Astronomy & Astrophysics

Active galactic nucleus X-ray luminosity function and absorption function in the Early Universe (3 ≤ *z* ≤ 6)[*]

E. Pouliasis[1][ID], A. Ruiz[1][ID], I. Georgantopoulos[1], F. Vito[2][ID], R. Gilli[2], C. Vignali[3,2], Y. Ueda[4], E. Koulouridis[1][ID],
M. Akiyama[5][ID], S. Marchesi[3,2,6][ID], B. Laloux[1,7][ID], T. Nagao[8], S. Paltani[9], M. Pierre[10], Y. Toba[11,12,8],
M. Habouzit[13,14][ID], B. Vijarnwannaluk[5,12], and C. Garrel[10]

1. IAASARS, National Observatory of Athens, Ioannou Metaxa and Vasileos Pavlou, GR-15236 Athens, Greece
   e-mail: epouliasis@noa.gr
2. INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Gobetti 93/3, 40129 Bologna, Italy
3. Università di Bologna, Dip. di Fisica e Astronomia "A. Righi", Via P. Gobetti 93/2, 40129 Bologna, Italy
4. Department of Astronomy, Kyoto University, Kitashirakawa-Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan
5. Astronomical Institute, Tohoku University, 6-3 Aramaki, Aoba-ku, Senda 980-8578, Japan
6. Department of Physics and Astronomy, Clemson University, Kinard Lab of Physics, Clemson, SC 29634, USA
7. Centre for Extragalactic Astronomy, Department of Physics, Durham University, Durham, UK
8. Research Center for Space and Cosmic Evolution, Ehime University, 2-5 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan
9. Department of Astronomy, University of Geneva, ch. d'Écogia 16, CH-1290 Versoix, Switzerland
10. AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, 91191 Gif-sur-Yvette, France
11. National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
12. Academia Sinica Institute of Astronomy and Astrophysics, 11F of Astronomy-Mathematics Building, AS/NTU, No.1, Section 4, Roosevelt Road, Taipei 10617, Taiwan
13. Zentrum für Astronomie der Universität Heidelberg, ITA, Albert-Ueberle-Str. 2, 69120 Heidelberg, Germany
14. Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany

ABSTRACT

The X-ray luminosity function (XLF) of active galactic nuclei (AGN) offers a robust tool to study the evolution and the growth of the supermassive black-hole population over cosmic time. Owing to the limited area probed by X-ray surveys, optical surveys are routinely used to probe the accretion in the high-redshift Universe $z \geq 3$. However, optical surveys may be incomplete because they are strongly affected by dust redenning. In this work we derive the XLF and its evolution at high redshifts ($z \geq 3$) using a large sample of AGN selected in different fields with various areas and depths covering a wide range of luminosities. Additionally, we put the tightest yet constraints on the absorption function in this redshift regime. In particular, we used more than 600 soft X-ray selected (0.5−2 keV) high-z sources in the *Chandra* deep fields, the *Chandra* COSMOS Legacy survey, and the XMM-XXL northern field. We derived the X-ray spectral properties for all sources via spectral fitting, using a consistent technique and model. To model the parametric form of the XLF and the absorption function, we used a Bayesian methodology, allowing us to correctly propagate the uncertainties for the observed X-ray properties of our sources and also the absorption effects. The evolution of XLF is in agreement with a pure density evolution model similar to what is witnessed at optical wavelengths, although a luminosity-dependent density evolution model cannot be securely ruled out. A large fraction (∼60%) of our sources are absorbed by column densities of $N_H \geq 10^{23}$ cm^{-2}, while ∼17% of the sources are Compton-Thick. Our results favour a scenario where both the interstellar medium of the host and the AGN torus contribute to the obscuration. The derived black hole accretion rate density is roughly in agreement with the large-scale cosmological hydrodynamical simulations, if one takes into account the results that the X-ray AGN are hosted by massive galaxies, while it differs from that derived using JWST data. The latter could be due to the differences in the AGN and host-galaxy properties.

Key words. galaxies: active – galaxies: high-redshift – galaxies: luminosity function, mass function – quasars: supermassive black holes – early Universe – X-rays: galaxies

1. Introduction

In the last several years it has become clear that most galaxies have a supermassive black hole (SMBH) in their centre (e.g. Kormendy & Ho 2013). SMBHs are active, and are known as active galactic nuclei (AGN) if there is material falling towards the centre of their galaxies. In most cases the in-falling material creates a geometrically thin, optically thick (Shakura & Sunyaev 1973) accretion disk producing copious amounts of radiation in the extreme ultraviolet (EUV) part of the spectrum. Alternatively, the accretion may be radiatively inefficient, channelling large amounts to the production of jets. All AGN produce intense X-ray radiation that equals roughly a few per cent of the bolometric luminosity (Lusso et al. 2012; Duras et al. 2020). The X-ray emission is believed to originate from Compton up-scattering of the UV accretion disk photons (e.g. Haardt & Maraschi 1991) on a hot electron corona with a mean temperature of $kT_e \sim 65 \pm 10$ keV (Akylas & Georgantopoulos 2021; Kamraj et al. 2022).

It is evident that the ubiquitous X-ray emission provides one of the most robust ways to detect AGN (Brandt & Alexander 2015) because X-ray wavelengths are hardly affected by obscuration (Burlon et al. 2011; Hickox & Alexander 2018; Georgantopoulos & Akylas 2019; Mountrichas et al. 2020; Georgakakis et al. 2020; Toba et al. 2022). Moreover, in most cases the contaminating star formation contributes a small fraction of the X-ray emission. The deepest X-ray observations in the *Chandra* deep field (Luo et al. 2017) have revealed a surface density of about 30 000 sources per square degree, where AGN form the vast majority of these sources. In contrast, the optical AGN surveys, for example the much shallower ($g < 22.5$ mag) Sloan Digital Sky Survey (SDSS, Pâris et al. 2018), reach a surface density of fewer than 200 luminous broad-line AGN per square degree.

The *Chandra* (Weisskopf et al. 2000) and *XMM-Newton* (Jansen et al. 2001) X-ray surveys in the $2-10$ keV band, have allowed us to study in detail the AGN demographics up to redshifts of $z \sim 3-4$ (e.g. Ueda et al. 2003, 2014; Aird et al. 2015; Ranalli et al. 2015; Miyaji et al. 2015; Buchner et al. 2015; Georgakakis et al. 2017; Peca et al. 2023; Laloux et al. 2023). The AGN luminosity function can be well described by a double power law which evolves with redshift according to a luminosity-dependent model (e.g. Ueda et al. 2014). According to this model the AGN evolution follows a "cosmic downsizing" pattern in the sense that the most luminous AGN ($\log L_X(2-10$ keV$)[$erg s$^{-1}] = 45-47$) formed first ($z \sim 2$), while the less luminous AGN ($\log L_X(2-10$ keV$)[$erg s$^{-1}] = 42-43$) have the peak of their redshift distribution at redshifts below one (e.g. Ueda et al. 2014; Aird et al. 2015). The above authors compare the evolution of the galaxy star formation rate density and the black hole accretion density (BHAD), as derived from the X-ray luminosity function. Although both peak at $z \sim 2$, at higher redshifts ($z = 4-5$) the BHAD presents a much stronger decline over redshift compared to the star formation rate density (SFRD). Thus, galaxy growth may precede the build up of their central SMBHs in the Early Universe (Aird et al. 2015). Alternatively, the SMBH may have formed massive enough, and thus they do not need high accretion rates to reach the local $M_{BH}-M\star$ relation. However, the X-ray data are quite scarce at these redshifts, and therefore this result awaits confirmation.

At redshifts higher than $z = 3$, X-ray surveys have harvested limited AGN samples because of the limited sky area covered. The density of high-redshift AGN is very low (~ 1 Gpc^{-3}, De Rosa et al. 2014) and therefore large areas need to be probed. The X-ray luminosity function at high redshifts ($3 < z < 6$) was derived by Vito et al. (2014, 2018), and Georgakakis et al. (2015). Vito et al. (2018) used about 100 X-ray sources from the CDF-N and CDF-S and focused on the faint end of the luminosity function, finding a very high fraction of obscured sources $\sim 0.6-0.8$ with column densities $\log N_H[$cm$^{-2}] \geq 23$. The ongoing all-sky extended ROentgen Survey with an Imaging Telescope Array (eROSITA, Predehl et al. 2021) is expected to facilitate the search for high-redshift AGN owing to its large grasp (field of view multiplied by effective area). After four years of operation, a few examples of very luminous $z > 6$ AGN have been identified with the eROSITA detector (Wolf et al. 2021, 2023; Medvedev et al. 2020), the majority of them radio-loud. The X-ray telescope on board the Niels Gehrels Swift mission has provided yet another $z > 6$ AGN (Barlow-Hall et al. 2023). The serendipitous *XMM-Newton* catalogue (Webb et al. 2020) provides another rich resource for detecting high-redshift AGN. Until December 2022, 657 000 unique sources had been detected covering an area of about 1300 deg^2. The next release of the

XMM-Newton serendipitous source catalogue (Webb et al. 2023) is expected to provide photometric redshifts derived using deep optical photometry, for example from the Dark Energy Survey (Abbott et al. 2021). In the near future, deep near-IR data from the *Euclid* mission (Euclid Collaboration 2022) will help to provide even more accurate photometric redshifts at $z > 4$. Nevertheless, the redshift confirmation of X-ray selected sources still requires the spectroscopic follow-up of the optical counterpart. This task is particularly difficult at high redshift because of the faintness of the optical counterparts.

In contrast to the relatively limited advances at X-ray wavelengths, at redshifts $z > 3$ the optical surveys have discovered high numbers of broad-line AGN, thanks to the availability of wide-field (i.e. $\sim 10^4$ deg^2) optical/near-infrared (NIR) surveys, such as the SDSS (Jiang et al. 2016), the UKIRT Infrared Deep Sky Survey (UKIDSS, Mortlock et al. 2011), the Canada-France High-redshift Quasar Survey (CFHQS, Willott et al. 2010), and the Panoramic Survey Telescope & Rapid Response System (Bañados et al. 2018). These led to the discovery of more than 300 broad-line AGN at $z > 5.8$ (Fan et al. 2023) when the Universe was less than 1 Gyr old. The highest redshift AGN discovered by the above surveys was identified at $z = 7.642$ (Wang et al. 2021). Deep optical surveys with the Subaru Hyper Supreme Cam (HSC, Miyazaki et al. 2018) such as the HSC Subaru Strategic Plan Survey (Aihara et al. 2018, 2019) allowed the detection of AGN at redshifts $z = 3-6$ at much fainter (>3 mag) absolute magnitudes (Akiyama et al. 2018; Niida et al. 2020; Matsuoka et al. 2022). The optical luminosity function decreases rapidly above redshifts $z \sim 3$. The drop in the AGN density is consistent with a pure density evolution model (McGreer et al. 2013; Castellano et al. 2023). Recently, the launch of the *James Webb* Space Telescope (JWST) allowed the detection of faint AGN up to redshifts of ~ 10 (e.g. Kocevski et al. 2023; Yang et al. 2023; Castellano et al. 2023; Bogdan et al. 2023). However, it is likely that the optical/UV luminosity function may be affected by large amounts of dust attenuation. When Lusso et al. (2023) convert the UV luminosity function to the X-ray band, they find that the X-ray luminosity function in the redshift range $z = 3-6$ is almost an order of magnitude higher than the optical. In this high-redshift regime it is likely that both the obscuring torus and the interstellar medium contribute to the obscuration Gilli et al. (2022). Interestingly, recent JWST mid-IR observations indicate that even the X-ray surveys may be affected by Compton-Thick obscuration. Yang et al. (2023), using the Mid-Infrared Instrument (MIRI) on board JWST, suggest a BHAD that is ~ 0.5 dex higher than the X-ray results at $z = 3-5$. At even higher redshifts the obscuration should be even higher. The BLUETIDES large volume cosmological simulations (Ni et al. 2020) show that at $z > 7$ a large fraction of AGN ($0.6-1$) could be heavily obscured by column densities $N_H \geq 10^{23}$ cm^{-2}.

Here, we visit anew the X-ray luminosity function at high redshifts ($z = 3-6$). We combine the most sensitive observations in X-rays in the *Chandra* deep fields with the COSMOS-Legacy 2 deg^2 *Chandra* observations and the large area of the *XMM-Newton*/XXL survey (25 deg^2). Our sample is the largest ever assembled in X-rays. It contains over 600 sources, including 100 and 30 sources above redshift $z = 4$ and $z = 5$, respectively. The selection of the high-z sample is presented in Sect. 2. In Sect. 3 we analyse the X-ray properties of our sample, and in Sect. 4 we explain the methodology and the models we used to constrain the X-ray luminosity function and the absorption function. Section 5 presents the best-fit parameters of our models, while Sect. 6 compares our results with other X-ray studies
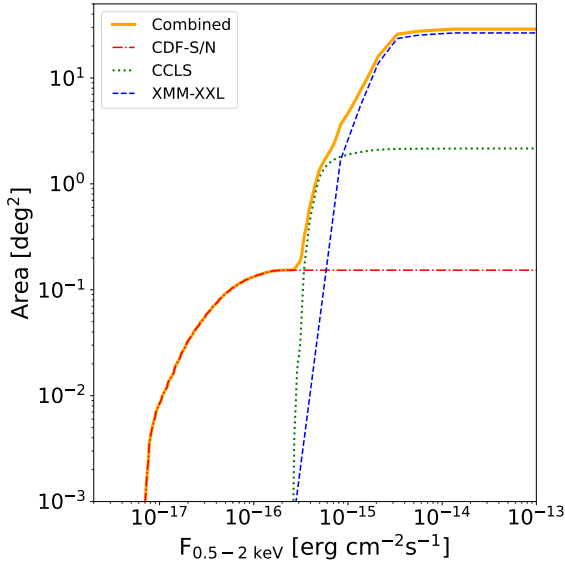
Fig. 1. X-ray sensitivity curves presented individually for the CDF-S/N, CCLS, and XMM-XXL-N fields. The total area curve is shown with the orange solid line.

in the literature and also the predicted values coming from the theoretical simulations. Then we constrain the space density and the black hole accretion density. In Sect. 7 we summarise the results. Throughout the paper we assume a standard ΛCDM cosmology with $H_0 = 70 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$ (Komatsu et al. 2009).

2. Sample selection

We derive the X-ray luminosity function in the rest-frame $2-10 \, \mathrm{keV}$ band and the absorption function of $z \geqslant 3.0$ AGN. Since the observed $0.5-2 \, \mathrm{keV}$ band corresponds to a rest-frame $2-8 \, \mathrm{keV}$ band at a redshift $z = 3$ and to the $3-12 \, \mathrm{keV}$ band at a redshift of $z = 5$, we construct our high-z sample using the soft-band detected sources. We selected our sample using three different X-ray surveys: the *Chandra* deep fields: South (Luo et al. 2017) and North (Xue et al. 2016), the *Chandra* COSMOS Legacy Survey (CCLS, Civano et al. 2016) and the northern region of the *XMM-Newton* XXL survey (Pierre et al. 2016, XMM-XXL-N). These surveys cover various sky areas and depths, allowing for the compilation of a high-z data set with the highest possible completeness with respect to luminosity, redshift and absorption column density ranges. The sensitivity-area curves for these surveys are shown in Fig. 1. The three surveys probe a large range of fluxes, allowing us to cover luminosities that span four orders of magnitude. Below, we give a brief description of the high-redshift AGN selection used in each field.

2.1. X-ray selected AGN from the XMM-XXL northern field

XMM-XXL-N covers an area of about $25 \, \mathrm{deg}^2$ at a depth of $6 \times 10^{-15} \, \mathrm{erg \, cm^{-2} \, s^{-1}}$ (at 3σ) in the soft band ($0.5-2 \, \mathrm{keV}$). Parts of this area have been observed in the framework of the XMM-SERVS survey ($5.4 \, \mathrm{deg}^2$, Chen et al. 2018) and the Subaru/*XMM-Newton* Deep Survey ($1.3 \, \mathrm{deg}^2$, Ueda et al. 2008), with sensitivity limits in the soft band of 1.7×10^{-15} and $6 \times 10^{-16} \, \mathrm{ergs \, cm^{-2} \, s^{-1}}$, respectively. We used the internal release of the XMM-XXL catalogue obtained with the V4.2 XXL pipeline that contains in total 15 547 X-ray sources. Restricting our area to the Hyper Suprime-Cam Subaru Strategic Program (HSC-

SSP, Miyazaki et al. 2018) coverage, we ended up with 10 232 soft-band detected sources. This reduced the area to $\sim 21 \, \mathrm{deg}^2$. In Pouliasis et al. (2022a), we present a catalogue of high-z AGN using the HSC colour–colour diagrams which are based on the Lyman break (drop-out) techniques. However, the above sample is limited to sources with $z \geq 3.5$ due to the lack of coverage in the u band. Furthermore, the Lyman-break method may miss a fraction of the high-redshift sources due to the different morphological selections or because of the redenned or host-galaxy dominated colours in the case of the obscured AGN (Le Fèvre et al. 2005; Paltani et al. 2007; Boutsia et al. 2021; Vijarnwannaluk et al. 2022). Thus, in order to increase the completeness of our sample we run the LePHARE algorithm (Arnouts et al. 1999; Ilbert et al. 2006) for all the soft-band X-ray sources in XMM-XXL-N. In Appendix A, we provide the whole procedure followed to estimate the photometric redshifts and gather all the available spectroscopic information. For the photo-z sample, we found that the fraction of outliers is $\eta = 20.9\%$ and the scatter between spectroscopic and photometric redshift is $\sigma_{\mathrm{NMAD}} = 0.07$. These statistics are similar or slightly better compared to previous X-ray studies (e.g. Salvato et al. 2022; Vijarnwannaluk et al. 2022). The final XMM-XXL-N sample of high-z sources includes 70 sources with spectroscopic redshifts and 438 sources that have a probability of more than 20% to be at $z \geq 3$ in their PDF(z). Among them, 319 sources have photometric redshifts with $z \geq 3$. Taking into account the sum of the PDF(z) of all sources at $z \geq 3$ in addition to the sources with spectroscopic redshift, the effective number of high-z sources is 390.7.

2.2. X-ray selected AGN from the CCLS

CCLS covers an area of $2.2 \, \mathrm{deg}^2$ with a mosaic of $\sim 180 \, \mathrm{ks}$ *Chandra* pointings, for a total observing time of about 4.6 Ms, reaching a depth of $2.2 \times 10^{-16} \, \mathrm{erg \, cm^{-2} \, s^{-1}}$ in the soft X-ray band ($0.5-2 \, \mathrm{keV}$). Marchesi et al. (2016a) provided optical and infrared identifications for the whole sample of 4016 X-ray sources in the CCLS and obtained photometric redshifts using the LePHARE SED fitting code. Marchesi et al. (2016b) built a sample of 174 sources within redshift $3 < z < 6.85$. 147 of them are detected in the soft band and were used in our work. Among them, 81 have available spectroscopic redshift. Additionally, we included lower-redshift sources which have a probability of more than 20% to lie at high redshift. The effective number of soft-band detected high-z sources is 151.2 in the CCLS field.

2.3. X-ray selected AGN from the Chandra deep fields

The *Chandra* deep fields (CDF), namely CDF-South and CDF-North, cover a total area of $0.15 \, \mathrm{deg}^2$. These are the deepest observations available in X-rays so far, reaching a depth of $6.4 \times 10^{-18} \, \mathrm{erg \, cm^{-2} \, s^{-1}}$ and $1.2 \times 10^{-17} \, \mathrm{erg \, cm^{-2} \, s^{-1}}$ in the soft band, respectively. In this work, we made use of the high-redshift AGN catalogue produced by Vito et al. (2018). They have gathered all the information for redshift estimates and provided a catalogue with X-ray sources having a probability larger than 20% of being in the high-redshift regime, $z \geq 3$. They only focused on the central regions of both fields, where vignetting effects and distortions of the point spread function (PSF) are minimal, and hence reaching an effective exposure above 1 Ms. The final area used was reduced ($\sim 330 \, \mathrm{arcmin}^2$ and $\sim 215 \, \mathrm{arcmin}^2$ in CDF-S and CDF-N, respectively). Since newer spectroscopic surveys and photometric data became available, we searched whether there are new derived spectra for these sources. In the south, we cross-matched (using the optical

Table 1. Number of sources in different redshift bins for each field and ensemble.

Field	$3 \le z \le 4$		$4 \le z \le 5$		$5 \le z \le 6$		$3 \le z \le 6$	
	z_{spec}	Total	z_{spec}	Total	z_{spec}	Total	z_{spec}	Total
CDF-S/N	38	75.1	1	11.6	1	2.6	40	89.3
CCLS	70	128.5	9	18.3	2	4.4	81	151.2
XMM-XXL-N	64	326.3	5	40.3	1	24.1	70	390.7
Combined	172	529.9	15	70.2	4	31.1	191	631.2

Notes. The number of sources with spectroscopic redshifts are reported, and also the total number of high-z sources (i.e. the spec-z sample plus the sum of PDF(z) of the photometric redshift sample in the redshift range $3.0 \le z \le 6.0$).
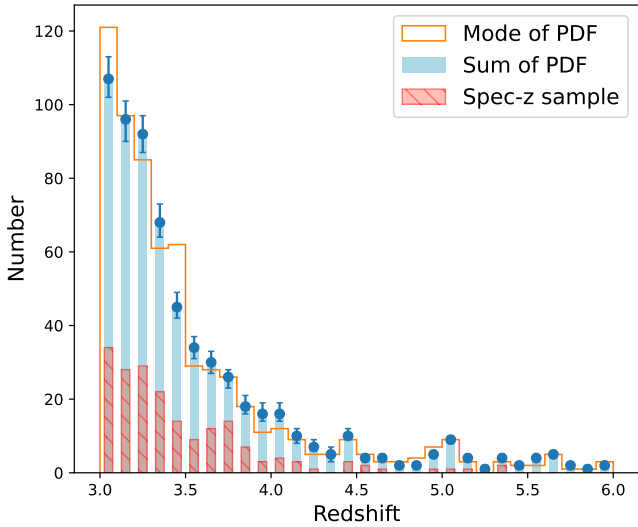
Fig. 2. Redshift distribution of our sample. The blue bars correspond to the sum of the PDF(z). The PDF(z) of sources with available spectroscopic redshift are represented by a Delta function centred at the spec-z value. The orange line represents the redshift histogram when taking into account only the mode of the PDF(z) for each source. The red hatched bars represent the spec-z sample.

counterparts of the X-ray sources and a search radius of $1''$) the X-ray catalogue with the final data release (DR4) of the VANDELS deep public ESO spectroscopic survey (Pentericci et al. 2018; McLure et al. 2018; Garilli et al. 2021) and also with the MUSE (Multi Unit Spectroscopic Explorer) *Hubble* Ultra Deep Field (HUDF) survey (Inami et al. 2017). Regarding CDF-North, we examined the sources with several catalogues in the literature (e.g. Momcheva et al. 2016). At the end, using the new spectroscopic information, we updated the previous photometric redshifts for twelve high-z sources. Ten out of twelve sources remained in the high-z regime, while we excluded two sources for which the new spec-z was below $z < 3$. In addition, we included in the sample six new high-z sources. In total, our sample contains 93 sources with $z \ge 3$. Among them, 40 sources have spectroscopic redshifts and 53 sources have photometric redshift estimations. Taking into account the full PDF(z) of the sources that lack spectroscopic information, in addition to the spec-z sample, the effective number of high-z AGN is 89.3.

2.4. Summary of the high-z sample

Table 1 summarises the final numbers of high-z AGN in different redshift bins selected through our analysis. We also report the numbers derived in each field individually. There are in total

191 sources with secure spectroscopic redshift at $z \ge 3$, 19 sources at $z \ge 4$ and four sources with redshift $z \ge 5$. Using the derived probability density function of the photometric redshifts, PDF(z), we were able to select additionally 438 sources where the maximum of the PDF(z) is above 3. However, a closer look at the PDF(z) of these sources reveals that, while for most of them the probability is concentrated around a narrow redshift range, a non-negligible number of sources show broad or doubled-peaked distributions. Hence, basing our selection only on the maximum probability value would be too restrictive, leaving out of our sample a number of sources with a high likelihood of being at high redshift. We therefore decided to include all sources for which P($z \ge 3$) > 0.2 (i.e. there is at least a 20% chance for the source to be at high redshift). This threshold has been adopted by Vito et al. (2018) to prevent including sources that show long tails with extremely low probability. A higher cutoff, such as 0.5, excludes about 10% of the effective number of sources that are equally distributed in the redshift range of our analysis. Even in this case, the differences in the XLF and the absorption function results are negligible. Furthermore, since luminosity priors have been applied in the photometric redshift estimation procedure in all fields, we are confident that the probability threshold used does not overestimate the number counts in the high-z regime. Finally, taking into account all the above, the effective number count is ~631.2. Our sample of X-ray selected sources is the largest to date in the Early Universe ($3 \le z \le 6$).

Figure 2 presents the redshift distribution of the sample used in our analysis that is the sum of the PDF(z) of each source. The PDF(z) of sources with available spectroscopic redshift are represented by a Delta function centred at the spec-z value. The uncertainties correspond to 1-σ confidence level. For reference, we overplot the redshift histogram only considering the peaks of the PDF(z) in addition to the spec-z sample. The agreement between the two distributions is very good at all redshifts indicating that the majority of the photometric redshifts show narrow peaks in their PDF(z). The red-hatched histogram shows the distribution of the spec-z sample.

3. X-ray properties of the high-z AGN sample

In order to calculate the X-ray luminosity and absorption functions, we need accurate estimations of the X-ray properties of the sources in our sample. In particular, we need to estimate the hydrogen column density, N_{H}, and the intrinsic, absorption-corrected luminosity in the rest-frame $2-10$ keV band, L_X. Previous studies of the surveys used for building our high-z sample presented their own X-ray analysis for most of these sources (e.g. Vito et al. 2018; Marchesi et al. 2016c,a; Liu et al. 2017). However, each study used their own methodology (different models and fitting methods) and, in some cases, the X-ray properties for
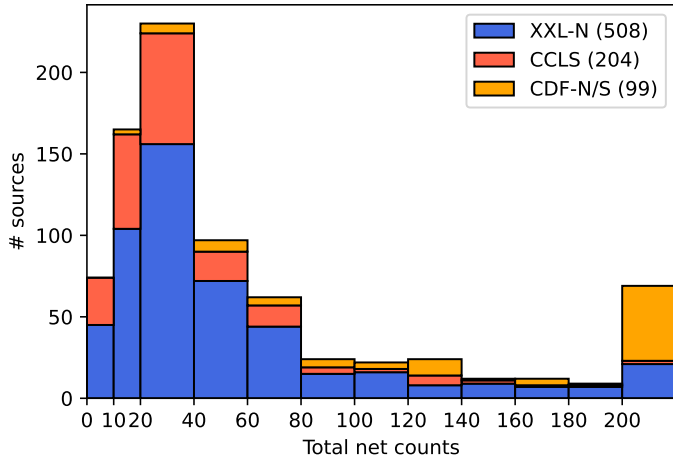
Fig. 3. Distribution of net counts (i.e. total counts minus background counts) for all the sources included in our spectral analysis. For the *XMM-Newton* spectra the counts correspond to the 0.3–10 keV interval, while for the *Chandra* spectra the counts correspond to 0.3–8 keV. The last bin contains all sources with more than 200 counts.

low-count sources are not derived via spectral analysis, but using hardness-ratios (HR). The latter, although useful for identifying highly absorbed sources, introduces high uncertainties, particularly in the N_H estimation and hides the underlying assumptions to convert between HR and N_H. However, the use of Cash statistic, which takes into account the Poisson nature of the data, in combination with modern, advanced fitting methods, such as Bayesian X-ray analysis (BXA, Buchner et al. 2014), handles low-count spectra accurately. Therefore, we have extracted X-ray spectra for all the sources in our sample and analysed them using an homogeneous approach.

Figure 3 shows the distribution of total net counts for the X-ray spectral products we obtained following the procedures described in Sect. 3.1. The distribution shows the total number of counts, background subtracted, for each source included in our final sample. In the case of sources with multiple spectra from different observations and/or cameras, all counts are added. The plot shows a significant fraction of sources with fewer than 20 net counts. As stated above, the BXA-based methodology we outline in Sect. 3.2 allows a rigorous statistical treatment even in the case of low-count sources, while extracting the maximum amount of information from the available observational data. For such spectra, most of the model parameters are not well constrained and their posterior distributions follow the selected initial priors. Since we use broad, non-informative priors, we do not impose any restrictive values and they are treated as the remaining sources. The large uncertainties of the spectral properties of these sources are then propagated to derived quantities like the X-ray luminosity, and fully taken into account in our methodology for estimating the luminosity function.

3.1. Spectral extraction

XMM-XXL-N. The XXL survey (Pierre et al. 2016) was built via a mosaic of multiple *XMM-Newton* observations with a high degree of overlapping between observations. This means that a single X-ray source can be observed multiple times. To analyse the X-ray spectra in the XXL sample, we used all the pointings contributing to a given source position. In particular, for each *XMM-Newton* observation and for each EPIC camera (PN, MOS1, MOS2) available, we extracted the source and back-

ground spectra, and their corresponding ancillary response files (ARFs) and response matrix files (RMFs). All observations were reprocessed using the *XMM-Newton* Science Analysis System (SAS, version 19) with the most updated calibration up to the date of analysis. The spectra were extracted following the standard procedure outlined by the SAS documentation. We used the `eregionanalyse` SAS task in order to calculate optimal elliptical source extraction regions centred at the position of each source. The orientation and eccentricity of the ellipse is defined by the PSF at the given position in the detector, and the final area of the region is calculated to maximise the signal-to-noise ratio (S/N) of the spectrum. Background spectra were extracted in circular regions of 30 arcsec centred at positions 1.5 arcmin away from the source. The exact position of the background region was selected to maximise the number of good pixels in the region (after masking areas outside the detector, bad pixels and other nearby detected sources) and to be as close as possible to the detector column of the source. For each source, all available spectra are not co-added, but instead they are fitted at once in the spectral fitting procedure outlined below in Sect. 3.2.

CCLS. We used CIAO 4.13 (Fruscione et al. 2006) for extracting the *Chandra* X-ray spectra for sources in the CCLS field, following the Laloux et al. (2023) methodology to optimise the S/N for each spectrum. A source extraction region is defined as a circular area at the position of the source. Different radii are tested, each corresponding to an encircled energy fraction (EEF) radius from 50% to 95%. The background region is defined as an annulus of width 17.5 arcsec with an inner radius 2.5 arcsec larger than the source region. Contributions from nearby sources were removed from the background region. The S/N is calculated for each EEF value from the number of counts in both regions, and the region with the maximum S/N was selected. For each source, the spectra from all individual *Chandra*/ ACIS-I observations were extracted with the `specextract` task and then combined using the `combine_spectra` task that also combines the ARF and RMF matrices.

CDF-S/N. For the sources in the CDF-S/N fields, we used the X-ray spectra from Vito et al. (2018) who extracted the *Chandra* spectra using the CIAO `ACISextract` package (Broos et al. 2010) following a similar procedure to Luo et al. (2017) and Xue et al. (2016). For the additional high-z sources identified in this work, we extracted the spectra following the procedure outlined for the CCLS field.

3.2. Spectral analysis

We fitted the X-ray spectra of the sources in our high-z sample using Sherpa 4.14.0 (Freeman et al. 2001; Burke et al. 2021) and BXA 4.1.1. BXA is a Python package that connects a nested-sampling Monte Carlo algorithm (Skilling 2004) as implemented in `UltraNest` (Buchner 2021), with the fitting software Sherpa, allowing a fully Bayesian approach for the X-ray spectral analysis. In this approach, the estimated background emission is not subtracted, but modelled using the method presented in Simmonds et al. (2018), who did a principal component analysis (PCA) of archival data for different X-ray missions. We applied the *Chandra*/*XMM-Newton* PCA models for fitting the background spectrum using a standard Cash minimisation with a Levenberg–Marquardt algorithm. Once a reasonable fit is obtained, the model parameters are fixed and the normalisation re-scaled to the area and effective exposure time of the source extraction region. This background model is incorporated into the total model used for fitting the source spectrum. To take into
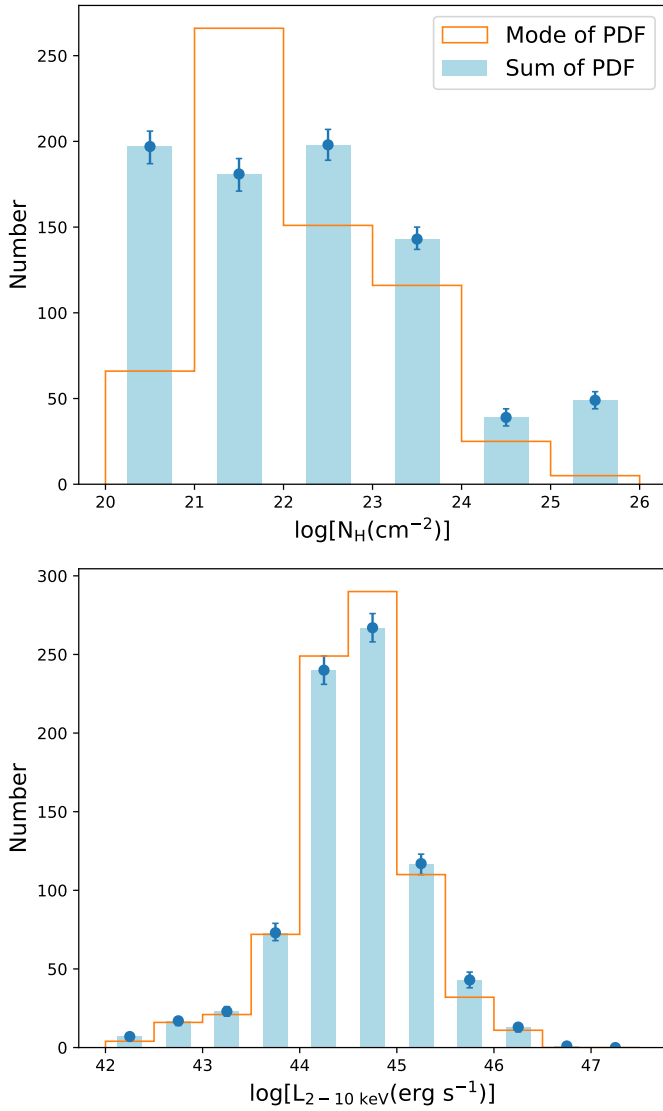
Fig. 4. Distributions of the hydrogen column density (top panel) and the 2−10 keV absorption corrected luminosity (bottom panel) for the high-z sample. The blue bars correspond to the sum of the probability density functions, while the orange lines represent the histograms of the properties when taking into account only the nominal values (mode of the posterior probability distributions).

account small discrepancies between the background spectrum in the extraction region and the background in the source region, we treated the normalisation of the background component in the source spectrum modelling as a free parameter, using a log-normal prior with mean equal to the scaled normalisation value and a dispersion of 50% of the value of the scaled normalisation.

We modelled the source emission using `UXClumpy` (Buchner et al. 2019), corrected with a multiplicative absorption component (`TBABS`, Wilms et al. 2000) to take into account the Galactic N_H along the line-of-sight of the source[1]. `UXClumpy` gives the reprocessed X-ray emission of the central AGN engine (a power law with an exponential cut-off) by a clumpy torus, where individual high-density gas clouds are distributed following a toroidal geometry. It includes three components: the transmitted emission through the absorbing clouds, the reflected

emission (including fluorescent lines) and the warm back-scattered emission containing mostly the incident power law from unobscured sight-lines. This model is suitable for both type 1 and type 2 AGN, since low inclination angles allow a direct view of the central, unabsorbed emission.

Our model has five free parameters: the $\log N_H$ along the line-of-sight of the observer, the photon index of the direct X-ray emission (Γ), the inclination angle of the torus (θ) with respect to the line-of-sight of the observer, the logarithm of the normalisations for the direct emission and the log of the fraction of the scattering component with respect to the direct emission ($\log f_s$). We used flat priors for $\log N_H$ (with limits between 20–26), θ (0–90°), $\log f_s$ (between −5 and −1.3, i.e. up to a 5% contribution for the scattered emission) and the normalisation. For Γ, we assumed a Gaussian prior with mean 1.95 and standard deviation 0.15, following the typical distribution of Γ expected for AGN (e.g. Nandra & Pounds 1994). In the case of sources with photometric redshifts, we followed the procedure of Ruiz et al. (2021) and treated the redshift as an additional free parameter using as a prior the corresponding PDF given by the photo-z estimation software. We kept the remaining parameters of `UXClumpy` fixed during the fit, at the default values of the model. The energy range we used for the fit was 0.3−10 keV for *XMM-Newton* spectra and 0.3−8 keV for *Chandra* spectra. We evaluated the goodness of our spectral fits using the Cash-based test proposed by Kaastra (2017). For a detailed presentation of this method, as well as a discussion of the reliability of our results and the informational gain we obtained for the parameters, see Appendix B.

One of the advantages of using BXA is that the final result of the fitting process is the full posterior distribution of the free parameters of our model, conserving all possible degeneracies and correlations between the parameters. Errors for each parameter can be calculated from the posterior via marginalisation. Derived quantities like observed fluxes or absorption corrected luminosities can be calculated using the full posterior, allowing us to correctly propagate the correlations between parameters and an accurate estimation of the errors of these derived quantities. In practical terms, the resulting BXA posteriors are equivalent to Monte Carlo Marko chains and they can be analysed using the tools already available in X-ray fitting software like Sherpa or XSPEC.

In Fig. 4, we show the absorption column density (upper) and the absorption-corrected X-ray luminosity (lower) distributions of our final high-z sample. We plot the sum of the posterior probability distribution functions of individual sources coming directly from the spectral analysis along with their uncertainties. The uncertainties of the derived parameters besides on the photon statistics also depend on the overall spectral shape and on the accuracy of the adopted redshift measurements. For comparison, we also show the histograms when considering only the nominal values (mode of the PDF) with the highest probability of the spectral analysis. The most prominent difference is found in the number density in the $\log N_H[\text{cm}^{-2}] = 20-23$ range. Using the sum of the posterior distributions, the number of sources in the first bins are of almost equal probability. This is something expected if we consider that at these high redshifts it is not possible to constrain the absorption below $\log N_H[\text{cm}^{-2}] = 23$ due to the shifted spectrum to higher energies. In contrast, if one uses the mode of the posterior distribution for each source would loose information, and hence, this would result in wrong conclusions. Furthermore, in the last bin of the N_H distribution there is an excess of sources when using the full posteriors that can be attributed to the tails at high values.

---

[1] We use `gdpyc` (Ruiz 2018) to calculate the Galactic N_H from the Leiden/Argentine/Bonn (LAB) survey (Kalberla et al. 2005).
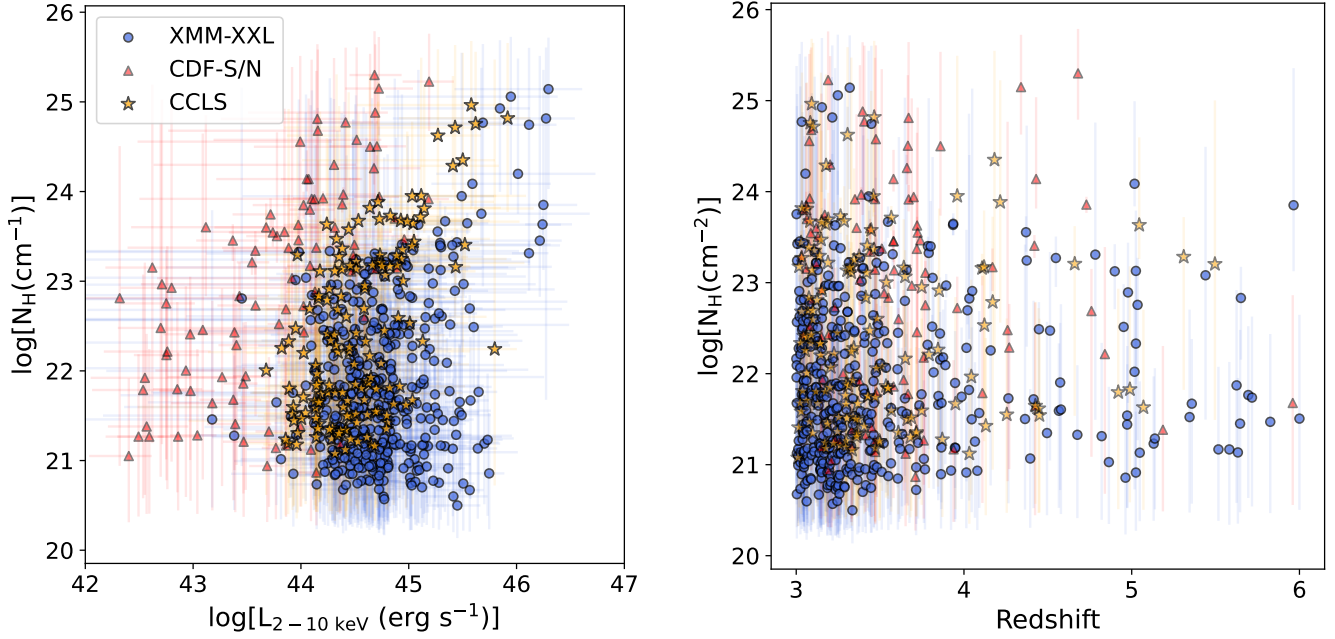
Fig. 5. Hydrogen column density vs X-ray absorption-corrected, rest-frame luminosity (left) and redshift (right) for the X-ray sources detected in the various fields, as indicated in the legend.

Figure 5 presents the distribution of the high-z sources (selected in the different X-ray surveys) on the intrinsic X-ray luminosity (absorption-corrected) versus redshift plane (left panel) and on the absorption column density versus intrinsic X-ray luminosity plane (right). Each data point corresponds to the most probable values of the 3-dimensional X-ray luminosity, column density and redshift probability distribution function of individual sources inferred from the X-ray spectral fits. This figure demonstrates the necessity of combining various surveys with a wide range of areas and depths that are complementary to each other to map with the highest possible completeness the full AGN population. The catalogue of all sources used in our analysis along with the derived spectral properties is only available in electronic form at the CDS.

4. X-ray luminosity and absorption functions

4.1. Survey area function

The sensitivity of X-ray surveys is not homogeneous, but decreases strongly towards faint fluxes (see Fig. 1). This introduces complex observational biases against sources with high N_H, high redshift or lower X-ray luminosities (e.g. highly absorbed sources show lower X-ray fluxes so they are less likely to be detected). Such biases must be quantified if we want an accurate estimation of the luminosity and absorption functions for the intrinsic population of the high-z AGN.

Using the UXCLUMPY model, we calculated the expected flux as a function of z, L_X and N_H assuming a constant value for the photon index, $\Gamma = 1.95$ (Nandra & Pounds 1994). The parameter space used here was within $\log L_X = 42-47$, $\log N_H = 20-26$ and $z = 3-6$ with 50 bins for each parameter[2]. The inclination angle was fixed to $i = 60$ deg and the torus opening to $\sigma = 30$ deg that is a mixture of type 1 ($\sigma = 15$ deg) and type 2 ($\sigma = 40$ deg) AGN (Buchner et al. 2019). Fixing these parame-

ters to different values does not affect significantly the detection probability of the sources. Then, we converted the fluxes into the corresponding area covered by each field by convolving with the area curves shown in Fig. 1. By normalising, we were able to obtain the probability of a source with specific column density and intrinsic luminosity being detected at a given redshift and survey.
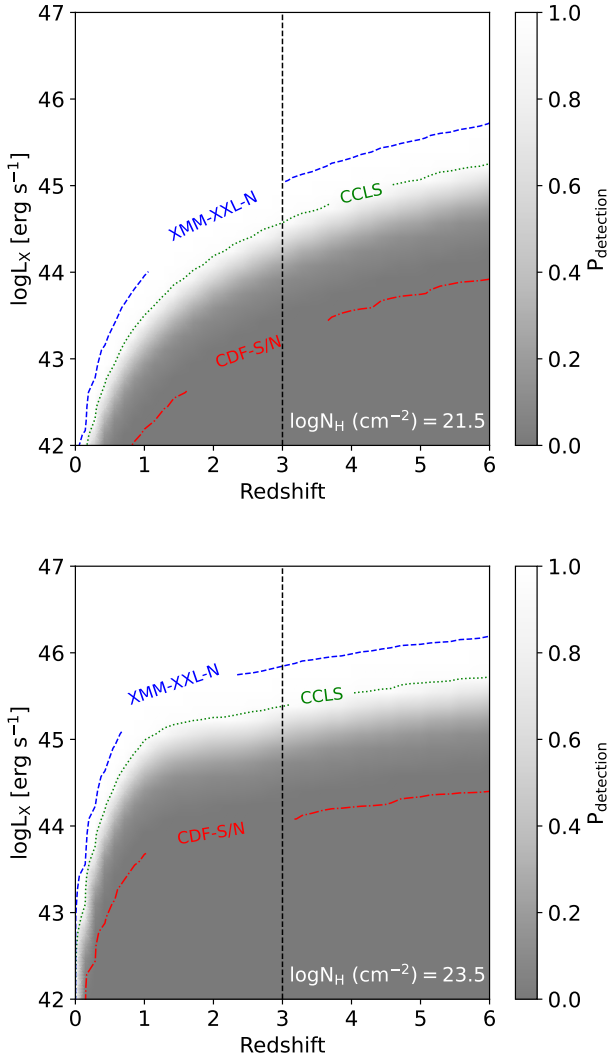
The upper (lower) panel in Fig. 6 shows the probability of detecting an unobscured (obscured) source with $\log N_H = 21.5$ (23.5) within the combined area of all fields (shaded region). For comparison, we show the sensitivity curves within the CDF-S/N (dashed-dotted line), CCLS (dotted line) and XMM-XXL-N (dashed line) fields individually. As expected, there is a rapid drop of the probability for sources at high redshifts and low luminosities. This trend is more obvious for the obscured population. For example, at a given redshift and luminosity, the probability of detecting a source with $\log N_H = 21.5$ is higher than considering higher column densities. Furthermore, the efficiency of detecting sources with lower intrinsic column densities and luminosities is higher in fields with deeper observations.

4.2. X-ray luminosity function

We define ϕ as the differential luminosity function of the AGN population in terms of $\log L_X$, since our high-z sample spans a wide luminosity range. By definition, ϕ is the number of sources N per comoving volume V and per logarithmic interval $\log L_X$ as a function of redshift, z, and luminosity, L_X:

$$\phi(L_X, z) = \frac{d\Phi(L_X, z)}{d\log L_X} = \frac{d^2 N(L_X, z)}{dV d\log L_X}. \tag{1}$$

We derived the analytical expression of the differential luminosity function by assuming a broken power law, which has been found to describe the shape well in the local and the nearby Universe (Maccacaro et al. 1983, 1984; Barger et al. 2005), and

---

[2] Along the paper, numerical values for the logarithm of N_H and L_X are always quoted in CGS units, i.e. $\log(N_H/\mathrm{cm}^{-2})$ and $\log(L_X/\mathrm{erg\,s}^{-1})$.

Fig. 6. Sensitivity maps of the total area as a function of redshift and intrinsic X-ray luminosity for a source with intrinsic column density of $\log N_{\rm H} = 21.5$ (upper) and $\log N_{\rm H} = 23.5$ (lower). For comparison, shown are the sensitivity maps at 99% within the CDFs, CCLS, and XMM-XXL-N independently.

is defined as:

$$\frac{{\rm d}\Phi(L_{\rm X}, z=0)}{{\rm d}\log L_{\rm X}} = A \times \left[\left(\frac{L_{\rm X}}{L_*}\right)^{\gamma_1} + \left(\frac{L_{\rm X}}{L_*}\right)^{\gamma_2} \right]^{-1}, \quad (2)$$

where A is a normalisation factor, L_* is the characteristic luminosity break, while γ_1 and γ_2 are the slopes of the power law before and after L_*, respectively (Miyaji et al. 2000; Hasinger et al. 2005).

In addition, we introduced the evolution of the luminosity function with the redshift testing different models. In particular, we adopted the pure density evolution model (PDE, Schmidt 1968) and the luminosity-dependent density evolution model (LDDE, Schmidt & Green 1983; Miyaji et al. 2000) that have been used extensively in the literature. Both can be expressed as:

$$\frac{{\rm d}\Phi(L_{\rm X}, z)}{{\rm d}\log L_{\rm X}} = \frac{{\rm d}\Phi(L_{\rm X}, z=0)}{{\rm d}\log L_{\rm X}} \times e(z) \quad (3)$$

where $e(z)$ is the factor that characterises the evolution with redshift and can be given as:

$$e(z) = \left(\frac{1+z}{1+z_c}\right)^{p_{\rm den}} \quad (4)$$

for the PDE model, while for the LDDE model there is an additional dependence on luminosity (Vito et al. 2014; Georgakakis et al. 2015) such that

$$e(z, L_{\rm X}) = \left(\frac{1+z}{1+z_c}\right)^{p_{\rm den}+\beta(\log L_{\rm X}-44)}. \quad (5)$$

The parameters $p_{\rm den}$ and β give the slope of the power law and the dependence on the luminosity, respectively. z_c is the critical redshift fixed at $z_c = 3$ (Vito et al. 2014; Georgakakis et al. 2015). The physical explanation of the PDE model is that the AGN population changes in numbers but its luminosity remains the same. Thus, only the normalisation of the XLF varies with redshift. The LDDE models assumes further that the variation of the AGN number density depends on the luminosity. This results in the change of the shape of the luminosity function.

For visualisation reasons, we also calculated the binned luminosity function of the high-z AGN by dividing the sample into luminosity, redshift and absorption column density bins. To construct the binned luminosity function, we used the Page & Carrera (2000) method that is an updated version of the $1/V_{\rm max}$ method (Schmidt 1968; Avni & Bahcall 1980). Thus, the binned luminosity function in a given range of redshift, luminosity, and hydrogen column density can be estimated by:

$$\phi(L_{\rm X}, z, N_{\rm H}) = \frac{\langle N \rangle}{\iiint \Omega(L_{\rm X}, z, N_{\rm H})\frac{{\rm d}V}{{\rm d}z}\,{\rm d}\log L_{\rm X}\,{\rm d}\log N_{\rm H}\,{\rm d}z}, \quad (6)$$

where $\langle N \rangle$ is the number of sources in the specific bin, ${\rm d}V/{\rm d}z$ is the differential comoving volume, and Ω is the survey sensitivity function defined in Sect. 4.1.

4.3. Absorption function

In Sect. 3, we derived the X-ray spectral properties of the individual high-z sources and presented the $\log N_{\rm H}$ distribution without taking into account observational biases (Fig. 4). Here, we formulate the intrinsic absorption distribution function of AGN taking into account the redshift and luminosity dependencies. Following the methodology of Ueda et al. (2003, 2014), Vijarnwannaluk et al. (2022), we model the $N_{\rm H}$ function, $f_{\rm abs}$, by combining flat-step functions for different $N_{\rm H}$ bins.

As mentioned in Sect. 3.2, our analysis of the *XMM-Newton* and *Chandra* spectra was limited in the low-energy range to $0.3\,{\rm keV}$ in the observer frame, which corresponds to a limit in the rest-frame of our sample of about $1.2-2\,{\rm keV}$. This means that we are not able to reliably constrain values of $\log N_{\rm H}$ below $\sim 22.5-23$, since absorption below these column densities only affects the X-ray spectrum below our observing energy range. Therefore, our absorption function is split into three $N_{\rm H}$ bins, as follows:

$$f_{\rm abs}(z, L_{\rm X}, N_{\rm H}) = \begin{cases} \frac{1}{3} - \frac{\varepsilon}{3(1+\varepsilon)}\psi(z, L_{\rm X}) & [20 \leq \log N_{\rm H} < 23] \\ \frac{\varepsilon}{1+\varepsilon}\psi(z, L_{\rm X}) & [23 \leq \log N_{\rm H} < 24] \\ \frac{f_{{\rm CTK},r}}{2}\,\psi(z, L_{\rm X}) & [24 \leq \log N_{\rm H} < 26]. \end{cases} \quad (7)$$

Table 2. Best-fit parameters of the XLF and absorption function in the cases of the LDDE and PDE models.

Parameter	Prior	Best value	
	XLF	LDDE	PDE
$\log A$	$-7, -3$	$-4.28^{+0.09}_{-0.10}$	$-4.28^{+0.09}_{-0.09}$
$\log L_*$	$42, 46$	$44.52^{+0.07}_{-0.07}$	$44.52^{+0.07}_{-0.07}$
γ_1	$-2, 2$	$0.29^{+0.08}_{-0.10}$	$0.21^{+0.08}_{-0.10}$
γ_2	$1, 6$	$2.38^{+0.15}_{-0.14}$	$2.23^{+0.14}_{-0.13}$
p_{den}	$-10, -3$	$-8.53^{+0.58}_{-0.65}$	$-7.35^{+0.39}_{-0.42}$
β	$-3, 5$	$2.18^{+0.79}_{-0.81}$	–
Absorption function			
ε	$0.1, 10$	$2.041^{+0.24}_{-0.24}$	$2.09^{+0.23}_{-0.23}$
a_2	$-2, 2$	$-0.28^{+1.57}_{-0.37}$	$-0.50^{+1.37}_{-0.21}$
ψ_3	$0, 5$	$3.66^{+0.95}_{-1.62}$	$3.99^{+0.74}_{-1.45}$
c	$-1, 1$	$0.28^{+0.42}_{-0.87}$	$0.39^{+0.32}_{-0.72}$
$f_{\mathrm{CTK},r}$	$0.03, 10$	$0.25^{+0.14}_{-0.22}$	$0.25^{+0.14}_{-0.23}$

Notes. The normalisation A and the break luminosity L_* are given in units of Mpc^{-3} and $\mathrm{erg\,s}^{-1}$, respectively.

Following the definitions of Vijarnwannaluk et al. (2022), the parameter ε is the ratio of sources with $23 \leq \log N_{\mathrm{H}} \leq 24$ to those with $22 \leq \log N_{\mathrm{H}} \leq 23$, while $f_{\mathrm{CTK},r}$ gives the relative ratio between Compton-Thick (CTK, $24 \leq \log N_{\mathrm{H}} \leq 26$) sources and absorbed Compton-Thin (CTN, $22 \leq \log N_{\mathrm{H}} \leq 24$) objects. The term $\psi(z, L_{\mathrm{X}})$ corresponds to the fraction of absorbed CTN AGN to the total number of AGN.

To compare with previous works (Ueda et al. 2014; Vijarnwannaluk et al. 2022), we use the following complex notation system where the redshift and luminosity dependence is contained in ψ, and it is parametrised using a linear dependence for $\log L_{\mathrm{X}}$:

$$\psi(z, L_{\mathrm{X}}) = \min(\psi_{\max}, \max(\psi_{43.75}(z) - c(\log L_{\mathrm{X}} - 43.75), \psi_{\min})), \tag{8}$$

where we used $\psi_{\min} = 0$ and $\psi_{\max} = 0.99$. The parameter c controls the luminosity dependence and $\psi_{43.75}(z)$ represents the absorption function of AGN at $\log L_{\mathrm{X}} = 43.75$ for a given redshift. $\psi_{43.75}(z)$ is well-constrain for $z < 2$ (Ueda et al. 2014), while above this redshift usually it is considered as a constant ($2 \leq z < 3$, Vijarnwannaluk et al. 2022). In this work, we define $\psi_{43.75}(z)$ for sources with $z \geq 3$, such as:

$$\psi_{43.75}(z \geq 3) = \psi_3 \times (1 + z)^{a_2}, \tag{9}$$

where ψ_3 is the absorption function at $\log L_{\mathrm{X}} = 43.75$ and $z = 3$, with a_2 being the evolution index. Both are free parameters to be determined from the analysis in the next section.

5. Fit and parameter estimations

In this work, we used Bayesian inference to estimate the parametric form of the X-ray luminosity function and the absorption function simultaneously. Given a data set of n observations, $D = \{d_i; i = 1, \dots, n\}$, and a model for the X-ray luminosity function defined by a set of parameters $\mathbf{\Theta}$, according to the Bayes'

theorem:

$$P(\mathbf{\Theta}|D) = \frac{P(D|\mathbf{\Theta})P(\mathbf{\Theta})}{P(D)}, \tag{10}$$

where $P(\mathbf{\Theta}|D)$, the posterior probability, is the probability of the selected model given the observational data; $\mathcal{L} = P(D|\mathbf{\Theta})$, the likelihood, is the probability of obtaining the observational data given the model; $P(\mathbf{\Theta})$, the prior, is the a priori probability for the parameters of the model, and $P(D) = \int P(\mathbf{\Theta}|D)\mathrm{d}\mathbf{\Theta}$ is the evidence of the model.

To derive the posterior probability distribution of the model parameters, we used the nested-sampling Monte Carlo algorithm MLFriends (Buchner 2016; Buchner & Bauer 2017), implemented in the UltraNest package. Nested sampling algorithms allow the posterior distribution of the model to be traced, given a data set, while at the same time calculating the Bayesian evidence. A direct estimate of the evidence is extremely useful for the comparison of the different XLF models via Bayes factors (see Sect. 6). Moreover, this Bayesian approach allows for a rigorous treatment of the uncertainties in the X-ray properties and photometric redshifts of the sources obtained in our X-ray analysis. During the inference process, we assumed flat priors for the model parameters, either uniform or log-uniform, that span a reasonably broad range of the parameter space according to previous studies in the literature (Vito et al. 2014; Vijarnwannaluk et al. 2022). In Table 2 we provide the minimum and maximum values allowed in the flat priors we used for the parameters of the XLF models.

Loredo (2004) proposed that the likelihood of observing a given data set can be constructed as the product of the probabilities of observing each individual source times the probability of not detecting any other source. Following the detailed derivation of Buchner et al. (2015; see also Aird et al. 2015; Georgakakis et al. 2015), the log-likelihood of this process can be written as:

$$\ln \mathcal{L}(\{d_i\}|\mathbf{\Theta}) =$$
$$- \lambda + \sum_i \ln \iiint P_i(L_{\mathrm{X}}, z, N_{\mathrm{H}}|\mathbf{\Theta}) \frac{\mathrm{d}V}{\mathrm{d}z} \mathrm{dlog}N_{\mathrm{H}} \, \mathrm{dlog}L_{\mathrm{X}} \, \mathrm{d}z. \tag{11}$$

The parameter λ is the expected number of observed sources for a Poisson process, given an XLF model with parameters $\mathbf{\Theta}$:

$$\lambda = \iiint \phi_{\mathrm{abs}}(L_{\mathrm{X}}, z, N_{\mathrm{H}}|\mathbf{\Theta})\Omega(L_{\mathrm{X}}, z, N_{\mathrm{H}}) \frac{\mathrm{d}V}{\mathrm{d}z} \mathrm{dlog}N_{\mathrm{H}} \, \mathrm{dlog}L_{\mathrm{X}} \, \mathrm{d}z, \tag{12}$$

where Ω is the survey sensitivity function calculated in Sect. 4.1 and $\phi_{\mathrm{abs}} = \phi \times f_{\mathrm{abs}}$. ϕ and f_{abs} are the luminosity and absorption functions defined by the parametrisations presented in Sects. 4.2 and 4.3, respectively.

The parameter P_i in Eq. (11) is given by:

$$P_i(L_{\mathrm{X}}, z, N_{\mathrm{H}}|\mathbf{\Theta}) = p(d_i|L_{\mathrm{X}}, z, N_{\mathrm{H}}) \, \phi_{\mathrm{abs}}(L_{\mathrm{X}}, z, N_{\mathrm{H}}|\mathbf{\Theta}) \, \Omega(L_{\mathrm{X}}, z, N_{\mathrm{H}}) \tag{13}$$

where $p(d_i|L_{\mathrm{X}}, z, N_{\mathrm{H}})$ is the probability of the source i being at redshift z with $N_{\mathrm{H}}, L_{\mathrm{X}}$ X-ray properties. This probability is given by the posterior probability distributions we obtained during our X-ray spectral analysis (Sect. 3). We have included in this term the sensitivity function of the survey Ω. While, according to Loredo (2004), this is formally incorrect, this term takes

into account the loss of information due to the different methods used for X-ray source detection and for the X-ray spectral analysis (see Appendix A of Buchner et al. 2015, for a detailed discussion of this issue). The integral involving P_i in Eq. (11) can be calculated using an importance sampling Monte Carlo integration technique (Kloek & van Dijk 1978; Press et al. 1992).

The integration limits used in Eqs. (11) and (12) are Aihara et al. (2019), Akaike (1974), [42,47] and [20,26] for the parameters z, $\log L_X$ and $\log N_H$, respectively. Overall, our combined parametrisation of the luminosity and the absorption functions have eleven free parameters when we use the LDDE model for XLF, and ten parameters for the PDE model. In Table 2 we report the best-fit parameter estimations with their uncertainties of the X-ray luminosity function and the absorption function derived adopting the LDDE and the PDE models. Figure 7 presents the one-dimensional (diagonal panels) and two-dimensional marginal posterior distributions for the PDE (purple) and LDDE (green) model parameters. As can be seen from the one-dimensional plots, only two parameters cannot be fully constrained: $f_{\mathrm{CTK},r}$ and a_2. Even though these parameters are well-constrained at 1σ, the 2σ contours are not closed.

Figure 8 shows the best-fitting intrinsic absorption function in the redshift range of our analysis. The boxes (points and errors) show the credible regions that correspond to 1σ confidence interval of the posterior probabilities for the PDE (LDDE) model. As can be seen further from Table 2, the absorption function parameters estimations are similar for both XLF models. We compare the intrinsic absorption function with the one derived in Vito et al. (2018). In order to do a proper comparison, we integrated their absorption function over the binning of our analysis (dashed lines). The fraction of heavily obscured sources ($\log N_H \geq 23$) agrees with our results within the uncertainties, even though their average values appear lower. Their derived fraction of sources with $20 \leq \log N_H \leq 23$ is about 50% that is 15% higher than in our case. Perhaps this is due to the fact that in their analysis the absorption function is normalised between $\log N_H = 20-25$ instead. Furthermore, we followed a different approach to compute the obscuration incompleteness that affect strongly the results.

In Fig. 9, we plot the total X-ray luminosity function at the full redshift range of our analysis ($3 \leq z \leq 6$) with the 1σ and 2σ uncertainties for the PDE (top) and LDDE (bottom) models. The shaded (empty) regions correspond to the XLF that includes any absorption level between $\log N_H = 20-26$ ($\log N_H = 20-24$). We overplot the binned luminosity function ($\log N_H = 20-26$). The binned data appear to overestimate the XLF at high luminosities compared to the models. This comes from the large uncertainties at $z > 5$ and high column densities (as can be seen in the next section at Fig. 10). For reference, we also show the results of previous X-ray studies (Ueda et al. 2014; Vito et al. 2014; Georgakakis et al. 2015; Peca et al. 2023) at $\log N_H = 20-24$. In the following section, we discuss further their similarities and differences with our results.

To quantify which of the two XLF models provides better fit to the data in the redshift range $3 \leq z \leq 6$, we used the information criteria: the Akaike Information Criterion (AIC, Akaike 1974) and the Bayesian Information criterion (BIC, Schwarz 1978) that have been used widely in the literature (e.g. Fotopoulou et al. 2016; Pouliasis et al. 2020). They take into account the complexity of the models in addition to the goodness of fit and prefer models with fewer parameters. The best model according to these is the LDDE model. However, taking the difference of the two models, we found DAIC = 4.5

and DBIC = 2.6 that suggest that we may not exclude the possibility of the PDE model to be the correct one. Furthermore, we used a Bayesian model comparison. In particular, for each model we used the Bayes factor (the ratio between the evidences) of each model compared to the one with the highest evidence. It is possible to provide a measure of the weight of the information that is included in the data in a favour of one model against the other. Thus, according the interpretation of the Bayes factor, models that have high values (>1) are rejected. In our analysis, the LDDE model has the highest evidence. The Bayes factor of the PDE model is $\Delta \log \zeta = 0.5$. With the Bayesian model comparison, we find that the two PDE and LDDE models represent the data equally well. In the following sections, we restrict ourselves to using only the PDE model as it has a lower number of parameters. However, we have to keep in mind that the LDDE model may represent equally well the observational data.

6. Discussion

6.1. X-ray luminosity function and comparison to previous constraints

In Fig. 10, we plot our X-ray luminosity function in several redshift bins with the 1σ and 2σ uncertainties (shaded regions) for the PDE model. We overplot the binned luminosity function and, for reference, we also show the results of the X-ray luminosity function derived in previous X-ray studies from the literature. In the upper panels, we compare our results with the parametric forms of the XLF derived by Ueda et al. (2014), Vito et al. (2014), Georgakakis et al. (2015), Peca et al. (2023) in the column density interval of $\log N_H = 20-24$. Vito et al. (2014) and Georgakakis et al. (2015) selected 141 and ~300 AGN, respectively, in the redshift range $3.0 \geq z \geq 5.0$ from various fields, while Ueda et al. (2014) gathered a sample of more than 4000 X-ray sources in the redshift range of $0.01 < z < 5$. More recently, Peca et al. (2023) used a large sample of luminous ($\log L_X \geq 44$) AGN selected in the Stripe 82X field (LaMassa et al. 2013a,b, 2016) that covers an area of 31.3 deg^2. We find a slightly higher number of sources compared to the aforementioned studies, especially at the knee of the luminosity function and towards the bright end. At lower luminosities ($\log L_X \leq 44$), our results are consistent with the XLFs derived by Vito et al. (2014) and Ueda et al. (2014).

In the first two lower panels of Fig. 10, we present our XLF integrated over $\log N_H = 20-26$ along with the works of Buchner et al. (2015) and Vito et al. (2018). Buchner et al. (2015) studied a sample of about 2000 AGN selected from partly the same fields as in our case in the full redshift range $0.01 < z < 7$. Their XLF are derived in the redshift range $3.2 \leq z \leq 4.0$ and $4.0 \leq z \leq 7.0$, respectively. In the first redshift bin ($3.0 \leq z \leq 3.6$), our results agree with the work of Buchner et al. (2015) in the faint and bright end. However, in the break luminosity we find higher number of sources. The Vito et al. (2018) binned luminosity falls below our XLF. This is probably due to the fact that they have selected AGN only in the *Chandra* deep fields, missing this way the population of luminous ($\log L_X \geq 43$) AGN. In the redshift range $3.6 \leq z \leq 6.0$, our XLF is in a better agreement with Vito et al. (2018) within the uncertainties, while the number counts of Buchner et al. (2015) appear to overpredict the AGN density. In their study, they provided confidence regions at each redshift intervals instead of an analytical XLF form. Moreover, we compare our results to the XLF predictions derived by the Ananna et al. (2019) AGN
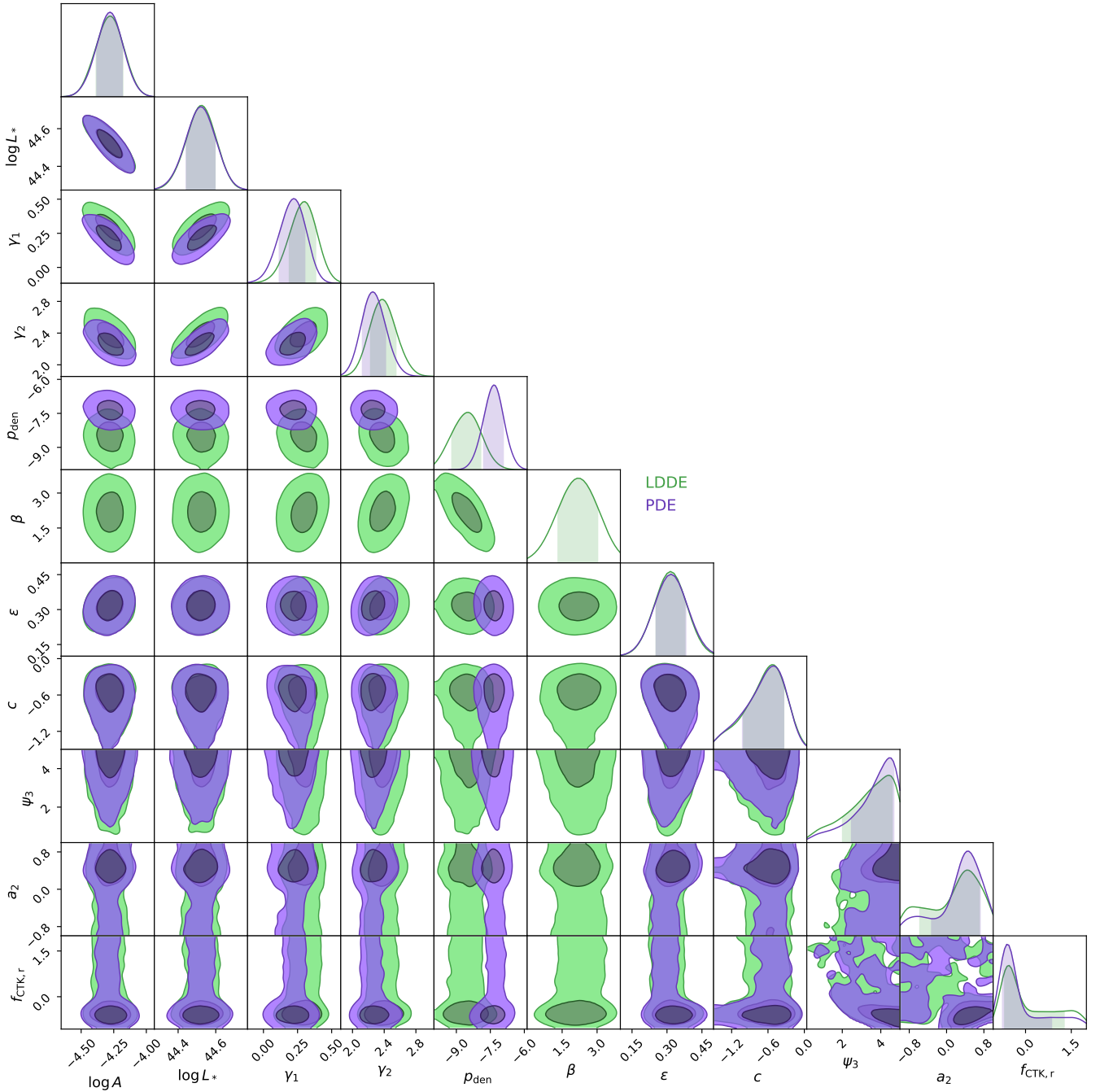
Fig. 7. One-dimensional (diagonal panels) and 2D marginal posterior distributions for the PDE (purple) and LDDE (green) model parameters. The shaded areas in the 2D posterior distributions correspond to 1σ and 2σ confidence levels (2D values, i.e. 39% and 86%, respectively). The shaded areas for the 1D posteriors correspond to 1σ confidence levels.

population synthesis model in both column density bins. When we consider the full AGN population ($\log N_H = 20-26$), our results are in a very good agreement at $\log L_X \geq \sim 44$, while they estimate higher values in the faint end. On the other hand, in the $\log N_H = 20-24$ bin, their predicted XLF follows the one derived by Ueda et al. (2014) that is lower compared to our results. This probably comes from the fact that in their model the Compton-Thick fraction is as high as 50% compared to the $\sim 17\%$ predicted in our analysis (Sect. 6.3).

To assess the agreement with recent observational constraints at the bright end of the function, we compare our results to the works of Wolf et al. (2021) and Barlow-Hall et al.

(2023). Wolf et al. (2021) identified a single source at $z = 5.81$ in the eROSITA Final Equatorial-Depth Survey (eFEDS, Brunner et al. 2022) that covers an area of $140 \deg^2$. Barlow-Hall et al. (2023) found a source at $z = 6.31$ in the Extragalactic Serendipitous *Swift* Survey (ExSeSS, Delaney et al. 2023) catalogue. They provided as well an upper limit at higher luminosities. In Fig. 10 (bottom right), we plot our best-fit PDE model extrapolated at $z = 6.05$ and integrated over $\log N_H = 20-26$. The derived binned luminosity function of the aforementioned studies are consistent with our results within the uncertainties. This result suggests that our derived parameteric X-ray luminosity function may hold at even higher redshifts.
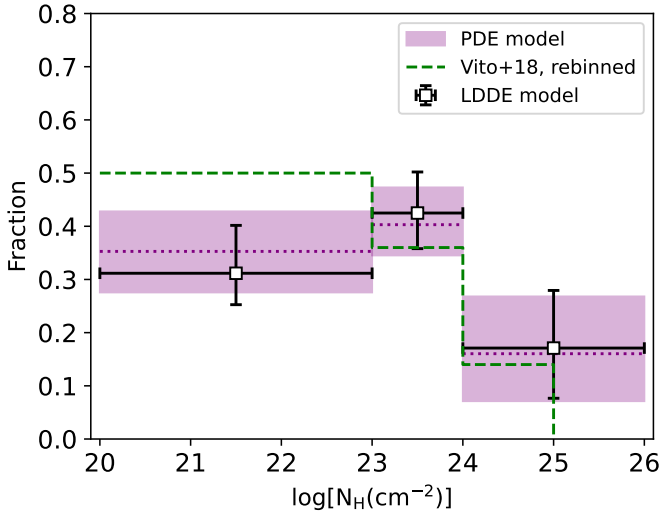
Fig. 8. Best-fitting intrinsic absorption function in the redshift range $3 \leq z \leq 6$ for the PDE model (dotted line). The boxes show the credible regions that correspond to 1σ confidence intervals of the posterior probabilities. The points represent the best-fitting results when using the LDDE model along with the 1σ uncertainties. For reference, shown are the results of Vito et al. (2018) re-scaled to the bins of our analysis (dashed line). The solid line presents their original absorption function; it is normalised between $\log N_{\rm H} = 20-25$.

6.2. Space density

In Fig. 11, we plot the comoving space density of AGN versus redshift in three X-ray luminosity bins as indicated on the top of each panel. The parametric number density was calculated by integrating the XLF over X-ray luminosity and hydrogen column density for the PDE model. The binned space density is represented with the data points. The upper and lower panels correspond to the space density when integrating over $\log N_{\rm H} = 20-24$ and $\log N_{\rm H} = 20-26$, respectively. We compare our results with previous X-ray studies (Marchesi et al. 2016b; Vito et al. 2018; Peca et al. 2023) and also with the predictions of the Ananna et al. (2019) model and that derived by Gilli et al. (2007) that shows a strong decline of the space desnity at high redshifts. Comparing our results with the space density derived by Marchesi et al. (2016b), we find a higher space density by a factor of 4–6. This difference can be ascribed to the fact that they did not take into account the obscuration incompleteness as it has been done in our analysis. We obtain similar results when we compare with the work of Peca et al. (2023). Even though they have corrected for obscuration effects, it is possible that obscured sources are missing due to the shallower Stripe-82X (8.7×10^{-16} erg cm^{-2} s^{-1} in the soft band) and even the correction for the area is insufficient. Furthermore, in the redshift range $3 < z < 4$ the Stripe-82X does not include many sources, while above $z > 4$ the space density is an extrapolation. Regarding the Vito et al. (2018) space density, in the lower luminosity bins, we find a comparable number of sources within the uncertainties, while for the most luminous AGN we predict a higher space density. This is probably due to the sparse sample at these luminosities in the *Chandra* deep fields, since the majority of the sources have $\log L_{\rm X} \leq 43$. Finally, we find higher number density by a factor of 3–4 compared to the predictions of the LDDE model with an exponential decline by Gilli et al. (2007). Regarding the Ananna et al. (2019) model, in the $\log N_{\rm H} = 20-24$ bin we find higher values at bright luminosities, while this difference diminishes going to the fainter end. For the full AGN popula-
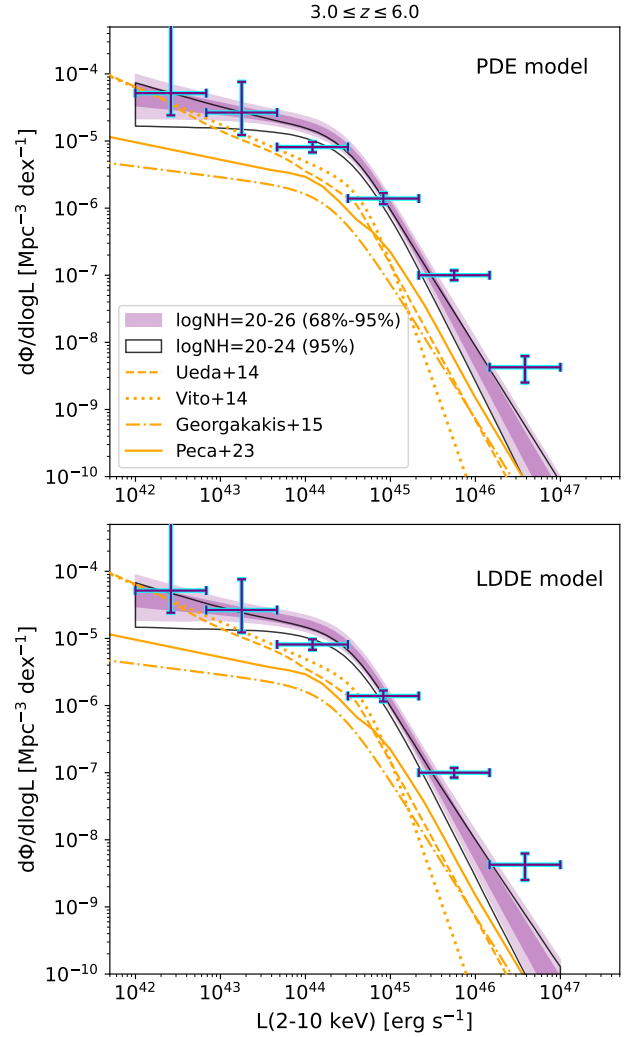
Fig. 9. X-ray luminosity function in the redshift range $3.0 \leq z \leq 6.0$ for the PDE (top) and LDDE model (bottom). The shaded regions represent the 68% and 95% confidence intervals of the XLF when integrating over $20 \leq \log N_{\rm H} \leq 26$, while the black lines show the XLF when integrating over $20 \leq \log N_{\rm H} \leq 24$. The points show the binned luminosity function. For comparison, shown are the XLFs derived by previous X-ray studies in the column density range $\log N_{\rm H} = 20-24$. The best-fitting models of these studies are evaluated at the mean redshift of each bin.

tion ($\log N_{\rm H} = 20-26$), the space density is in agreement within the uncertainties at bright luminosities. In the faintest bin, their model predicts higher values.

Figure 12 (upper panels) compares our results with the predicted number densities from large-scale hydrodynamical cosmological simulations (Habouzit et al. 2021, 2022). Taking into account the full AGN population predicted by simulations ($M_* > 10^9 \, M_\odot$), we find a large discrepancy with our results at lower luminosities. However, in Pouliasis et al. (2022b) we derived the stellar masses of the high-z X-ray sample selected in CCLS, XMM-XXL-N and the eFEDS fields and found that all the high-z X-ray AGN reside in galaxies with stellar masses above $10^{10} \, M_\odot$, with the majority being at $\geq 5 \times 10^{10} \, M_\odot$. For the CDF-S/N sample, many studies also suggest that the majority of the high-z X-ray AGN have host galaxies with $M_\star > 10^{10} \, M_\odot$ (Santini et al. 2015; Yang et al. 2017; Circosta et al. 2019). Thus, in order to compare properly the space density of the X-ray selected AGN to the one derived by the simulations, we have to restrict the
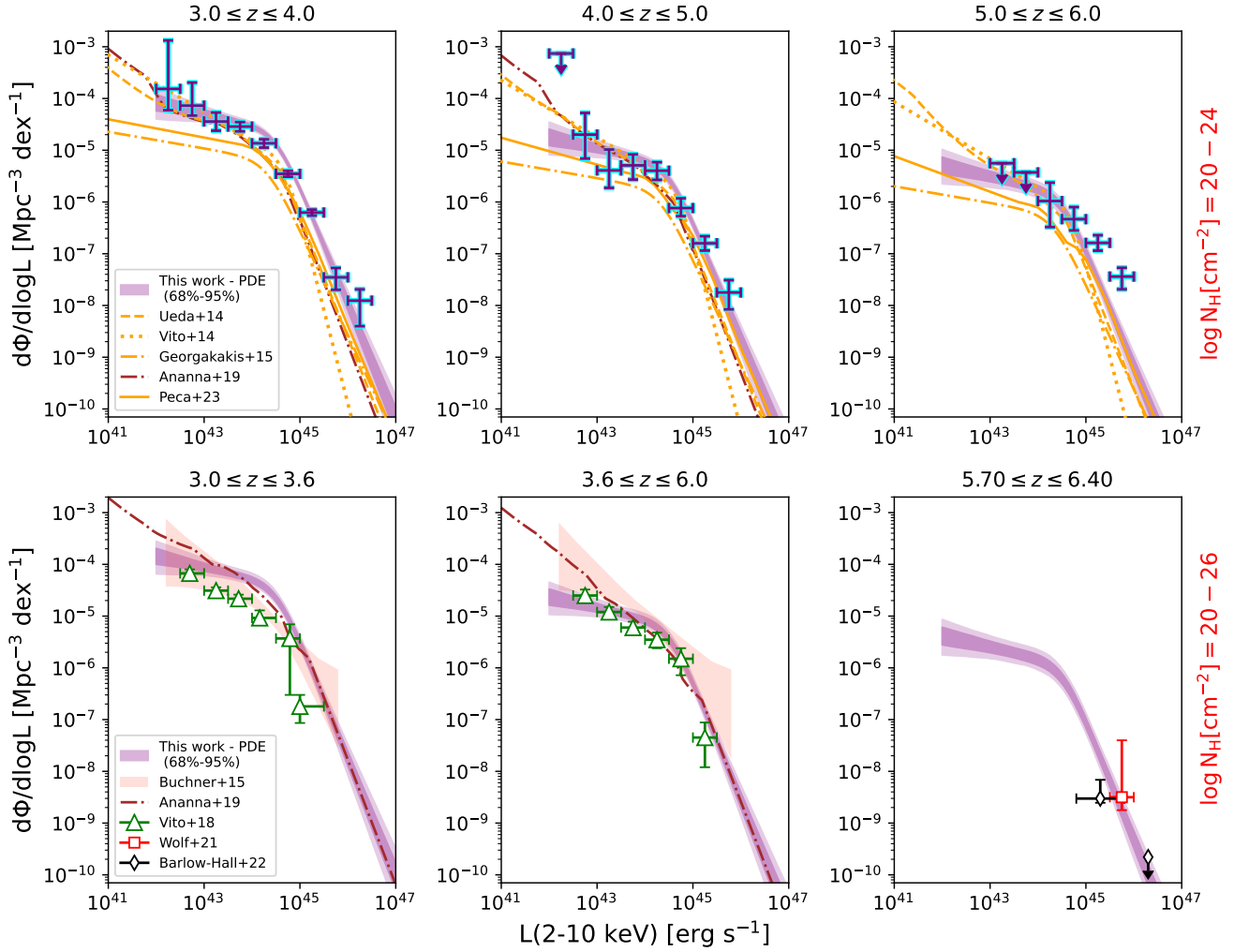
Fig. 10. Best-fitting PDE model in several redshift bins computed by integrating the XLF over redshift and column density. The purple shaded regions represent the 68% and 95% confidence intervals of the model, while the purple data points indicate the binned luminosity function. Our results are compared with the parametric forms of the XLF derived by previous X-ray studies (Ueda et al. 2014; Vito et al. 2014; Georgakakis et al. 2015; Peca et al. 2023) in the column density interval of $\log N_H = 20-24$ (upper panels). Also shown is our XLF integrating over $\log N_H = 20-26$ (lower panels) to the results of Buchner et al. (2015), Vito et al. (2018), Wolf et al. (2021), Barlow-Hall et al. (2023). The XLF of Buchner et al. (2015) in the first two panels are derived in the redshift range $3.2 \leq z \leq 4.0$ and $4.0 \leq z \leq 7.0$, respectively. Our PDE model at the last panel is extrapolated at $z = 6.05$. The brown dash-dotted lines show the predicted XLF derived by Ananna et al. (2019).

sample of Habouzit et al. (2022) to sources with stellar masses of $M_* > 10^{10}\,M_\odot$ (lower panels of Fig. 12).

In such a case, in the lowest luminosity bin ($\log L_X = 42-43$), our results (both shape and normalisation) agree very well with all the predictions coming from all the different simulations. In the luminosity bin $\log L_X = 43-44$, there is a discrepancy by a factor of around 10 compared to all simulations, even though the shapes are consistent. As mentioned above, the majority of the sources in Pouliasis et al. (2022b) have host galaxies with stellar masses greater than $5 \times 10^{10}\,M_\odot$. Hence, a higher stellar-mass cut in the simulations could diminish this difference. For the most luminous AGN ($\log L_X = 44-45$), our space density is consistent qualitatively with all the simulations but the EAGLE. Here, it is worth mentioning that the number densities from the simulations span at least half a dex, which highlights the large uncertainties in the number of AGN produced in the simulations and the complexity of modelling the black-hole and galaxy physics. Furthermore, these simulations do not include any correction for obscuration of any kind. A more thorough analysis taking into account all the selection

biases and using the same physical parameter range between the simulations and the observed X-ray population is necessary. All the above point towards the fact that either the simulations over-predict the AGN in low-mass galaxies below $10^{10}\,M_\odot$ or the current X-ray surveys are not able to detect them. In the first case, this could be due to too massive black-holes in these simulated galaxies, too efficient BH accretion, or too weak supernova feedback (or even AGN feedback). The feedback explanation has been also suggested by Vito et al. (2018). In Sect. 6.4, we investigate further the second point (do X-rays miss the AGN population hosted by low-mass galaxies?) by comparing the BHAD derived in X-rays and in the mid-IR wavelengths through JWST data.

6.3. Intrinsic obscuration fraction

The obscured fraction of the AGN population is important to understand how obscured and unobscured sources evolve through cosmic time and to study the accretion history of the super-massive black holes (Hiroi et al. 2012; Iwasawa et al. 2012).
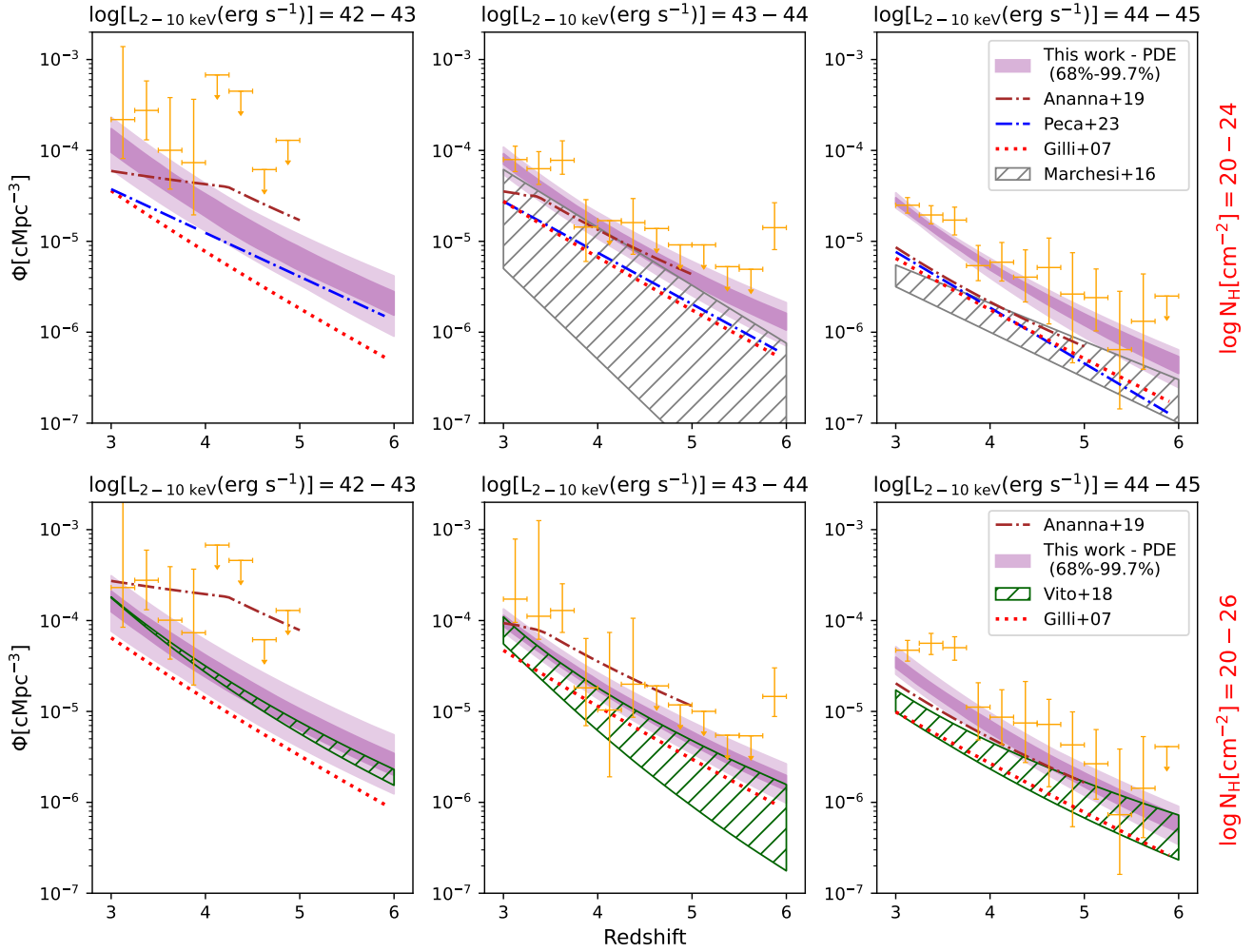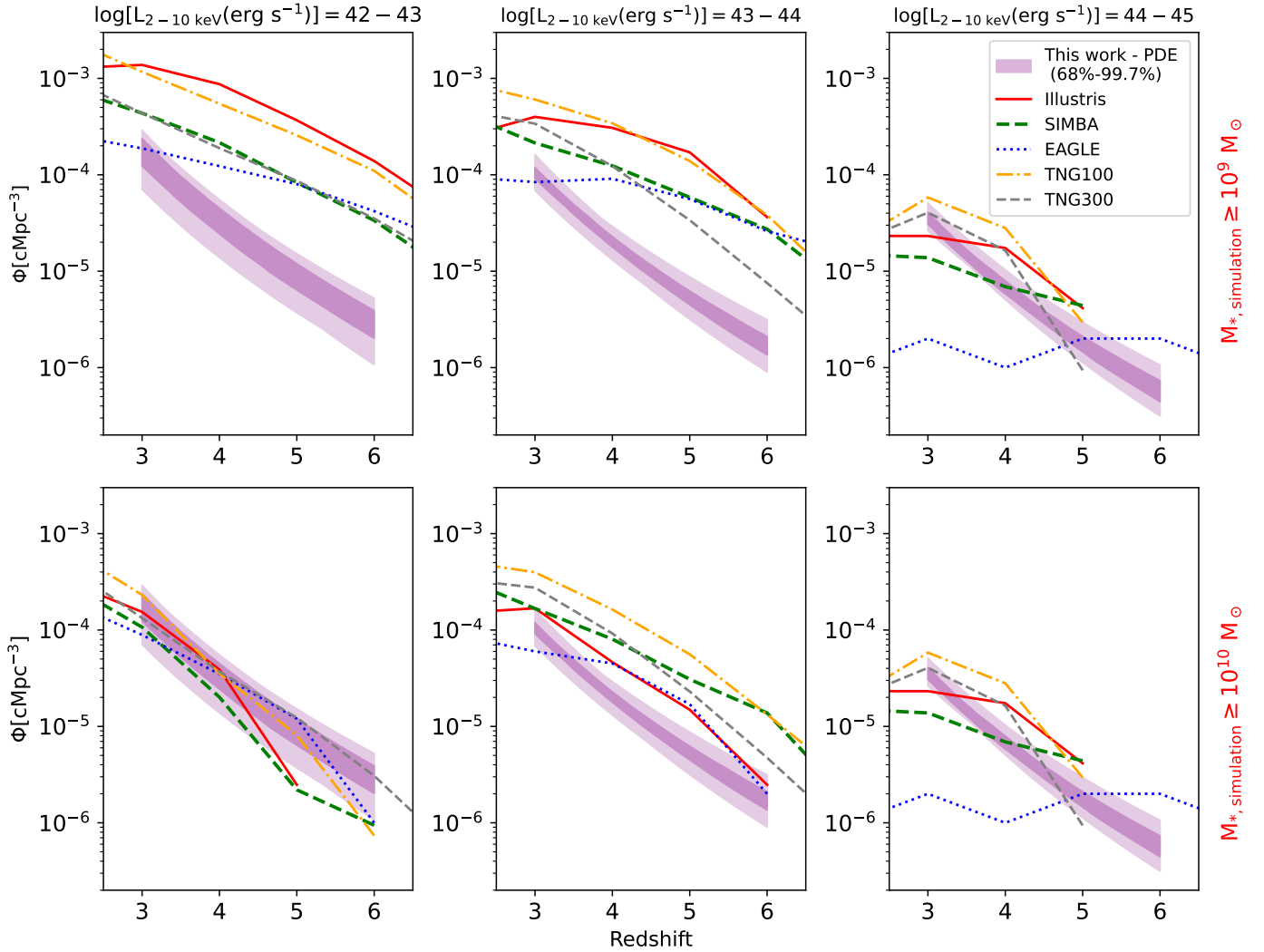
Fig. 11. Redshift evolution of the AGN space density in three luminosity for $\log N_{\mathrm{H}} = 20-24$ (upper panels) and $\log N_{\mathrm{H}} = 20-26$ (lower panels). The shaded regions represent the 68% and 99.7% confidence intervals of the best-fitting PDE model, while the points represent the binned luminosity function. For comparison, the results of Marchesi et al. (2016b), Vito et al. (2018), and Peca et al. (2023) are overplotted. The dotted red lines indicate the LDDE model predictions with an exponential decline (with a power-law decay) by Gilli et al. (2007). The brown dash-dotted lines show the predicted XLF derived by Ananna et al. (2019).

Heavily obscured sources by large column of gas ($\log N_{\mathrm{H}} \geq 24$) play a key role both in the determination of the evolutionary models (Alexander & Hickox 2012; Ricci et al. 2017) and in the population synthesis models to constrain the shape of the cosmic X-ray background (Gilli et al. 2007; Ananna et al. 2019). Due to the suppression of the spectrum at low energies because of obscuration, it is important to account properly for any observational biases to study this class of heavily obscured AGN. Using the intrinsic N_{H} distribution derived from the minimisation method described in Sect. 5, we find that the intrinsic fraction of Compton-Thick sources over the whole population is $F_{\mathrm{CTK}} = 0.17^{-0.09}_{+0.07}$ in the redshift range $3 < z < 6$. Comparing our results to the works of Burlon et al. (2011), Aird et al. (2015), Buchner et al. (2015), Masini et al. (2018), Georgantopoulos & Akylas (2019), Laloux et al. (2023) at lower redshifts, we find a constant Compton-Thick fraction from the local Universe up to redshift $z = 6$. In a following paper (Pouliasis et al., in prep.) we will present exclusively the number density and properties of the Compton-Thick population.

In Sect. 3, we presented the observed distribution of the hydrogen column density of the combined sample. The observed obscured fraction of sources with $\log N_{\mathrm{H}} > 23$ is 28.7% in

the redshift range $3 \leq z \leq 6$ compared to the overall population when using the full N_{H} posterior distributions. Though, the intrinsic fraction of obscured AGN ($\log N_{\mathrm{H}} > 23$) is $61^{+8}_{-7}\%$. As mentioned in the previous section, this fraction of obscured AGN involves host galaxies mainly with $M_* > 10^{10}\,M_\odot$. Among this obscured population, the relative fraction of Compton-Thick AGN with $\log N_{\mathrm{H}}[\mathrm{cm}^{-2}] \geq 24$ is $F_{\mathrm{CTK,relative}} = 0.25^{+0.14}_{-0.23}$. In Fig. 13, we plot the obscured fraction with $\log N_{\mathrm{H}}(\mathrm{cm}^{-2}) > 23$ as a function of the intrinsic X-ray luminosity. We overplot the fraction derived by Vito et al. (2018) in different luminosity bins. Despite the large uncertainties in the high-luminosity end, the results are consistent with each other indicating a flat fraction of ~68% across the hard X-ray luminosity. Similarly, we find a negligible dependence of the obscured fraction over cosmic time in the redshift range $3.0 \leq z \leq 6.0$ for the same luminosity bins.

In Fig. 14, we present our results in addition to the obscured fractions given in previous studies in the literature (Liu et al. 2017; Vito et al. 2018; Vijarnwannaluk et al. 2022; Signorini et al. 2023) at lower redshifts and having the same definition of the obscured fraction ($\log N_{\mathrm{H}}(\mathrm{cm}^{-2}) > 23$). We show only the AGN population with $43.5 \leq \log L_X \leq 44.5$ to compare

Fig. 12. Redshift evolution of the AGN space density in three luminosity bins, as indicated at the top of the panels. The shaded regions represent the 68% and 99.7% confidence intervals of the best-fitting LDDE model. The lines represent the space densities derived from different cosmological simulations (Habouzit et al. 2022). The upper and lower panels correspond to systems with stellar mass higher than $M_* > 10^9 M_\odot$ and $M_* > 10^{10} M_\odot$, respectively.

properly with the aforementioned studies. Liu et al. (2017), combining data from the 7 Ms CDF-S and the CCLS surveys, studied the obscuration properties of their sample and concluded that there is an evolution in the obscuration fraction from $z = 1$ up to $z = 3$. Vito et al. (2018) found that the obscured fraction defined as N_{23-25}/N_{20-25} is almost constant at $f_{abs} = 0.5$ in the redshift range within $z = 3-5$ without any correction for the obscuration incompleteness. Applying the obscuration completeness, they found a value of $f_{abs} = 0.65$ that is similar to our results. For the AGN population with luminosities around $\log L_X \sim 44$, they found an obscured fraction of $\sim 0.66\%$ and $\sim 0.73\%$ for the luminosity bins $43.5 \leq \log L_X \leq 44.0$ and $44.0 \leq \log L_X \leq 44.5$, respectively. Vijarnwannaluk et al. (2022), using the deep multi-wavelength data of the XMM-SERVS catalogue inside the XMM-LSS region (Chen et al. 2018), suggested that 50% of the sources are obscured at the cosmic noon ($z = 2-3$). More recently, Signorini et al. (2023) studied the spectral properties of X-ray selected AGN in the $J1030$ *Chandra* field. They derived also the obscuration fraction of $\log L_X \sim 44$ AGN in the redshift range $0.8 \leq z \leq 2.8$ with an average value of $f_{abs} = 0.5-0.6$.

With the grey shaded regions we show the extrapolation of the absorption function below and above the redshift range of our analysis. For redshifts lower than $z = 3$, as we discussed in Sect. 4.3, we have adopted the formulations of Ueda et al. (2014) and Vijarnwannaluk et al. (2022). Our predicted obscuration fraction is in agreement within the uncertainties to the previous works. In total, there is clear redshift evolution of the obscuration fraction of the AGN population at redshifts below the interval of our analysis.

This evolution can be driven by larger gas reservoirs observed in high-z AGN (Carilli & Walter 2013). Indeed, D'Amato et al. (2020), studying a sample of high-redshift AGN with $\log N_H > 23$, found that the contribution of the interstellar medium (ISM) of the host galaxies to the obscured fraction of the AGN is comparable to that estimated by X-ray or infrared estimations as in Circosta et al. (2019). This may suggest that the host galaxy of the AGN can strongly affect the obscuration fraction, and especially, at high redshifts. The column density of the ISM at $z > 3$ may be even 100 times larger than in the local Universe and could reach values close to the Compton-Thick regime at $z > 6-8$. Such behavior has already been observed
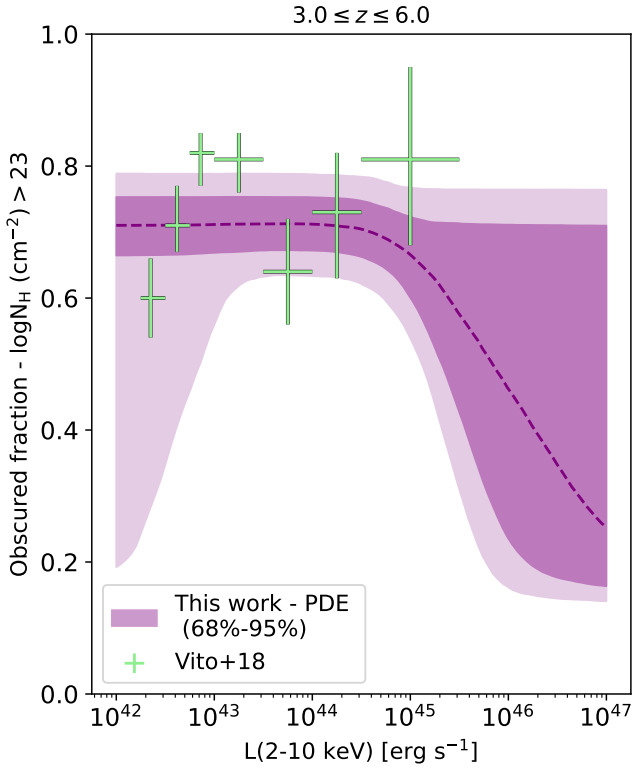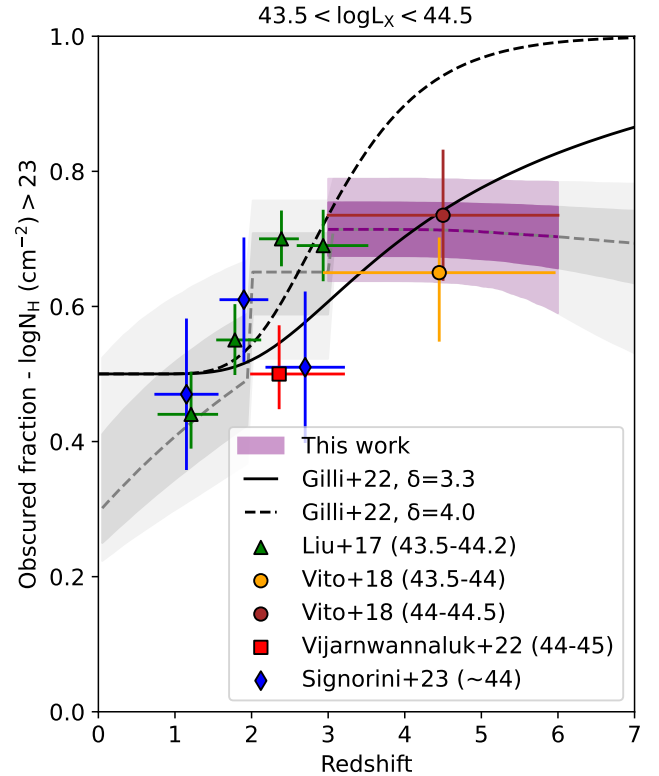
Fig. 13. Obscured fraction with $\log N_H \geq 23$ vs. X-ray luminosity. Our results are shown with shaded regions that represent the 68% and 95% confidence intervals of the best-fitting absorption function. For reference, the results of Vito et al. (2018) are shown.

Fig. 14. Obscured fraction with $\log N_H \geq 23$ vs. redshift. Our results are shown as the purple shaded regions that represent the 68% and 95% confidence intervals of the best-fitting absorption function. Also shown are the extrapolation (grey shaded regions) of the absorption function below and above the redshift range of our analysis. For reference, shown are the results of Liu et al. (2017), Vito et al. (2018), Vijarnwannaluk et al. (2022), and Signorini et al. (2023), which span a wide range of redshifts. The luminosity ranges used in these studies are given in brackets (see legend in inset). The lines are the predictions of the obscuration fraction originated on both ISM and torus components (Gilli et al. 2022) for different values of the obscuration evolutionary factor, δ.

at $z > 2$ where the ISM column density is consistent with the X-ray N_H values (Gilli et al. 2022). Gilli et al. (2022) modelled analytically at the same time both the large-scale (ISM) and small-scale (AGN torus) obscuration assuming clouds of different sizes, masses, and surface densities.

Comparing with different X-ray surveys, they concluded that the median ISM column density evolves with $N_{H,ISM} \propto (1 + z)^\delta$ where $\delta = 3.3$. In addition, they found that the evolution of the characteristic cloud surface density with redshift ($\Sigma_{c,*} \propto (1+z)^\gamma$), can be described with $\gamma = 2$. Signorini et al. (2023) including their data and allowing for different values of the evolutionary factors found that $\gamma = 2$ and $\delta = 4$ best reproduce the data. In Fig. 14, we show as well the estimations of the obscuration fraction predicted by the best Gilli et al. (2022) model with $\gamma = 2$ and torus opening angle of 60 deg. The latter was adopted since we are focused here on luminous AGN ($\log L_X \sim 44$). The lines represent different values of the evolutionary factor δ. Our results are consistent with an evolution of $\delta = 3.3$ similarly to Gilli et al. (2022). Hence, the large sample of high-z AGN used in our analysis is compatible with a scenario where the AGN obscuration fraction is dependent on the evolving ISM obscuration across cosmic time. However, we notice that the extrapolation of our obscured fraction towards higher redshifts ($z \sim 6$) diverges from the ISM model by Gilli et al. (2022). Larger samples of AGN, though, are needed at these redshifts (e.g. Sect. 7) to shed light on this context.

6.4. Black-hole accretion rate density

The black hole accretion rate density (BHAD) is fundamental to characterise effectively the growth of the AGN population.

In this section, we derive the BHAD over cosmic time from our updated XLF and compare it to the predictions of theoretical models and the BHAD derived from several XLFs in the literature. Firstly, we converted the XLF into the bolometric luminosity function (BLF) using the luminosity-dependent bolometric correction by Duras et al. (2020). In particular, we used the following equation:

$$\frac{d\Phi}{d\log L_{bol}} = \frac{d\Phi}{d\log L_X} \times \frac{d\log L_X}{d\log L_{bol}}, \tag{14}$$

where $\frac{d\Phi}{d\log L_{bol}}$ and $\frac{d\Phi}{d\log L_X}$ are the bolometric and X-ray luminosity functions, respectively, and $\frac{d\log L_X}{d\log L_{bol}}$ is the derivative to convert the XLF to BLF. Then, we may compute the BHAD from the BLF:

$$BHAD = \frac{1-\epsilon}{\epsilon c^2} \times \int_{43}^{49} L_{bol} \frac{d\Phi}{d\log L_{bol}} d\log L_{bol}, \tag{15}$$

where c is the speed of light and ϵ is the radiative efficiency to convert the energy into mass. In this work, we adopt $\epsilon = 0.1$ similarly to previous works in the literature (Hopkins et al. 2007). The integral was calculated in the bolometric luminosity range $43 < \log L_{bol} < 49$, corresponding to X-ray luminosities of
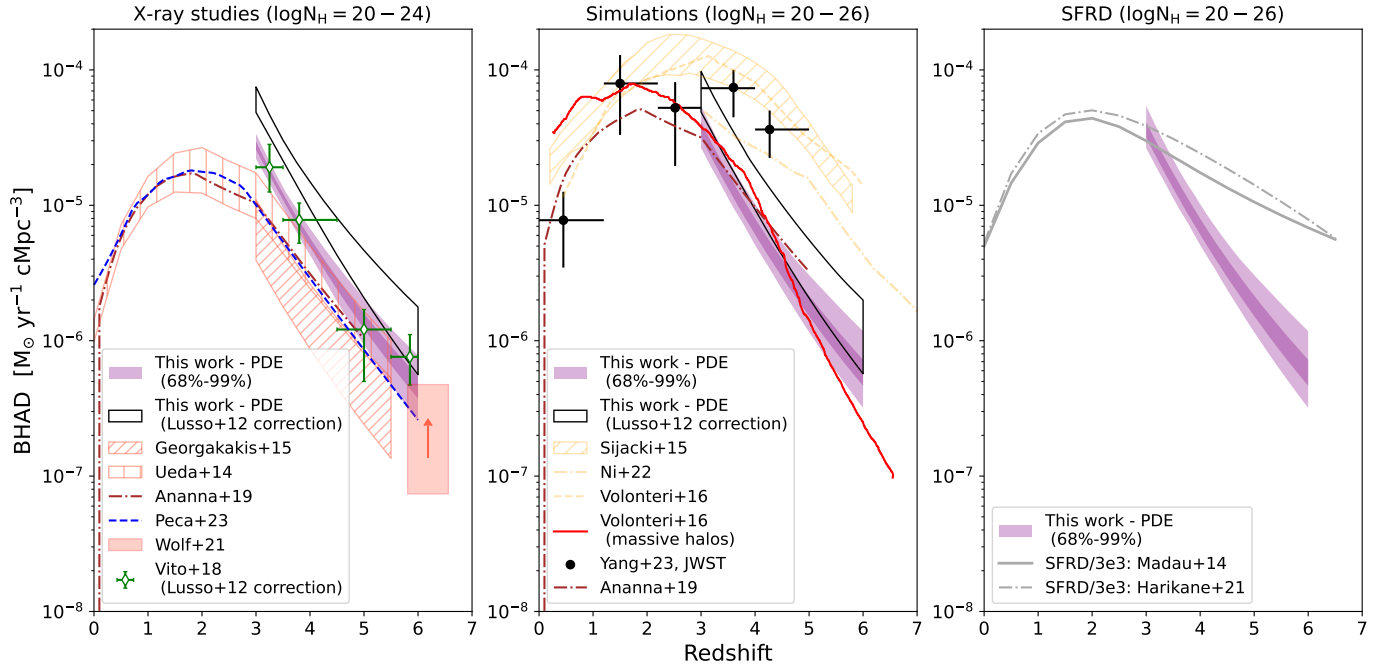
Fig. 15. Redshift evolution of the black hole accretion rate density. The shaded regions represent the 68% and 99% confidence intervals of the BHAD using the best-fitting PDE model. For comparison, shown are (from left to right) our results with previous X-ray studies (Georgakakis et al. 2015; Ueda et al. 2014; Vito et al. 2018; Wolf et al. 2021; Peca et al. 2023), with simulations (Volonteri et al. 2016; Ni et al. 2022; Sijacki et al. 2015), and with the star formation rate density scaled down by a factor of 3000 (Madau & Dickinson 2014; Harikane et al. 2022). The black regions in the first two panels indicate the 99% confidence interval of the BHAD adopting the bolometric correction by Lusso et al. (2012) for direct comparison to the results of Vito et al. (2018). The black points in the middle panel correspond to the JWST results by Yang et al. (2023). The brown dash-dotted line shows the predicted XLF derived by Ananna et al. (2019).

$42 \lesssim \log L_X \lesssim 47$ using the bolometric correction (L_{bol}/L_X) from Duras et al. (2020).

In Fig. 15, we show the BHAD based on our derived XLF as a function of redshift. The uncertainties were calculated using the output of the Bayesian analysis of the XLF. In addition, we have included the uncertainties of the Duras et al. (2020) correction parameters. We compare our results (left panel of Fig. 15) for $\log N_H = 20-24$ with the BHAD computed from the XLF derived in previous X-ray studies (Georgakakis et al. 2015; Ueda et al. 2014; Vito et al. 2018; Wolf et al. 2021; Peca et al. 2023) using the method described above. Since Vito et al. (2018) used the Lusso et al. (2012) bolometric correction, we show as well our BHAD adopting the latter correction (black region). Our BHAD is higher by a factor of about 2–4 compared to all the previous X-ray studies, while Wolf et al. (2021) derived an upper limit of the BHAD that agrees with ours at high redshifts. Furthermore, we compare our BHAD with the one predicted by Ananna et al. (2019). As mentioned in the previous sections, in the $\log N_H = 20-26$ bin, our results agrees within the uncertainties with their model. However, in the $\log N_H = 20-26$ we find a much higher value. This difference can be ascribed to the fact that in their analysis the predicted Compton-Thick fraction is ~50% that is much higher compared to our value $(F_{CTK} = 0.17^{-0.09}_{+0.07})$.

In the middle panel of Fig. 15, we compare our derived BHAD with the ones predicted by various large-scale cosmological simulations (Volonteri et al. 2016; Ni et al. 2022; Sijacki et al. 2015). When we consider the full AGN population predicted in simulations, we find approximately one order of magnitude difference. However, as discussed in Sect. 6.2, the X-ray sample consists of systems with stellar masses of $\geq 10^{10} M_\odot$. Restricting the simulated AGN only to massive sys-

tems, this difference should be decreased. Indeed, using the predictions of Volonteri et al. (2016) for massive dark matter halos $(\geq 5 \times 10^{11} M_\odot)$ that roughly correspond to a cut in the stellar mass of the host galaxies at $3 \times 10^{10} M_\odot$ (Dubois et al. 2015), the BHAD coming from X-rays is in a very good agreement with the simulations. So, in order to compare properly the BHAD derived from X-ray studies and the simulations, one should be very conscious of the selection biases (e.g. stellar mass, obscuration).

One explanation for this difference in BHAD is that simulations may produce too much accretion in small systems. On the other hand, Yang et al. (2023), using mid-infrared observations from the Mid-Infrared Instrument (MIRI) on board JWST, found that the BHAD is about 0.5 dex higher than that of previous X-ray studies, while it is more consistent with the simulations. If we compare it with our results at $z \geq 3$, we find a difference of about 0.8 dex and 0.5 dex when using the Duras et al. (2020) and Lusso et al. (2012) bolometric correction, respectively. They ascribe this difference to the fact that X-rays may miss a large population of heavily obscured AGN and hence the simulations overpredict the number counts. If this is the case, it means that even our correction of the obscuration incompleteness is not sufficient enough.

However, in Maiolino et al. (2023) and Lyu et al. (2023) the vast majority of the JWST IR selected AGN in the redshift range $3.0 \leq z \leq 6.0$ are hosted by low-mass galaxies $(<10^{10} M_\odot)$. The latter study also suggests that the low-mass galaxy population is comparable to that of the high-mass galaxies at least in the cosmic noon. We argue here that the X-ray population may miss a small fraction of heavily obscured AGN compared to the SED selected AGN, though the large discrepancies in BHAD arise from the different AGN populations selected in each selection method, and especially, different host-galaxy properties.

A straightforward test would be to directly compare the host-galaxy properties between our sample and the JWST high-z AGN (Yang et al. 2023). Since the stellar mass estimates of the JWST sources are not available, we examine this issue by using the relation between the stellar mass of the host and the black-hole accretion rate (BHAR). Yang et al. (2017, 2018) have shown that BHAR strongly depends on the stellar mass of the host galaxy. Hence, we compare the BHAR of the high-z JWST sources in Yang et al. (2023) with the SED fitting results of the X-ray detected sources in Pouliasis et al. (2022b). For the latter, we used similar approach to the former. We find that the X-ray detected sources have log BHAR $= -1, +3$, while the mid-IR AGN ranges between log BHAR $= -3, -0.5$. This supports further our argument that X-ray studies are focused on identifying AGN hosted by larger-mass galaxies, while the mid-IR observations unveil in a complementary manner the low-mass regime. Furthermore, by comparing the bolometric luminosities, we find significant differences between the selection methods. In particular, using the Duras et al. (2020) bolometric correction, we derived a median value of log $L_{bol} = 45.9 \pm 0.75$. Yang et al. (2023), using the disk luminosity as a proxy for the angle-averaged bolometric luminosity, found log $L_{bol} = \sim 44.5$ in the redshift range $z = 3 - 8$ (their Fig. 6). In addition, Lyu et al. (2023) derived a similar value of log $L_{bol} = \sim 44.6$ in the redshift range $z = 3 - 4$.

Finally, we compare the BHAD evolution with the star formation rate density (SFRD) evolution. Above redshift two there is a decline at both BHAD and SFRD. Many simulations (Habouzit et al. 2021; Zhang et al. 2023) have shown that at higher redshifts ($z > 3$), the ratio BHAD/SFRD drops, indicating that the AGN growth evolves more rapidly than the galaxies. On the right panel of Fig. 15, we plot our derived BHAD with the SFRD predicted by Madau & Dickinson (2014) and Harikane et al. (2022) scaled down by a factor of 3000 to match approximately our BHAD at $z = 3$. The shape of the SFRD from the aforementioned studies is qualitatively consistent with those of the simulations. We found that the BHAD/SFRD ratio drops (as in simulations) from 4×10^{-4} at $z = 3$ to 7×10^{-7} at $z = 6$. These results are in contrast to the findings of Yang et al. (2023), where they found that the BHAD/SFRD ratio increases at $z \geq 3$.

7. Summary and conclusions

We have built the largest sample (>600) of X-ray-selected AGN at high redshifts ($z > 3$). The sample is compiled using fields with different areas and sensitivity depths (CDF-S, CDF-N, CCLS, XMM-XXL-N). About one-third of our sources have spectroscopic redshifts available. In this paper we place tight constraints on the X-ray luminosity and the absorption functions. The advantage of this work is that we derived the X-ray spectral properties (luminosities, absorbing column densities) using a Bayesian technique in a consistent way for all sources taking into account the photometric redshift uncertainties. Our main results can be summarised as follows:

– The XLF is described by a double power law, which evolves according to a pure density evolution model similar to what is witnessed in optical wavelengths. However, a luminosity-dependent density evolution model cannot be securely ruled out.

– The vast majority of our sources are heavily obscured with column densities log $N_H > 23$. Our results confirm previous findings from Vito et al. (2018) in the *Chandra* deep fields. We find no luminisity dependence on the obscured fration, while by combining our results with those at lower redshifts,

we find an evolution of the obscuration with redshift. This could be explained by ISM obscuration in addition to that of the torus (e.g. Gilli et al. 2022).

– The BHAD derived through our analysis is higher compared to previous X-ray studies, while it is roughly in agreement with the simulations that use a cut in the stellar mass of the host galaxies. After comparison with the BHAD derived using JWST data, we conclude that IR-selected AGN involve sources with lower bolometric luminosities hosted by galaxies with $\leq 10^{10}\ M_\odot$, while X-rays probe AGN hosted by larger systems.

To date the eROSITA all-sky survey has detected almost a million X-ray sources (Merloni et al. 2024). At the same time, *XMM-Newton* has detected about 650 thousand serendipitous sources covering $\sim 1300\ deg^2$ (Webb et al. 2023), while the latest *Chandra* Source Catalog Version 2.1 (CSC 2.1) will include about 400 thousand unique X-ray sources detected by the *Chandra* X-ray observatory (Evans et al. 2010, 2020). All of the above surveys are a treasure trove for studies of high-redshift AGN. These surveys are highly complementary. Owing to the large area covered, eROSITA is expected to detect the most luminous AGN occupying the bright end of the luminosity function. On the other hand, *XMM-Newton* and *Chandra* will explore luminosities near and below the break of the luminosity function, respectively. Because of the combination of their energy band-passes and the flux depth probed, eROSITA will detect primarily unobscured AGN, while *XMM-Newton* and *Chandra* will probe more efficiently the obscured AGN population at high redshift.

Because of the faintness of the optical counterparts of the high-redshift AGN, it is clear that excellent quality optical data are necessary in order to fully exploit the high redshift content of these surveys. To have a deeper knowledge on the physical properties governing the growth and evolution of supermassive black holes over cosmic time, large AGN samples are needed to map those populations missed by the current X-ray missions (low-luminosity obscured sources at medium-high redshifts). In the future, ESA's Advanced Telescope for High-ENergy Astrophysics (ATHENA, Nandra et al. 2013) X-ray mission will revolutionise the studies of high-redshift AGN, owing to its unprecedented sensitivity. In addition, the next-generation X-ray telescopes, such as the Advanced X-ray Imaging Satellite (AXIS, Mushotzky et al. 2019; Marchesi et al. 2020), the Survey and Time-domain Astrophysical Research eXplorer (STAR-X[3]), a Medium Explorer mission selected by NASA, and the High Energy X-ray Probe (HEX-P, Madsen et al. 2018, 2023) will offer new opportunities to detect and study obscured sources at high redshifts.

---

[3] http://star-x.xraydeep.org/

References

Abbott, T. M. C., Adamów, M., Aguena, M., et al. 2021, ApJS, 255, 20
Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4
Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 71, 114
Aird, J., Coil, A. L., Georgakakis, A., Nandra, K., & Pérez-González, P. G. 2015, MNRAS, 451, 1892
Akaike, H. 1974, IEEE Trans. Automat. Control, 19, 716
Akiyama, M., Ueda, Y., Watson, M. G., et al. 2015, PASJ, 67, 82
Akiyama, M., He, W., Ikeda, H., et al. 2018, PASJ, 70, S34
Akylas, A., & Georgantopoulos, I. 2021, A&A, 655, A60
Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12
Alexander, D. M., & Hickox, R. C. 2012, New Astron. Rev., 56, 93
Ananna, T. T., Salvato, M., LaMassa, S., et al. 2017, ApJ, 850, 66
Ananna, T. T., Treister, E., Urry, C. M., et al. 2019, ApJ, 871, 240
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, AJ, 156, 123
Avni, Y., & Bahcall, J. N. 1980, ApJ, 235, 694
Bañados, E., Venemans, B. P., Mazzucchelli, C., et al. 2018, Nature, 553, 473
Barger, A. J., Cowie, L. L., Mushotzky, R. F., et al. 2005, AJ, 129, 578
Barlow-Hall, C. L., Delaney, J., Aird, J., Evans, P. A., & Watson, M. G. 2023, MNRAS, 519, 6055
Bogdan, A., Goulding, A., Natarajan, P., et al. 2023, ArXiv e-prints [arXiv:2305.15458]
Boutsia, K., Grazian, A., Fontanot, F., et al. 2021, ApJ, 912, 111
Brandt, W. N., & Alexander, D. M. 2015, A&ARv, 23, 1
Broos, P. S., Townsley, L. K., Feigelson, E. D., et al. 2010, ApJ, 714, 1582
Brunner, H., Liu, T., Lamer, G., et al. 2022, A&A, 661, A1
Buchner, J. 2016, Stat. Comput., 26, 383
Buchner, J. 2021, J. Open Source Softw., 6, 3001
Buchner, J. 2022, RNAAS, 6, 89
Buchner, J., & Bauer, F. E. 2017, MNRAS, 465, 4348
Buchner, J., & Boorman, P. 2023, ArXiv e-prints [arXiv:2309.05705]
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
Buchner, J., Brightman, M., Nandra, K., Nikutta, R., & Bauer, F. E. 2019, A&A, 629, A16
Burke, D., Laurino, O., Wmclaugh, et al. 2021, https://doi.org/10.5281/zenodo.5554957
Burlon, D., Ajello, M., Greiner, J., et al. 2011, ApJ, 728, 58
Carilli, C. L., & Walter, F. 2013, ARA&A, 51, 105
Castellano, M., Fontana, A., Treu, T., et al. 2023, ApJ, 948, L14
Chen, C. T. J., Brandt, W. N., Luo, B., et al. 2018, MNRAS, 478, 2132
Circosta, C., Vignali, C., Gilli, R., et al. 2019, A&A, 623, A172
Civano, F., Marchesi, S., Comastri, A., et al. 2016, ApJ, 819, 62
Coil, A. L., Blanton, M. R., Burles, S. M., et al. 2011, ApJ, 741, 8
Cool, R. J., Moustakas, J., Blanton, M. R., et al. 2013, ApJ, 767, 118
D'Amato, Q., Gilli, R., Vignali, C., et al. 2020, A&A, 636, A37
Dauser, T., Falkner, S., Lorenz, M., et al. 2019, A&A, 630, A66
De Rosa, G., Venemans, B. P., Decarli, R., et al. 2014, ApJ, 790, 145
Delaney, J. N., Aird, J., Evans, P. A., et al. 2023, MNRAS, 521, 1620
Dubois, Y., Volonteri, M., Silk, J., et al. 2015, MNRAS, 452, 1502
Duras, F., Bongiorno, A., Ricci, F., et al. 2020, A&A, 636, A73

Edge, A., Sutherland, W., Kuijken, K., et al. 2013, Messenger, 154, 32
Euclid Collaboration (Scaramella, R., et al.) 2022, A&A, 662, A112
Evans, I. N., Primini, F. A., Glotfelty, K. J., et al. 2010, ApJS, 189, 37
Evans, I. N., Primini, F. A., Miller, J. B., et al. 2020, Am. Astron. Soc. Meet. Abstr., 235, 154.05
Fan, X., Bañados, E., & Simcoe, R. A. 2023, ARA&A, 61, 373
Fotopoulou, S., Buchner, J., Georgantopoulos, I., et al. 2016, A&A, 587, A142
Freeman, P., Doe, S., & Siemiginowska, A. 2001, in Astronomical Data Analysis, eds. J. L. Starck, & F. D. Murtagh, SPIE Conf. Ser., 4477, 76
Fruscione, A., McDowell, J. C., Allen, G. E., et al. 2006, in CIAO: Chandra's Data Analysis System, eds. D. R. Silva, & R. E. Doxsey, SPIE Conf. Ser., 6270, 62701V
Garilli, B., Guzzo, L., Scodeggio, M., et al. 2014, A&A, 562, A23
Garilli, B., McLure, R., Pentericci, L., et al. 2021, A&A, 647, A150
Gelman, A., Carlin, J. B., Stern, H. S., et al. 2014, Bayesian Data Analysis (Chapman & Hall)
Georgakakis, A., Aird, J., Buchner, J., et al. 2015, MNRAS, 453, 1946
Georgakakis, A., Aird, J., Schulze, A., et al. 2017, MNRAS, 471, 1976
Georgakakis, A., Ruiz, A., & LaMassa, S. M. 2020, MNRAS, 499, 710
Georgantopoulos, I., & Akylas, A. 2019, A&A, 621, A28
Gilli, R., Comastri, A., & Hasinger, G. 2007, A&A, 463, 79
Gilli, R., Norman, C., Calura, F., et al. 2022, A&A, 666, A17
Haardt, F., & Maraschi, L. 1991, ApJ, 380, L51
Habouzit, M., Li, Y., Somerville, R. S., et al. 2021, MNRAS, 503, 1940
Habouzit, M., Somerville, R. S., Li, Y., et al. 2022, MNRAS, 509, 3015
Harikane, Y., Ono, Y., Ouchi, M., et al. 2022, ApJS, 259, 20
Hasinger, G., Miyaji, T., & Schmidt, M. 2005, A&A, 441, 417
Hickox, R. C., & Alexander, D. M. 2018, ARA&A, 56, 625
Hiroi, K., Ueda, Y., Akiyama, M., & Watson, M. G. 2012, ApJ, 758, 49
Hoaglin, D. C., Mosteller, F., & Tukey, J. W. 1983, Understanding Robust and Exploratory Data Analysis (Wiley)
Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, ApJ, 654, 731
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Ilbert, O., Capak, P., Salvato, M., et al. 2008, ApJ, 690, 1236
Inami, H., Bacon, R., Brinchmann, J., et al. 2017, A&A, 608, A2
Iwasawa, K., Gilli, R., Vignali, C., et al. 2012, A&A, 546, A84
Jansen, F., Lumb, D., Altieri, B., et al. 2001, A&A, 365, L1
Jarvis, M. J., Bonfield, D. G., Bruce, V. A., et al. 2013, MNRAS, 428, 1281
Jiang, L., McGreer, I. D., Fan, X., et al. 2016, ApJ, 833, 222
Kaastra, J. S. 2017, A&A, 605, A51
Kalberla, P. M. W., Burton, W. B., Hartmann, D., et al. 2005, A&A, 440, 775
Kamraj, N., Brightman, M., Harrison, F. A., et al. 2022, ApJ, 927, 42
Kloek, T., & van Dijk, H. K. 1978, Econometrica, 46, 1
Kocevski, D. D., Onoue, M., Inayoshi, K., et al. 2023, ApJ, submitted [arXiv:2302.00012]
Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330
Kormendy, J., & Ho, L. C. 2013, ARA&A, 51, 511
Kullback, S., & Leibler, R. A. 1951, Ann. Math. Stat., 22, 79
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
Laloux, B., Georgakakis, A., Andonie, C., et al. 2023, MNRAS, 518, 2546
LaMassa, S. M., Urry, C. M., Glikman, E., et al. 2013a, MNRAS, 432, 1351
LaMassa, S. M., Urry, C. M., Cappelluti, N., et al. 2013b, MNRAS, 436, 3581
LaMassa, S. M., Urry, C. M., Cappelluti, N., et al. 2016, ApJ, 817, 172
Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599
Le Fèvre, O., Paltani, S., Arnouts, S., et al. 2005, Nature, 437, 519
Le Fèvre, O., Cassata, P., Cucciati, O., et al. 2013, A&A, 559, A14
Liske, J., Baldry, I. K., Driver, S. P., et al. 2015, MNRAS, 452, 2087
Liu, T., Tozzi, P., Wang, J.-X., et al. 2017, ApJS, 232, 8
Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M., et al. 2003, PASP, 115, 897
Loredo, T. J. 2004, in Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, eds. R. Fischer, R. Preuss, & U. V. Toussaint, AIP Conf. Ser., 735, 195
Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2017, ApJS, 228, 2
Lusso, E., Comastri, A., Simmons, B. D., et al. 2012, MNRAS, 425, 623
Lusso, E., Valiante, R., & Vito, F. 2023, The Dawn of Black Holes (Springer Nature Singapore), 1
Lyu, J., Alberts, S., Rieke, G. H., et al. 2023, ApJ, accepted [arXiv:2310.12330]
Maccacaro, T., Gioia, I. M., Avni, Y., et al. 1983, ApJ, 266, L73
Maccacaro, T., Gioia, I. M., & Stocke, J. T. 1984, ApJ, 283, 486
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Madsen, K. K., Harrison, F., Broadway, D., et al. 2018, in Space Telescopes and Instrumentation 2018: Ultraviolet to Gamma Ray, eds. J. W. A. den Herder, S. Nikzad, & K. Nakazawa, SPIE Conf. Ser., 10699, 106996M
Madsen, K. K., García, J. A., Stern, D., et al. 2023, ArXiv e-prints [arXiv:2312.04678]

Maiolino, R., Scholtz, J., Curtis-Lake, E., et al. 2023, A&A, submitted [arXiv:2308.01230]
Marchesi, S., Civano, F., Elvis, M., et al. 2016a, ApJ, 817, 34
Marchesi, S., Civano, F., Salvato, M., et al. 2016b, ApJ, 827, 150
Marchesi, S., Lanzuisi, G., Civano, F., et al. 2016c, ApJ, 830, 100
Marchesi, S., Gilli, R., Lanzuisi, G., et al. 2020, A&A, 642, A184
Masini, A., Civano, F., Comastri, A., et al. 2018, ApJS, 235, 17
Matsuoka, Y., Iwasawa, K., Onoue, M., et al. 2022, ApJS, 259, 18
Mauduit, J. C., Lacy, M., Farrah, D., et al. 2012, PASP, 124, 714
McGreer, I. D., Jiang, L., Fan, X., et al. 2013, ApJ, 768, 105
McLure, R. J., Pentericci, L., Cimatti, A., et al. 2018, MNRAS, 479, 25
McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2013, Messenger, 154, 35
Medvedev, P., Sazonov, S., Gilfanov, M., et al. 2020, MNRAS, 497, 1842
Meng, X.-L. 1994, Ann. Stat., 22, 1142
Menzel, M.-L., Merloni, A., Georgakakis, A., et al. 2016, MNRAS, 457, 110
Merloni, A., Lamer, G., Liu, T., et al. 2024, A&A, 682, A34
Miyaji, T., Hasinger, G., & Schmidt, M. 2000, A&A, 353, 25
Miyaji, T., Hasinger, G., Salvato, M., et al. 2015, ApJ, 804, 104
Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, PASJ, 70, S1
Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27
Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, Nature, 474, 616
Mountrichas, G., Georgantopoulos, I., Ruiz, A., & Kampylis, G. 2020, MNRAS, 491, 1727
Mushotzky, R., Aird, J., Barger, A. J., et al. 2019, BAAS, 51, 107
Nandra, K., & Pounds, K. A. 1994, MNRAS, 268, 405
Nandra, K., Barret, D., Barcons, X., et al. 2013, ArXiv e-prints [arXiv:1306.2307]
Ni, Y., Di Matteo, T., Gilli, R., et al. 2020, MNRAS, 495, 2135
Ni, Y., Di Matteo, T., Bird, S., et al. 2022, MNRAS, 513, 670
Niida, M., Nagao, T., Ikeda, H., et al. 2020, ApJ, 904, 89
Page, M. J., & Carrera, F. J. 2000, MNRAS, 311, 433
Paltani, S., Le Fèvre, O., Ilbert, O., et al. 2007, A&A, 463, 873
Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, A&A, 613, A51
Peca, A., Cappelluti, N., Urry, C. M., et al. 2023, ApJ, 943, 162
Pentericci, L., McLure, R. J., Garilli, B., et al. 2018, A&A, 616, A174
Pierre, M., Pacaud, F., Adami, C., et al. 2016, A&A, 592, A1
Pouliasis, E., Mountrichas, G., Georgantopoulos, I., et al. 2020, MNRAS, 495, 1853
Pouliasis, E., Georgantopoulos, I., Ruiz, A., et al. 2022a, A&A, 658, A175
Pouliasis, E., Mountrichas, G., Georgantopoulos, I., et al. 2022b, A&A, 667, A56
Predehl, P., Andritschke, R., Arefiev, V., et al. 2021, A&A, 647, A1
Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, Numerical Recipes in FORTRAN. The Art of Scientific Computing (Cambridge University Press)
Ranalli, P., Georgantopoulos, I., Corral, A., et al. 2015, A&A, 577, A121
Ricci, C., Trakhtenbrot, B., Koss, M. J., et al. 2017, ApJS, 233, 17

Ruiz, A. 2018, https://doi.org/10.5281/zenodo.1482888
Ruiz, A., Corral, A., Mountrichas, G., & Georgantopoulos, I. 2018, A&A, 618, A52
Ruiz, A., Georgantopoulos, I., & Corral, A. 2021, A&A, 645, A74
Saha, T., Markowitz, A. G., & Buchner, J. 2022, MNRAS, 509, 5485
Salvato, M., Hasinger, G., Ilbert, O., et al. 2009, ApJ, 690, 1250
Salvato, M., Ilbert, O., Hasinger, G., et al. 2011, ApJ, 742, 61
Salvato, M., Wolf, J., Dwelly, T., et al. 2022, A&A, 661, A3
Santini, P., Ferguson, H. C., Fontana, A., et al. 2015, ApJ, 801, 97
Schmidt, M. 1968, ApJ, 151, 393
Schmidt, M., & Green, R. F. 1983, ApJ, 269, 352
Schwarz, G. 1978, Ann. Stat., 6, 461
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Signorini, M., Marchesi, S., Gilli, R., et al. 2023, A&A, 676, A49
Sijacki, D., Vogelsberger, M., Genel, S., et al. 2015, MNRAS, 452, 575
Simmonds, C., Buchner, J., Salvato, M., Hsu, L. T., & Bauer, F. E. 2018, A&A, 618, A66
Skilling, J. 2004, in Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, eds. R. Fischer, R. Preuss, & U. V. Toussaint, AIP Conf. Ser., 735, 395
Taylor, M. B. 2005, in Astronomical Data Analysis Software and Systems XIV, eds. P. Shopbell, M. Britton, & R. Ebert, ASP Conf. Ser., 347, 29
Toba, Y., Liu, T., Urrutia, T., et al. 2022, A&A, 661, A15
Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, ApJ, 598, 886
Ueda, Y., Watson, M. G., Stewart, I. M., et al. 2008, ApJS, 179, 124
Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, ApJ, 786, 104
Vijarnwannaluk, B., Akiyama, M., Schramm, M., et al. 2022, ApJ, 941, 97
Vito, F., Gilli, R., Vignali, C., et al. 2014, MNRAS, 445, 3557
Vito, F., Brandt, W. N., Yang, G., et al. 2018, MNRAS, 473, 2378
Volonteri, M., Dubois, Y., Pichon, C., & Devriendt, J. 2016, MNRAS, 460, 2979
Wang, F., Yang, J., Fan, X., et al. 2021, ApJ, 907, L1
Webb, N. A., Coriat, M., Traulsen, I., et al. 2020, A&A, 641, A136
Webb, N. A., Carrera, F. J., Schwope, A., et al. 2023, Astron. Nachr., 344, e20220102
Weisskopf, M. C., Tananbaum, H. D., Van Speybroeck, L. P., & O'Dell, S. L. 2000, in X-Ray Optics, Instruments, and Missions III, eds. J. E. Truemper, & B. Aschenbach, SPIE Conf. Ser., 4012, 2
Willott, C. J., Albert, L., Arzoumanian, D., et al. 2010, AJ, 140, 546
Wilms, J., Allen, A., & McCray, R. 2000, ApJ, 542, 914
Wolf, J., Nandra, K., Salvato, M., et al. 2021, A&A, 647, A5
Wolf, J., Nandra, K., Salvato, M., et al. 2023, A&A, 669, A127
Xue, Y. Q., Luo, B., Brandt, W. N., et al. 2016, ApJS, 224, 15
Yang, G., Chen, C. T. J., Vito, F., et al. 2017, ApJ, 842, 72
Yang, G., Brandt, W. N., Vito, F., et al. 2018, MNRAS, 475, 1887
Yang, G., Caputi, K. I., Papovich, C., et al. 2023, ApJ, 950, L5
Zhang, H., Behroozi, P., Volonteri, M., et al. 2023, MNRAS, 518, 2123

Appendix A: XMM-XXL-N redshift estimations

In the XMM-XXL northern field, we used an internal release obtained with the V4.2 XXL pipeline that has been used in Pouliasis et al. (2022a,b). This version is expected to be superseded by the final catalogue, V4.3. The initial catalogue contains 15547 X-ray sources. Restricting our sample to point-like sources only detected in the soft band, we ended up with 13610 X-ray sources. Since our main goal was to use the HSC data to increase the completeness of the identifications and the photometric redshift completeness, we restricted the X-ray sample into the HSC coverage. Thus, the area of the XMM-XXL-N field that overlaps with the HSC data (after excluding the HSC masked areas) is about 21 deg^2. In this area, there are 10232/13610 (~75%) X-ray sources detected in the soft band. In the next sections, we present the methodology and the data we used to determine the redshift of the sources.

A.1. Spectroscopic redshifts

The XMM-XXL-N field has been covered by several spectroscopic surveys targeting extra-galactic sources that include both AGN and galaxy samples. The majority of the targets are pre-selected in the optical or UV wavelengths, but there are many surveys dedicated only to X-ray selected sources. In our analysis, we used the spectroscopic information gathered by the HSC team and was already associated with the HSC photometric catalogue. This catalogue contains spectroscopic redshifts from the PRIsm MUlti-object Survey (PRIMUS Coil et al. 2011; Cool et al. 2013) in the sub-region XMM-LSS (~ 2.88 deg^2) of the XXL-N, the Galaxy And Mass Assembly (GAMA, Liske et al. 2015), the VIMOS VLT Deep Survey (VVDS, Le Fèvre et al. 2013), the VIMOS Public Extragalactic Survey (VIPERS, Garilli et al. 2014) and the latest data release of (SDSS-DR16, Ahumada et al. 2020). SDSS-DR16 that is the fourth release of the Sloan Digital Sky Survey IV and includes the previous data releases DR12 and DR14 (Alam et al. 2015; Pâris et al. 2018). Additionally, we made use of the spectroscopic catalogues of X-ray detected sources derived in Menzel et al. (2016) and Akiyama et al. (2015). For the latter, the spectroscopic catalogues were matched to the optical positions in our sample with a radius of 1 ″. Initially, we selected all spectroscopic redshifts that have secure measurements according to the flags provided in each catalogue. Then, we included in the spectroscopic sample sources with less reliable spectroscopic redshift estimations but that do agree with the photometric redshifts (next section) within $|\Delta z|/(1+z_{\rm spec}) > 0.10$ following Marchesi et al. (2016a) methodology. We ended up with 3673 sources with available spectroscopic information (1414 point-like and 2259 extended sources) with 70 of them being at $z > 3$.

A.2. Photometric redshifts

For the X-ray sources that do not have available spectroscopic information, we derived the photometric redshifts using the LePHARE SED fitting algorithm (Arnouts et al. 1999; Ilbert et al. 2006). LePHARE is able to provide the best fit among different templates using a χ^2 minimisation procedure between the observed and the model photometry of the SEDs. As proposed in many previous studies (e.g. Salvato et al. 2022; Ilbert et al. 2008), it is important to run LePHARE with different templates depending on the morphological type of a source (point-like or extended) to avoid degeneracies between the parameters of the models. Hence, for point-like sources we used a library of templates similar to that applied by Salvato et al. (2022), while for the X-ray sources that have an extended morphology in the optical images, we used a library of templates as in Salvato et al. (2009, 2011). In addition, we applied different luminosity priors for these samples. In particular, following previous studies (e.g. Ananna et al. 2017, and references therein), we used an absolute magnitude of $-24 < M_{HSC_g} < -8$ and $-30 < M_{HSC_g} < -20$ for the extended and point-like sources, respectively.

To construct the observed SEDs of the X-ray sources, we used the photometry described in Pouliasis et al. (2022a). Furthermore, we complemented this data set with deeper observations in the optical, near-infrared and mid-infrared wavelengths. In particular, we added the deep layer of the optical HSC-SSP PDR2 catalogue (Aihara et al. 2019) and the latest data releases of the near-infrared VISTA Hemisphere Survey (DR5, McMahon et al. 2013), VIKING (DR5, Edge et al. 2013), VISTA Deep Extragalactic Observations (VIDEO) Survey (Jarvis et al. 2013) and the DXS and UDS surveys of the UKIRT Infrared Deep Sky Survey (DR11, Lawrence et al. 2007). In addition, we have included in our analysis deeper mid-infrared data from the latest data releases of the Spitzer SWIRE (Lonsdale et al. 2003) and the SERVS (Mauduit et al. 2012) surveys.

To have an estimation of the photometric redshift accuracy, we used the sub-sample of the X-ray sources with available spectroscopic information. In particular, we used the traditional statistical indicators: the normalised median absolute deviation $\sigma_{\rm NMAD}$ (Hoaglin et al. 1983; Salvato et al. 2009; Ruiz et al. 2018) and the percentage of the catastrophic outliers η (Ilbert et al. 2006; Laigle et al. 2016) that can be defined as follows:

$$\sigma_{\rm NMAD} = 1.4826 \times median\left(\frac{|\Delta z - median(\Delta z)|}{1 + z_{\rm spec}}\right) \quad (A.1)$$

and

$$\eta(\%) = \frac{N_{\rm outliers}}{N_{\rm total}} \times 100, \quad (A.2)$$

where $\Delta z = z_{\rm phot} - z_{\rm spec}$, $N_{\rm total}$ is the total number of sources and $N_{\rm outliers}$ is the number of the outliers. An object is defined as an outlier if it has $|\Delta z|/(1 + z_{\rm spec}) > 0.15$. Figure A.1 shows the comparison between spectroscopic and photometric redshift estimations for the samples of point-like (upper) and extended (lower) sources.

We obtained η=20.9% and $\sigma_{\rm NMAD}$=0.07 for the point-like sources and η=17% and $\sigma_{\rm NMAD}$=0.07 for the extended sources. Breaking down the sample, we find that the point-like sources have a higher accuracy at higher redshifts ($z > 2$) compared to the extended sources, and, vice versa. Also, the LePHARE algorithm performs equally well at all magnitudes for the extended sample, but regarding the point-like sources there is an increase in the fraction of outliers and the scatter at the very faint magnitudes ($HSC_i \geq 24$). However, at these magnitudes the number of sources is very low compared to the bright ones. Furthermore, when considering only the spec-z sample at high redshift ($z \geq 3$), the accuracy significantly improves with η=19%. This fraction decreases to η =~ 10% for sources with $z > 3.5$ up to zero for sources with $z > 4$. This is an improvement compared to previous works, such as Salvato et al. (2022) where they found η=27.3% for their $z \geq 3$ sample.
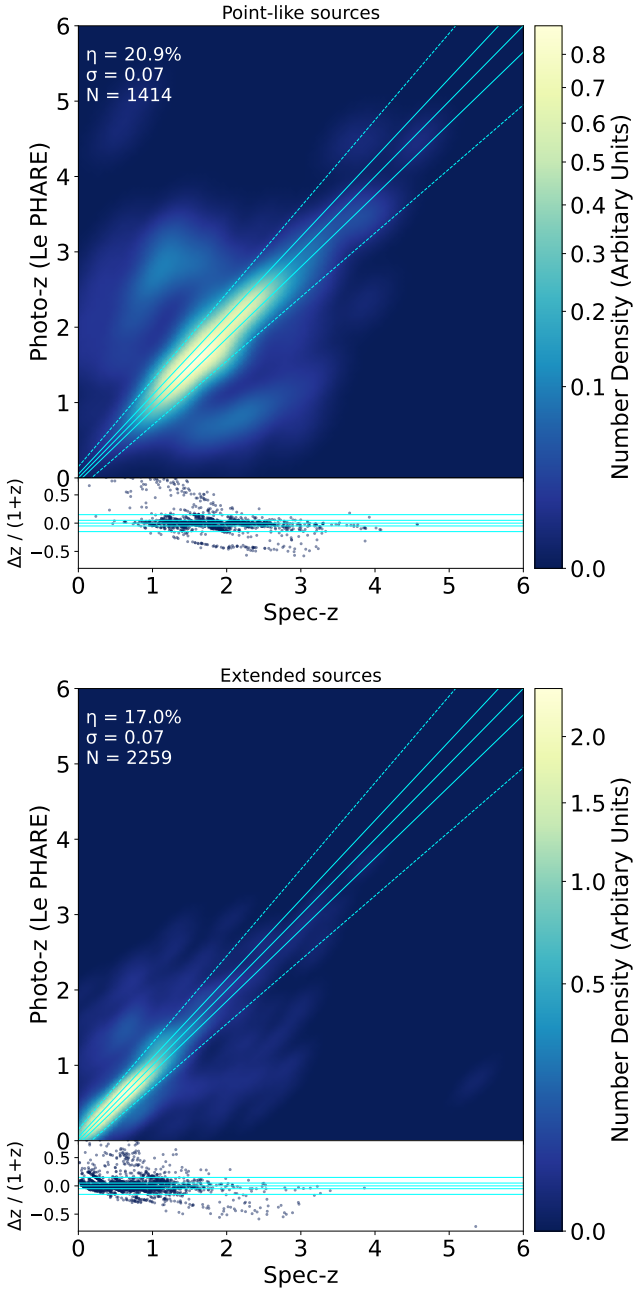
Fig. A.1. Photometric vs spectroscopic redshifts for the X-ray sources that have available spec-z information. The dotted lines represents the limits of the catastrophic outliers.

Appendix B: Goodness of fit, informational gain and reliability of X-ray spectral fits

In this appendix, we present in detail the methodology we followed to evaluate the goodness of our X-ray spectral fit analysis (App. B.1). We also explore the informational gain we obtained for the model parameters (App. B.2) and the reliability of our results (App. B.3), particularly in the case of low-count spectra.

B.1. Goodness of fit

The goodness of fit (GoF) evaluates how well the selected model in a fitting procedure reproduces the observed data. While no statistical test is able to confirm that the model is correct, if the

**Table B.1.** Input parameters for the spectral model used for the reliability simulations.

Parameter	Value
Galactic absorption	
Galactic N_H	3×10^{22} cm^{-2}
UXClumpy	
log N_H	21.5, 23.5, 24.5
phoindex	1.95
ecut	300 keV
torsigma	28 deg
ctkcover	0.4
theta_inc	18.5 deg
redshift	3.5
normalisation[a]	6.174×10^{-5}
scattering fraction	0.01

Notes. [a] Normalisation value selected such that log $L_X = 45$.

GoF does not reach a certain significance level, the model can be rejected. The Cash statistic we used for our BXA X-ray spectral fits does not provide a GoF test (unlike the χ^2 statistics, for example). However, Kaastra (2017) proposed a Cash-based test for model testing: for a given model, it is possible to calculate the expected Cash value (C_e) and its variance (C_v). For a selected significance level, it is then possible to define an interval centred around C_e. If the Cash value for the best-fit model is outside this interval, then the model is rejected. We performed this test for the X-ray spectral fits for all sources in our sample. Using a 90% significance level, the model was not rejected for all sources.

Alternatively, a robust, fully Bayesian approach for estimating the GoF is through posterior predictive checks (see e.g. Chapter 6 of Gelman et al. 2014). The basic idea of the method is to do a random sampling of the posterior distribution obtained during the fitting procedure, and then create simulated data for this random sampling. This simulated data set can be compared to the actual observed data set via visual or statistical methods. For example a qualitative, visual test can be done via QQ-plots (Buchner et al. 2014; Buchner & Boorman 2023), where the cumulative count distributions of the data and the model are compared. In the top panels of Fig. B.1, we show the differential QQ-plots using the BXA results for a *Chandra* source selected from our sample. The plots show the difference for the normalised cumulative distribution of counts between the data and the model, as a function of the observed energy. The grey shaded areas show the one- and two-sigma percentiles for 5000 simulated spectra using a sampling of the UXClumpy model parameters from the BXA posterior distribution, and the corresponding instrumental set-up for the observed spectrum (same response matrices and exposure times) and background model. The red, shaded areas in the left panel correspond to the results of the actual posterior distribution. If the results for the observed data show strong deviations from the area defined by the simulations in the QQ-plots, then the model should be rejected.

A quantitative check can be done using posterior predictive p-values (Meng 1994; Gelman et al. 2014). For each simulation in the posterior sampling, the corresponding Cash statistic is calculated. A p-value can be calculated by comparing the Cash distribution of the simulations with the distribution for the actual, observed data. If the obtained p-value is below a preselected significance level, then the model should be rejected. The lower panels of Fig. B.1 show the Cash distributions for the
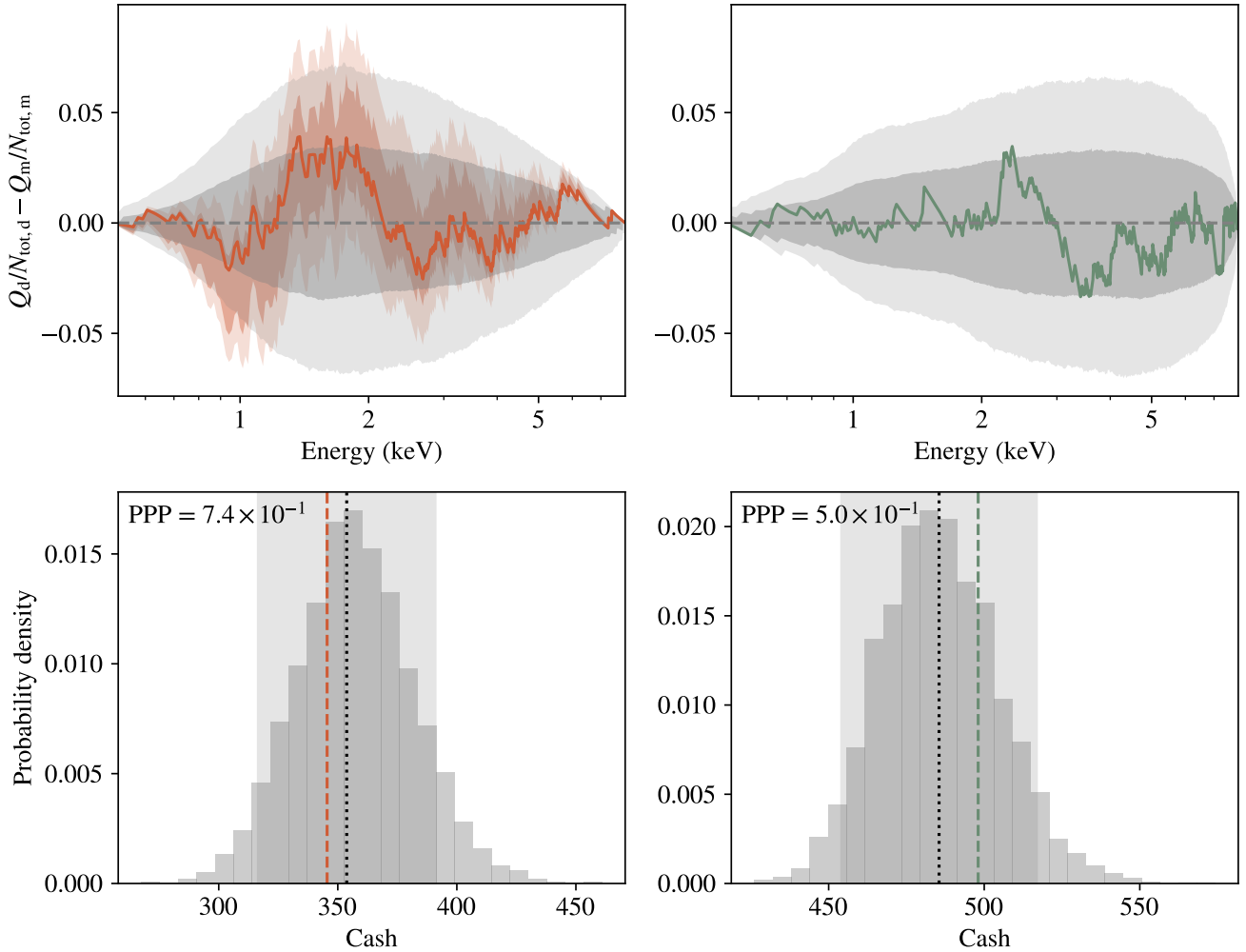
Fig. B.1. Example of posterior predictive checks for the *Chandra* source CCLS LID 460. The panels on the left show the results for the source data, while the right panels correspond to the background data. **Top panels:** Differential QQ-plots. The grey shaded areas show the one-sigma and two-sigma percentiles for 5000 X-ray spectral simulations using a random sampling of the posterior distribution. The solid lines correspond to the best-fit model obtained using BXA, and the red shaded areas are the one-sigma and two-sigma percentiles from the posterior distribution. **Bottom panels:** Grey histograms showing the distribution of the Cash statistic for the simulations. The vertical dashed lines show the Cash value for our best-fit model. The vertical dotted lines and the vertical shaded areas show, respectively, the expected Cash value (C_e) for the best-fit model and the corresponding 90% confidence region estimated using the [Kaastra (2017)](#) method. The values in the upper left corner of each panel are the posterior predictive p-values estimated using the simulations.

5000 simulations (grey histograms) and the Cash statistics for the best-fit model (vertical dashed lines). The numbers in the upper left corner of each panel correspond to the two-sided posterior predictive p-value we estimated. The vertical grey shaded areas show the 90% confidence intervals predicted by [Kaastra (2017)](#), and the black, dotted vertical lines are the corresponding expected Cash values.

We compared the results of the [Kaastra (2017)](#) method and the posterior predictive checks for a subset of sources, including low- and high-count spectra, and in all cases we found consistent results between both methods. Posterior predictive checks are computationally very expensive. Given the size of our sample we decided to use only the [Kaastra (2017)](#) method for the whole sample.

B.2. Informational gain

In this section, we investigate the knowledge we obtained for different X-ray spectral model parameters through our BXA spec-

tral fits. This is in particular interesting in the case of low-count spectra, where in principle the model parameters are not tightly constrained. To this end, we have used the Kullback-Leibler divergence (D_{KL}, [Kullback & Leibler 1951](#)), which is a statistic that quantifies the difference between the posterior and prior distributions for a given parameter (see Sect. [5](#)). We followed the method presented in [Buchner (2022)](#) for the numerical calculation of D_{KL}. As a rough, intuitive way of understanding the informational gain obtained for a given parameter, if $D_{KL} = 0.5, 1, 2$, then the shrinking factors between a Gaussian prior and a Gaussian posterior are, respectively, ~ 2, ~ 3 and ~ 6.5 (see Fig. 1 of [Buchner 2022](#)).

We calculated the D_{KL} for three parameters: $\log N_H$, the photon index Γ, and the observed 0.5-2 keV flux. While the latter is not a direct parameter of the UXClumpy model, it is a derived quantity directly linked with the overall normalisation of the model. We compared the derived posterior distribution of the X-ray flux with a flat prior covering the full range of observed fluxes in our sample.
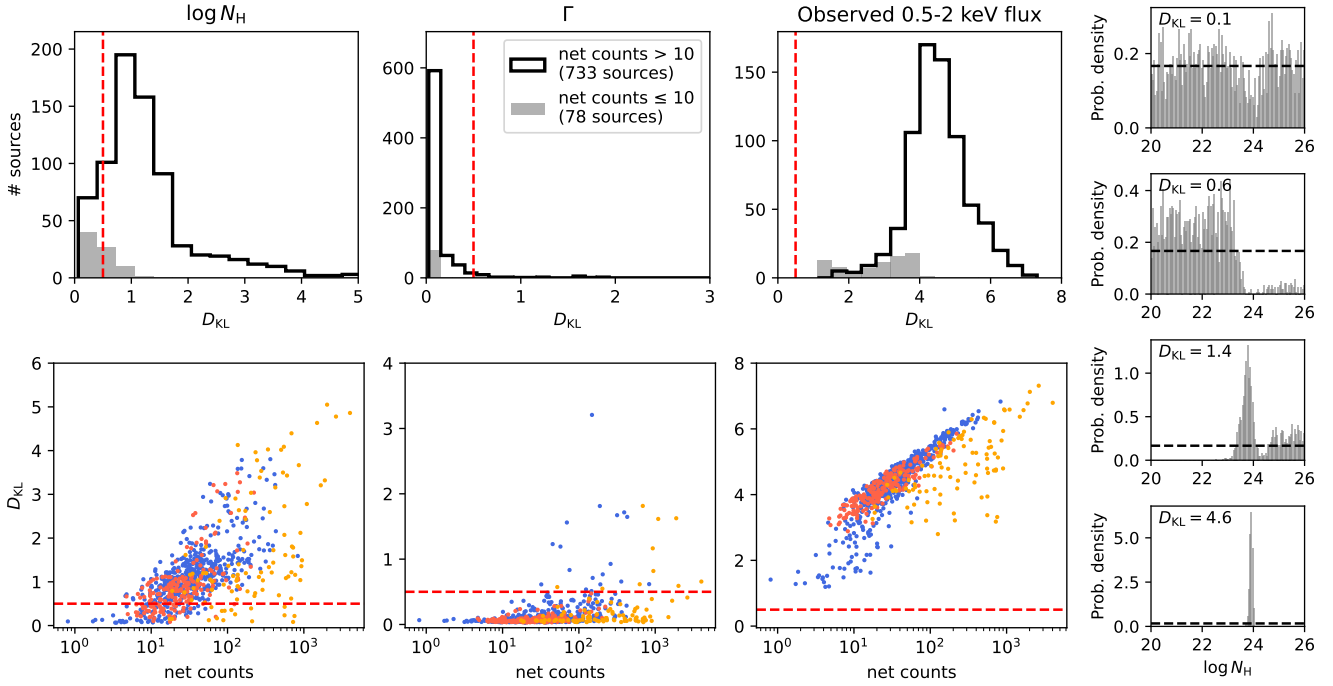
Fig. B.2. Informational gain obtained for our X-ray fitting results. **Top row:** Distribution of the Kullback-Leibler divergence for hydrogen column density, photon index, and observed X-ray flux. The solid black lines show the distribution for sources with more than ten net counts in their X-ray spectra; the grey shaded areas show the distribution for sources with fewer than ten net counts. The vertical dashed red lines show $D_{KL} = 0.5$. **Bottom row:** D_{KL} vs net counts for the same three parameters in the top row. The symbols are colour-coded according to the source survey (blue: XXL-North; red: CCLS; orange: CDF). The horizontal dashed red lines again show $D_{KL} = 0.5$. **Right column:** Four examples of the $\log N_H$ posterior distribution (grey histograms) obtained from our X-ray spectral fits, for different D_{KL} values. The horizontal dashed lines show the flat prior distribution assumed in our analysis.

Figure B.2 show our estimates of the informational gain we obtained for our X-ray fitting results. The panels in the top row show the D_{KL} distribution for the three parameters of interest, divided between sources with equal to or fewer than ten net counts (grey histograms), and more than ten counts (solid, black lines). The lower panels show the D_{KL} estimates versus the net spectral counts for each source in our sample.

The plots show how for example the informational gain for the X-ray fluxes is high ($D_{KL} > 1$) for all sources, including those with low-count spectra. This is expected, since even a detection barely above the sensitivity limit of the survey is enough to put a strong constrain in the flux of the source. We can also see how the informational gain for Γ is very limited, with most sources showing $D_{KL} \leq 0.5$, and it shows no strong dependence with the number of counts. There are two major reasons for this result: on one hand, the prior we selected for the photon index correspond to the Γ distribution for the overall population of AGN; on the other hand, the degeneracy between N_H and Γ adds a dispersion level to the posterior distribution of Γ that is difficult to reduce given the redshifts of our sources and the observed energy range of the spectral data.

In the case of the hydrogen column density, our results show that for most of the sources with spectral counts above 10, the informational gain is significant, with $D_{KL} > 0.5$. For sources with very low counts (< 5) there is no gain, the posterior and prior distributions are very similar. It is nevertheless interesting that for sources between about five and ten counts some of them show a significant informational gain, demonstrating than in such cases the posterior distribution can be constrained to a certain level. In the next section we explore how reliable are these constrains for $\log N_H$ in the case of low-count spectra.

B.3. Reliability

In this section, we explored how the number of counts available in the X-ray spectra affects the reliability of the fitting results, particularly in the low-count regime. To this end, we have followed a methodology similar to the one presented in Peca et al. (2023). Using spectral simulations we evaluated the performance of the BXA fitting results. We simulated *XMM-Newton* EPIC-PN spectra[4] assuming the UXClumpy model we used for our spectral analysis. We used three different values for the absorption level, with $\log N_H = 21.5, 23.5, 24.5$, and five exposure time values: 2, 5, 10, 20 and 50 ks. The remaining parameters of the model were kept at fixed values, as show in Table B.1. For each set of model parameters and exposure times, we run 50 simulations. Each simulation was then fitted using the same model and priors used in our BXA fitting procedure (Sect. 3.2). To quantify the reliability of the method, we calculated the match percentage for $\log N_H$ and $\log L_X$; in other words, the percentage of simulations were the 90% credible interval estimated using the posterior distribution includes the input value of the parameter for the simulated spectra.

Figure B.3 shows our match percentage results for the hydrogen column density and the intrinsic X-ray luminosity (black circles) for the three $\log N_H$ values used in the simulations, at different values in the observed counts in the simulated spectra. The overall match percentage (grey triangles) for both parameters is $\gtrsim 95\%$. The $\log N_H$ reliability for low or CT absorption is quite high, very close to 100%, even when the counts in the spectrum are below five counts. As we showed in the previous section,

---

[4] We used the response and ancillary matrices provided by SIXTE, an X-ray observation simulation software (Dauser et al. 2019).

when the spectral counts are very low the posterior distribution are poorly constrained compared with the initial prior, so the estimated confidence intervals are very large. In general, for low or CT absorption level the posteriors are not strongly constrained, giving at best an upper or lower limit estimate, respectively. For $\log N_H = 23.5$ the match percentage is lower. Giving the redshift and observed energy range of the simulations, at this absorption level the posterior are better constrained, with narrower confidence intervals. The drop in reliability is most significant between 10 and 50 counts, but even in this case the match percentage is close or above 90%.

A more detailed analysis would be needed for a complete characterisation of the reliability of the UXClumpy model using BXA spectral fits (see e.g. Saha et al. 2022), including a full exploration of the parameter space and more realistic simulations taking into account the effects of the background emission and different instrumental set-ups. Nevertheless, the results we presented here show enough evidence to conclude that no significant biases affect our estimations of the hydrogen column density or the intrinsic X-ray luminosity, even when low-count spectra are considered.
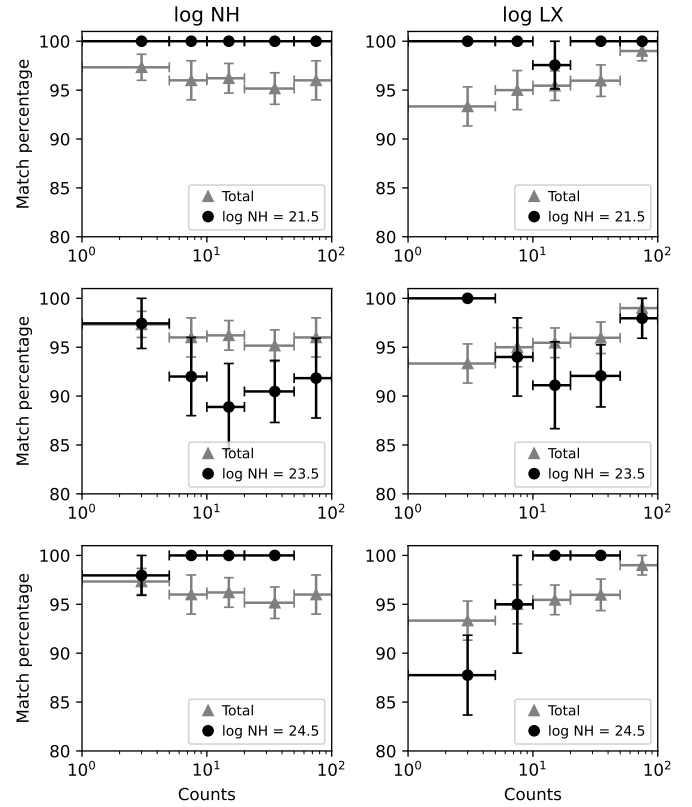
Fig. B.3. Dependence of the match percentage with the number of counts in the X-ray spectra for the hydrogen column density (left column) and the intrinsic 2-10 keV luminosity (right column). The top, middle, and bottom rows correspond to the results for sources with $\log N_H = 21.5, 23.5, 24.5$, respectively (black circles). In all panels also included are the results for the total sample, without splitting in $\log N_H$ (grey triangles). The error bars correspond to one-sigma uncertainties, calculated through bootstrapping.