al. (2014) for the definition of flags. For UDSz, see McLure et al. (2013) for the definition of flags. For the 3D-HST catalog, we only select redshifts with  $\sigma_z/(1+z) \leq 0.003$  thus no redshift flags are included.
- (15) Column 180–183: Photometric redshift, the associated upper and lower uncertainties, and the photometric redshift quality parameter ( $Q_z$ ). The photometric redshift measurements are limited to the 1 deg<sup>2</sup> area with forced-photometry from Nyland et al. (2017). See §5.3 for details.
- (16) Column 184: Flag set to 1 if the X-ray source is classified as an AGN based the “broad-line” flag in the original redshift catalog. Flag set to 2 if the X-ray source is classified as an AGN based on its X-ray luminosity, where  $L_X > 3 \times 10^{42} \text{ s}^{-1}$ . Flag set to 3 if the X-ray source is classified as an AGN based on the large X-ray-to-optical or X-ray-to-near-IR flux ratios, see Fig 18. Flag set to 0 if the source is not classified as an AGN, see §5.4 for details.

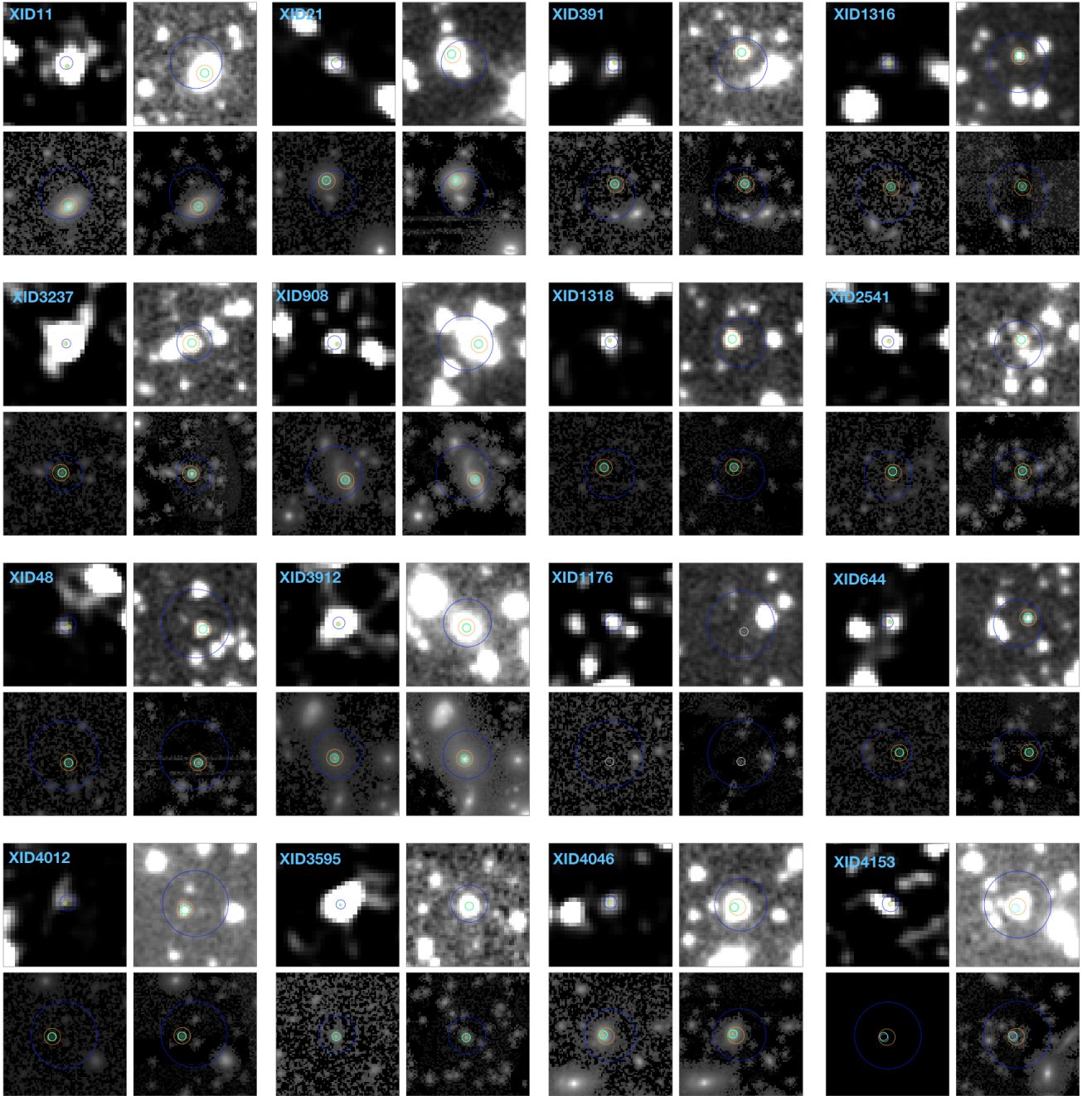
### Multiwavelength properties for additional counterparts

In our source catalog, there are 947 X-ray sources with two  $LR \geq LR_{th}$  counterparts and the second highest  $LR$  counterpart also satisfies  $LR \geq 0.5LR_{primary}$ . The highest  $LR$  counterparts are considered as “primary” with properties reported in Columns 113–179. Here we report the multiwavelength properties of the “secondary” counterparts in Columns 185–251, which are identical as Columns 113–179 except the additional “SECONDARY\_” prefixes. There are also 20 X-ray sources with three  $LR \geq LR_{th}$  counterparts, where the secondary and the tertiary counterparts both satisfy the  $LR \geq 0.5LR_{primary}$  criterion. The multiwavelength properties of the secondary counterparts for these 20 sources are also reported in Columns 185–251. The properties for the tertiary counterparts are reported in Columns 252–318, which are identical to Columns 113–179 except the additional “TERTIARY\_” prefixes.

### Postage-stamp images

For each source, we also generate postage-stamp images in the X-ray, mid-IR, near-IR, and optical wavelengths. For illustration, we show a random collection of 16 sources in Figure A1.





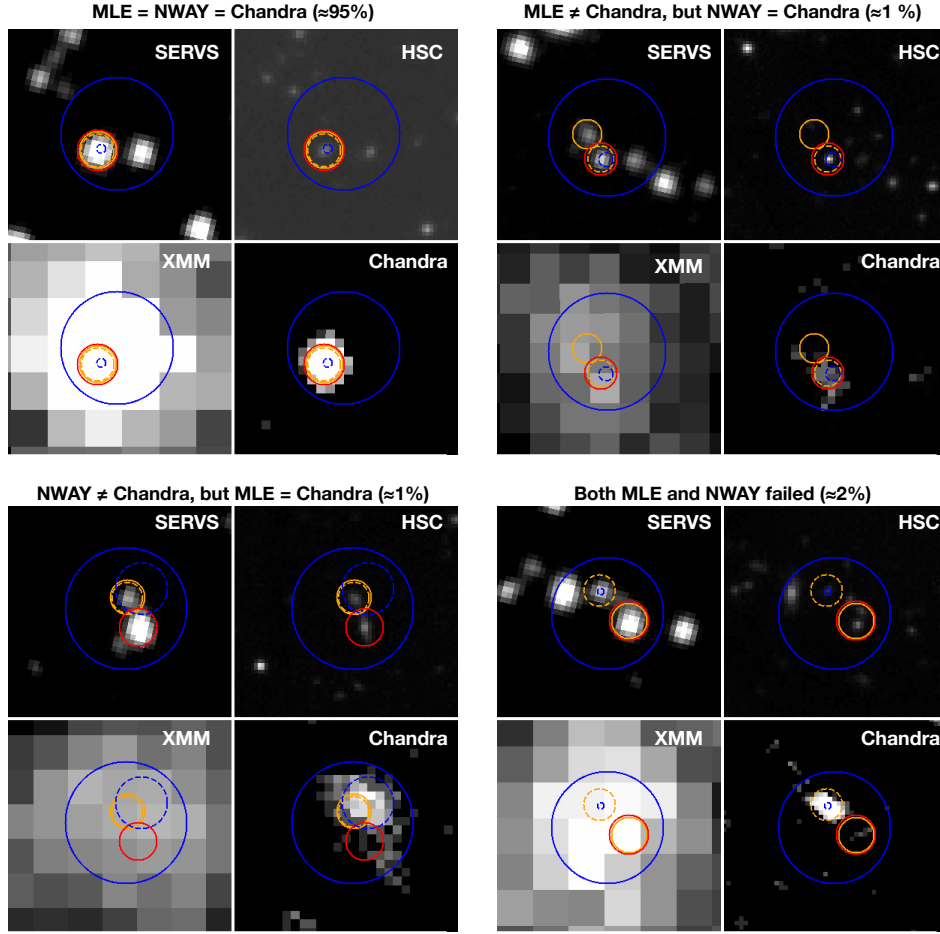
**Figure A1.** Postage-stamp images for 16 randomly selected X-ray sources in our catalog. For each source, we show (1) X-ray image (upper-left panel) with the unique source ID from Table A. (2) mid-IR image from SERVS (upper-right panel) (3) near-IR image from VIDEO (lower-left panel) (4) optical image from HSC-PDR1 (lower-right panel). Due to the large pixel size, the X-ray image for each source is set at  $2' \times 2'$ . For the OIR images, the sizes are set at  $0.5' \times 0.5'$ . In each image, the X-ray position is marked as the blue circle with the corresponding 99.73% X-ray positional uncertainty radius. The position of the most-probable mid-IR SERVS counterpart is marked as the orange circle with a  $2''$  radius. The positions of VIDEO, CFHTLS, and HSC-PDR1 counterparts are marked as  $1''$  circles of green, cyan, and white colors, respectively. The entire set of postage-stamp images are available in the electronic version.

## APPENDIX B: SUPPLEMENTARY MULTIWAVELENGTH MATCHING RESULTS WITH NWAY BAYESIAN CATALOG MATCHING METHOD

We supplement the MLE matching results with the Bayesian catalog matching tool, NWAY (Salvato et al. 2017).<sup>16</sup> The

<sup>16</sup> <https://github.com/JohannesBuchner/nway>.





**Figure B1.** Illustrations of the comparison between the matching results using *XMM-Newton* positions or *Chandra* positions for four X-ray sources in our sample. Within the panel for each source, we show images from SERVS [3.6 $\mu$ m] (top-left), HSC *i*-band (top-right), *XMM-Newton* 0.5–10 keV (bottom-left), and *Chandra* 0.5–7 keV (bottom-right). X-ray positions are marked as blue circles with a 99.73% error radius, with the *XMM-Newton* positions marked using solid lines and the *Chandra* positions marked using dashed lines. SERVS counterparts identified with the MLE method are marked as orange circles with a 2'' radius, solid lines mark the counterparts of the *XMM-Newton* positions, and dashed lines mark the counterparts of the *Chandra* positions. SERVS counterparts of the *XMM-Newton* positions identified using NWAY are shown as the red circles. For the vast majority of *XMM-Newton* sources with *Chandra* counterparts from CSC, our counterpart-matching results are identical to the results obtained using *Chandra* coordinates and positional uncertainties.

fundamental difference between the Bayesian approach and the likelihood-ratio approach is that the Bayesian approach makes use of the distance and magnitude priors from multiple catalogs simultaneously to select the most probable counterpart in all catalogs considered. The details of NWAY's matching methodology is described in Appendix B of Salvato et al. (2017).

NWAY computes three informative quantities for deciding the most probable match,  $p_{\text{single}}$ ,  $p_{\text{any}}$  and  $p_i$ , where each possible counterpart has a different  $p_{\text{single}}$  value based on their distance from the *XMM-Newton* position. This value could be weighted by the priors supplied (e.g.,  $q(m)$  and  $n(m)$  in Eq. 6 are similar to a magnitude prior). In our case,  $p_{\text{single}}$  is the posterior probability for a counterpart to be correctly associated with the X-ray source based on the angular separation from the X-ray position weighted by the magnitude distribution prior, and the surface densities of the X-ray and OIR catalogs. For each X-ray source,

$p_{\text{single}}$  of all possible counterparts are considered to compute a single  $p_{\text{any}}$  value, which stands for the posterior probability of the X-ray source having any correct counterparts (i.e.,  $p_{\text{any}} = 0$  if there are no OIR counterparts within the search radius of the X-ray source.) The last quantity,  $p_i$ , is the relative probability of a possible counterpart being the correct match. For an X-ray source with multiple possible counterparts, the counterpart with the highest  $p_i$  ( $p_i$  (Best)) is considered to be the most probable match and is assigned the MATCH\_FLAG = 1 flag by NWAY. Counterparts with  $p_i$  higher than 50% of  $p_i$  (Best) are also flagged by NWAY as MATCH\_FLAG = 2.

Similar to our MLE approach, we make use of the *Chandra* sources in the XMM-LSS field to compute the priors of the expected counterparts. We use the “auto” functionality of NWAY with a 1.5'' search radius for defining the “real” counterparts. In addition to the magnitude priors, we also include an additional prior based on the *Spitzer*

IRAC color from SERVS,  $[3.6\mu\text{m}]/[4.5\mu\text{m}]$ . Since the majority of our X-ray sources are expected to be AGNs, the distinct  $[3.6\mu\text{m}]/[4.5\mu\text{m}]$  mid-IR color of luminous AGNs will provide additional discerning powers. For a small number of sources, this additional prior is useful for discerning two adjacent SERVS sources with comparable magnitudes (see the top-right panel of Fig. B1 for illustration).

After computing the magnitude and IRAC color priors using the *Chandra* sources, we run NWAY on the full X-ray catalog with a search radius of  $10''$ . All four OIR catalogs are considered simultaneously. We report the multiwavelength matches with `MATCH_FLAG=1,2` in a separate table in the Appendix (Table B) supplementary to the MLE matching results.

Here we also discuss how we estimated the reliability of NWAY matching results. Since NWAY matches all four OIR catalogs simultaneously, we cannot determine the spurious matching rates for the “associated” and “isolated” populations as we did for estimating the spurious matching rates for MLE results using Monte Carlo simulations. Salvato et al. (2017) suggest that the NWAY matching reliability can be determined by a  $p_{\text{any}}$  threshold, which is chosen based on re-running NWAY on randomly shifted “fake” X-ray catalogs. However, this approach is equivalent to estimating the spurious matching rates for the isolated population using the Broos et al. (2007) method, which is usually much higher than the results obtained with the two-population approach (see Broos et al., 2007, Xue et al., 2011, and §5.2 for details). Therefore, we do not adopt any  $p_{\text{any}}$  thresholds for the NWAY matching results. The NWAY matching results can still be assessed by investigating the CSC-matched subsample of 223 X-ray sources, and we find that the difference between the matching results obtained using *Chandra* and *XMM-Newton* positions with NWAY are similar with the MLE results described in §5.2.

We also use the 223 *Chandra*-detected subsample as a baseline for comparing matching results obtained using NWAY or the MLE methods. We focus only on comparing the SERVS counterparts, as the vast majority of MLE matching results are decided based on the primary counterparts from SERVS. We confirm that all *Chandra* sources have the same SERVS matching results using MLE and NWAY. Therefore, we can use the *Chandra* results obtained with MLE to assess the matching reliability of both MLE and NWAY matching results with *XMM-Newton* positions. We show examples of such comparisons in Fig. B1. We find 95% of the sources have the same matching results from MLE, NWAY, and *Chandra*. A small fraction (two sources) of MLE matching results do not agree with those of *Chandra* but could be recovered by NWAY. On the other hand, two of the NWAY matching results do not agree with the *Chandra* results but could be identified by MLE. Five of the *Chandra* sources have different SERVS counterparts than both MLE and NWAY results. As demonstrated in Fig B1, these five sources have multiple counterparts with comparable magnitudes and similar spatial separations from the *XMM-Newton* position. This suggests MLE and NWAY perform similarly for finding SERVS counterparts.

When further scrutinizing the 95% sources with identical SERVS counterparts from MLE, NWAY, and *Chandra*, we find that NWAY cannot identify the correct counterparts in the other three OIR catalogs for a small fraction ( $\sim 10\%$ )

of sources. For instance, one of the X-ray sources has a reliable SERVS counterpart identified by both NWAY and MLE. For the SERVS counterpart, there is only one VIDEO source within the  $0.5''$  positional error circle of SERVS. For the MLE approach described in §5.1, the VIDEO source is associated to the correct SERVS counterpart. However, NWAY does not consider this VIDEO sources to be among the most probable combination of counterparts from all four OIR catalogs that were being matched simultaneously. This is likely due to how NWAY computes  $p_i$ . When multiple OIR catalogs are taken into account simultaneously,  $p_i$  represents the relative probability of counterparts from all OIR catalogs being the correct match. In this example, the VIDEO counterpart has an unlikely magnitude according to the VIDEO magnitude prior therefore including the VIDEO source as a correct match would result in a lower  $p_i$  compared to the case where the VIDEO source is excluded from the matched counterparts. Similar mismatches are also found when comparing the NWAY and MLE matching results for the full *XMM-Newton* catalog. We note that NWAY do not have this behavior when no magnitude or color priors are used. However, without the inclusion of magnitude and color priors NWAY could only rely on the distance-based priors, therefore losing the critical discerning powers for matching *XMM-Newton* sources to the dense OIR catalogs. Further corroborating the Bayesian method’s effectiveness of counterpart-matching with multiple OIR catalogs is beyond the scope of this work. Therefore, we list the NWAY matching results as-is in Table B and we consider only the MLE matching results listed in Table A when exploring the multiwavelength properties of the X-ray sources reported in this work.

The matching results obtained using NWAY are shown in B, and the descriptions of columns are listed below. Only the counterparts with `MATCH_FLAG`  $\geq 1$  are included. Similar to MLE matching results, some of the X-ray sources have multiple probable counterparts. In this table, the same X-ray source can have multiple counterparts and the information of each counterpart is given in an independent row. Similar to columns 138–208 of Table A, properties from SERVS, VIDEO, CFHTLS, and HSC-PDR1 are marked with prefixes SERVS\_, VIDEO\_, CFHT\_, and HSC\_, respectively. Null values are marked as -99 throughout the table.

- (1) Column 1: The unique source ID (XID) assigned to the X-ray source.
- (2) Column 2: The posterior probability of the X-ray source to have any correct counterparts,  $p_{\text{any}}$ , for each X-ray source.
- (3) Column 3: The relative probability of a counterpart to be the correct match,  $p_i$ .
- (4) Columns 4–11: RA and DEC of the counterpart in each OIR catalog.
- (5) Column 12–15: The original Object ID of the counterpart from each OIR catalog.
- (6) Columns 16–19: Separation to the X-ray position from the counterpart in each OIR catalog.
- (7) Columns 20–23: SERVS  $3''$  aperture photometry and the associated uncertainties in the  $3.6\mu\text{m}$  and  $4.5\mu\text{m}$  bands.
- (8) Columns 24–31: VIDEO PSF photometry and uncertainties in AB magnitude in the  $Y$ ,  $J$ ,  $H$ , and  $K_s$  bands.
- (9) Columns 32–41: CFHTLS PSF photometry and uncertainties in AB magnitude in the  $u$ ,  $g$ ,  $r$ ,  $i$ ,  $z$  bands.

- (10) Columns 42–51: HSC cmodel photometry and uncertainties in AB magnitude in the  $g$ ,  $r$ ,  $i$ ,  $z$ ,  $y$  bands.
- (11) Columns 52: Matching flag, `MATCH_FLAG`. For the most probable counterparts the flag is set to 1. For other counterparts that are almost as likely as the most probable counterpart (i.e., with  $p_i \geq p_{-i}$  (Best)), the flag is set to 2.

**Table A.** The main X-ray source catalog with a selection of columns. Empty or null values are marked as -99. The numbers listed on the second row of this table is the column number of the full X-ray catalog with 208 columns. See Appendix A for a detailed descriptions of each column. This table is available in its entirety in machine-readable form online.

XID (1)	RA (2)	DEC (3)	XPOSERR (4)	FB_DET_ML (16)	FB_EXP (19)	FB_BKG (31)	FB_SCTS (43)	FB_FLUX (93)	HR (95)	LX (98)	MLE_FLAG_RE (109)	MLE_CATALOG (113)	ZBEST (117)	ZSOURCE (178)	CLASS (184)
XMM00000	34.200218	-4.035255	1.44	19.0	59076.2	1.74	83.04	$8.30 \times 10^{-15}$	-99	-99	False	SERVS	-99	-99	
XMM00001	34.200713	-4.933734	1.45	63.0	61051.8	1.0	82.03	$7.93 \times 10^{-15}$	-99	$8.98 \times 10^{43}$	True	SERVS	1.82	UDSz	
XMM00002	34.201454	-5.556716	1.96	16.4	29731.6	0.8	29.64	$5.31 \times 10^{-15}$	-99	$2.46 \times 10^{42}$	True	SERVS	0.459	VIPERS	
XMM00003	34.201466	-4.499315	1.5	23.3	72553.8	1.76	72.37	$5.32 \times 10^{-15}$	-99	$1.37 \times 10^{43}$	True	SERVS	0.959	PRIMUS	
XMM00004	34.201949	-4.555523	0.93	316.8	87846.9	1.81	351.91	$2.67 \times 10^{-14}$	-0.43	$9.57 \times 10^{42}$	True	SERVS	0.41	SDSS	
XMM00005	34.202636	-5.690719	1.66	16.5	26430.1	1.01	52.23	$1.30 \times 10^{-14}$	-99	$1.69 \times 10^{44}$	True	CFHTLS	1.932	VIPERS	
XMM00006	34.203276	-4.315289	1.55	29.2	107957.7	1.79	65.42	$2.94 \times 10^{-15}$	-99	-99	True	SERVS	-99	-99	
XMM00007	34.203748	-5.433790	1.77	11.3	78270.8	1.54	41.87	$4.71 \times 10^{-15}$	-99	-99	True	VIDEO	-99	-99	
XMM00008	34.203820	-4.595275	1.17	114.8	83485.0	1.49	168.25	$1.21 \times 10^{-14}$	-0.48	$2.05 \times 10^{42}$	True	VIDEO	0.294	SDSS	
XMM00009	34.204668	-5.378238	1.35	57.1	93769.9	1.4	101.92	$6.80 \times 10^{-15}$	-99	-99	False	SERVS	-99	-99	
XMM00010	34.204772	-4.520794	1.51	32.0	77664.5	1.51	71.54	$5.70 \times 10^{-15}$	-99	-99	True	SERVS	-99	-99	
XMM00011	34.206732	-4.469321	1.25	81.2	63040.8	1.74	132.09	$1.58 \times 10^{-14}$	-99	$2.61 \times 10^{42}$	True	SERVS	0.291	SDSS	
XMM00012	34.207426	-4.585313	1.54	17.3	94203.8	1.85	67.16	$4.22 \times 10^{-15}$	-99	$6.80 \times 10^{41}$	True	SERVS	0.289	PRIMUS	0
XMM00013	34.208245	-5.295083	1.28	66.6	91655.9	1.22	124.63	$7.18 \times 10^{-15}$	-0.32	-99	True	SERVS	-99	-99	3
XMM00014	34.209443	-4.012790	1.34	34.8	63642.5	1.92	106.5	$9.30 \times 10^{-15}$	-99	-99	True	SERVS	-99	-99	3
XMM00015	34.209479	-4.028269	0.88	349.3	68619.6	1.94	426.86	$4.12 \times 10^{-14}$	-0.12	$1.26 \times 10^{44}$	True	SERVS	1.031	SDSS	1
XMM00016	34.209496	-4.421868	1.48	89.5	23041.3	0.22	76.5	$2.31 \times 10^{-14}$	-99	0.0	True	SERVS	0.0	SDSS	0
XMM00017	34.209799	-4.328703	1.12	84.6	115477.7	2.71	194.51	$1.32 \times 10^{-14}$	-99	$2.17 \times 10^{42}$	True	SERVS	0.291	SDSS	1
XMM00018	34.209992	-4.563673	1.14	187.9	88839.2	1.41	181.45	$1.38 \times 10^{-14}$	-0.25	-99	True	SERVS	-99	-99	3
XMM00019	34.210413	-3.890166	1.53	37.9	44730.1	0.95	68.48	$7.55 \times 10^{-15}$	-99	-99	True	CFHTLS	-99	-99	3
XMM00020	34.210863	-5.410119	1.76	11.6	89369.8	1.53	43.01	$2.63 \times 10^{-15}$	-99	-99	True	SERVS	-99	-99	3
XMM00021	34.211423	-5.194363	1.09	104.2	85216.5	1.93	207.87	$1.36 \times 10^{-14}$	-0.36	$1.58 \times 10^{42}$	True	SERVS	0.249	SDSS	0
XMM00022	34.211535	-3.892733	1.62	-99	-99	-99	-99	-99	-99	-99	False	HSC	-99	-99	3
XMM00023	34.211690	-5.313163	1.44	34.3	96622.2	1.24	83.2	$4.54 \times 10^{-15}$	-99	$4.33 \times 10^{42}$	True	SERVS	0.627	VIPERS	2
XMM00024	34.213193	-4.355747	1.64	11.1	100517.8	2.9	54.4	$4.40 \times 10^{-15}$	-99	-99	True	SERVS	-99	-99	3
XMM00025	34.214208	-5.417995	1.75	10.7	88372.6	1.56	43.75	$5.35 \times 10^{-15}$	-99	-99	True	SERVS	-99	-99	3
XMM00026	34.215117	-4.937388	1.82	13.5	64067.0	1.07	38.16	$5.06 \times 10^{-15}$	-99	-99	True	SERVS	-99	-99	3
XMM00027	34.215205	-4.388650	1.73	16.5	67996.2	2.29	45.43	$3.98 \times 10^{-15}$	-99	$4.70 \times 10^{41}$	True	SERVS	0.251	PRIMUS	0
XMM00028	34.215240	-3.918303	1.62	15.1	33523.0	0.94	56.51	$6.71 \times 10^{-15}$	-99	-99	True	SERVS	-99	-99	3
XMM00029	34.217465	-4.087917	1.48	22.2	82858.6	1.96	75.51	$6.05 \times 10^{-15}$	-99	$7.65 \times 10^{42}$	True	SERVS	0.707	PRIMUS	2
XMM00030	34.217481	-5.601158	1.71	17.5	24629.1	1.0	47.25	$1.06 \times 10^{-14}$	-99	$7.02 \times 10^{43}$	True	VIDEO	1.44	SDSS	1

**Table B.** The NWAY matching results with a selection of columns. Empty or null values are marked as -99. See Appendix B for a detailed descriptions of each column.

XID (1)	PANY (2)	PI (3)	SERVS_ID (12)	SERVS_MAG1 (20)	VIDEO_ID (13)	VIDEO_KSMAG (30)	CFHT_ID (14)	CFHT_IMAG (38)	HSC_ID (15)	HSC_IMAG (46)	MATCH_FLAG (52)
XMM00000	0.99	0.96	172727	20.5	644245983190	21.08	1114_031255	21.93	37484971320960508	22.04	1
XMM00001	0.94	1.0	-99	-99	644245946363	17.51	1123_223188	19.29	37484833882002514	19.22	1
XMM00002	1.0	0.99	174016	17.13	644245975018	17.08	1114_024092	17.85	-99	-99	1
XMM00003	0.99	0.6	172557	19.57	644245974364	19.35	1114_023481	20.58	37485108759912843	20.5	1
XMM00004	0.99	0.97	130124	18.41	644245971200	18.41	1114_021145	19.78	37484833882008818	19.85	1
XMM00005	0.99	0.98	-99	-99	-99	-99	1123_211260	20.41	37484692148084612	20.56	1
XMM00006	0.99	1.0	-99	-99	-99	-99	1123_212886	20.98	37484692148084849	21.09	1
XMM00007	1.0	0.99	159785	19.4	644245968189	20.16	1114_018595	21.27	37485108759911057	21.16	1
XMM00008	0.15	0.37	-99	-99	644245967227	19.97	-99	-99	-99	-99	1
XMM00009	0.94	1.0	-99	-99	644246423098	19.85	1114_015512	20.02	-99	-99	1
XMM00010	0.56	0.67	-99	-99	644245971580	21.66	1114_021242	22.87	37484971320960299	21.79	1



## REFERENCES

- Aihara H., et al., 2017, eprint arXiv:1702.08449
- Bradshaw E. J., et al., 2013, *Mon. Not. R. Astron. Soc.*, 433, 194
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *Astrophys. Journal*, Vol. 686, Issue 2, Artic. id. 1503-1513, pp. (2008)., 686
- Brandt W. N., Alexander D. M., 2015, *Astron. Astrophys. Rev.*, 23, 1
- Broos P. S., Feigelson E. D., Townsley L. K., Getman K. V., Wang J., Garmire G. P., Jiang Z., Tsuboi Y., 2007, *Astrophys. J. Suppl. Ser.*, 169, 353
- Broos P. S., et al., 2011, *Astrophys. J. Suppl. Vol. 194, Issue 1, Artic. id. 2, 19 pp. (2011).*, 194
- Brusa M., et al., 2007, *Astrophys. J. Suppl. Ser.*, 172, 353
- Cappelluti N., et al., 2007, *Astrophys. J. Suppl. Ser.*, 172, 341
- Cappelluti N., et al., 2009, *Astron. Astrophys.*, 497, 635
- Chiappetti L., et al., 2005, *Astron. Astrophys.*, 439, 413
- Civano F., et al., 2016, *Astrophys. J.*, 819, 62
- Coil A. L., et al., 2010, *Astrophys. Journal*, Vol. 741, Issue 1, Artic. id. 8, 15 pp. (2011)., 741
- Evans I. N., et al., 2010, *Astrophys. J. Suppl. Ser.*, 189, 37
- Garilli B., et al., 2014, *Astron. Astrophys.*, 562, A23
- George M. R., et al., 2011, *Astrophys. J.*, 742, 125
- Hudelot P., et al., 2012, VizieR On-line Data Cat. II/317. Orig. Publ. SPIE Conf. 2012, 2317
- Jarvis M. J., et al., 2013, *Mon. Not. R. Astron. Soc.*, 428, 1281
- Kelson D. D., et al., 2014, *Astrophys. J.*, 783, 110
- Kim M., Wilkes B. J., Kim D., Green P. J., Barkhouse W. A., Lee M. G., Silverman J. D., Tananbaum H. D., 2007, *Astrophys. J.*, 659, 29
- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *Mon. Not. R. Astron. Soc.*, 457, 4340
- LaMassa S. M., et al., 2016, *Astrophys. J.*, 817, 172
- Le Fevre O., et al., 2013, *Astron. Astrophys.*, 559, A14
- Liu Z., et al., 2016, *Mon. Not. R. Astron. Soc.*, 459, 1602
- Lonsdale C. J., et al., 2003, *Publ. Astron. Soc. Pacific*, Vol. 115, Issue 810, pp. 897-927., 115, 897
- Luo B., et al., 2010, *Astrophys. J. Suppl. Ser.*, 187, 560
- Luo B., et al., 2017, *Astrophys. J. Suppl. Ser.*, 228, 2
- Mauduit J. C., et al., 2012, *Publ. Astron. Soc. Pacific*, Vol. 124, Issue 917, pp. 714 (2012)., 124, 714
- McLure R. J., et al., 2013, *Mon. Not. R. Astron. Soc.*, 428, 1088
- Meneux B., et al., 2009, *Astron. Astrophys.*, 505, 463
- Menzel M. L., et al., 2016, *Mon. Not. R. Astron. Soc.*, 457, 110
- Momcheva I. G., et al., 2016, *Astrophys. J. Suppl. Ser.*, 225, 27
- Nyland K., et al., 2017, *Astrophys. J. Suppl. Ser.*, 230, 9
- Pacaud F., et al., 2006, *Mon. Not. R. Astron. Soc.*, 372, 578
- Pierre M., et al., 2016, *Astron. Astrophys.*, 592, A1
- Pineau F.-X., et al., 2017, *Astron. Astrophys.*, 597, A89
- Ranalli P., et al., 2013, *Astron. Astrophys.*, 555, A42
- Ranalli P., et al., 2015, *Astron. Astrophys.*, 577, A121
- Rosen S. R., et al., 2016, *Astron. Astrophys.*, 590, A1
- Salvato M., et al., 2017, eprint arXiv:1705.10711
- Silverman J. D., et al., 2009, *Astrophys. J.*, 695, 171
- Skelton R. E., et al., 2014, *Astrophys. J. Suppl. Ser. Vol. 214, Issue 2, Artic. id. 24, 49 pp. (2014).*, 214
- Skibba R. A., et al., 2014, *Astrophys. J.*, 784, 128
- Sutherland W., Saunders W., 1992, *Mon. Not. R. Astron. Soc.*, 259, 413
- Tanaka M., et al., 2012, *Publ. Astron. Soc. Japan*, 64, 22
- Ueda Y., et al., 2008, *Astrophys. J. Suppl. Ser.*, 179, 124
- Vaccari M., 2016, Proc. "The many Facet. extragalactic radio Surv. Towar. new Sci. challenges" (EXTRA-RADSUR2015). 20-23 Oct. 2015. Bol. Italy. Online <http://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=267>, id.27
- Watson M. G., et al., 2008, *Astron. Astrophys.*, 493, 339
- Xue Y. Q., et al., 2011, *Astrophys. J. Suppl. Ser.*, 195, 10
- Yang G., et al., 2014, *Astrophys. J. Suppl. Ser.*, 215, 27
- Yang G., et al., 2016, *Astrophys. Journal*, Vol. 831, Issue 2, Artic. id. 145, 20 pp. (2016)., 831
- de la Torre S., et al., 2010, *Mon. Not. R. Astron. Soc. Vol. 409, Issue 2, pp. 867-872.*, 409, 867

This paper has been typeset from a T<sub>E</sub>X/L<sup>A</sup>T<sub>E</sub>X file prepared by the author.