Identification of high-redshift X-ray active galactic nuclei in the 4XMM-DR11 serendipitous catalogue using DES data

Comparative analysis with optically selected QSOs

E. Pouliasis[1],[*] , A. Ruiz[1] , I. Georgantopoulos[1], A. Akylas[1], N. A. Webb[2] , F. J. Carrera[3] , S. Mateos[3],
A. Nebot[4] , M. G. Watson[5] , F. X. Pineau[4], and C. Motch[4]

[1] IAASARS, National Observatory of Athens, Ioannou Metaxa and Vasileos Pavlou 15236 Athens, Greece
[2] IRAP, Université de Toulouse, CNRS, CNES, 9 avenue du Colonel Roche, 21028 Toulouse, France
[3] Instituto de Física de Cantabria (CSIC-Universidad de Cantabria), Avenida de los Castros, 39005 Santander, Spain
[4] Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, 67000 Strasbourg, France
[5] School of Physics & Astronomy, University of Leicester, University Road, Leicester LE1 7RH, UK

ABSTRACT

X-rays provide a robust method in identifying active galactic nuclei (AGN). However, in the high-redshift Universe ($z \geq 3.0$), their space density is relatively low and due to the small areas covered by X-ray surveys, the selected AGN are poorly sampled. Deep optical and infrared (IR) data are essential for locating counterparts and determining redshifts. In this work, we leverage the *XMM-Newton* 4XMM-DR11 serendipitous catalogue (1240 deg^2), alongside the extensive optical Dark Energy Survey (DES, 5000 deg^2) and the near-IR VISTA Hemisphere Survey (VHS) to select one of the largest high-redshift X-ray AGN samples to date. Our analysis is focussed on the overlapping area of these surveys, covering about 185 deg^2. In addition, we aspire to compare the properties of the X-ray AGN with optically selected QSOs. For sources without spectroscopic data (~80%), we estimated the photometric redshifts using both spectral energy distribution (SED) fitting and machine-learning algorithms. Among the ~65 000 X-ray sources in the 4XMM-DES-VHS area, we ended up with 833 $z \geq 3.5$ AGN (11% having spec-z information) with a high level of reliability, along with a fraction of outliers of $\eta \leq 10\%$. The sample completeness is ~90%, driven by the depth of DES data. Only ~10% of the X-ray selected AGN are also optical QSOs and vice versa. Our findings indicate an observed absorbed fraction (log N$_H$ [cm^{-2}] ≥ 23) of 20–40% for the X-ray AGN, significantly higher than that of optical QSOs. X-ray AGN exhibit fainter observed optical magnitudes and brighter mid-IR magnitudes than optical QSOs. Their median rest-frame SED shapes differ notably with optical QSOs being dominated by AGN emission in the UV-optical wavelengths. While the median SEDs of X-ray AGN suggest extinction in the UV-optical range, individual sources exhibit a wide range of spectral shapes, indicating significant diversity within the population. This analysis supports the notion that X-ray-selected and optically selected AGN represent distinct and complementary populations.

Key words. methods: data analysis – galaxies: active – galaxies: high-redshift – quasars: supermassive black holes –
early Universe – X-rays: galaxies

1. Introduction

Almost all massive galaxies in the Universe host a supermassive black hole (SMBH) in their center (Magorrian et al. 1998; Kormendy & Kennicutt 2004; Filippenko & Ho 2003; Barth et al. 2004; Greene & Ho 2004, 2007; Dong et al. 2007; Greene et al. 2008). When matter from the galaxies starts to accrete onto the SMBH, an enormous amount of energy is released across the electromagnetic spectrum (from radio emission up to X- and γ-rays). This constitutes the characteristic signature of active galactic nuclei (AGN).

Despite the difference in the physical scale between the SMBH and the galaxy spheroid that is about nine times orders of magnitude, there is a tight correlation between the masses of the SMBH and the galaxy bulge (Silk & Rees 1998; Magorrian et al. 1998; Ferrarese & Merritt 2000; Gebhardt et al. 2000). In addition, many studies suggest a close interaction between the growth and evolution of galaxies and that of SMBHs

(Silk & Rees 1998; Granato et al. 2004; Di Matteo et al. 2005; Hopkins et al. 2006, 2008; Menci et al. 2008). In particular, there is cumulative evidence for a similar cosmic evolution of the SMBH accretion and the star-formation (SF) rate density, both peaking at cosmic noon (redshift z = 1–3, Madau & Dickinson 2014; Aird et al. 2015; Harikane et al. 2022). However, at higher redshifts (z = 4−5), the AGN growth (in terms of black hole accretion density; i.e. total black hole growth rate per comoving volume) presents a much stronger decline over redshift compared to the SF rate density (SFRD; i.e. the amount of stars formed per year in a unit of cosmological volume). Thus, galaxy growth may have preceded the build up of their central SMBHs in the early Universe (Aird et al. 2015). Alternatively, the SMBH may have been massive enough when it formed and would not require high accretion rates to reach the local $M_{BH} - M_\star$ relation. Additionally, it is possible that a population of primarily obscured AGN at high redshifts remains undetected, which could contribute to the observed decline in AGN growth. Some simulations (Ni et al. 2020; Habouzit et al. 2021; Ni et al. 2022; Gilli et al. 2022) have indeed suggested the existence of such a population, which could

[*] Corresponding author: epouliasis@noa.gr

help reconcile the observed trends in SMBH accretion and SFRD at high redshifts.

Most theoretical galaxy evolutionary models today predict a regulating mechanism between the AGN power and the SF of the host galaxy, leading to the formulation of the AGN-galaxy co-evolution paradigm (e.g. Hopkins et al. 2007; Lapi et al. 2014, 2016). In this scenario, merging systems are responsible for intense SF activity. The large fraction of gas reservoirs, then, is fuelled towards the SMBH and it initiates AGN activity. This phase is characterised by the growth of both the stellar mass of the galaxy and the SMBH. Since there are large quantities of gas, the AGN emission is absorbed and the sources appear as obscured AGN that last for a long period of time. There are several studies indicating that a larger fraction of obscured AGN are found in mergers or post-merger galaxies (Georgakakis et al. 2008; Satyapal et al. 2014; Lanzuisi et al. 2015; Ellison et al. 2019; Secrest et al. 2020). X-ray observations of optically dual-AGN in late-state mergers also showed this trend (e.g. De Rosa et al. 2023). More recent studies with the James Webb Space Telescope (JWST) support further the merging scenario. For example, Bonaventura et al. (2024) studied the morphology of the host galaxies of AGN with different obscuration levels and concluded that when going from the unobscured and moderately obscured to highly obscured AGN samples, the fraction of disturbed galaxies increases from 63 to 95%. As the SF and AGN activity consumes, heats up, and expels gas, the AGN becomes less obscured and it eventually becomes powerful enough to push away the surrounding material and turn into an unobscured type 1 AGN (Ciotti & Ostriker 1997; Hopkins et al. 2006; Hopkins & Hernquist 2006; Somerville et al. 2008; Blecha et al. 2018).

However, these theoretical models are in contrast to the standard unification model, where the AGN obscuration depends only on the inclination angle related to the line of sight (e.g. Antonucci 1993; Netzer 2015; Ramos Almeida & Ricci 2017). In addition, there are several studies that suggest that the obscured phase of the AGN is not the beginning in this scenario; the dust-obscured AGN are not, in fact, young quasars. For example, Georgantopoulos et al. (2023) compared the relative ages of the host galaxies of the absorbed and unabsorbed AGN and found that the absorbed AGN lie in intermediate-age galaxies between the young and old populations.

Hence, the physical processes governing the AGN-galaxy relationships are not yet fully understood. Therefore, to validate the several evolutionary models of BHs and whether AGN affect their host galaxy properties, it is crucial to have a complete (as much as possible) census of AGN demographics. In recent decades, even though a huge progress has been made in the construction of the BH demographics (especially at $z \leq 3$), there is a large fraction (~75%) of highly obscured AGN predicted by the X-ray background (XRB) synthesis models that are currently remain undetected in X-rays (Barchiesi et al. 2021). The identification and classification of such sources becomes even more complicated in the high-z regime, since the obscuration (both the obscuring medium within the AGN, the so-called torus, and the larger-scale dust of the galaxy) increases with redshift, reaching a plateau at z ~ 3−4 (Gilli et al. 2022; Signorini et al. 2023; Pouliasis et al. 2024).

At redshifts higher than $z = 3$, the expected density of luminous QSOs ($M_{1450Å,rest−frame} \leq −24$) is very low (~1 Gpc^{-3}, De Rosa et al. 2014). Therefore, large areas need to be probed. So far, X-ray surveys have harvested limited high-z AGN samples because of the limited sky area covered. In particular, using both Chandra and XMM-Newton observations, several studies (e.g. Vito et al. 2014; Georgakakis et al. 2015; Marchesi et al. 2016a;

Vito et al. 2018; Khorunzhev et al. 2019; Pouliasis et al. 2022a; Peca et al. 2023) have identified a few hundred $z > 3$ X-ray sources. In addition, the ongoing all-sky extended ROentgen Survey with an Imaging Telescope Array (eROSITA, Predehl et al. 2021) is expected to facilitate the search for high-redshift AGN owing to its large grasp (field-of-view multiplied by effective area). After four years of operation, a few examples of very luminous $z > 6$ AGN have been identified with the eROSITA detector, with the majority of them also being radio-loud (e.g. Medvedev et al. 2020; Wolf et al. 2021, 2023, 2024). The X-ray telescope onboard the *Gehrels/SWIFT* mission has provided yet another $z > 6$ AGN (Barlow-Hall et al. 2023). By combining more than 600 Chandra and XMM-Newton high-z AGN, Pouliasis et al. (2024) derived the most up to date X-ray luminosity function (XLF) indicating that about 60% of the sources are obscured with hydrogen column densities of $N_H \geq 10^{23}$ cm^{-2}.

Although their derived AGN space density and black hole accretion rate density (BHAD) are in agreement with the large-scale cosmological hydrodynamical simulations (e.g. Volonteri et al. 2016; Habouzit et al. 2022), when considering the findings that X-ray AGN are hosted by massive galaxies, these results contrast with those obtained from mid-IR JWST data. For example, Yang et al. (2023) used JWST observations and identified AGN through spectral energy distribution (SED) decomposition with a significantly higher density. In particular, the BHAD of the JWST-selected AGN is approximately three times larger than that of the X-ray-selected AGN. This discrepancy suggests the existence of an X-ray AGN population that may be missing due to either heavy obscuration or low intrinsic luminosity. However, more recently a similar JWST-MIRI study by Lyu et al. (2024), which utilised deeper ancillary data and more accurate redshifts (~40% spectroscopic) over a larger field, reported a significantly lower AGN number density of ~6.3 arcmin^{-2}, compared to the ~14 arcmin^{-2} found by Yang et al. (2023). Furthermore, Lyu et al. (2024) suggested that the latter higher AGN density may be due to contamination by low-mass galaxies and methodological differences, such as the treatment of PAH emission.

Furthermore, the ability of the optical/near-IR surveys to scan large areas in the sky (i.e. ~10^4 deg^2) led to the discovery of a large sample of high-z broad-line QSOs increasing the diversity between the AGN populations in the early Universe. Shallow surveys such as SDSS (Jiang et al. 2016), the UKIRT Infrared Deep Sky Survey (UKIDSS, Mortlock et al. 2011), the Canada-France High-redshift Quasar Survey (CFHQS, Willott et al. 2010), and the Panoramic Survey Telescope & Rapid Response System (Bañados et al. 2018) played a significant role in selecting more than 400 confirmed spectroscopically high-z ($z \gtrsim 5.5$) QSOs (Fan et al. 2023). However, only (~10%) have been observed with deep X-ray observations and, out of those, about 70% have been detected (Vito et al. 2019; Lusso et al. 2023; Wolf et al. 2024). Deeper optical surveys with the Subaru Hyper Supreme Cam, HSC, (Miyazaki et al. 2017) such as the HSC Subaru Strategic Plan Survey (Aihara et al. 2018, 2019) allowed for the detection of AGN at redshifts of $z = 3−6$ at much fainter (>3 mag) absolute magnitudes (Akiyama et al. 2018; Niida et al. 2020; Matsuoka et al. 2022), compared to previous SDSS surveys. More recently, Yang & Shen (2023) released a sample of optically selected QSO candidates using the DES catalogue with high reliability (~95%). Among them, there are more than 40 000 optically selected QSO candidates at $z \geq 3$ in an area of 5000 deg^2.

Besides the limited sky areas covered by the X-rays, another obstacle faced by such studies is the lack of deep optical and/or near-IR photometry suitable for identifying the correct
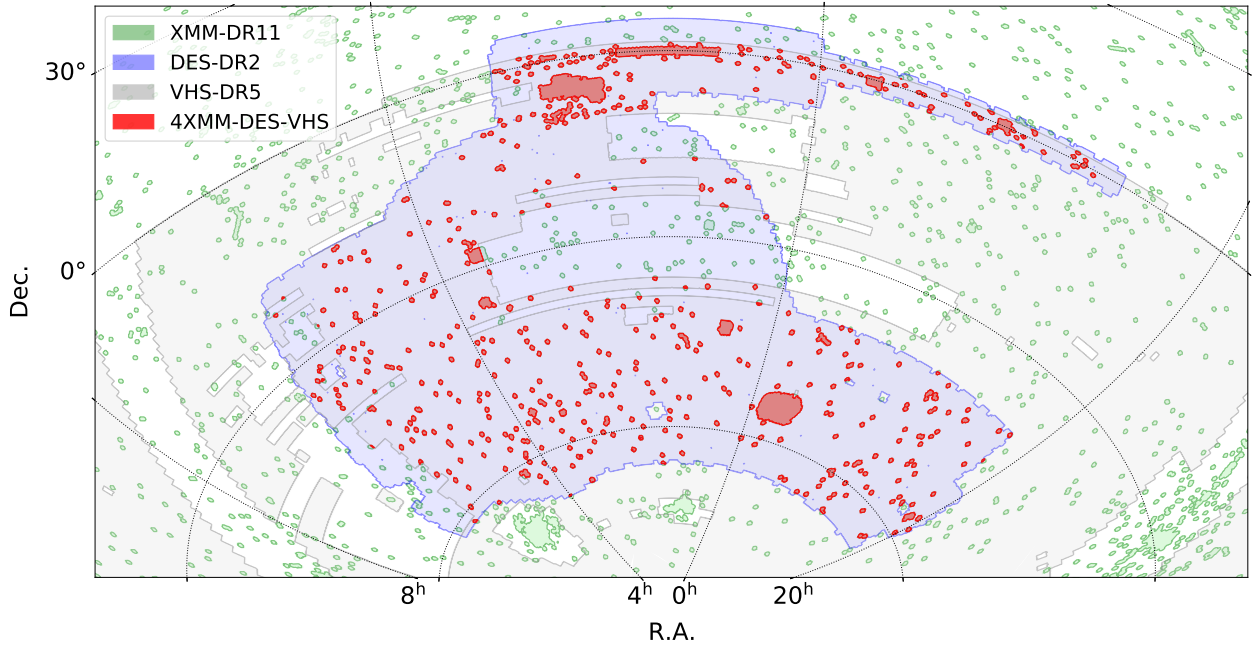
Fig. 1. Survey footprint of the 4XMM-DES-VHS catalogue (red). In addition, we separately show the coverage of the 4XMM-DR11, DES-DR2, and VHS-DR5 surveys in the sky area of our interest.

optical/NIR counterpart of the X-ray sources and deriving their photometric redshifts through SED fitting techniques. In this work, we aim to select high-z X-ray AGN using the wide-area 4XMM-DR11 catalogue (Webb et al. 2020) and compare this sample to the optically selected QSOs by Yang & Shen (2023). In particular, we aim to investigate whether these two AGN populations are intrinsically different. The 4XMM-DR11 catalogue covers approximately 1240 deg^2 of the sky, while the second data release of the Dark Energy Survey (DES-DR2, Abbott et al. 2021) spans around 5000 deg^2, reaching faint magnitudes down to mag$_i \sim 25$. Additionally, the VISTA Hemisphere Survey (VHS, McMahon et al. 2013) has been included in this analysis. The combined overlapping area between the X-ray and optical surveys, along with VHS, covers roughly 185 deg^2, providing an excellent opportunity to directly compare X-ray AGN to optically selected QSOs.

The data used in our analysis are presented in Sect. 2. Sections 3 and 4 describe the procedure followed to construct a reliable high-z X-ray sample discussing its completeness. In Sect. 5, we discuss the X-ray properties of the high-z sample along with the multi-wavelength properties derived using the SED fitting and we compare the observational properties of the X-ray AGN, with respect to the optically selected broad-line QSO sample. In Sect. 6, we summarise our results. Throughout the paper, we assume a ΛCDM cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$ (Komatsu et al. 2009).

2. Data

2.1. 4XMM-DES-VHS catalogue

In this work, we used the 4XMM-DR11[1] catalogue that contains all the serendipitous X-ray sources from the European Space Agency's (ESA) *XMM-Newton* observatory. 4XMM-DR11 includes data from 12 210 *XMM-Newton* observations obtained

by the European Photon Imaging Camera (EPIC, Strüder et al. 2001; Turner et al. 2001), which were made between 2000 February 3 and 2020 December 17, covering an energy range of 0.2–12 keV. The catalogue has been produced through the *XMM-Newton* Survey Science Centre (SSC) on behalf of ESA. In total, there are 895 415 detections in these observations. The 'slim' version of the catalogue contains 602 543 unique X-ray sources, where multiple detections of the same physical source appear only once. For the majority of the analysis in this paper, we utilised the 'slim' version of the catalogue. However, we utilised the detection catalogue to calculate the total area curve and derive the cumulative number counts of the X-ray sources (Sect. 4.1), as well as to determine the X-ray spectral properties of the high-z sample (Sect. 5.1).

The classification of X-ray sources used in this work was performed by the XMM2ATHENA project[2]. Following the classification work using a newly developed Naive Bayes classifier, presented in Tranin et al. (2022), the XMM2ATHENA team further refined the algorithm and applied it to the 4XMM catalogue (Webb et al. 2020). This updated algorithm is capable of identifying sources as an AGN or galaxy, extragalactic and galactic X-ray binaries (XRBs), stars, cataclysmic variables (CVs), and extended sources, enabling the distinction between Galactic and extragalactic sources. To enhance reliability, the classification was augmented with data from multi-wavelength catalogues and validated through manual screening and testing with sources of known types. Among the sources with a secure classification (99.9%), we selected only those flagged as AGN or QSOs, representing nearly 95% of the total X-ray sources. The false-positive rate of sources misclassified as AGN is estimated to be approximately 5%. However, in the high-z regime, this percentage increases, as discussed further in the next section. The sample of point-like X-ray-selected AGN and QSOs inside the 4XMM-DES-VHS footprint consists of 75 782 sources.

---

[1] http://xmmssc.irap.omp.eu/Catalogue/4XMM-DR11/4XMM_DR11.html

[2] http://xmm-ssc.irap.omp.eu/xmm2athena/wp-content/uploads/2023/04/DeliverableD8.2.pdf
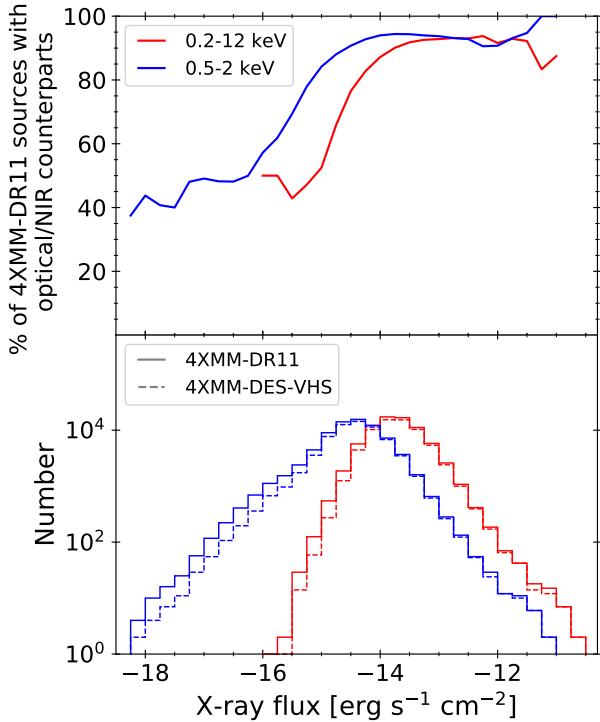
Fig. 2. Upper panel: percentage of 4XMM-DR11 sources with an optical or near-IR counterpart as a function of X-ray flux in the soft (0.5–2 keV, blue) and the full band (0.2–12 keV, red). Lower panel: flux distributions in the soft band and full bands of the 4XMM-DR11 and 4XMM-DES-VHS samples.

For the optical photometry, we used the second data-release of the dark energy survey (DES-DR2, Abbott et al. 2021) that contains photometry in five broad bands (*g, r, i, z,* and *Y*), while for the NIR photometry (*J, H, Ks* bands), we made use of the fifth release of the VISTA Hemisphere Survey (VHS-DR5, McMahon et al. 2013). We only used the DES-DR2 bands with reliable photometry: $Flags \leq 4$ (well-behaved objects), $imaflags_{iso} = 0$ (no missing or flagged pixels in the source in all single epoch images), and $N_{iter} = 0$ that ensures no missing imaging data. In addition, the 4XMM-DES-VHS footprint (Fig. 1) has taken into account the DES DR2 masking regions (Abbott et al. 2021), which exclude sources near bright stars that may be affected by diffraction spikes or high-background counts, near bleed trails, and artificial satellite streaks. The selection of the optical/NIR counterparts of the X-ray sources was achieved under the framework of the H2020 XMM2ATHENA project (Webb et al. 2023) using the ARCHES statistical analysis (Motch et al. 2017) for multi-wavelength cross-correlation and cross-identification based on astrometric criteria (Pineau et al. 2017). By selecting the counterparts with the highest probability being the correct one, we ended up with 66 798 X-ray sources with an optical and/or NIR counterpart. The completeness with respect to the whole X-ray catalogue inside the DES/VHS footprint is 88%.

In Fig. 2 (upper panel), we show the percentage of the 4XMM-DR11 sources with an optical/NIR counterpart versus the X-ray flux in the full band (0.2–12 keV, red). The lower panel presents the flux distributions of the 4XMM-DR11 and 4XMM-DES-VHS samples. It is important to note here that the 4XMM catalogue used in our analysis contains detections with a significance greater than ~3σ in the full band, which corresponds to a maximum likelihood of 6 (Webb et al. 2020). For
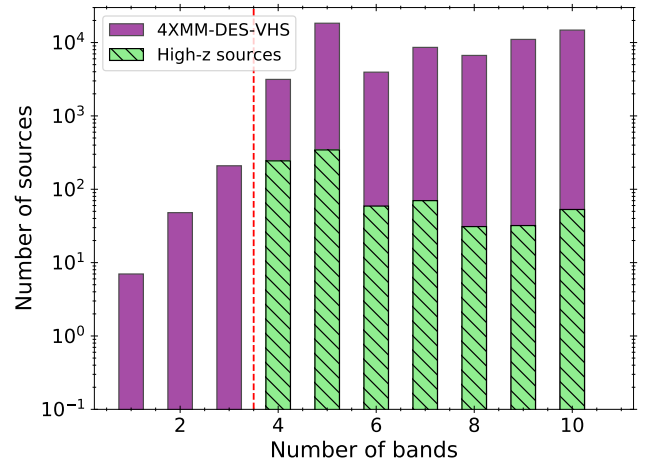

Fig. 3. Number of sources versus number of photometric bands for the 4XMM-DES-VHS (purple) and the final high-z (hatched, Sect. 4) samples. The vertical dotted line corresponds to the cut at N ≥ 4 we used in our analysis.

73% of the sources in our sample, the significance is greater than 5σ (maximum likelihood of 14). For reference, we also present the percentage of the soft band (0.5–2 keV, blue) sources with an optical/NIR counterpart that was used for the high-z sample incompleteness estimation (Sect. 4.2). At the bright end ($S_{0.5-2 \text{ keV}} > 10^{-15}$ erg s^{-1} cm^{-2}, which represents 85% of 4XMM-DR11), the completeness is about 90%, however at fainter fluxes it drops down to 50%. Regarding the tail of sources with a soft X-ray flux below 1×10^{-16} erg cm^{-2} s^{-1}, these constitute less than 2% of the full sample and for more than half of these, the significance remains above 5σ. As mentioned above, the 4XMM catalogue was build upon the full band detections, hence we expect some hard sources to have very faint fluxes in the soft band.

Furthermore, we complemented the multi-wavelength catalogue with mid-IR photometry from the Wide-field Infrared Survey Explorer mission (WISE; Wright et al. 2010). In particular, we cross-matched the optical/NIR counterparts to the X-ray sources with the ALLWISE source catalogue (Cutri et al. 2013) using a radius of 2″ following Fotopoulou et al. (2016). More information on the datasets and the cross-matching procedure will be available in Ruiz et al., (in prep.). We acknowledge that no explicit correction for the different resolutions was applied during the positional cross-matching of photometric data points. Although this may introduce some uncertainties in identifying the correct MIR counterparts, particularly for blended sources in WISE images, the overall effect on the photometric redshift solutions is expected to be minimal. However, we estimated that only about 5.4% of WISE sources are likely blended with multiple counterparts within the 2″ matching radius, based on a self-cross-match of the optical/NIR catalogue. This low blending fraction minimises the potential impact on the photo-z solutions, as the majority of sources remain uncontaminated by resolution-related effects. Finally, to construct reliable SEDs with the available photometry, to derive the photometric redshifts, and also to decompose the AGN and the galaxy components, we required for all sources to have at least four bands available. Figure 3 presents the number of sources versus the number of bands available for the 4XMM-DES-VHS sample. The vertical line corresponds to the threshold on number of bands we set (N ≥ 4). This cut excludes almost 5% of the 4XMM-DES-VHS sample. The majority of the sources (94%) do have detections in

all five DES bands. This stands in contrast to the infrared (IR) bands, where 60% (52%) of the sources have at least one near-IR (mid-IR) band available. This mainly is due to the different depths of the optical compared to near- and mid-IR (NIR and MIR) photometry. Hence, we ended up with a sample of 66 524 X-ray sources in the 4XMM-DES-VHS footprint to be used in the selection of high-z sources.

2.2. Spectroscopic redshifts

Inside the 4XMM-DES-VHS area, there is a plethora of optical spectroscopic information originating from different surveys targeting extra-galactic sources, including both AGN and galaxies. For this work, we made use of the Millions of Optical-Radio/X-ray Associations (MORX) Catalogue (Flesch 2016, 2024). In particular, we used the second version of the MORX catalogue that includes 3 115 575 optical objects with radio or X-ray associations. About 20% of the MORX sources have a spectroscopic redshift. The 4XMM-DR13 catalogue (Webb et al. 2020) is one out of the several X-ray surveys used to build this catalogue. We cross-matched this dataset to the 4XMM-DES-VHS catalogue using the optical/NIR coordinates and a searching radius of 1″. Among the 32 353 MORX associations, there are 13 424 X-ray sources with available spectroscopy.

2.3. Photometric redshifts

In Ruiz et al., (in prep.), we derived photometric redshifts for the 4XMM-DES-VHS sources using both machine learning (ML) algorithms and SED fitting methods. Specifically, we applied the Trees for Photo-z (TPZ) algorithm (Carrasco Kind & Brunner 2013), which is based on random forest techniques, and the LePHARE code (Arnouts et al. 1999; Ilbert et al. 2006). For LePHARE, it is essential to run the code with different SED templates depending on the morphological type of the source (point-like or extended) to prevent parameter degeneracies in the models (e.g. Pouliasis et al. 2024). To classify the sources according to their morphological type, we used the parameter EXTENDED_COADD provided in the DES-DR2 catalogue. Following Abbott et al. (2021), we selected point-like and extended sources using EXTENDED_COADD < 2 and EXTENDED_COADD ≥ 2, respectively. In addition, we applied the appropriate luminosity priors in each case. In particular, we used an absolute magnitude of $-23 < M_{DESg} < -8$ for the extended and $-30 < M_{DESg} < -20$ for the point-like sources (Pouliasis et al. 2024, and references therein). To assess the accuracy of photometric redshifts derived from TPZ and LePHARE, we used the sub-sample of X-ray sources with available spectroscopic redshifts (Sect. 2.2), from which we calculated the percentage of outliers, η (Ilbert et al. 2006; Laigle et al. 2016), defined as:

$$\eta(\%) = \frac{N_{outliers}}{N_{total}} \times 100, \qquad (1)$$

where N_{total} is the total number of sources and $N_{outliers}$ is the number of the outliers. An object is defined as an outlier if it has $|\Delta z|/(1 + z_{spec}) > 0.15$, where $\Delta z = z_{phot} - z_{spec}$. Figure 4 illustrates the percentage of outliers as a function of redshift for both point-like and extended sources using the TPZ and LePHARE algorithms. For each spectroscopic redshift, we calculate the percentage of outliers within the range $z_{spec} \pm 0.2$. This range is used to evaluate the number of outliers within a window around the spectroscopic redshift. It is independent of the $|\Delta z|$
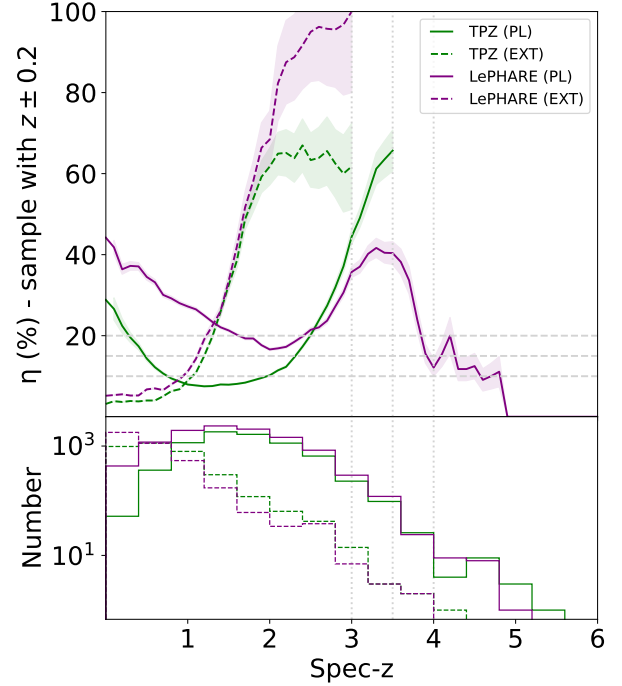

Fig. 4. Percentage of outliers versus spectroscopic redshift (upper panel). The solid (dashed) lines correspond to point-like (extended) sources for the TPZ (green) and LePHARE (purple) results. Shaded regions represent the uncertainty in the percentage of outliers at each spec-z bin, calculated as the percentage of outliers divided by $\sqrt{N}$, where N is the number of sources in each bin. Larger sample sizes result in narrower shaded regions. The bottom panel shows the distribution of the training sample with spectroscopic redshifts in each case.

definition, which measures the difference between photometric and spectroscopic redshifts for individual sources.

ML algorithms have proven highly effective and accurate for estimating photometric redshifts using broad-band photometry, but they require large training sets of sources with spectroscopic redshifts. Consequently, ML algorithms like TPZ are very powerful at $z < 3$, but their performance diminishes beyond this redshift due to a lack of spec-z data. To overcome this limitation, SED fitting algorithms such as LePHARE can be used, which is why we decided to rely on LePHARE's photometric redshift estimates for all X-ray sources without spec-z information. However, in the next section we make use of the secure and accurate TPZ results to filter out low-z galaxies from the LePHARE high-z sample, ensuring a more reliable selection of high-redshift sources.

3. Selection of the high-z sample

3.1. Identification of high-z sources

Using the MORX spectroscopic catalogue mentioned in Sect. 2.2, we selected all the X-ray sources with $z \geq 3.5$. Out of 13 424 X-ray sources with spectroscopic redshift, there are 92 AGN at $z \geq 3.5$. Among them, 30 sources are found at $z \geq 4$, while three sources are at $z \geq 5$. For sources that lack spectroscopic information, we used the photometric redshift estimations.

The threshold of $z = 3.5$ was set for the selection of high-redshift sources to have a low percentage of outliers in the photometric-redshift sample. In Fig. 5, we plot the percentage
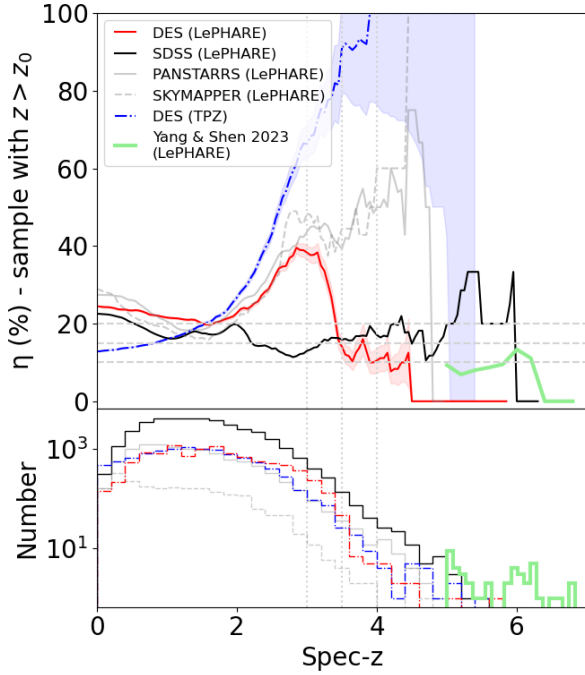
Fig. 6. Redshift distribution of the 4XMM-DES-VHS sample used in our analysis (gray filled). We highlight the spectroscopic and photometric redshift samples as indicated in the legend.

Fig. 5. Percentage of outliers versus spectroscopic redshift (upper panel) for the point-like sources. At each redshift, we calculate the percentage of outliers for all sources above this redshift. The bottom panel show the distribution of the training samples with available spectroscopic information. The performance in the DES field using TPZ (blue-dashed-dotted) and LePHARE (red-solid) is compared across different redshifts. For reference, we show the performance of LePHARE algorithm for the 4XMM-DR13 sources using different optical photometry adopted from Ruiz et al., (in prep.), such as SDSS (black-solid), PANSTARRS (gray-solid) and SKYMAPPER (gray-dashed). Finally, we present the LePHARE performance of $z \geq 5$ sources of Yang & Shen (2023).

of outliers versus the spectroscopic redshift (upper panel, red line). At each spec-z value, we estimate the percentage of outliers for all the sources above this redshift. The bottom panel shows the histogram of the spec-z sources. Around $z = 3$, we observe a noticeable bump corresponding to an increased percentage of outliers, which is expected due to template degeneracies. This issue is primarily caused by the absence of the u-band in the optical (DES) photometry. For comparison, we also show the results for 4XMM-DR13 sources with SDSS photometry (Ruiz et al., in prep.), which includes the u-band, and the bump disappears. Consequently, in the absence of the u-band, one should target sources beyond $z \sim 3.2$ to reliably select high-redshift objects. For sources with $z > 3.5$, DES provides statistically higher accuracy compared to SDSS due to its greater depth (the median 5σ depths of DES-DR2 and SDSS in i-band are $i_{DES} = 24.5$ mag and $i_{SDSS} = 22.2$ mag, respectively). To further emphasise this point, we also present results using PAN-STARRS (Chambers et al. 2016) and Skymapper (Wolf et al. 2018; Onken et al. 2019) optical photometry (Ruiz et al., in prep.). Both of these surveys, like DES, lack the u-band but are also much shallower. In addition to the bump observed around $z = 3$, we see an increasing trend in the percentage of outliers for $z > 3.5$, highlighting the importance of deep optical data.

For the DES dataset, the redshift accuracy improves significantly at $z > 3.5$, with the outlier fraction dropping below 12%. This percentage further decreases to approximately 10% for sources with $z > 4$ and approaches zero for sources with $z > 5$.
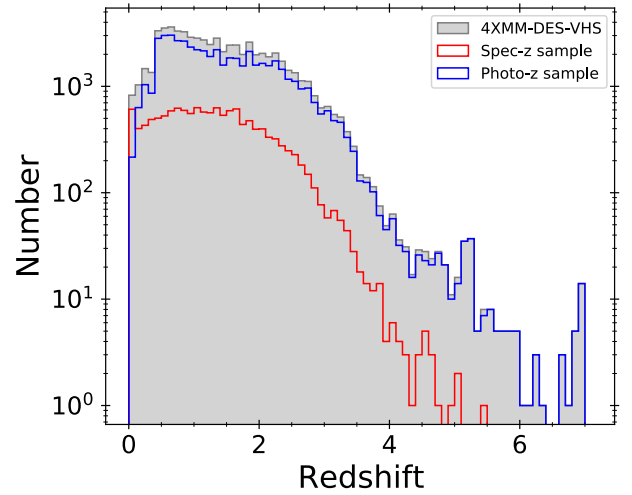
This behaviour is expected, as at higher redshifts ($z > 4$), low-redshift galaxies are less likely to contaminate the sample due to the spectrum shift. However, since our spectroscopic sample for $z > 5$ is limited, we followed a similar approach to Saxena et al. (2024) to evaluate the performance of the LePHARE algorithm in this regime. In particular, they used the CIRCLEZ algorithm to derive photometric redshifts for X-ray selected AGN and relied on an optical QSO sample for training in the absence of high-z X-ray spectroscopic data. Similarly, we tested the performance of LePHARE at $z > 5$ using the Yang & Shen (2023) sample. Specifically, we ran LePHARE on 43 sources with $z > 5$, of which only four were identified as outliers. Figure 5 shows the fraction of outliers versus redshift for this sample, where the outlier fraction is approximately 10%, dropping further at $z > 6$. This represents an upper limit, as the sample consists of optically selected QSOs dominated by central engine emission. In contrast, as discussed later, we expect a lower outlier fraction for X-ray-selected sources, where the host galaxy emission is more prominent. At the end, out of the 66 524 X-ray sources in our sample, there are 972 sources with $z \geq 3.5$ obtained by the LePHARE algorithm without spectroscopic information.

3.2. Low-z galaxy interlopers

From the percentage of outliers for the test sample discussed above, we would expect less than 12% of the sources with photometric high redshifts to be low-z galaxy interlopers. As noted earlier, ML algorithms for photometric redshift estimation are highly accurate for z < 3 (Fig. 5, left). To address this, we utilised the most reliable TPZ results, where reliability is defined by the peak strength (PS) value of the probability density function of redshift (Ruiz et al. 2018). Hence, we used all sources with PS > 0.7 (Ruiz et al., in prep.) to identify and exclude these interlopers from the high-redshift sample. This threshold was chosen, as per the study of Ruiz et al. (2018), to balance the trade-off between sample purity and completeness. While this criterion does not entirely eliminate the possibility of secondary redshift solutions at lower redshift, it significantly reduces their likelihood, thereby enhancing the reliability of our high-redshift sample.

In total, we identified 161 X-ray sources, which accounts for about 17% of the photometric high-z sample, which serves as a lower limit. Figure 6 shows the redshift distribution for the

whole sample (shaded gray area). The red and blue histograms represent the spectroscopic and photometric redshift samples, respectively. In the high-redshift regime, there are two abnormal overdensities at $z \sim 5.2$ and $z \sim 6.9$. We examined all the information available and found that the first one is due to stellar objects misclassified as AGN. The second overdensity is mainly due to imaging issues and/or sources being near very bright objects that passed our criteria initially Sect. 2. The next two sections describe the removal of these objects from our high-z sample.

3.3. Stellar contamination

Although we initially selected the AGN/QSO class for our analysis, we expect about 5% of these sources to be stars and misclassified as an AGN/QSO (Tranin et al. 2022). In order to remove the remaining stellar population contaminating our high-z AGN sample (first overdensity at $z \sim 5.2$), we utilised multiple criteria for sources with point-like morphology in the optical images. Firstly, we used the *Gaia* source catalogue (Gaia Collaboration 2016, 2023) and removed all the sources with high proper motion (PM $\geq$ 5). To take into account the uncertainties (pmra_error and pmdec_error) and the correlation (pmra_pmdec_corr) of the PM components along the equatorial coordinates we followed the definition of Carnerero et al. (2023):

$$PM_{corr} = \sqrt{\frac{\alpha^2 + \beta^2 - 2\alpha\beta\gamma}{1 - \gamma^2}} \geq 5, \qquad (2)$$

where α = pmra/pmra_error, β = pmdec/pmdec_error, and γ = pmra_pmdec_corr. In addition, we excluded sources with $\log(PM_{corr}) - \log(F_X/F_r) \geq 1.5$, where F_X/F_r represents the X-ray (0.5–2 keV band) to optical flux ratio. This criterion is based on the $\log(PM_{corr})$ and $\log(F_X/F_r)$ plane, where stellar and AGN populations are distinct. The value of 1.5 corresponds to the optimal separation between these two populations.

However, the *Gaia* matches to our catalogue consist of less than 25% of the whole 4XMM-DES-VHS sample. Hence, for sources without *Gaia* information, we made use of the probability of a source being a star from the XMM2ATHENA classification algorithm, PbaC1 (Tranin et al. 2022), and then the difference in the χ^2 between the best fit of AGN/QSO versus stellar templates. For the latter, we ran the LePHARE algorithm with stellar templates and calculated the difference as $\Delta\chi^2 = \chi^2_{AGN} - \chi^2_{star}$. A negative $\Delta\chi^2$ indicates that the AGN model provides a better fit to the data compared to the stellar model, while a positive or small $\Delta\chi^2$ suggests that the stellar model is preferred or both models fit equally well. The significance of the $\Delta\chi^2$ is assessed by comparing it to a threshold, beyond which we can confidently reject the stellar model in favor of the AGN model. To define the thresholds for both these criteria, in Fig. 7 (top panel) we plot $\log(PM_{corr}) - \log(F_X/F_r)$ versus $\log(PM_{corr})$. The dashed lines correspond to the *Gaia* criteria mentioned above to select a reliable stellar sample. The stellar population is located in the top right corner of this plot (shaded region). The bottom panel of Fig. 7 corresponds to the high-z sample. Based on these, we set a cut-off at $\Delta\chi^2 = 5$ (black squares) and also a logPbaC1 = −1.25. Sources that have values greater than these are considered to be stars. By using all four criteria, we were able to exclude in total 1854 stars, that is, about 2% of the point-like sample. Among them, there are 106 sources that were falsely identified as AGN with a photometric redshift of $z \geq 3.5$. The false-positive rate of the AGN class for
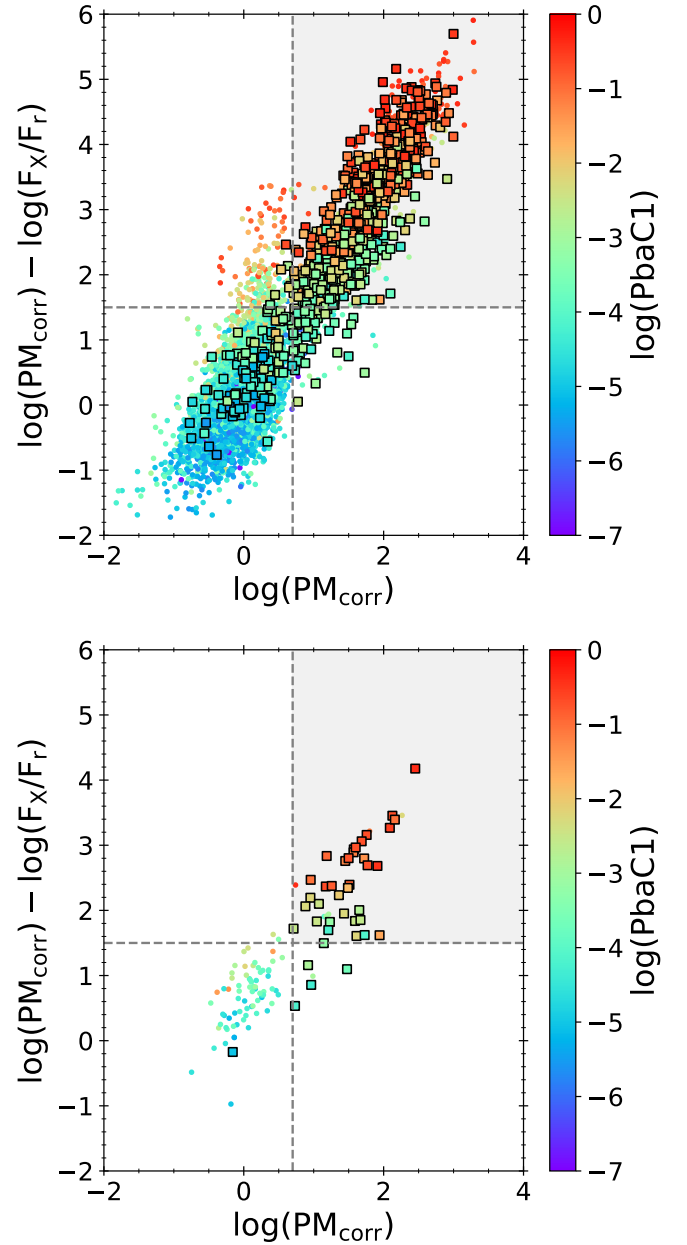

Fig. 7. $\log(PM_{corr}) - \log(F_X/F_r)$ versus the corrected *Gaia* proper motion in logarithmic scale, $\log(PM_{corr})$, where F_X/F_r is the X-ray to optical flux ratio. The dashed lines correspond to the *Gaia* criteria to select stars. The sources are colour-coded with the probability of a source being a star from the classification algorithm (Tranin et al. 2022), the PbaC1 class. The squares indicate sources with $\Delta\chi^2 \geq 5$, where $\Delta\chi^2$ is the difference in the χ^2 between the best fit of AGN/QSO versus stellar templates. The upper and lower panels correspond to the total and the high-z samples, respectively.

the whole 4XMM sample was expected to be about 5%. This value was derived from the performance of the classification algorithm on a well-characterised validation sample, as detailed in the XMM2ATHENA project[3]. However, in the high-redshift regime, this increased to 10.8%, highlighting the limitations of the algorithm in the case of rare sources such as high-z AGN.

---

[3] http://xmm-ssc.irap.omp.eu/xmm2athena/wp-content/uploads/2023/04/DeliverableD8.2.pdf

Table 1. Final sample of X-ray sources inside the 4XMM-DES-VHS footprint in different redshift bins after excluding stellar objects (Sect. 3.3) and sources with unreliable photometry (Sect. 3.4).

Redshift bins	Number with spec-z	Total number
Full sample	13 424 (20.2%)	64 529
$z \geqslant 3.5$	92 (11%)	833
$z \geqslant 4$	30 (9.3%)	323
$z \geqslant 5$	3 (5.4%)	56
$z \geqslant 6$	0	7
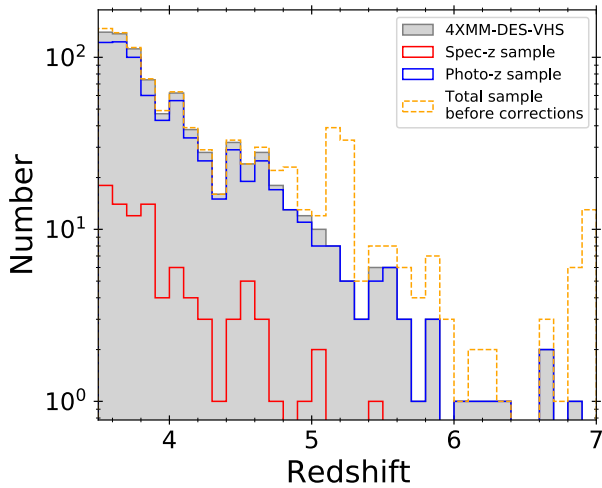
Fig. 8. Redshift distribution of the final high-z sample (gray filled). We highlight the spectroscopic and photometric redshift samples as indicated in the legend. The dotted orange histogram shows the initial selection of high-z sources before removing stellar objects and sources with unreliable photometry.

3.4. Visual inspection of the images

Section 2 describes the steps we took to obtain a clean sample by removing sources with unreliable photometry. Despite applying strict criteria, such as excluding objects near bright sources affected by diffraction spikes, some problematic photometry persisted in the high-z sample, leading to an artificially enhanced number of sources, particularly at $z > 6$. To address this, we visually inspected the images of all high-z sources across all bands (DES, VHS, WISE). This process enabled us to remove 32 sources that were strongly affected by nearby very bright stars or satellite trails, as their photometric redshift estimates were biased in one or more bands. The majority of these sources had falsely estimated photometric redshifts above $z \geqslant 6$.

4. Final sample of the high-z X-ray sources

Through our analysis, we selected in total 833 reliable X-ray sources in the early Universe ($z \geq 3.5$). Among them, there are 92 sources (~11%) with available spectroscopic information. In Table 1, we summarise the number of 4XMM-DES-VHS sources in different redshift intervals. Figure 8 shows the redshift distribution of the final high-z X-ray sample. In the same plot, we also display the samples separately, distinguishing among those with photometric and spectroscopic redshifts. For comparison, we also present the distribution of the sample presented in Sect. 3.1,

before the corrections in Sections 3.2–3.4. The complete table of high-z X-ray sources is available at the CDS.

4.1. Cumulative number counts

To verify that the number of high-z sources selected in our analysis aligns with the ones predicted by different models (e.g. Gilli et al. 2007), we calculated the cumulative number counts (logN-logS relation) and made a direct comparison with the predictions. To calculate the logN-logS, we need an estimation of the area curve of the 4XMM within the DES-VHS footprint. The XMM Survey Science Center provides EPIC sensitivity maps in the 0.2–12 keV band for each observation included in the 4XMM catalogue (see Sect. 9 of Webb et al. 2020 for details on the calculation)[4]. A sensitivity map can be converted into an area curve by calculating the cumulative number of pixels in the map with a sensitivity limit above a certain X-ray flux. Since theoretical estimates of the logN-logS and previous observational results are usually presented in the 0.5–2 keV range, we converted the sensitivity maps into the same energy band by assuming the spectral model used for the calculation of the maps: an absorbed power law with photon index of $\Gamma = 1.42$ and a column density of $N_H = 1.7 \times 10^{20}$ cm^{-2}. The choice of the Γ and N_H follows the approach used in Webb et al. (2020) for constructing EPIC sensitivity maps. This spectral model is motivated by previous studies (e.g. Lumb et al. 2002) that characterised the unresolved extragalactic X-ray background, which is primarily composed of faint AGN. The photon index of 1.42 is a standard value representing the average spectrum of these sources, making it a reasonable choice for estimating sensitivity across the survey area. The chosen N_H value corresponds to a typical low level of Galactic absorption, ensuring that the sensitivity calculations are not significantly biased by intervening gas and dust. This standard approach allows for a consistent comparison with previous observational and theoretical logN-logS distributions. The overall footprint of the 4XMM in the DES-VHS area can be quite complex, with multiple overlapping XMM-Newton observations. However, by using the 4XMM detections catalogue for our logN-logS calculation, we do not need to consider this issue. In such case, each observation can be treated as an independent survey and the total area curve can be estimated as the sum of the area curve for each observation. Our final area curve for the 4XMM-DES-VHS includes 1985 XMM-Newton observations.

As a sanity check of our method, we calculated the total logN-logS of the survey in the 0.5–2 keV band, using our estimated area curve and the 106 790 4XMM-DR11 detections included in that sky region. The 0.5–2 keV flux for each detection was calculated by adding the band 2 (0.5–1 keV) and band 3 (1–2 keV) EPIC fluxes (`EP_2_FLUX` and `EP_3_FLUX` columns from the 4XMM catalogue). Figure 9 shows the cumulative numbers derived using the described method (orange shaded region) compared to previous studies. In particular, we over-plot the logN-logS trends derived in the 2XMM (Mateos et al. 2008), the 3XMM-XXL and 4XMM-XXL surveys (Chiappetti et al. 2018; Pouliasis et al. 2022a), the 7 Ms Chandra deep field south (CDFS, Luo et al. 2017), and the Stripe 82X field (LaMassa et al. 2016). The number counts agrees very well with the aforementioned studies indicating that the area curve obtained in our analysis is accurate.

By using the logN-logS predicted by the X-ray background synthesis model of Gilli et al. (2007)[5] and our area curve for
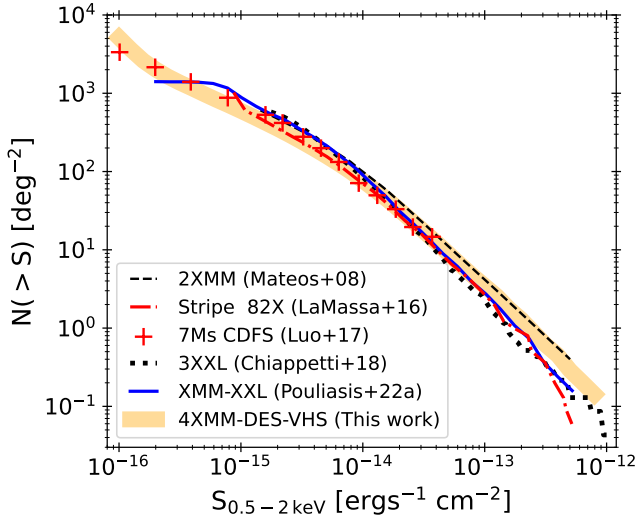
---

[4] http://xmm-catalog-dev.irap.omp.eu/~mcoriat/shared2/epfluxmap/
[5] http://www.bo.astro.it/~gilli/counts.html

Fig. 9. Cumulative number counts for the 4XMM-DES-VHS catalogue are presented with the orange shaded area highlighting the 1σ error in the soft 0.5–2 keV band. For reference, we compared with the number counts derived by Mateos et al. (2008), LaMassa et al. (2016), Luo et al. (2017), Chiappetti et al. (2018) and Pouliasis et al. (2022a).

the 4XMM-DES-VHS survey, we estimated a total of 1681 X-ray detections in the 3.5–7 redshift range for this survey. The 833 unique 4XMM sources in our high-redshift sample correspond to 1066 detections. The number of high redshift sources we identified is roughly of the same order a magnitude, but significantly lower.

Figure 10 shows the cumulative source distribution of the high-z sample. We compared our results to the predictions of the X-ray background synthesis model by Gilli et al. (2007), which is based on the optical luminosity function parametrised with a luminosity-dependent density evolution (LDDE) model and an exponential decline at high-z (solid line). We also show the number counts from the mock catalogue of X-ray selected AGN generated by Marchesi et al. (dotted line, 2020), based on the X-ray luminosity function (XLF) by Vito et al. (2014), which assumes a pure density evolution (PDE, Schmidt 1968). For reference, we include the number counts derived by Marchesi et al. (2016b) using a z ≥ 3 sample from the COSMOS Legacy survey. The lower panel of Fig. 10 shows the ratio between our data and the different studies, with the shaded blue area representing the (1σ) uncertainties in our measurements.

The number counts derived for the 4XMM-DES-VHS catalogue are systematically lower than the predictions from the LDDE model by Gilli et al. (2007) and the PDE-based mock catalogue of Marchesi et al. (2020) across the entire flux range. At fainter fluxes ($S_{0.5-2\text{keV}} < 10^{-15}$ erg s^{-1} cm^{-2}), our results show a significant deficit compared to these models by a factor of up to 2–3. At the bright end, our measurements are in relatively closer agreement but still fall slightly below both models. A similar discrepancy is observed when comparing to the results of Marchesi et al. (2016b). These differences are likely due to the incompleteness of our high-z sample at faint fluxes, as discussed in the next section.

4.2. Completeness

Figure 11 (upper panel) shows the normalised histogram of the DES i-band magnitude for the high-z X-ray AGN compared to the full DES sample inside the 4XMM-DES-VHS footprint. The
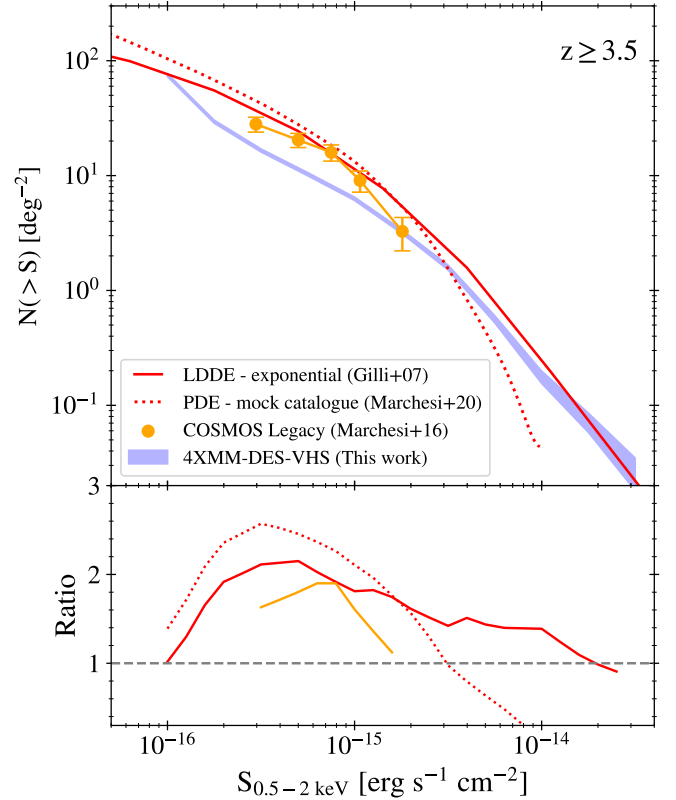

Fig. 10. Cumulative number counts for the high-z sample in the redshift bin z ⩾ 3.5 (upper panel). The shaded blue area represent the 1σ uncertainties in our measurements. The solid line indicates the LDDE model predictions with an exponential decline by Gilli et al. (2007), while the dotted line shows the mock catalogue of Marchesi et al. (2020) based on the XLF by Vito et al. (2014). For reference, we include the data points derived by Marchesi et al. (2016b). The lower panel presents the ratio between our data and the different models.

vertical line indicates the mode of the DES i-band magnitude distribution for the latter at i$_{\text{DES}}$ = 24.3 mag. Yang & Shen (2023), using the deeper (i$_{\text{HSC,limit}}$ = 26.3 mag) Hyper-Suprime Camera PDR2 deep/ultradeep data (HSC-PDR2, Aihara et al. 2019), derived the detection efficiency curves for all bands. The detection efficiency is defined as the fraction of the HSC objects that have a match in the DES DR2 catalogue (using a matching radius of 1″). In the i-band, the 95% completeness detection limit is at i$_{\text{DES}}$ = 24.0 mag, while at i$_{\text{DES}}$ = 25 mag this value drops to 40%. Since our high-z sample peaks at about i$_{\text{DES}}$ = 24.3 mag, we expect to miss sources at fainter magnitudes.

To quantify the detection efficiency of the high-z 4XMM-DES-VHS sources, we compared our sample to the high-z sample derived in the XMM-XXL-Northern field using the HSC data (Pouliasis et al. 2024). The latter study utilised the stacked XMM-Newton catalogue, which combines overlapping observations to achieve greater sensitivity and depth compared to the standard 4XMM-DR11 catalogue used in our analysis. Specifically, the stacked catalogue reaches a depth of 6×10^{-15} erg cm^{-2} s^{-1} (at 3σ) in the soft band (0.5−2keV), whereas the standard catalogue has a sensitivity limit of ∼4 × 10^{-14} erg cm^{-2} s^{-1} at the same significance level. This difference in depth allows the stacked catalogue to probe fainter X-ray sources, particularly in the high-redshift regime.

Following a similar procedure as in Yang & Shen (2023), but only considering sources with z ≥ 3.5, we matched the
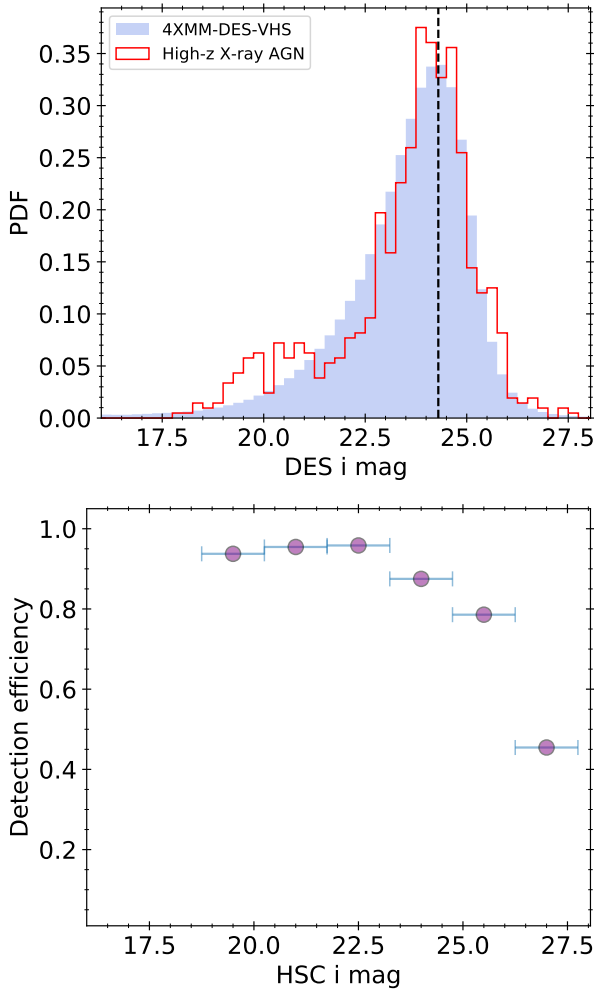
Fig. 11. Upper panel: normalised histograms of the high-z X-ray AGN selected through our analysis (red) and the DES sources inside the 4XMM-DES-VHS footprint (blue) as a function of the DES *i*-band magnitude. The vertical line indicates the mode of the DES *i*-band magnitude distribution for the latter at $i_{DES} = 24.3$ mag. Lower panel: fraction of high-z X-ray HSC sources that have a match with the DES DR2 catalogue.

DES DR2 sources to the HSC data and restricted our sample to sources with X-ray fluxes above the limit of the 4XMM-DR11 catalogue. We then calculated the detection efficiency of the high-z sources in several magnitude bins, as shown in Fig. 11 (lower panel). As expected, the completeness remains relatively constant at 90–95% up to $i_{DES} = 25$ mag and then drops sharply. The total completeness with respect to the HSC high-z sample is about 86%. Due to this incompleteness arising from the DES magnitude limit, we miss sources with faint X-ray fluxes, as seen in Fig. 2. In particular, at fluxes fainter than $f_{0.5-2\ \mathrm{keV}} < 10^{-15}$ erg s^{-1} cm^{-2}, we miss about 20–40% of the sources. This explains the underestimation of the logN-logS relation at faint fluxes (Fig. 10). The estimate of missed sources is derived from the decline in the fraction of 4XMM-DES-VHS sources with optical/near-IR counterparts (Fig. 2).

5. Comparison with the optically selected QSOs

One of the main goals of this work is to study the observational properties of the X-ray sources compared to the optically
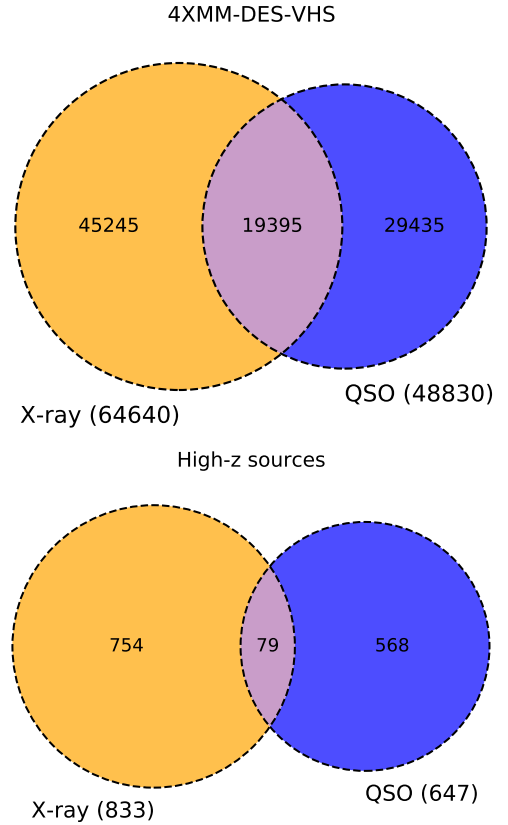

Fig. 12. Venn diagram of the X-ray sample used in our analysis and the optically selected QSO candidates for the total (upper panel) and the high-z (lower panel) samples inside the 4XMM-DES-VHS area.

selected QSOs in the early Universe. For the latter, we used a catalogue of 1.4 million photometrically-selected QSO candidates (Yang & Shen 2023) in the southern hemisphere over the 5000 deg^2 DES DR2 survey. The QSO candidates are expected to be broad-line (unobscured) AGN with a 94.7.% reliability. Inside the 4XMM-DES-VHS footprint, there are 48 830 photometrically-selected QSOs. Among them, there are 686 sources with a redshift of $z \geq 3.5$. As can be seen from the Venn diagram (Fig. 12), when comparing all the sources inside the area of our interest, about 30% of the X-ray AGN are also broad-line AGN candidates, while ~40% of the QSOs have been detected in X-rays. Concerning the high-z samples, there is only a small portion of X-ray sources that have been selected in the QSO sample and vice versa. In particular, only 9.5% of the high-z X-ray AGN are also optically selected QSOs. On the other hand, 12.5% of the QSOs are X-ray detected. In the next subsections, we compare the X-ray properties, the magnitude distributions, and the rest-frame SED shapes of both X-ray and optically selected QSOs.

5.1. X-ray properties

In this section, we present a brief view of the X-ray properties of our final high-z sample and, in particular, focusing on their X-ray absorption properties. To this end, we used the catalogues of X-ray spectral properties built by the XMM2ATHENA project for the 4XMM sources[6]. These catalogues fit a simple X-ray spectral model (an absorbed power-law) to the 4XMM detections and derived estimates and uncertainties for the line-of-sight
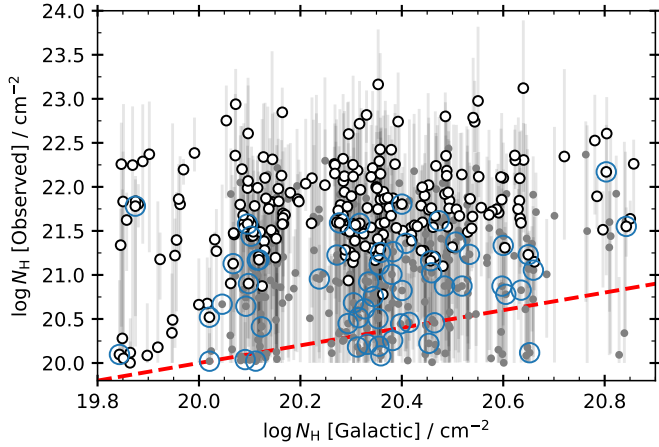
---

[6] http://xmm-ssc.irap.omp.eu/xmm2athena/catalogues/

Fig. 13. Galactic N_H versus the observed N_H for our high-z sample, estimated from the XMM2ATHENA 4XMM catalogs of X-ray spectral properties. The red, dashed line shows the one-to-one relation. Grey, solid circles are sources with an observed N_H consistent with the galactic N_H. The black, open circles show sources with N_H robustly above the Galactic value (the lower limit of the 90% confidence interval is higher than the Galactic N_H). The blue circles mark the optically selected QSO identified by Yang & Shen (2023) also included in our sample. The plot only includes sources where our estimated 90% confidence interval for the observed log N_H is narrower than 3 dex (501 out of 801 sources).

Fig. 14. Estimated distribution of the rest-frame Hydrogen column density for our high-z sample. The black, solid histogram corresponds to sources with a significant observed N_H (i.e. above the Galactic value) and corrected to rest-frame values. The hatched histogram shows our estimated number of unabsorbed sources. The blue histogram shows the corresponding distributions for the optically selected QSO by Yang & Shen (2023) included in our sample.

hydrogen column density, photon index, and 0.2–12 keV flux. The fits were done without any knowledge of redshift, so in the case of redshift dependent properties (e.g. log N_H), the quoted values in the catalogue should be considered as 'observed' quantities and should be corrected to rest-frame if possible. For a detailed description of these catalogues, we refer to Viitanen et al., (in prep.).

We first searched for counterparts of our sample in the D6.1 XMM2ATHENA catalogue, which contains X-ray spectral properties for the 319 565 detections, with at least one X-ray spectrum in the XMM-Newton archive. We found information for 161 detections, corresponding to 125 sources in our sample. After rejecting 27 detections flagged as bad fits in the catalogue (most are rejected due to bad background fits), we ended up with 104 sources (134 detections) with spectral properties from D6.1. These detections have a median net count of 90 and a signal-to-noise ratio (S/N) of approximately 9. For the remaining objects in our sample, we used the D6.3 XMM2ATHENA catalogue, which includes information for all the detections in the 4XMM. In this case, the X-ray spectral properties were derived by constructing low-resolution spectra using the 4XMM count-rates in the five standard XMM-Newton energy bands. We found 840 detections for the remaining 729 sources in our sample. We rejected 37 detections flagged either as potentially spurious or as bad fits, ending up with a total of 698 sources (805 detections) with data from D6.3. This subset has a median net count of 30 and a S/N of 6.

We calculated the average log N_H value and confidence interval for those sources with multiple detections. Finally, we rejected sources where our estimated confidence interval was larger than 3 dex, since we considered that no meaningful information about the X-ray absorption could be inferred in such cases. In all, from our sample of 833 high-redshift sources, we
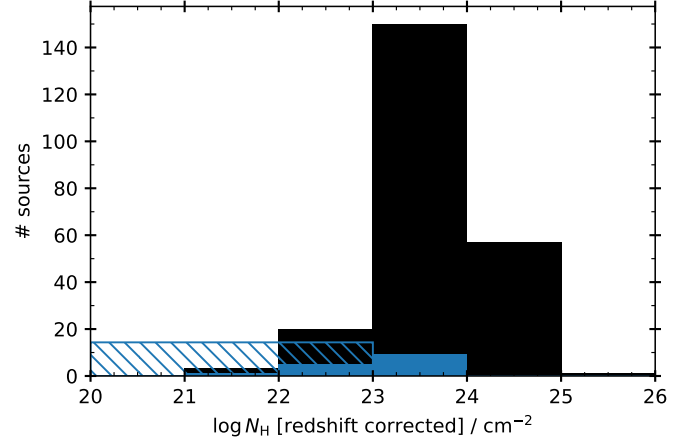
were able to obtain reasonably robust estimates of X-ray spectral parameters for 501 sources[7].

Our goal is to have a crude estimate of the fraction of sources that show signs of X-ray absorption. Since the spectral models used for the spectral fits in the XMM2ATHENA catalogues do not include the Galactic absorption, we must take this into consideration. Figure 13 shows our estimated observed log N_H average for the 501 high-z sources against the Galactic N_H towards their line of sight using the LAB Survey results (Kalberla et al. 2005). The grey circles are sources where the lower end of the log N_H confidence interval is below the Galactic absorption. We consider those sources (270 objects) as unabsorbed, since the observed N_H is consistent with the Galactic. On the other hand, the open white circles in the plot show sources with log N_H significantly above the Galactic absorption (231 objects). For these sources, we can roughly estimate the rest-frame N_H by taking into account the redshift information, using the following expression:

$$N_H(z) = N_H(z = 0) \times (1 + z)^{2.7}. \tag{3}$$

The solid black histogram in Fig. 14 shows the redshift-corrected log N_H distribution for these sources. Given the sensitivity range of XMM-Newton and the redshift of our sources, it is expected that their rest-frame N_H are above 10^{23} cm^{-2}, since only sources with absorption roughly above that level would show significant absorption in their XMM-Newton spectra. We considered sources to be as X-ray absorbed at log $N_H \geq 23$. The black hatched histogram shows the number of sources we considered as unabsorbed since they do not show an observed N_H above the Galactic level. We just show them uniformly distributed between 20 and 23.

In total, out of the 501 sources with a reasonable N_H estimate, we classified 293 of them as X-ray unabsorbed and 208 as X-ray absorbed. The observed fraction of X-ray absorbed sources

---

[7] If we apply a more conservative criterion and select sources with a confidence interval no larger than 2 dex, the size of the final sample is reduced to 354 sources. This however does not affect significantly our conclusions about the fraction of absorbed sources.
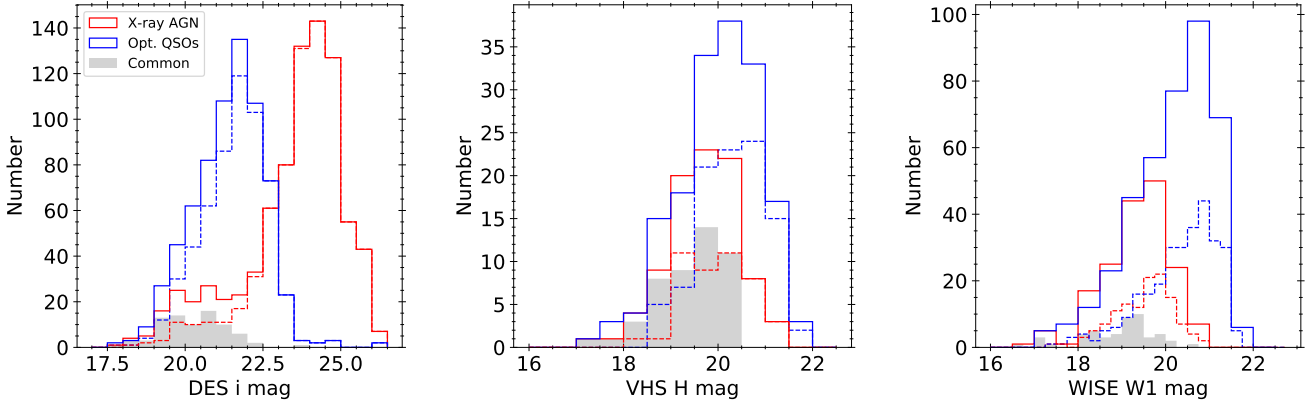
Fig. 15. Magnitude distributions of the X-ray AGN selected through our analysis (red) compared to the optically selected QSO sample (blue) for the DES optical *i*-band (left), the VHS near-IR *H* band (middle), and the WISE mid-IR *W1* band (right). The shaded histograms represent the common sources between the two samples, while the dashed lines correspond to the unique AGN in each selection method.

would be $40 \pm 2\%$. The uncertainties correspond to one-sigma confidence interval, calculated through bootstrapping. We note however that the uncertainties in the observed log N_H for a major number of our sources can be quite large (>1 dex), which makes our estimate of the fraction of absorbed sources uncertain. When we applied a more conservative criteria for the identification of absorbed sources (e.g. the lower limit of the confidence limit for the redshift corrected log N_H is above 23), we found 95 sources, corresponding to $20 \pm 2\%$ of X-ray absorbed sources. We reran our pipeline using only the sources with spec-z information and the derived fractions of obscured AGN remain the same. For a better identification of X-ray absorbed sources, a detailed spectral analysis ought to be carried out, but that is beyond the scope of this work. Nevertheless, this crude estimate is in a reasonable agreement with previous studies. For example, Pouliasis et al. (2024) found an observed fraction of X-ray absorbed sources at high-z ($3 \le z \le 6$) in the XMM-XXL-Northern field (with an average depth comparable to the 4XMM) of 28.7%.

It is also interesting to compare our results for the total sample with the optically selected QSOs that are in common with our sample. In Fig. 13 the blue circles mark 58 sources identified as QSO in the DES survey (Yang & Shen 2023). The figure already suggests that the optical QSOs have observed log N_H towards the lower end of the distribution. By using the identification criteria for X-ray absorbed sources we presented above, we found nine absorbed sources (~15%) out of the 58. When applying the more conservative criterion, only one QSO is classified as absorbed. As expected, the fraction of X-ray absorbed sources within optically selected QSOs seems to be significantly lower than in our total sample.

5.2. Rest-frame SEDs

Figure 15 presents the X-ray AGN (red) and optically selected QSO (blue) distributions of the DES *i*-band (left), VHS *H*-band (middle), and WISE *W1*-band (right) magnitudes. The common sources between the two samples are shown also with the gray shaded histogram, while we overplot the X-ray AGN that are not optically selected QSOs and vice-versa (dashed lines). The X-ray population covers the faint end of the DES *i*-band distribution, while the QSOs populate the bright one. One should take into account in this comparison that the selection function of the QSOs is biased towards the bright end in the optical. From the optical to mid-IR wavelengths, this trend is reversed;
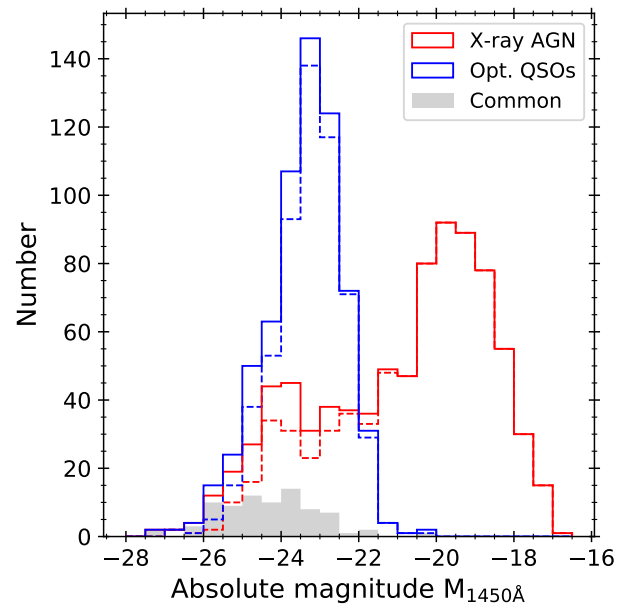

Fig. 16. Absolute magnitude distribution at 1450 Å of the X-ray AGN selected through our analysis (red) compared to the optically selected QSO sample (blue). The shaded histograms represent the common sources between the two samples, while the dashed lines correspond to the unique AGN in each selection method.

the QSOs have fainter *W1* bands compared to the X-ray sample. We also examined the absolute magnitude distribution at 1450 Å for the same samples (Fig. 16). The absolute magnitudes were calculated after rerunning the LePhare algorithm using the final redshifts of the high-z sources. The QSO sample spans a range of approximately -26 to -22, with a median value of $M_{1450Å} = -23.3$, while the X-ray AGN cover a much wider magnitude range, namely, from -26 to -18. However, the majority of the X-ray sources are at lower luminosities, with a median absolute magnitude of $M_{1450Å} = -20.3$. This confirms the observed trend that X-ray AGN tend to be fainter in the optical compared to optically selected QSOs, reflecting the bias of optical selection towards the brightest sources.

To better understand the above patterns, we aim to compare the rest-frame averaged SEDs of both the QSO and the X-ray AGN samples. We ran the CIGALE code (Yang et al. 2020, 2022)

Table 2. Models and their parameter space used by CIGALE for the SED fitting of the high-z sources.

Parameter	Value
Star formation history: delayed SFH with optional exponential burst)	
Age of the main stellar population in Myr	50,100,200,500,1000,1500, 2000, 3000, 4000, 5000
e-folding time of the main stellar population model in Myr, τ_{main}	50,100,200, 500, 700, 1000, 2000, 3000, 4000, 5000
Age of the late burst in Myr, age_{burst}	20
Mass fraction of the late burst population, f_{burst}	0.0, 0.005, 0.01, 0.015, 0.02, 0.05, 0.10, 0.15, 0.18, 0.20
e-folding time of the late starburst population model in Myr, τ_{burst}	50.0
Stellar population synthesis model	
Single Stellar Population Library	Bruzual & Charlot (2003)
Initial Mass Function	Salpeter (1955)
Metallicity	0.02 (Solar)
Nebular emission	
Ionisation parameter ($\log U$)	−2.0
Fraction of Lyman continuum escaping the galaxy (f_{esc})	0.0
Fraction of Lyman continuum absorbed by dust (f_{dust})	0.0
Line width (FWHM) in km/s	300.0
Dust attenuation: modified attenuation law Charlot & Fall (2000)	
V-band attenuation in the interstellar medium in mag, Av_{ISM}	0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4
Dust template: Dale et al. (2014)	
AGN fraction	0.0
Alpha slope, α	2.0
AGN models from Stalevski et al. (2016) (SKIRTOR)	
Average edge-on optical depth at 9.7 micron (t)	3.0, 7.0
Power-law exponent that sets radial gradient of dust density (pl)	1.0
Index that sets dust density gradient with polar angle (q)	1.0
Angle measured between the equatorial plane and edge of the torus in degrees (oa)	40
Ratio of outer to inner radius, R_{out}/R_{in}	20
Fraction of total dust mass inside clumps (M_{cl})	97%
Inclination angle in degrees (i)	30, 70
AGN fraction	0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.99
Extinction in polar direction in mag, E(B-V)	0.0, 0.2, 0.4
Emissivity of the polar dust	1.6
Temperature of the polar dust (K)	100.0
The extinction law of polar dust	SMC

Notes. Type-2 AGN have an inclination $i = 70°$, while type-1 AGN have $i = 30°$. The extinction in polar direction, E(B-V), included in the AGN module, accounts for the possible extinction in type-1 AGN, due to polar dust. The AGN fraction is measured as the AGN emission relative to IR luminosity (1–1000 μm).

to model and fit the SEDs of both the X-ray sample and the optically selected QSOs. CIGALE is a multi-component SED fitting algorithm that fits the observational data of the sources to the theoretical models, and has been widely used in the literature (e.g. Pouliasis et al. 2020). We used all the available photometry to construct the SEDs, from optical to mid-IR, and we created a grid that models both the galaxy and the AGN emission.

In brief, we used the stellar population synthesis model of Bruzual & Charlot (2003), assuming the initial mass function (IMF) by Salpeter (1955) and a constant solar metallicity (Z = 0.02). We used the dust emission templates by Dale et al. (2014) without AGN emission and the dust extinction law by Charlot & Fall (2000). For the AGN component, we adopted the AGN templates presented in Stalevski et al. (2012, 2016, SKIRTOR). In addition, for the star formation history (SFH), we used a delayed SFH with the functional form SFR $\propto$ t $\times$ exp($-t/\tau$)

that includes a star formation burst no longer than $\tau = 20$ Myr. We did not include the X-ray module in the SED-fitting process, as it requires intrinsic absorption-corrected X-ray fluxes, which cannot be reliably determined for our sample. Furthermore, since we are comparing the averaged SEDs of X-ray-detected AGN with optically selected QSOs (most of which lack X-ray detections), excluding X-ray fluxes ensures a consistent and unbiased comparison between the two populations. The list of the models and their parameters used in our analysis are given in Table 2. With this configuration, we were able to fit the observational data with more than 650 million models. In Fig. 17, we show three SED examples representative of the objects of our samples used in this work (e.g. X-ray obscured and unobscured AGN and an optically selected QSO). In order to have results as reliable as possible, we kept all the sources that have low reduced χ^2 (χ_r^2) that is indicative of the goodness of the SED fitting
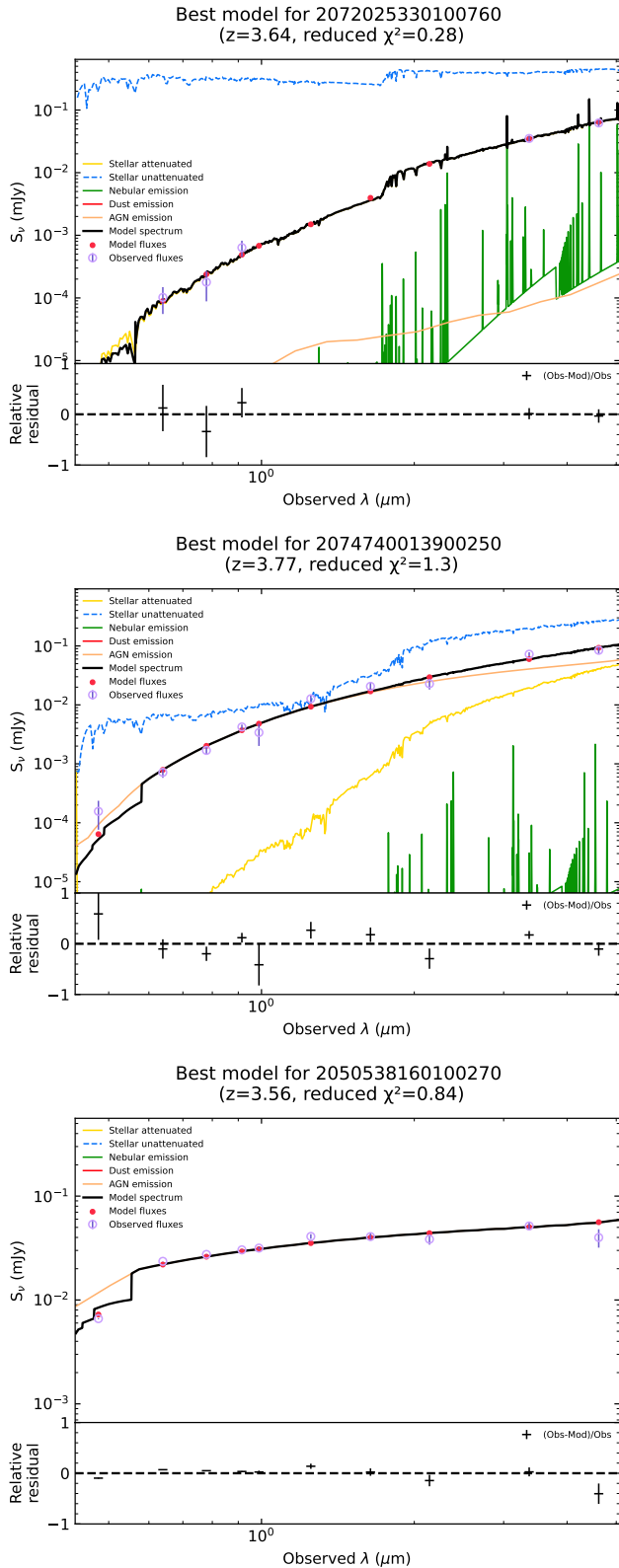
Fig. 17. Example SED fits of an X-ray absorbed AGN (top), an X-ray unabsorbed AGN (middle), and an optically selected QSO (bottom). The dust emission is plotted in red, the AGN component in orange, the attenuated (unattenuated) stellar component is shown with the yellow (blue) solid (dashed) line, while the green lines shows the nebular emission. The total flux is represented with black colour. Below each SED, we plot the relative residual fluxes versus the wavelength.

process. Therefore, we excluded sources that have $\chi_r^2 > 5$. This value has been used in previous studies (e.g. Mountrichas et al. 2019; Pouliasis et al. 2022b; Georgantopoulos et al. 2023) and is based on visual inspection of the SEDs. Following this criterion, we excluded 36 X-ray AGN and 137 optically selected QSO candidates from further analysis. We examined whether the poorly fitted SEDs correlate with the number of available photometric points, magnitudes, or redshift, but we found no clear dependence.

In CIGALE, the obscuration is primarily constrained by the inclination angle (i), with typical values of 30 degrees for type 1 AGN and 70 degrees for type 2 AGN (e.g., Mountrichas et al. 2022; Koutoulidis et al. 2022). Additionally, for type-1 AGN, it is possible to also account for dust extinction in the polar direction. The optical depth at 9.7 microns $\tau_{9.7}$ is not strongly constrained by the SED, as indicated by the flat relationship between the true and estimated values (Yang et al. 2020, 2022). Therefore, the classification into type 1 and type 2 AGN is based on the inclination angle, rather than $\tau_{9.7}$, with the algorithm showing a robustness of 70–85%, compared to spectroscopic samples in the literature (e.g. Mountrichas et al. 2021). According to this criterion, 90% of the QSO sample are classified as type 1, according to the SED fitting verifying the QSO selection criteria of Yang & Shen (2023).

The two plots in Fig. 18 provide a comparative analysis of the SEDs for different AGN populations, specifically optically selected QSOs and X-ray-selected sources. These plots illustrate the efficacy of X-ray-selected samples in identifying AGN with varying levels of obscuration and showcase the distinct SED shapes of each population across the electromagnetic spectrum.

The upper left panel compares the median SEDs of QSOs (not X-ray detected, blue) and X-ray only-selected sources (red) across the rest-frame wavelength spectrum. The common sources in the two samples are presented in yellow. These SEDs are constructed by median-combining the individual SEDs of the AGN in each sample, with the shaded regions marking the 25th to 75th percentile range to represent dispersion at each wavelength. For the QSO population, the SED in the UV-optical region is dominated by emission from the accretion disk, which is characteristic of an unobscured AGN. This strong UV-optical emission, observed as a peak in the QSO SED, reflects minimal dust absorption and is a key indicator of the active central engine. In contrast, the X-ray-selected sources show a marked difference in SED shape. The UV-optical part of their spectrum appears significantly extincted, due to the presence of dust and gas obscuring the central engine. X-ray observations are particularly effective in detecting AGN across a broad range of obscuration levels, allowing for the identification of both heavily and lightly obscured sources, as evidenced by the absorption in this wavelength region for the X-ray population. In the IR part of the spectrum, above 1 μm, the SED of the X-ray-selected AGN shows a noticeably higher level of emission compared to QSOs. This suggests that X-ray-selected AGN contain a larger amount of dust. The median SED for common sources sits between the QSO and X-ray AGN SEDs across much of the wavelength range. In the UV-optical part, common sources show evidence of accretion disk emission, similar to QSOs, indicating that they include unobscured AGN. In the IR range, they show higher dust emission than QSOs, though still less than the purely X-ray-selected sources.

The lower left panel in Fig. 18 illustrates the ratio of host-galaxy to total emission across wavelengths, providing insights
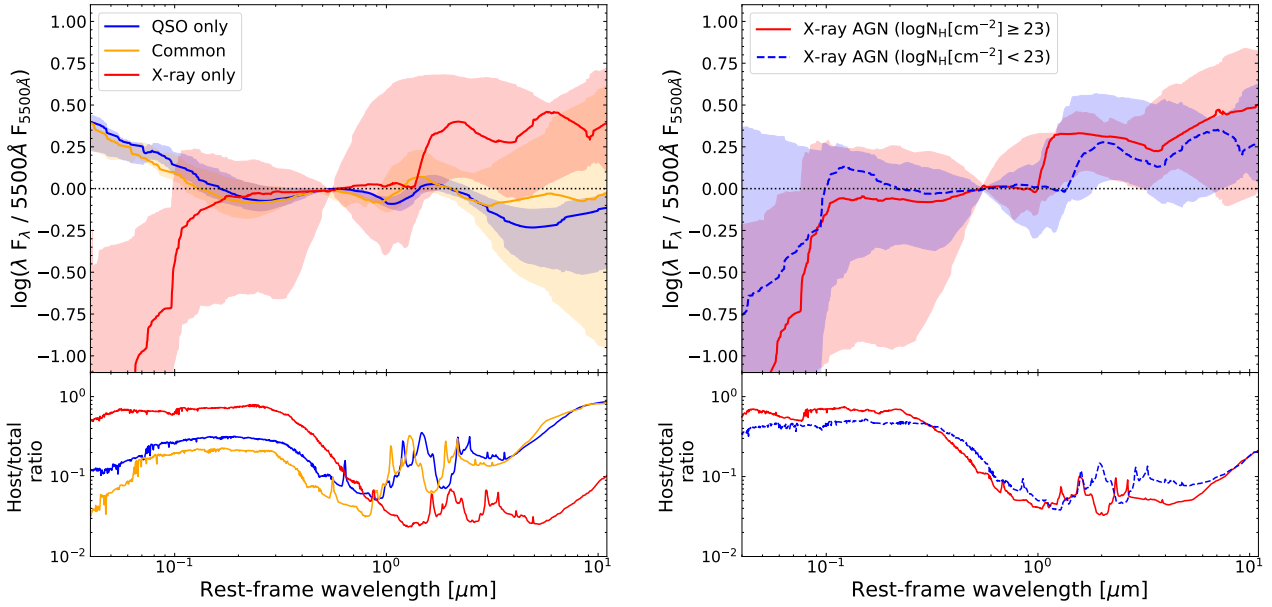
Fig. 18. Left: median SEDs (solid lines) of the X-ray (red) and optically (blue) selected AGN, including the common sources (yellow). Right: median SEDs (solid lines) of the X-ray selected AGN with $\log N_H$ (cm^{-2}) < 23 (blue) and $\log N_H$ (cm^{-2}) $\geq$ 23 (red). The shaded regions correspond to the 25th and 75th percentiles of the dispersion of the templates at each wavelength. The bottom panel shows the ratio of the host-galaxy to the total emission. All individual SEDs are normalised at 5500 Å.

into the contributions from the AGN and host galaxy for each population. For the QSO-only sample, this ratio reveals as expected that the central engine overwhelmingly dominates the total emission, with minimal influence from the host galaxy. This aligns with the strong accretion disk emission observed in the UV-optical range, where the AGN outshines any host-galaxy contributions, resulting in a low host-galaxy to total emission ratio. In contrast, for X-ray only-selected sources, the emission is more balanced, with substantial contributions from both the AGN and the host galaxy. This balance reflects the impact of obscuration, which dims the central AGN and allows the host galaxy to be more visible. Consequently, the host-galaxy to total emission ratio is higher, yielding a flatter profile across wavelengths than seen in QSOs.

The right panel of Fig. 18 further explores these trends by comparing the median SEDs of high-z X-ray sources, specifically contrasting absorbed ($\log N_H \geq 23$) and unabsorbed ($\log N_H < 23$) AGN. In line with expectations, the median UV-optical emission of absorbed AGN is heavily obscured, whereas the unabsorbed AGN show less absorption and consequently slightly stronger emission in this range. The difference in X-ray obscured and unobscured AGN in the IR part of the spectrum does not seem so strong taking the median values. Also, when considering the 25–75 percentiles of the distributions (represented by the shaded regions in the plot), there is substantial overlap between absorbed and unabsorbed AGN. This overlap suggests that despite the general trends, individual AGN can exhibit a wide range of SED shapes within each category, highlighting diversity within each population.

Hence, these plots demonstrate that X-ray-selected samples are highly effective in identifying AGN with varying levels of obscuration, capturing differences in SED shapes across both unobscured and obscured populations. The clear accretion disk emission in QSOs and common sources contrasts with the extinction observed in X-ray-selected sources in the UV-optical region, while the higher IR emission of X-ray AGN points to more substantial dust content.

6. Summary and conclusions

In this work, we made use of the 4XMM-DR11 catalogue (1240 deg^2) that contains all the serendipitous X-ray sources from the *XMM-Newton* space telescope, in combination with the deep optical second release of the Dark Energy Survey (DES-DR2, 5000 deg^2) to select the largest sample so far of high-redshift ($z \geq 3.5$) X-ray selected AGN. Our main goal in this work is to study the properties of the high-z X-ray AGN and compare them to the high-z optically selected QSOs. Our main results can be summarised as follows:

– In the overlapping area of about 185 deg^2 between the 4XMM-DR11, DES-DR2, and VHSD-DR5 surveys, there are 64 529 X-ray sources with a reliable optical or IR counterpart. For sources without spectroscopic information (about 80%), we used the photometric redshifts derived using both SED fitting and machine-learning algorithms. After removing spurious sources, low-redshift galaxy interlopers and stellar objects contaminating our sample, we ended up with 833 high-z X-ray AGN with high reliability. In particular, the percentage of outliers between the spectroscopic and photometric redshifts is $\eta \leq 12\%$ for $z \geq 3.5$. Among the high-z X-ray AGN, there are 92 sources (11%) with available spectroscopic information.

– We were able to derive the logN-logS relation in the redshift bin $z \geq 3.5$ at fluxes between $f_{0.5-2\,\text{keV}} = 10^{-16} - 10^{-14}$ erg s^{-1} cm^{-2}. Our number counts are systematically lower than predictions from LDDE and PDE models, with a discrepancy of up to a factor of 2–3. These discrepancies are caused mainly by the incompleteness of our sample due to the sensitivity limits of the optical and near-IR bands.

– We compared the X-ray selected sample with the optically selected broad-line QSO candidates in the same area. Only a small fraction ($\sim 10\%$) of the high-z X-ray selected AGN are also optical QSOs and vice versa, which is in contrast with the 35% of the full sample. This result agrees with previous studies at high redshifts.

- By examining the X-ray properties of the sources, we found that among the X-ray sources the observed absorbed fraction ($\log N_H$ [cm^{-2}] ≥ 23) is ~20–40%. This fraction is significantly lower (2–15%) for the optical QSOs that are detected in X-rays (12.5%).
- In comparing the magnitude distributions and rest-frame SEDs of QSOs and X-ray AGN, the X-ray AGN are observed to have fainter optical and brighter mid-IR magnitudes than optically selected QSOs. Their rest-frame SED shapes also differ: optical QSOs are dominated by AGN emission in the UV and optical range with little host-galaxy influence, while X-ray AGN show significant absorption in these wavelengths for both absorbed and unabsorbed sources. Common sources serve as an intermediate group, showing moderate obscuration, with AGN emission more pronounced, as compared to X-ray-selected sources.

We conclude that the differences among the AGN selection methods (optical, X-rays) are not due to selection or survey biases, but on the differences of the intrinsically AGN or host-galaxy properties. Hence, it is crucial to use wide X-ray surveys (such as 4XMM) to uncover both the obscured (type 2) AGN population and the type 1 that are complementary to the optically selected QSOs. Future, deeper and wider surveys, such as *Euclid*, will increase the optical/IR identification of the X-ray sources and the completeness of the high-z AGN sources.

Data availability

The table with the final high-z X-ray-selected AGN is available at the CDS via anonymous ftp to `cdsarc.cds.unistra.fr` (`130.79.128.5`) or via `https://cdsarc.cds.unistra.fr/viz-bin/cat/J/A+A/697/A123`

References

Abbott, T. M. C., Adamów, M., Aguena, M., et al. 2021, ApJS, 255, 20
Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4
Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 71, 114
Aird, J., Coil, A. L., Georgakakis, A., Nandra, & Pérez-González, P. G. 2015, MNRAS, 451, 1892
Akiyama, M., He, W., Ikeda, H., et al. 2018, PASJ, 70, S34
Antonucci, R. 1993, ARA&A, 31, 473
Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540
Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, AJ, 156, 123
Bañados, E., Venemans, B. P., Mazzucchelli, et al. 2018, Nature, 553, 473
Barchiesi, L., Pozzi, F., Vignali, C., et al. 2021, PASA, 38, e033
Barlow-Hall, C. L., Delaney, J., Aird, J., Evans, & Watson, M. G. 2023, MNRAS, 519, 6055
Barth, A. J., Ho, L. C., Rutledge, R. E., & Sargent, W. L. W. 2004, ApJ, 607, 90
Blecha, L., Snyder, G. F., Satyapal, S., & Ellison, S. L. 2018, MNRAS, 478, 3056

Bonaventura, N., Lyu, J., Rieke, G. H., et al. 2024, ArXiv e-prints [arXiv:2401.07863]
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Carnerero, M. I., Raiteri, C. M., Rimoldini, L., et al. 2023, A&A, 674, A24
Carrasco Kind, M., & Brunner, R. J. 2013, MNRAS, 432, 1483
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints [arXiv:1612.05560]
Charlot, S., & Fall, S. M. 2000, ApJ, 539, 718
Chiappetti, L., Fotopoulou, S., Lidman, C., et al. 2018, A&A, 620, A12
Ciotti, L., & Ostriker, J. P. 1997, ApJ, 487, L105
Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products by R. M. Cutri et al.
Dale, D. A., Helou, G., Magdis, G. E., et al. 2014, ApJ, 784, 83
De Rosa, G., Venemans, B. P., Decarli, et al. 2014, ApJ, 790, 145
De Rosa, A., Vignali, C., Severgnini, P., et al. 2023, MNRAS, 519, 5149
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604
Dong, X., Wang, T., Yuan, W., et al. 2007, ApJ, 657, 700
Ellison, S. L., Viswanathan, A., Patton, D. R., et al. 2019, MNRAS, 487, 2491
Fan, X., Bañados, E., & Simcoe, R. A. 2023, ARA&A, 61, 373
Ferrarese, L., & Merritt, D. 2000, ApJ, 539, L9
Filippenko, A. V., & Ho, L. C. 2003, ApJ, 588, L13
Flesch, E. W. 2016, PASA, 33, e052
Flesch, E. W. 2024, Open J. Astrophys., 7, 6
Fotopoulou, S., Pacaud, F., Paltani, S., et al. 2016, A&A, 592, A5
Gaia Collaboration (Prusti, T., et al.) 2016, A&A, 595, A1
Gaia Collaboration (Vallenari, A., et al.) 2023, A&A, 674, A1
Gebhardt, K., Bender, R., Bower, G., et al. 2000, ApJ, 539, L13
Georgakakis, A., Nandra, K., Yan, R., et al. 2008, MNRAS, 385, 2049
Georgakakis, A., Aird, J., Buchner, J., et al. 2015, MNRAS, 453, 1946
Georgantopoulos, I., Pouliasis, E., Mountrichas, G., et al. 2023, A&A, 673, A67
Gilli, R., Comastri, A., & Hasinger, G. 2007, A&A, 463, 79
Gilli, R., Norman, C., Calura, F., et al. 2022, A&A, 666, A17
Granato, G. L., De Zotti, G., Silva, L., Bressan, A., & Danese, L. 2004, ApJ, 600, 580
Greene, J. E., & Ho, L. C. 2004, ApJ, 610, 722
Greene, J. E., & Ho, L. C. 2007, ApJ, 670, 92
Greene, J. E., Ho, L. C., & Barth, A. J. 2008, ApJ, 688, 159
Habouzit, M., Li, Y., Somerville, R. S., et al. 2021, MNRAS, 503, 1940
Habouzit, M., Somerville, R. S., Li, Y., et al. 2022, MNRAS, 509, 3015
Harikane, Y., Ono, Y., Ouchi, M., et al. 2022, ApJS, 259, 20
Hopkins, P. F., & Hernquist, L. 2006, ApJS, 166, 1
Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, ApJS, 163, 1
Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, ApJ, 654, 731
Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008, ApJS, 175, 356
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90
Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841
Jiang, L., McGreer, I. D., Fan, X., et al. 2016, ApJ, 833, 222
Kalberla, P. M. W., Burton, W. B., Hartmann, Bajaja, E., & Morras. 2005, A&A, 440, 775
Khorunzhev, G. A., Burenin, R. A., Sazonov, S. Y., et al. 2019, Astron. Lett., 45, 411
Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330
Kormendy, J., & Kennicutt, Jr., R. C. 2004, ARA&A, 42, 603
Koutoulidis, L., Mountrichas, G., Georgantopoulos, I., Pouliasis, E., & Plionis, M. 2022, A&A, 658, A35
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
LaMassa, S. M., Urry, C. M., Cappelluti, N., et al. 2016, ApJ, 817, 172
Lanzuisi, G., Ranalli, P., Georgantopoulos, I., et al. 2015, A&A, 573, A137
Lapi, A., Raimundo, S., Aversa, R., et al. 2014, ApJ, 782, 69
Lapi, A., Aversa, R., & Danese, L. 2016, in The Interplay between Local and Global Processes in Galaxies, eds. S. F. Sanchez, C. Morisset, & G. Delgado-Inglada, 7
Lumb, D. H., Warwick, R. S., Page, M., & De Luca, A. 2002, A&A, 389, 93
Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2017, ApJS, 228, 2
Lusso, E., Valiante, R., & Vito, F. 2023, The Dawn of Black Holes (Singapore: Springer Nature), 1
Lyu, J., Alberts, S., Rieke, G. H., et al. 2024, ApJ, 966, 229
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, AJ, 115, 2285
Marchesi, S., Civano, F., Salvato, M., et al. 2016a, ApJ, 827, 150
Marchesi, S., Lanzuisi, G., Civano, F., et al. 2016b, ApJ, 830, 100
Marchesi, S., Gilli, R., Lanzuisi, G., et al. 2020, A&A, 642, A184
Mateos, S., Warwick, R. S., Carrera, F. J., et al. 2008, A&A, 492, 51
Matsuoka, Y., Iwasawa, K., Onoue, M., et al. 2022, ApJS, 259, 18
McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2013, The Messenger, 154, 35
Medvedev, P., Sazonov, S., Gilfanov, M., et al. 2020, MNRAS, 497, 1842
Menci, N., Fiore, F., Puccetti, S., & Cavaliere, A. 2008, ApJ, 686, 219

Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2017, PASJ, 70, S1
Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, Nature, 474, 616
Motch, C., Carrera, F., Genova, F., et al. 2017, ASP Conf. Ser., 512, 165
Mountrichas, G., Georgakakis, A., & Georgantopoulos, I. 2019, MNRAS, 483, 1374
Mountrichas, G., Buat, V., Georgantopoulos, I., et al. 2021, A&A, 653, A70
Mountrichas, G., Buat, V., Yang, G., et al. 2022, A&A, 667, A145
Netzer, H. 2015, ARA&A, 53, 365
Ni, Y., Di Matteo, T., Gilli, R., et al. 2020, MNRAS, 495, 2135
Ni, Y., Di Matteo, T., Bird, S., et al. 2022, MNRAS, 513, 670
Niida, M., Nagao, T., Ikeda, H., et al. 2020, ApJ, 904, 89
Onken, C. A., Wolf, C., Bessell, M. S., et al. 2019, PASA, 36, e033
Peca, A., Cappelluti, N., Urry, C. M., et al. 2023, ApJ, 943, 162
Pineau, F. X., Derriere, S., Motch, C., et al. 2017, A&A, 597, A89
Pouliasis, E., Mountrichas, G., Georgantopoulos, I., et al. 2020, MNRAS, 495, 1853
Pouliasis, E., Georgantopoulos, I., Ruiz, A., et al. 2022a, A&A, 658, A175
Pouliasis, E., Mountrichas, G., Georgantopoulos, I., et al. 2022b, A&A, 667, A56
Pouliasis, E., Ruiz, A., Georgantopoulos, I., et al. 2024, A&A, 685, A97
Predehl, P., Andritschke, R., Arefiev, V., et al. 2021, A&A, 647, A1
Ramos Almeida, C., & Ricci, C. 2017, Nat. Astron., 1, 679
Ruiz, A., Corral, A., Mountrichas, G., & Georgantopoulos, I. 2018, A&A, 618, A52
Salpeter, E. E. 1955, ApJ, 121, 161
Satyapal, S., Secrest, N. J., McAlpine, W., et al. 2014, ApJ, 784, 113
Saxena, A., Salvato, M., Roster, W., et al. 2024, A&A, 690, A365
Schmidt, M. 1968, ApJ, 151, 393
Secrest, N. J., Ellison, S. L., Satyapal, S., & Blecha, L. 2020, MNRAS, 499, 2380

Signorini, M., Marchesi, S., Gilli, R., et al. 2023, A&A, 676, A49
Silk, J., & Rees, M. J. 1998, A&A, 331, L1
Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, MNRAS, 391, 481
Stalevski, M., Fritz, J., Baes, M., Nakos, T., & Popović, L. Č. 2012, MNRAS, 420, 2756
Stalevski, M., Ricci, C., Ueda, Y., et al. 2016, MNRAS, 458, 2288
Strüder, L., Briel, U., Dennerl, K., et al. 2001, A&A, 365, L18
Taylor, M. B. 2005, ASP Conf. Ser., 347, 29
Tranin, H., Godet, O., Webb, N., & Primorac, D. 2022, A&A, 657, A138
Turner, M. J. L., Abbey, A., Arnaud, M., et al. 2001, A&A, 365, L27
Vito, F., Gilli, R., Vignali, C., et al. 2014, MNRAS, 445, 3557
Vito, F., Brandt, W. N., Yang, G., et al. 2018, MNRAS, 473, 2378
Vito, F., Brandt, W. N., Bauer, F. E., et al. 2019, A&A, 630, A118
Volonteri, M., Dubois, Y., Pichon, C., & Devriendt, J. 2016, MNRAS, 460, 2979
Webb, N. A., Coriat, M., Traulsen, I., et al. 2020, A&A, 641, A136
Webb, N. A., Carrera, F. J., Schwope, A., et al. 2023, Astron. Nachr., 344, e20220102
Willott, C. J., Albert, L., Arzoumanian, D., et al. 2010, AJ, 140, 546
Wolf, C., Onken, C. A., Luvaul, L. C., et al. 2018, PASA, 35, e010
Wolf, J., Nandra, K., Salvato, M., et al. 2021, A&A, 647, A5
Wolf, J., Nandra, K., Salvato, M., et al. 2023, A&A, 669, A127
Wolf, J., Salvato, M., Belladitta, S., et al. 2024, A&A, 691, A30
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Yang, Q., & Shen, Y. 2023, ApJS, 264, 9
Yang, G., Boquien, M., Buat, V., et al. 2020, MNRAS, 491, 740
Yang, G., Boquien, M., Brandt, W. N., et al. 2022, ApJ, 927, 192
Yang, G., Caputi, K. I., Papovich, C., et al. 2023, ApJ, 950, L5