Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **473,** 2378–2406 (2018) doi:10.1093/mnras/stx2486
Advance Access publication 2017 September 26

High-redshift AGN in the *Chandra* Deep Fields: the obscured fraction and space density of the sub-L_* population

F. Vito,[1,2★] W. N. Brandt,[1,2,3] G. Yang,[1,2] R. Gilli,[4] B. Luo,[5,6] C. Vignali,[7,4]
Y. Q. Xue,[8,9] A. Comastri,[4] A. M. Koekemoer,[10] B. D. Lehmer,[11] T. Liu,[8,9,12]
M. Paolillo,[13,14,15] P. Ranalli,[16] D. P. Schneider,[1,2] O. Shemmer,[17] M. Volonteri[18]
and J. Wang[8,9]

[1] *Department of Astronomy & Astrophysics, 525 Davey Lab, The Pennsylvania State University, University Park, PA 16802, USA*
[2] *Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA*
[3] *Department of Physics, The Pennsylvania State University, University Park, PA 16802, USA*
[4] *INAF – Osservatorio Astronomico di Bologna, Via Gobetti 93/3, I-40129 Bologna, Italy*
[5] *School of Astronomy and Space Science, Nanjing University, Nanjing 210093, China*
[6] *Key Laboratory of Modern Astronomy and Astrophysics (Nanjing University), Ministry of Education, Nanjing, Jiangsu 210093, China*
[7] *Dipartimento di Fisica e Astronomia, Università degli Studi di Bologna, Via Gobetti 93/2, I-40129 Bologna, Italy*
[8] *CAS Key Laboratory for Research in Galaxies and Cosmology, Department of Astronomy, University of Science and Technology of China, Hefei, Anhui 230026, China*
[9] *School of Astronomy and Space Science, University of Science and Technology of China, Hefei 230026, China*
[10] *Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*
[11] *Department of Physics, University of Arkansas, 226 Physics Building, 825 West Dickinson Street, Fayetteville, AR 72701, USA*
[12] *Astronomy Department, University of Massachusetts, Amherst, MA 01003, USA*
[13] *Dip. di Fisica Ettore Pancini, University of Naples 'Federico II', C.U. Monte SantAngelo, Via Cinthia, I-80126 Naples, Italy*
[14] *INFN Sezione di Napoli, Via Cinthia, I-80126 Napoli, Italy*
[15] *Agenzia Spaziale Italiana – Science Data Center, Via del Politecnico snc, I-00133 Roma, Italy*
[16] *Lund Observatory, Department of Astronomy and Theoretical Physics, Lund University, Box 43, SE-22100 Lund, Sweden*
[17] *Department of Physics, University of North Texas, Denton, TX 76203, USA*
[18] *Institut d'Astrophysique de Paris, Sorbonne Universitès, UPMC Univ Paris 6 et CNRS, UMR 7095, 98 bis bd Arago, F-75014 Paris, France*

Accepted 2017 September 22. Received 2017 September 15; in original form 2017 June 23

ABSTRACT

We investigate the population of high-redshift ($3 \le z < 6$) active galactic nuclei (AGN) selected in the two deepest X-ray surveys, the 7 Ms *Chandra* Deep Field-South and 2 Ms *Chandra* Deep Field-North. Their outstanding sensitivity and spectral characterization of faint sources allow us to focus on the sub-L_* regime ($\log L_X \lesssim 44$), poorly sampled by previous works using shallower data, and the obscured population. Taking fully into account the individual photometric-redshift probability distribution functions, the final sample consists of ≈102 X-ray-selected AGN at $3 \le z < 6$. The fraction of AGN obscured by column densities $\log N_H > 23$ is ∼0.6–0.8, once incompleteness effects are taken into account, with no strong dependence on redshift or luminosity. We derived the high-redshift AGN number counts down to $F_{0.5-2\,\mathrm{keV}} = 7 \times 10^{-18}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$, extending previous results to fainter fluxes, especially at $z > 4$. We put the tightest constraints to date on the low-luminosity end of AGN luminosity function at high redshift. The space density, in particular, declines at $z > 3$ at all luminosities, with only a marginally steeper slope for low-luminosity AGN. By comparing the evolution of the AGN and galaxy densities, we suggest that such a decline at high luminosities is mainly driven by the underlying galaxy population, while at low luminosities there are hints of an intrinsic evolution of the parameters driving nuclear activity. Also, the black hole accretion rate density and star formation rate density, which are usually found to evolve similarly at $z \lesssim 3$, appear to diverge at higher redshifts.

Key words: methods: data analysis – surveys – galaxies: active – galaxies: evolution – galaxies: high-redshift – X-rays: galaxies.

★E-mail: fvito@psu.edu

1 INTRODUCTION

Supermassive black holes (SMBHs) and their hosting galaxies are broadly recognized to influence the evolution of each other over cosmic time. This 'co-evolution' is reflected by the tight relations between the masses of SMBH and the properties of host galaxies in the nearby Universe, such as masses and velocity dispersions of the bulges (e.g. Magorrian et al. 1998; Ferrarese & Merritt 2000; Marconi & Hunt 2003) and the broadly similar evolution of the star formation and black hole accretion densities in the last ∼10 Gyr (e.g. Aird et al. 2015), although the details of this interplay are still not well known (see e.g. Kormendy & Ho 2013, and references therein). Studying galaxies and SMBH in the early Universe, where these relations could be set in place, would boost our knowledge of how SMBH and galaxies formed and evolved. However, while galaxy properties have been traced as far back in time as $z \sim 8$–10 (e.g. Bouwens et al. 2015), our knowledge of SMBH is limited to later times.

Only ∼90 accreting SMBHs, shining as active galactic nuclei (AGN), have been identified at $z > 6$ (e.g. Bañados et al. 2016; Wang et al. 2017) and are usually found to have masses of the order of 1–10 billion solar masses (e.g. Mortlock et al. 2011; Wu et al. 2015). The presence of such massive black holes a few 10^8 yr after the big bang challenges our understanding of SMBH formation and growth in the early Universe, one of the major issues in modern astrophysics (e.g. Reines & Comastri 2016, and references therein). Different classes of theories have been proposed to explain the formation of the BH seeds that eventually became SMBH, the two most promising ones involving 'light seeds' ($M \sim 10^2\,\mathrm{M}_{\odot}$), as remnants of the first Pop III stars, and 'heavy seeds' ($M \sim 10^{4-6}\,\mathrm{M}_{\odot}$), perhaps formed during the direct collapse of giant pristine gas clouds (e.g. Haiman 2013; Johnson & Haardt 2016; Volonteri et al. 2016b, and references therein). To match the masses of SMBH discovered at $z > 6$, all such models require continuous nearly Eddington-limited or even super-Eddington accretion phases during which the growing SMBH is plausibly buried in material with large column densities, even exceeding the Compton-thick level (e.g. Pacucci et al. 2015). However, these objects represent the extreme tail of the underlying distribution (in terms of both mass and luminosity) and are not representative of the overall population.

X-ray surveys are the most suitable tools for investigating the evolution of the bulk of the AGN population up to high redshift: being less affected by absorption and galaxy dilution, they provide cleaner and more complete AGN identification with respect to optical/IR surveys (Brandt & Alexander 2015, and references therein). Over the last two decades, several works have focused on the properties and evolution of X-ray selected, $z > 3$ AGN in wide (e.g. Brusa et al. 2009; Civano et al. 2011; Hiroi et al. 2012; Marchesi et al. 2016) and deep (e.g. Vignali et al. 2002; Fiore et al. 2012; Vito et al. 2013; Giallongo et al. 2015; Weigel et al. 2015; Cappelluti et al. 2016) surveys performed with *Chandra* and *XMM–Newton*, or using combinations of different surveys (e.g. Kalfountzou et al. 2014; Vito et al. 2014; Georgakakis et al. 2015). Common findings amongst such works are (1) a decline of the space density of luminous ($\log L_{\mathrm{X}} \gtrsim 44$) AGN proportional to $(1+z)^d$ with $d \sim -6$ (similar to the exponential decline of the space density of optically selected quasars, e.g. McGreer et al. 2013), and (2) a larger fraction of obscured AGN than that usually derived at lower redshifts, particularly at moderate-to-high luminosities (e.g. Aird et al. 2015; Buchner et al. 2015).

However, most of the low-luminosity ($\log L_{\mathrm{X}} \lesssim 43$) and $z \gtrsim 4$ AGN are missed even by the deepest surveys, leading to discrepant results amongst different studies. For instance, the evolution of the space density of low-luminosity, X-ray-detected AGN is largely unconstrained, while Georgakakis et al. (2015) reported an apparent strong flattening of the faint end of the AGN X-ray luminosity function (XLF) at $z > 3$, Vito et al. (2014) found that the decline of the space density of low-luminosity AGN is consistent with that of AGN with higher luminosities. Moreover, Giallongo et al. (2015), using detection techniques that search for clustering of photons in energy, space and time, reported the detection of several faint AGN, resulting in a very steep XLF faint end (see also Fiore et al. 2012). These results also have strong impact on the determination of the AGN contribution to cosmic reionization (e.g. Madau & Haardt 2015). Moreover, the typical obscuration levels in these faint sources remain unknown, although hints of a decrease of the obscured AGN fraction with decreasing luminosity (for $\log L_{\mathrm{X}} \lesssim 44$) at high-redshift have been found (e.g. Aird et al. 2015; Buchner et al. 2015; Georgakakis et al. 2015). This relation is the opposite trend to that found at low redshift (e.g. Aird et al. 2015; Buchner et al. 2015), where the obscured AGN fraction shows a clear anticorrelation with AGN luminosity. Finally, the very detection of faint $z > 5$ AGN in deep X-ray surveys is debated (e.g. Vignali et al. 2002; Giallongo et al. 2015; Weigel et al. 2015; Cappelluti et al. 2016; Parsa et al. 2017).

The recently completed 7 Ms *Chandra* Deep Field-South (CDF-S; Luo et al. 2017) observations provide the deepest X-ray view of the early Universe, reaching a flux limit of $F_{0.5-2\,\mathrm{keV}} = 6.4 \times 10^{-18}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$. Moreover, the catalogue of X-ray sources in the second deepest X-ray survey to date (limiting flux $F_{0.5-2\,\mathrm{keV}} = 1.2 \times 10^{-17}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$), the 2 Ms *Chandra* Deep Field-North (CDF-N; Alexander et al. 2003), was recently re-analysed by Xue et al. (2016) with the same detection procedure applied to the CDF-S, which provides detections for more real sources. Therefore, the two deepest *Chandra* fields allow us now to study high-redshift, faint AGN using homogeneous data sets. In Vito et al. (2016), we applied a stacking technique to CANDELS (Grogin et al. 2011; Koekemoer et al. 2011) selected galaxies to study the X-ray emission from individually undetected sources in the 7 Ms CDF-S, finding that the emission is probably mostly due to X-ray binaries (XRBs) rather than nuclear accretion, and concluding that most of the SMBH growth at $3.5 < z < 6.5$ occurred during bright AGN phases. In this paper, we combine the 7 Ms CDF-S and 2 Ms CDF-N data to study the X-ray properties of X-ray-detected AGN at $z > 3$, with a particular focus on low-luminosity sources ($\log L_{\mathrm{X}} \lesssim 44$), which are best sampled by deep, pencil-beam surveys. Taking fully into account the probability distribution functions (PDFs) of the photometric redshifts for sources lacking spectroscopic identifications (see e.g. Marchesi et al. 2016 for a similar use of the photometric redshifts), the final sample consists of ≈102 X-ray-detected AGN at $3 \leq z < 6$. The number of sources contributing to this sample with their PDF(z) is 118. We performed a spectral analysis on our sample assuming the X-ray spectra are well represented by power-law emission subjected to Galactic and intrinsic absorption. The spectral analysis allowed us to take into account the full probability distribution of the intrinsic column densities. We also considered the probability distribution of the count rates of X-ray-detected sources and applied a correction to mitigate the Eddington bias. The flux (and hence luminosity) probability distributions were derived by applying for each source the proper response matrices and the conversion factors between count rate and flux, which depend on the observed spectral shape. We present the trends of the obscured AGN fraction with redshift and luminosity, the number counts, and the space-density evolution of $3 < z < 6$

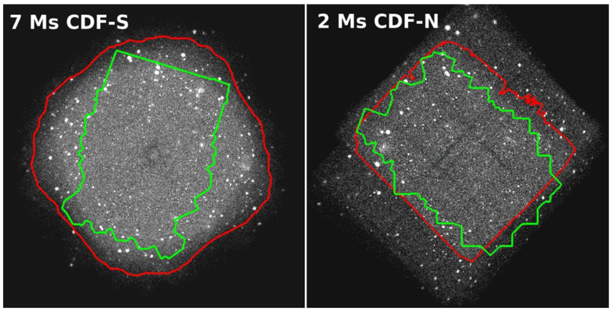

Figure 1. 0.45° × 0.45° images of the 7 Ms CDF-S (left-hand panel) and 2 Ms CDF-N (right-hand panel) in the 0.5–2 keV band. Red regions encompass the areas with effective exposure >1 Ms that are used in this work. Green polygons denote the CANDELS surveys in these fields.

AGN. Throughout this paper, we will use a $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$ cosmology and we will assume Galactic column densities of $\log N_{\mathrm{H}} = 0.9 \times 10^{20}$ and $1.6 \times 10^{20}\,\mathrm{cm^{-2}}$ along the line of sight of CDF-S and CDF-N, respectively. Errors and upper limits are quoted at the 68 per cent confidence level, unless otherwise noted.

2 THE SAMPLE

2.1 AGN parent sample and redshifts in the 7 Ms CDF-S

We selected a sample of X-ray detected, $z > 3$ AGN in the 7 Ms CDF-S,[1] the deepest X-ray survey to date, from the Luo et al. (2017, hereafter L17) catalogue, which also provides multiwavelength identifications and spectroscopic and photometric redshifts for the X-ray sources. In particular, photometric redshifts were collected from Luo et al. (2010), Rafferty et al. (2011), Hsu et al. (2014, hereafter H14), Skelton et al. (2014, hereafter S14), Santini et al. (2015) and Straatman et al. (2016, hereafter S16). Each X-ray source can therefore be associated with up to six different photometric redshifts. We considered only sources located in the area (∼330 arcmin^2, red region in the left-hand panel of Fig. 1) of the survey where the effective exposure is ≥1 Ms, in order to exclude the outskirts of the field, where the point spread function (PSF) distortions and the effects of the vignetting affect the quality of the X-ray data and the optical identification rate and accuracy. Moreover, the inner region of the CDF-S is covered by the deepest optical/IR observations (green region), which are essential to derive highly reliable spectroscopic and photometric redshifts. With this selection, our parent sample in the CDF-S consists of 952 out of the 1008 X-ray sources in L17.

We adopted the L17 definitions for the spectroscopic redshift quality. Following L17, we associated with each X-ray source a spectroscopic redshift if it is defined as 'secure' or 'insecure' but in agreement within $\frac{|z_{\mathrm{phot}}-z_{\mathrm{spec}}|}{1+z_{\mathrm{spec}}} < 15$ per cent with at least one photometric redshift. Using more conservative criteria, such as requiring that the spectroscopic redshift agrees with at least 2 or >50 per cent of the available photometric redshifts, would have no effect on the final sample of $z > 3$ AGN. The photometric redshifts used to validate the 'insecure' spectroscopic redshifts at $z > 3$ are of good quality, with a 68 per cent confidence level $\Delta z = 0.03$–0.2.

If the requirements for using the spectroscopic redshift are not satisfied, or if a source lacks spectroscopic identification, we assumed a photometric redshift from amongst those available. H14, S14 and S16 provide the PDFs of their photometric redshifts. We define a priority order amongst these catalogues by estimating the accuracy of the photometric redshifts:

$$\frac{|\Delta z|}{1+z} = \frac{|z_{\mathrm{phot}} - z_{\mathrm{spec}}|}{1+z_{\mathrm{spec}}}, \qquad (1)$$

where z_{phot} is the peak of the photometric-redshift PDF of each X-ray source in the L17 catalogue with 'secure' spectroscopic redshift, and the normalized median absolute deviation, defined as

$$\sigma_{\mathrm{NMAD}} = 1.48 \times \mathrm{Med}\left(\frac{|\Delta z - \mathrm{Med}(\Delta z)|}{1+z_{\mathrm{spec}}}\right). \qquad (2)$$

We found median values of $|\Delta z|/(1 + z_{\mathrm{spec}})$ of 0.009, 0.009 and 0.007, and $\sigma_{\mathrm{NMAD}} = 0.010$, 0.011 and 0.008 using the photometric redshifts from H14, S14 and S16, respectively. A similar assessment of the photometric redshift accuracy for the sample of high-redshift sources is presented in Section 2.3.

We also estimated the accuracy of the confidence intervals provided by the PDFs by computing the fraction of sources whose spectroscopic redshift is included in the 68 per cent confidence interval provided by its PDF (defined as the narrowest redshift interval where the integrated redshift probability is 0.68). If the PDFs provided accurate confidence intervals, that fraction would be 0.68, while we found 0.49, 0.50 and 0.63 for H14, S14 and S16, respectively, reflecting a general mild underestimate of the confidence intervals, hence of the photometric-redshift errors. This effect could be due to underestimating the errors of the fitted photometric data (e.g. see section 5.3 in Yang et al. 2014). We found indeed that the most accurate confidence intervals are provided by S16, who addressed in detail this issue by employing an empirical technique to derive more accurate photometric errors than those usually provided by detection software such as SExtractor. The reported fractions refer to the particular comparative spectroscopic sample, i.e. X-ray-selected galaxies, and are expected to be different considering the entire galaxy samples in those works. The PDFs are usually derived by fitting the observed spectral energy distribution (SED) with models M of galactic emission varying the redshift as $\mathrm{PDF}(z) \propto \exp(-0.5\chi^2(z))$, where $\chi^2(z) = \sum_i \frac{(M_i(z)-\mathrm{SED}_i)^2}{\sigma_i^2}$ is the test statistic of the fit, and the index i represents the different photometric bands. If the photometric errors σ_i are underestimated, the resulting PDFs will be too sharp and their confidence intervals will be underestimated as well. In this case, more accurate confidence intervals can be obtained by multiplying the photometric errors by a factor of α, which represents the average underestimating factor of the photometric errors amongst the used bands, or, equivalently, by using the 'corrected' distribution $\mathrm{PDF}^{\mathrm{corr}}(z) = \mathrm{PDF}^{\mathrm{input}}(z)^{\frac{1}{\alpha^2}}$, where α^2 is computed empirically such that the 68 per cent considered interval provided by their PDFs encompasses the associated spectroscopic redshift in 68 per cent of the sample. This procedure is equivalent to empirically 'correcting' (i.e. increasing) the photometric errors following $\sigma^{\mathrm{corr}} = \alpha \times \sigma^{\mathrm{input}}$. We obtained $\alpha^2 = 5.2$, 4.4 and 1.5 for H14, S14 and S16, respectively, and use the 'corrected' PDFs hereafter.

All of these tests led us to adopt the photometric redshift from S16 as first choice. The photometric redshifts from H14 and S14 have

---

[1] The integrated X-ray emission from high-mass and low-mass XRB in a galaxy can reach luminosities of $\log L_{\mathrm{X}} \approx 42$ (i.e. Lehmer et al. 2016). At $z > 3$, the flux limit of the 7 Ms CDF-S corresponds to $\log L_{\mathrm{X}} \gtrsim 42$. Therefore, we will consider all of the X-ray sources at $z > 3$ to be AGN and discuss the possible level of contamination from XRB in Section 6.

similar accuracy, but H14 provide the redshifts for the entire Extended CDF-S (E-CDF-S), while S14 (as well as S16) is limited to the GOODS-S/CANDELS region. We therefore used the photometric redshifts from H14 and S14 as second and third choices, respectively.[2] Amongst the 21 remaining sources with no entries in the above-considered photometric-redshift catalogues, two sources have a photometric redshift from Rafferty et al. (2011). We approximated their PDF(z), not provided by that work, as normalized Gaussian functions centred on the nominal redshift and with σ equal to the 1σ error, separately for the positive and negative sides. Given the very low number of sources for which we adopted the Rafferty et al. (2011) redshifts (two, and only one will be included in the final sample), the approximation we used to describe their PDF(z) does not have any effect on the final results.

Nineteen sources out of the 952 X-ray sources in the parent sample (1.9 per cent) remain with no redshift information. Most of them (14/19) are not associated with any counterpart from catalogues at different wavelengths. The five sources with an optical counterpart but no redshift information have been assigned a flat PDF. We preferred not to use redshift priors based on optical magnitudes, as these sources are part of a very peculiar class of objects (i.e. extremely X-ray faint sources), whose redshift distribution is not well represented by any magnitude-based distribution derived for different classes of galaxies. We did not include the 14 X-ray sources with no optical counterparts in this work, as a significant number of them are expected to be spurious detections. In fact, L17 estimated the number of spurious detections in the entire 7 Ms CDF-S main catalogue to be ~19 and argued that, given the superb multiwavelength coverage of the CDF-S and sharp *Chandra* PSF, resulting in high-confidence multiwavelength identifications, X-ray sources with multiwavelength counterparts are extremely likely to be real detections. Therefore, most of the spurious detections are expected to be accounted for by X-ray sources with no optical counterpart. This conclusion is especially true considering that most of the unmatched sources lie in the CANDELS/GOODS-S field, where the extremely deep multiwavelength observations would likely have detected their optical counterpart if they were true sources. We also checked their binomial no-source probability (P_B, i.e. the probability that the detected counts of a source are due to a background fluctuation), provided by L17 for all the sources in their catalogue. Thirteen out of the 14 sources with no counterparts have $P_B \gtrsim 10^{-4}$, close to the detection threshold of $P_B = 0.007$ (L17). For comparison, ~80 per cent of the sources with optical counterparts have $P_B < 10^{-4}$. The notable exception is XID 912, which has no optical counterpart, but $P_B \sim 10^{-13}$ and ≈77 net counts. L17 suggested that XID 912 is a off-nuclear X-ray source associated with a nearby low-redshift galaxy, detected in optical/IR observations.

Table 1. (1) Redshift. (2) Number of X-ray sources that were assigned a spectroscopic redshift or (3) a photometric redshift from the different catalogues. At $3 \le z < 6$, we considered only the fraction of PDF(z) in that redshift interval. (4) Median accuracy of the photometric redshifts and (5) $\sigma_{\rm NMAD}$, both computed for X-ray sources with secure spectroscopic redshifts. (6) Number of sources with a multiwavelength counterpart and no redshift information, which are assigned a flat PDF(z). (7) Number of sources with no multiwavelength counterpart, which are not included in the high-redshift sample, as most of them are expected to be spurious detections. (8) Total number of sources.

Sample (1)	$z_{\rm spec}$ (2)	$z_{\rm phot}$ (3)	Med($\frac{\lvert\Delta z\rvert}{1+z_s}$) (4)	$\sigma_{\rm NMAD}$ (5)	no z (6)	no cp. (7)	$N_{\rm tot}$ (8)
			CDF-S				
All	616	317	0.008	0.009	5	14	952
$3 \le z < 6$	18	51.5	0.015	0.014	1.5	–	71.0
			CDF-N				
All	309	166	0.026	0.031	2	8	485
$3 \le z < 6$	10	20.0	0.074	0.040	0.6	–	30.6
			CDF-S + CDF-N				
All	925	483	0.013	0.017	7	22	1437
$3 \le z < 6$	28	71.5	0.032	0.039	2.1	–	101.6

We fixed the PDFs of sources with spectroscopic redshift to zero everywhere but at the spectroscopic redshift where PDF($z_{\rm spec}$) = 1. All the PDFs are normalized such as

$$\int_0^{10} {\rm PDF}(z){\rm d}z = 1, \quad (3)$$

where the redshift range was chosen to be the same as in S16. The fraction of the PDF for the ith source at $z_1 \le z < z_2$, i.e. the probability the ith source is at $z_1 \le z < z_2$ is

$$P^i(z_1 \le z < z_2) = \int_{z_1}^{z_2} {\rm PDF}^i(z)\,{\rm d}z. \quad (4)$$

Table 1 summarizes the number of X-ray sources in the CDF-S that are associated with spectroscopic, photometric or no redshifts.

2.2 AGN parent sample and redshifts in the 2 Ms CDF-N

Xue et al. (2016, hereafter X16) presented an updated version (683 sources in total) of the original 2 Ms CDF-N catalogue (Alexander et al. 2003), applying the same detection procedure used in the CDF-S (Xue et al. 2011, L17) and providing multiwavelength identifications and spectroscopic and photometric redshifts from the literature. In particular, photometric redshifts were collected from S14 and Yang et al. (2014, hereafter Y14). Both of these works provide the PDF of the photometric redshifts. We adopted the spectroscopic redshifts collected from X16, as they considered only those marked as 'secure' in the original works.

For X-ray sources lacking spectroscopic redshift, we followed the procedure described in Section 2.1 to define a priority order amongst the two used photometric catalogues. In particular, we derived median values of $|\Delta z|/(1 + z_{\rm spec}) = 0.026$ both for S14 and Y14, $\sigma_{\rm NMAD} = 0.025$ and 0.035 and $\alpha^2 = 9.2$ and 3.0 for S14 and Y14, respectively. The α^2 values mean that the photometric redshift PDFs from Y14 account for the redshift uncertainties more realistically than the S14 ones: this behaviour can be explained again by considering that Y14 applied an empirical method to estimate the photometric errors more reliably than those provided by standard

---

[2] With two exceptions, relevant to the purpose of this work: (1) XID 638 has no spectroscopic redshift, a photometric redshift from S16 $z \sim 3.10$ and a photometric redshift from H14 $z \sim 2.64$. A significant iron line is detected in the X-ray spectrum (see also Liu et al. 2017). If it is produced by neutral iron at 6.4 keV, which is usually the strongest line, the observed line energy is consistent with the H14 redshift. Even in the case of completely ionized iron, the line is consistent with a redshift $z < 3$ at 90 per cent confidence level. (2) XID 341 has a photometric redshift $z = 5.05$ in S16, which is inconsistent with the visible emission in the GOODS-S B and V bands. Therefore, for these two sources, we adopted the H14 redshifts instead of the S16 solutions.

detection software. We therefore adopted the photometric redshifts from Y14 and S14 as first and second choice, respectively.

As in Section 2.1, we considered only the sources in the area covered by ≥1 Ms effective exposure (∼215 arcmin^2; red contour in the right-hand panel of Fig. 1, almost coincident with the CANDELS survey in that field) as the parent sample (485 sources). Amongst them, 35 sources (∼7 per cent) have no redshift information. This non-identification rate is mostly due to the method used in X16 to match the X-ray sources with the entries in the photometric-redshift catalogues. First, for each X-ray source they identified a primary multiwavelength counterpart using a likelihood-ratio procedure (Luo et al. 2010). They then matched the coordinates of the primary counterparts with the photometric-redshift catalogues using a 0.5 arcsec radius. However, the positional uncertainties and offsets amongst the different photometric catalogues can be comparable to the utilized matching radius. We therefore visually inspected the positions of all the X-ray sources on the CANDELS/GOODS-N images[3] (Grogin et al. 2011; Koekemoer et al. 2011) and could match most of the X-ray sources with no redshift information in X16 with an entry in one of the considered photometric-redshift catalogues. The visual multiwavelength matching allowed us also to associate a spectroscopic redshift from Y14 (which included only high-quality spectroscopic redshifts) with several sources with only photometric or no redshifts in X16.

The resulting number of X-ray sources with no redshift information is 12 (∼2.5 per cent). As in Section 2, we excluded sources with no multiwavelength counterpart (8 sources) from this analysis, as most of them are expected to be spurious. Two out of the remaining four sources with counterparts but no redshifts in the considered catalogues have photometric-redshift entries in the CANDELS/GOODS-N catalogue (Kodra et al. in preparation). Their PDF(z) lies entirely at $z < 3$, thus we excluded these sources from the high-redshift sample. Finally, we assigned flat PDF(z) over the range of $z = 0$–10 to the only two remaining sources with multiwavelength counterparts but no redshifts. Table 1 summarizes the number of X-ray sources in the CDF-N that are associated with spectroscopic, photometric or no redshifts.

2.3 The sample of high-redshift AGN in the *Chandra* Deep Fields

We checked the photometric redshift accuracy by plotting in Fig. 2 the $\Delta z/(1 + z_{\rm spec})$ and the $\sigma_{\rm NMAD}$ for sources with secure spectroscopic redshifts (see also Table 1). The $\sigma_{\rm NMAD}$ is computed in a shifting interval of redshift with variable width such to include 10 sources (separately for the positive and negative sides). The photometric redshift for each AGN is chosen following the priority order described in Sections 2.1 and 2.2. We considered only sources within the area covered by ≥1 Ms exposure. The scatter increases slightly at $z \sim 2.5$ (by a factor of ∼2–3), but the photometric redshift accuracy does not appear to deteriorate dramatically at high redshift.

Fig. 3 presents the redshift distributions of the sources at $z > 3$ in the two deep fields, considering their PDF(z). At $z > 6$, the source statistics are poor, and sources with no redshift information (red line), which carry little information, constitute a significant fraction of the total number of sources. We therefore chose to limit our investigation to the redshift range of $3 \leq z < 6$. To prevent the inclusion of several sources whose PDFs show long tails of extremely low probability at high redshift, we also required that the probability of a source to be in the considered redshift range (computed as in equation 4) is $P(3 \leq z < 6) > 0.2$. The number of sources satisfying this criterion is $N_{\rm hz} = 118$. Integrating their PDFs in the redshift range of $3 \leq z < 6$, the number of 'effective' sources included in the sample is $N^{\rm eff}_{\rm hz} \simeq 101.6$. The cut $P(3 \leq z < 6) > 0.2$ results in the rejection of ∼2 effective sources. Table 1 reports the breakdown of the redshift type for sources at $3 \leq z < 6$. Table 2 provides the redshift along with the 68 per cent confidence interval, its origin and the probability of the source to be at $3 \leq z < 6$ according to its PDF for each source in the high-redshift sample.

In principle, equation (4) measures only the uncertainty of the photometric redshift. The probability for a source to be at $z_1 \leq z < z_2$ can be computed by weighting equation (4) with the XLF, in order to take into account the intrinsic variation of the AGN space density with redshift, that can introduce systematic errors

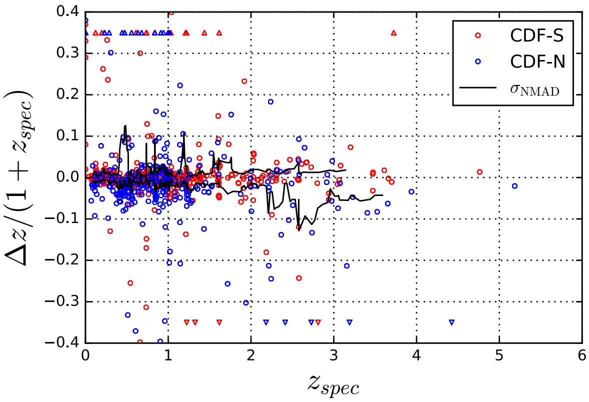

Figure 2. Photometric redshift accuracy for sources with secure spectroscopic redshift. The lines are the normalized median absolute deviations ($\sigma_{\rm NMAD}$) computed in a shifting redshift interval with variable width such to include 10 sources separately for the positive and negative sides. Points at $\Delta z/(1 + z_{\rm spec}) > 0.4$ (<-0.4) are plotted as upward-pointing (downward-pointing) triangles at $\Delta z/(1 + z_{\rm spec}) = 0.35$ (-0.35).

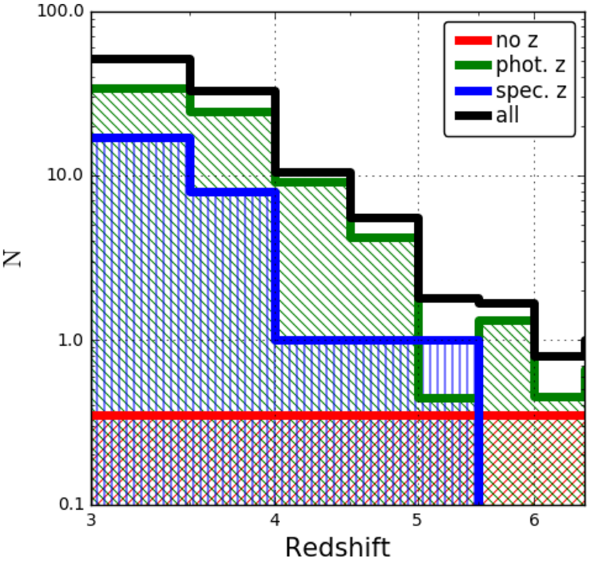

Figure 3. Redshift distribution of the X-ray sources in the CDF-S and CDF-N at $z > 3$. Different classes of sources are represented by lines of different colours.

---

[3] https://archive.stsci.edu/prepds/candels/

Table 2. Main properties of the high-redshift sources.

ID	z	Ref	P	Band	N_H	$\log F^{\rm obs}_{0.5-2\,\rm keV}$	$\log F^{\rm intr}_{0.5-2\,\rm keV}$	$\log L^{\rm obs}_{2-10\,\rm keV}$	$\log L^{\rm intr}_{2-10\,\rm keV}$	Net counts	
			($3 \le z < 6$)		(10^{22} cm^{-2})	(erg cm^{-2} s^{-1})	(erg cm^{-2} s^{-1})	(erg s^{-1})	(erg s^{-1})	(0.5–2 keV)	(2–7 keV)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
						CDF-S					
8	$3.13^{+3.28}_{-0.48}$	P3	0.55	S	59^{+24}_{-18}	$-15.56^{+0.07}_{-0.08}$	$-15.59^{+0.08}_{-0.08}$	$44.43^{+0.20}_{-0.23}$	$44.41^{+0.17}_{-0.27}$	99.2	119.4
9	$1.93^{+1.39}_{-0.20}$	P3	0.44	S	11^{+8}_{-5}	$-15.31^{+0.05}_{-0.08}$	$-15.32^{+0.06}_{-0.07}$	$44.04^{+0.16}_{-0.13}$	$44.03^{+0.15}_{-0.14}$	140.7	95.4
10	$2.93^{+0.85}_{-0.40}$	P3	0.32	S	3^{+2}_{-2}	$-15.40^{+0.06}_{-0.11}$	$-15.43^{+0.08}_{-0.09}$	$43.76^{+0.09}_{-0.13}$	$43.73^{+0.09}_{-0.14}$	70.1	38.9
12	3.153	S2	1.00	H	135^{+69}_{-46}	$-16.59^{+0.21}_{-0.75}$	<-17.28	$44.06^{+0.12}_{-0.18}$	$42.21^{+1.32}_{-0.22}$	12.4	89.0
25	-1	N	0.30	S	<74	$-15.65^{+0.08}_{-0.13}$	$-15.69^{+0.08}_{-0.15}$	$43.63^{+0.44}_{-0.20}$	$43.60^{+0.46}_{-0.18}$	75.0	88.3
29	$3.33^{+0.32}_{-0.29}$	P4	0.99	S	<1	$-15.63^{+0.05}_{-0.10}$	$-15.65^{+0.05}_{-0.11}$	$43.47^{+0.12}_{-0.13}$	$43.45^{+0.13}_{-0.12}$	119.0	44.4
35	-1	N	0.30	S	25^{+29}_{-17}	$-15.61^{+0.09}_{-0.09}$	$-15.63^{+0.08}_{-0.11}$	$43.97^{+0.21}_{-0.20}$	$43.94^{+0.20}_{-0.22}$	98.9	67.8
53	3.202	S1	1.00	S	<22	$-16.41^{+0.18}_{-0.36}$	<-16.97	$42.74^{+0.20}_{-0.40}$	<42.30	29.8	40.7
84	$3.12^{+0.35}_{-0.59}$	P1	0.70	S	49^{+14}_{-13}	$-16.07^{+0.06}_{-0.11}$	$-16.10^{+0.06}_{-0.12}$	$43.75^{+0.14}_{-0.19}$	$43.71^{+0.15}_{-0.19}$	80.4	98.5
88	-1	N	0.30	S	83^{+19}_{-26}	$-15.98^{+0.06}_{-0.08}$	$-16.01^{+0.06}_{-0.09}$	$44.19^{+0.12}_{-0.27}$	$44.16^{+0.14}_{-0.25}$	99.3	195.6
99	$4.24^{+0.05}_{-0.36}$	P1	0.99	S	30^{+34}_{-28}	$-15.91^{+0.14}_{-0.19}$	<-16.19	$43.55^{+0.22}_{-0.26}$	<43.32	41.3	47.8
101	$4.34^{+0.02}_{-0.61}$	P1	1.00	S	68^{+12}_{-22}	$-16.11^{+0.09}_{-0.16}$	$-16.20^{+0.18}_{-0.27}$	$43.87^{+0.24}_{-0.24}$	$43.78^{+0.22}_{-0.37}$	51.2	110.8
109	$3.58^{+0.03}_{-0.04}$	P1	1.0	S	28^{+5}_{-6}	$-16.48^{+0.15}_{-0.23}$	<-16.68	$43.14^{+0.16}_{-0.25}$	<42.96	144.9	150.9
110	$3.08^{+0.38}_{-0.08}$	P1	0.87	S	<1	$-16.54^{+0.15}_{-0.24}$	<-16.74	$42.55^{+0.19}_{-0.24}$	<42.40	22.8	4.4
121	$3.42^{+0.03}_{-0.05}$	P1	1.00	S	68^{+14}_{-21}	$-16.27^{+0.09}_{-0.13}$	$-16.33^{+0.11}_{-0.16}$	$43.69^{+0.16}_{-0.22}$	$43.62^{+0.16}_{-0.27}$	50.4	65.9
126	$3.13^{+0.07}_{-0.03}$	P1	0.95	S	<28	$-15.84^{+0.13}_{-0.23}$	<-16.55	$43.32^{+0.20}_{-0.22}$	<42.67	39.9	53.1
133	3.474	S1	1.00	H	141^{+73}_{-50}	<-16.96	<-17.39	$43.61^{+0.13}_{-0.22}$	$42.23^{+0.93}_{-0.15}$	4.0	56.3
165	$3.55^{+0.02}_{-0.02}$	P1	1.00	S	16^{+6}_{-6}	$-15.94^{+0.07}_{-0.09}$	$-15.96^{+0.06}_{-0.11}$	$43.52^{+0.09}_{-0.13}$	$43.5^{+0.08}_{-0.15}$	71.9	34.2
185	$3.09^{+0.04}_{-0.03}$	P1	0.87	S	39^{+24}_{-24}	$-16.70^{+0.11}_{-0.19}$	$-16.82^{+0.16}_{-0.29}$	$42.86^{+0.22}_{-0.37}$	$42.71^{+0.25}_{-0.45}$	18.7	22.6
189	$3.65^{+0.07}_{-0.04}$	P1	1.00	H	107^{+25}_{-33}	<-16.97	<-17.31	$43.38^{+0.09}_{-0.16}$	$43.30^{+0.12}_{-0.19}$	7.1	34.5
214	3.74	S2	1.00	S	26^{+1}_{-2}	$-14.81^{+0.00}_{-0.03}$	$-14.82^{+0.01}_{-0.03}$	$44.81^{+0.02}_{-0.03}$	$44.81^{+0.02}_{-0.03}$	1377.4	1213.8
223	$3.76^{+0.06}_{-0.17}$	P1	1.00	S	<14	$-16.46^{+0.11}_{-0.15}$	$-16.55^{+0.13}_{-0.22}$	$42.80^{+0.11}_{-0.19}$	$42.71^{+0.15}_{-0.23}$	27.9	0.0
238	$5.96^{+0.01}_{-0.04}$	P2	1.00	S	<66	$-16.29^{+0.09}_{-0.17}$	$-16.37^{+0.14}_{-0.21}$	$43.42^{+0.15}_{-0.17}$	$43.35^{+0.18}_{-0.24}$	40.5	7.1
248	$3.44^{+0.22}_{-0.13}$	P1	1.00	S	39^{+5}_{-7}	$-15.55^{+0.04}_{-0.06}$	$-15.56^{+0.04}_{-0.06}$	$44.18^{+0.07}_{-0.10}$	$44.17^{+0.07}_{-0.10}$	186.9	192.9
276	$3.12^{+0.02}_{-0.01}$	P1	1.00	S	15^{+3}_{-3}	$-15.90^{+0.04}_{-0.06}$	$-15.91^{+0.04}_{-0.06}$	$43.51^{+0.05}_{-0.08}$	$43.50^{+0.05}_{-0.08}$	122.9	9.2
299	$3.79^{+0.05}_{-0.05}$	P1	1.00	S	10^{+15}_{-9}	$-16.63^{+0.10}_{-0.16}$	$-16.71^{+0.12}_{-0.19}$	$42.67^{+0.13}_{-0.18}$	$42.59^{+0.15}_{-0.21}$	19.4	21.0
329	$3.47^{+0.10}_{-0.10}$	P1	1.00	S	32^{+33}_{-29}	$-16.69^{+0.14}_{-0.20}$	<-16.74	$42.61^{+0.31}_{-0.25}$	<42.77	17.2	17.1
337	3.660	S1	1.00	S	98^{+7}_{-9}	$-16.08^{+0.04}_{-0.09}$	$-16.10^{+0.05}_{-0.08}$	$44.16^{+0.07}_{-0.10}$	$44.14^{+0.08}_{-0.09}$	78.3	221.4
351	$3.86^{+0.14}_{-0.54}$	P1	0.97	S	74^{+9}_{-23}	$-16.28^{+0.08}_{-0.14}$	$-16.34^{+0.10}_{-0.17}$	$43.71^{+0.17}_{-0.19}$	$43.65^{+0.15}_{-0.26}$	50.8	73.2
356	$3.38^{+0.01}_{-0.19}$	P1	0.99	S	<8	$-16.38^{+0.07}_{-0.14}$	$-16.43^{+0.08}_{-0.15}$	$42.76^{+0.10}_{-0.13}$	$42.71^{+0.10}_{-0.15}$	31.9	31.4
366	$3.67^{+0.53}_{-0.18}$	P1	1.00	S	81^{+14}_{-15}	$-16.15^{+0.06}_{-0.10}$	$-16.18^{+0.07}_{-0.10}$	$43.96^{+0.14}_{-0.14}$	$43.93^{+0.13}_{-0.16}$	66.8	131.2
368	$3.21^{+0.09}_{-0.08}$	P1	1.00	S	<1	$-16.57^{+0.10}_{-0.15}$	$-16.65^{+0.11}_{-0.19}$	$42.52^{+0.10}_{-0.19}$	$42.44^{+0.11}_{-0.20}$	20.9	0.0
414	$3.58^{+0.11}_{-0.10}$	P1	1.00	H	$-1{**}$	<-17.28	<-17.41	$42.71^{+0.20}_{-0.36}$	<42.61	0.3	9.7
416	3.470	S1	1.00	S	<1	$-16.63^{+0.10}_{-0.19}$	$-16.75^{+0.19}_{-0.27}$	$42.52^{+0.10}_{-0.19}$	$42.40^{+0.15}_{-0.26}$	18.0	0.4
433	$4.11^{+0.15}_{-0.12}$	P1	1.00	S	<1	$-16.80^{+0.13}_{-0.23}$	<-16.86	$42.55^{+0.14}_{-0.24}$	<42.51	12.9	0.3
459	$3.91^{+0.05}_{-0.04}$	P1	1.00	H	166^{+80}_{-64}	<17.20	<-17.43	$43.34^{+0.17}_{-0.25}$	<43.11	0.3	24.4
462	$3.88^{+0.41}_{-0.78}$	P1	0.95	H	51^{+21}_{-21}	$-16.24^{+0.19}_{-0.44}$	<-17.12	$43.92^{+0.16}_{-0.25}$	<42.69	25.6	63.5
464	$4.76^{+0.02}_{-0.10}$	P2	1.00	S	<16	$-15.67^{+0.04}_{-0.07}$	$-15.68^{+0.05}_{-0.06}$	$43.81^{+0.06}_{-0.08}$	$43.80^{+0.06}_{-0.08}$	160.5	10.1
472	$4.16^{+0.02}_{-0.17}$	P1	1.00	H	110^{+317}_{-94}	<-16.97	<-17.28	$43.14^{+0.17}_{-0.24}$	<43.00	7.0	19.2
490	$4.73^{+0.06}_{-0.08}$	P1	1.00	S	71^{+9}_{-8}	$-16.05^{+0.04}_{-0.07}$	$-16.06^{+0.04}_{-0.07}$	$44.03^{+0.07}_{-0.08}$	$44.01^{+0.06}_{-0.10}$	86.8	123.3
500	$3.15^{+0.04}_{-0.02}$	P1	1.00	S	40^{+89}_{-38}	<-16.79	<-17.19	<42.83	<42.46	10.4	16.6
503	$3.14^{+0.03}_{-0.07}$	P1	0.99	S	<5	<-16.93	<-17.19	<42.20	<41.95	7.0	7.6
517	3.256	S2	1.00	S	<1	$-16.24^{+0.06}_{-0.11}$	$-16.27^{+0.06}_{-0.12}$	$42.85^{+0.07}_{-0.10}$	$42.82^{+0.07}_{-0.11}$	44.1	24.0
521	$2.96^{+0.07}_{-0.07}$	P1	0.41	S	21^{+60}_{-19}	$-16.89^{+0.18}_{-0.33}$	<-17.11	$42.34^{+0.39}_{-0.36}$	$42.60^{+0.15}_{-0.19}$	13.0	27.7
527	$3.76^{+0.05}_{-0.29}$	P1	1.00	S	<1	$-16.55^{+0.10}_{-0.15}$	$-16.63^{+0.12}_{-0.19}$	$42.68^{+0.12}_{-0.17}$	$42.60^{+0.15}_{-0.19}$	22.5	0.0
539	4.762	S1	1.00	H	102^{+49}_{-47}	$-16.48^{+0.18}_{-0.42}$	<-17.09	$43.83^{+0.13}_{-0.27}$	<42.83	27.0	61.2
551	3.700	S1	1.00	S	76^{+4}_{-5}	$-15.71^{+0.03}_{-0.05}$	$-15.72^{+0.03}_{-0.05}$	$44.37^{+0.05}_{-0.05}$	$44.37^{+0.04}_{-0.06}$	209.8	441.4

Table 2 – *continued*

ID (1)	z (2)	Ref (3)	P $(3 \leq z < 6)$ (4)	Band (5)	N_H $(10^{22}\,\mathrm{cm}^{-2})$ (6)	$\log F^{\mathrm{obs}}_{0.5-2\,\mathrm{keV}}$ $(\mathrm{erg\,cm^{-2}\,s^{-1}})$ (7)	$\log F^{\mathrm{intr}}_{0.5-2\,\mathrm{keV}}$ $(\mathrm{erg\,cm^{-2}\,s^{-1}})$ (8)	$\log L^{\mathrm{obs}}_{2-10\,\mathrm{keV}}$ $(\mathrm{erg\,s^{-1}})$ (9)	$\log L^{\mathrm{intr}}_{2-10\,\mathrm{keV}}$ $(\mathrm{erg\,s^{-1}})$ (10)	Net counts (0.5–2 keV) (11)	Net counts (2–7 keV) (12)
580	$3.17^{+0.19}_{-0.12}$	P1	0.97	S	42^{+10}_{-8}	$-16.34^{+0.06}_{-0.08}$	$-16.36^{+0.05}_{-0.10}$	$43.40^{+0.11}_{-0.13}$	$43.38^{+0.12}_{-0.12}$	45.1	66.6
617	$3.58^{+0.20}_{-0.13}$	P1	1.00	S	<1	$-15.77^{+0.03}_{-0.07}$	$-15.78^{+0.03}_{-0.07}$	$43.41^{+0.05}_{-0.08}$	$43.40^{+0.06}_{-0.07}$	136.0	69.1
619	$3.67^{+0.04}_{-0.09}$	P1	1.00	S	68^{+18}_{-23}	$-16.85^{+0.10}_{-0.20}$	$-16.97^{+0.15}_{-0.22}$	$43.12^{+0.19}_{-0.27}$	$42.99^{+0.20}_{-0.32}$	12.3	23.2
622	$3.33^{+0.09}_{-0.17}$	P1	1.00	S	<1	$-16.50^{+0.11}_{-0.19}$	$-16.62^{+0.18}_{-0.41}$	$42.62^{+0.14}_{-0.20}$	$42.49^{+0.20}_{-0.44}$	24.5	0.0
623	$3.58^{+0.02}_{-0.03}$	P1	1.00	S	27^{+10}_{-10}	$-16.43^{+0.07}_{-0.12}$	$-16.47^{+0.08}_{-0.12}$	$43.17^{+0.13}_{-0.14}$	$43.13^{+0.11}_{-0.18}$	36.7	34.5
640	$3.77^{+0.02}_{-0.03}$	P1	1.00	S	16^{+18}_{-15}	$-16.68^{+0.10}_{-0.16}$	$-16.75^{+0.10}_{-0.19}$	$42.65^{+0.16}_{-0.18}$	$42.58^{+0.18}_{-0.18}$	17.7	3.5
657	$3.58^{+0.17}_{-0.10}$	P1	1.00	S	115^{+11}_{-19}	$-16.59^{+0.09}_{-0.14}$	$-16.66^{+0.10}_{-0.17}$	$43.76^{+0.13}_{-0.19}$	$43.69^{+0.15}_{-0.20}$	26.2	79.6
662	$4.84^{+0.05}_{-0.12}$	P1	1.00	S	<54	$-16.82^{+0.14}_{-0.24}$	<-16.90	$42.71^{+0.18}_{-0.27}$	<42.68	13.3	2.6
692	$3.42^{+0.13}_{-0.08}$	P1	1.00	H	117^{+56}_{-51}	<-17.14	<-17.37	$42.84^{+0.18}_{-0.31}$	<43.22	3.6	9.7
714	$3.48^{+0.03}_{-0.09}$	P1	1.00	S	<1	$-16.25^{+0.08}_{-0.10}$	$-16.29^{+0.07}_{-0.13}$	$42.91^{+0.07}_{-0.13}$	$42.87^{+0.08}_{-0.13}$	44.4	0.0
723	3.045	S1	1.00	S	59^{+4}_{-6}	$-16.02^{+0.05}_{-0.05}$	$-16.03^{+0.05}_{-0.05}$	$43.88^{+0.05}_{-0.10}$	$43.87^{+0.05}_{-0.10}$	93.0	188.1
746	3.064	S1	1.00	S	52^{+1}_{-2}	$-15.14^{+0.02}_{-0.05}$	$-15.15^{+0.03}_{-0.04}$	$44.69^{+0.03}_{-0.05}$	$44.68^{+0.03}_{-0.05}$	689.2	1281.4
758	$3.08^{+0.01}_{-0.03}$	P1	1.00	S	<36	<-17.12	<-17.37	<42.13	<41.92	8.7	13.5
760	3.35	S2	1.00	S	74^{+2}_{-5}	$-15.111^{+0.03}_{-0.03}$	$-15.11^{+0.02}_{-0.04}$	$44.94^{+0.03}_{-0.05}$	$44.94^{+0.03}_{-0.05}$	763.5	999.8
774	3.61	S1	1.00	S	7^{+1}_{-1}	$-14.71^{+0.03}_{-0.04}$	$-14.71^{+0.02}_{-0.05}$	$44.63^{+0.04}_{-0.03}$	$44.63^{+0.03}_{-0.04}$	1162.8	691.5
788	3.193	S2	1.00	S	2^{+1}_{-1}	$-15.11^{+0.03}_{-0.03}$	$-15.11^{+0.02}_{-0.04}$	$44.01^{+0.02}_{-0.05}$	$44.01^{+0.02}_{-0.05}$	646.8	381.8
811	3.471	S1	1.00	S	<1	$-15.37^{+0.02}_{-0.05}$	$-15.37^{+0.02}_{-0.05}$	$43.77^{+0.03}_{-0.04}$	$43.77^{+0.02}_{-0.05}$	312.9	194.9
853	$3.72^{+0.13}_{-0.14}$	P1	1.00	S	<2	$-16.21^{+0.16}_{-0.29}$	<-16.81	$43.05^{+0.18}_{-0.35}$	<42.60	37.0	69.7
859	$2.88^{+0.13}_{-0.09}$	P1	0.22	S	<3	$-16.71^{+0.17}_{-0.30}$	<-16.93	$42.38^{+0.21}_{-0.37}$	<42.34	16.4	1.6
873	$3.76^{+0.04}_{-0.06}$	P1	1.00	S	<1	$-16.25^{+0.10}_{-0.12}$	$-16.31^{+0.11}_{-0.16}$	$42.98^{+0.10}_{-0.13}$	$42.92^{+0.11}_{-0.17}$	41.4	0.0
876	3.470	S2	1.00	S	11^{+1}_{-1}	$-14.57^{+0.01}_{-0.01}$	$-14.57^{+0.01}_{-0.01}$	$44.82^{+0.01}_{-0.02}$	$44.82^{+0.01}_{-0.02}$	2211.9	1482.2
885	$3.08^{+0.02}_{-0.10}$	P1	0.78	S	95^{+10}_{-10}	$-16.18^{+0.07}_{-0.10}$	$-16.21^{+0.06}_{-0.13}$	$44.02^{+0.10}_{-0.14}$	$43.98^{+0.12}_{-0.13}$	64.4	201.9
901	$3.51^{+0.25}_{-0.12}$	P1	1.00	S	<43	$-16.63^{+0.17}_{-0.32}$	<-16.94	$42.65^{+0.25}_{-0.34}$	<42.43	19.8	14.7
908	$3.60^{+0.05}_{-0.20}$	P1	1.00	H	>23	$-16.94^{+0.61}_{-0.55}$	<-17.20	$43.64^{+0.27}_{-0.26}$	<43.01	7.5	52.5
921	3.082	S1	1.00	S	<4	$-15.33^{+0.02}_{-0.07}$	$-15.34^{+0.03}_{-0.06}$	$43.73^{+0.05}_{-0.04}$	$43.73^{+0.03}_{-0.07}$	347.6	204.8
926	$4.27^{+0.06}_{-1.5}$	P1	0.53	S	<7	$-15.61^{+0.05}_{-0.05}$	$-15.62^{+0.04}_{-0.07}$	$43.76^{+0.06}_{-0.06}$	$43.75^{+0.05}_{-0.08}$	182.2	116.4
939	-1	N	0.30	S	<1	$-16.00^{+0.10}_{-0.11}$	$-16.05^{+0.10}_{-0.16}$	$43.47^{+0.12}_{-0.36}$	$43.38^{+0.15}_{-0.36}$	62.8	15.0
940	$3.00^{+0.26}_{-0.19}$	P1	0.63	S	35^{+7}_{-10}	$-16.00^{+0.06}_{-0.09}$	$-16.03^{+0.07}_{-0.09}$	$43.64^{+0.11}_{-0.16}$	$43.61^{+0.13}_{-0.14}$	91.7	97.4
962	$4.68^{+2.25}_{-2.25}$	P3	0.45	S	141^{+49}_{-48}	$-16.40^{+0.15}_{-0.31}$	<-16.95	$44.01^{+0.26}_{-0.43}$	<43.54	35.2	130.2
965	$3.64^{+0.11}_{-0.80}$	P3	0.65	S	<1	$-15.79^{+0.06}_{-0.10}$	$-15.81^{+0.06}_{-0.11}$	$43.36^{+0.10}_{-0.13}$	$43.33^{+0.10}_{-0.14}$	104.0	40.2
971	$2.17^{+1.24}_{-0.59}$	P3	0.35	S	36^{+8}_{-13}	$-15.61^{+0.02}_{-0.05}$	$-15.16^{+0.02}_{-0.05}$	$44.73^{+0.11}_{-0.28}$	$44.73^{+0.10}_{-0.29}$	536.7	513.2
974	$3.23^{+0.18}_{-0.13}$	P3	0.92	S	9^{+10}_{-8}	$-15.76^{+0.08}_{-0.10}$	$-15.79^{+0.07}_{-0.13}$	$43.44^{+0.11}_{-0.13}$	$43.41^{+0.11}_{-0.14}$	105.6	71.5
977[a]	-1	N	0.3	S	89^{+34}_{-29}	$-16.13^{+0.10}_{-0.18}$	<-16.20	$44.04^{+0.23}_{-0.29}$	<44.03	48.6	128.0
990	3.724	S1	1.00	S	62^{+64}_{-54}	$-16.06^{+0.17}_{-0.28}$	<-16.99	$43.35^{+0.58}_{-0.24}$	<42.86	41.0	61.3
CDF-N											
31	$3.45^{+0.21}_{-0.36}$	P5	0.95	S	<1	$-15.66^{+0.08}_{-0.12}$	$-15.71^{+0.08}_{-0.15}$	$43.48^{+0.12}_{-0.13}$	$43.43^{+0.11}_{-0.17}$	44.4	16.6
81	3.19	S3	1.00	S	<1	$-15.34^{+0.07}_{-0.10}$	$-15.37^{+0.07}_{-0.10}$	$43.72^{+0.07}_{-0.10}$	$43.69^{+0.07}_{-0.10}$	74.3	25.9
112	$3.15^{+0.18}_{-0.18}$	P5	0.85	S	<14	$-16.15^{+0.13}_{-0.14}$	$-16.24^{+0.13}_{-0.22}$	$43.00^{+0.15}_{-0.19}$	$42.90^{+0.18}_{-0.23}$	20.7	8.1
129	3.938	S3	1.00	S	<1	$-15.43^{+0.07}_{-0.09}$	$-15.46^{+0.07}_{-0.10}$	$43.84^{+0.07}_{-0.09}$	$43.81^{+0.06}_{-0.11}$	63.7	24.3
133	$3.06^{+0.46}_{-0.36}$	P5	0.63	S	93^{+16}_{-19}	$-16.34^{+0.15}_{-0.23}$	<-16.57	$43.84^{+0.21}_{-0.28}$	<43.62	15.0	41.0
142	$2.95^{+0.08}_{-0.13}$	P5	0.23	S	135^{+24}_{-25}	$-16.28^{+0.13}_{-0.18}$	$-16.41^{+0.14}_{-0.23}$	$44.24^{+0.23}_{-0.25}$	$44.11^{+0.22}_{-0.31}$	14.6	55.0
158	$3.19^{+0.01}_{-0.02}$	P2	1.00	S	54^{+8}_{-10}	$-15.74^{+0.08}_{-0.11}$	$-15.68^{+0.08}_{-0.12}$	$44.10^{+0.12}_{-0.14}$	$44.06^{+0.11}_{-0.16}$	46.4	72.2
177	$3.44^{+0.14}_{-0.34}$	P5	0.98	S	33^{+14}_{-12}	$-16.09^{+0.09}_{-0.15}$	$-16.16^{+0.11}_{-0.14}$	$43.55^{+0.17}_{-0.21}$	$43.48^{+0.18}_{-0.21}$	27.5	29.3
195	-1	N	0.30	S	<2	$-16.02^{+0.11}_{-0.13}$	$-16.09^{+0.11}_{-0.14}$	$43.46^{+0.13}_{-0.35}$	$43.38^{+0.15}_{-0.33}$	28.3	12.9
196	$3.24^{+0.02}_{-0.02}$	P2	1.00	S	85^{+10}_{-9}	$-16.06^{+0.12}_{-0.13}$	$-16.13^{+0.10}_{-0.16}$	$44.08^{+0.12}_{-0.19}$	$44.00^{+0.13}_{-0.18}$	29.9	96.9
201	$4.43^{+0.23}_{-1.33}$	P5	0.73	S	-1[b]	<-16.78	<-17.33	<42.98	$44.06^{+0.12}_{-0.19}$	3.4	15.5
207	3.652	S3	1.00	S	<1	$-15.09^{+0.05}_{-0.08}$	$-15.11^{+0.05}_{-0.08}$	$44.10^{+0.05}_{-0.08}$	$42.21^{+0.30}_{-0.15}$	242.6	80.7
227	$4.26^{+0.03}_{-0.07}$	P2	1.00	S	-1[b]	$-16.70^{+0.24}_{-0.31}$	<-17.00	$42.71^{+0.40}_{-0.55}$	<43.18	6.5	3.9

Table 2 – *continued*

ID (1)	z (2)	Ref (3)	P $(3 \leq z < 6)$ (4)	Band (5)	N_H $(10^{22}\,\mathrm{cm}^{-2})$ (6)	$\log F^{\mathrm{obs}}_{0.5-2\,\mathrm{keV}}$ $(\mathrm{erg\,cm^{-2}\,s^{-1}})$ (7)	$\log F^{\mathrm{intr}}_{0.5-2\,\mathrm{keV}}$ $(\mathrm{erg\,cm^{-2}\,s^{-1}})$ (8)	$\log L^{\mathrm{obs}}_{2-10\,\mathrm{keV}}$ $(\mathrm{erg\,s^{-1}})$ (9)	$\log L^{\mathrm{intr}}_{2-10\,\mathrm{keV}}$ $(\mathrm{erg\,s^{-1}})$ (10)	Net counts (0.5–2 keV) (11)	Net counts (2–7 keV) (12)
229	3.413	S4	1.00	S	<1	$-15.43^{+0.06}_{-0.09}$	$-15.46^{+0.06}_{-0.09}$	$43.70^{+0.06}_{-0.09}$	$43.67^{+0.06}_{-0.09}$	111.8	48.8
257	$2.95^{+0.29}_{-0.44}$	P5	0.35	S	60^{+37}_{-32}	$-16.56^{+0.14}_{-0.22}$	$-16.74^{+0.19}_{-0.30}$	$43.24^{+0.32}_{-0.43}$	$43.04^{+0.37}_{-0.47}$	9.6	16.7
290	4.424	S3	1.00	S	<1	$-16.24^{+0.10}_{-0.17}$	$-16.33^{+0.11}_{-0.18}$	$43.14^{+0.11}_{-0.16}$	$43.05^{+0.12}_{-0.17}$	17.7	9.1
293	$3.96^{+0.17}_{-0.30}$	P5	1.00	S	<8	$-15.32^{+0.06}_{-0.07}$	$-15.33^{+0.06}_{-0.07}$	$43.98^{+0.07}_{-0.12}$	$43.96^{+0.08}_{-0.11}$	115.0	66.9
297	3.23	S3	1.00	S	<1	$-16.09^{+0.12}_{-0.12}$	$-16.16^{+0.10}_{-0.16}$	$42.99^{+0.10}_{-0.15}$	$42.02^{+0.11}_{-0.15}$	22.3	10.8
310	$3.06^{+0.36}_{-0.23}$	P5	0.70	S	<41	$16.67^{+0.19}_{-0.32}$	<-16.93	$42.52^{+0.30}_{-0.40}$	<42.47	6.6	6.5
315	-1	N	0.30	S	<1	$-16.61^{+0.16}_{-0.25}$	$-16.86^{+0.21}_{-0.51}$	$42.78^{+0.23}_{-0.36}$	$42.46^{+0.31}_{-0.51}$	7.5	2.8
320	$3.06^{+0.26}_{-0.20}$	P5	0.69	S	<1	$-16.73^{+0.19}_{-0.28}$	<-16.91	$42.37^{+0.21}_{-0.31}$	<42.23	5.8	5.5
330	5.186	S3	1.00	S	<1	$-15.53^{+0.07}_{-0.12}$	$-15.57^{+0.07}_{-0.12}$	$44.00^{+0.07}_{-0.12}$	$43.96^{+0.07}_{-0.12}$	84.4	31.4
363	$3.28^{+0.05}_{-0.05}$	P2	1.00	S	<1	$-16.19^{+0.14}_{-0.17}$	$-16.31^{+0.14}_{-0.21}$	$42.92^{+0.15}_{-0.17}$	$42.80^{+0.15}_{-0.21}$	15.8	3.9
373	$3.36^{+0.34}_{-0.29}$	P5	0.84	S	52^{+31}_{-31}	$-16.76^{+0.19}_{-0.27}$	<-16.92	$42.98^{+0.25}_{-0.51}$	<42.85	5.9	6.5
388	$2.94^{+0.08}_{-0.12}$	P5	0.31	S	<1	$-15.49^{+0.07}_{-0.12}$	$-15.53^{+0.08}_{-0.13}$	$43.56^{+0.08}_{-0.13}$	$43.52^{+0.09}_{-0.14}$	56.5	0.0
390	$3.19^{+0.28}_{-0.29}$	P5	0.85	H	204^{+53}_{-46}	<-16.89	<-17.39	$43.38^{+0.07}_{-0.13}$	$43.30^{+0.09}_{-0.16}$	3.1	31.5
404	$3.15^{+0.04}_{-0.13}$	P5	0.91	S	<1	$-14.67^{+0.03}_{-0.06}$	$-14.67^{+0.03}_{-0.06}$	$44.37^{+0.04}_{-0.06}$	$44.36^{+0.04}_{-0.06}$	555.4	220.8
424	$4.43^{+0.13}_{-0.26}$	P5	1.00	H	166^{+38}_{-40}	$-16.37^{+0.18}_{-0.36}$	<-17.17	$43.34^{+0.17}_{-0.25}$	$42.32^{+0.58}_{-0.11}$	13.1	62.2
428	$3.03^{+0.09}_{-0.11}$	P5	0.55	S	18^{+2}_{-3}	$-15.07^{+0.04}_{0.06-}$	$-15.08^{+0.04}_{-0.06}$	$44.36^{+0.05}_{-0.08}$	$44.35^{+0.05}_{-0.08}$	258.7	188.1
439	3.406	S3	1.00	S	27^{+73}_{-26}	$-16.54^{+0.17}_{-0.28}$	$-16.81^{+0.27}_{-0.46}$	$42.76^{+0.41}_{-0.35}$	$42.49^{+0.46}_{-0.51}$	8.5	15.8
498	$4.42^{+0.31}_{-0.24}$	P5	0.99	S	23^{+6}_{-6}	$-15.33^{+0.06}_{-0.09}$	$-15.35^{+0.05}_{-0.10}$	$44.36^{+0.09}_{-0.12}$	$44.34^{+0.09}_{-0.12}$	149.2	86.2
502	3.157	S3	1.00	S	<1	$-15.56^{+0.08}_{-0.13}$	$-15.61^{+0.10}_{-0.10}$	$43.50^{+0.08}_{-0.13}$	$43.45^{+0.08}_{-0.13}$	73.5	29.8
504	$3.36^{+0.10}_{-0.09}$	P5	1.00	H	831^{+123}_{-537}	<-16.31	<-17.57	$43.14^{+0.17}_{-0.24}$	$42.26^{+0.59}_{-0.13}$	9.8	30.8
509	$3.15^{+0.12}_{-0.57}$	P5	0.47	S	<15	$-16.26^{+0.18}_{-0.26}$	<-16.93	$42.88^{+0.21}_{-0.32}$	<42.30	15.1	3.2
526	$4.17^{+0.10}_{-0.13}$	P5	1.00	S	15^{+7}_{-6}	$-15.84^{+0.07}_{-0.10}$	$-15.88^{+0.08}_{-0.10}$	$43.71^{+0.09}_{-0.14}$	$43.67^{+0.09}_{-0.15}$	89.0	48.0
545	3.524	S3	1.00	S	<11	$-15.27^{+0.05}_{-0.07}$	$-15.28^{+0.05}_{-0.07}$	$43.95^{+0.07}_{-0.09}$	$43.93^{+0.08}_{-0.08}$	132.4	46.8
572	$3.77^{+0.03}_{-0.01}$	P2	1.00	S	10^{+2}_{-2}	$-15.49^{+0.08}_{-0.13}$	$-15.54^{+0.10}_{-0.17}$	$43.94^{+0.08}_{-0.14}$	$43.89^{+0.11}_{-0.17}$	288.9	156.2

(1) ID from L17 or X16; (2) nominal redshifts (corresponding to the peaks of the PDFs) and uncertainties corresponding to the narrowest redshift intervals in which the integral of the PDFs is 0.68; (3) reference for the redshifts and PDFs: P1, P2, P3, P4 and P5 refer to photometric redshifts from S16, S14, H14, Rafferty et al. (2011) and Y14, respectively. S1 and S2 stand for spectroscopic redshift from the collection presented in L17, flagged as 'secure' and 'insecure' (but upgraded as described in Section 2.1), respectively. S3 and S4 refer to spectroscopic redshift from X16 and Y14 (for sources for which the multiwavelength matching has been improved via the visual inspection of X-ray and optical images, see Section 2.2), respectively. N refers to no redshift and, in such a case, column 2 is assigned a value of -1 and a flat PDF(z) is assumed. (4) Fraction of the PDFs at $3 \leq z < 6$ (see equation 4); (5) primary detection band (S: soft band, H: hard band; see Section 3.1); (6), (7), (8), (9) and (10): best estimates of the intrinsic column density, soft-band flux, 2–10 keV intrinsic luminosity and 68 per cent confidence-level uncertainties, respectively. For fluxes and luminosities, both the observed (i.e. not applying the Eddington bias correction in equation 5; columns 7 and 9) and the intrinsic (i.e. for which the Eddington bias correction has been applied; columns 8 and 10) values are reported.

Notes. [a]XID 977 has an insecure spectroscopic redshift in L17, but no photometric redshifts. Here, we conservatively discard that spectroscopic redshift and treat this source as spectroscopically unidentified.

[b]For these sources, we assumed a flat distribution in column density, as their spectral quality was too poor to perform a spectral analysis. (11) and (12): net counts in the soft and hard bands, respectively.

(e.g. Georgakakis et al. 2015). However, the general high quality of the photometric redshifts we used (see Sections 2.1 and 2.2, and Table 1), due to the availability of deep multiband data in the fields we considered, is reflected in narrow PDF(z). In fact, the PDF(z) of the spectroscopically unidentified sources included in our sample have a median 1σ uncertainty of 0.23. Such narrow PDF(z) would be at most sensitive to the 'local' XLF. For continuity reasons, the XLF cannot vary dramatically in a narrow redshift range and the applied weights are therefore expected to be negligible. This statement may not be true in cases of broader PDF(z), which would be therefore sensitive to the space-density variations over larger redshift intervals.

To quantify this effect in our case, we repeated the procedure followed to derive the redshift distribution of the sample, after weighting the PDF(z) with the Ueda et al. (2014) XLF. We also considered additional cases in which the Ueda et al. (2014) XLF is modified at high redshift to match the results of Vito et al. (2014) and Georgakakis et al. (2015). The new redshift distributions are barely distinguishable from Fig. 3, with 1.7–5.9 per cent (depending on the particular XLF used) of the sample shifting below $z = 3$. Including in this computation sources with no redshift information (i.e. for which a flat PDF(z) is assumed), we found that 2.5–7.7 per cent of the samples shift to low redshift. Moreover, the weighting is sensitive in particular to the assumed shape and evolution of the XLF faint end at high redshift, which is currently highly uncertain. For instance, using models with a steeper XLF faint end (e.g. Giallongo et al. 2015) would move a fractional number (<1) of sources from $z \lesssim 3$ to $z > 3$, slightly increasing our sample size. Given the small effect obtained applying the XLF weighting and the source of additional uncertainty represented by the particular XLF assumed at high redshift, we will use equation (4) to measure the intrinsic redshift probability distribution.

We also made a basic assessment of the effects of the expected presence of redshift outliers in our high-redshift sample (i.e. intrinsically low-redshift sources scattered to high redshift because of the catastrophic failure of their photometric redshifts, and hence PDF(z)). We considered only X-ray sources with a high-quality spectroscopic redshift $z_{\rm spec} < 3$, assigned to them a photometric redshift following the priority order described in Sections 2.1 and 2.2, and computed the fraction of redshift outliers, defined as sources for which $|z_{\rm spec} - z_{\rm phot}|/(1 + z_{\rm spec}) > 0.15$ (see e.g. Luo et al. 2010), that would be counted in the high-redshift sample according to their photometric redshift (i.e. PDF(z)). Considering this fraction and the number of X-ray-selected, spectroscopically unidentified sources with photometric redshift $z_{\rm phot} < 3$, we estimated that ≈4 sources in our high-redshift sample (≈4 per cent) are such low-redshift interlopers. The outlier fraction may be higher for spectroscopically unidentified sources, since these are typically fainter than the control sample we used, but our results are not qualitatively affected even considering an outlier fraction of ≈10 per cent.

Our sample includes a few sources studied in detail in previous, dedicated works, like the two Compton-thick candidate AGN in the CDF-S at $z = 4.762$ (Gilli et al. 2014, ID 539 in Table 2) and $z = 3.70$ (Comastri et al. 2011, ID 551). We also include the three spectroscopically-identified $z > 4$ AGN in the CDF-N analysed by Vignali et al. (2002): ID 290 at $z = 4.424$, ID 330 at $z = 5.186$, which is also the X-ray source with the highest spectroscopic redshift in the *Chandra* deep fields, and ID 526 at $z = 4.137$.[4]

3 SPECTRAL ANALYSIS AND PARAMETER DISTRIBUTIONS

X-ray spectra of the sources in the $z > 3$ sample were extracted with ACIS EXTRACT (Broos et al. 2010), as described in L17 and X16. Following Vito et al. (2013), we analysed the individual X-ray spectra with XSPEC[5] v12.9.0 (Arnaud 1996) assuming an absorbed power-law model with fixed $\Gamma = 1.8$ (a typical value for X-ray selected AGN, e.g. Liu et al. 2017), Galactic absorption (Kalberla et al. 2005) and intrinsic absorption (XSPEC model *wabs* × *zwabs* × *powerlaw*), for a total of two free parameters (the intrinsic column density, $N_{\rm H}$, which assumes solar metallicity, and the power-law normalization), in the energy range of 0.5–7 keV. More complex parametrizations are precluded by the general paucity of net counts characterizing the sources in the sample, which, being at high redshift, are typically faint. Appendix A describes our check through spectral simulations that the simplicity of the assumed spectral model does not affect the best-fitting column densities at $\log N_{\rm H} \gtrsim 23$ (see Section 5 for discussion). In Appendix B, we study the photon index of a subsample of unobscured sources. The *W* statistic,[6] based on the Cash statistic (Cash 1979) and suitable for Poisson data with Poisson background, was used to fit the model to the unbinned spectra. If a spectroscopic redshift is assigned to a source, the fit was performed at that redshift, otherwise we used a grid of redshifts (with step $\Delta z = 0.01$), obtaining for each source a set of best-fitting models as a function of redshift. In Section 3.1, we derive the probability distribution of net count rate for each source, including a correction for Eddington bias. In Section 3.2, we convolve the set of best-fitting models with the redshift PDF for each source to derive the probability distribution of best-fitting column density. The fluxes of the sources are derived from the observed count rates in Section 3.3, where we also present the log*N*-log*S* of the sample. In Section 3.4, we combine the fluxes and column densities to derive the intrinsic X-ray luminosities. Results from spectral analysis are reported in Table 2.

3.1 Count rates, sky-coverage and Eddington bias

The fundamental quantity measured by an X-ray detector is the count rate of a source, which can then be converted into energy flux assuming a spectral shape and using the detector response matrices. Absorption affects strongly the spectral shape, in particular in the observed soft band (0.5–2 keV). Therefore, different levels of absorption affecting sources with the same count rate result in different measured fluxes. Usually, when performing statistical studies of populations of astrophysical objects detected in X-ray surveys, weights are assigned to each source to take into account the different sensitive areas of the survey at different fluxes. The curves of the area sensitive to a given flux (namely, the sky-coverage) for the 7 Ms CDF-S and 2 Ms CDF-N are shown in Fig. 4 (where we considered only the area covered by ≥1 Ms effective exposure, as per Section 2.1) for our two detection bands (see below). Since the source-detection algorithm used by L17 and X16 operates in the count rate space, sources with the same fluxes in the detection energy band can be detected over different areas, if their spectral shapes are different (i.e. the same count rate can correspond to different fluxes). This is usually not taken into account, and an average spectral shape (usually a power-law with $\Gamma = 1.4$) is assumed for all the sources, which can introduce errors up to a factor of several in the weights associated with each source (Vito et al. 2014).

We will therefore remain in the count rate space, transforming count rates to energy fluxes, when needed, by assuming every time the specific spectral shape and response files suitable for a source. In doing so we do not assume any a priori spectral shape to compute the weights related to the sky-coverage for our sources. Following Lehmer et al. (2012), we derived the probability distribution of the count rate for one source as

$$P({\rm CR}) \propto \left(\frac{T}{T+b}\right)^{s} \left(1-\frac{T}{T+b}\right)^{b} \frac{{\rm d}N}{{\rm dCR}} \tag{5}$$

where

$$T = t_{\rm exp}\,{\rm CR}\,f_{\rm PSF} + b\,\frac{A_{\rm s}}{A_{\rm b}}. \tag{6}$$

In these equations, s and b are the numbers of counts detected in the source and background regions, which have areas of $A_{\rm s}$ and $A_{\rm b}$, respectively, $t_{\rm exp}$ is the exposure time and $f_{\rm PSF}$ is the fraction of *Chandra* PSF corresponding to $A_{\rm s}$ at the position of the source. The parameter T is therefore the expected total (i.e. source plus background) number of counts detected in the $A_{\rm s}$ region for a source with a net count rate of CR. The first two factors of equation (5) derive from the binomial probability of observing s counts in the $A_{\rm s}$ region, given the expected number T and $b\frac{A_{\rm s}}{A_{\rm b}}$ background counts. The last factor is the X-ray source number counts (differential number of X-ray sources per unit count rate) derived by L17 using the 7 Ms CDF-S data set. This factor includes our knowledge of the intrinsic

---

[4] However, the quality of the optical spectrum of the latter source, presented in Barger et al. (2002), is low, and Y14 and X16 discarded that spectroscopic redshift in favour of a photometric redshift $z = 4.17^{+0.10}_{-0.13}$ from Y14, which we adopted too, following the procedure described in Section 2.2.

[5] https://heasarc.gsfc.nasa.gov/xanadu/xspec/

[6] https://heasarc.gsfc.nasa.gov/xanadu/xspec/manual/XSappendixStatistics.html

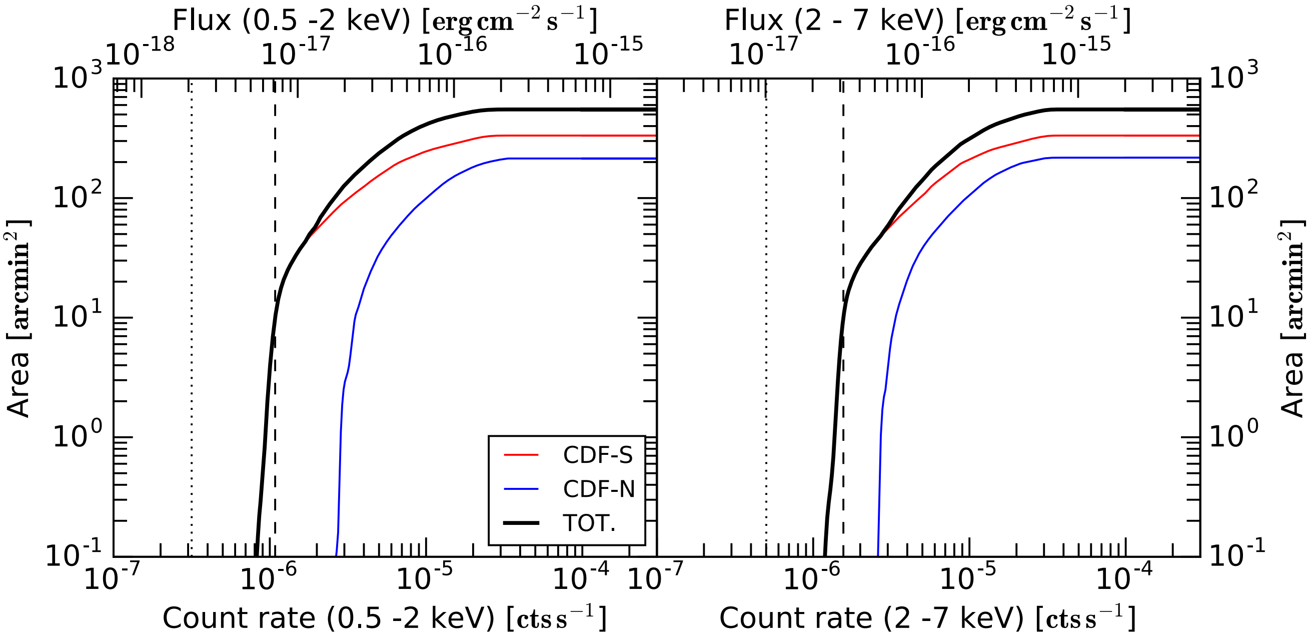

Figure 4. Area covered by ≥1 Ms effective exposure sensitive to a given minimum count rate (lower x-axis) or flux (upper x-axis, assuming a power-law spectral shape with $\Gamma = 1.4$) for the CDF-S (red curve), CDF-N (blue curve) and for the combination of the two surveys (black curve), in the soft band (left-hand panel) and hard band (right-hand panel). Different spectral assumptions would change only the flux axis. The dotted vertical line marks the minimum *intrinsic* count rate we assumed for the sources in our sample (equation 7). The vertical dashed lines indicate the minimum count rate corresponding to a 10 arcmin^2 sensitive area, which we consider the observed count rate limit for a source to be detected (see equation 8) to avoid large correction factors due to the small area sensitive at lower count rates.

distribution in flux (or, in this case, count rate) of the AGN population, and accounts for the Eddington bias (see Lehmer et al. 2012 for discussion), expected for our sample of faint sources. Count rate probability distributions are preferentially computed in the soft band, where the *Chandra*/ACIS sensitivity peaks. For sources undetected in the soft band and detected in the hard band (2−7 keV) in the X-ray catalogues, we computed the count rates in the hard band. L17 provide AGN number counts for both bands. The number of sources in our sample detected in the soft (hard) band is 108 (14), as flagged in Table 2.[7]

The $\frac{\mathrm{d}N}{\mathrm{d}CR}$ factor diverges for very faint fluxes, which can be explained by considering that the L17 model is constrained only down to the 7 Ms CDF-S flux limit. For this reason, we set a hard lower limit to the *intrinsic* count rate, $\mathrm{CR}_{\mathrm{lim}}^{\mathrm{intr}} = 10^{-6.5}$ and $10^{-6.3}$cts s^{-1} for the soft and hard bands, respectively, a factor of ∼0.5 dex fainter than the nominal count rate limit of the survey (dotted vertical lines in Fig. 4) in those bands and normalized P(CR) to unity above that limit:

$$\int_{\mathrm{CR}_{\mathrm{lim}}^{\mathrm{intr}}}^{\infty} P(\mathrm{CR})\,\mathrm{dCR} = 1; \tag{7}$$

i.e. we assume that the intrinsic count rate of a detected source is at most ∼0.5 dex lower than the 7 Ms CDF-S count rate limit. This approach allows us to apply the Eddington bias correction and obtain convergent probabilities. The small area covered by the 7 Ms CDF-S at fluxes close to that limit would result in unreasonably large weights for very small count rates. To prevent this bias, we applied another cut, $\mathrm{CR}_{\mathrm{lim}}^{\mathrm{obsc}}$, corresponding to the count rate at which the sensitive area is, following Lehmer et al. (2012), ≥10 arcmin^2 in the two bands (dashed vertical lines in Fig. 4). Finally, we define the weighting factor

$$\Omega(\mathrm{CR}) = \begin{cases} \frac{\mathrm{MAX(sky\text{-}coverage)}}{\mathrm{sky\text{-}coverage(CR)}} & \text{if CR} > \mathrm{CR}_{\mathrm{lim}}^{\mathrm{obsc}} \\ 0 & \text{otherwise,} \end{cases} \tag{8}$$

where sky-coverage(CR) is the black curve in Fig. 4 in each band, i.e. the sum of the sky-coverages of the CDF-S and CDF-N (computed by L17 and X16, respectively), and the maximum sensitive area covered by the survey with ≥1 Ms effective exposure is MAX(sky-coverage) ∼ 551 arcmin^2. Using the combined sky-coverage, we assume that each source could in principle be detected in both surveys, i.e. we adopt the coherent addition of samples of Avni & Bahcall (1980).

3.2 Intrinsic column-density distribution

Once the best-fitting model for a source at a given redshift had been obtained, we derived the observed best-fitting logN_{H} distribution, $P(\log N_{\mathrm{H}}|z)$, through the XSPEC *steppar* command. The probabilities are normalized such that

$$\int_{\log N_{\mathrm{H}}=20}^{25} P(\log N_{\mathrm{H}}|z)\,\mathrm{d}\log N_{\mathrm{H}} = 1. \tag{9}$$

The intrinsic column density probability *at a given redshift* is then weighted by the PDF(z) at high redshift to derive the intrinsic

---

[7] Two sources in our sample, XID 500 and 503, in the 7 Ms CDF-S catalogue are reported as full-band detections only. Given the small number of such objects, for the purposes of this work we consider them as soft-band detections.

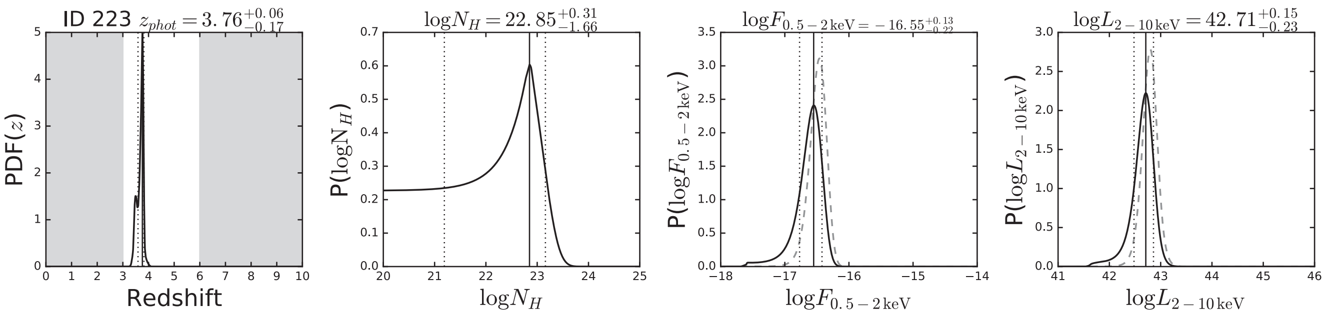

Figure 5. From left to right: PDFs of redshift, column density (equation 10), flux (equation 16) and luminosity (equation 20) for ID 223. The area not shadowed in the first panel shows the redshift range of this work and the part of the PDF(z) used to compute the probability distributions of the other parameters. In each panel, the solid and dotted vertical lines mark the position of the nominal value and the 68 per cent confidence interval of the relative parameter. Dashed grey curves in the last two panels are the probability distribution of flux and luminosity if the correction for Eddington bias is not applied in equation (5). Probability distributions for all sources will be made available online. Most of the distributions are well behaved like those shown here, while a few have more complex (i.e. broad or with double peaks) shapes. The complex functions are produced by the combined effects of the measured PDF(z), and, especially for faint sources, the uncertainties on $N_{\rm H}$ and flux, and the area-correction (i.e. the Ω) factors.

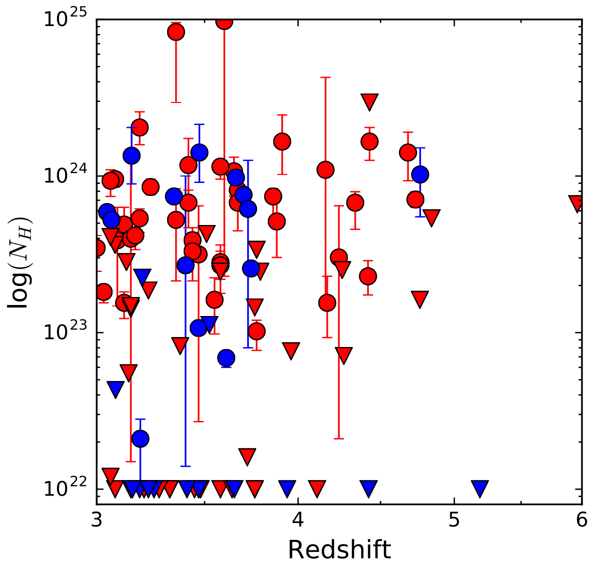

Figure 6. Best-fitting column density versus redshift for sources with spectroscopic (blue symbols) and photometric (red symbols) redshift. Upper limits are displayed as downward-pointing triangles.

N_H probability distribution for each source:

$$P(\log N_{\rm H}) = \frac{\int_{z=3}^{6} P(\log N_{\rm H}|z)\,{\rm PDF}(z)\,{\rm d}z}{P(3 \le z < 6)}, \tag{10}$$

where $P(3 \le z < 6)$ is defined in equation (4). Note that equation (10) is normalized to unity, in order to compute confidence intervals and estimate uncertainties.

For each source, we report in Table 2 the $\log N_{\rm H}$ that maximizes equation (10) (i.e. our best estimate of the intrinsic column density) and the 68 per cent confidence level uncertainties, corresponding to the narrowest $\log N_{\rm H}$ interval in which the integrated $P(\log N_{\rm H})$ is 68 per cent. Fig. 5 presents the PDF(z) and $P(\log N_{\rm H})$ for a source as an example. The PDF of redshift, column density, flux (see Section 3.3) and luminosity (see Section 3.4) for every source in the sample are made available with the online version of this paper. Fig. 6 presents the best-fitting column density plotted against the nominal redshift of each source (i.e. the spectroscopic redshift or the redshift corresponding to the peak of PDF(z)).

We then derived the intrinsic distribution of column density of our sample as

$$N^{\rm tot}(\log N_{\rm H}) = \sum_{i=1}^{N_{\rm hz}} P^i(\log N_{\rm H})\,\Omega^i\,P^i(3 \le z < 6), \tag{11}$$

where

$$\Omega^i = \int_{{\rm CR}_{\rm lim}^{\rm intr}}^{\infty} \Omega({\rm CR})\,P^i({\rm CR})\,{\rm dCR}. \tag{12}$$

and the sum is performed over the $N_{\rm hz} = 118$ sources that contribute to the high-redshift sample. CR and ${\rm CR}_{\rm lim}^{\rm intr}$ are defined in the soft or hard bands for sources detected in the respective bands, as discussed in Section 3.1. The $P^i(3 \le z < 6)$ term in equation (11) keeps track of the actual probability of the ith source to be at high redshift.

Fig. 7 shows the binned $N^{\rm tot}(\log N_{\rm H})$. Errors have been computed through a bootstrap procedure: we created a list of $N_{\rm hz}$ sources randomly chosen amongst the $N_{\rm hz}$ sources in our sample, allowing for repetition (i.e. one source can appear in the new list 0, 1 or multiple times). We then recomputed $N^{\rm tot}(\log N_{\rm H})$. Repeating this procedure 1000 times, we derived 1000 distributions of $N^{\rm tot}(\log N_{\rm H})$ or, equivalently, at each $\log N_{\rm H}$ we derived 1000 values of probability. The 68 per cent confidence interval at each $\log N_{\rm H}$ has been computed by sorting the corresponding 1000 probabilities and selecting the 16th and 84th quantile values. Liu et al. (2017) presented an X-ray spectral analysis of the 7 Ms CDF-S sources with >80 net counts. Fig. 7 is consistent with the column-density distribution presented in that work in an overlapping redshift bin (z = 2.4–5), in spite of the different assumed spectral models and performed analysis. Also, the priority we assigned to the different photometric-redshift catalogues is different than that used by Liu et al. (2017), resulting in some sources having different redshifts.

The flattening at $\log N_{\rm H} \lesssim 23$ is due to the photoelectric cut-off detection limit. Indeed, the determination of the best-fitting $N_{\rm H}$ for an X-ray spectrum is largely driven by the detection of the photoelectric cut-off, which is a function of both redshift and column density. In particular, the photoelectric cut-off shifts to higher energies for increasing $N_{\rm H}$ and to lower energies for increasing redshift. Thus, at high redshift, unless a source is heavily obscured, the photoelectric cut-off is shifted outside the *Chandra* bandpass and, for small-to-moderate numbers of net counts, it is virtually impossible to determine the intrinsic column density. In Vito et al. (2013), we estimated the minimum column density that could be constrained

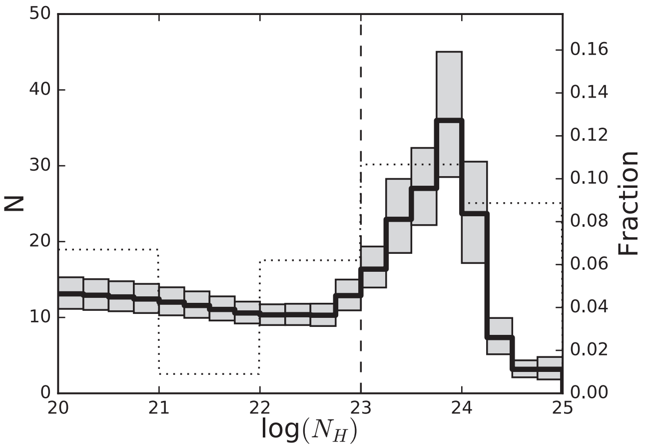

Figure 7. Binned intrinsic (i.e. weighted by the redshift probability distribution functions and sky-coverages associated with the individual sources) distribution of column density of our sample (equation 11). Uncertainties (at 68 per cent confidence level; grey regions) have been computed through a bootstrap procedure. The vertical dashed line represents the threshold used in this work to separate obscured and unobscured AGN at high redshift. The flatness of the distribution below log$N_H \approx 23$ is due to the lack of information in X-ray spectra at high redshift useful to constrain such low levels of obscuration, as discussed in Section 3.2. The dotted histogram is the column-density distribution assumed by Gilli et al. (2007), which will be used in Section 4.

for a typical (∼100 net counts) AGN in the 4 Ms CDF-S at $z \sim 3.5$ to be log$N_{\rm H} \approx 23$.

We stress that the flattening at log$N_{\rm H} \lesssim 23$, due to the lack of information in X-ray spectra useful to constrain such low values of obscuration at high redshift, should not be considered real. We can derive the probability of a source having log$N_{\rm H} \lesssim 23$, but not the exact probability distribution in column density bins below this threshold, resulting in a flat behaviour. According to these considerations, we will adopt log$N_{\rm H} = 23$ as the threshold between obscured and unobscured sources. This choice is further reinforced by the spectral simulations we performed in Appendix A.

3.3 Flux distribution and number counts

The observed flux in one band of a source at a given redshift and characterized by a given intrinsic column density is defined as

$$F_{\rm X}({\rm CR}, \log N_{\rm H}, z) = \frac{\rm CR}{f_{\rm X,CR}(\log N_{\rm H}, z)}, \tag{13}$$

where $f_{\rm X,\,CR}(\log N_{\rm H}, z)$, computed with XSPEC, is the conversion factor between count rate and flux, which depends on the response files associated with the source, and its observed-frame spectral shape (i.e. $N_{\rm H}$ and z, given that the photon index is fixed). *CR* is the count rate in the same energy band. According to the change of variables in equation (13) (e.g. Bishop 2006), we can derive the probability distribution density of flux from the known P(CR) (equation 5):

$$\begin{aligned} &P(F_{\rm X}({\rm CR}, \log N_{\rm H}, z)) \\ &\quad = P(F_{\rm X}({\rm CR}, \log N_{\rm H}, z) f_{\rm X,CR}(\log N_{\rm H}, z)) f_{\rm X,CR}(\log N_{\rm H}, z) \\ &\quad = P({\rm CR}) f_{\rm X,CR}(\log N_{\rm H}, z), \end{aligned} \tag{14}$$

where the probability density in the first term is defined in the flux space, while the probability densities in the second and third term are defined in the count rate space. The Jacobian of the transformation, $f_{\rm X,\,CR}(\log N_{\rm H}, z)$, conserves the integrated probability during the change of variable, i.e.

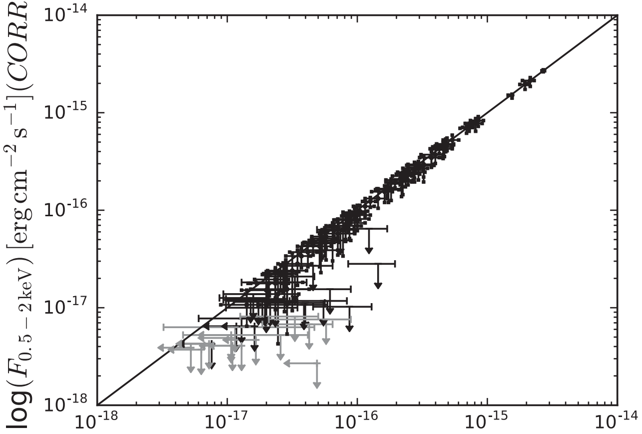

Figure 8. Best estimates of the soft-band flux obtained by integrating equation 16, with relative 68 per cent confidence level uncertainties, corrected for Eddington bias (see equation 5), compared to the uncorrected (i.e. observed) values. Grey symbols represent soft-band undetected sources, which are shown for completeness, for which the count rate (and therefore flux) in the hard band has been used everywhere else in this work.

$$\int_{F_{\rm X,1}}^{F_{\rm X,2}} P(F_{\rm X}){\rm d}F_{\rm X} = \int_{CR_1}^{CR_2} P({\rm CR}) f_{\rm X,CR}{\rm dCR}, \tag{15}$$

where $F_{\rm X,\,1}$ and $F_{\rm X,\,2}$ are linked to CR$_1$ and CR$_2$, respectively, through equation (13).

We can therefore define the probability density distribution in the $\{F_{\rm X}, \log N_{\rm H}, z\}$ parameter space as

$$P(F_{\rm X}, \log N_{\rm H}, z) = \frac{P(F_{\rm X}({\rm CR}, \log N_{\rm H}, z))\, P(\log N_{\rm H}|z)\, {\rm PDF}(z)}{P(3 \le z < 6)}. \tag{16}$$

Equation (16) includes the correction for Eddington-bias applied in equation (5) and $P(F_{\rm X}, \log N_{\rm H}, z)$ is normalized to unity. The flux probability distribution ($P(F_{\rm X})$) of each source is derived by integrating equation (16) over the considered redshift and column density ranges. Fig. 5 displays the flux probability distribution for a source as an example. Table 2 reports for each source the soft-band flux value that maximizes $P(F_{\rm X})$ and the 68 per cent confidence level uncertainties, corresponding to the narrowest interval containing 68 per cent of the total probability. If less than 68 per cent of P(CR) of a source lies above the count rate limit, we report the upper limit on $P(F_{\rm X})$. For consistency, Table 2 lists the soft-band flux also for sources detected in the hard-band only, for which, however, the hard-band flux has been used to derive the luminosity in Section 3.4. Fig. 8 presents the $F_{0.5\text{–}2\,{\rm keV}}$ estimated by applying and not applying the Eddington-bias correction (i.e. the last factor in equation 5), which, as expected, causes the slight bend at faint fluxes in Fig. 8.

Considering only soft-band detected sources, we derived the intrinsic soft-band flux distribution (which includes the sky-coverage correction) as

$$N^{\rm tot}(F_{0.5\text{–}2\,{\rm keV}}) = \sum_{i=1}^{N_{\rm hz}^{\rm sb}} [P^i(3 \le z < 6)$$

$$\int_{z=3}^{6} \int_{\log N_{\rm H}=20}^{25} P^i(F_{\rm X}, \log N_{\rm H}, z)\, \Omega(F_{\rm X}, \log N_{\rm H}, z)\, {\rm d}\log N_{\rm H}\, {\rm d}z], \tag{17}$$

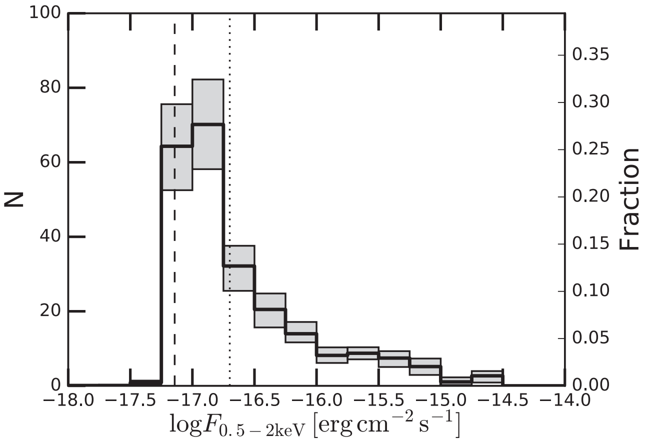

Figure 9. Binned intrinsic (i.e. weighted by the redshift probability distribution functions and sky-coverages associated with the individual sources) distribution of flux for our subsample of soft-band detected AGN (equation 17) and relative uncertainties (at 68 per cent confidence level; grey bars). The dashed and dotted vertical lines represent the flux limits (assuming $\Gamma = 1.4$) of the CDF-S (corresponding to a sensitive area $A = 10\,\mathrm{arcmin}^2$, i.e. dashed line in Fig. 4) and the CDF-N, respectively.

where $N^{\rm sb}_{\rm hz} = 108$ is the number of soft-band detected sources contributing to the high-redshift sample (see Section 3.1). In equation (17), the weighting factor is expressed as a function of the integrating variables, following standard rules for a change of variables, as

$$\begin{aligned}\Omega(F_{\rm X}, \log N_{\rm H}, z) &= \Omega(F_{\rm X}({\rm CR}, \log N_{\rm H}, z)\, f_{\rm X,CR}(\log N_{\rm H}, z))\\ &= \Omega({\rm CR}).\end{aligned} \qquad (18)$$

This change of variables allows us to account for the specific spectral shape of each source, as discussed in Section 3.1, since the same observed flux can correspond to different count rates for different values of N_H and z, which, according to our simple spectral model, describe completely the spectral shape.

Fig. 9 shows the binned $N^{\rm tot}(F_{0.5-2\,\rm keV})$ with 68 per cent confidence intervals derived with a bootstrapping procedure, similarly to what was done in Section 3.2 and Fig. 7. We note that, as we remained in the count rate space, the distributions extend slightly below the nominal flux-limit, since the assumed count rate limit corresponds to slightly different fluxes for different observed spectral shapes and response matrices (see Section 3.1).

3.3.1 AGN logN-logS at $3 \le z < 6$

The cumulative version of equation (17), divided by the total area in deg^2, is the log*N*-log*S* relation of our sample. The 68 per cent confidence region derived through the bootstrap procedure is compared with results from previous works at $z > 3$ and $z > 4$ in Fig. 10. We derived the $z > 4$ log*N*-log*S* by repeating the procedures described in this section, using $z = 4$ as the lower limit for the considered redshift range. Bright and rare sources are not sampled properly by the pencil-beam, deep surveys such as those used in this work. Therefore, Fig. 10 displays the curves up to the flux at which the median expected number of sources from the bootstrap analysis is 1 in the area considered in this work ($\sim$550 arcmin^2, see Fig. 4, corresponding to $\sim$6.5 deg^{-2} sources).

Our results are in good agreement with previous measurements at $\log F_{0.5-2\,\rm keV} \gtrsim -16$, which is the flux regime better sampled by wide surveys. At fainter fluxes, where we can exploit the excellent sensitivity of the *Chandra* deep fields, our results are consistent with previous results that used data from the 4 Ms CDF-S (Lehmer et al. 2012; Vito et al. 2014) and with the Ueda et al. (2014) curve. Lehmer et al. (2012) presented the number counts derived in the 4 Ms CDF-S down to slightly fainter fluxes than those reached in this work. This is due to the less-conservative approach they used to compute the sensitivity of the 4 Ms CDF-S compared to that used by L17 for the 7 Ms CDF-S, which we adopted here.

At $z > 4$, we can push the AGN number counts down to $\log F_{0.5-2\,\rm keV} \approx -17$. While our curve agrees with previous results at bright fluxes and with the Gilli, Comastri & Hasinger (2007) log*N*-log*S* reasonably well, it excludes very steep number counts such as those reported by Fiore et al. (2012) at faint fluxes. The selection in Fiore et al. (2012) made use of pre-determined optical

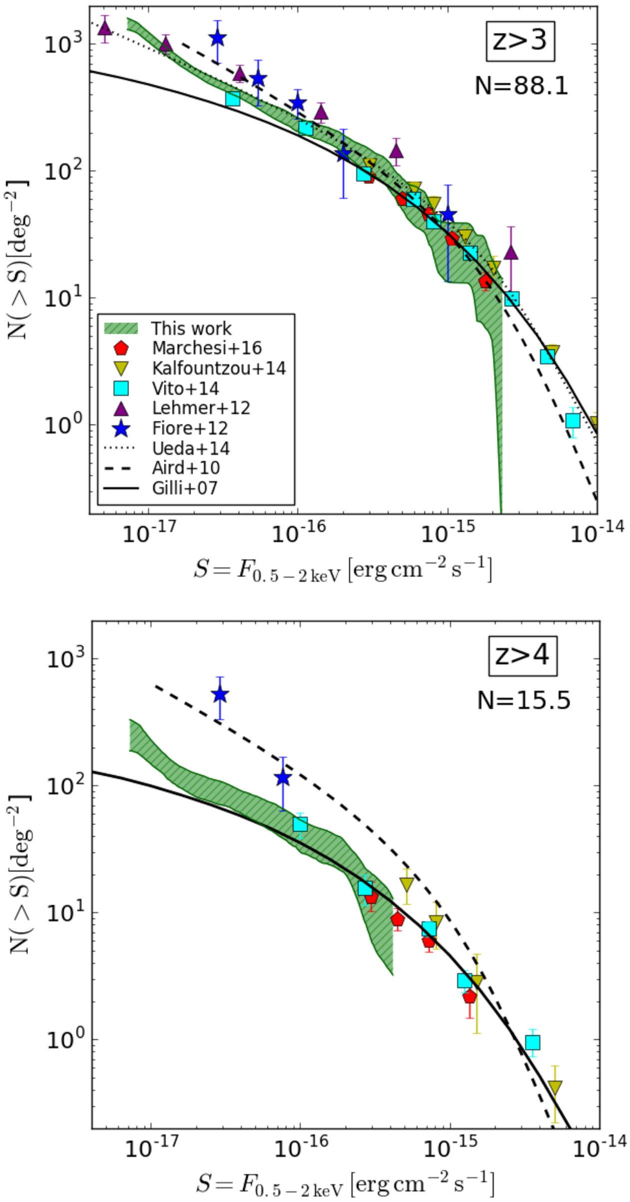

Figure 10. Confidence region (at the 68 per cent level) of the log*N*-log*S* of our subsample of soft-band detected, $z > 3$ (top panel) and $z > 4$ (bottom panel) AGN, compared with results from previous works. The numbers of soft-band-detected sources, weighted by the PDF(z), included in these figures are also reported.

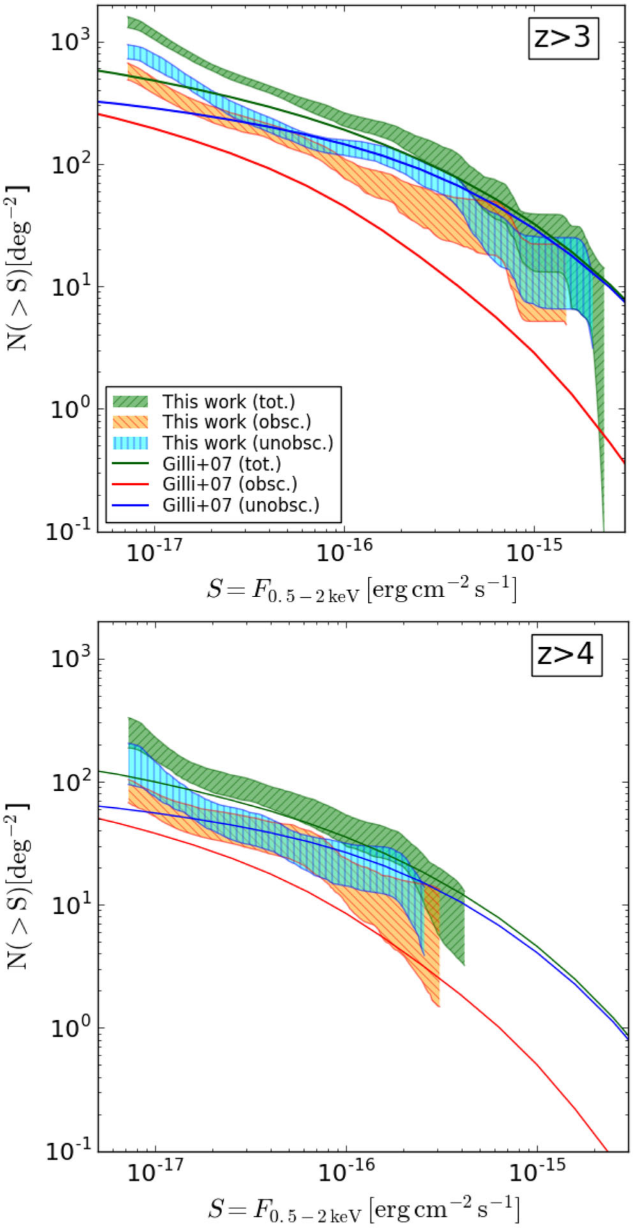

Figure 11. Confidence region (at the 68 per cent level) of the log*N*-log*S* of our subsample of soft-band detected, $z > 3$ (top panel) and $z > 4$ (bottom panel) AGN, for the total sample (green stripes), obscured (orange stripes) and unobscured (cyan stripes) AGN. The results are compared to the expectations of the Gilli et al. (2007) X-ray background synthesis model, using the same N_{H} threshold to define the obscuration-based subsamples.

positions of galaxies and adaptive X-ray detection bands, and it is therefore less conservative than a blind X-ray detection. However, about half of the 4 Ms CDF-S sources reported in that work at $z > 4$ were not detected even in the deeper 7 Ms exposures, leaving doubts about their detection significance (especially considering that the Fiore et al. (2012) analysis is plausibly affected by Eddington bias). This issue, together with the different, more recent photometric redshifts we used, which shift some of the Fiore et al. (2012) to lower redshifts, accounts completely for the discrepancies in Fig. 10.

By changing the limits of the integral of column density in equation (17), we derived separately the log*N*-log*S* for obscured ($23 \le \log N_{\mathrm{H}} < 25$) and unobscured ($20 \le \log N_{\mathrm{H}} < 23$) AGN (Fig. 11). The Gilli et al. (2007) X-ray background synthesis model underestimates the number of obscured AGN at high redshift and the steepening of our number counts of unobscured sources at $\log F_{0.5\text{–}2\,\mathrm{keV}} < -16.5$ (see Fig. 11). Such steepening could be due to the population of star-forming galaxies, which begin providing a significant contribution to the number counts at the faintest fluxes probed by this work (e.g. Lehmer et al. 2012).

3.4 Luminosity distribution

The intrinsic (i.e. absorption-corrected) rest-frame 2–10 keV luminosity (L_{X}) can be derived from the observed flux as

$$L_{\mathrm{X}}(F_{\mathrm{X}}, \log N_{\mathrm{H}}, z) = \frac{F_{\mathrm{X}}}{l_{\mathrm{X}}(\log N_{\mathrm{H}}, z)} \tag{19}$$

where the $l_{\mathrm{X}}(\log N_{\mathrm{H}}, z)$ factor, computed with XSPEC for each source, depends on the observed-frame spectral shape (i.e. N_{H} and z, since the photon index is fixed). F_{X} is the flux in the soft or hard band, as derived in Section 3.3, for sources detected in the soft band or only in the hard band, respectively, as marked in column 5 of Table 2. The l_{X} factors are computed in both bands and applied accordingly for each source.

Similarly to Section 3.3, we derived for each source the hard-band intrinsic luminosity probability density distribution as

$$P(L_{\mathrm{X}}, \log N_{\mathrm{H}}, z) = \frac{P(L_{\mathrm{X}}(F_{\mathrm{X}}, \log N_{\mathrm{H}}, z))}{P^{i}(3 \le z < 6)}, \tag{20}$$

where

$$\begin{aligned} P(L_{\mathrm{X}}(F_{\mathrm{X}}, \log N_{\mathrm{H}}, z)) &= P(L_{\mathrm{X}}(F_{\mathrm{X}}, \log N_{\mathrm{H}}, z)\, l_{\mathrm{X}}(\log N_{\mathrm{H}}, z)) \\ &\quad \times l_{\mathrm{X}}(\log N_{\mathrm{H}}, z). \end{aligned} \tag{21}$$

$P(L_{\mathrm{X}}, \log N_{\mathrm{H}}, z)$ is normalized to unity. Equation (20) includes the correction for Eddington-bias applied in equation (5). However, we checked that neglecting this correction results in a similar luminosity distribution, with a slight decrease in the number of $\log L_{X} < 42.5$ sources, balanced by an increase at higher luminosities, as expected. The luminosity probability distribution ($P(L_{\mathrm{X}})$) of each source is derived by integrating equation (20) over the considered redshift and column-density ranges. Fig. 5 presents $P(L_{\mathrm{X}})$ for a source as an example. Table 2 reports for each source the luminosity value that maximizes $P(L_{\mathrm{X}})$ (i.e. our best estimate of the luminosity of each source) and the 68 per cent confidence level uncertainties, corresponding to the narrowest interval containing 68 per cent of the total probability. If less than 68 per cent of $P(\mathrm{CR})$ of a source lies above the count rate limit, we report the upper limit of $P(L_{\mathrm{X}})$.

We derived the luminosity distribution of our sample as

$$\begin{aligned} N^{\mathrm{tot}}(L_{\mathrm{X}}) = \sum_{i=1}^{N_{\mathrm{hz}}} \int_{z=3}^{6} \int_{\log N_{\mathrm{H}}=20}^{25} & P^{i}(L_{\mathrm{X}}, \log N_{\mathrm{H}}, z) \\ & \times \Omega(L_{\mathrm{X}}, \log N_{\mathrm{H}}, z)\, \mathrm{d}\log N_{\mathrm{H}}\, \mathrm{d}z \end{aligned} \tag{22}$$

where, similar to equation (18),

$$\begin{aligned} \Omega(L_{\mathrm{X}}, \log N_{\mathrm{H}}, z) &= \Omega(L_{\mathrm{X}}(F_{\mathrm{X}}, \log N_{\mathrm{H}}, z)\, l_{\mathrm{X}}(\log N_{\mathrm{H}}, z)) \\ &= \Omega(F_{\mathrm{X}}, \log N_{\mathrm{H}}, z) = \Omega(\mathrm{CR}). \end{aligned} \tag{23}$$

Fig. 12 shows the binned $N^{\mathrm{tot}}(L_{\mathrm{X}})$ with 68 per cent confidence intervals derived from a bootstrapping procedure, similar to what was done in Section 3.2 and Fig. 7.

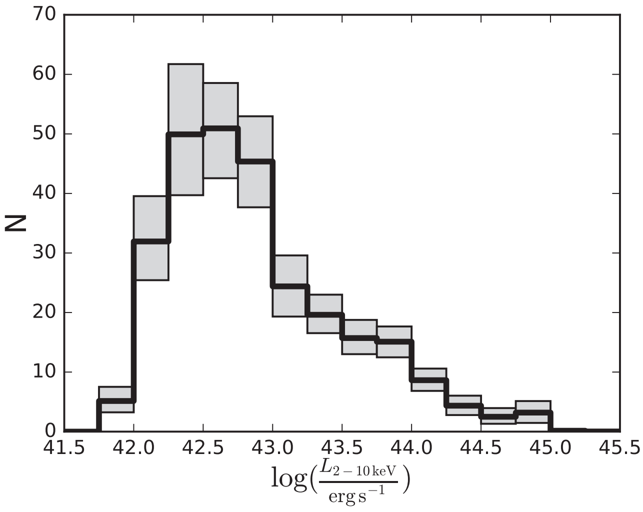

Figure 12. Binned intrinsic (i.e. weighted by the redshift probability distribution functions and sky-coverages associated with the individual sources) distribution of hard-band luminosity of our sample (equation 22) and relative uncertainties (at 68 per cent confidence level, grey bars).

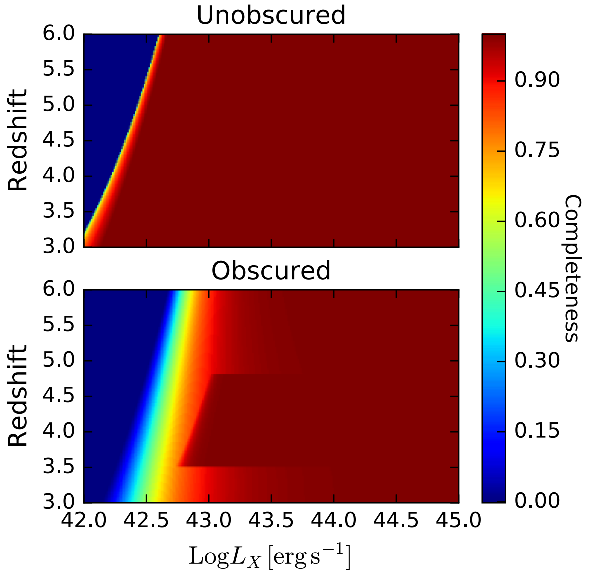

Figure 13. Detection completeness at the 7 Ms CDF-S depth as a function of redshift and luminosity for obscured (lower panel) and unobscured (upper panel) sources. An intrinsic column density distribution similar to Fig. 7 is assumed, as described in Section 4, where the dark thick stripe at $\log L_\mathrm{X} \sim 43$–$43.5$ and $z \sim 3.5$–4.5 is also discussed.

4 OBSCURATION-DEPENDENT INCOMPLETENESS

The flux (or count rate) limit of a survey is effectively a limit in luminosity, redshift, and column density, as these are the physical parameters defining the flux (assuming our simple spectral model). Therefore, a low-luminosity AGN at a given redshift can be detected only if its column density is below a certain level, causing an obscuration-dependent incompleteness that affects the low-luminosity regime by favouring the detection of unobscured AGN (see also section 5.1.2 of Liu et al. 2017). Such incompleteness usually appears in X-ray studies of flux-limited AGN samples as a lack of detected sources in the low-luminosity and heavy-obscuration region of the $L_X - N_H$ parameter space.

It is worth noting that this effect is not corrected by the Ω(CR) factor, which compensates for the different areas sensitive to different observed fluxes for *detected* sources, irrespective of their intrinsic properties (e.g. obscuration). In fact, while representing a potentially significant fraction of the population, low-luminosity and heavily obscured AGN are not detected at all, and are therefore not accounted for when deriving intrinsic properties (e.g. space density) of the low-luminosity population. This effect is particularly important when two obscuration-based subsamples are defined to compute the obscured AGN fraction, F_obsc, which would be biased towards lower values, and its trends with redshift and luminosity (e.g. Gilli et al. 2010). Therefore, before deriving F_obsc as a function of redshift and luminosity in Section 5, in this section we assess the effect of the obscuration-dependent incompleteness and derive suitable corrections.

In order to evaluate this bias, we computed the observed count rate (in both the soft and hard bands) corresponding to the parameters (z, L_X, N_H):

$$\mathrm{CR}(z, L_\mathrm{X}, N_\mathrm{H}) = L_\mathrm{X}\, l_\mathrm{X}(z, N_\mathrm{H})\, f_\mathrm{X,CR}(z, N_\mathrm{H}), \tag{24}$$

where the factors $f_\mathrm{X,\,CR}$ and l_X are defined in Sections 3.3 and 3.4, respectively. We then assigned a value of 1 (detection) or 0 (non-detection) to this triplet of coordinates if the resulting count rate is larger or smaller, respectively, than the count rate limit in the 7 Ms CDF-S in at least one of the two considered bands. By comparing the observed count rate with the count rate limit of the survey, we are assuming that all the sensitivity dependence on position is properly taken into account by the Ω factor.

The completeness level in the z–L_X plane is shown in Fig. 13 by computing the mean over the N_H axis separately for $\log N_\mathrm{H} = 23$–25 and $\log N_\mathrm{H} = 20$–23, weighted over an assumed intrinsic column-density distribution $P(\log N_\mathrm{H})$, which we here considered to be represented by Fig. 7. At $\log L_\mathrm{X} \lesssim 43$, the two classes of objects are characterized by very different completeness levels: in the upper panel of Fig. 13, the detection of unobscured sources is complete up to $z \sim 6$ for $\log L_\mathrm{X} \gtrsim 42.5$, while lower-luminosity AGN can be detected only at lower redshifts. Since the photoelectric cut-off for $\log N_\mathrm{H} \lesssim 23$ shifts close to or below the lower-limit of the *Chandra* bandpass, for unobscured sources the completeness is almost independent of the column-density distribution. For this reason, the transition between 0 per cent and 100 per cent completeness is sharp, as it depends only on redshift and luminosity.[8] The lower panel shows that for $\log N_\mathrm{H} = 23$–25 at a given redshift the transition is smoother and occurs at larger luminosities. In fact, contrary to the case of unobscured sources, the flux of an obscured source depends strongly on the particular N_H value: more heavily obscured sources can be detected only at higher luminosities than sources affected by milder obscuration. The darker stripe at $\log L_\mathrm{X} \sim 43$–$43.5$ and $z \sim 3.5$–4.5 is due to the inclusion of the hard band; heavily-obscured sources in those redshift and luminosity intervals are more easily detected in the hard band than in the soft band.

By projecting Fig. 13 over the luminosity axis (i.e. averaging over the redshift range), we derived the completeness curve as a function of luminosity (solid lines in Fig. 14). The averaging is weighted assuming an intrinsic redshift distribution characterized

---

[8] This result again confirms that at $z > 3$ column densities of $\log N_\mathrm{H} \lesssim 23$ do not sensibly affect X-ray spectra, and therefore cannot be correctly identified. As a consequence, a column density threshold larger than the usual $\log N_\mathrm{H} = 22$ value must be used to define obscured AGN.

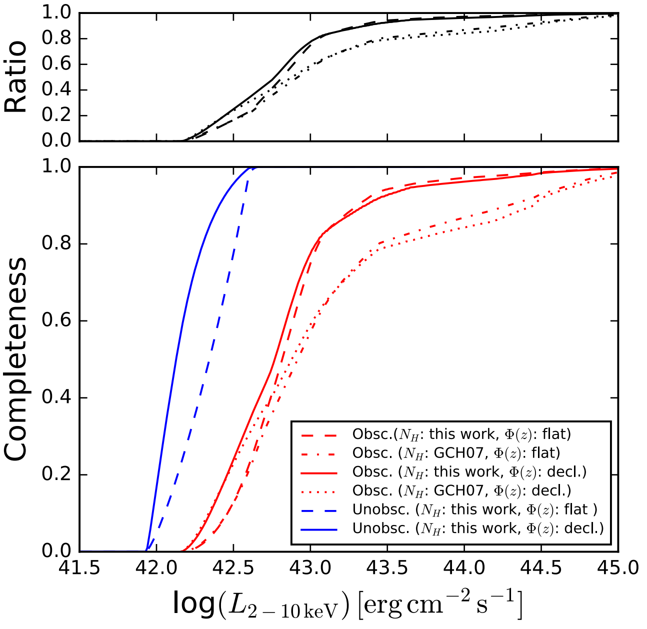

Figure 14. Detection completeness at the 7 Ms CDF-S depth as a function of luminosity for obscured (red lines) and unobscured (blue lines) sources, assuming different intrinsic distributions of column density (the one presented in this work in Fig. 7 and the one used in Gilli et al. 2007) and redshift (flat or more realistically declining with redshift as $\propto 10^{-6(1+z)}$; e.g. Hiroi et al. 2012; Vito et al. 2013), as described in Section 4. For the unobscured case, the curves corresponding to the two column-density distributions are almost indistinguishable and therefore are not plotted for clarity. The solid lines have been used to derive the correction factors applied to Figs 16 and 18. The upper panel presents the ratio of the completeness of obscured and unobscured sources.

by a decline in the space density of high-redshift AGN proportional to $(1+z)^{-6}$ (e.g. Hiroi et al. 2012; Vito et al. 2014, solid lines). The unweighted projection corresponds to a flat distribution in redshift and is shown for completeness as dashed lines in Fig. 14. The specific redshift distribution has a small effect on the displayed curves: this behaviour is due to the negative curvature of absorbed X-ray spectra, which causes inverted K-corrections at increasing redshift.

To assess the dependence on the particular choice of column-density distribution, we also assumed the one from Gilli et al. (2007), in which, in particular, Compton-thick AGN are more numerous (see Fig. 7). While the completeness curve for unobscured sources is not sensitive to the particular choice of $P(\log N_{\mathrm{H}})$ (we do not show the curve corresponding to the Gilli et al. 2007 case for clarity), the completeness curve for obscured AGN is more severely affected by the Gilli et al. (2007) $P(\log N_{\mathrm{H}})$.

The bias affecting the obscured AGN fraction is linked to the relative completeness levels of the obscured and unobscured subsamples as a function of redshift and luminosity. To visualize this effect, the upper panel in Fig. 14 shows the ratio of the completeness characterizing obscured and unobscured subsamples as a function of luminosity. The ratio varies rapidly from $\log L_{\mathrm{X}} = 42$–43, rising from zero to ≈ 0.8. This is the regime in which the incompleteness effects are strongest: at low intrinsic luminosities, the presence or absence of even moderate obscuration is crucial for observing a flux above or below the survey sensitivity limit. At higher luminosities, only the most-obscured systems cannot be detected, and incompleteness is less severe.

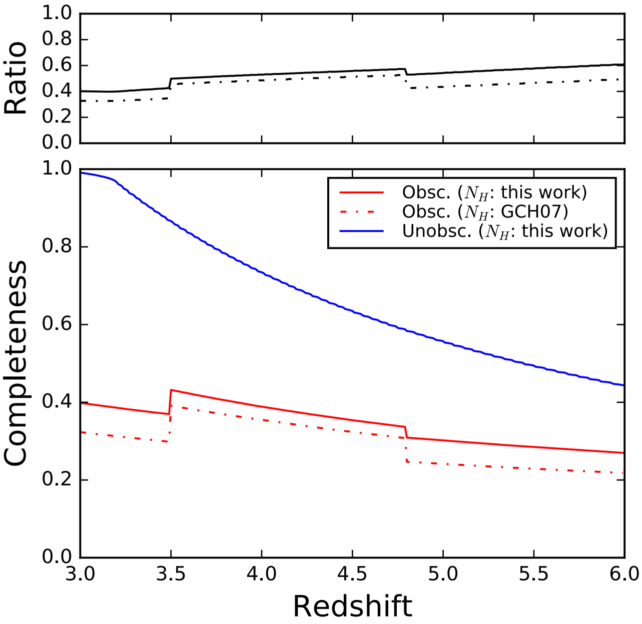

Figure 15. Detection completeness at the 7 Ms CDF-S depth as a function of redshift for obscured (red lines) and unobscured (blue line) sources, assuming different intrinsic distributions of column density and the luminosity function of Vito et al. (2014). The solid line has been used to derive the correction factors applied to Figs 16 and 21. The upper panel shows the ratio of the completeness of obscured and unobscured sources.

Similarly, we derived the completeness as a function of redshift (Fig. 15) by projecting Fig. 13 over the redshift axis and assuming an intrinsic luminosity distribution (i.e. luminosity function). We used the pure density evolution model of Vito et al. (2014). The upper panel of Fig. 15 shows that the relative strength of incompleteness for obscured and unobscured sources does not significantly vary with redshift, but the evolution is slightly stronger for unobscured sources: this behaviour is due to the inverse *K*-correction characterizing obscured X-ray sources, which helps the detection of these systems at higher redshifts.

5 OBSCURED AGN FRACTION

In this section, we derive the obscured AGN fraction F_{obsc} as a function of redshift, flux and luminosity for our sample of high-redshift AGN. We define the obscured AGN fraction as a function of a parameter *x*, where *x* is redshift, flux or luminosity in the next sections, as

$$F_{\mathrm{obsc}}(x) = \frac{N_{\mathrm{obsc}}(x)}{N_{\mathrm{tot}}(x)}, \tag{25}$$

where $N_{\mathrm{tot}}(x) = N_{\mathrm{obsc}}(x) + N_{\mathrm{unobs}}(x)$ is the total number of observed sources, and $N_{\mathrm{obsc}}(x)$ and $N_{\mathrm{unobs}}(x)$ are the numbers of obscured and unobscured sources, respectively, as a function of *x*. Since all of these are intrinsic numbers, as shown in the following sections, equation (25) is equivalent (modulo a volume term) to computing the ratio of space densities of obscured and unobscured AGN, as is usually done to derive F_{obsc} (e.g. Buchner et al. 2015; Aird et al. 2015; Georgakakis et al. 2015).

As discussed in Section 3.2 and following Vito et al. (2013, 2014), throughout this work we used $\log N_{\mathrm{H}} = 23$ as the column-density threshold dividing obscured and unobscured AGN at $z > 3$. This value is higher than the commonly adopted threshold of $\log N_{\mathrm{H}} = 22$ in X-ray studies of the AGN population (e.g. Ueda et al. 2014; Aird

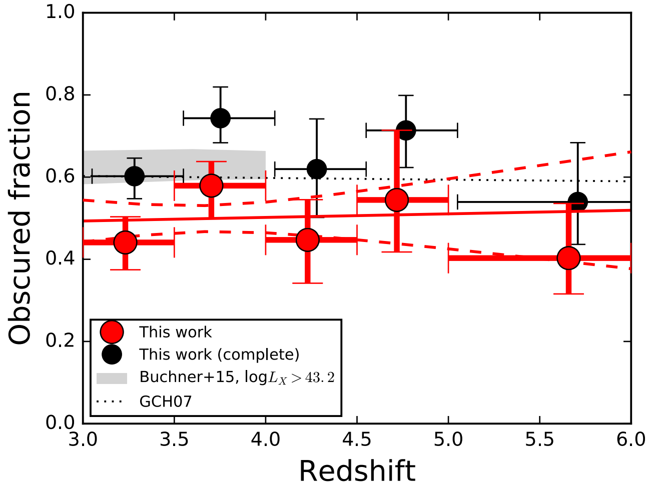

Figure 16. Binned estimates of the obscured AGN fraction as a function of redshift ($F_{\rm obsc}(z)$; red circles). The red solid and dashed lines show the best-fitting linear model and 68 per cent confidence level region, respectively. The continuous line is not the fit to the points: line and points are different estimates of the obscured AGN fraction as a function of redshift derived using different (parametric and non-parametric) methods. Grey regions encompass the $F_{\rm obsc}(z)$ from Buchner et al. (2015) for luminous AGN. Black circles, slightly shifted for visual purposes, indicate the non-parametric estimate of $F_{\rm obsc}(z)$, where a correction for the obscuration-dependent incompleteness has been applied (see Section 4). The dotted line is the predictions from the X-ray background synthesis model of Gilli et al. (2007).

et al. 2015; Buchner et al. 2015), which is also more in agreement with optical classification (e.g. Merloni et al. 2014). However, such low levels of obscuration are extremely difficult to detect at high redshift, where the photoelectric cut-off in X-ray spectra shifts below the energy band probed by X-ray observatories. The redshifted cut-off, together with the typical limited spectral quality of high-redshift sources, leads to a general overestimate of obscuration for low values of $N_{\rm H}$, as can be seen in Appendix A. This effect is especially important when considering the obscured AGN fraction and its trends with other quantities, such as redshift and luminosity. We therefore prefer to define $\log N_{\rm H} = 23$, as the minimum column density of (heavily) obscured AGN at high redshift and discuss the dependence of (heavily) obscured AGN on redshift and luminosity. Unfortunately, this choice complicates the comparison with previous works utilizing more standard definitions. In Appendix C, we show that neglecting the full probability distributions of redshift and spectral parameters would lead to different results than those presented in the following subsections.

5.1 Obscured AGN fraction versus redshift

For each source i in the high-redshift sample, we define

$$N_{\rm obsc}(z) = \sum_{i=1}^{N_{\rm hz}} \int_{23}^{25} P_i(\log N_{\rm H}|z)\,{\rm PDF}_i(z)\,\Omega^i\,{\rm d}\log N_{\rm H}$$

$$N_{\rm unobs}(z) = \sum_{i=1}^{N_{\rm hz}} \int_{20}^{23} P_i(\log N_{\rm H}|z)\,{\rm PDF}_i(z)\,\Omega^i\,{\rm d}\log N_{\rm H}. \qquad (26)$$

Red circles in Fig. 16 indicate the obscured AGN fraction as a function of z in five redshift bins with errors computed through a bootstrapping procedure. We also report the results by Buchner et al. (2015), derived from their space densities of luminous ($\log L_{\rm X} > 43.2$) AGN with column densities $\log N_{\rm H} = 20$–23 and $\log N_{\rm H} = 23$–26, and show the predictions of the X-ray background synthesis model of Gilli et al. (2007).

To derive an independent and parametric estimate of the dependency of $F_{\rm obsc}$ on redshift, we performed a Bayesian analysis[9] of $F_{\rm obsc}(z)$, assuming a linear model of the form $F^{\rm m}_{\rm obsc} = q_z + m_z(z - 3)$. We applied a flat prior to q_z and a minimally informative prior to m_z, such that $P(m_z) \propto (1 + m_z^2)^{3/2}$ (e.g. VanderPlas 2014 and references therein). The latter is preferred over a flat prior on β, which would weight steep slopes more.[10] We also associated a null probability to parameter pairs resulting in unphysical $F_{\rm obsc}$ values (i.e. negative or larger than unity) in the redshift interval probed by the data.

By construction, our sources can be assigned to one of the two obscuration-defined classes (obscured or unobscured AGN) only with a certain probability, since a probability distribution of $N_{\rm H}$ is associated with each source. The outcome of this approach is that the number of sources in the two classes (i.e. the number of 'successes' and 'failures') is fractional (see equation 26). Therefore, the use of the Binomial distribution for describing the data likelihood in this case is not formally correct. However, we can use the Beta distribution

$$B(x; \alpha, \beta) \propto x^{\alpha-1}(1 - x)^{\beta-1} \qquad (27)$$

as a formally correct expression of the data likelihood in one redshift bin, with $\alpha - 1 = N_{\rm obsc}(z)$, $\beta - 1 = N_{\rm unobs}(z)$, and $x = F^{\rm m}_{\rm obsc}(z)$.[11] The total likelihood of the data under a particular set of model parameters is therefore derived multiplying equation (27) over the entire redshift range.

The resulting best-fitting values are $q_z = 0.50^{+0.05}_{-0.05}$ and $m_z = 0.00^{+0.06}_{-0.06}$. The red dashed lines in Fig. 16 enclose the 68 per cent confidence interval of 10^6 MCMC realizations derived using this parametric approach, and the solid line is the median value. The parametric and non-parametric representations, derived independently, agree well, showing the robustness of our approach. The observed obscured AGN fraction is found in Fig. 16 to be flat at $z = 3$–5 at $F_{\rm obsc} \approx 0.5$. $F_{\rm obsc}$ is usually found to increase with redshift up to $z \approx 2$–2.5, and then saturates at higher redshifts (e.g. Treister & Urry 2006; Hasinger 2008; Iwasawa et al. 2012; Buchner et al. 2015; Liu et al. 2017). We probe such behaviour up to $z \sim 6$. Black circles in Fig. 16 are derived by correcting the red points using the solid lines in Fig. 15, causing $F_{\rm obsc}(z)$ generally to increase slightly, with no strong dependence on redshift, to values ≈0.65, consistent with the results by Buchner et al. (2015) at high luminosities ($\log L_{\rm X} > 43.2$) up to $z \sim 4$. We could push the investigation of $F_{\rm obsc}(z)$ down to lower luminosities.

5.2 Obscured AGN fraction versus soft-band flux

Considering only the soft-band detected sources, we defined $N_{\rm unobs}(F_{\rm X})$ and $N_{\rm obsc}(F_{\rm X})$ from equation (17) by limiting the

---

[9] We used the EMCEE PYTHON package (http://dan.iel.fm/emcee/current/), which implements the Affine Invariant Markov chain Monte Carlo (MCMC) Ensemble sampler by Goodman & Weare (2010).

[10] This effect can be visualized considering that, while β can be any real number, half of the plane is covered by $-1 < \beta < 1$. Giving equal weight to values in and out of this range would in principle preferentially select $|\beta| > 1$, hence steep slopes. However, we checked that using a flat prior on m_z does not affect significantly the results.

[11] An assessment of the covariance between z and L_X is discussed in Section 5.3.

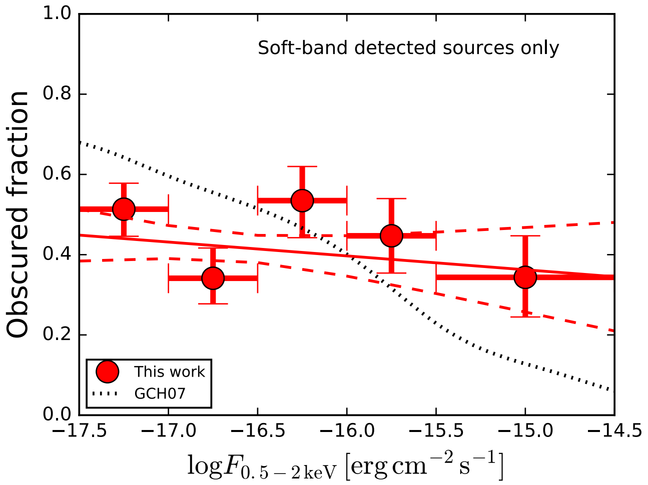

Figure 17. Binned estimates of the obscured, soft-band detected AGN fraction as function of flux (red circles). Red continuous and dashed lines are the best-fitting linear model and 68 per cent confidence level region, respectively. The prediction of the Gilli et al. (2007) X-ray background synthesis model are indicated by the dotted black line.

integral over column density to the ranges $\log N_{\mathrm{H}} = 20$–23 and 23–25, respectively, and summing only the soft-band-detected sources.

Similar to the process described in Section 5.1, we derived the parametric (assuming a linear model $F_{\mathrm{obsc}} = q_{F_{\mathrm{X}}} + m_{F_{\mathrm{X}}}(\log F_{\mathrm{X}} - (-17))$ and non-parametric obscured AGN fraction as a function of flux. The results are shown in Fig. 17 and compared with the predictions of the Gilli et al. (2007) X-ray background synthesis model. The resulting best-fitting values are $q_{F_{\mathrm{X}}} = 0.43^{+0.04}_{-0.04}$ and $m_{F_{\mathrm{X}}} = -0.04^{+0.06}_{-0.06}$.

5.3 Obscured AGN fraction versus luminosity

We defined $N_{\mathrm{unobs}}(L_{\mathrm{X}})$ and $N_{\mathrm{obsc}}(L_{\mathrm{X}})$ from equation (22) by limiting the integral over column density to the ranges $\log N_{\mathrm{H}} = 20$–23 and 23–25, respectively. Fig. 18 presents the obscured AGN fraction as a function of intrinsic, 2–10 keV luminosity with errors computed with a bootstrapping procedure (red symbols). As this non-parametric description cannot be well parametrized by a simple linear model; in this section, we allowed the slope to change above a characteristic luminosity, i.e.:

$$F_{\mathrm{obsc}}(L_{\mathrm{X}}) = \begin{cases} q_L + m_L(\log L - \log L_*) & \text{if } \log L \le \log L_* \\ q_L + n_L(\log L - \log L_*) & \text{if } \log L > \log L_*, \end{cases} \tag{28}$$

where all the parameters (q_L, m_L, n_L and $\log L_*$) are allowed to vary. The best-fitting model derived with the Bayesian analysis is shown in Fig. 18. The most noticeable feature is the drop of $F_{\mathrm{obsc}}(L_{\mathrm{X}})$ at low luminosities. However, this trend can be ascribed to a selection effect described in detail in Section 4, where tentative corrections are also provided. The best-fitting values describing the relation at high-luminosities are $q_L = 0.79^{+0.06}_{-0.06}$, $m_L = -0.06^{+0.08}_{-0.09}$ and $\log L_* = 42.89^{+0.09}_{-0.09}$.

We compare the high fraction of heavily obscured AGN ($F_{\mathrm{obsc}} \approx 0.7$–0.8) we derived at high luminosities ($\log L_{\mathrm{X}} > 43$) with previous findings by Aird et al. (2015) and Buchner et al. (2015), where obscured AGN are defined in a more ‘standard’ way as those characterized by $\log N_{\mathrm{H}} = 22$–24. The column-density threshold used to separate obscured and unobscured AGN in their works, $\log N_{\mathrm{H}} = 22$, while useful for comparison purposes with results at lower redshift, is not suitable at high redshift, where the photoelectric cut-off shifts at low energies, close to or even below the lower energy boundary of the *Chandra* bandpass and is therefore poorly constrained, as discussed in Section 7 and Appendix A. For completeness, we also show the results of Georgakakis et al. (2015), where the obscured

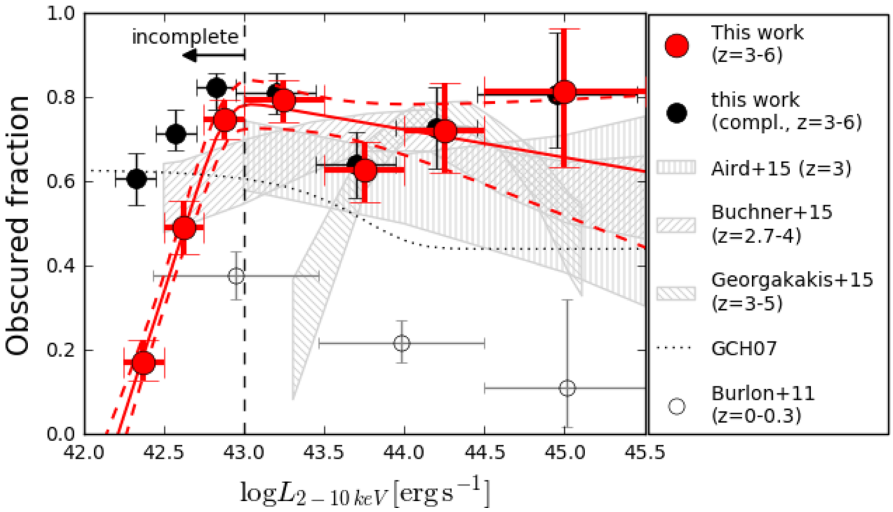

Figure 18. Binned estimates of the obscured AGN fraction as a function of intrinsic hard-band luminosity ($F_{\mathrm{obsc}}(L_{\mathrm{X}})$, red circles). Solid and dashed lines represent the best-fitting model (as defined in equation 28) and 68 per cent confidence level region, respectively, as derived through a Bayesian analysis. Grey-dashed regions encompass the $F_{\mathrm{obsc}}(L_{\mathrm{X}})$ derived by Aird et al. (2015), Buchner et al. (2015) and Georgakakis et al. (2015), as labelled in the figure. The curves are plotted for slightly different redshift ranges, as reported in the legend. Obscured AGN are defined as those with $\log N_{\mathrm{H}} = 22$–24 in Aird et al. (2015) and Buchner et al. (2015), while we require $\log N_{\mathrm{H}} > 23$. Georgakakis et al. (2015) derived their obscured AGN fraction by comparing X-ray and UV luminosity functions. The dotted black line indicates the prediction of the X-ray background synthesis model of Gilli et al. (2007). We also plot the results derived by Burlon et al. (2011) in the local Universe, as grey empty circles to show the strong evolution of F_{obsc} from low to high redshift. The Burlon et al. (2011) points use the same column-density threshold, as we assumed to define obscured and unobscured sources, and are therefore directly comparable with our points. Black circles (slightly shifted for visual purposes) show the non-parametric estimate of $F_{\mathrm{obsc}}(L_{\mathrm{X}})$, where a correction for the obscuration-dependent incompleteness has been applied (see Section 4).

AGN fraction is derived by comparing X-ray and UV luminosity functions, and therefore is not directly comparable to the other purely X-ray defined curves.

While at $\log L_{\rm X} > 43$ incompleteness effects are negligible, they dominate the observed trend of $F_{\rm obsc}$ at lower luminosities in Fig. 18. Applying the corrections discussed in Section 4 results in values, represented by the black circles in Fig. 18, consistent with those at $\log L_{\rm X} > 43$ (i.e. $F_{\rm obsc} \approx 0.7$–0.8), although at $\log L_{\rm X} < 43$ a slightly decreasing trend of $F_{\rm obsc}$, down to ≈ 0.6, with decreasing luminosity remains visible. However, we caution that the applied corrections depend quite strongly on the particular intrinsic column density and luminosity distributions. Also, the possible detection of star-forming galaxies, which have typically steep spectra, at the lowest luminosities probed by this work ($\log L_{\rm X} \approx 42$–$42.5$) may decrease $F_{\rm obsc}$ in such regime. We therefore do not consider such trend to be significant.

The curve of Buchner et al. (2015) appears to suffer less from similar issues at low luminosities, although they reported a decrease of $F_{\rm obsc}(L_{\rm X})$ at low luminosities. The difference with our observed result is probably due to the different procedure used; Buchner et al. (2015) applied a Bayesian procedure that disfavours strong variations of $F_{\rm obsc}$ over close redshift and luminosity bins. At high redshift, the faint regime is not well sampled, probably causing the obscured AGN fraction at low luminosities to be dominated by the priors, i.e. to be similar to the value at lower redshift and/or higher luminosity, where incompleteness issues are less severe, alleviating this issue. The slight decrease of $F_{\rm obsc}$ towards low luminosity, which they ascribed to different possible physical mechanisms, may be at least partly due to incompleteness effects.

A strong anti-correlation between $F_{\rm obsc}$ and luminosity is usually found at low redshifts. For instance, grey empty symbols in Fig. 18 represent the fraction of AGN with $\log N_{\rm H} > 23$ derived by Burlon et al. 2011 at $z < 0.3$. This behaviour appears not to hold at high redshift, or at least to be much less evident. Comparing our points with the Burlon et al. (2011), we note the positive evolution of $F_{\rm obsc}$ from the local Universe to high redshift, which is stronger at high luminosities. In Vito et al. (2014), where we derived similar results from a combination of different X-ray surveys, we ascribed the larger fraction of luminous obscured AGN at high redshift than at low redshift to the larger gas fractions of galaxies at earlier cosmic epochs (e.g. Carilli & Walter 2013), which can cause larger covering factors and/or longer obscuration phases. This description is especially true if luminous AGN are preferentially triggered by wet-merger episodes (e.g. Di Matteo, Springel & Hernquist 2005; Menci et al. 2008; Treister et al. 2012), whose rate is expected to be higher in the early Universe. In this case, chaotic accretion of gas on to the SMBH can produce large covering factors.

Figs 16 and 18 do not account for the covariance between the $F_{\rm obsc} - z$ and $F_{\rm obsc} - L_{\rm X}$ trends, due to our data set being flux limited. This could in principle bias the results when investigating $F_{\rm obsc}$ separately as a function of one of the two parameters. To check for this possible effect, we derived the trend of $F_{\rm obsc}$ with redshift separately for two luminosity bins, $\log L_{\rm X} = 43$–43.5 and 43.5–45, chosen not to suffer significantly from incompleteness and to include approximately the same number of sources. For both subsamples, $F_{\rm obsc}$ is consistent with being flat at similar values, reassuring us that this result does not strongly depend on the intrinsic trend with redshift. We also repeated the procedure dividing the sample in two redshift bins and deriving the trend of $F_{\rm obsc}$ with luminosity in each of them. The results are consistent with those reported above in this section within the (larger) uncertainties.

6 EVOLUTION OF THE HIGH-REDSHIFT AGN POPULATION

6.1 The AGN XLF at high redshift

The AGN XLF can be derived from equation (22) by integrating over narrower redshift intervals and dividing the results by the volume sampled by the surveys in each redshift bin. Fig. 19 presents the XLFs in two redshift bins, chosen to include approximately the same numbers of sources. As stated in Section 3.4, ignoring the correction for Eddington bias does not change significantly the luminosity distribution of our sample, and, as a consequence, the derived XLFs. In particular, the slope of the faint end does not significantly vary when applying or neglecting such a correction. Our XLFs are fairly consistent with previous observational results from Aird et al. (2010), Ueda et al. (2014) and Vito et al. (2014, 2016). They are also in agreement with Buchner et al. (2015) at $z < 3.6$, which do not provide an analytical expression for the XLF, but rather confidence regions in redshift intervals. At higher redshifts, the Buchner et al. (2015) region exceeds our results. Similarly, the simulated XLFs by Habouzit, Volonteri & Dubois (2017) appear to overpredict the density of low-luminosity AGN.

Giallongo et al. (2015) applied a detection procedure based on searching for clustering of photons in energy, space and time, at the pre-determined optical positions of CANDELS-detected galaxies (see also Fiore et al. 2012), which allowed the authors to push the *Chandra* sensitivity beyond the limit reachable by blind detection methods (i.e. with no previous knowledge of the positions of galaxies). The large number of low-luminosity AGN at high redshift detected by Giallongo et al. (2015) suggests that AGN could have an important role in cosmic reionization (e.g. Madau & Haardt 2015; but see also Ricci et al. 2017).

Giallongo et al. (2015) derived the UV LF of their sample of X-ray detected, high-redshift AGN candidates by deriving the absolute UV magnitude from the apparent optical magnitude in the filter closest to rest-frame 1450 Å at the redshift of each source. We transformed their UV LF into an X-ray LF (see Fig. 19) by assuming a SED shape $F_\nu \propto \nu^{-0.5}$ between 1450 and 2500 Å (as in Georgakakis et al. 2015), the Lusso et al. (2010) α_{ox} and $\Gamma = 1.8$ for the X-ray spectrum (see also Vito et al. 2016). Our results do not support the very steep LFs derived by Giallongo et al. (2015). Parsa et al. (2017) recently disputed the very detection of the faint $z > 4$ AGN in Giallongo et al. (2015). Moreover, Cappelluti et al. (2016), applying a similar procedure which makes use of pre-determined optical positions, did not find such a large number of $z > 4$ sources. Other issues plausibly affect the assessment of the population of faint, high-redshift AGN, such as the uncertainties in the photometric redshifts, the expected presence of Eddington bias (which leads to overestimating the number of detected faint sources), and the XRBs contribution to X-ray emission at faint fluxes (e.g. Lehmer et al. 2012). Moreover, the uncertainties related to the assumed UV/X-ray spectral slope (discussed by Giallongo et al. 2015) may affect the conversion between UV magnitude and X-ray luminosity, although reasonable values for that conversion cannot produce the tension between their slope of the AGN XLF faint end and our results.

At the low luminosities probed by the 7 Ms CDF-S ($\log L_{\rm X} = 42 - 43$), X-ray emission from XRB may provide a non-negligible contribution (e.g. Bauer et al. 2002; Ranalli, Comastri & Setti 2003, 2005; Lehmer et al. 2010; Mineo, Gilfanov & Sunyaev 2012). The green dashed line in Fig. 19 (right-hand panel) is the XLF of star-forming galaxies at $z = 4$, where the emis-

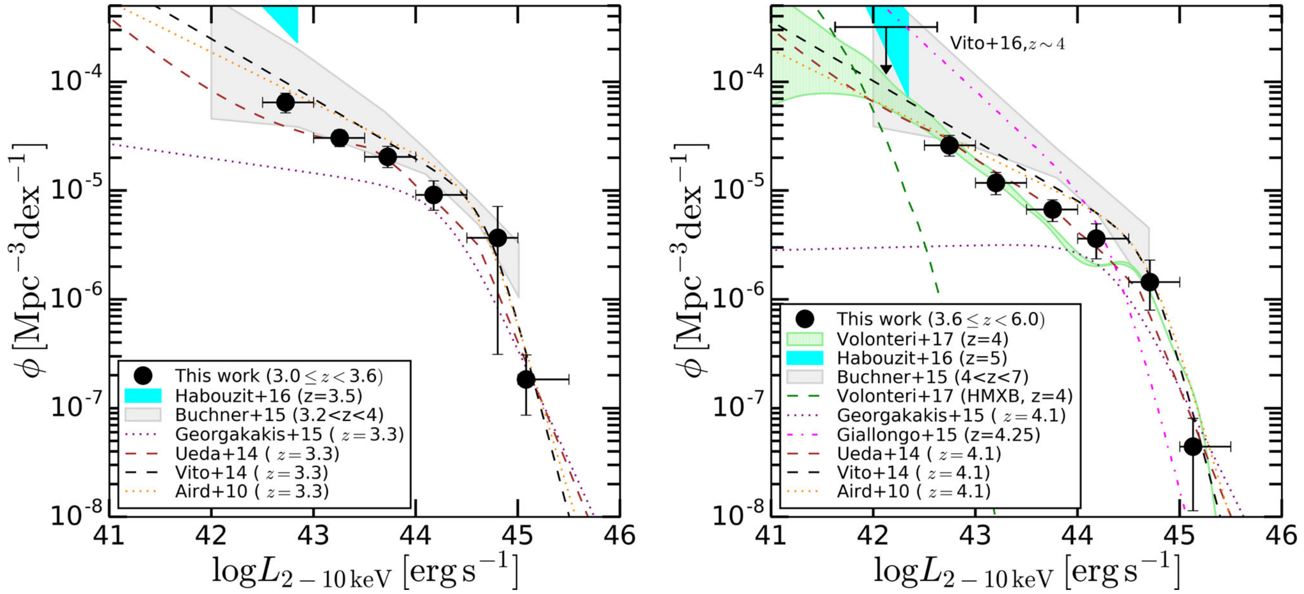

Figure 19. XLFs of AGN at $z = 3$–3.6 (left-hand panel) and $z = 3.6$–6 (right-hand panel), compared with results from previous observational (Aird et al. 2010; Ueda et al. 2014; Vito et al. 2014; Buchner et al. 2015; Georgakakis et al. 2015; Giallongo et al. 2015; extrapolated to lower luminosities than those probed by the original works) and theoretical (Habouzit et al. 2017; Volonteri et al. 2017) works at similar redshifts. In the right-hand panel, the downward-pointing arrow is the upper limit derived by Vito et al. (2016) through a stacking analysis of the 7 Ms CDF-S data set, and the dashed green line is the expected XLF of HMXBs at $z = 4$ (from Volonteri et al. 2017). Our results exclude extremely steep or flat slopes of the faint end of the AGN XLF.

sion is mostly due to high-mass XRB, from the model by Volonteri et al. (2017, see in particular their section 2.2 for assumptions and caveats). Briefly, they mapped the stellar-mass functions of Song et al. (2016) into an XLF through the main sequence of star-forming galaxies (Salmon et al. 2015) and the SFR-L_X scaling relation by Fragos et al. (2013, model 269), which Lehmer et al. (2016) demonstrated to be the best representation at $z = 0$–2.5. At $\log L_X \gtrsim 42.5$, the luminosity limit we applied to our luminosity functions, as the $42 \lesssim \log L_X \lesssim 42.5$ regime, was strongly affected by incompleteness; the density of star-forming galaxies is a factor of $\gtrsim$10 lower than the AGN density. We conclude that at such luminosities, our sample is not significantly contaminated by star-forming galaxies. At lower luminosities, the XLF of star-forming galaxies is comparable to, or even exceeds, the AGN XLF, in agreement with Vito et al. (2016), who found that the X-ray emission in galaxies individually undetected in the 7 Ms CDF-S is mostly due to XRB. Similar conclusions are reached by scaling the star-formation rate functions (SFRF; i.e. the comoving space densities of galaxies per unit star-formation rate) from Smit et al. (2012) and Gruppioni et al. (2015), at $z \approx 3.5$ and 4, respectively, into XLF using the best-fitting, redshift-dependent Fragos et al. (2013) and Lehmer et al. (2016) relations.

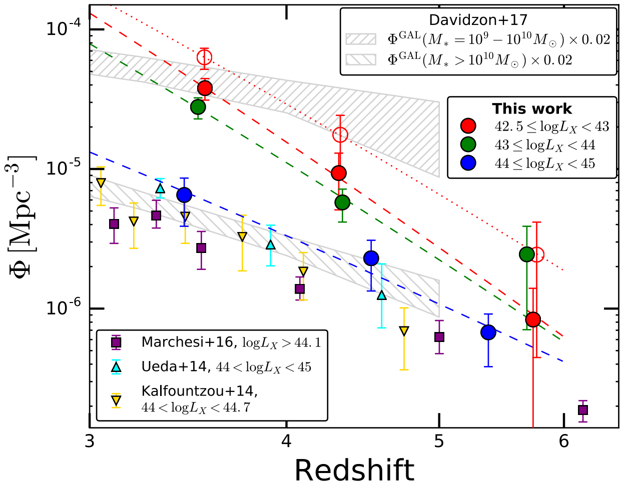

Figure 20. Comoving space density of three luminosity classes of AGN, divided in three redshift bins ($z = 3$–4, 4–5, 5–6). The points are located at the weighted average redshift of each subsample. For the low-luminosity bin, the empty circles indicate the results derived applying the correction factors described in Section 4. At higher luminosities, the corrections have negligible effects, and are not shown for clarity. Since high-luminosity sources are not well sampled by the deep, pencil-beam fields used in this work, we report previous results in similar luminosity bins derived analysing data from wider fields. The dashed curves are the best-fitting models described in Section 6. The dotted curve is the best-fitting model for the low-luminosity bin, after the correction for incompleteness. The space density of galaxies in two different mass regimes from Davidzon et al. (2017), rescaled by an arbitrary factor of 0.02, is shown as a grey stripe.

6.2 The evolution of the AGN space density and comparison with the galaxy population

Integrating the XLF over luminosity, we derived the comoving space density of AGN in three luminosity bins (Fig. 20). The decline of the space density in Fig. 20 appears to be steeper for low-luminosity AGN than for luminous sources. However, incompleteness significantly affects the detection of low-luminosity AGN, as can be seen in Fig. 13. In fact, applying the incompleteness correction discussed in Section 4 to the density of low-luminosity AGN (empty circles in Fig. 20), a slightly flatter behaviour is derived. To quantify the decline of the space density, we fitted a model $\Phi(z) = A((1+z)/(1+z_0))^p$, where $z_0 = 3$, to the points through a simple χ^2 minimization procedure, before (dashed lines in Fig. 20) and after the correction for incompleteness was applied. Fig. 20 displays the best-fitting model corrected for incompleteness only

Table 3. Best-fitting parameters describing the space-density evolution with redshift for the luminosity-based and obscuration-based subsamples (see Section 6).

Subsample	A (Mpc^{-3})	p
All	$(2.47 \pm 0.57) \times 10^{-4}$	-7.59 ± 0.69
$42.5 \le \log L_X < 43$	$(1.31 \pm 0.08) \times 10^{-4}$	-9.57 ± 0.32
$43 \le \log L_X < 44$	$(7.96 \pm 2.70) \times 10^{-5}$	-8.85 ± 1.71
$44 \le \log L_X < 45$	$(1.33 \pm 0.38) \times 10^{-5}$	-6.17 ± 0.87
Obscured	$(1.17 \pm 0.43) \times 10^{-4}$	-8.12 ± 1.11
Unobscured	$(1.44 \pm 0.34) \times 10^{-4}$	-8.25 ± 0.77
	After correction for incompleteness	
All	$(4.22 \pm 1.19) \times 10^{-4}$	-7.08 ± 0.77
$42.5 \le \log L_X < 43$	$(1.80 \pm 0.02) \times 10^{-4}$	-8.15 ± 0.36
$43 \le \log L_X < 44$	$(8.28 \pm 2.94) \times 10^{-5}$	-8.93 ± 1.82
$44 \le \log L_X < 45$	$(1.35 \pm 0.37) \times 10^{-5}$	-6.16 ± 0.84
Obscured	$(2.87 \pm 1.05) \times 10^{-4}$	-7.66 ± 1.14
Unobscured	$(1.37 \pm 0.26) \times 10^{-4}$	-6.58 ± 0.62

for the low-luminosity bin (dotted line), as at higher luminosities incompleteness is negligible. Table 3 summarizes the best-fitting values for the total sample and the three luminosity-based subsamples.

The evolution of the space density of luminous ($\log L_X \gtrsim 44$) AGN in Fig. 20 is consistent with previous results by, e.g., Hiroi et al. (2012), Kalfountzou et al. (2014), Vito et al. (2014) and Marchesi et al. (2016) (i.e. $\Phi \propto (1+z)^p$ with $p \sim -6$, see Table 3). Luminous ($\log L_X > 44$) AGN are usually hosted by massive ($M_* \gtrsim 10^{10} M_\odot$) galaxies (e.g. Xue et al. 2010; Yang et al. 2017). Fig. 20 shows that the space densities of luminous AGN and massive galaxies (from Davidzon et al. 2017) evolve with similar slopes, suggesting that the decline of the space density of $\log L_X > 44$ AGN is merely driven by the evolution of the galaxy number density. This interpretation is further supported by the recent findings of Yang et al. (submitted), who derive a positive evolution of the *average* black hole accretion rate in massive galaxies from low redshift up to $z \approx 3$, followed by a steady behaviour up to $z \approx 4$. This trend would be in contrast with the well-established decline of the space density of luminous AGN if the number of massive galaxies itself did not decrease.

Thanks to the unmatched sensitivity of the *Chandra* deep fields, in particular the 7 Ms CDF-S, coupled with the deep multiwavelength coverage, which allows a nearly complete multiwavelength identification of the X-ray sources, we could push the investigation of the evolution of the AGN space density down to $\log L_X = 42.5$ and up to $z = 6$. We found a slightly steeper decline for low-luminosity sources than for high-luminosity AGN, even after applying our fiducial corrections for incompleteness (Fig. 20 and Table 3). However, such corrections depend on the assumed intrinsic distribution in redshift, column density and luminosity (see Section 4). Moreover, the fit we performed is an approximation (although widely used in literature), as the error bars of the points in Fig. 20, are derived through a bootstrap procedure, which formally returns a measurement of the dispersion of the bootstraped distribution. For these reasons, we do not consider the best-fitting values reported in Table 3 as definitive evidence, but rather a suggestion, for a differential evolution of the AGN space density in different luminosity regimes. However, the strong flattening of the XLF faint-end found by Georgakakis et al. (2015) is ruled out by Fig. 19. Future X-ray missions, discussed in Section 7.2, will shed light on the evolution of the population of low-luminosity AGN at higher redshifts than those probed in this work.

Intriguingly, the evolution of the space density of low-luminosity AGN is also steeper than the density evolution of both low-mass and high-mass galaxies (see Fig. 20), as derived by Davidzon et al. (2017). This discrepancy requires an evolution of one or more of the physical parameters driving black hole accretion for low-luminosity AGN, other than the mere evolution of the number of galaxies. For instance, a steeper evolution of the AGN density than galaxy density can be expected if the duty cycle and/or occupation fraction decrease at increasing redshift, especially for low-mass galaxies.

6.3 XLF and space density of obscured and unobscured AGN

Fig. 21 (left-hand panel) shows the luminosity functions separately for obscured and unobscured sources, obtained by performing the integral over column density in equation (22) over the ranges $\log N_H = 23$–25 and 20–23, respectively, compared with the XLFs of Hasinger, Miyaji & Schmidt (2005, used by Gilli et al. 2007 for their X-ray background synthesis model) in the same N_H bins. We also report the results from Aird et al. (2015), where obscured and unobscured AGN are defined as those obscured by column densities $\log N_H = 22$–24 and 20–22, respectively. Points in the first luminosity bin (empty circles) are heavily affected by the obscuration-dependent incompleteness discussed in Section 4, as is evident by the drop of the density of obscured AGN with respect to the unobscured subsample. Fig. 21 (right-hand panel) displays the evolution of the space density of obscured and unobscured AGN. Table 3 summarizes the best-fitting values for the obscured and unobscured subsamples. We found no significant difference in the evolution of the space density of obscured and unobscured AGN. This conclusion is consistent with Fig. 16, where the obscured AGN fraction is flat with redshift. After correcting for incompleteness, a slightly shallower decline of the space density of unobscured AGN is derived, because of the faster decrease of detection completeness with increasing redshift for unobscured sources than for obscured sources (see Fig. 15). However, the uncertainties due to the procedure of correcting for incompleteness and the relatively small size of the sample prevent us from drawing strong conclusions.

6.4 Black hole accretion rate density

The black hole accretion rate density (BHAD) is defined as

$$\Psi(z) = \int \frac{(1-\epsilon)}{(\epsilon c^2)} L_{\mathrm{bol,AGN}} \phi(L_{\mathrm{bol,AGN}}, z) \mathrm{d}\log L_{\mathrm{bol,AGN}}, \quad (29)$$

where $L_{\mathrm{bol,\,AGN}} = K_{\mathrm{bol}}(L_{\mathrm{X,\,AGN}}) L_{\mathrm{X,\,AGN}}$ is the AGN bolometric luminosity, $K_{\mathrm{bol}}(L_{\mathrm{X,\,AGN}})$ is the bolometric correction, ϵ is the radiative efficiency (fixed here to $\epsilon = 0.1$) and $\phi(L_{\mathrm{bol,\,AGN}}, z)$ is the AGN bolometric luminosity function, which can be derived from the AGN XLF through the bolometric correction. We assumed $K_{\mathrm{bol}}(L_{\mathrm{X,\,AGN}})$ from Lusso et al. (2012). Fig. 22 presents the BHAD derived from our sample as filled black circles. We used the same luminosity bins of Vito et al. (2016) in order to facilitate the comparison between the two works. In Vito et al. (2016), we derived the BHAD due to X-ray undetected $z = 3.5$–6.5 galaxies in the 7 Ms CDF-S through a stacking procedure. Empty circles and squares present the BHAD for the entire sample of galaxies and for the half of most massive ones, respectively, under the most optimistic assumption that all the signal is due to accretion on to SMBH (i.e. the signal due to XRB is negligible). By comparing these points with the BHAD of a preliminary sample of high-redshift, X-ray-detected AGN in the 7 Ms CDF-S, we concluded that most of the BH growth at $z = 3.5$–6.5 occurs during the bright and fast (relative to galaxy lifespan) AGN phase,

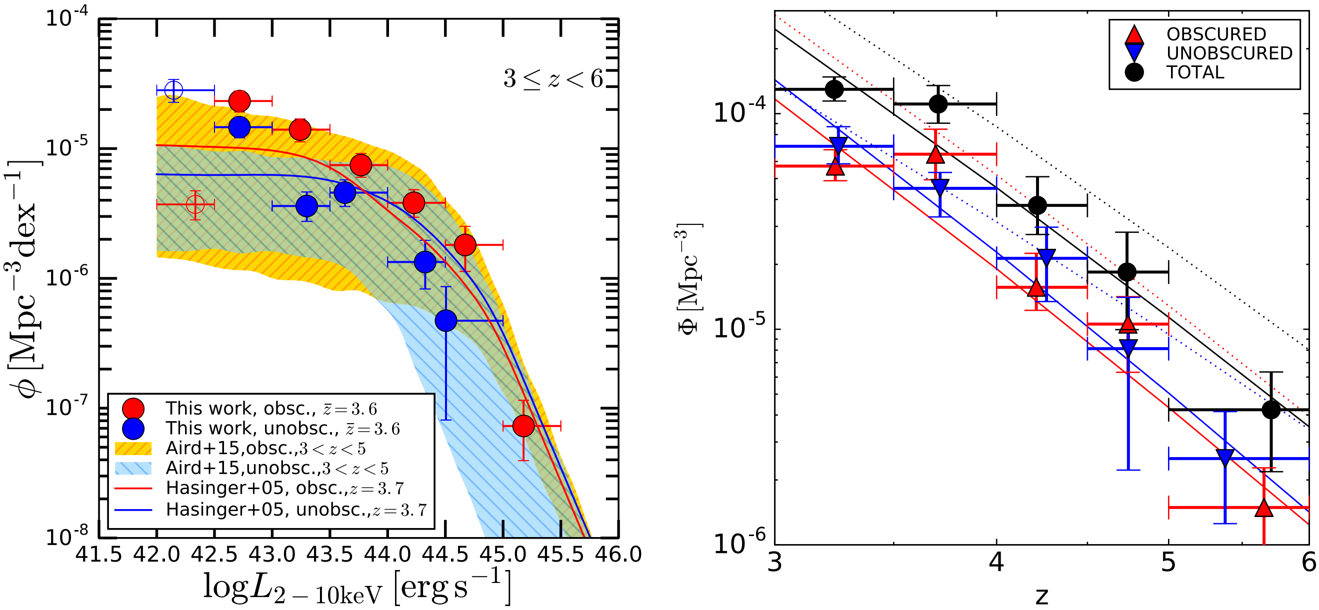

Figure 21. Left-hand panel: XLFs of obscured (red circles) and unobscured (blue circles) AGN at z = 3–6. The open circles are the binned points in the faintest luminosity bin, and show how incompleteness affects differentially the two subsamples at those luminosities (see Section 4). Symbols are centred at the weighted average luminosity of each bin. Red and blue lines are the XLFs of obscured ($\log N_{\mathrm{H}} > 23$) and unobscured ($\log N_{\mathrm{H}} < 23$) AGN, respectively, by Hasinger et al. (2005). Orange and cyan regions represent the areas covered by luminosity functions of $\log N_{\mathrm{H}} > 22$ and <22 AGN, respectively, at $3 \leq z \leq 5$ from Aird et al. (2015). The different column density thresholds assumed to define the obscuration-based subsamples are discussed in the text. Right-hand panel: evolution of the comoving space density with redshift for obscured (red symbols) and unobscured (blue symbols) subsamples. Those symbols are slightly offset for visualization purposes. Black circles represent the total sample. Solid lines are the best-fitting linear models obtained by simple χ^2 minimization. To demonstrate the effects of the obscuration-dependent incompleteness (Section 4), we present the best-fitting linear models corrected for such incompleteness as dotted lines.

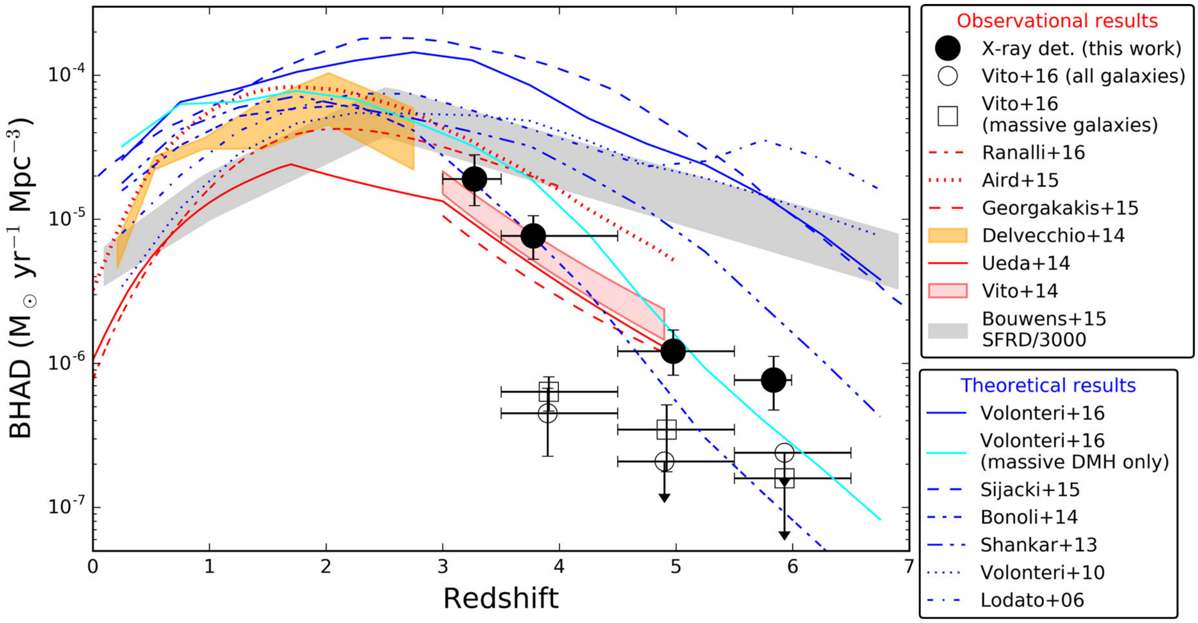

Figure 22. BHAD derived from our sample of X-ray-detected AGN (filled black symbols), compared with previous observational and theoretical results from literature. We also show the BHAD resulting from the stacking analysis of a sample of individually undetected galaxies (Vito et al. 2016), under the most optimistic assumption that all the stacked X-ray signal is due to accretion on to SMBH. Empty circles and squares correspond to the entire sample and the half containing the most massive galaxies, respectively. The grey stripe is the star formation rate density of Bouwens et al. (2015), rescaled by a factor of 3000.

with negligible contribution from continuous, low-rate accretion. In this work, we refine the sample of X-ray-detected, high-redshift AGN and are therefore able to present a more accurate comparison.

Fig. 22 displays the BHAD corresponding to the XLFs of Ranalli et al. (2016), Aird et al. (2015), Georgakakis et al. (2015), Ueda et al. (2014) and Vito et al. (2014), converted into BHAD as described above,[12] and that derived by Delvecchio et al. (2014) exploiting *Herschel* data. We also add the BHAD predicted by simulations (Lodato & Natarajan 2006; Volonteri 2010; Shankar, Weinberg & Miralda-Escudé 2013; Bonoli, Mayer & Callegari 2014; Sijacki et al. 2015; Volonteri et al. 2016a) under different assumptions of BH seed mass and growth mechanisms. Except for Sijacki et al. (2015), who directly reported the BHAD, we derived the curves by differentiating the BH mass densities presented in those works.

Simulations tend to derive BHAD exceeding by about one order of magnitude the observed values at $z > 3$ (see Fig. 22). This discrepancy is usually ascribed to the contribution of continuous, low-rate accretion, difficult to detect even in deep surveys. In Vito et al. (2016), we used a stacking procedure to demonstrate that such contribution is negligible compared to the BHAD due to X-ray-detected AGN (see also e.g. Pezzulli et al. 2017). With the refined sample of high-redshift AGN of this work, we confirmed this result up to $z \sim 6$ and suggest that simulations tend to overestimate the BHAD at high redshift.

A similar issue is known to affect the faint end of the AGN LF at high redshift, with simulations finding steeper slopes than the observed ones. Volonteri et al. (2016a) suggested that this discrepancy could be due to supernovae feedback (i.e. the removal of gas in the galaxy responsible of both nuclear accretion and star formation as a consequence of supernovae activity) being underestimated in low-mass haloes in simulations. Considering only massive ($M > 5 \times 10^{11} M_{\odot}$) dark-matter haloes, where supernovae feedback has little effect due to the deeper galactic gravitational potential well, improves the agreement with observations. This statement is also true for the BHAD, as can be seen in Fig. 22, where the results of Volonteri et al. (2016a) are presented without any mass cut (continuous blue curve) and for massive haloes only (cyan curve).

The common evolution of star formation rate density (SFRD) and the BHAD at $z \lesssim 2.5$ is generally considered to be a manifestation of the BH–galaxy co-evolution. However, these two quantities are found to evolve with different slopes at higher redshifts (e.g. Silverman et al. 2008; Aird et al. 2015, see e.g. Haiman, Ciotti & Ostriker 2004 for a theoretical interpretation), mainly through extrapolations of the AGN XLF. In this work, we can probe this behaviour, whose physical causes are difficult to identify, down to the faint flux limits of the *Chandra* deep fields. Most of the BHAD we derived at high redshift is due to luminous ($\log L_{\rm X} > 44$) AGN, which contribute 70−90 per cent to the black circles in Fig. 22; therefore, the uncertainties on the density of low-luminosity AGN have no strong impacts in this respect. In particular, the contribution of accretion in individually undetected galaxies is negligible (empty black symbols from Vito et al. 2016 in Fig. 22). Indeed, the evolution the BHAD is similar to that of the space density of luminous AGN, both declining by a factor of $\gtrsim$10 from $z = 3$ to $z = 6$. In contrast, the evolution of the SFRD is more similar to that of low-mass galaxies than massive systems, both declining by a factor of ∼3 in the same redshift range (cf. Figs 22 and 20). This discrepancy may mark a breakdown of the BH–galaxy co-evolution at high redshift.

7 FUTURE PROSPECTS

7.1 Spectroscopic follow-up of high-redshift AGN candidates

One of the main sources of uncertainty affecting our knowledge of the demography of high-redshift AGN is the persistently low spectroscopic identification rates of the candidates. The deep *Chandra* surveys allow sampling of the bulk of the population at high redshift, which is constituted of low luminosity and/or obscured sources. However, these objects are typically faint in the optical/NIR bands (e.g. ≈80 per cent of our sources located in the CANDELS/GOODS-S area have $H > 24$), and are therefore difficult to follow up spectroscopically with current facilities in order to confirm their high-redshift nature. While photometric redshifts of good quality are available in the fields we considered, and despite the careful treatment of their PDF, they may introduce non-negligible uncertainties in the final results (e.g. see section 4.3 of Georgakakis et al. 2015). This issue is less problematic for shallower X-ray surveys, which sample AGN with brighter optical/NIR counterparts, but miss the sub-population of low-luminosity AGN.

Current facilities that could be used effectively to increase the spectroscopic completeness of our sample include *ALMA*[13] (e.g. Walter et al. 2016), the Keck observatory[14] and *VLT/MUSE*.[15] The last, in particular, with 27 h of integration time in the *Hubble* Ultra Deep Field, has already been able to obtain spectroscopic redshifts for sources below the *Hubble* detection limit (Bacon et al. 2015). In the near future, facilities like *James Webb Space Telescope (JWST)*,[16] and, in the somewhat longer term, the Extremely Large Telescopes (with $\gtrsim$30 m diameter mirrors) will provide the required spectroscopic sensitivity to perform ultradeep spectroscopic follow-up campaigns aimed at identifying high-redshift AGN candidates with extremely high efficiency (see e.g. the discussion in Brandt & Alexander 2015).

7.2 The case for *Athena* and *Lynx*

The understanding of the SMBH formation and growth mechanisms in the early Universe requires the collection of much larger samples of X-ray-selected AGN and galaxies at higher redshifts and lower luminosities than those probed by current X-ray surveys. Moreover, while *Chandra* and *XMM–Newton*, jointly with large-area optical/IR surveys, are successfully improving our knowledge of the bright quasar population up to $z \approx 7$ (e.g. Brandt & Vito 2017), the bulk of the AGN population, constituted by low luminosity and possibly obscured systems, are currently completely missed beyond $z \approx$ 5–6 even by the deepest X-ray surveys available, and *Chandra* observations substantially deeper than the 7 Ms CDF-S are not foreseeable. The *Athena* X-ray Observatory (Nandra et al. 2013)[17], which will be launched in ∼2028, is expected to detect hundreds of

---

[12] Aird et al. (2015) provided directly their estimate of BHAD, but used the bolometric corrections from Hopkins, Richards & Hernquist (2007) instead of the Lusso et al. (2012) values we assumed. The slightly higher normalization of the Aird et al. (2015) curve with respect to the other observational results reflects this difference and can be therefore considered a measure of the uncertainty due to bolometric corrections.

[13] http://www.almaobservatory.org/

[14] http://www.keckobservatory.org/

[15] http://muse-vlt.eu/

[16] https://jwst.stsci.edu/

[17] http://www.the-athena-X-ray-observatory.eu/

$\lesssim L_*$ AGN at $z > 6$ (Aird et al. 2013), thanks to its survey power a factor of ~100 faster than *Chandra* or *XMM–Newton*, boosting our knowledge of the high-rate accretion phases, which likely dominate SMBH growth at high redshift (Vito et al. 2016).

Simulations suggest that different combinations of the physical parameters governing SMBH formation and early growth (e.g. seed-mass and Eddington ratio distributions, occupation fraction, feedback mode, and their dependence on the host-galaxy and environment properties) produce different shapes of the faint-end ($\log L_X \lesssim 42$) of the AGN LF. The green-striped region in Fig. 19 (right-hand panel) is the locus of the XLF models of Volonteri et al. (2017) at $z = 4$, under different assumptions. In particular, the upper bound is derived using a minimum initial BH mass of 10^2 $M_\odot$ and occupation fraction equal to unity, while for the lower bound a minimum initial mass of $10^5 M_\odot$ and a modified occupation fraction from Habouzit et al. (2017), flagged as 'H17low1' in Volonteri et al. (2017). Different combinations of parameters are mostly encompassed by these boundaries. The models are normalized to the observed points, which explain the small spread at $\log L_X \gtrsim 43$. However, at lower luminosities, which even the currently deepest surveys cannot sample, different combinations of parameters produce significantly different XLF shapes. The spread of the XLF shapes in this faint regime is expected to increase at higher redshifts (see fig. 12 of Volonteri et al. 2017). A tremendous step forward towards understanding the early growth of SMBH will be provided by *Lynx* (also known as *X-ray Surveyor*; Weisskopf et al. 2015), which, with its *Chandra*-like angular resolution and ≈25 times higher sensitivity, will reach an expected flux limit of $\log F \approx -18.5$ in a blind 4 Ms survey, sufficient to probe the faint end of the XLF down to $\log L_X \approx 40$–42 at $z \lesssim 10$, providing unprecedented constraints on the combinations of parameters driving SMBH formation and early growth.

The constraints derived by X-ray telescopes will be complemented by deep optical/IR observations provided by future facilities such as *JWST* and *WFIRST*[18], necessary to identify the X-ray sources. Recently, several works have focused on the multiwavelength properties of emission from BHs during the early accretion phases. For instance, Volonteri et al. (2017) proposed a colour-colour selection based on *JWST* bands to separate AGN from normal galaxies at high redshift. In particular, the characteristic shape of the SED of accreting direct-collapse BH (with initial $M_{\rm BH} = 10^4$–10^5 $M_\odot$) has been used by Natarajan et al. (2017) to propose a *JWST* colour-colour method to select such objects. In contrast, optical/IR emission from accreting light-seeds BH ($M_{\rm BH} = 10^2$–10^3 $M_\odot$) probably will be outshined by stellar emission in the young host galaxies. The joint effort of the next generation X-ray observatories and optical/IR telescopes therefore will open a completely new window on the cosmic epoch where SMBH formed.

8 CONCLUSIONS

We have investigated the population of $3 \le z < 6$ AGN detected in the two deepest X-ray fields, the 7 Ms CDF-S and the 2 Ms CDF-N, selected on the basis of a careful assessment of the photometric-redshift uncertainties for sources lacking spectroscopic identification. For each source, uncertainties on redshift, column density, flux and luminosity were accounted for by using the full probability distributions of such parameters, and a statistical correction for Eddington bias was applied. Our main results are the following:

(i) We derived the best-constrained column-density distribution of moderate-luminosity obscured AGN at high redshift, finding it to peak at $\log N_{\rm H} \sim 23.5$–24 (see Section 3.2).

(ii) Comparing the mean photon index of X-ray spectra of high-redshift, $L_X < L_*$ AGN with results derived for luminous quasars, we found no significant variation with redshift and luminosity (see Appendix Section B).

(iii) The obscured AGN fraction ($F_{\rm obsc}$), where the column-density threshold to define obscured versus unobscured AGN is set to $\log N_{\rm H} = 23$, is consistent with being flat with redshift at observed values of ≈0.5–0.6. Once corrected for incompleteness, we derive $F_{\rm obsc} \approx 0.6$–0.7. This result agrees with previous works finding a saturation of this quantity at $z \gtrsim 2$, after a positive evolution at $z \approx 0$–2 (see Section 5.1).

(iv) Thanks to the sensitivity of the *Chandra* deep fields, we could probe the number counts of high-redshift AGN down to fainter fluxes, especially at $z > 4$, than those reached by wider and shallower surveys. Comparing the $\log N$-$\log S$ of our sample with the predictions from the Gilli et al. (2007) X-ray background synthesis model, which was constrained at lower redshifts, separately for obscured and unobscured AGN, we found an excess of obscured sources, likely caused by the positive evolution of $F_{\rm obsc}$ from the local Universe to high redshift (see Section 3.3.1).

(v) We found $F_{\rm obsc} \approx 0.7$–0.8 at $\log L_X \gtrsim 43$, somewhat larger, but still in agreement with other works performed at similar redshifts. In contrast with previous results, derived especially at low redshift, there is no significant anticorrelation between $F_{\rm obsc}$ and luminosity at $\log L_X \gtrsim 43$, suggesting a stronger positive evolution of $F_{\rm obsc}$ towards earlier cosmic times for luminous AGN than for moderate-luminosity sources (see Section 5.3).

(vi) At low luminosities ($\log L_X \lesssim 43$), the observed behaviour of $F_{\rm obsc}$ is strongly affected by sample incompleteness: low luminosity, obscured sources are more difficult to detect than unobscured sources, biasing the observed result towards low values of $F_{\rm obsc}$. We suggest that recent claims of a turn-down of $F_{\rm obsc}$ towards low luminosities at high redshift may be at least partly due to similar incompleteness issues. By assuming intrinsic distributions of sources in $N_{\rm H}$ and L_X space, we applied tentative corrections that led $F_{\rm obsc}$ to be fairly consistent with the values at higher luminosities, although a possible slight decrease down to values of ≈0.6 is still visible (see Sections 5.3 and 4).

(vii) The space density of luminous ($\log L_X \gtrsim 44$) AGN declines from $z = 3$ to $z = 6$, as $(1 + z)^d$, with $d \approx -6$, in agreement with previous results. The XLF in the redshift range probed by this work does not show significant steepening of the faint end, which has been recently linked with a possible significant contribution of low-luminosity AGN to cosmic reionization. Our results tend to exclude this hypothesis. By contrast, the space density of AGN with low-to-moderate luminosity appears to decline slightly faster than for high-luminosity AGN. However, uncertainties due to incompleteness might affect this result. The evolution of the AGN XLF faint end at high redshift is particularly important for placing constraints on the combination of the physical parameters driving the formation and early growth of SMBHs, one of the goals of future X-ray observatories (see Sections 6 and 7.2).

(viii) The space densities of luminous AGN and massive galaxies show a similar trend with redshift, suggesting that the evolution of luminous AGN is driven by the underlying galaxy population. By contrast, the space density of low-luminosity AGN declines with increasing redshift faster than the space density of both high-mass and low-mass galaxies. This difference suggests an evolution in the physical parameters regulating nuclear activity, such as duty cycle

---

[18] https://wfirst.gsfc.nasa.gov/

and occupation fraction. We also derived the BHAD at $z = 3$–6, and noted that, while the BHAD and SFRD track each other at $z \approx 0$–2.5, at higher redshifts the BHAD decreases with a faster rate than the SFRD (see Section 6).

ACKNOWLEDGEMENTS

We are grateful to the referee, A. Georgakakis, for useful comments that helped in clarifying several aspects of this work. FV, WNB and GY acknowledge support from *Chandra* X-ray Center grant GO4-15130A, the PennState ACISInstrument Team Contract SV4-74018 (issued by the *Chandra* X-ray Center, which is operated by the Smithsonian Astrophysical Observatoryfor and on behalf of NASAunder contract NAS8-03060), and the V.M. Willaman Endowment. BL acknowledge support from the National Natural Science Foundation of Chinagrant 11673010 and the Ministry of Science and Technologyof China grant 2016YFA0400702. YQX acknowledges support from the 973 Program (2015CB857004), NSFC-11473026, NSFC-11421303, the CASFrontier Science Key Research Program (QYZDJ-SSW-SLH006), and the Fundamental Research Funds for the Central Universities. We warmly thank Harry Ferguson and Dritan Kodra for providing PDFs of CANDELS photometric-redshifts, and Iary Davidzon, Carlotta Gruppioni, and Melanie Habouzit for providing useful data. The Guaranteed Time Observations for the CDF-N included here were selected by the ACISInstrument Principal Investigator, Gordon P. Garmire, currently of the Huntingdon Institute for X-ray Astronomy, LLC, which is under contract to the Smithsonian Astrophysical Observatory; Contract SV2-82024.

REFERENCES

Ai Y., Dou L., Fan X., Wang F., Wu X.-B., Bian F., 2016, ApJ, 823, L37
Aird J. et al., 2010, MNRAS, 401, 2531
Aird J. et al., 2013, preprint (arXiv:1306.2325)
Aird J., Coil A. L., Georgakakis A., Nandra K., Barro G., Pérez-González P. G., 2015, MNRAS, 451, 1892
Alexander D. M. et al., 2003, AJ, 126, 539
Arnaud K. A., 1996, in Jacoby G. H., Barnes J., eds, ASP Conf. Ser. Vol. 101, Astronomical Data Analysis Software and Systems V. Astron. Soc. Pac., San Francisco, p. 17
Avni Y., Bahcall J. N., 1980, ApJ, 235, 694
Bacon R. et al., 2015, A&A, 575, A75
Bañados E. et al., 2016, ApJS, 227, 11
Barger A. J., Cowie L. L., Brandt W. N., Capak P., Garmire G. P., Hornschemeier A. E., Steffen A. T., Wehner E. H., 2002, AJ, 124, 1839
Bauer F. E., Alexander D. M., Brandt W. N., Hornschemeier A. E., Vignali C., Garmire G. P., Schneider D. P., 2002, AJ, 124, 2351
Bishop C. M., 2006, Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, New York
Bonoli S., Mayer L., Callegari S., 2014, MNRAS, 437, 1576
Bouwens R. J. et al., 2015, ApJ, 803, 34
Brandt W. N., Alexander D. M., 2015, A&AR, 23, 1
Brandt W. N., Vito F., 2017, Astron. Nachr., 338, 241
Brightman M. et al., 2013, MNRAS, 433, 2485
Brightman M. et al., 2016, ApJ, 826, 93
Broos P. S., Townsley L. K., Feigelson E. D., Getman K. V., Bauer F. E., Garmire G. P., 2010, ApJ, 714, 1582
Brusa M. et al., 2009, ApJ, 693, 8
Buchner J. et al., 2015, ApJ, 802, 89
Burlon D., Ajello M., Greiner J., Comastri A., Merloni A., Gehrels N., 2011, ApJ, 728, 58
Cappelluti N. et al., 2016, ApJ, 823, 95
Carilli C. L., Walter F., 2013, ARA&A, 51, 105
Cash W., 1979, ApJ, 228, 939
Civano F. et al., 2011, ApJ, 741, 91
Comastri A. et al., 2011, A&A, 526, L9
Davidzon I. et al., 2017, A&A, 605, A70
Delvecchio I. et al., 2014, MNRAS, 439, 2736
Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604
Fanali R., Caccianiga A., Severgnini P., Della Ceca R., Marchese E., Carrera F. J., Corral A., Mateos S., 2013, MNRAS, 433, 648
Farrah D., Priddey R., Wilman R., Haehnelt M., McMahon R., 2004, ApJ, 611, L13
Ferrarese L., Merritt D., 2000, ApJ, 539, L9
Fiore F. et al., 2012, A&A, 537, A16
Fragos T. et al., 2013, ApJ, 764, 41
Georgakakis A. et al., 2015, MNRAS, 453, 1946
Giallongo E. et al., 2015, A&A, 578, A83
Gilli R., Comastri A., Hasinger G., 2007, A&A, 463, 79
Gilli R., Comastri A., Vignali C., Ranalli P., Iwasawa K., 2010, X-ray Astronomy 2009; Present Status, Multi-Wavelength Approach and Future Perspectives, 1248, 359
Gilli R. et al., 2014, A&A, 562, A67
Goodman J., Weare J., 2010, Communications in Applied Mathematics and Computational Science, 5
Grogin N. A. et al., 2011, ApJS, 197, 35
Gruppioni C. et al., 2015, MNRAS, 451, 3419
Habouzit M., Volonteri M., Dubois Y., 2017, MNRAS, 468, 3935
Haiman Z., 2013, in Wiklind T., Mobasher B., Bromm V., eds, Astrophysics and Space Science Library, Vol. 396, The First Galaxies. Springer-Verlag, Berlin, p. 293
Haiman Z., Ciotti L., Ostriker J. P., 2004, ApJ, 606, 763
Hasinger G., 2008, A&A, 490, 905
Hasinger G., Miyaji T., Schmidt M., 2005, A&A, 441, 417
Hiroi K., Ueda Y., Akiyama M., Watson M. G., 2012, ApJ, 758, 49
Hopkins P. F., Richards G. T., Hernquist L., 2007, ApJ, 654, 731
Hsu L.-T. et al., 2014, ApJ, 796, 60 (H14)
Iwasawa K. et al., 2012, A&A, 546, A84
Johnson J. L., Haardt F., 2016, PASA, 33, e007
Just D. W., Brandt W. N., Shemmer O., Steffen A. T., Schneider D. P., Chartas G., Garmire G. P., 2007, ApJ, 665, 1004
Kalberla P. M. W., Burton W. B., Hartmann D., Arnal E. M., Bajaja E., Morras R., Pöppel W. G. L., 2005, A&A, 440, 775
Kalfountzou E., Civano F., Elvis M., Trichas M., Green P., 2014, MNRAS, 445, 1430
Koekemoer A. M. et al., 2011, ApJS, 197, 36
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Lehmer B. D., Alexander D. M., Bauer F. E., Brandt W. N., Goulding A. D., Jenkins L. P., Ptak A., Roberts T. P., 2010, ApJ, 724, 559
Lehmer B. D. et al., 2012, ApJ, 752, 46
Lehmer B. D. et al., 2016, ApJ, 825, 7
Liu T. et al., 2017, ApJS, 232, 8
Lodato G., Natarajan P., 2006, MNRAS, 371, 1813
Luo B. et al., 2010, ApJS, 187, 560
Luo B. et al., 2017, ApJS, 228, 2 (L17)
Lusso E. et al., 2010, A&A, 512, A34
Lusso E. et al., 2012, MNRAS, 425, 623
Madau P., Haardt F., 2015, ApJ, 813, L8
Magorrian J. et al., 1998, AJ, 115, 2285
Marchesi S. et al., 2016, ApJ, 827, 150
Marconi A., Hunt L. K., 2003, ApJ, 589, L21
McGreer I. D. et al., 2013, ApJ, 768, 105
Menci N., Fiore F., Puccetti S., Cavaliere A., 2008, ApJ, 686, 219
Merloni A. et al., 2014, MNRAS, 437, 3550
Mineo S., Gilfanov M., Sunyaev R., 2012, MNRAS, 419, 2095
Moretti A. et al., 2014, A&A, 563, A46
Mortlock D. J. et al., 2011, Nature, 474, 616
Murphy K. D., Yaqoob T., 2009, MNRAS, 397, 1549
Nandra K. et al., 2013, preprint (arXiv:1306.2307)
Nanni R., Vignali C., Gilli R., Moretti A., Brandt W. N., 2017, A&A, 603, A128

Natarajan P., Pacucci F., Ferrara A., Agarwal B., Ricarte A., Zackrisson E., Cappelluti N., 2017, ApJ, 838, 117
Pacucci F., Ferrara A., Volonteri M., Dubus G., 2015, MNRAS, 454, 3771
Page M. J., Simpson C., Mortlock D. J., Warren S. J., Hewett P. C., Venemans B. P., McMahon R. G., 2014, MNRAS, 440, L91
Parsa S., Dunlop J. S., McLure R. J., 2017, MNRAS, preprint (arXiv:1704.07750)
Pezzulli E., Valiante R., Orofino M. C., Schneider R., Gallerani S., Sbarrato T., 2017, MNRAS, 466, 2131
Rafferty D. A., Brandt W. N., Alexander D. M., Xue Y. Q., Bauer F. E., Lehmer B. D., Luo B., Papovich C., 2011, ApJ, 742, 3
Ranalli P., Comastri A., Setti G., 2003, A&A, 399, 39
Ranalli P., Comastri A., Setti G., 2005, A&A, 440, 23
Ranalli P. et al., 2016, A&A, 590, A80
Reines A. E., Comastri A., 2016, PASA, 33, e054
Ricci F., Marchesi S., Shankar F., La Franca F., Civano F., 2017, MNRAS, 465, 1915
Risaliti G., Young M., Elvis M., 2009, ApJ, 700, L6
Salmon B. et al., 2015, ApJ, 799, 183
Santini P. et al., 2015, ApJ, 801, 97
Shankar F., Weinberg D. H., Miralda-Escudé J., 2013, MNRAS, 428, 421
Shemmer O., Brandt W. N., Vignali C., Schneider D. P., Fan X., Richards G. T., Strauss M. A., 2005, ApJ, 630, 729
Shemmer O., Brandt W. N., Netzer H., Maiolino R., Kaspi S., 2006, ApJ, 646, L29
Shemmer O., Brandt W. N., Netzer H., Maiolino R., Kaspi S., 2008, ApJ, 682, 81
Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, MNRAS, 452, 575
Silverman J. D. et al., 2008, ApJ, 679, 118
Skelton R. E. et al., 2014, ApJS, 214, 24 (S14)
Smit R., Bouwens R. J., Franx M., Illingworth G. D., Labbé I., Oesch P. A., van Dokkum P. G., 2012, ApJ, 756, 14
Song M. et al., 2016, ApJ, 825, 5
Straatman C. M. S. et al., 2016, ApJ, 830, 51 (S16)
Trakhtenbrot B. et al., 2017, MNRAS, 470, 800
Treister E., Urry C. M., 2006, ApJ, 652, L79
Treister E., Schawinski K., Urry C. M., Simmons B. D., 2012, ApJ, 758, L39
Ueda Y., Akiyama M., Hasinger G., Miyaji T., Watson M. G., 2014, ApJ, 786, 104
VanderPlas J., 2014, preprint (arXiv:1411.5018)
Vignali C., Bauer F. E., Alexander D. M., Brandt W. N., Hornschemeier A. E., Schneider D. P., Garmire G. P., 2002, ApJ, 580, L105
Vignali C., Brandt W. N., Schneider D. P., Kaspi S., 2005, AJ, 129, 2519
Vito F. et al., 2013, MNRAS, 428, 354
Vito F., Gilli R., Vignali C., Comastri A., Brusa M., Cappelluti N., Iwasawa K., 2014, MNRAS, 445, 3557
Vito F. et al., 2016, MNRAS, 463, 348
Volonteri M., 2010, A&AR, 18, 279
Volonteri M., Dubois Y., Pichon C., Devriendt J., 2016a, MNRAS, 460, 2979
Volonteri M., Habouzit M., Pacucci F., Tremmel M., 2016b, in Kaviraj S., ed., Proc. IAU Symp. 319, Galaxies at High Redshift and Their Evolution Over Cosmic Time, p. 72
Volonteri M., Reines A., Atek H., Stark D. P., Trebitsch M., 2017, preprint (arXiv:1704.00753)
Walter F. et al., 2016, ApJ, 833, 67
Wang F. et al., 2017, ApJ, 839, 27
Weigel A. K., Schawinski K., Treister E., Urry C. M., Koss M., Trakhtenbrot B., 2015, MNRAS, 448, 3167
Weisskopf M. C., Gaskin J., Tananbaum H., Vikhlinin A., 2015, in Hudec R., Pina L., eds., Proc. SPIE, Conf. Ser. Vol. 9510, EUV and X-ray Optics: Synergy between Laboratory and Space IV. SPIE, Bellingham
Wu X.-B. et al., 2015, Nature, 518, 512
Xue Y. Q. et al., 2010, ApJ, 720, 368
Xue Y. Q. et al., 2011, ApJS, 195, 10
Xue Y. Q., Luo B., Brandt W. N., Alexander D. M., Bauer F. E., Lehmer B. D., Yang G., 2016, ApJS, 224, 15 (X16)
Yang G. et al., 2014, ApJS, 215, 27
Yang G. et al., 2017, ApJ, 842, 72

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

online_figures

APPENDIX A: CHECKING THE ACCURACY OF THE X-RAY SPECTRAL ANALYSIS

In Section 3, we assumed a simple spectral model, an absorbed power law with photon index fixed to $\Gamma = 1.8$, as a representation of the X-ray spectra of the sources in our high-redshift sample, and performed a spectral analysis to derive the probability distribution of the intrinsic column density. Although the use of more complex spectral models is generally precluded by the limited number of counts in the X-ray spectra of the typically faint AGN in our sample, we check in this appendix for potential biases arising from this assumption using X-ray spectral simulations.

We assumed that the intrinsic X-ray spectra in cases of column densities $\log N_{\rm H} > 22$ are well represented by the MYTorus spectral model[19] (Murphy & Yaqoob 2009), which fully accounts for the transmission and scattered components, as well as the production of the iron K edge and K_α and K_β emission lines. We fixed the Galactic absorption to $\log N_{\rm H} = 20$ (the characteristic value towards the CDF-S; Kalberla et al. 2005), $\Gamma = 1.8$, the inclination angle $\Theta = 90^\circ$ (i.e. edge-on configuration), and the normalization of the scattered and line components to unity with respect to the transmitted component. The MYTorus model does not allow lower values of column densities, and therefore for $\log N_{\rm H} < 22$ we used the XSPEC *plcabs* model, which also takes into account Compton scattering. We assumed three values of observed soft-band flux, $\log F = -17$, -16 and -15, corresponding to sources close to the flux limit of the 7 Ms CDF-S, faint sources and bright sources (in deep X-ray surveys), respectively. We also considered three different redshifts ($z = 3$, 4 and 5) and eight different values of column density (from $\log N_{\rm H} = 21$ to 24.5 in steps of 0.5 dex). For each of the resulting 72 combinations of these parameters, we simulated 10 X-ray source and background spectra with a 6 Ms effective exposure (typical of the CDF-S), allowing for Poissonian fluctuations, using the *fakeit* command in XSPEC. The background spectrum is sampled from the real CDF-S background. We then fitted the simulated spectra with our simple model and derived the best-fitting column density and its errors at the 90 per cent confidence level.

Fig. A1 shows the best-fitting column densities against the simulated ones for the different redshift-flux combinations. The output values in cases of large simulated column densities are consistent with the input ones. When decreasing the simulated flux and increasing the redshift, low values of column density ($\log N_{\rm H} \lesssim 22.5$ and, less severely, $\log N_{\rm H} \sim 23$) become increasingly difficult to constrain. In particular, the output column densities of sources simulated with $\log N_{\rm H} \leq 22$ are never constrained to be lower than that value, which is the widely used threshold separating obscured and

---

[19] http://www.mytorus.com/

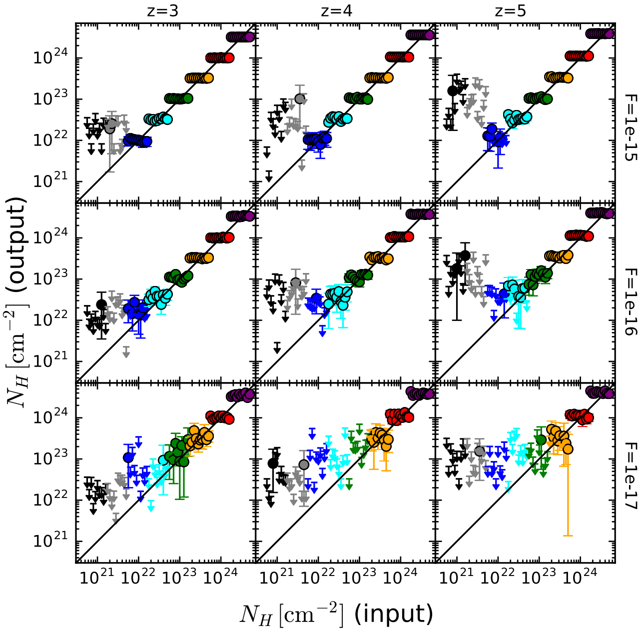

Figure A1. Best-fitting column density and 90 per cent CL errors derived by fitting a simple absorbed power-law model plotted against the simulated value, for different redshifts (columns) and fluxes (rows), as marked in the figure. Spectra are simulated using the MYTorus model and 8 different values of column density (in the range log$N_{\rm H}$ = 20–24.5, in steps of 0.5 dex, colour coded in the figure) as described in Appendix A. The log$N_{\rm H}$ positions are slightly shifted for presentation purposes. Black lines mark the 1:1 relations.

unobscured AGN at low redshift. This result is due to the photoelectric cut-off shifting close to, and even below, the low-energy spectral limit of the *Chandra* bandpass. Small differences in the intrinsic photon index (within the expected range of $\Gamma = 1.7$–2.0) do not affect significantly the output column densities. We therefore conclude that the spectral model we assumed, despite its simplicity, is a useful and appropriate approximation, at least for the redshift and flux ranges probed by this work.

APPENDIX B: THE X-RAY POWER-LAW PHOTON INDEX OF HIGH-REDSHIFT, SUB-L_* AGN

The photon index (Γ) of the primary power-law emission in AGN X-ray spectra can be used to estimate the Eddington ratio ($\lambda_{\rm Edd} = L/L_{\rm Edd}$) of a source through the Γ–$\lambda_{\rm Edd}$ relation (e.g. Shemmer et al. 2006, 2008; Risaliti, Young & Elvis 2009; Brightman et al. 2013, 2016; Fanali et al. 2013, but see the caveats discussed in Trakhtenbrot et al. 2017), according to which higher $\lambda_{\rm Edd}$ would produce steeper Γ. This relation is likely produced by the coupling between the accretion disc and the hot corona, and can therefore be used to investigate whether the accretion mode depends on the Eddington ratio at different redshifts.

The spectral analysis presented by Liu et al. (2017) on bright (>80 net counts) X-ray-selected AGN in the 7 Ms CDF-S reveals a median value of $\Gamma = 1.82 \pm 0.15$, typical of Seyfert galaxies. Evidence for steep photon indexes in a few luminous Type-I QSOs at $z > 6$ have been found by Farrah et al. (2004), Page et al. (2014) (whose results are disputed by Moretti et al. 2014) and Ai et al. (2016). However, studies of samples of $z = 4$–7 Type-I QSOs did not find a significant evolution of the average Γ in luminous QSOs with redshift (e.g. Shemmer et al. 2005; Vignali et al. 2005; Nanni et al. 2017). Flatter photon indexes ($\Gamma \approx 1.1$–2.0) have been derived by Vignali et al. (2002) for three spectroscopically confirmed $z > 4$, moderate-luminosity AGN in the CDF-N, probably due to the presence of intrinsic absorption.

Table B1. Best-fitting parameters of the spectral analysis described in Section B.

ID	Sample size	Γ	$N_{\rm H}$ (10^{22} cm^{-2})	log$L_{\rm X}$ (erg s^{-1})
		Individual fit to bright sources		
29	1	$2.14^{+0.95}_{-0.59}$	<7	43.52
207	1	$1.76^{+0.28}_{-0.14}$	<3	44.10
229	1	$1.62^{+0.33}_{-0.23}$	<4	43.70
293	1	$1.39^{+0.48}_{-0.33}$	<13	43.97
330	1	$2.43^{+0.83}_{-0.64}$	<19	44.20
404	1	$1.63^{+0.16}_{-0.12}$	<1	44.37
617	1	$1.79^{+0.41}_{-0.33}$	<9	43.45
774	1	$1.57^{+0.11}_{-0.11}$	3^{+2}_{-2}	44.57
788	1	$1.59^{+0.14}_{-0.14}$	<5	44.03
811	1	$1.51^{+0.18}_{-0.15}$	<2	43.77
921	1	$1.68^{+0.30}_{-0.26}$	<8	43.78
926	1	$1.36^{+0.39}_{-0.25}$	<11	43.68
965	1	$1.63^{+1.35}_{-0.49}$	<44	43.35
		Joint fit of faint sources		
–	3	$1.89^{+0.68}_{-0.46}$	–	42–43
–	6	$2.06^{+0.38}_{-0.32}$	–	43–43.75
		Joint fit of all sources		
–	3	$1.89^{+0.68}_{-0.46}$	–	42–43
–	11	$1.78^{+0.18}_{-0.17}$	–	43–43.75
–	9	$1.65^{+0.06}_{-0.06}$	–	43.75–44.75

Here, we compare the photon indexes of $L \lesssim L_*$ AGN with results derived for much more luminous objects. We selected a subsample of *bona fide* unobscured sources at $z = 3$–6 from Table 2, requiring that the best-fitting column density is log$N_{\rm H} < 23$ at the 68 per cent confidence level and a signal-to-noise ratio (SNR) > 3 in the 0.5–7 keV band. We performed a more detailed spectral analysis than that reported in Section 3, by allowing both the intrinsic column density and the photon index to vary. Fig. B1 (blue squares) shows the resulting best-fitting values of Γ as a function of intrinsic luminosity for sources with >110 net counts in the 0.5–7 keV band. This threshold was chosen in order to obtain reasonable uncertainties on the best-fitting Γ for individual AGN.

For the fainter sources, we derived the average photon indexes by performing a joint spectral analysis in luminosity bins chosen to include approximately the same number of sources. During the fitting, the column density was left free to vary for each spectrum, while the photon index parameters were linked amongst all sources, in order to derive an average value. Red circles in Fig. B1 present the results of this joint analysis for sources with <110 net counts. Finally, we performed the joint spectral analysis in three luminosity bins, including both bright and faint objects (black points). Fit results are reported in Table B1 for the individual fit of bright sources and for the joint fit of faint and all (bright and faint) sources.

Fig. B1 presents as continuous and dashed grey lines the median Γ and the standard deviation of the Γ distribution ($\Gamma = 1.82 \pm 0.15$), respectively, derived by Liu et al. (2017) for sources with >80 net

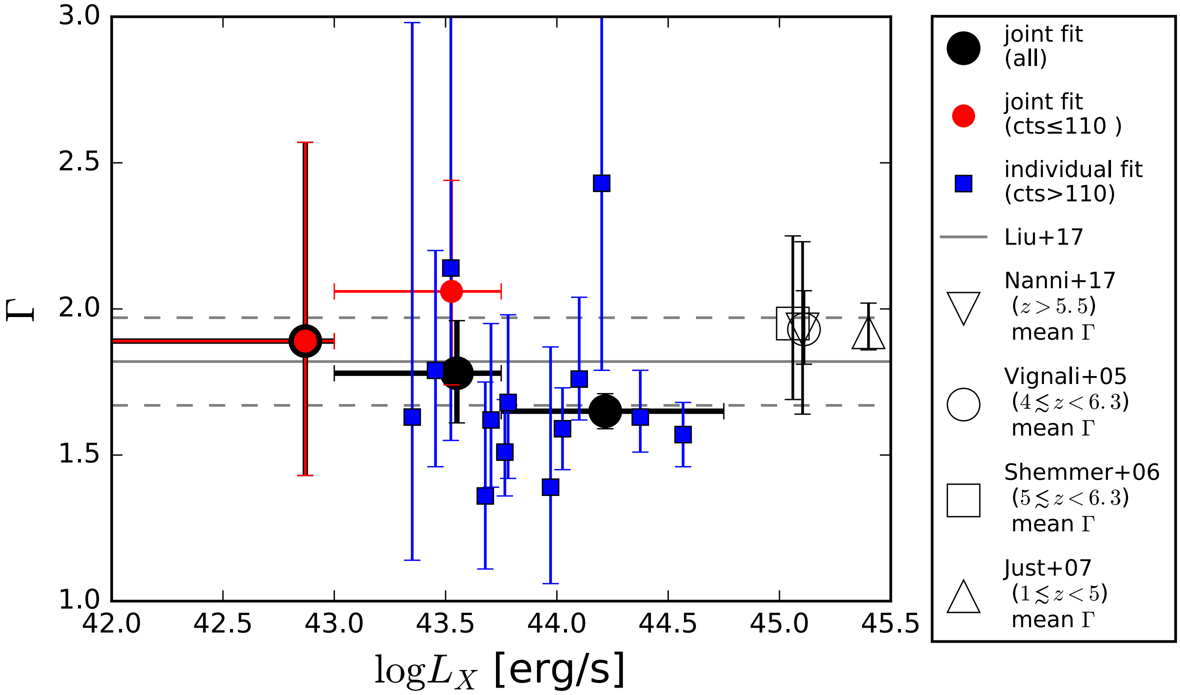

Figure B1. Best-fitting photon indexes for individual, bright sources with >110 net counts in the 0.5–7 keV band (blue squares), and the average values derived from a joint spectral analysis including all sources (black points) and only those with <110 net counts (red points), in three luminosity bins. No bright source lies in the first luminosity bin, where therefore the black and red points are coincident. Grey solid and dashed lines mark the median Γ derived by Liu et al. (2017) in the 7 Ms CDF-S and its standard deviation, respectively. We also show the average Γ of samples of luminous, optically selected Type-I QSOs in different redshift intervals.

counts in the 7 Ms CDF-S. Our sources are on average broadly consistent with such values, justifying the use of $\Gamma = 1.8$ for all sources in Section 3. The photon indexes of the luminous sources in our sample appear to be on average slightly flatter. This result is probably driven by the small number of luminous AGN considered here, which is due to the limited area covered by the deep surveys we used. Moreover, the best-fitting average Γ in the highest luminosity bin is quite strongly driven by the 1–2 brightest sources, which have $\Gamma \approx 1.6$ and alone provide the 38–59 per cent of the net counts for the joint fitting in this luminosity bin. Neither of these two sources are detected in radio bands, and, in general, discarding the minority of sources (2 out of 23) detected in radio surveys does not change the results.

Our findings at $42 < \log L_X \lesssim 44$ are compared in Fig. B1 with the average Γ of optically selected, luminous ($\log L_X > 45$) Type-I QSOs at low-to-intermediate redshift ($z =$ 1–5, Just et al. 2007) and $z =$ 4–7 (Vignali et al. 2005; Shemmer et al. 2006; Nanni et al. 2017). Our results are consistent with both the median value of Γ derived by Liu et al. (2017) for similar luminosities at all redshifts and the average values of luminous QSOs up to $z \approx 7$.

APPENDIX C: THE EFFECT OF NEGLECTING THE FULL PROBABILITY DISTRIBUTION OF SPECTRAL PARAMETERS

In Section 3, we derived the full probability distribution of flux, and luminosity by convolving the probability distribution of count rate and redshift, and using the proper count-rate-to-flux conversion for each source. Such information has been used in Section 5 to derive the dependence of the obscured AGN fraction on redshift, flux and luminosity.

To check the importance of applying this careful but computationally expensive procedure, we repeat the analysis in Section 5 by considering only the nominal value of parameters for each source, according to the following steps. (1) A single redshift ($z^{\rm peak}$), corresponding to the spectroscopic redshift or the peak of PDF(z), was assigned to each source. We did not include objects with no redshift information. (2) We adopted the values corresponding to the peak of the count rate (CR$^{\rm peak}$) and column-density ($N_{\rm H}^{\rm peak}$) distributions. (3) Applying the proper conversion factor for $z^{\rm peak}$ and CR$^{\rm peak}$, we derived single values of flux ($F^{\rm peak}$) and luminosity ($L^{\rm peak}$). Finally, following the procedures described in Section 5, we derive $F_{\rm obsc}$ as a function of redshift, flux and luminosity for our sample (Fig. C1). Larger fractions of obscured AGN are in general derived by this procedure, affecting in particular the trends with redshift and flux. This may be at least in part due to an overestimation of the best-fitting column density affecting intrinsically unobscured sources, due to statistical fluctuations in low-quality X-ray spectra discussed in section 3.3 of Vito et al. (2013). When considering the full probability distribution of column density, this issue is greatly alleviated. We also checked that, considering the $F^{\rm peak}$ and $L^{\rm peak}$ values, the number counts resulted not to be strongly different, while the XLF tends to be flatter at low luminosities. We again ascribe this behaviour to the tendency of overestimating the column densities, which has the effect of shifting the obscuration-corrected luminosities to larger values.

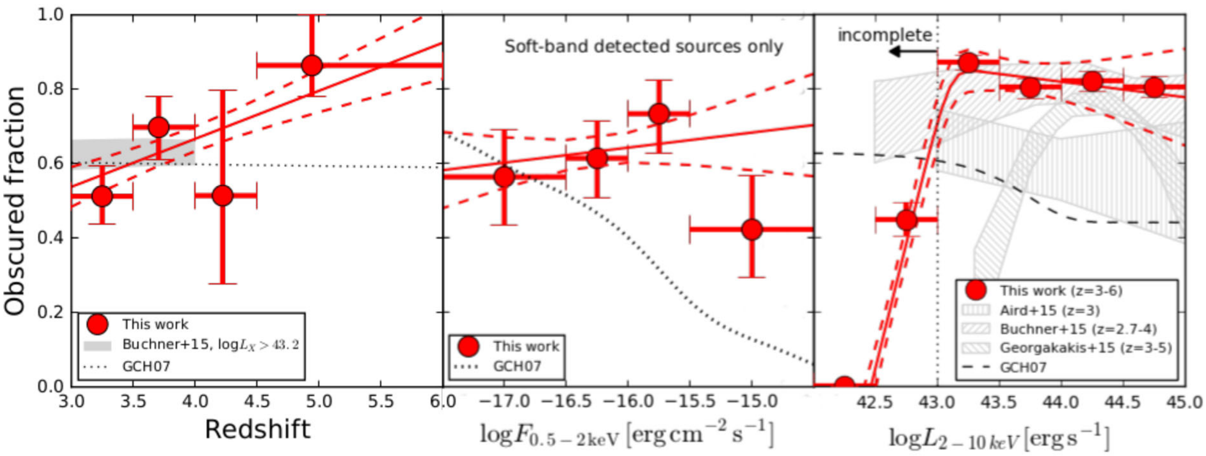

Figure C1. From left to right, obscured AGN fraction as a function or redshift, flux and luminosity, derived applying the procedure described in Appendix C. These figures, derived considering only the nominal values of redshift, count rate and column density for each X-ray source, are to be compared with the results in Section 5.

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