

## The XXL Survey

### VI. The 1000 brightest X-ray point sources<sup>★,★★</sup>

S. Fotopoulou<sup>1</sup>, F. Pacaud<sup>2</sup>, S. Paltani<sup>1</sup>, P. Ranalli<sup>3,4</sup>, M. E. Ramos-Ceja<sup>2</sup>, L. Faccioli<sup>5</sup>, M. Plionis<sup>6,7,3</sup>, C. Adami<sup>8</sup>, A. Bongiorno<sup>9</sup>, M. Brusa<sup>10,11</sup>, L. Chiappetti<sup>12</sup>, S. Desai<sup>13,14</sup>, A. Elyiv<sup>11,16</sup>, C. Lidman<sup>17</sup>, O. Melnyk<sup>19,18</sup>, M. Pierre<sup>5</sup>, E. Piconcelli<sup>9</sup>, C. Vignali<sup>11,10</sup>, S. Alis<sup>20</sup>, F. Ardila<sup>21</sup>, S. Arnouts<sup>22</sup>, I. Baldry<sup>23</sup>, M. Bremer<sup>24</sup>, D. Eckert<sup>1</sup>, L. Guennou<sup>25</sup>, C. Horellou<sup>26</sup>, A. Iovino<sup>27</sup>, E. Koulouridis<sup>5,3</sup>, J. Liske<sup>28</sup>, S. Maurogordato<sup>29</sup>, F. Menanteau<sup>30,31</sup>, J. J. Mohr<sup>13,14,15</sup>, M. Owers<sup>32</sup>, B. Poggianti<sup>33</sup>, E. Pompei<sup>34</sup>, T. Sadibekova<sup>5</sup>, A. Stanford<sup>35</sup>, R. Tuffs<sup>36</sup>, and J. Willis<sup>37</sup>

(Affiliations can be found after the references)

Received 18 September 2015 / Accepted 12 February 2016

#### ABSTRACT

**Context.** X-ray extragalactic surveys are ideal laboratories for the study of the evolution and clustering of active galactic nuclei (AGN). Usually, a combination of deep and wide surveys is necessary to create a complete picture of the population. Deep X-ray surveys provide the faint population at high redshift, while wide surveys provide the rare bright sources. Nevertheless, very wide area surveys often lack the ancillary information available for modern deep surveys. The XXL survey spans two fields of a combined 50 deg<sup>2</sup> observed for more than 6Ms with *XMM-Newton*, occupying the parameter space that lies between deep surveys and very wide area surveys; at the same time it benefits from a wealth of ancillary data.

**Aims.** This paper marks the first release of the XXL point source catalogue including four optical photometry bands and redshift estimates. Our sample is selected in the 2–10 keV energy band with the goal of providing a sizable sample useful for AGN studies. The limiting flux is  $F_{2-10\text{keV}} = 4.8 \times 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$ .

**Methods.** We use both public and proprietary data sets to identify the counterparts of the X-ray point-like sources by means of a likelihood ratio test. We improve upon the photometric redshift determination for AGN by applying a Random Forest classification trained to identify for each object the optimal photometric redshift category (passive, star forming, starburst, AGN, quasi-stellar objects (QSO)). Additionally, we assign a probability to each source that indicates whether it might be a star or an outlier. We apply Bayesian analysis to model the X-ray spectra assuming a power-law model with the presence of an absorbing medium.

**Results.** We find that the average unabsorbed photon index is  $\langle \Gamma \rangle = 1.85 \pm 0.40$  while the average hydrogen column density is  $\log \langle N_{\text{H}} \rangle = 21.07 \pm 1.2 \text{ cm}^{-2}$ . We find no trend of  $\Gamma$  or  $N_{\text{H}}$  with redshift and a fraction of 26% absorbed sources ( $\log N_{\text{H}} > 22$ ) consistent with the literature on bright sources ( $\log L_{\text{x}} > 44$ ). The counterpart identification rate reaches 96.7% for sources in the northern field, 97.7% for the southern field, and 97.2% in total. The photometric redshift accuracy is 0.095 for the full XMM-XXL with 28% catastrophic outliers estimated on a sample of 339 sources.

**Conclusions.** We show that the XXL-1000-AGN sample number counts extended the number counts of the COSMOS survey to higher fluxes and are fully consistent with the Euclidean expectation. We constrain the intrinsic luminosity function of AGN in the 2–10 keV energy band where the unabsorbed X-ray flux is estimated from the X-ray spectral fit up to  $z = 3$ . Finally, we demonstrate the presence of a supercluster size structure at redshift 0.14, identified by means of percolation analysis of the XXL-1000-AGN sample. The XXL survey, reaching a medium flux limit and covering a wide area, is a stepping stone between current deep fields and planned wide area surveys.

**Key words.** catalogs – surveys – galaxies: active – X-rays: general

### 1. Introduction

Supermassive black holes in the centres of galaxies participate in the evolution of their hosts. This is demonstrated by the relationships between their masses and the host galaxy properties, such as bulge luminosity or bulge stellar velocity dispersion (Magorrian et al. 1998; Gebhardt et al. 2000; Ferrarese & Merritt 2000). These relations show that the

phenomena that form the galactic stellar mass and make the black holes grow are connected, although the precise physical mechanism that shapes them is still unclear.

Black hole growth is observable throughout the universe in the objects collectively called active galactic nuclei (AGN). These objects accrete galactic matter in an accretion disk which produces intense radiation before penetrating inside the central black hole. The connection between stellar mass and black hole mass ultimately results from the shared history of star formation and accretion, which follow the same pattern of steady increase until a peak redshift at  $z \sim 1-2$ , followed by a steep decline until present times (e.g. Hopkins et al. 2006a; Silverman et al. 2008; Watson et al. 2009). In addition, star formation and accretion seemingly share the same anti-hierarchical scenario, moving from massive galaxies to less massive galaxies

\* Based on observations obtained with *XMM-Newton*, an ESA science mission with instruments and contributions directly funded by ESA Member States and NASA. Based on observations made with ESO Telescopes at the La Silla and Paranal Observatories under programme ID 089.A-0666 and LP191.A-0268.

\*\* A copy of the XXL-1000-AGN Catalogue is available at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](http://cdsarc.u-strasbg.fr) (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/592/A5>

(Hirschmann et al. 2012). Different versions of negative feedback affecting both star formation and accretion and thereby regulating the growth of stellar and black hole masses have been proposed. Observations of activity in galaxies in the green valley, transiting from the blue cloud to the red sequence, put strong constraints on these mechanisms (Schawinski et al. 2009).

The role of black holes in galaxy evolution calls for a systematic census of AGN in order to study the distribution of their parameters (e.g. luminosity, accretion rate or black hole mass) and their evolution, in order to relate them to the stellar properties. Because of the diversity of the AGN phenomenon, largely due to the presence of obscuring torus on parsec-scales in the immediate vicinity of the AGN that strongly affects their observational properties, there is no unique way to build AGN samples, and each method is subject to some level of incompleteness. Numerous studies of these distributions have therefore been performed using AGN samples selected with different means, such as optical spectroscopy (e.g. Bongiorno et al. 2007; Masters et al. 2012; Ross et al. 2013), X-rays (e.g. Aird et al. 2010; Ueda et al. 2014; Miyaji et al. 2015), infrared (e.g. Han et al. 2012).

Among the different tools to select AGN, X-ray surveys are among the most efficient. With the exception of extended sources, AGN are often X-ray bright, at least above 2 keV, and the vast majority of high-latitude X-ray sources turn out to be AGN down to extremely faint fluxes around  $10^{-17}$  erg s $^{-1}$  cm $^{-2}$  in the 0.5–2 keV band where a significant population of normal galaxies appears (Lehmer et al. 2012). Several important X-ray surveys have been conducted, from the widest all-sky surveys performed by ROSAT (Voges et al. 1999), ASCA (Ueda et al. 2001, 2005; Nandra et al. 2003), BeppoSAX (Fiore et al. 2001; Verrecchia et al. 2007), and MAXI (Ueda et al. 2011) to the deepest small-field surveys like the COSMOS field with *XMM-Newton* (Hasinger et al. 2007) and *Chandra* (Elvis et al. 2009) or the *Chandra* Deep Fields (Giacconi et al. 2001; Alexander et al. 2003; Xue et al. 2011). Thanks to its large collecting area, *XMM-Newton* is able to efficiently cover large sky areas reaching at the same time medium flux depth. For example, the Hard Bright Serendipitous Survey (HBSS) covers 25 deg $^2$  reaching a flux limit of  $F_{4.5-10\text{keV}} = 7 \times 10^{-14}$  erg s $^{-1}$  cm $^2$  (Della Ceca et al. 2004). The XMM-LSS survey (Pierre et al. 2006), a survey with *XMM-Newton* covering contiguously 11 deg $^2$ , is a good compromise between sky area and depth which has provided several results on the environmental properties of AGN (Elyiv et al. 2012; Melnyk et al. 2013; Koulouridis et al. 2014, 2016, Paper XII), showing in particular different behaviors between objects having soft and hard X-ray spectra.

AGN surveys also efficiently probe the large-scale structure of the Universe. The clustering pattern of AGN and its evolution can provide important constraints on the mass of their dark matter halo hosts, leading to somewhat conflicting results depending on the AGN selection (Coil et al. 2007, 2009; Koutoulidis et al. 2013), and shed light on the influence of the environment on the nuclear activity. Furthermore, the dependence of AGN clustering on luminosity can also provide important constraints on the AGN triggering mechanism and on studies of AGN-galaxy coevolution, since different AGN fueling modes make distinct predictions for the environment of galaxies that host AGN (Shankar et al. 2009; Fanidakis et al. 2013b). The AGN triggering mechanism adopted by models of galaxy formation are either major mergers for the most luminous AGN (Di Matteo et al. 2005; Hopkins et al. 2006b; Marulli et al. 2009), for which only weak luminosity dependence on clustering is expected

(Hopkins et al. 2005; Bonoli et al. 2009), or secular disk instabilities for the lowest luminosity AGN (Hopkins & Hernquist 2006; Bournaud et al. 2011), for which luminosity dependence on clustering should be strong. Observational indications of varying strength for such dependencies have been reported in the literature (Plionis et al. 2008; Krumpe et al. 2010; Cappelluti et al. 2010; Koutoulidis et al. 2013; Fanidakis et al. 2013a).

The Ultimate *XMM-Newton* Extragalactic X-ray survey, or XXL (Pierre et al. 2016, Paper I), is an extension of the XMM-LSS survey at the same depth, but covering 50 deg $^2$  in two 25 deg $^2$  fields XXL-N (RA = 02h20, Dec = –5d00) and XXL-S (RA = 23h30, Dec = –55d00). While the XXL Survey has been primarily designed to build a consistent sample of galaxy clusters for cosmology (Pacaud et al. 2016, Paper II), an immediate by-product of the survey is the identification of numerous point sources, most of which will turn out to be AGN. In this paper, number VI of the first XXL release, we introduce the XXL point source catalogue and present the basic properties of the 1000 brightest objects.

The paper outline is as follows. In Sect. 2 we describe the X-ray observations for XXL, namely the source detection, the extraction, and the analysis of the X-ray spectra. While X-rays are a powerful tool used to uncover nuclear activity in galaxies, additional information about the AGN and their galaxy hosts are needed. In Sect. 3 we give an overview of the photometric data sets available in the XXL field assembled from both publicly available and proprietary surveys, and we describe the creation of the photometric catalogue and the counterpart assignment for the X-ray detections. We also give a summary of the photometric redshift estimation (photo- $z$  hereafter), and the use of the Random Forest method to classify our sources.

In Sect. 5 we present the results of the X-ray spectral analysis also as a function of redshift and the median SEDs for our sample. In Sect. 6, we put our sample within the context of the cosmic web, and we present the observed number counts, the determination of the X-ray AGN luminosity function up to redshift  $z = 3$ , and the percolation analysis in 10 Mpc and 25 Mpc. In Sect. 7 we describe the released catalogue that accompanies this paper. Finally, in Sect. 8 we close with our results and the future prospects of the XMM-XXL survey.

## 2. X-ray observations

In this section we describe the X-ray source detection, the definition of the XXL-1000-AGN sample, and the X-ray spectral analysis. We model the source X-ray spectra as a power-law distribution with the presence of an absorbing medium, determining both the photon index ( $\Gamma$ ) and the hydrogen column density ( $N_{\text{H}}$ ).

### 2.1. Source detection and sample selection

The primary goal of XXL is the accurate detection of clusters of galaxies, which appear as extended emission in the 0.5–2.0 keV energy band. To this end, we use the dedicated pipeline described in Pacaud et al. (2006) (XAMIN). As a by-product, the accurate detection of point-like sources is also possible. XAMIN was used previously in the XMM-LSS survey, the pilot of XXL, and was described in detail in Chiappetti et al. (2013). In this work we are using XAMIN 3.3, which uses the latest calibration files for the *XMM-Newton* observations.

XAMIN proceeds in three stages: (1) images from the three EPIC detectors are combined and a smoothed image is obtained using a multiresolution wavelet algorithm tuned to the

**Table 1.** Summary of 1000 brightest XXL AGN detections, fluxes refer to XAMIN 3.3 determined values.

Sample	Area (deg <sup>2</sup> )	$F_{0.5-2\text{keV},\text{min}}$ (erg s <sup>-1</sup> cm <sup>-2</sup> )	$F_{2-10\text{keV},\text{min}}$ (erg s <sup>-1</sup> cm <sup>-2</sup> )	$N_{\text{det}}$	$N_{\text{Xspec}}$	$N_{\text{ctp}}$	$N_{\text{spec-z}}$	$N_{\text{photo-z}}$	$z_{\text{median}}$
XXL-N	26.9	$1.35 \times 10^{-15}$	$4.86 \times 10^{-14}$	558	481	540	339	201	0.621
XXL-S	23.6	$1.83 \times 10^{-15}$	$4.86 \times 10^{-14}$	442	424	432	250	182	0.637
XXL-1000-AGN	50.5	$1.35 \times 10^{-15}$	$4.86 \times 10^{-14}$	1000	902	972	589	383	0.632

**Notes.** Column 1: sample name, Col. 2: area of each XXL field, Cols. 3–4: minimum flux detected in the 0.5–2 keV and 2–10 keV energy band, respectively, Col. 5: number of detected sources, Col. 6: number of sources with X-ray spectra, Col. 7: number of sources with identified optical counterparts, Col. 8: number of sources with spectroscopic redshift, Col. 9: number of sources with photometric redshifts when spectroscopic redshift is not available, Col. 10: median redshift of sample using spec-z when available and photo-z otherwise.

low-count Poisson regime (Starck & Pierre 1998); (2) source detection is then performed on this smoothed image via Sextractor (Bertin & Arnouts 1996) and a list of candidate sources is produced; and (3) a maximum likelihood (ML) fit based on C–statistic (Cash 1979) is performed for each candidate source; only sources with a detection likelihood from the ML fit  $>15$  are considered significant, which corresponds to 0.5 spurious sources within 10' of the field of view (FOV) (Pacaud et al. 2006). This process is performed independently for the soft (0.5–2 keV) and hard (2–10 keV) bands.

The final stage of the catalogue creation is the ingestion of the source list in the database. At this stage the association of the soft and hard bands is performed. The complete procedure is described in Chiappetti et al. (2013); here we briefly describe the main points of the process for completeness. We use a search radius of 10'' to match sources between the two detection bands. We also allow for a source to be below the ML fit  $<15$ , at most in one band. If a source is present in more than one pointing, we favour sources present in good pointings ( $>7$  ks exposure time and background level  $<1.5 \times 10^{-5}$  cts s<sup>-1</sup> pixel<sup>-1</sup>, quality flag 0), over sources originating from pointings with low exposure time ( $>3$  ks exposure time, quality flag 1) and/or<sup>1</sup> high background level (background level  $<4.5 \times 10^{-5}$  cts s<sup>-1</sup> pixel<sup>-1</sup>, quality flag 2). If a source is found in more than one good pointing, we favour the detection that has the smaller off-axis angle.

Finally, in order to define the count-to-flux conversion factors we assume a single power-law spectrum with fixed photon index  $\Gamma = 1.7$  and hydrogen column density  $N_{\text{H}} = 2.6 \times 10^{20}$  cm<sup>-2</sup>. The following conversion factors, corresponding to 1 ct s<sup>-1</sup>, are used: for the PN camera  $1.5 \times 10^{-12}$  erg s<sup>-1</sup> cm<sup>-2</sup> and  $7.9 \times 10^{-12}$  erg s<sup>-1</sup> cm<sup>-2</sup> for the 0.5–2 keV and 2–10 keV energy bands, respectively, while for the MOS cameras the corresponding factors are  $5 \times 10^{-12}$  erg s<sup>-1</sup> cm<sup>-2</sup> and  $23 \times 10^{-12}$  erg s<sup>-1</sup> cm<sup>-2</sup> (see also Table 3 in Chiappetti et al. 2013). Each source is assigned a unique flux in the soft (0.5–2 keV) and hard (2–10 keV) bands by taking the average flux from all respective PN and MOS observations.

Since the XXL survey was designed for cosmological studies using X-ray selected galaxy clusters as probes, the identification of extended X-ray sources is of central importance in our collaboration. The definition of extended sources is based on the 0.5–2 keV energy band and it is detailed in Paper II. Briefly, a detection enters the extended candidate list when it has an extent greater than 5'' and extension likelihood greater than 15. This list is then split into two categories: C1 for objects with extension likelihood greater than 33 and detection likelihood greater

than 32 and C2 for the remaining candidates. The C1 class is mostly free from contamination, while the C2 class is 50% contaminated by spurious detections or blended point-like sources (see Paper II for more details).

In the first release of our catalogue we present the brightest 1000 X-ray point sources, consisting of detections belonging to neither C1 nor C2. By merging the source lists of the two fields, we selected the 1000 brightest X-ray point-like sources based on the 2–10 keV flux estimated by the XAMIN 3.3 and assuming a fixed photon index of  $\Gamma = 1.7$ . We chose to select our sources in the 2–10 keV band since we are mainly interested in AGN. We refer to this sample throughout the paper as XXL-1000-AGN. Even though the sample contains a small number of unidentified objects (2.8%), stars (2.3%), and normal galaxies (3%), the vast majority of the sources have an estimated luminosity of  $L_{2-10\text{keV}} > 10^{42}$  erg s<sup>-1</sup> ( $>90\%$ ), the typical threshold used to separate AGN from starburst galaxies.

In Table 1 we give an overview of the XXL-1000-AGN sample. The minimum flux, as estimated by the pipeline, is  $F_{0.5-2\text{keV}} = 1.35 \times 10^{-15}$  erg s<sup>-1</sup> cm<sup>-2</sup> and  $F_{2-10\text{keV}} = 4.85 \times 10^{-14}$  erg s<sup>-1</sup> cm<sup>-2</sup> for the soft and hard bands, respectively. Column  $N_{\text{Xspec}}$  gives the number of sources with good-quality X-ray spectra (signal-to-noise ratio  $S/N > 3$ ; see Sect. 2.3). Column  $N_{\text{ctp}}$  gives the number of identified counterparts using either optical or near infrared images (see Sect. 4.2). Columns  $N_{\text{spec-z}}$  and  $N_{\text{photo-z}}$  give the number of sources with spectroscopic redshift (spec-z) and photometric redshift (photo-z) determination, respectively. Finally, from  $N_{\text{median}}$  we see that the median redshift of this flux limited sample is approximately 0.63.

## 2.2. Spectral extraction

For each source in the XXL-1000-AGN sample, a circular extraction region and an annulus-shaped background region were computed. The radii were chosen in order to maximise the expected S/N of the spectrum, as described below.

Since XXL is a mosaic of many overlapping *XMM-Newton* pointings, any source may be present on more than one of them and at different off-axis angles. For the spectral analysis, we make use of all pointings. However, our pipeline analyses each pointing individually, which does not guarantee that a detection is available on all pointings for each source. Therefore, the expected total counts  $C$  available for spectral analysis from all exposures were estimated as

$$C = A \times (C_{\text{B,ref}} + C_{\text{CD,ref}}), \quad (1)$$

where  $C_{\text{B,ref}}$  and  $C_{\text{CD,ref}}$  are the counts in the reference exposure for the 0.5–2 and 2–10 keV bands, respectively, and  $A$  is a factor which corrects for the presence of other exposures in addition to

<sup>1</sup> When both conditions of low exposure time and high background level are met, quality level 3 is assigned.



the reference one. The expected total background surface brightness was also defined in the same way.

The  $A$  factor depends on the number of overlapping exposures, on their relative length, and on their relative overlap (through the effective area of the EPIC CCDs, which depends on the off-axis angle<sup>2</sup>). For simplicity, we used  $A = 1.5$  for all sources, which is justified in the following cases which we assumed as representative of sources in the XXL mosaic:

- a source is detected in the reference pointing within 3' of the boresight position, and in a second pointing, with the same length, at 9'–10' off-axis;
- an exposure has been repeated, e.g. because of high background. Therefore, a source is detected in the reference pointing, and then in another one at the same off-axis angle, but the clean time on the non-reference exposure will be shorter.

Had we computed  $A$  for each source individually, the expected variation on  $A$  would only have had a minor effect on the radii of the extraction regions. A non-optimal radius would either have included more background (if the radius had been larger than optimal) or cause some flux loss (if the radius had been smaller than optimal). However, for the bright sources discussed in this paper, such an effect is likely negligible.

The spectral extraction regions were computed using the `autoregions` software (Ranalli et al. 2015) which for every source  $i$

- defines the source regions as circles of radius  $r_i$  and the background regions as annuli of inner and outer radii  $1.5r_i$  and  $2r_i$ , respectively;
- given the expected source counts and background surface brightness, finds the  $r_i$  which maximises the expected S/N of the spectrum;
- checks for neighbour sources. If there are any within 12'', source  $i$  is considered confused and dropped from the sample. Else, for any neighbour  $j$  within 30'' of source  $i$ , it excises a circular area of radius  $r_j$  from both the source and background regions of  $i$ .

The spectra were finally extracted using the `cdms-extract` software (Ranalli et al. 2015), which for any pair of source  $i$  and pointing  $k$  (treating PN, MOS1, MOS2 exposures individually)

- checks if source  $i$  is present in the FOV of  $k$ ;
- checks that no more than 40% of the flux of  $i$  is lost in chip gaps, missing MOS1 CCDs, and borders of the FOV, otherwise the  $(i,k)$  pair is dropped;
- extracts source and background spectra, and computes responses using the standard XMM SAS tools (`evselect`, `rmfgen`, `arfgen`).

### 2.3. Spectral analysis setup

For each detected source we produced their spectral products (PHA and response files) per camera and per exposure as described in the previous section. In addition, a background region was associated to each source and its spectral products were extracted. Thus, we obtain a varying number of source spectra and corresponding background spectra for each source. Because of

the relatively shallow depth of the XXL survey, spectra generally have few counts, even for the brightest sources. This is especially true for background spectra, meaning that they are subject to large relative uncertainties. We therefore do not subtract the background spectra from the source+background spectra, but we use them as source-free fields to constrain the background parameters.

Following the approach of Leccardi & Molendi (2008), we model the background in all observations (source and source-free) using two components: X-ray background (XB) and non-X-ray background (NXB). The XB itself consists of three components: the cosmic X-ray background (CXB), the Local Bubble emission, and the Galactic Halo (see Fig. B.1 in Leccardi & Molendi 2008). We model the first with a power law and the last two with a thermal plasma emission (`apec` model in `xspec`). The slope of the CXB component is fixed to the value of 1.46 (De Luca & Molendi 2004), but its normalisation is allowed to vary for any given sky position (source and corresponding source-free regions are fixed to the same normalisation). The shape of the Local-Bubble and Galactic Halo emissions are fixed, but their normalisations are determined using ROSAT data. We extracted from HEASARC's X-Ray Background Tool<sup>3</sup> the ROSAT spectra from regions covering the XXL-N and XXL-S areas (Snowden et al. 1997). The normalisations of the two components were fixed by fitting the emission models (including CXB) over these two areas. Depending on the instrument, the NXB consists of a combination of several emission lines typically resulting from internal fluorescence and continuum components phenomenologically represented with a combination of power laws and/or a blackbody. The CXB and NXB models and parameters were derived from detailed studies of the *XMM-Newton*/EPIC background (De Luca & Molendi 2004; Leccardi & Molendi 2008; Kuntz & Snowden 2008; Eckert et al. 2014).

Considering the very complex data sets, background models, and the large number of sources that prevent manual analysis of each source separately, we adopt here a Bayesian approach. We avoid fitting the models in the traditional sense with the goal of providing point estimates for all model parameters, which would include the parameters of the CXB and NXB. Instead, we build credible intervals for the parameters of interest and marginalise over the parameters of the CXB and NXB, which are considered as nuisance parameters for this work. Non-informative priors are selected for all parameters (using the Jeffreys prior in the case of norm parameters). We determine the Bayesian evidence using the nested sampling algorithm (Skilling 2004). For this purpose, we use the BXA `xspec` interface (Buchner et al. 2014) of the MultiNest implementation (Feroz & Hobson 2008; Feroz et al. 2009, 2013) of this algorithm. MultiNest directly builds a posterior sample from the full distribution, which we marginalise over the nuisance parameters to recover the joint posterior distribution of the parameters of interest (the source's parameters).

We perform a fit appropriate for AGN, assuming a power law absorbed by our Galaxy, allowing the presence of absorption in the host galaxy at the redshift of the source (`phabs*zphabs*powerlaw` in `xspec`); therefore, the redshift of the host galaxy absorber needs to be determined (see Sect. 4). It is either fixed to the spectroscopic redshift when available, or let free in the 68% confidence interval of the photometric redshift. The full 0.5–12 keV energy range is used, and the allowed range for the intrinsic photon index is  $\Gamma = [1, 3]$ , while

<sup>2</sup> See the *XMM-Newton* Users' Handbook, Sect. 3.2.2.2 and Fig. 13; [http://xmm.esa.int/external/xmm\\_user\\_support/documentation/uhb/effareaoffaxis.html](http://xmm.esa.int/external/xmm_user_support/documentation/uhb/effareaoffaxis.html)

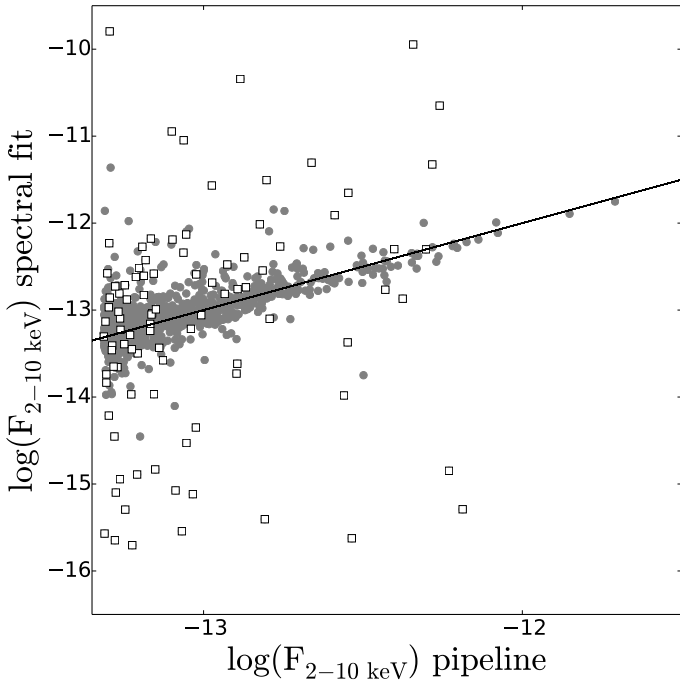
<sup>3</sup> <http://heasarc.gsfc.nasa.gov/cgi-bin/Tools/xraybg/xraybg.pl>



**Table 2.** X-ray spectral quality classes defined based on the XMM pointing quality and the signal-to-noise ratio (S/N) of the spectrum.

Quality class		N	$F_{XAMIN}/F_{fit}$			Flag	PN counts			MOS1 counts			MOS2 counts		
Pointing	S/N		med	mean	$\sigma$		min	max	med	min	max	med	min	max	med
0	>6	547	1.00	1.0	0.3	1	25	11194	268	13	2626	105	6	2651	107
0/1/2	>4	281	1.01	1.1	0.5	2	8	1879	103	3	244	43	6	415	40
0/1/2	3–4	75	1.07	2	3	3	6	246	43	2	49	23	4	81	20
0/1/2	<3	96	0.78	52	191	4	3	177	16	1	34	6	1	55	5
NoData	–	1	–	–	–	5	–	–	–	2	2	–	1	1	–

**Notes.** The XMM pointing quality classes are: 0 = good quality (>7 ks exposure time and  $<1.5 \times 10^{-5}$  cts  $s^{-1}cm^{-2}$  background level), 1 = low exposure time (>3 ks), 2 = high background  $<4.5 \times 10^{-5}$  cts  $s^{-1}cm^{-2}$ ). Pointings with quality class 3 corresponding to both low exposure time and high background were not used for spectrum extraction.



**Fig. 1.** Comparison between the 2–10 keV flux estimated by XAMIN 3.3 assuming a power-law spectrum with  $\Gamma = 1.7$ ,  $N_H = 2.6 \times 10^{20} cm^{-2}$ , and the spectral fit value (grey points). The black line shows the one-to-one relation. The open squares denote the sources that have low-quality X-ray spectra, (flag = 4, see Table 2).

for the intrinsic absorption  $\log N_H = [19, 24.5]$ . We include the 10–12 keV range, as it efficiently constrains the continuum part of the NXB. We rebin all spectra individually so that they have at least three counts in each bin to avoid bins with zero counts. The Galactic hydrogen column density was determined from the hydrogen maps of [Dickey & Lockman \(1990\)](#). The resulting fixed Galactic absorption, modelled with the `phabs` `xspec` model, is applied to all components except the Local Bubble emission and the NXB. All components except the source are scaled to the area of the extraction region. As the NXB does not have an astrophysical origin, we did not apply the effective area correction (ARF) on the NXB models; however, we applied the redistribution matrix (RMF) to take into account the instrument resolution and redistribution.

In the released catalogue we provide a flag (see Table 2) showing the combined quality of the X-ray pointing and the S/N of the X-ray spectrum. From Table 2, we see that sources with S/N above four have the highest quality spectra. Figure 1 shows the comparison between the 2–10 keV flux determined

by XAMIN 3.3 and the spectral fit. The black line shows the one-to-one relation. The scatter of the points around the one-to-one relation is due to (1) the combined quality of the XMM pointing and the extracted spectrum and (2) the real spectrum of the source. Both are expected to cause a scatter of the estimated flux around the one-to-one relation since lower quality pointings introduce noise in the spectra, while the real spectrum of the observed sources will deviate from the power law of constant photon index assumed in the pipeline. In Table 2, we summarise the median, mean, and standard deviation of the ratio  $F_{2-10 keV, pipe}/F_{2-10 keV, fit}$ . The scatter between  $F_{2-10 keV, pipe}$  and  $F_{2-10 keV, fit}$  actually seems to increase as we move to fainter sources, but no systematic offset is easily observed. The open squares show sources with  $S/N < 3$  (flag = 4). The challenging nature of flag = 4 sources is the result of a combination of factors including the low quality of the X-ray spectrum, the difficulty in constraining the background, and the uncertainty on the redshift of the source. We suggest avoiding the use of these sources in a source-by-source scientific analysis based on the X-ray spectral fits. These sources are not used in Sect. 5.1 where the results of the X-ray spectral analysis are discussed in detail.

In what follows, unless otherwise stated, we also consider sources with flag 3 since estimated flux shows a slightly larger scatter ( $\sigma = 3$ ) than the pipeline, but not a significant offset (flux ratio median = 1.07). In Table 2 we also present the observed counts for PN, MOS1, and MOS2 split according to the quality of the X-ray spectra. For the highest quality spectra (flag = 1) the PN counts range from 25 to 10<sup>4</sup> counts, while the median observed counts are 268, 105, and 107 for PN, MOS1, and MOS2, respectively. Only one source belongs to the category with flag = 5, for which it was not possible to extract spectral data on any of the pointings. The sample considered in the rest of the paper (flag < 4) has median counts  $C_{PN} = 175$ ,  $C_{MOS1} = 70$ , and  $C_{MOS2} = 65$ . Further results of the X-ray spectral analysis are discussed in Sect. 5.1.

### 3. Multiwavelength observations

The XXL fields benefit from ancillary photometric observations ranging from the ultraviolet to infrared wavelengths consisting of both private and public surveys. A summary of all observations available across the electromagnetic spectrum targeting the XXL field is given in Paper I, Table 2. Here, we describe briefly the data sets used for this work. In Tables 3 and 4 we gather the survey information we used to construct the spectral energy distributions (SEDs) of sources in the XXL-N and XXL-S, respectively. The limiting magnitudes refer to the third quantile of the aperture magnitude distribution.

**Table 3.** XXL-N: ancillary photometric data set used in this work and measured limiting magnitudes.

Telescope	Survey	Version	Limiting magnitude					
GALEX	AIS	GR6/7	FUV	NUV				
	DIS		22.18	22.47				
	GI		24.57	24.18				
			23.51	23.69				
CFHT	W1	T0007	<i>u</i>	<i>g</i>	<i>r</i>	<i>i</i>	<i>y<sup>b</sup></i>	<i>z</i>
	D1		24.96	25.30	24.74	24.39	24.50	23.40
	WA	PI	26.26	26.62	26.34	26.08	25.61	25.21
	WB		–	24.82	24.49	–	–	22.86
	WC		–	24.98	24.52	–	–	23.55
SDSS	DR10	–	24.88	–	–	–	–	
			20.73	19.77	19.92	19.96	–	20.13
VISTA	VHS	DR2	<i>z</i>	<i>Y</i>	<i>J</i>	<i>H</i>	<i>K</i>	
	VIKING	DR1	–	20.85	20.98	20.70	20.27	
	VIDEO	DR3	22.81	22.10	21.50	21.36	21.25	
WIRcam		PI	25.45	24.65	24.65	24.25	23.90	
UKIDSS	DXS	DR10	–	–	–	–	22.20	
	UDS		–	–	22.37	–	22.17	
			–	–	24.95	24.37	24.64	
IRAC		PI	3.6 $\mu$ m	4.5 $\mu$ m				
			21.50	21.34				
WISE	ATLAS	ALLWISE	<i>W1</i>	<i>W2</i>	<i>W3</i>	<i>W4</i>		
			20.38	20.35	18.85	17.50		

**Notes.** Magnitudes given in AB. As limiting magnitude we quote the third quantile of the respective magnitude distribution. CFHT replacement *i*-filter.

**Table 4.** XXL-S: ancillary photometric data set used in this work and measured limiting magnitudes.

Telescope	Survey	Version	Limiting magnitude			
GALEX	AIS	GR6/7	FUV	NUV		
	MIS		22.42	22.82		
	GII		–	23.61		
			23.92	24.03		
BCS DECAM		PI	<i>g</i>	<i>r</i>	<i>i</i>	<i>z</i>
			24.14	24.06	23.23	21.68
			25.73	25.78	25.6	24.87
VISTA	VHS	DR2	<i>J</i>	<i>H</i>	<i>K</i>	
			21.1	20.77	20.34	
IRAC	SSDF		3.6 $\mu$ m	4.5 $\mu$ m		
			21.5	21.45		
WISE	ATLAS	ALLWISE	<i>W1</i>	<i>W2</i>	<i>W3</i>	<i>W4</i>
			20.32	20.35	18.68	17.26

**Notes.** Magnitudes given in AB. As limiting magnitude we quote the third quantile of the respective magnitude distribution.

### 3.1. XXL-N and XXL-S

Observations that are in common for the two XXL fields:

- *GALEX* Galaxy Evolution Explorer<sup>4</sup> (*GALEX*) released the final mission catalogue (GR6/7) with full sky coverage in the far ultraviolet ( $\lambda_{\text{FUV}} = 1516 \text{ \AA}$ ) and near ultraviolet ( $\lambda_{\text{NUV}} = 2267 \text{ \AA}$ ). We retrieved the relevant images through the Mikulski Archive for Space Telescopes<sup>5</sup> (MAST). The full XXL-N field is covered by one or more of the following surveys: the All-Sky Imaging Survey (AIS), the

Deep Imaging Survey (DIS), and guest investigator programs (GI). The DIS and GI surveys mainly focus on the XMM-LSS area. The XXL-S field is fully covered by *GALEX* GR6/7 with AIS, Medium Imaging Survey (MIS), and/or GI programs.

- *VISTA* Three public European Southern Observatory (ESO) large programme surveys have observed the XXL field in the near infrared (*z*, *Y*, *J*, *H*, *K* filters: 0.8–2.1  $\mu$ m) with the Visible and Infrared Survey Telescope for Astronomy<sup>6</sup> (*VISTA*) in various depths. XXL-N is covered by all three

<sup>4</sup> <http://www.galex.caltech.edu/>

<sup>5</sup> <http://archive.stsci.edu/>

<sup>6</sup> <http://www.eso.org/public/teles-instr/surveytelescopes/vista/>

surveys: VISTA Hemisphere Survey (VHS, PI: R., McMahon), VISTA Kilo-degree Infrared Galaxy Survey (VIKING, PI: W. Sutherland), and VISTA Deep Extragalactic Observations Survey (VIDEO, PI: M. Jarvis), while the XXL-S is covered only by the VHS survey.

- *IRAC* Both XXL fields have been targeted with the Infrared Array Camera (IRAC) on board the *Spitzer* Space Telescope<sup>7</sup> providing imaging at 3.6  $\mu\text{m}$  and 4.5  $\mu\text{m}$  over the whole area (XXL-N; PI: M. Bremer, XXL-S; SPT-*Spitzer* Deep Field, SSDF, Ashby et al. 2014).
- *WISE* The Wide-field Infrared Survey Explorer<sup>8</sup> (WISE) is an all-sky mission observing the sky from 3.4  $\mu\text{m}$  to 22  $\mu\text{m}$ . In this work we use the ALLWISE data release<sup>9</sup>, which provides imaging for the whole XXL field in all four WISE filters.

### 3.2. XXL-N

Observations covering exclusively the XXL-N area:

- *CFHT* The Canada-France-Hawaii Telescope<sup>10</sup> (CFHT) has observed the parts of the XXL-N area as part of the CFHT-Legacy Survey<sup>11</sup> (CFHTLS) in two configurations a) wide area (*WI* field) and b) deep field (*D1* field) observed with MEGACam, providing imaging in the filters:  $u, g, r, i/y, z$  (3800–8800 Å). In this work we use the T0007 data release downloaded from Terapix<sup>12</sup>. We complemented the coverage of CFHT with three additional 1 deg<sup>2</sup> fields to the north of *WI*. The *WA* and *WB* fields were observed in the  $g, r, z$  filters while the *WC* field was observed in the  $g$  filter (PI: M. Pierre). Additionally, XXL-N has been observed with the WIRcam camera on CFHT in the  $K_s$  band (2.2  $\mu\text{m}$ , VIPERS-Multi-Lambda Survey). Details on the observations and data reduction are provided in Moutard et al. (2016a,b).
- *SDSS* The Sloan Digital Sky Survey<sup>13</sup> (SDSS) provides observations in the whole XXL-N area in the filters:  $u, g, r, i, z$  (3800–8800 Å). In this work we use the DR10 data release (Ahn et al. 2014).
- *UKIDSS* The UKIRT Infrared Deep Sky Survey<sup>14</sup> (UKIDSS, Dye et al. 2006) has two fields overlapping with XXL-N: the Ultra Deep Survey (UDS), and the Deep Extragalactic Survey (DXS) targeting the XMM-LSS area in  $J, H, K_s$  and  $J, K_s$  bands, respectively. Here we use the DR10 data release downloaded from the WFCAM Science Archive<sup>15</sup>.

### 3.3. XXL-S

Observations covering exclusively the XXL-S area:

- *BCS* The Blanco Cosmology Survey<sup>16</sup> (BCS; Desai et al. 2012) has targeted an area of 80 deg<sup>2</sup> which overlaps with

<sup>7</sup> <http://ssc.spitzer.caltech.edu/>

<sup>8</sup> <http://wise.ssl.berkeley.edu/>

<sup>9</sup> <http://wise2.ipac.caltech.edu/docs/release/allwise/>

<sup>10</sup> <http://www.cfht.hawaii.edu/>

<sup>11</sup> <http://www.cfht.hawaii.edu/Science/CFHTLS/>

<sup>12</sup> [http://terapix.iap.fr/rubrique.php?id\\_rubrique=268](http://terapix.iap.fr/rubrique.php?id_rubrique=268)

<sup>13</sup> <http://www.sdss.org/>

<sup>14</sup> The UKIDSS project is defined in Lawrence et al. (2007). UKIDSS uses the UKIRT Wide Field Camera (WFCAM, Casali et al. 2007). The photometric system is described in Hewett et al. (2006), and the calibration is described in Hodgkin et al. (2009). The pipeline processing and science archive are described in Hambly et al. (2008).

<sup>15</sup> <http://surveys.roe.ac.uk/wsa/>

<sup>16</sup> <http://www.usm.uni-muenchen.de/BCS/>

the XXL-S field using the MOSAIC II imager at the Cerro Tololo Inter-American Observatory<sup>17</sup> (CTIO). The observations cover the  $g, r, i, z$  bands (4850–9000Å). The images used in this work were analysed as described in Menanteau et al. (2009).

- *DECAM* The Dark Energy Camera<sup>18</sup> (DECAM) is the successor of the MOSAIC II camera at CTIO. We observed the XXL-S field (PI: C. Lidman) in the  $g, r, i, z$  bands (4850–9000Å), but at deeper depth compared to BCS (see Table 4). The details of the observations and data reduction are in Gardiner et al. (in prep.). The stacked images used in this work were performed as described in Desai et al. (2012).

We retrieved all the images available for these surveys. All the optical and near-infrared images were rescaled to zero-point 30 for consistency and ease during the photometry extraction. GALEX, IRAC, and WISE data already have homogeneous zero-points (per survey) so there was no need for rescaling.

## 4. Multiwavelength catalogue

The photometry of publicly available catalogues is usually a generic extraction which suffers from false detections, non-detected real objects, blended photometry, etc. To make the most out of the existing observations, we performed our own photometric extraction. Since AGN are rare objects (we have about 10<sup>4</sup> X-ray detections in the 50 deg<sup>2</sup> of XXL compared to ~10<sup>6</sup> optically detected galaxies in the same area), it is crucial that the photometry is 1) accurate, in terms of photometric calibration; and 2) complete, in terms of detected sources in each filter. At the same time, the computation of photometric redshift for AGN requires special treatment, particularly true in the case of bright X-ray sources (Salvato et al. 2011). The detailed procedure of the photometry pipeline and the photometric redshift estimation is outside the scope of this work and it will be presented in detail in a companion paper (Fotopoulou et al., in prep.).

### 4.1. Photometry

Similarly to the *XMM-Newton* pointings, the ancillary multi-wavelength observations are organised in tiles corresponding to the FOV of each telescope. Since the photometry is not obtained simultaneously, different observing conditions will cause variations to the PSF and the zero-point of each image. We extracted the photometry per filter and per tile for each survey. Each catalogue is then corrected for Galactic extinction according to the Schlegel et al. (1998) maps. The aperture magnitudes are corrected to total from the difference between the 3'' aperture magnitude and the mag<sub>auto</sub> estimated with SExtractor (Bertin & Arnouts 1996). For each survey and filter (see Tables 3 and 4) we built a source catalogue keeping track of the PSF size, aperture correction, and zero-point calibration for each tile separately.

In order to merge the detections with a given survey, it is necessary to remove duplicate detections due to overlapping tiles. For each tile, we identify the distance of the nearest neighbour for each source. We then concatenate the individual tile catalogues and search for sources that lie within that radius. If there are positive matches, they are considered as one source

<sup>17</sup> <http://www.ctio.noao.edu/noao/>

<sup>18</sup> <http://www.ctio.noao.edu/noao/content/dark-energy-camera-decam>



and the best is kept in the catalogue. We consider as “best” the source with the smallest photometric uncertainty. With this procedure we create the catalogue of primary detections for each filter.

In order to create the multiwavelength catalogue, we impose the condition that a source must be detected in at least two filters. Since our optical and near-IR images have a typical seeing of  $0.8''$  it is sufficient to use positional matching to associate the sources detected on our images. We adopted a matching radius  $0.7''$ . From this step we identify matched sources and single-band detections. As a second step, we match GALEX, IRAC, and WISE catalogues within  $1''$  of the multiwavelength catalogue from the previous step and to the individual single-detection catalogues.

#### 4.2. X-ray counterpart association

Source association is trivial when the image PSF is small ( $<1''$ ) for which a positional match usually suffices. This is the case for our optical and near-IR. However, in order to ensure a correct association between X-ray detections and sources detected in the optical we must employ a statistical approach. We chose to use the method of Sutherland & Saunders (1992) which has been used frequently in the literature. The basis of the method lies in the fact that X-ray sources are rare events; bright optical sources are also rare events, so the observation of an X-ray source and a bright optical source in the same region of the sky is considered a non-random event.

Since the available optical observations are highly inhomogeneous we have to search for counterparts in multiple catalogues. We start from the brightest catalogues (i.e. SDSS), which usually cover a larger area on the sky, and we use successively surveys with deeper photometric observations available (i.e. CFHTLS-D1). Following the procedure used in previous XMM surveys (Brusa et al. 2005, 2007; Rovilos et al. 2011) we use the optical  $i$  band as a starting point and successively we use the  $K$  band, and  $3.6\ \mu\text{m}$  in cases where a counterpart is not found. With this method, we are able to identify 540 counterparts out of 558 X-ray detected sources for the XXL-N and 432 out of the 442 X-ray detected sources for the XXL-S. In total we reach a 97.2% identification rate for the XXL-1000-AGN.

#### 4.3. Redshifts and classification

A significant effort has been made to gather a large number of spectroscopic redshifts for the XXL sources. We use a combination of publicly available data (e.g. SDSS<sup>19</sup>) and data obtained through collaborations with other consortia, namely with the VIMOS Public Extragalactic Redshift Survey<sup>20</sup> (VIPERS; Guzzo et al. 2014) and Galaxy and Mass Assembly<sup>21</sup> (GAMA; Driver et al. 2009, 2011; Hopkins et al. 2013; Liske et al. 2015). Additionally, through the efforts of our collaboration, the XXL benefits from more than 20 000 high-quality spectra. As is the case for the photometric observations, the XXL-N field also has a better spectroscopic coverage (Adami et al., in prep). For the XXL-N there is a large compilation of spectra for normal galaxies, while in the XXL-S through targeted observations pursued by our collaboration, more than 3000 spectra were obtained to study X-ray selected AGN Lidman et al. (2016, Paper XIV).

<sup>19</sup> <https://www.sdss3.org/dr10/>

<sup>20</sup> <http://vipers.inaf.it/>

<sup>21</sup> <http://www.gama-survey.org/>

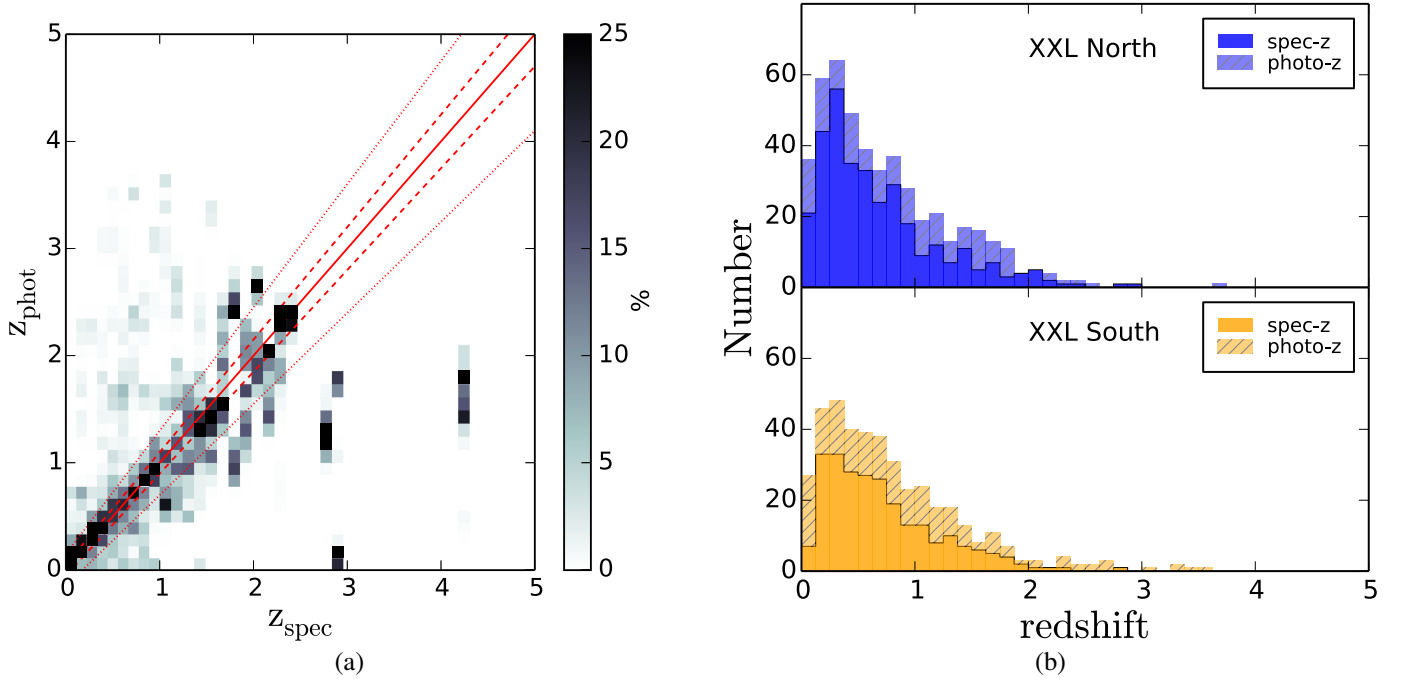
The XXL-1000-AGN sample has 540 high-quality spectroscopic redshifts. For the remaining 44.4% of the sample, we computed photometric redshifts adopting the following procedure. Building on the legacy of the COSMOS survey on photometric for AGN described in detail in Salvato et al. (2009, 2011); instead of three classes – namely normal galaxies, AGN, quasi-stellar objects (QSO), – we increased the number to five classes: passive, star forming, starburst, AGN, QSO. We used the COSMOS templates to estimate a photometric redshift solution for a given class. More specifically, we used templates 1–7 from Ilbert et al. (2009) to describe the passive class, templates 8–18 for the star forming class, and templates 19–31 for the starburst class. For the active galaxies from Salvato et al. (2009) we used the pure QSO templates with their extension to the UV to describe the QSO class and the hybrid templates to describe the AGN class. The same set-up of extinction laws and redshift step was used as stated in these papers. Namely, we searched for the best fit model within each class in the redshift range  $z = (0, 6)$  with  $z_{\text{step}} = 0.01$ . The extinction laws of Calzetti et al. (2000) was applied to the star forming and starburst classes, while the extinction law of Prevot et al. (1984) was applied to the QSO and AGN classes. Therefore, for each object we have five photo- $z$  estimates with the corresponding probability distribution function (PDF). According to the COSMOS schema, X-ray flux, morphology (point-likeness), and variability are used to distinguish between the three categories: normal, AGN, QSO. Here instead we use an extended attribute set consisting of 84 attributes:

- All colour combinations using the bands,  $g, r, i, z, J, H, K, 3.6\ \mu\text{m}$ ;
- FWHM scaled to the corresponding PSF;
- Half-light radius scaled to the corresponding PSF;
- X-ray flux;
- Hardness ratio;
- $\chi^2$  values of the five model categories fitted to the data.

For the classification we used a Random Forest classifier (Breiman 2001) through the `sci-kit.learn` package in Python. We used first the spectroscopic sample from the whole XXL field keeping only sources with high-quality spec- $z$  determination. We split the sample into three equal parts: training, test, validation. Each sample consists of about 8000 galaxies (including AGN/QSO). The training sample is used by the classifier to create the decision trees and the test sample is used again by the classifier to estimate the quality of the classification. In order to label our sources as belonging in one of the five categories, we use the proximity of the photo- $z$  solution to the true spec- $z$  value. The validation sample is never seen by the classifier, and it is used to assess the performance once the Random Forest is created. The forest can be easily saved and applied to all sources in our field.

We used two more Random Forest classifiers to assign probability for a source to be (1) a star or (2) an outlier. Going through the same procedure as described above of class assignment-training-testing, we used the spectroscopically identified stars for the star classifier (1600 objects) and the sources that had photo- $z$  solution in all five categories  $|z_{\text{spec}} - z_{\text{phot}}| > 0.15 \cdot (1 + z_{\text{spec}})$  as outliers (3000 objects). These classifiers are able to predict the correct class for about 98% and 96% of the respective test sample of the star and outlier class.

Figure 2a shows the photo- $z$  performance of the XXL-1000-AGN sample, using the validation sample, i.e. the sample that was never seen by our classifier during training. In order to visualise all the information included in the redshift PDF, we create spec- $z$  slices,  $\Delta z_{\text{spec}} = 0.1$ . We then take the PDFs of all sources



**Fig. 2.** **a)** Spectroscopic versus photometric redshift for the 339 XXL-1000-AGN sources not included in the Random Forest training. The plot is a 2D histogram of the stacked photo- $z$  probability distribution functions and the colourbar shows the probability enclosed in each cell for a given spectroscopic redshift slice ( $\Delta z = 0.1$ ). The solid red line is the diagonal, the dashed and dotted lines are  $0.10 \times (1 + z_{\text{spec}})$  and  $0.15 \times (1 + z_{\text{spec}})$  respectively. Sources outside the dotted line are considered outliers. **b)** Redshift distribution of XXL-N (blue) and XXL-S (orange) fields. The solid histograms show the distribution of spectroscopic redshift, the stacked hatched bars show the distribution of additional photometric redshift when spectroscopic redshifts are not available. The combined height shows the total redshift distribution in each field.

that fall in a specific bin. Using the same binning  $\Delta z_{\text{phot}} = 0.1$  on the photo- $z$  axis, we integrate the PDF in a given bin. The integral gives the probability of each source falling in that particular bin cell. Since our sample consists of independent sources, we add the calculated probabilities of each source within a given bin. We then divide the cell value by the number of sources present in the spec- $z$  slice. The grey scale colourbar shows the percentage of this probability. In the ideal case of perfect photometric redshift estimates, our cells would have 100% probability along the diagonal and 0% elsewhere. This is not achievable using only broadband photometry. Instead, as seen in Fig. 2a, there are areas above and below the diagonal with low probability values up to  $z = 2.5$ .

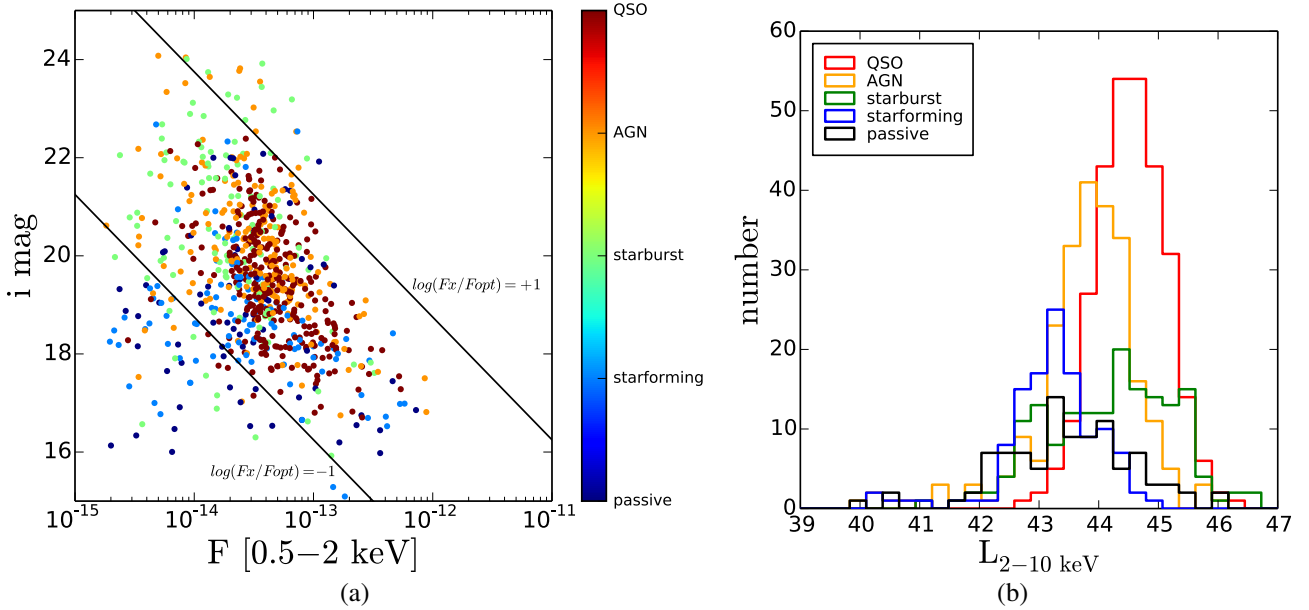
To compare these data with previous results we use the normalised median absolute deviation (NMAD) as an estimator of the accuracy defined as  $\sigma_{\text{NMAD}} = 1.48 \times |z_{\text{phot}} - z_{\text{spec}}| / (1 + z_{\text{spec}})$ . The accuracy in the validation sample of 339 sources is  $\sigma_{\text{NMAD}} = 0.095$  with  $\eta = 28.3\%$  catastrophic outliers (i.e. percentage of sources with  $|z_{\text{phot}} - z_{\text{spec}}| / (1 + z_{\text{spec}}) > 0.15$ ). As expected, very luminous X-ray sources with photometry observed not simultaneously in all bands is affected by intrinsic variability, which in turn makes the photometric redshift estimation challenging. Once we exclude sources that have a probability of being outliers of more than 0.2% estimated using our classifier, the accuracy is  $\sigma_{\text{NMAD}} = 0.071$  with  $\eta = 16.3\%$ , but the sample has been reduced by half.

Figure 2b shows the redshift distribution of the XXL-1000-AGN in the XXL-N (top) and the XXL-S (bottom). The solid histograms represent the spec- $z$  sources, while the stacked hatched bars show the additional photo- $z$ . We see that the photo- $z$  distribution follows loosely the spec- $z$  distribution.

#### 4.4. Classification results

Our classifier returns the optimal photometric class per source for 70% of the sources in the test sample. We note that the scope of the current set-up is not to identify correctly the exact nature of the source, but rather to identify the class that will give us the optimal photo- $z$  solution. Therefore, misclassification between normal galaxy classes would not necessarily cause a dramatic effect on the photo- $z$  estimates, if (a) enough photometric bands are available and (b) characteristic features such as the Balmer break are present in the SED. Figure 3a shows the  $f_x/f_{\text{opt}}$  diagram (Maccacaro et al. 1988; Lehmann et al. 2001; Hornschemeier et al. 2003; Brusa et al. 2005; Fotopoulou et al. 2012). The AGN and QSO are expected to lie largely between  $\log(f_x/f_{\text{opt}}) = \pm 1$ , calculated as  $\log(f_x/f_{\text{opt}}) = \log f_x + I/2.5 + 5.5$  (black lines). The colour-coding denotes the class assigned by the Random Forest. We see that the AGN (orange) and QSO (red) lie between the two lines, while at the same time the galaxies labelled as passive objects (dark blue) lie below the  $\log(f_x/f_{\text{opt}}) = -1$  line and star forming (cyan) and starburst (green) objects appear to have lower X-ray fluxes than AGN and QSO in agreement with the observation of Salvato et al. (2011) in the COSMOS survey. This diagram gives us additional confidence on the classifier's results.

In Fig. 3b we plot the histogram of the intrinsic  $L_{2-10\text{keV}}$ , splitting the sample according to the Random Forest class assignment. To calculate the  $L_{2-10\text{keV}}$  we used the absorption corrected flux estimated by the X-ray spectrum with the corresponding photon index using spectroscopic redshift when available or photometric redshift otherwise. We observe that the two dominant classes are indeed QSO (red) and AGN (orange). Even by using information such as broadband colours and morphology estimates, our classifier correctly identified QSOs, which a



**Fig. 3. a)**  $F_X/F_{opt}$  diagram for the XXL-1000-AGN. The majority of the objects classified as quasi-stellar object (QSO)/AGN (red and orange dots) fall between the  $\log(F_X/F_{opt}) = \pm 1$  lines, while passive objects (dark blue) and stars are gathered in the bottom left corner of the plot. **b)** X-ray luminosity per class. QSO classified objects have the highest 2–10 keV luminosity. The colour-coding indicates the photo- $z$  class assigned by the Random Forest in both plots: QSO – red, AGN – orange, starburst – green, star forming – blue, passive – black.

posteriori we find to be high-luminosity objects ( $\log L_{xpeak} \sim 44.5$ ). Similarly, AGNs show  $\log L_{xpeak} \sim 43.9$ , while the star forming and passive galaxies, show  $\log L_{xpeak} \sim 43.0$ . Interestingly, the starburst population shows a broad distribution in  $\log L_x$  extending up to high luminosities. This population includes objects that show strong absorption in the blue part of the SED, while at the same time enhanced infrared emission. Therefore, they are indeed AGN as we concluded by the X-ray luminosity, but the photometric redshift is best constrained by a starburst galaxy template.

In Appendix A we present a subsample of the XXL-1000-AGN sources. For each of the Random Forest categories – QSO, AGN, starburst, star forming, and passive – we show the X-ray spectrum, the multiwavelength SED along with the best fit model for the photo- $z$  solution (red line) and the star model (grey line). Additionally, we show a single filter image and a false colour image. The green dashed circle is centred at the X-ray position, while the chosen counterpart is marked with a red circle.

## 5. XXL-1000-AGN multiwavelength properties

In this section we discuss further the XXL-1000-AGN X-ray spectral analysis results and SED fitting. We show that on average, the XXL-1000-AGN sample comprises unabsorbed sources both in the X-rays ( $\log N_H \sim 10^{21} \text{ cm}^{-2}$ ) and the optical ( $E(B-V) < 0.1$ ), with photon index  $\langle \Gamma \rangle = 1.85$ , and average intrinsic luminosity  $\log L_x \sim 44$ . We also provide a recipe for creating an expected  $N_H$  distribution, given a hardness ratio value.

### 5.1. X-ray spectral properties

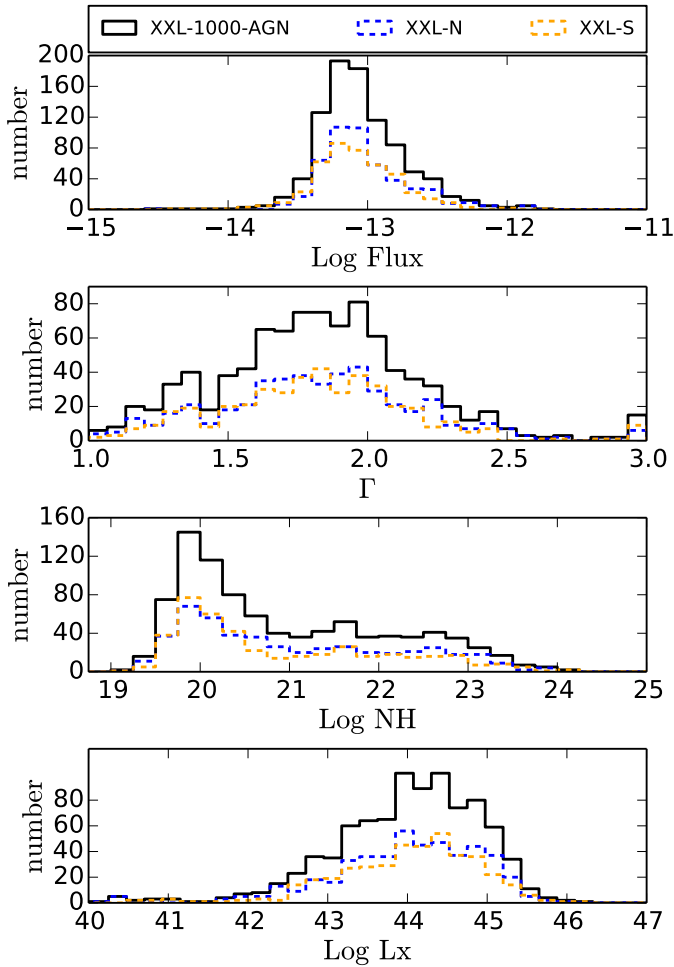
The selection of the XXL-1000-AGN sample was based on the 2–10 keV estimated by our pipeline, where the spectrum of a source is considered to be a power law with a universal slope of  $\Gamma = 1.7$ . With the brightest sample we are in a position to more accurately determine the shape of the spectrum as described in

Sect. 2.3. For this work we adopt a power law with absorbing medium as a sufficient description of the X-ray spectrum. We find that there are ten sources that would clearly require an extra modelling component for the soft X-ray emission, but seven out of these ten sources are associated with stars based on visual inspection of the SEDs and optical images.

In Fig. 4 we show the distribution of the X-ray spectral parameters determined by the fitting (top to bottom),  $\log$  Flux,  $\Gamma$ ,  $\log N_H$ , and the distribution of the intrinsic luminosity  $L_x$ . The photon index  $\Gamma$  is in the range [1.0, 3.0] during the fitting. We find an average value of  $\langle \Gamma \rangle = 1.85 \pm 0.4$ , in agreement with previous observations (Mainieri et al. 2002, 2007; Tozzi et al. 2006; Buchner et al. 2014; Corral et al. 2015). The relative uncertainties associated with our fitting results vary as a function of X-ray spectral quality flag (as defined in Sect. 2.3). We find that the 2–10 keV flux shows relative uncertainty of 10%, 20%, and 35% for classes 1–3. Similarly, the photon index shows relative uncertainty of 5%, 10%, and 20%, for classes 1–3. The hydrogen column density is the least well constrained parameter with relative uncertainty of about 80% for class 1. Only  $\sim 10\%$  of our sources have  $N_H$  relative uncertainty below 30%. This is a combination of low count rate X-ray spectra and the fact that we are incorporating the uncertainty interval of the photometric redshift estimation in the fitting process. However, using the full probability distribution function in our scientific analysis we are able to propagate correctly the knowledge (or lack thereof) of the intrinsic absorption.

Figure 5a shows the estimated photon index  $\Gamma$ , as a function of redshift (upper panel) and intrinsic luminosity (lower panel). The points show the values obtained for each source in our sample (XXL-N: blue points, XXL-S: orange points). In order to investigate if there is any trend of photon index with redshift or luminosity, we also plot boxplots. We use five bins in redshift and luminosity and show the median of the photon index per bin (red line). The boxes enclose 50% of the distribution, while the dashed lines extend to the minimum and maximum values. The





**Fig. 4.** Panels from *top to bottom*: distribution of logarithms of the intrinsic flux determined by the absorbed power law, the photon index  $\Gamma$ , the logarithm of the hydrogen column density, the logarithm of the intrinsic luminosity in the 2–10 keV energy band. In all panels we show the distribution for XXL-N (blue lines) and XXL-S (orange lines) separately. The black line shows the XXL-1000-AGN sample.

red dots are considered outliers<sup>22</sup>. Both panels show that there is no significant change on the median  $\Gamma$  value. Similar findings are also reported in Piconcelli et al. (2005), Shemmer et al. (2005).

Equivalently, in Fig. 5b we investigate the presence of any possible trends of  $\log N_{\text{H}}$  with redshift and intrinsic luminosity. We do not find any trends of change in the median value of  $\log N_{\text{H}}$  within the XXL-1000-AGN sample. There is an ongoing debate in the literature regarding the fraction of obscured AGN and its dependence on redshift and/or luminosity. Studies in the literature<sup>23</sup> find that the fraction of obscured objects evolves with both redshift and luminosity (La Franca et al. 2005; Akyas et al. 2006; Gilli et al. 2007; Della Ceca et al. 2008; Hasinger 2008; Aird et al. 2015; Buchner et al. 2015). Nevertheless, Merloni et al. (2014) have shown the impact of the method used to estimate the obscuration (X-rays versus optical classification). With the current XXL-1000-AGN we do not detect a strong decline in the fraction of obscured AGN with luminosity. We obtained an observed fraction of

approximately 26% obscured objects ( $\log N_{\text{H}} > 10^{22} \text{ cm}^{-2}$ ) in XXL-1000-AGN, which is consistent with the fraction of obscured high-luminosity objects ( $\log L_x > 44$ ) reported in the deeper X-ray surveys noted previously, and in rough agreement with earlier wide area XMM-surveys (Piconcelli et al.; Perola et al.; see also Gilli et al. 2007, Fig. 16). This number is not surprising if we consider that the fraction of absorbed sources is the highest (up to 60–80%) for low-luminosity objects at higher redshifts. The flux limit of the XXL-1000-AGN simply does not allow this parameter space to be covered.

*Hardness ratio distribution.* Hardness ratios (HR) are used commonly in the literature as a rough estimator of the absorption in the absence of good quality X-ray spectra. For example, in Ueda et al. (2003) the authors use hardness ratio estimates to take into account absorption effects in their estimation of the X-ray luminosity function. We wish to derive a simple relationship between the observed hardness ratio and the true  $\log N_{\text{H}}$ . In Fig. 6a, we show the comparison between the hardness ratio computed from the PN count rate and the  $\log N_{\text{H}}$  estimated from the spectra with quality flag = 1–2. We use only PN count rate here since the sensitivity of PN is higher than the MOS instrument. We have verified, however, that the plot is similar in the case of HR estimated from the MOS count rate. The hardness ratio is calculated from the count rate (CR) as

$$\text{HR} = \frac{\text{CR}_{2-10\text{keV}} - \text{CR}_{0.5-2\text{keV}}}{\text{CR}_{0.5-2\text{keV}} + \text{CR}_{2-10\text{keV}}} \quad (2)$$

Typically,  $\text{HR} \sim -0.5$  corresponds to unabsorbed or moderately absorbed sources ( $\log N_{\text{H}} < 22$ ) over a broad range of redshifts. Higher values of HR correspond to higher absorption systems, while  $\text{HR} = 1.0$  denotes systems that are only detected in the 2–10 keV energy band. We observe a good agreement between the HR and the  $\log N_{\text{H}}$ , albeit with a large scatter and some outliers owing to the degeneracy between the determination of photon index  $\Gamma$ , absorption  $N_{\text{H}}$ , and redshift effects. Panel b of Fig. 6, shows the  $\log N_{\text{H}}$  histogram for four bins of HR. The dashed histograms show the  $\log N_{\text{H}}$  distribution when the median of the PDF is used as a point estimate of the hydrogen column density. In order to incorporate the full uncertainty estimation on the  $N_{\text{H}}$  parameter as determined by our X-ray spectral fitting, we summed the individual PDFs. The result is normalised to an area of one (Fig. 6 solid lines). In Appendix B we provide the combined  $\log N_{\text{H}}$  PDF for the XXL-1000-AGN sample (black lines) for the four hardness ratio bins. These curves can be used to draw random values of  $\log N_{\text{H}}$  when only the HR is available.

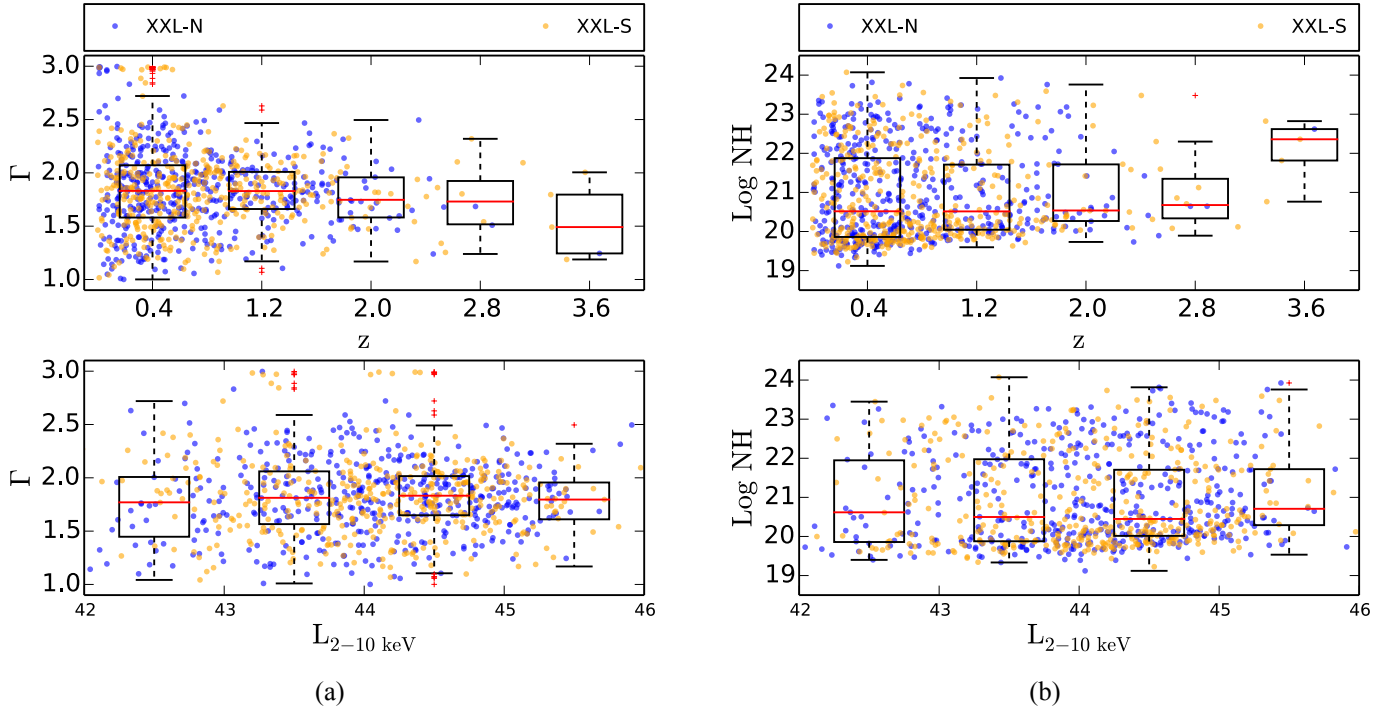
## 5.2. Median SEDs far-UV to mid-IR

The galaxy – AGN coevolution has been a very active research field in the past decades. One of the main components still under debate in the literature involving observations, simulations, and theory is the interplay between the AGN and the host galaxy. The universal presence of supermassive black holes in the centres of galaxies with bulges and the scaling relation between black hole mass and bulge luminosity, point in the direction of a shared history.

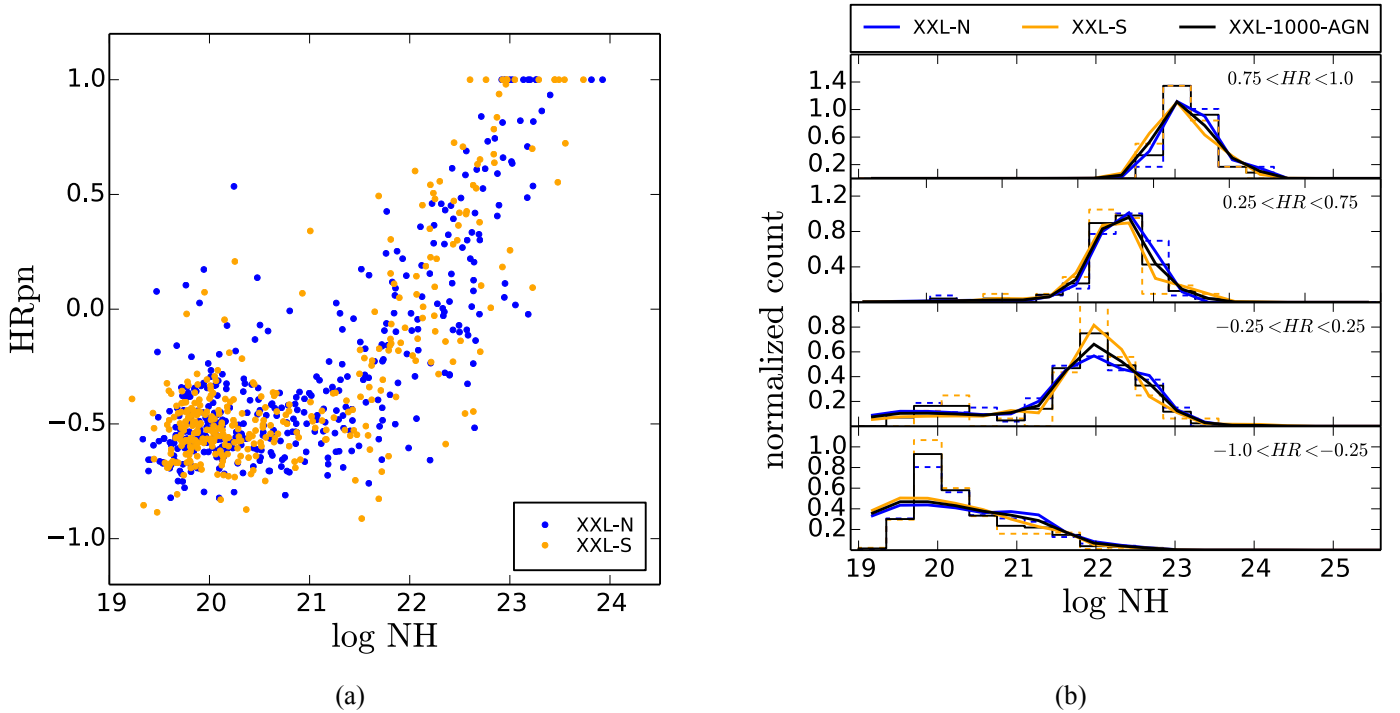
In Fig. 7a we show the median SEDs from the ultraviolet to the mid-infrared for XXL-N. We split the SED according to the intrinsic absorption  $\log N_{\text{H}}$  estimated from the X-ray spectrum. We see that the SEDs of objects with low intrinsic absorption  $\log N_{\text{H}} < 21$  appear to have QSO-like SEDs. At higher amounts

<sup>22</sup> For a boxplot diagram we consider as outliers the points that deviate from the best fit Gaussian distribution within each bin.

<sup>23</sup> Taking into account the available cosmological volume in order to avoid biasing the fraction in favour of the unabsorbed population.



**Fig. 5.** Photon index  $\Gamma$  **a)** and hydrogen column density  $\log N_H$  **b)** as a function of redshift (*top*) and luminosity (*bottom*). Splitting the sample into bins, we show with boxplots the median (red line) and 50% of the distribution within each bin.

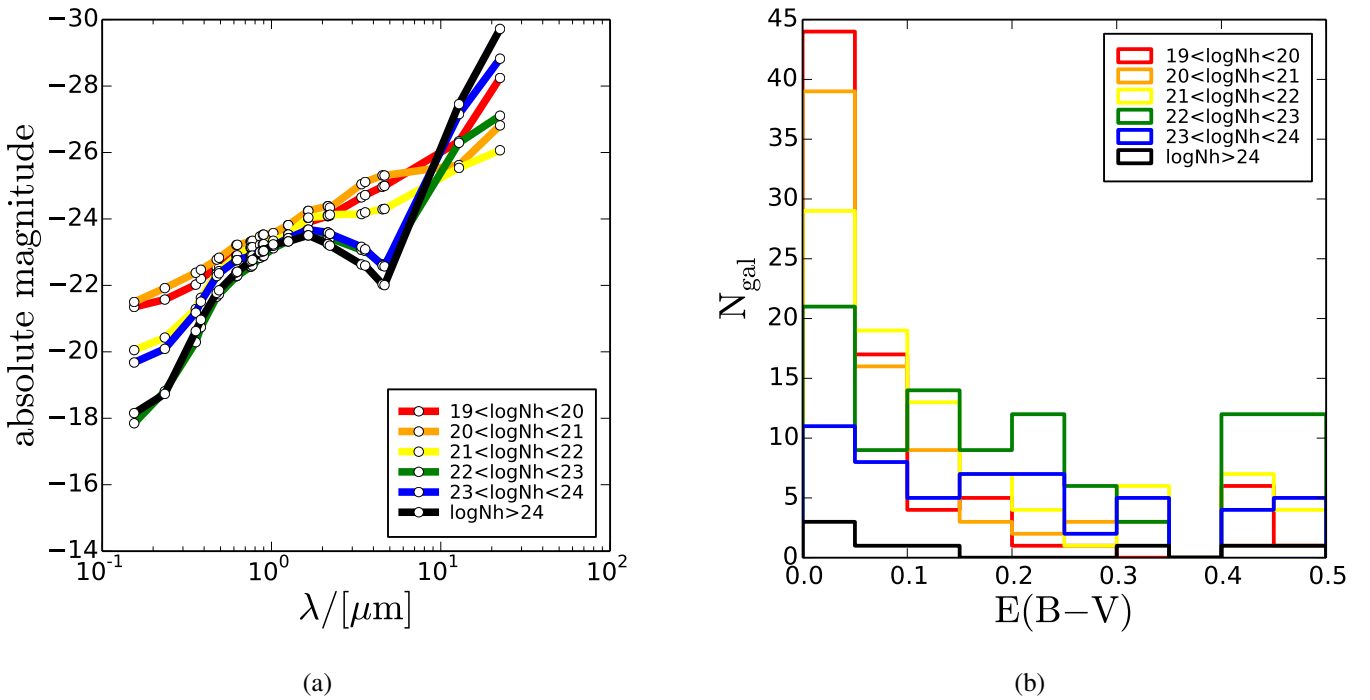


**Fig. 6. a)** Hardness ratio versus  $\log N_H$  for the XXL-N (blue points) and XXL-S (orange points); the hardness ratio is in good agreement with the estimated  $\log N_H$  from the X-ray spectra. **b)**  $\log N_H$  for four hardness ratio bins. The normalised dashed histograms represent the observed values in the XXL-N (blue) and XXL-S (orange) and the combined XXL-1000-AGN sample (solid black histogram). The solid lines show the combined  $\log N_H$  PDF of the sources in each bin (see Appendix B for the tabulated values of these curves).

of absorption ( $21 < \log N_H < 22$ ) the central engine does not outshine the host galaxy and the SED starts to have features that are expected in normal galaxies, namely the stellar bump at about  $1\mu\text{m}$ . At higher levels of absorption, the host galaxy resembles a passive galaxy, but with a very enhanced mid-infrared emission.

According to the unified model of AGN, this is the radiation that is absorbed and re-emitted by the torus close to the AGN.

For comparison in Fig. 7b we show the estimated absorption  $E(B - V)$  from the SED model fitting, using the same split according to the  $\log N_H$  estimated from the X-ray spectrum. In



**Fig. 7.** **a)** Median rest-frame SEDs for X-ray counterparts from UV to mid-infrared for XXL-N sources and **b)** Values of optical absorption  $E(B-V)$  estimated from the broadband SEDs and split into  $N_H$  bins determined from the X-ray spectrum, ranging from unabsorbed (red) to Compton-thick sources (black).

general, we observe that systems that are less absorbed in the X-rays also appear less absorbed in the optical, showing small  $E(B-V)$  values (less than 0.1). For moderate and higher hydrogen column densities, where the central engine is not powerful enough to outshine the galaxy, we see no correlation between the absorption close to the black hole and the absorption in the galaxy. This picture is consistent with observing an AGN seeing the host galaxy face-on and the torus aligned with the host galaxy. In this case, both optical and X-ray radiation escape the system with minimal losses due to absorption.

In Fig. 8 we show the median rest-frame host galaxy SED ranging from the mid-IR to the X-rays, split into four X-ray luminosity bins. The solid lines correspond to XXL-N sources, while the dashed lines to XXL-S sources. In order to appreciate the large scatter in host galaxy emission, we show with the individual sources that comprise the most luminous sample ( $45 < \log L_{2-10\text{keV}} < 46$ , grey lines). This plot clearly shows the impact of the presence of a luminous central engine on the broadband SED. Comparing the two extreme cases of the most luminous AGN ( $45 < \log L_x < 46$ , black lines) with the least luminous ( $42 < \log L_x < 43$ , blue lines), we see the enhancement of the emission both in the blue part of the SED ( $\nu \sim 10^{15}$  Hz) originating from the accretion disk close to the black hole at the centre of the galaxy and in the near infrared  $\nu \sim 10^{14}$  Hz) originating from the dusty torus (see also Lusso et al. 2011).

## 6. XXL-1000-AGN in the cosmic web

In this section we explore the XXL-1000-AGN sample within the context of the cosmic web by means of number counts, evolution of the X-ray luminosity function, and large-scale structure analysis. In this section we include the flag = 4 sources in the analysis, since the sensitivity of the survey is defined

according to the XAAMIN 3.3 detections independently of the X-ray spectra.

### 6.1. Number counts

The XXL survey with the large contiguous areas and medium X-ray flux limit fills the sparsely explored parameter space between deep and all-sky surveys (see Paper I and Brandt & Alexander 2015, for a review). Using the 2–10 keV flux determined by XAMIN 3.3 we estimated the cumulative number counts for the XXL survey using the equation

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

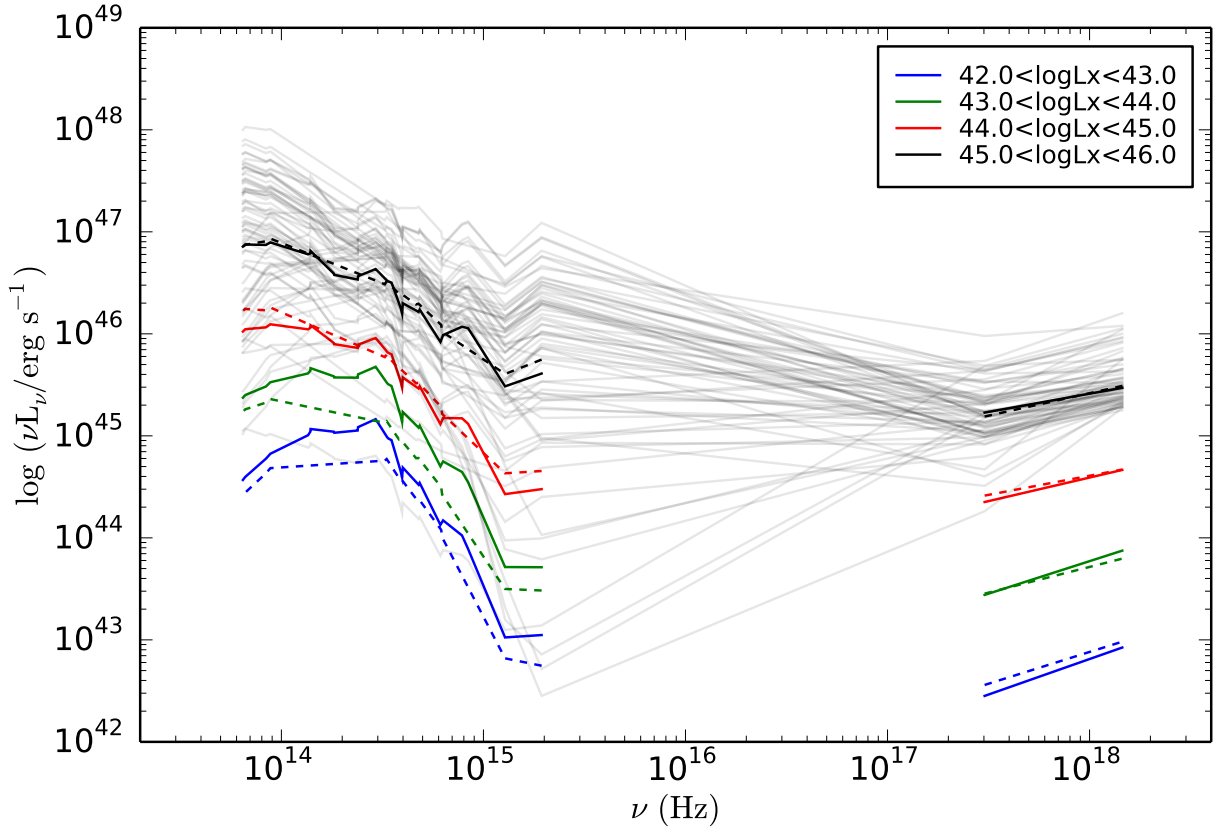
and the associated uncertainty is estimated by

$$\sigma = \sqrt{\sum_{i=1}^N \left(\frac{1}{\Omega_i}\right)^2}, \quad (4)$$

where  $\Omega$  the corresponding area that is sensitive to the flux limit  $S$ . The area curve for the XXL is determined following the same approach as for XMM-LSS described in Elyiv et al. (2012) estimated for the source detection of XAMIN 3.3 (Fig. 9a). In Fig. 9b we show the number counts estimated using the XXL-1000-AGN sample (XXL-N: blue circles, XXL-S: orange squares). The XXL-1000-AGN number counts are in good agreement with the observed number counts from deeper and narrower XMM-Newton surveys such as COSMOS (2 deg<sup>2</sup>, Cappelluti et al. 2007), H-ATLAS (16 deg<sup>2</sup>, Ranalli et al. 2015, Fig. 8), XMM-CDFS (0.2 deg<sup>2</sup>, Ranalli et al. 2013).

The XXL-1000-AGN sample has a brighter flux limit ( $S_{2-10\text{keV}} = 4.8 \times 10^{-14}$  erg s cm<sup>-2</sup>) than the characteristic flux break observed in deeper surveys (e.g.





**Fig. 8.** Median rest-frame SEDs for X-ray counterparts from X-rays to mid-infrared, split into X-ray luminosity bins. Solid lines refer to XXL-N sources, dashed lines to XXL-S sources. The grey lines show the SEDs of all sources in the  $45 < \log L_x < 46$  bin as an example of the large diversity of AGN hosts.

at  $S_{2-10\text{keV}} = 1.5 \times 10^{-15} \text{ erg s cm}^{-2}$  in Cappelluti et al. (2007). Therefore, a single power law is sufficient to describe the observed distribution. The weighted linear least squares fit on both XXL-N and XXL-S points gives  $\log N = (-1.6 \pm 0.12) \log S - (20 \pm 2)$ , consistent with the expectation of Euclidean number counts (i.e.  $N(>S) \propto S^{3/2}$ ).

The same slope has been reported in Ueda et al. (1999) using a sample of 44 sources detected in the 2–10 keV energy band in the ASCA Large Sky Survey (ALSS). Similar findings have been reported in the Fiore et al. (2001) using 147 *BeppoSAX* detections over  $85 \text{ deg}^2$  in the 4.5–10 keV energy band. Contrary, Baldi et al. (2002), one of the earliest attempts to constrain the bright end of the 2–10 keV number count distribution with *XMM-Newton* observations, using 495 detections in the HELAS2XMM survey reported a sub-Euclidean slope. Similarly, Mateos et al. (2008) found a discrepancy with the ASCA counts at the bright end ( $F_{2-10\text{keV}} > 1 \times 10^{-13} \text{ erg s}^{-1} \text{ cm}^{-2}$ ) of the  $\log N - \log S$  distribution in 2XMMi sample ( $\sim 9000$  sources). They concluded that the observed deviation cannot be explained by cross-calibration uncertainties alone.

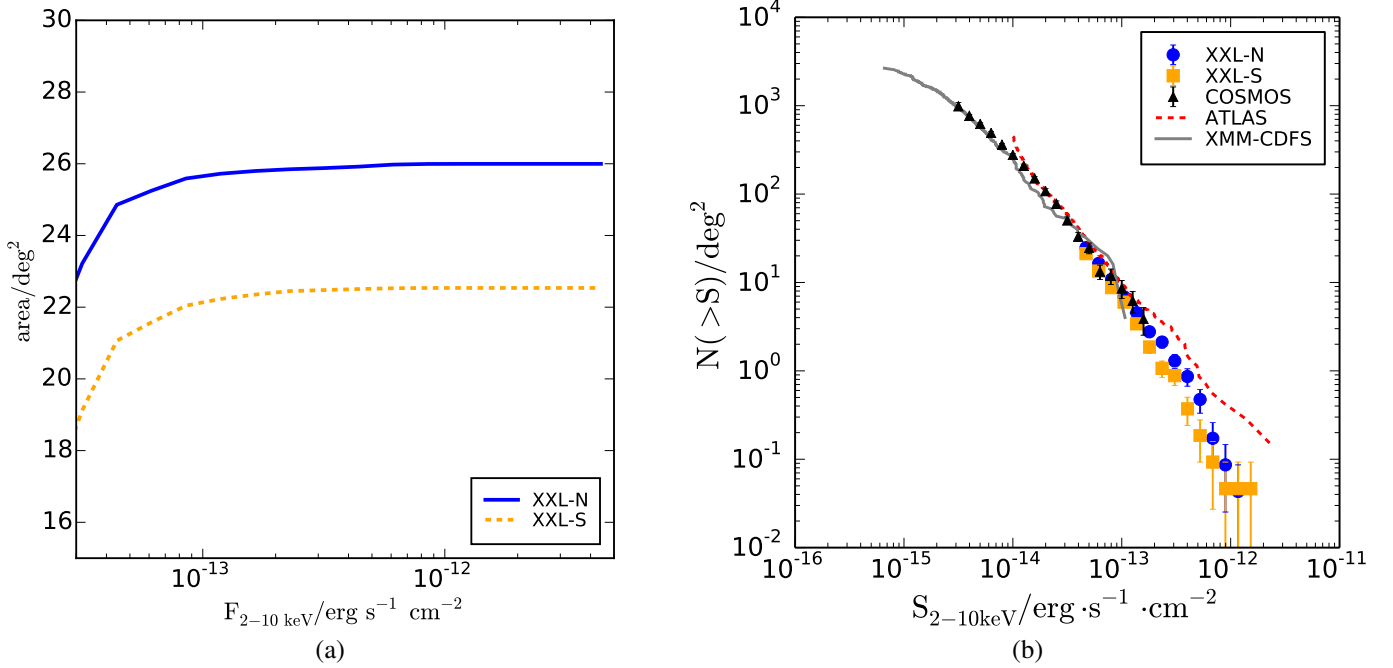
## 6.2. 2–10 keV luminosity function

The AGN X-ray luminosity function (XLF) traces the growth of supermassive black holes throughout the history of the Universe. Usually, because of the low number density of AGN, independent observations are combined coherently to create a comprehensive survey (Avni & Bahcall 1980). This approach also ensures that the result is not affected by cosmic variance. Numerous deep pencil beam X-ray surveys both with XMM and

*Chandra* have explored the faintest and furthest AGN reaching depths, for example of  $F_{2-10\text{keV}} \sim 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$ . Nevertheless, authors still rely on the ASCA catalogues of Akiyama et al. (2003) and Ueda et al. (2005) to introduce the rarest and brightest objects in their samples.

In Fig. 11 we show the coverage of the luminosity – redshift plane of the XXL-1000-AGN sample (black points) compared to a few current X-ray surveys – red: XMM-COSMOS (Cappelluti et al. 2007; Brusa et al. 2010), dark blue: AEGIS (Nandra et al. 2015), cyan: Lockman Hole (LH, Brunner et al. 2008; Fotopoulou et al. 2012), green: XMM-CDFS (Ranalli et al. 2015; Hsu et al. 2014), pink: CDFN (Alexander et al. 2003), and orange: XMM medium sensitivity survey (XMS Barcons et al. 2007). The orange shaded area shows the anticipated coverage of the full XXL catalogue using a provisional flux limit of  $F_{2-10\text{keV}} = 3 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ . The yellow area shows the coverage by the eRosita all-sky survey, which is expected to reach a flux limit of  $F_{2-10\text{keV}} = 2 \times 10^{-13} \text{ erg s}^{-1} \text{ cm}^{-2}$  after four years of observations (Merloni et al. 2012). The XXL-1000-AGN sample with its wide area coverage can help to set constraints on the faint end of the AGN XLF at redshifts less than 0.5, around the characteristic break luminosity  $\log L_0 \sim 44$  at redshifts  $0.5 < z < 1.0$ , and at the bright end at redshifts  $z > 1.0$ .

Early works on the 2–10 keV XLF, using samples of  $\sim 10^2$  sources, have shown that there is a strong evolution with redshift (e.g. La Franca et al. 2002; Ueda et al. 2003; La Franca et al. 2005; Ebrero et al. 2009; Aird et al. 2010), while the exact behaviour of the XLF, particularly at high redshift ( $z > 2$ ) has been heavily debated. More recent



**Fig. 9.** **a)** Area detection efficiency as a function of 2–10 keV X-ray flux for the XXL-1000-AGN sample. These curves were used for the calculation of the number counts and the AGN luminosity function. **b)**  $\log N - \log S$  for the XXL-1000-AGN in the XXL-N field (blue circles) and in the XXL-S field (orange squares). The black points are the number counts of XMM-COSMOS, the red line shows the cumber counts from the H-ATLAS, and the grey line the number counts of XMM-CDFS.

works in the literature have used numerous samples of AGN (1000–4000 objects) to constrain the AGN XLF and its evolution up to redshift  $z = 5$  (Ueda et al. 2014; Aird et al. 2015; Miyaji et al. 2015; Buchner et al. 2015), utilising several approaches in treating the intrinsic absorption  $N_{\text{H}}$ , which – depending on the redshift – can affect the observed 2–10 keV flux significantly. They confirm previous studies of the lower redshift Universe showing that the XLF is adequately described by a broken power law which evolves with redshift, showing a decline in the number density of AGN high redshift and low luminosities observed both in the soft (0.5–2.0 keV) and hard (2.0–10.0 keV) X-ray bands. (Miyaji et al. 2000; Ueda et al. 2003; Hasinger et al. 2005; La Franca et al. 2005; Ebrero et al. 2009; Aird et al. 2010).

In this work, we present the 2–10 keV intrinsic XLF where the unabsorbed flux is estimated by the power-law continuum emission corrected for absorption (see Sect. 2.3). In Sect. 5.1 we show that the XXL-1000-AGN sample includes mostly unabsorbed sources. However, the determination of  $\log N_{\text{H}}$  remains challenging especially for the spectra with lower counts (flag = 4, 9.6% of the sample). In the following analysis we include all the sources in our sample bearing in mind that the determination of the XLF is subject to these uncertainties.

We estimate the shape and evolution of the XLF adopting the luminosity dependent density evolution (LDDE) model (Miyaji et al. 2000) using the Ueda et al. (2003) parametrisation. Fotopoulou et al. (2016) have shown that the parametrisation of Ueda et al. (2003), thanks to its simplicity, is more favourable for samples at least up to  $z = 4$  than the more complex Ueda et al. (2014) model, which incorporates enough flexibility to account for an exponential number density cut-off at redshift  $z > 3$ . The LDDE model we used is given by

$$\frac{d\phi(L, z)}{d \log L} = \frac{d\phi(L, z = 0)}{d \log L} \times e(L, z). \quad (5)$$

The local XLF is modelled as a broken power-law distribution

$$\frac{d\phi(L, z = 0)}{d \log L} = \frac{N}{\left(\frac{L}{L_0}\right)^{\gamma_1} + \left(\frac{L}{L_0}\right)^{\gamma_2}}, \quad (6)$$

where  $L_0$ , is the luminosity at which the “break” occurs and  $\gamma_1$  and  $\gamma_2$  are the slopes of the power-law distributions below and above  $L_0$ . The evolution factor with  $z$  and  $L$  is given by

$$e(z, L) = \frac{(1 + z_c)^{p_1} + (1 + z_c)^{p_2}}{\left(\frac{1+z}{1+z_c}\right)^{-p_1} + \left(\frac{1+z}{1+z_c}\right)^{-p_2}} \quad (7)$$

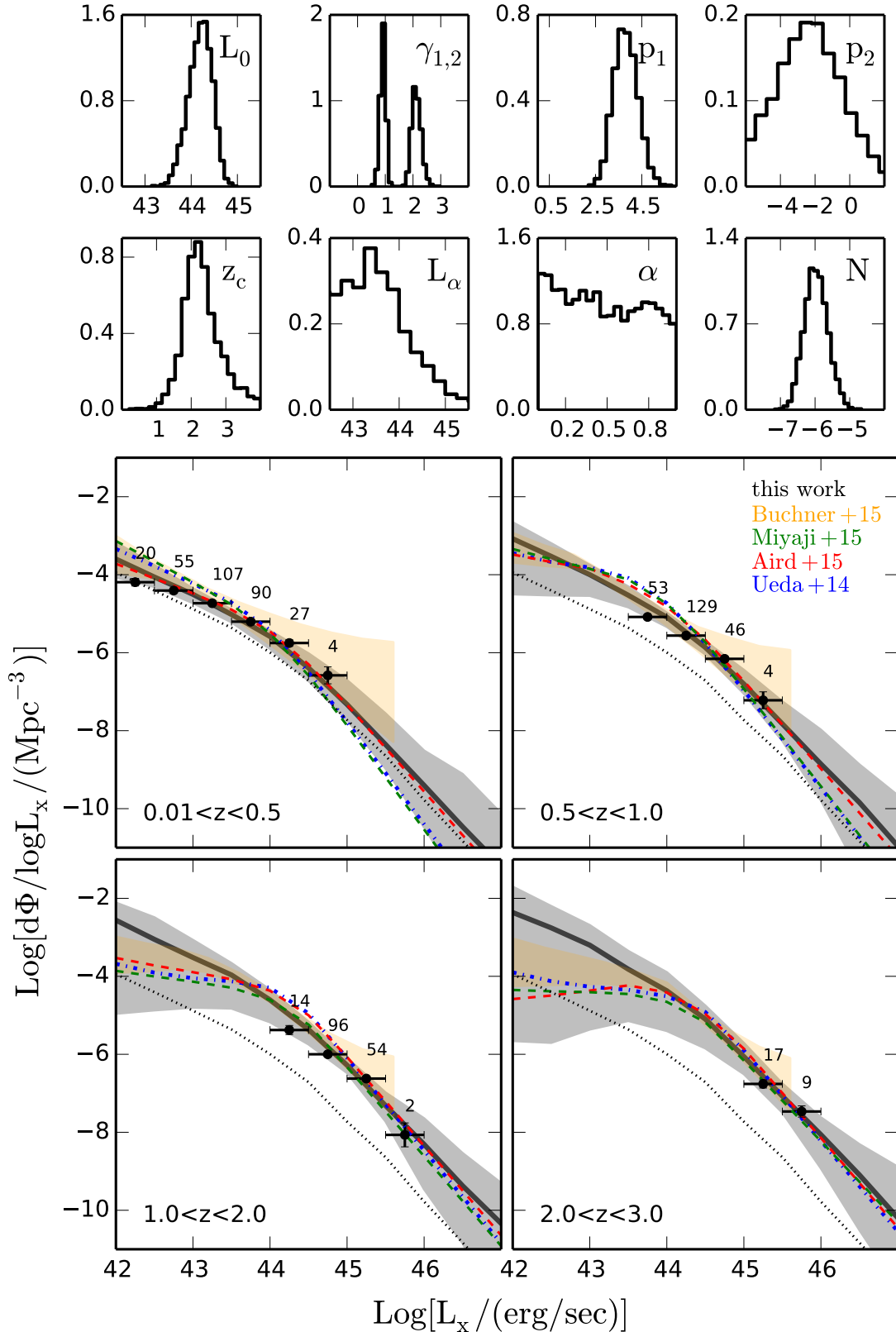
with

$$z_c(L) = \begin{cases} z_c^* & L \geq L_a \\ z_c^* \times \left(\frac{L}{L_a}\right)^a & L < L_a \end{cases}. \quad (8)$$

Similarly to Aird et al. (2010) and Fotopoulou et al. (2016) we used spectroscopic redshifts when available and the probability distribution function of the photometric redshift estimates for sources without spec- $z$ . For the parameter estimation we used PyMultiNest<sup>24</sup> (Buchner et al. 2014), the python interface of MultiNest (Feroz & Hobson 2008; Feroz et al. 2009, 2013). MultiNest performs Nested Sampling introduced by Skilling (2004) and it is able to explore the posterior even in the case of multimodal distributions.

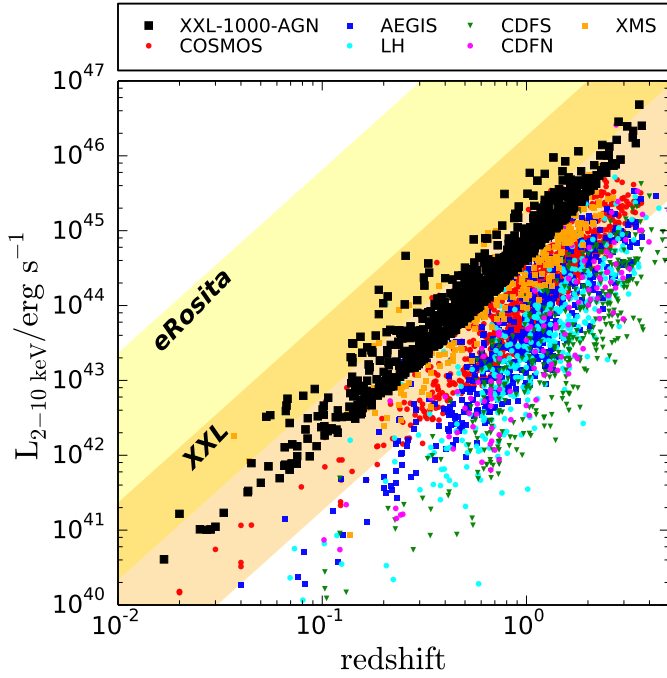
In Fig. 10 we show the XLF estimated using the XXL-1000-AGN sample. The top panels show the posteriors of all model parameters. Parameters  $\gamma_1$  and  $\gamma_2$  are symmetric in Eq. (6); therefore, they have the same double-peaked posterior distribution (shown here only once for simplicity). We note that when using a Bayesian approach, it is not necessary to fix any of the model parameters that are not constrained by our data. This is the case for parameter  $\alpha$ , which describes the decline of the faint end of

<sup>24</sup> <https://github.com/JohannesBuchner/PyMultiNest>



**Fig. 10.** X-ray luminosity function from the XXL-1000-AGN sample. The *upper panels* show the marginalised posterior distribution of the parameters in luminosity dependent density evolution (LDDE) model. The *bottom panels* show the estimated XLF as a function of luminosity in four redshift bins. The black solid line shows the mode of the distribution and the grey shaded area encloses the 90% credible interval. The dotted black line is our estimated  $z = 0$  XLF shown for reference in all plots. The black points are estimated using the  $1/V_{\max}$  method in good agreement with our modelling. The numbers above the points show the number of objects per bin. They also show the parameter space occupied by the XXL-1000-AGN sample. The orange area and dashed lines show 2–10 keV XLF estimates from the literature (orange: Buchner et al. 2015; green: Miyaji et al. 2015; red: Aird et al. 2015; blue: Ueda et al. 2014).





**Fig. 11.** Luminosity redshift plane for X-ray fields in the 2–10 keV energy band. The black points show the XXL-1000-AGN sample. The orange shaded area shows the parameter space covered by the full XXL catalogue. The yellow shaded area shows the expectation for the four-year all-sky coverage by eRosita. The XXL survey is a stepping stone between current deep surveys (coloured points) and future all-sky surveys.

the XLF at redshifts above  $z_c$ . The posterior distribution resembles the flat prior that we used during the parameter estimation. Since we are using samples drawn from the multidimensional posterior to estimate the XLF, the uncertainty on any of the parameters is naturally incorporated in the uncertainty budget of the XLF (grey shaded area). In Table 5, we give the mean of the posterior distribution and the standard deviation for each parameter. The bottom panels of Fig. 10 show the resulting XLF using the LDDE model estimated with our sample as a function of luminosity and in four redshift bins. The black dashed line shows the estimated local luminosity function (Eq. (6)) for reference. The red lines show the mode of the luminosity function distribution, while the grey shaded area shows the 90% credible interval. As an independent method, we show the binned estimates calculated using the  $1/V_{\max}$  method (Schmidt 1968; Page & Carrera 2000). The numbers above the points show how many objects belong in each bin. We see that the two methods agree rather well within the uncertainties.

The XLF from this work is in good agreement with recent estimates of the AGN XLF (Buchner et al. 2015; Miyaji et al. 2015; Aird et al. 2015; Ueda et al. 2014, yellow area, green, blue, and red lines in Fig. 10, respectively). We note that XLFs are in good agreement when compared within the parameter space that can be constrained by the data set used for the estimation, while large discrepancies, for example the faint end slope at  $z > 2$ , appear when extrapolating the models.

### 6.3. Large-scale structure

Galaxy formation theories show that the action of gravitational instability on a perturbed fluctuation background gives rise to a wealth of large-scale structures, the interconnections

**Table 5.** Best fit values of the luminosity dependent density evolution model.

Parameter	Mean	$\sigma$	Parameter	Mean	$\sigma$
$\log L_0$	44.18	0.26	$z_c$	2.30	0.5
$\gamma_1$	0.91	0.10	$\log L_\alpha$	43.00	1.0
$\gamma_2$	2.13	0.17	$\alpha$	0.48	0.29
p1	3.80	0.5	$\log N$	-6.00	0.3
p2	-3.00	2.0			

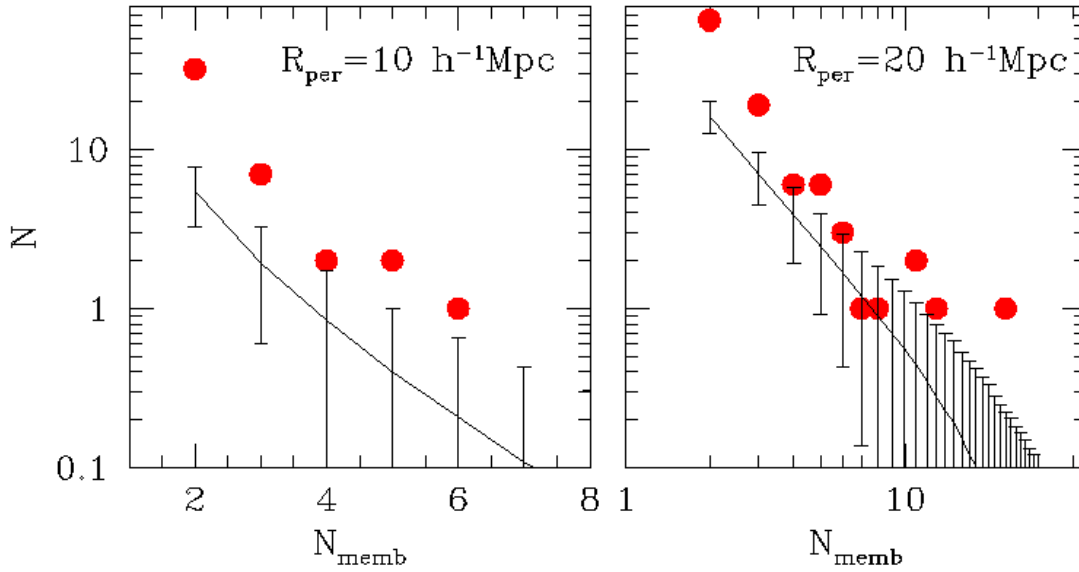
of which provide the so-called cosmic web or cosmic foam (Bond et al. 1996). The complex and interconnected structures of the cosmic web come in the variety of groups and clusters of galaxies, filaments, walls, and voids, which are low-density regions filling most of the volume of the Universe (e.g. van de Weygaert & Bond 2008). Observationally, it has been found that when using galaxies as tracers of the large-scale structure, the wealth of structures in the Universe can clearly be identified (for recent studies, see Parihar et al. 2014; James et al. 2007). AGN have also been used to trace the large-scale structure by using a variety of techniques, among which the two-point correlation function and BAOs (e.g. Cappelluti et al. 2012; Hütsi et al. 2014, and references therein) or their number counts (e.g. Dai et al. 2015).

In this section we present a preliminary analysis of the X-ray AGN large-scale structures by applying the friends-of-friends algorithm in order to identify compact or loose groupings of AGN and then we assess their significance by using extended Monte Carlo simulations.

We use all the X-ray AGN for which we have either spectroscopic (589) or photometric (383) redshifts. In Fig. 12 we present as red points the multiplicity function of AGN structures as a function of AGN members and for two percolation radii ( $R_{\text{per}} = 10$  and  $20 h^{-1}$  Mpc).

We also provide by the continuous line the mean random expectation from the application of our friends-of-friends algorithm on 10 000 random realisations having the same number of points, angular selection function, and redshift distribution as the real X-ray AGN. This is achieved by using the same celestial coordinates as the real data, but randomising the redshift of each AGN such that all observed redshifts are assigned to random points while no redshift is duplicated. Furthermore, we smooth the redshifts using a Gaussian kernel with zero mean and  $\sigma = 0.1$  (we verified that varying  $\sigma$  between 0.05 and 0.2 provides equivalent results). It is evident that the most significant X-ray AGN structures are those with two or three members (for both percolation radii).

For the  $R_{\text{per}} = 20 h^{-1}$  Mpc case we also identify one significant large structure containing 23 AGN, which has a diameter of  $81 h^{-1}$  Mpc, which leads us to consider it as a candidate supercluster of X-ray AGN. The centre of mass of this structure has celestial coordinates  $(\alpha, \delta) = (34.80^\circ, -5.17^\circ)$  and a mean redshift  $\langle z \rangle = 0.14$ , and it is related to the most abundant (XLSSC-*b*) supercluster of XXL bright clusters (see Paper II). The significance of this AGN structure can be appreciated not only by the probability of observing such an AGN membership, as provided by the comparison with the random expectation in Fig. 12, but also by the compactness of this structure. In Fig. 13 we present the distribution of diameters of all the structures with 23 members that have been found in 10 000 Monte Carlo simulations. Only in 3% of the cases is the diameter of the random structures as small as that of the observed structure. We have also performed the same analysis using only sources with spectroscopic



**Fig. 12.** Multiplicity function of AGN for two percolation radii,  $R_{\text{per}} = 10 h^{-1} \text{ Mpc}$  **a)** and  $R_{\text{per}} = 20 h^{-1} \text{ Mpc}$  **b)**. The grey lines show the expected multiplicity of a random field with the same smoothed redshift distribution as the data. We see significant structures with 2–3 members at both radii and one significant structure consisting of 23 members for  $R_{\text{per}} = 20 h^{-1} \text{ Mpc}$ .

redshifts. The overdensity is also present in the restricted sample, which contains 16 members.

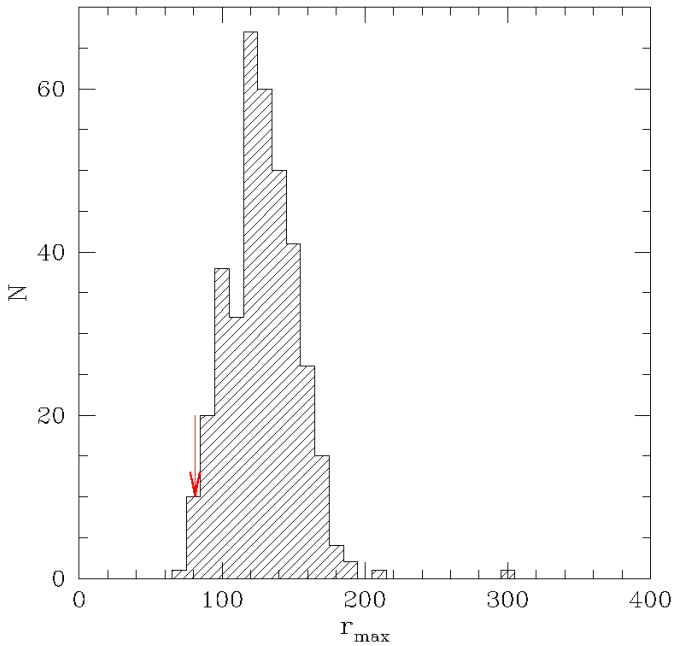
We conclude that the spatial distribution of XXL AGN shows a wealth of significant large-scale structures, some of which are of supercluster size. The detected AGN supercluster, but also other compact structures of various memberships, will be investigated in detail in a forthcoming paper.

## 7. Released catalogue description

With this paper we release the XXL-1000-AGN sample, the 1000 brightest X-ray point-like sources detected in the XMM-XXL field. The fluxes estimated by XAMIN 3.3 (Cols. 4–5) assume  $\Gamma = 1.7$ , and  $N_{\text{H}} = 2.6 \times 10^{20} \text{ cm}^{-2}$ . The count-to-flux conversion factors are given in Sect. 2.1. For the quantities estimated by the spectral fit (Cols. 6–17) we provide the mode of the marginalised posterior distribution, the 68% credible interval around the mode, and the median of the distribution (see Sect. 2.3). All magnitudes and associated uncertainties are measured with Sextractor. They originate from SDSS/CFHTLS for XXL-N or BCS/DECam for XXL-S. They are given in AB, measured in a fixed circular aperture of  $3''$ , and they are corrected to total and for Galactic extinction. All fluxes are given in  $\text{erg s}^{-1} \text{ cm}^{-2}$ , magnitudes in AB. The contents of the catalogue are as follows:

- (1) Xcatname: unique source identification;
- (2) Xra: X-ray point source right ascension;
- (3) Xdec: X-ray point source declination;
- (4) Bflux: 0.5–2 keV flux (Xamin 3.3);
- (5) CDFlux: 2–10 keV flux (Xamin 3.3);
- (6)  $F_{\text{mode}}$ : 2–10 keV intrinsic (unabsorbed) flux, mode of PDF;
- (7)  $F_{\text{l}}$ : 2–10 keV intrinsic (unabsorbed) flux, lower bound of 68% credible interval;
- (8)  $F_{\text{h}}$ : 2–10 keV intrinsic (unabsorbed) flux, upper bound of 68% credible interval;
- (9)  $F_{\text{median}}$ : 2–10 keV intrinsic (unabsorbed) flux, median of PDF;

- (10)  $\Gamma_{\text{mode}}$ : photon index, mode of PDF;
- (11)  $\Gamma_{\text{l}}$ : photon index, lower bound of 68% credible interval;
- (12)  $\Gamma_{\text{h}}$ : photon index, upper bound of 68% credible interval;
- (13)  $\Gamma_{\text{median}}$ : photon index, median of PDF;
- (14)  $N_{\text{H,mode}}$ : hydrogen column density, mode of PDF;
- (15)  $N_{\text{H,l}}$ : hydrogen column density, lower bound of 68% credible interval;
- (16)  $N_{\text{H,h}}$ : hydrogen column density, upper bound of 68% credible interval;
- (17)  $N_{\text{H,median}}$ : hydrogen column density, median of PDF;
- (18) Xflag: X-ray spectrum quality flag (see Sect. 2.3, Table 2);
- (19) CtpRa: counterpart right ascension;
- (20) CtpDec: counterpart declination;
- (21)  $g$ :  $g$ -band magnitude;
- (22)  $g_{\text{err}}$ :  $g$ -band magnitude uncertainty;
- (23)  $r$ :  $r$ -band magnitude;
- (24)  $r_{\text{err}}$ :  $r$ -band magnitude uncertainty;
- (25)  $i$ :  $i$ -band magnitude;
- (26)  $i_{\text{err}}$ :  $i$ -band magnitude uncertainty;
- (27)  $z$ :  $z$ -band magnitude;
- (28)  $z_{\text{err}}$ :  $z$ -band magnitude uncertainty;
- (29) photo\_origin: parent photometric survey. A four-character code “griz”, where each of the four characters can assume the value “–” in the case of missing data in the band,  $S$  or  $C$  (for SDSS or CFHTL, XXL-N), or  $B$  or  $D$  (for BCS or DECam, XXL-S);
- (30) zspec: spectroscopic redshift, when available;
- (31) zspec\_origin: spectroscopic redshift parent survey
  - AAT: Lidman et al. (2016),
  - Akiyama et al. (2015),
  - GAMA: Hopkins et al. (2013),
  - LDSS03: Adami et al. (2011),
  - SDSSDR12 Menzel et al. (2016),
  - SDSS DR10,
  - Simpson et al. (2006),
  - Simpson et al. (2012),
  - Stalin et al. (2010),



**Fig. 13.** Radius distribution of overdensities with 23 members estimated in 10000 Monte Carlo simulations. The red arrow shows the radius of the overdensity with 23 members identified in the data. The compactness of the structure ( $81 h^{-1}$  Mpc) makes it a candidate supercluster of AGN.

- VIPERS: Garilli et al. (2014),
- VVDS: Le Fèvre et al. (2013),
- WHT: Paper XII.
- (32) zphot: photometric redshift;
- (33) zphot\_l: lower bound of 68% credible interval;
- (34) zphot\_h: upper bound of 68% credible interval;
- (35) zphot\_class: classification for best photo-z;
  - passive = 1,
  - star forming = 2,
  - starburst = 3,
  - AGN = 4,
  - QSO = 5,
- (36) Pstar: probability that the source is a star;
- (37) Poutlier: probability that a source has  $|z_{\text{phot}} - z_{\text{spec}}| > 0.15 \cdot (1 + z_{\text{spec}})$ .

An example page of the final source catalogue is shown in Table C.1. A full printable version is available. The full catalogue is available as a queryable database table `XXL_1000_AGN` via the XXL Master Catalogue browser<sup>25</sup>. A copy is also deposited at the Centre de Données astronomiques de Strasbourg (CDS).

## 8. Conclusions

We presented the XXL-1000-AGN sample, the first release of the XXL point source catalogue, consisting of the 1000 brightest sources, selected in the 2–10 keV energy band. The flux limit of our catalogue is  $F_{2-10\text{keV}} = 4.8 \times 10^{-14}$  erg s<sup>-1</sup> cm<sup>-2</sup>.

- We built a multiwavelength catalogue creating SEDs from the far-ultraviolet to the mid-infrared and assigned the counterparts to the X-ray detections using the likelihood

ratio technique. We retrieve counterparts for 97% of our sources.

- Using machine learning classification, we assign the optimal class describing the broadband SEDs which improved the photometric redshift estimates.
- Modelling the X-ray spectra with an absorbed power-law distribution we find that the average photon index is  $\langle \Gamma \rangle = 1.85 \pm 0.40$ . Our sample is dominated by unabsorbed sources (80% with  $N_{\text{H}} < 10^{22}$  cm<sup>-2</sup>) with average hydrogen column density  $\log \langle N_{\text{H}} \rangle = 21.0 \pm 1.3$ .
- We present the median observed SED (AGN combined with the host galaxy emission) and show that the X-ray absorption, which was estimated from X-ray spectra, shows the same general trend as the host galaxy absorption estimated by the SEDs. Low absorption in the X-rays in general corresponds to low absorption in the optical. This is consistent with the picture of the unified model of AGN where we observe the galaxy face-on.
- The XXL X-ray number counts are fully consistent with the Euclidean expectation and agree with previous deep (CDFs, COSMOS) and wide (H-ATLAS) *XMM-Newton* surveys.
- We present the best fit parameters for the LDDE XLF model up to  $z = 3$ . The XXL-1000-AGN sample poses constraints on the low luminosity – low redshift, medium luminosity – medium redshift X-ray luminosity function, in good agreement with recent estimates using deeper X-ray observations.
- An application of the friends-of-friends algorithm at  $10 h^{-1}$  Mpc and  $20 h^{-1}$  Mpc percolation radii shows significant structures with 2–3 members. Additionally, the analysis of  $R_{\text{per}} = 20 h^{-1}$  Mpc suggests the presence of a candidate supercluster of AGN with 23 members at redshift  $z = 0.14$ . The same result is retrieved using only sources with spec-z.
- We release the catalogue of the XXL-1000-AGN sample with positions, flux estimates from both the pipeline and the X-ray spectra, optical magnitudes, and redshift information (spec-z 60% and photo-z 40%).

With the first data release, we provide a significant number of AGN detected with *XMM-Newton*, of comparable quality to modern deep X-ray surveys. In future publications we will expand the analysis presented in this work to the full XXL catalogue containing an unprecedented number of  $\sim 10^4$  (X-ray) point-like 2–10 keV detected sources analysing the fully combined XMM pointings to reach maximum depth. A unique advantage of XXL collaboration is the combined study of X-ray point-like sources and X-ray detected galaxy clusters in great numbers, which will allow the study of AGN with respect to their environment and, vice-versa, the study of the impact of AGN on clusters. The full catalogue is planned to be released in incremental flux limits.

*Acknowledgements.* We thank the anonymous referee for the useful comments on our work. F.P. and M.R.-C. acknowledge support from DiG Transregio Programme TR33. P.R. acknowledges a grant from the Greek General Secretariat of Research and Technology in the framework of the programme Support of Postdoctoral Researchers. M.B. acknowledges support from the European Union’s FP7 under grant agreement 321913 (CIG, “SMBH evolution through cosmic time”). O.M. is grateful for the financial support provided by the European Union Seventh Framework Programme (FP7 2007–2013), grant agreement n° 291823 Marie Curie FP7-PEOPLE-2011-COFUND (The new International Fellowship Mobility Programme for Experienced Researchers in Croatia – NEWFELPRO, project “AGN environs in XXL”, Grant Agreement #83). SA acknowledges a post-doctoral fellowship from TUBITAK-BIDEB through 2219 program. XXL is an international project based around an XMM Very Large Programme surveying two 25 deg<sup>2</sup> extragalactic fields at a depth of  $5 \times 10^{-15}$  erg cm<sup>-2</sup> s<sup>-1</sup> in the [0.5–2] keV band for point-like sources. The

<sup>25</sup> <http://cosmosdb.iasf-milano.inaf.it/XXL/>



XXL website is <http://irfu.cea.fr/xxl>. Multiband information and spectroscopic follow-up of the X-ray sources are obtained through a number of survey programmes, summarised at <http://xxlmultiwave.pbworks.com/>. Based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/IRFU, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at Terapix available at the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS. Based on observations obtained with WIRCam, a joint project of CFHT, Taiwan, Korea, Canada, France, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique of France, and the University of Hawaii. This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. This paper uses data from the VISTA Hemisphere Survey ESO programme ID: 179.A-2010 (PI. McMahon). This publication has made use of data from the VIKING survey from VISTA at the ESO Paranal Observatory, programme ID 179.A-2004. Based on data products from observations made with ESO Telescopes at the La Silla or Paranal Observatories under ESO programme ID 179.A-2006 (VIDEO survey). Data processing has been contributed by the VISTA Data Flow System at CASU, Cambridge and WFAU, Edinburgh. This work is based in part on data obtained as part of the UKIRT Infrared Deep Sky Survey. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the US Department of Energy Office of Science. The SDSS-III web site is <http://www.sdss3.org/>. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University. This work is partly based on observations made with the *Spitzer* Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. Some of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. Support for MAST for non-HST data is provided by the NASA Office of Space Science via grant NNX09AF08G and by other grants and contracts. Based in part on data acquired through the Australian Astronomical Observatory, under programs A/2013A/018 and A/2013B/001, and on observations at Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory (NOAO Prop. IDs 2013A-0618 and 2015A-0618), which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. This project used data obtained with the Dark Energy Camera (DECam), which was constructed by the Dark Energy Survey (DES) collaboration. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência, Tecnologia e Inovação, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the University of Edinburgh, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität München and the associated Excellence

Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, the Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University.

## References

- Adami, C., Mazure, A., Pierre, M., et al. 2011, *A&A*, **526**, A18  
 Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2014, *ApJS*, **211**, 17  
 Aird, J., Nandra, K., Laird, E. S., et al. 2010, *MNRAS*, **401**, 2531  
 Aird, J., Coil, A. L., Georgakakis, A., et al. 2015, *MNRAS*, **451**, 1892  
 Akiyama, M., Ueda, Y., Ohta, K., Takahashi, T., & Yamada, T. 2003, *ApJS*, **148**, 275  
 Akiyama, M., Ueda, Y., Watson, M. G., et al. 2015, *PASJ*, **67**, 82  
 Akylas, A., Georgantopoulos, I., Georgakakis, A., Kitsionas, S., & Hatziminaoglou, E. 2006, *A&A*, **459**, 693  
 Alexander, D. M., Bauer, F. E., Brandt, W. N., et al. 2003, *AJ*, **126**, 539  
 Ashby, M. L. N., Stanford, S. A., Brodwin, M., et al. 2014, *ApJS*, **212**, 16  
 Avni, Y., & Bahcall, J. N. 1980, *ApJ*, **235**, 694  
 Baldi, A., Molendi, S., Comastri, A., et al. 2002, *ApJ*, **564**, 190  
 Barcons, X., Carrera, F. J., Ceballos, M. T., et al. 2007, *A&A*, **476**, 1191  
 Bertin, E., & Arnouts, S. 1996, *A&AS*, **117**, 393  
 Bond, J. R., Kofman, L., & Pogosyan, D. 1996, *Nature*, **380**, 603  
 Bongiorno, A., Zamorani, G., Gavignaud, I., et al. 2007, *A&A*, **472**, 443  
 Bonoli, S., Marulli, F., Springel, V., et al. 2009, *MNRAS*, **396**, 423  
 Bournaud, F., Dekel, A., Teyssier, R., et al. 2011, *ApJ*, **741**, L33  
 Brandt, W. N., & Alexander, D. M. 2015, *A&ARv*, **23**, 1  
 Breiman, L. 2001, *Machine Learning*, **45**, 1  
 Brunner, H., Cappelluti, N., Hasinger, G., et al. 2008, *A&A*, **479**, 283  
 Brusa, M., Comastri, A., Daddi, E., et al. 2005, *A&A*, **432**, 69  
 Brusa, M., Zamorani, G., Comastri, A., et al. 2007, *ApJS*, **172**, 353  
 Brusa, M., Civano, F., Comastri, A., et al. 2010, *ApJ*, **716**, 348  
 Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, *A&A*, **564**, A125  
 Buchner, J., Georgakakis, A., Nandra, K., et al. 2015, *ApJ*, **802**, 89  
 Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, *ApJ*, **533**, 682  
 Cappelluti, N., Hasinger, G., Brusa, M., et al. 2007, *ApJS*, **172**, 341  
 Cappelluti, N., Ajello M., Burlon D. 2010, *ApJ*, **716**, L209  
 Cappelluti, N., Allevalo, V., & Finoguenov, A. 2012, *Adv. Astron.*, **2012**, 1  
 Casali, M., Adamson, A., Alves de Oliveira, C., et al. 2007, *A&A*, **467**, 777  
 Cash, W. 1979, *ApJ*, **228**, 939  
 Chiappetti, L., Clerc, N., Pacaud, F., et al. 2013, *MNRAS*, **429**, 1652  
 Coil, A. L., Hennawi, J. F., Newman, J. A., Cooper, M. C., & Davis, M. 2007, *ApJ*, **654**, 115  
 Coil, A. L., Georgakakis, A., Newman, J. A. et al. 2009, *ApJ*, **701**, 1484  
 Corral, A., Georgantopoulos, I., Watson, M. G., et al. 2015, *A&A*, **576**, A61  
 Dai, X., Griffin, R. D., Kochanek, C. S., Nugent, J. M., & Bregman, J. N. 2015, *ApJS*, **218**, 8  
 De Luca, A., & Molendi, S. 2004, *A&A*, **419**, 837  
 Della Ceca, R., Maccacaro, T., Caccianiga, A., et al. 2004, *A&A*, **428**, 383  
 Della Ceca, R., Caccianiga, A., Severgnini, P., et al. 2008, *A&A*, **487**, 119  
 Desai, S., Armstrong, R., Mohr, J. J., et al. 2012, *ApJ*, **757**, 83  
 Dickey, J. M., & Lockman, F. J. 1990, *ARA&A*, **28**, 215  
 Di Matteo, T., Springel, V., & Hernquist, L. 2005, in *Growing Black Holes: Accretion in a Cosmological Context*, eds. A. Merloni, S. Nayakshin, & R. A. Sunyaev, 340  
 Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, *Astron. Geophys.*, **50**, 12  
 Driver, S. P., Hill, D. T., Kelvin, L. S., et al. 2011, *MNRAS*, **413**, 971  
 Dye, S., Warren, S. J., Hambly, N. C., et al. 2006, *MNRAS*, **372**, 1227  
 Eckert, D., Molendi, S., Owers, M., et al. 2014, *A&A*, **570**, A119  
 Ebrero, J., Carrera, F. J., Page, M. J., et al. 2009, *A&A*, **493**, 55  
 Elvis, M., Civano, F., Vignali, C., et al. 2009, *ApJSS*, **184**, 158  
 Elyiv, A., Clerc, N., Plionis, et al. 2012, *A&A*, **537**, A131  
 Fanidakis, N., Georgakakis, A., Mountrichas, G., et al. 2013a, *MNRAS*, **435**, 679  
 Fanidakis, N., Macció, A. V., Baugh, C. M., Lacey, C. G., & Frenk, C. S. 2013b, *MNRAS*, **436**, 315  
 Feroz, F., & Hobson, M. P. 2008, *MNRAS*, **384**, 449  
 Feroz, F., Hobson, M. P., & Bridges, M. 2009, *MNRAS*, **398**, 1601  
 Feroz, F., Hobson, M. P., Cameron, E., & Pettitt, A. N. 2013, *ArXiv e-prints* [[arXiv:1306.2144](https://arxiv.org/abs/1306.2144)]  
 Ferrarese, L. & Merritt, D. 2000, *AJ*, **539**, L9  
 Fiore, F., Giommi, P., Vignali, C., et al. 2001, *MNRAS*, **327**, 771  
 Fotopoulou, S., Salvato, M., Hasinger, G., et al. 2012, *ApJS*, **198**, 1  
 Fotopoulou, S., Buchner, J., Georgantopoulos, I., et al. 2016, *A&A*, **587**, A142  
 Garilli, B., Guzzo, L., Scodreggio, M., et al. 2014, *A&A*, **562**, A23  
 Gebhardt, K., Kormendy, J., Ho, L. C., et al. 2000, *ApJ*, **543**, L5  
 Giacomini, R., Rosati, P., Tozzi, P., et al. 2001, *ApJ*, **551**, 624



- Gilli, R., Comastri, A., & Hasinger, G. 2007, *A&A*, 463, 79
- Guzzo, L., Scodreggio, M., Garilli, B., et al. 2014, *A&A*, 566, A108
- Hambly, N. C., Collins, R. S., Cross, N. J. G., et al. 2008, *MNRAS*, 384, 637
- Han, Y., Dai, B., Wang, B., Zhang, F., & Han, Z. 2012, *MNRAS*, 423, 464
- Hasinger, G. 2008, *A&A*, 490, 905
- Hasinger, G., Miyaji, T., & Schmidt, M. 2005, *A&A*, 441, 417
- Hasinger, G., Cappelluti, N., Brunner, H., et al. 2007, *ApJSS*, 172, 29
- Hewett, P. C., Warren, S. J., Leggett, S. K., & Hodgkin, S. T. 2006, *MNRAS*, 367, 454
- Hirschmann, M., Somerville, R. S., Naab, T., & Burkert, A. 2012, *MNRAS*, 426, 237
- Hodgkin, S. T., Irwin, M. J., Hewett, P. C., & Warren, S. J. 2009, *MNRAS*, 394, 675
- Hopkins P. F., & Hernquist L. 2006, *ApJS*, 166, 1
- Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2005, *ApJ*, 630, 716
- Hopkins, P. F., Somerville, R. S., Hernquist, L., et al. 2006a, *ApJ*, 652, 864
- Hopkins, P. F., Hernquist, L., Cox, T. J., Robertson, B., & Springel, V. 2006b, *ApJS*, 163, 1
- Hopkins, A. M., Driver, S. P., Brough, S., et al. 2013, *MNRAS*, 430, 2047
- Hornschemeier, A. E., Bauer, F. E., Alexander, D. M., et al. 2003, *AJ*, 126, 575
- Hsu, L.-T., Salvato, M., Nandra, K., et al. 2014, *ApJ*, 796, 60
- Hütsi, G., Gilfanov, M., Kolodzig, A., & Sunyaev, R. 2014, *A&A*, 572, A28
- Ilbert, O., Capak, P., Salvato, M., et al. 2009, *ApJ*, 690, 1236
- James, J. B., Lewis, G. F., & Colless, M. 2007, *MNRAS*, 375, 128
- Koulouridis, E., Plionis, M., Melnyk, O., et al. 2014, *A&A*, 567, A83
- Koulouridis, E., Poggianti, B., Altieri, B., et al. 2016, *A&A*, 592, A11 (XXL Survey, XII)
- Koutoulidis, L., Plionis, M., Georgantopoulos, I., Fanidakis, N. 2013, *MNRAS*, 428, 1382
- Krumpe, M., Miyaji, T., Coil, A. L. 2010, *ApJ*, 713, 558
- Kuntz, K. D., & Snowden, S. L. 2008, *A&A*, 478, 575
- La Franca, F., Fiore, F., Vignali, C., et al. 2002, *ApJ*, 570, 100
- La Franca, F., Fiore, F., Comastri, A., et al. 2005, *ApJ*, 635, 864
- Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, *MNRAS*, 379, 1599
- Leccardi, A., & Molendi, S. 2008, *A&A*, 486, 359
- Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, *A&A*, 559, A14
- Lehmann, I., Hasinger, G., Schmidt, M., et al. 2001, *A&A*, 371, 833
- Lehmer, B. D., Xue, Y. Q., Brandt, W. N., et al. 2012, *ApJ*, 752, 46
- Lidman, C., Ardila, F., Owers, M., et al. 2016, *PASA*, 33, e001 (XXL Survey, XIV)
- Liske, J., Baldry, I. K., Driver, S. P., et al. 2015, *MNRAS*, 452, 2087
- Lusso, E., Comastri, A., Vignali, C., et al. 2011, *A&A*, 534, A110
- Maccacaro, T., Gioia, I. M., Wolter, A., Zamorani, G., & Stocke, J. T. 1988, *ApJ*, 326, 680
- Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, *AJ*, 115, 2285
- Mainieri, V., Bergeron, J., Hasinger, G., et al. 2002, *A&A*, 393, 425
- Mainieri, V., Hasinger, G., Cappelluti, N., et al. 2007, *ApJS*, 172, 368
- Marulli, F., Bonoli, S., Branchini, E., et al. 2009, *MNRAS*, 396, 1404
- Masters, D., Capak, P., Salvato, M., et al. 2012, *ApJ*, 755, 169
- Mateos, S., Warwick, R. S., Carrera, F. J., et al. 2008, *A&A*, 492, 51
- Melnyk, O., Plionis, M., Elyiv, A., et al. 2013, *A&A*, 557, A81
- Menanteau, F., Hughes, J. P., Jimenez, R., et al. 2009, *ApJ*, 698, 1221
- Menzel, M.-L., Merloni, A., Georgakakis, A., et al. 2016, *MNRAS*, 457, 110
- Merloni, A., Predehl, P., Becker, W., et al. 2012, *ArXiv e-prints* [[arXiv:1209.3114](https://arxiv.org/abs/1209.3114)]
- Merloni, A., Bongiorno, A., Brusa, M., et al. 2014, *MNRAS*, 437, 3550
- Moutard, T., Arnouts, S., Ilbert, O., et al. 2016a, *A&A*, 590, A102
- Moutard, T., Arnouts, S., Ilbert, O., et al. 2016b, *A&A*, 590, A103
- Miyaji, T., Hasinger, G., & Schmidt, M. 2000, *A&A*, 353, 25
- Miyaji, T., Hasinger, G., Salvato, M., et al. 2015, *ApJ*, 804, 104
- Nandra, K., Georgantopoulos, I., Ptak, A., & Turner, T. J. 2003, *ApJ*, 582, 615
- Nandra, K., Laird, E. S., Aird, J. A., et al. 2015, *ApJS*, 220, 10
- Pacaud, F., Pierre, M., Refregier, A., et al. 2006, *MNRAS*, 372, 578
- Pacaud, F., Clerc, N., Giles, P. A., et al. 2016, *A&A*, 592, A2 (XXL Survey, II)
- Page, M. J., & Carrera, F. J. 2000, *MNRAS*, 311, 433
- Parihar, P., Vogeley, M. S., Gott, J. R., III, et al. 2014, *ApJ*, 796, 86
- Perola, G. C., Puccetti, S., Fiore, F., et al. 2004, *A&A*, 421, 491
- Pierre, M., Pacaud, F., Duc, P.-A., et al. 2006, *MNRAS*, 372, 591
- Pierre, M., Pacaud, F., Adami, C., et al. 2016, *A&A*, 592, A1 (XXL Survey, I)
- Piconcelli, E., Cappi, M., Bassani, L., Di Cocco, G., & Dadina, M. 2003, *A&A*, 412, 689
- Piconcelli, E., Jimenez-Bailón, E., Guainazzi, M., et al. 2005, *A&A*, 432, 15
- Plionis M., Rovilos M., Basilakos S., et al. 2008, *ApJ*, 674, L5
- Prevot, M. L., Lequeux, J., Prevot, L., et al. 1984, *A&A*, 132, 389
- Ranalli, P., Comastri, A., Vignali, C., et al. 2013, *A&A*, 555, A42
- Ranalli, P., Georgantopoulos, I., Corral, A., et al. 2015, *A&A*, 577, A121
- Ross, N. P., McGreer, I. D., White, M., et al. 2013, *ApJ*, 773, 14
- Rovilos, E., Fotopoulou, S., Salvato, M., et al. 2011, *A&A*, 529, A135
- Salvato, M., Hasinger, G., Ilbert, O., et al. 2009, *ApJ*, 690, 1250
- Salvato, M., Ilbert, O., Hasinger, G., et al. 2011, *ApJ*, 742, 61
- Schawinski, K., Virani, S., Simmons, B., et al. 2009, *ApJ*, 692, L19
- Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, *ApJ*, 500, 525
- Schmidt, M. 1968, *ApJ*, 151, 393
- Shankar, F., Weinberg, D. H., & Miralda-Escudé, J. 2009, *ApJ*, 690, 20
- Shemmer, O., Brandt, W. N., Vignali, C., et al. 2005, *ApJ*, 630, 729
- Silverman, J. D., Green, P. J., Barkhouse, W. A., et al. 2008, *ApJ*, 679, 118
- Simpson, C., Martínez-Sansigre, A., Rawlings, S., et al. 2006, *MNRAS*, 372, 741
- Simpson, C., Rawlings, S., Ivison, R., et al. 2012, *MNRAS*, 421, 3060
- Skilling, J. 2004, *AIP Phys. Conf. Ser.*, 735, 395
- Snowden, S. L., Egger, R., Freyberg, M. J., et al. 1997, *ApJ*, 485, 125
- Stalin, C. S., Petitjean, P., Srianand, R., et al. 2010, *MNRAS*, 401, 294
- Starck, J.-L., & Pierre, M. 1998, *A&AS*, 128, 397
- Sutherland, W., & Saunders, W. 1992, *MNRAS*, 259, 413
- Tozzi, P., Gilli, R., Mainieri, V., et al. 2006, *A&A*, 451, 457
- Ueda, Y., Takahashi, T., Ishisaki, Y., et al. 1999, *ApJ*, 524, L11
- Ueda, Y., Ishisaki, Y., Takahashi, T., et al. 2001, *ApJS*, 133, 1
- Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, *ApJ*, 598, 886
- Ueda, Y., Ishisaki, Y., Takahashi, T., et al. 2005, *ApJS*, 161, 185
- Ueda, Y., Hiroi, K., Isobe, N., et al. 2011, *PASJ*, 63, S937
- Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, *ApJ*, 786, 104
- van de Weygaert, R., & Bond, J. R. 2008, *A Pan-Chromatic View of Clusters of Galaxies and the Large-Scale Structure*, 740, 409
- Verrecchia, F., in't Zand, J. J. M., Giommi, P., et al. 2007, *A&A*, 472, 705
- Voges, W., Aschenbach, B., Boller, T., et al. 1999, *A&A*, 349, 389
- Xue, Y. Q., Luo, B., Brandt, W. N., et al. 2011, *ApJS*, 195, 10
- Watson, C. R., Kochanek, C. S., Forman, W. R., et al. 2009, *ApJ*, 696, 2206

<sup>1</sup> Department of Astronomy, University of Geneva, ch. d'Ecogia 16, 1290 Versoix, Switzerland  
e-mail: sotiria.fotopoulou@unige.ch

<sup>2</sup> Argelander Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, 53121 Bonn, Germany

<sup>3</sup> IAASARS, National Observatory of Athens, 15236 Penteli, Greece

<sup>4</sup> Lund Observatory, PO Box 43, 22100 Lund, Sweden

<sup>5</sup> Service d'Astrophysique AIM, CEA Saclay, 91191 Gif-sur-Yvette, France

<sup>6</sup> Aristotle University of Thessaloniki, Physics Department, 54124 Thessaloniki, Greece

<sup>7</sup> Instituto Nacional de Astrofísica Óptica y Electrónica, AP 51 y 216, 72000 Puebla, Mexico

<sup>8</sup> Université Aix Marseille, CNRS, LAM (Laboratoire d'Astrophysique de Marseille) 13388 UMR 7326, Marseille, France

<sup>9</sup> Osservatorio Astronomico di Roma (INAF), via Frascati 33, 00078 Monte Porzio Catone (Rome), Italy

<sup>10</sup> INAF-Osservatorio Astronomico di Bologna, via Ranzani 1, 40127 Bologna, Italy

<sup>11</sup> Dipartimento di Fisica e Astronomia, Università di Bologna, viale Bertini Pichat 6/2, 40127 Bologna, Italy

<sup>12</sup> INAF, IASF Milano, via Bassini 15, 20133 Milano, Italy

<sup>13</sup> Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, 81679 München, Germany

<sup>14</sup> Excellence Cluster Universe, Boltzmannstr. 2, 85748 Garching, Germany

<sup>15</sup> Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse 1, 85748 Garching, Germany

<sup>16</sup> Main Astronomical Observatory, Academy of Sciences of Ukraine, 27 Akademika Zabolotnoho St., 03680 Kyiv, Ukraine

<sup>17</sup> Australian Astronomical Observatory, PO Box 915, North Ryde NSW 1670, Australia

<sup>18</sup> Department of Physics, University of Zagreb, Bijenicka cesta 32, 10000 Zagreb, Croatia

<sup>19</sup> Astronomical Observatory, National Taras Shevchenko University of Kyiv, 3 Observatorna St., 04053 Kyiv, Ukraine

<sup>20</sup> Department of Astronomy and Space Sciences, Faculty of Science, Istanbul University, 34119 Istanbul, Turkey

<sup>21</sup> Department of Astronomy, University of Florida, Gainesville, FL 32611, USA

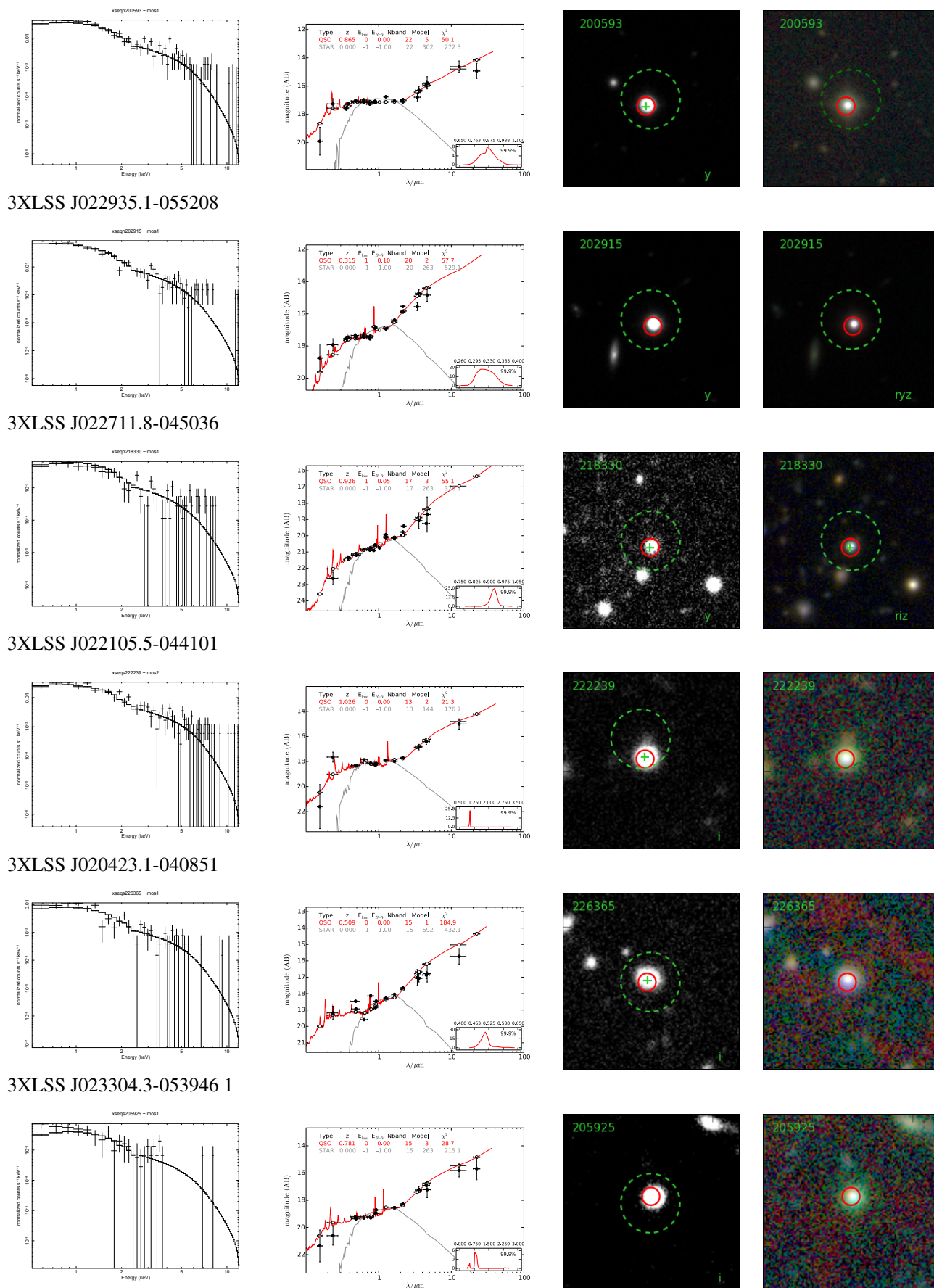
- <sup>22</sup> Laboratoire d'Astrophysique de Marseille, Pôle de l'Étoile, Site de Château-Gombert, 38 rue Frédéric Joliot-Curie, 13388 Marseille Cedex 13, France
- <sup>23</sup> Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK
- <sup>24</sup> H.H. Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, BS8 1TL, UK
- <sup>25</sup> Astrophysics and Cosmology Research Unit, University of KwaZulu-Natal, 4041 Durban, South Africa
- <sup>26</sup> Chalmers University of Technology Onsala Space Observatory, 439 92 Onsala, Sweden
- <sup>27</sup> INAF–Osservatorio Astronomico di Brera, via Brera, 28, 20159 Milano, Italy
- <sup>28</sup> Universität Hamburg, Hamburger Sternwarte, Gojenbergsweg 112, 21029 Hamburg, Germany
- <sup>29</sup> Laboratoire Lagrange, UMR 7293, Université de Nice Sophia Antipolis, CNRS, Observatoire de la Côte d'Azur, 06304 Nice, France
- <sup>30</sup> National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, 1205 W. Clark St., Urbana, IL 61801, USA
- <sup>31</sup> Department of Astronomy, University of Illinois at Urbana-Champaign, W. Green Street, Urbana, IL 61801, USA
- <sup>32</sup> Department of Physics and Astronomy, Macquarie University, NSW 2109, Australia and Australian Astronomical Observatory PO Box 915, North Ryde NSW 1670, Australia
- <sup>33</sup> INAF–Astronomical Observatory of Padova, vicolo dell'Osservatorio 5, 35122 Padova, Italy
- <sup>34</sup> European Southern Observatory, Alonso de Cordova 3107, Vitacura, Santiago de Chile, Chile
- <sup>35</sup> University of California, Davis, US
- <sup>36</sup> Astrophysics Department Max Planck Institut für Kernphysik, Saupfercheckweg 1, 69117 Heidelberg, Germany
- <sup>37</sup> Department of Physics and Astronomy, University of Victoria, 3800 Finnerty Road, Victoria, BC, Canada

## Appendix A: XXL-1000-AGN image cutouts

In this appendix we present example X-ray spectra, multiwavelength SEDs, and cutout images for a total of 30 sources from the XXL-1000-AGN sample. We group them according to the Random Forest classification: QSO, AGN, starburst, star forming, and passive. We note that the classification refers to the best class for the photo- $z$  estimation (see also Sect. 4.3). In each class we present three sources from the XXL-N field (top panels) and three sources from the XXL-S (bottom panels), selected randomly from our sample where the only constraint is that it must have a high S/N X-ray spectrum. From left to right the images are:

1. X-ray spectrum, either PN or MOS as stated on the image.
2. SED from the far-ultraviolet to mid-infrared including the best fit model (red line) and the star model (grey line). The black dots show the available observations, while the white dots show the model magnitudes at the best fit redshift. The insert in the SED plot shows 99.9% of the photo- $z$  PDF.
3. We also provide a single-filter cutout in the  $i$ -band (XXL-N:CFHTLS, XXL-S:BCS) unless otherwise stated.

The image cutouts are  $30'' \times 30''$  and the orientation is north-up, east-left. The dashed green circle is centred on the X-ray position and has a radius of  $5''$ . The circle shows the chosen counterpart and has a radius of  $1.5''$ .



**Fig. A.1.** Sources classified as QSO. The panels from *left to right* are: X-ray spectrum, SED from far-ultraviolet to mid-infrared with best fit photo- $z$  solution (red line) and star solution (grey line). The *inset* is the 99.9% PDF of the photo- $z$  solution. Single-filter image, three-colour composite.



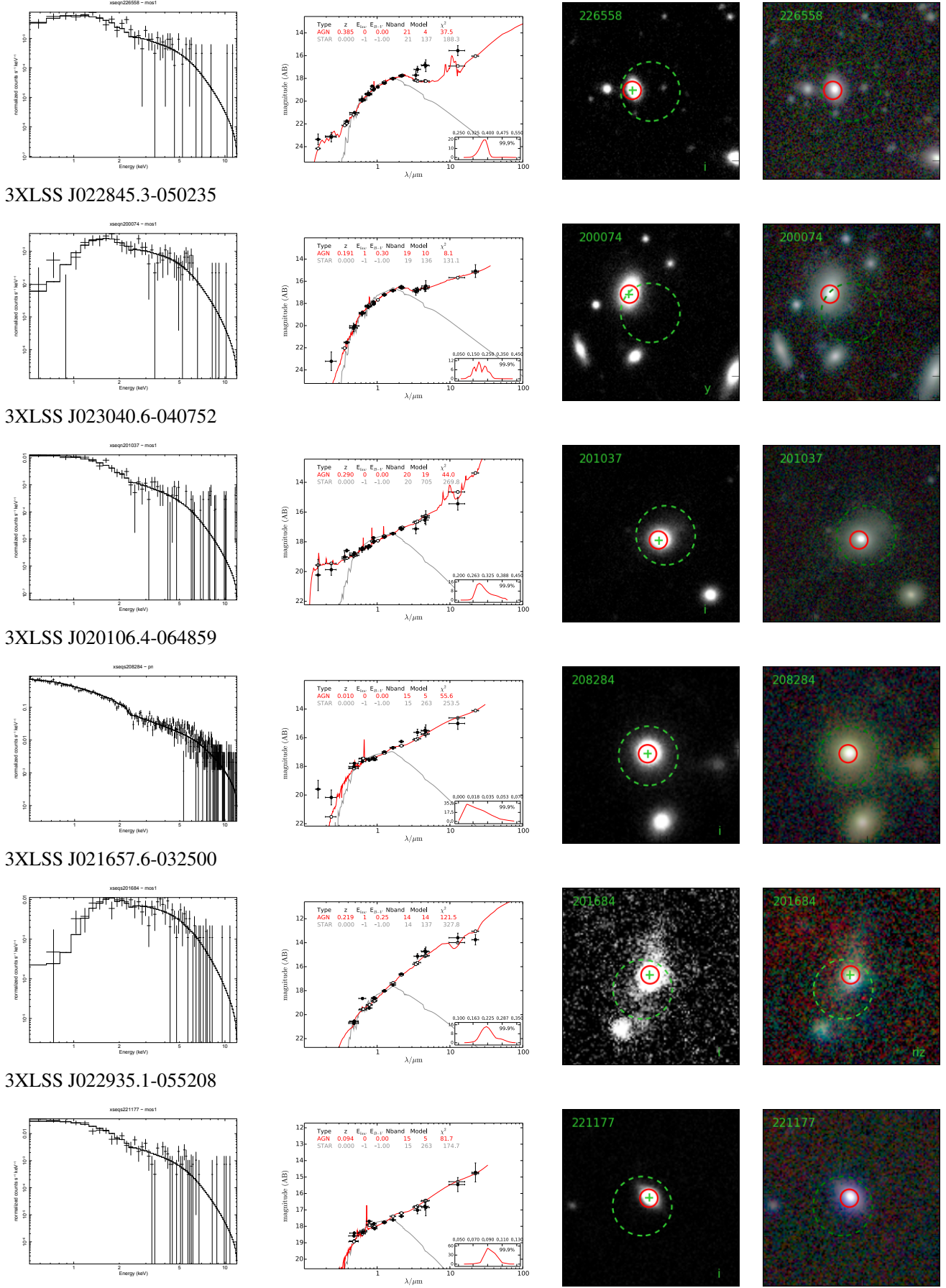


Fig. A.2. Sources classified as AGN. Panels as in Fig. A.1.

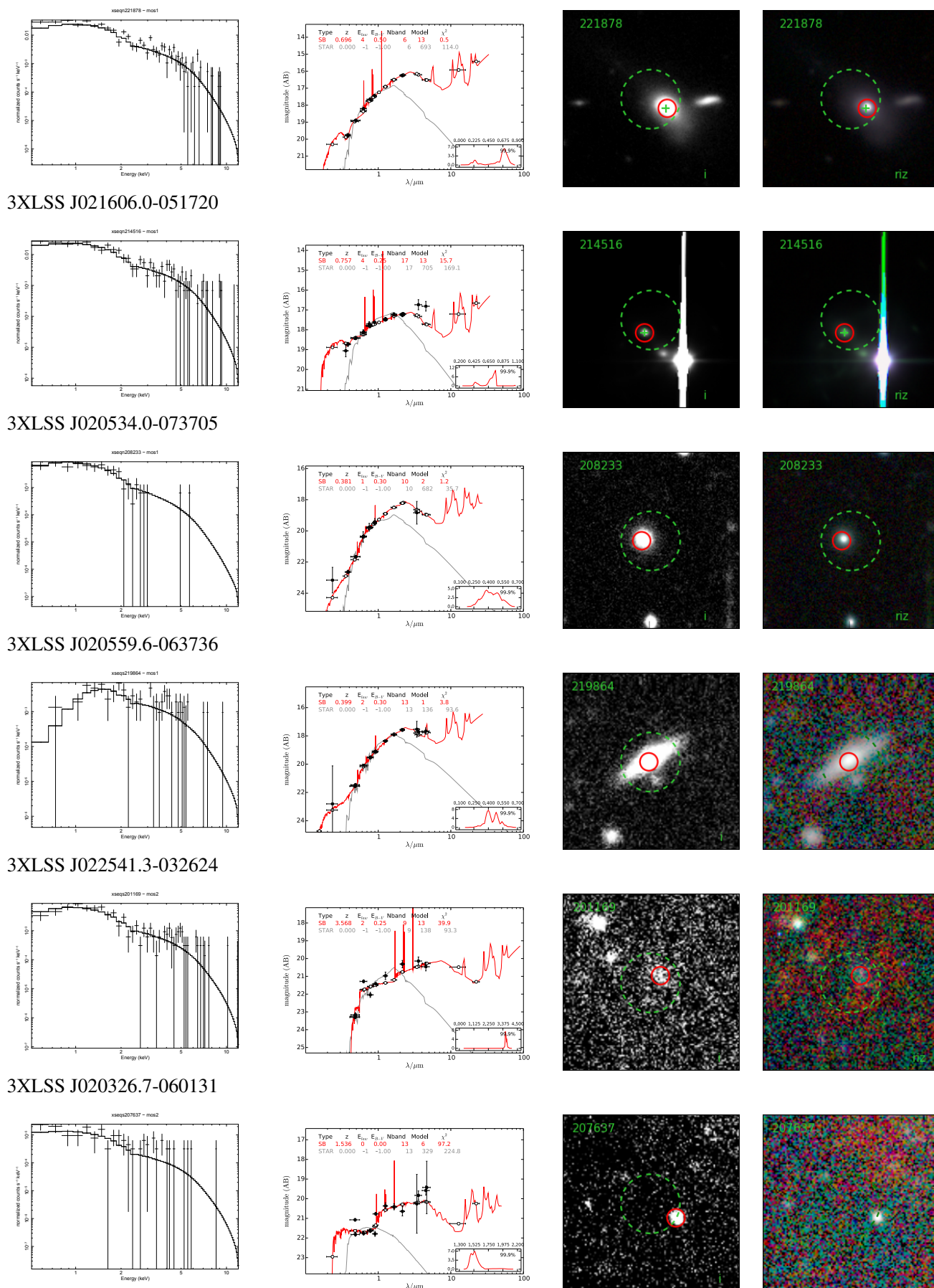


Fig. A.3. Sources classified as starburst. Panels as in Fig. A.1.

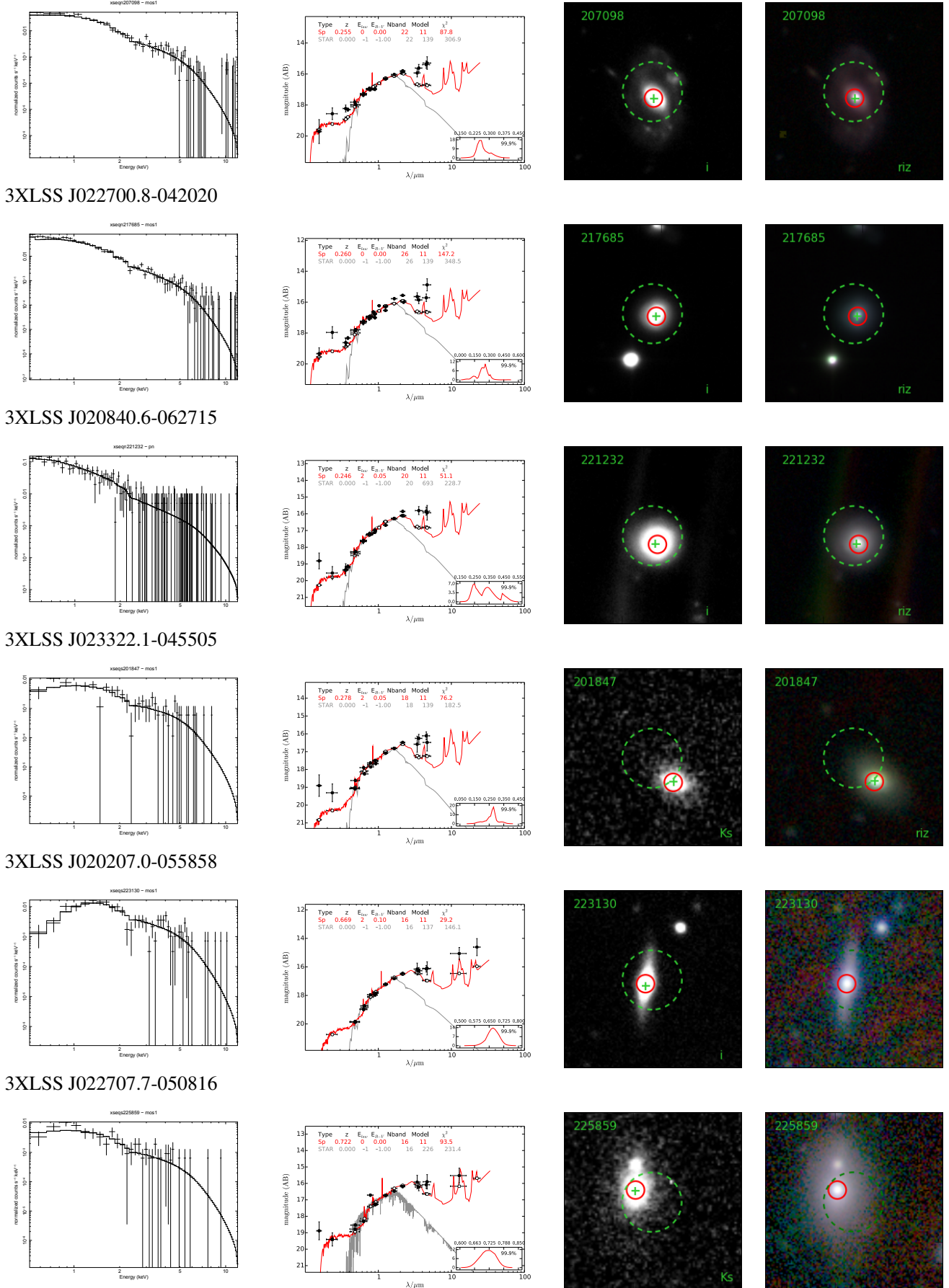


Fig. A.4. Sources classified as star forming. Panels as in Fig. A.1.



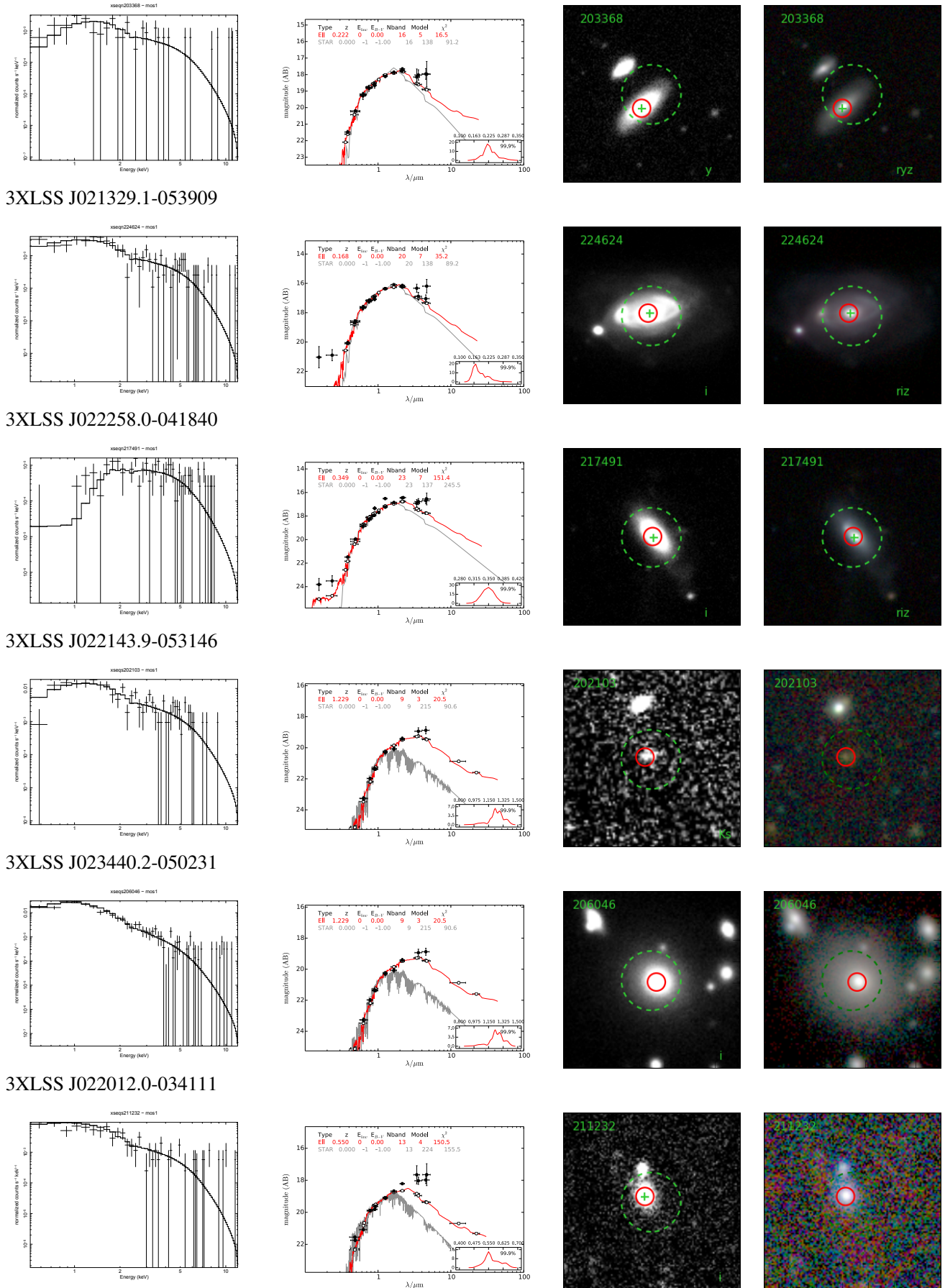


Fig. A.5. Sources classified as passive. Panels as in Fig. A.1.



**Appendix B: Probability density functions for  $\log N_{\text{H}}$** 

We provide the probability density functions (PDFs) of the  $\log N_{\text{H}}$  in four hardness ratio (HR) bins. These curves can be used to draw random values of the  $\log N_{\text{H}}$  when only HR is available for an X-ray source.

**Table B.1.** Tabulated values of the stacked  $\log N_{\text{H}}$  PDF for four HR bins as described in Sect. 5.1 and presented in Fig. 6b.

$\log N_{\text{H}}$	Hardness ratio bin			
	(-1.0, -0.25)	(-0.25, 0.25)	(0.25, 0.75)	(0.75, 1.00)
19.08	0.271	0.060	0.013	0.006
19.26	0.411	0.099	0.021	0.014
19.43	0.442	0.110	0.023	0.017
19.61	0.455	0.115	0.025	0.020
19.78	0.456	0.118	0.027	0.023
19.96	0.445	0.118	0.028	0.025
20.13	0.428	0.116	0.031	0.025
20.31	0.402	0.113	0.033	0.023
20.48	0.375	0.109	0.035	0.021
20.66	0.355	0.107	0.038	0.018
20.83	0.337	0.121	0.040	0.015
21.01	0.317	0.137	0.037	0.013
21.18	0.302	0.151	0.035	0.012
21.36	0.249	0.224	0.037	0.013
21.53	0.202	0.351	0.050	0.013
21.71	0.130	0.531	0.077	0.014
21.88	0.087	0.633	0.145	0.015
22.06	0.047	0.581	0.336	0.019
22.23	0.039	0.530	0.687	0.033
22.41	0.039	0.401	0.862	0.077
22.58	0.028	0.360	0.994	0.222
22.76	0.008	0.246	0.825	0.632
22.93	0.004	0.141	0.490	1.004
23.11	0.002	0.111	0.365	0.865
23.28	0.001	0.062	0.208	0.837
23.46	0.001	0.024	0.108	0.707
23.63	0.001	0.018	0.085	0.444
23.81	0.001	0.017	0.039	0.294
23.98	0.001	0.015	0.006	0.197
24.16	0.001	0.009	0.002	0.072
24.33	0.001	0.002	0.002	0.009
24.51	4.3E-4	5.5E-4	0.001	0.00
24.68	0.00	0.00	0.00	0.00

