A&A, 691, A136 (2024)
https://doi.org/10.1051/0004-6361/202451509
© The Authors 2024

Astronomy & Astrophysics

MAMBO: An empirical galaxy and AGN mock catalogue for the exploitation of future surveys

X. López-López[1,2,⋆], M. Bolzonella[1], L. Pozzetti[1], M. Salvato[3,4], L. Bisigello[5,6], A. Feltre[7], I. E. López[1], A. Viitanen[8,9], V. Allevato[10], A. Bongiorno[8], G. Girelli[11], J. Buchner[3], S. Charlot[12], F. Ricci[13,8], C. Schreiber[14], and G. Zamorani[1]

[1] INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
[2] Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy
[3] Max-Planck-Institut für extraterrestrische Physik, Giessenbachstr. 1, 85748 Garching, Germany
[4] Exzellenzcluster ORIGINS, Boltzmannstr. 2, 85748 Garching, Germany
[5] INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, 40129 Bologna, Italy
[6] Dipartimento di Fisica e Astronomia "G. Galilei", Università di Padova, Via Marzolo 8, 35131 Padova, Italy
[7] INAF – Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125, Firenze, Italy
[8] INAF – Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy
[9] Department of Physics and Helsinki Institute of Physics, Gustaf Hällströmin katu 2, 00014 University of Helsinki, Finland
[10] INAF – Osservartorio Astronomico di Capodimonte, Via Moiariello 16, 80131 Napoli, Italy
[11] Independent researcher, Bologna, Italy
[12] Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France
[13] Dipartimento di Matematica e Fisica, Università degli Studi Roma Tre, via della Vasca Navale 84, 00146 Roma, Italy
[14] IBEX Innovations, Sedgefield, Stockton-on-Tees TS21 3FF, UK

Received 15 July 2024 / Accepted 6 September 2024

ABSTRACT

Context. Current and future large surveys will produce unprecedented amounts of data. Realistic simulations have become essential for the design and development of these surveys, as well as for the interpretation of the results.
Aims. We present MAMBO, a flexible and efficient workflow to build empirical galaxy and active galactic nucleus (AGN) mock catalogues that reproduce the physical and observational properties of these sources.
Methods. We started with simulated dark matter (DM) haloes, to preserve the link with the cosmic web, and we populated them with galaxies and AGN using abundance matching techniques. We followed an empirical methodology, using stellar mass functions, host galaxy AGN mass functions, and AGN accretion rate distribution functions studied at different redshifts to assign, among other properties, stellar masses, the fraction of quenched galaxies, or the AGN activity (demography, obscuration, multiwavelength emission, etc.).
Results. As a proof test, we applied the method to a Millennium DM lightcone of 3.14 deg^2 up to a redshift of $z = 10$ and down to stellar masses of $\mathcal{M} \gtrsim 10^{7.5}\,M_\odot$. We show that the AGN population from the mock lightcone presented here reproduces with good accuracy various observables, such as state-of-the-art luminosity functions in the X-ray up to z∼7 and in the ultraviolet up to z∼5, optical/near-infrared colour–colour diagrams, and narrow emission line diagnostic diagrams. Finally, we demonstrate how this catalogue can be used to make useful predictions for large surveys. Using *Euclid* as a case example, we compute, among other forecasts, the expected surface densities of galaxies and AGN detectable in the *Euclid* H_E band. We find that *Euclid* might observe (on H_E only) about 10^7 and 8×10^7 type 1 and 2 AGN, respectively, and 2×10^9 galaxies at the end of its 14 679 deg^2 Wide survey, in good agreement with other published forecasts.

Key words. methods: numerical – methods: statistical – catalogs – surveys – galaxies: general – quasars: general

1. Introduction

In the upcoming years, 'full-sky' space and ground-based surveys, such as *Euclid* (Euclid Collaboration 2024g), MOONS (Cabral et al. 2020), *Rubin*/LSST (Ivezić et al. 2019), or eROSITA (Merloni et al. 2024), among others, will survey unprecedentedly large areas of the sky, gathering photometric and spectroscopic data for billions of galaxies and (at least) millions of active galactic nuclei (AGN). Synthetic data reproducing observed properties of astrophysical sources have become essential to enhance the scientific return from this new data (see e.g. Georgakakis et al. 2019, 2020; Comparat et al. 2020; Allevato et al. 2021; Bisigello et al. 2021; Euclid Collaboration 2024h). They are needed before the start of observations; for example, to define the selection bias, test data analysis pipelines, and develop and calibrate models and parameters for future analyses. Synthetic data can also help in understanding the observations; in particular, it can be used to estimate incompleteness and biases and to refine and validate hypotheses on the basis of the simulations.

Different methods can be used to generate mock catalogues (see Wechsler & Tinker 2018, for a review), ranging from physics-oriented models (hydrodynamic simulations or semi-analytic models) to data-oriented simulations (empirical models adopting observed scaling relations). Many of these

⋆ Corresponding author; xavier.lopezlopez2@unibo.it

Open Access article, published by EDP Sciences, under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
This article is published in open access under the Subscribe to Open model. Subscribe to A&A to support open access publication.

simulations are primarily focussed on cosmological purposes, such as performing idealised tests on key cosmological probes like clustering and weak lensing (e.g. Korytov et al. 2019; Ishiyama et al. 2021; Hadzhiyska et al. 2023). However, to generate reliable forecasts from these measurements, it is crucial to incorporate realistic galaxy and AGN properties to accurately account for observational selection and systematic effects.

Recently, large-scale simulations that accurately reproduce galaxy properties have been developed (e.g. Kovacs et al. 2022; Wechsler et al. 2022; Gu et al. 2024; Euclid Collaboration 2024b), often adopting semi-empirical or hybrid methodologies. These methods not only save computation time but also ensure that results align well with most observed scaling relations. In addition, thanks to their optimised computational efficiency, they can also be quickly adapted to implement new empirical relations following new discoveries. Our work is conducted within this framework: we have adopted a semi-empirical method to produce realistic samples of synthetic galaxies and AGN, a component often neglected in previous similar works, necessary to enhance the realism of our simulation. At the same time, we preserve the link with the cosmic web traced by dark matter (DM) haloes: this connection is fundamental for deriving cosmological forecasts and linking the visible properties of galaxies and AGN to the DM distribution.

The inclusion of AGN in these catalogues is of the utmost importance; supermassive black holes (SMBHs) are believed to exist at the cores of most galaxies, and their masses are known to correlate closely with the masses, velocity dispersions, and luminosities of the host galaxy bulges (Richstone et al. 1998; Gebhardt et al. 2000; Merritt & Ferrarese 2001; Kormendy & Ho 2013). Furthermore, galaxies hosting actively accreting SMBHs – AGN – frequently present characteristic observational features (both in photometric and spectroscopic observations) across the full electromagnetic spectrum. Reproducing these features in realistic empirical catalogues is essential.

Although the aim of this work is to produce mock near-infrared (NIR)/optical catalogues, we decided to implement AGN starting from X-ray-selected samples due to the efficiency of this selection in identifying AGN across a wide range of redshifts and luminosities, including those with high levels of nuclear obscuration ($22 < \log N_{\rm H}/{\rm cm}^{-2} < 24$) that may be overlooked in other bands (e.g. Padovani et al. 2017). Since X-ray AGN photons are generated through non-thermal processes near the central black hole, they serve as an ideal tracer of AGN activity (Mushotzky et al. 1993; Brandt & Hasinger 2005). Additionally, the X-ray band, especially for the brightest sources, experiences minimal contamination from the host galaxy, which predominantly emits at different wavelengths.

The structure of this paper is the following. In Sect. 2, we describe the methodology followed to create the galaxy mock catalogue that was used as a base for the consequent AGN mock. In Sect. 3, we describe in detail the methodology followed to populate the galaxy mock with AGN. Then, in Sect. 4, we validate the catalogue by comparing its outputs with observations that were not used for its calibration, and finally, in Sect. 5, we show one example of how this catalogue can be used to make predictions for future surveys, focussing on *Euclid* as a representative case.

Throughout this paper, a standard cosmology ($\Omega_{\rm m} = 0.3$, $\Omega_\Lambda = 0.7$ and $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$) has been assumed. The stellar masses are given in units of solar masses for a Chabrier initial mass function (Chabrier 2003).

2. Mock galaxy catalogue: MAMBO

MAMBO (Mocks with Abundance Matching in BOlogna) is a workflow designed to construct an empirical mock catalogue of galaxies that can reproduce with accuracy their physical properties and observables, such as rest-frame and observed magnitudes and spectral features. A detailed description of the methodology and a validation of the galaxy properties can be found in Girelli (2021) and is briefly summarised in this section. A version of MAMBO, prior to the inclusion of AGN, has been used in Euclid Collaboration (2024d) to forecast the reliability of *Euclid* algorithms at recovering redshifts and physical parameters (stellar masses and star-formation rates) from *Euclid* photometry. In the rest of this paper, we focus our attention on the inclusion of AGN into this workflow. A schematic view of the steps explained in this section, as well as in Sect. 3, is presented in the flowchart of Fig. 1.

MAMBO takes in as inputs a few quantities from cosmological N-body DM simulations: the DM halo mass, expressed as M_{200}[1] for main haloes and $M_{\rm infall}$[2] for subhaloes and orphans[3]; and the redshift, z, of each halo and subhalo. The sky coordinates RA and Dec are not necessary inputs, but can be used to reconstruct the cosmic web and derive clustering and environmental properties of different classes of objects.

The results presented in the following are based on a lightcone built by Henriques et al. (2015) using the Mock Map Facility (MoMaF, Blaizot et al. 2005) on the Millennium Simulation (Springel et al. 2005); namely, lightcone number 23, which was chosen because it presents a mass function that is the closest to the mean of all the 25 available lightcones. The lightcone spans from $z = 0$ to $z = 10$ and contains DM haloes with a minimum mass of $M_{\rm halo} \gtrsim 10^{10.24}\,M_\odot\,h^{-1}$, which corresponds to 20 DM particles, and it covers an area of 3.14 deg^2. However, the method can be applied to any simulated catalogue of DM haloes and subhaloes.

In the first step, a galaxy with stellar mass $\mathcal{M}$ is assigned to each DM halo by means of a stellar-to-halo mass relation (SHMR). The SHMR was derived using a subhalo abundance matching technique, and calibrated on the Millennium lightcones by means of observed stellar mass functions (SMFs) on the SDSS (York et al. 2000), COSMOS (McCracken et al. 2012) and CANDELS (Grogin et al. 2011) fields. A detailed description of the SHMR can be found in Girelli et al. (2020) and Girelli (2021).

Additionally, every galaxy in the mock is classified as passive/quiescent (Q) or star-forming (SF) in a probabilistic way, following the relative ratio of the blue and red populations in observed SMFs. For this, we used the following SMFs: at $z\sim0$, the SMF evaluated by Peng et al. (2010) on the SDSS survey and divided into passive and SF using the rest-frame $(U - B)$ colour; at $0.2 < z < 4$, the SMF by Ilbert et al. (2013), derived on the COSMOS field and classified into red/blue using the rest frame colour selection $(NUV - r)$ vs $(r - J)$ (Ilbert et al. 2010).

At $z \geq 4$, SMFs divided by SF/Q type are not available. Recent studies have tried to put constraints on this fraction (Merlin et al. 2017; Girelli et al. 2019; Xie et al. 2024), but still with large uncertainties. For this reason, the SF fraction

[1] Mass within the radius where the halo has an overdensity 200 times the critical density of the simulation.

[2] Subhalo mass at the time it was accreted to the host halo; the infall mass is considered a better tracer of the potential well and then it correlates with galaxy properties such as the stellar mass (e.g. Conroy et al. 2006; Moster et al. 2010).

[3] Subhaloes that have lost all or part of their mass.

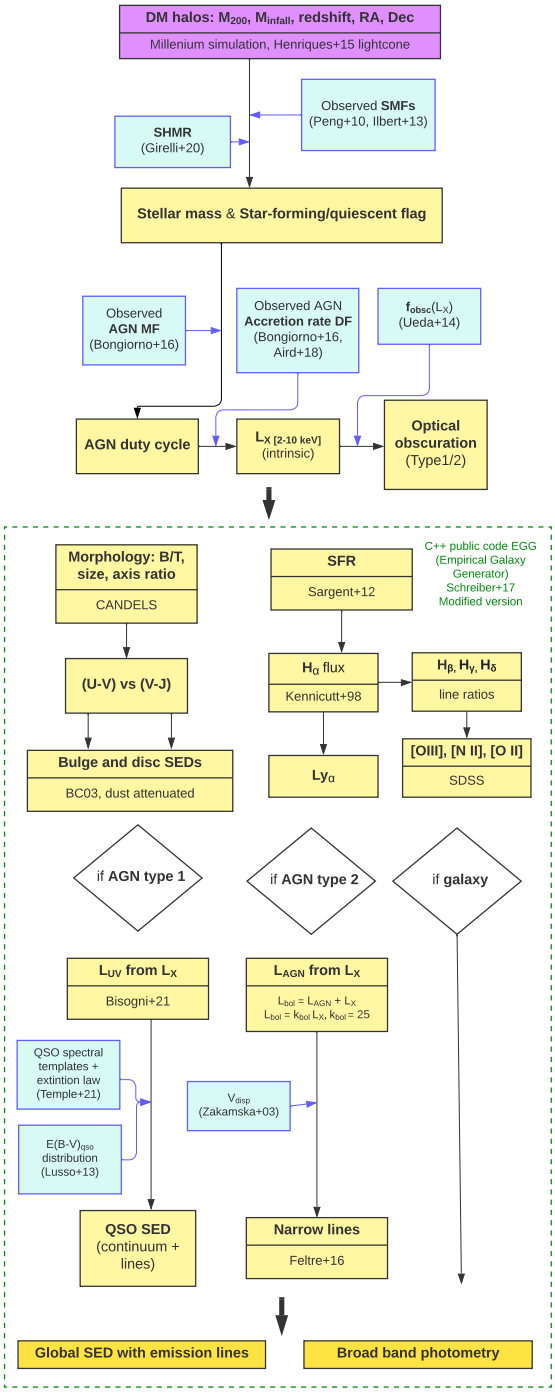

Fig. 1. Schematic view of the full workflow presented in this paper to create the galaxy and AGN mock catalogue. Yellow boxes represent the result of each step, while blue boxes show the necessary inputs. Steps within the dotted green box take place within the modified version of the public code EGG. A detailed description is given in Sects. 2 and 3.

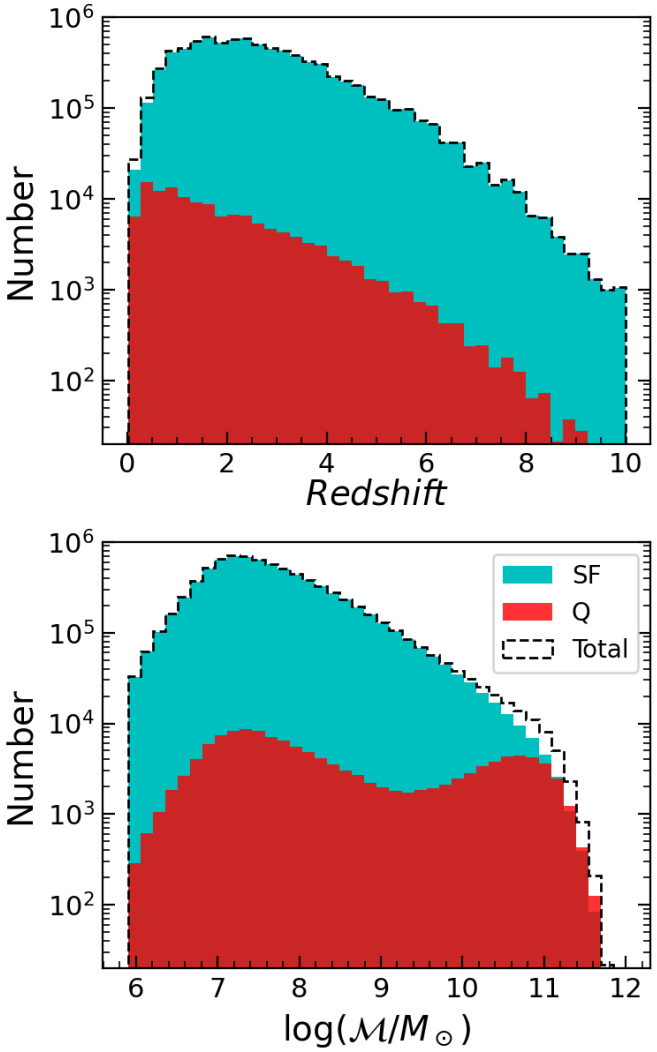

Fig. 2. Distribution of stellar mass and redshift of all the galaxies in the MAMBO lightcone used in this work, separated into the SF (blue) and quiescent (red) populations. The dashed black line shows the total (SF+Q) galaxy population.

was extrapolated from the results at lower redshifts, arriving at a maximum of $f_{SF} = 99\%$ at $z = 6$, which is kept constant at higher redshifts. This choice is motivated by the fact that at that redshift, the Universe is supposed to be too young (0.6 Gyr at $z = 6$) to contain any considerable fraction of quiescent galaxies. Recently, new results from JWST data are starting to find small samples of quiescent galaxies at $3 < z < 6$ (Carnall et al. 2023, 2024; Alberts et al. 2023), while Looser et al. (2024) reported the finding of a quiescent galaxy at $z = 7.3$. We plan to revise these assumptions once sufficient data becomes available to accurately constrain the statistics of this population.

We show in Fig. 2 the stellar mass and redshift distributions resulting from applying the SHMR and the SF/Q classification process described above. As it is visible from this figure, the galaxy catalogue is complete at least down to $\mathcal{M}{\sim}10^{7.5}\,M_{\odot}$, which is a consequence of the mass completeness of the DM halo catalogue. Additionally, it can be seen from this figure that the stellar mass distribution of quiescent galaxies follows the shape of a double Schechter function, as is often the case at redshifts where it is possible to observe also low mass galaxies (Drory et al. 2009; Peng et al. 2010; Ilbert et al. 2013; Davidzon et al. 2017; Weaver et al. 2023).

As a final step, other physical properties, as well as the photometry and spectra of the galaxies in the catalogue are retrieved with a modified version of the public code EGG (Empirical Galaxy Generator, Schreiber et al. 2017). EGG is a C++ code designed to generate an empirical mock catalogue of galaxies with realistic physical properties (such as the star formation rate, size, dust extinction, velocity dispersion, and emission line luminosities), where every galaxy is treated as a two component

system (bulge + disc). Additionally, the code produces the observed and rest-frame photometry in any desired band, and the redshifted (observer-frame) spectra from the ultraviolet (UV) to the submillimeter. The code has been calibrated purely by using empirical relations to produce realistic observable properties. In order to assign an SED (spectral energy distribution) to each galaxy, EGG selects an optical template from a prebuilt library of Bruzual & Charlot (2003) models, covering uniformly the observed part of the UVJ plane (Williams et al. 2009), which separates quiescent from SF galaxies. Infrared SEDs are instead derived from a set of libraries (Chary & Elbaz 2001; Magdis et al. 2012; Schreiber et al. 2017) aimed at reproducing dust emission, characterised by the values of infrared luminosity of dust and PAHs (polycyclic aromatic hydrocarbons) at $\lambda = [8, 1000]$ µm, the dust temperature, and the ratio of IR to 8 µm luminosity (IR8, Elbaz et al. 2011) for dust and PAH.

Afterwards, Gaussian emission lines with a given velocity dispersion[4] were assigned to the SEDs of the bulge and the disc components. EGG can take as an input the stellar mass, $\mathcal{M}$, redshift, z, and type (SF or quiescent) of each galaxy, or produce these quantities by randomly extracting z and $\mathcal{M}$ from the observed galaxy SMFs. We used EGG in the former configuration. More details about the way EGG works and about the modifications we made to the code are given in Sect. 3.5. It is worth stressing that the full MAMBO workflow can be applied to any DM simulation, and, similarly, the method presented in Sect. 3 to populate galaxies with AGN can be applied to any mock galaxy catalogue containing information about galaxy stellar mass, redshift, and galaxy type.

3. Adding AGN to the catalogue

In this section, we present the methodology adopted to populate our galaxy catalogue with AGN, which uses a completely empirical methodology comprising the following steps: *i)* we flagged each galaxy from the lightcone as hosting an AGN or not, following a probabilistic method that depends on $\mathcal{M}$ and z (Sect. 3.2); *ii)* every object flagged as AGN was assigned an intrinsic X-ray luminosity (Sect. 3.3); *iii)* we separated the AGN population into optically unobscured or obscured AGN (type 1 or type 2, respectively) by means of their intrinsic X-ray luminosity (Sect. 3.4); *iv)* we built the observed spectra and photometry of both type 1 and type 2 AGN with the help of photoionisation models of AGN narrow-line regions and parametric SEDs of typical quasi-stellar objects (QSOs), which model both the continuum and emission lines (see Sect. 3.5). A schematic view of this workflow is given in Fig. 1.

3.1. Specific accretion rate

The specific accretion rate ($\lambda_{\rm SAR} \propto L_{\rm X}/\mathcal{M}$)[5] is a directly measurable quantity that has been extensively used in the literature (Bongiorno et al. 2012; Aird et al. 2012; Georgakakis et al. 2014) as a proxy for the rate of accretion onto the SMBH ($\dot{M}_{\rm BH}$) relative to the stellar mass of the host galaxy. Furthermore, if a bolometric correction and an $M_{\rm BH}/\mathcal{M}$ scaling relation are assumed, the specific accretion rate can also be regarded as a proxy for the Eddington ratio of the SMBH, $\lambda_{\rm Edd} = L_{\rm bol}/M_{\rm BH}$ (Bongiorno et al. 2016; Aird et al. 2018).

In the following subsections, we use two specific accretion rate distribution functions (SARDFs) with different definitions. On the one hand, Bongiorno et al. (2016) defines the specific accretion rate as $\lambda_{\rm SAR} = L_{\rm X}/\mathcal{M}$. On the other hand, Aird et al. (2018) defines the specific black hole accretion rate ($\lambda_{\rm sBHAR}$) as the dimensionless quantity:

$$\lambda_{\rm sBHAR} = \frac{k_{\rm bol} L_{\rm X}}{1.3 \times 10^{38} {\rm erg\,s^{-1}} \times 0.002 \frac{\mathcal{M}}{M_\odot}}, \quad (1)$$

where $k_{\rm bol}$ is a bolometric correction factor ($L_{\rm bol} = k_{\rm bol} L_{\rm X}$), assumed to have a constant value of $k_{\rm bol} = 25$. This definition also assumes a constant scaling relation of $M_{\rm BH} = 0.002\mathcal{M}$ (Marconi & Hunt 2003, assuming also $\mathcal{M} \approx M_{\rm bul}$). Under these assumptions, $\lambda_{\rm sBHAR} = \lambda_{\rm Edd}$, and therefore an AGN accreting at 1% of the Eddington limit, would have $\lambda_{\rm sBHAR} \sim 10^{-2}$. It is possible to convert between the two definitions of the specific accretion rate as $\lambda_{\rm SAR} \approx 10^{34}\, \lambda_{\rm sBHAR}$. Therefore, with the same assumptions, the 1% of the Eddington limit corresponds to $\lambda_{\rm SAR} = 10^{32}\,{\rm erg\,s^{-1}}\,M_\odot^{-1}$. We note that both in Bongiorno et al. (2016) and in Aird et al. (2018), this was set as the lower limit to define a galaxy as hosting an AGN.

3.2. Fraction of AGN

In order to derive the probability of a galaxy with a given stellar mass, $\mathcal{M}$, and at a given redshift, z, hosting an active nucleus, $p\,({\rm AGN} \mid \mathcal{M}, z)$, we made use of the AGN host galaxy mass function (HGMF) derived by Bongiorno et al. (2016, hereafter B16). In B16, the authors studied a sample of 877 hard (2–10 keV) X-ray selected AGN from the XMM-COSMOS point-like source catalogue (Hasinger et al. 2007; Cappelluti et al. 2009; Brusa et al. 2010; Bongiorno et al. 2012) in the redshift range $0.3 < z < 2.5$ and with a limiting flux of $F_{\rm X} \sim 3 \times 10^{-15}\,{\rm erg\,s^{-1}\,cm^{-2}}$. This sample was also selected with a stellar mass limit of $\mathcal{M} > 10^{9.5}\,M_\odot$ and a specific accretion rate of $\lambda_{\rm SAR} > 10^{32}\,{\rm erg\,s^{-1}M_\odot^{-1}}$.

To derive the AGN HGMF and the SARDF, B16 corrected for the stellar mass incompleteness of the sample down to the above-mentioned limit in $\lambda_{\rm SAR}$. Additionally, B16 also accounted for the incompleteness due to the sources that were missed in the sample because of their high column density, $N_{\rm H}$. This was done considering column density values in the range of $20 < \log N_{\rm H}/{\rm cm^{-2}} < 24$, and therefore not including Compton-thick AGN, which are heavily obscured objects with $\log N_{\rm H}/{\rm cm^{-2}} > 24$. Although a significant fraction of AGN are expected to be CTK sources, their exact fraction is still a matter of debate (e.g. Buchner et al. 2015 found a constant fraction of $f_{\rm CTK} \sim 35\%$, independent of redshift and accretion luminosity, while Pouliasis et al. 2024 found a much lower fraction of $f_{\rm CTK} \sim 17\%$ for $3 \lesssim z \lesssim 6$ AGN). Furthermore, the redshift and luminosity-dependence of this fraction is yet not fully understood (e.g. Ricci et al. 2017a). Therefore, the inclusion of these sources is left to future work.

It is worth noting that the choice of a minimum value of $\lambda_{\rm SAR}$ sets the definition of AGN used in this work. In general, different criteria can be used to select AGN based on their X-ray emission. A common method is to select sources emitting above a threshold intrinsic X-ray luminosity (generally $L_{\rm X} > 10^{42}\,{\rm erg\,s^{-1}}$, e.g. Brandt & Alexander 2015). The AGN can also be selected according to the relative X-ray emission from non-nuclear mechanisms (such as X-ray binary stars and hot gas emission) with respect to the total $L_{\rm X}$ of the galaxy (e.g.

[4] We modified the original recipe in EGG with a mass-dependent $\sigma_{\rm gas}$ from Bezanson et al. (2018).

[5] Throughout this paper we use $L_{\rm X}$ and $F_{\rm X}$ to refer, respectively, to the intrinsic luminosity and observed flux in the hard ([2–10 keV]) X-ray band.

Birchall et al. 2020). It is well known that selecting AGN with these different criteria can lead to significant differences in their observed fraction (see e.g. Birchall et al. 2022), which should be taken into consideration when using this catalogue.

In B16, the authors derived the HGMF by jointly fitting the SMF and the SARDF (see Sect. 3.3), with the X-ray luminosity function (XLF) as an additional constraint. For this purpose, they used a maximum likelihood method to determine the HGMF and the SARDF as a bivariate distribution function of the stellar mass and a specific accretion rate, $\Psi(\mathcal{M}, \lambda_{\rm SAR}, z)$. As a result of this approach, the HGMF cannot be expressed as a simple analytic function. Instead, the authors provide an analytic approximation by performing a least-squares fit to the HGMF with a Schechter (Schechter 1976) function:

$$\Phi(\mathcal{M})\,\mathrm{d}\mathcal{M} = \Phi^{\star}\left(\frac{\mathcal{M}}{\mathcal{M}^{\star}}\right)^{\alpha} \exp\left(-\frac{\mathcal{M}}{\mathcal{M}^{\star}}\right) \mathrm{d}\left(\frac{\mathcal{M}}{\mathcal{M}^{\star}}\right), \quad (2)$$

evaluated at the centre of three redshift bins ($0.3 < z < 0.8$, $0.8 < z < 1.5$ and $1.5 < z < 2.5$). The best-fit parameters of this fit are given in Table 2 of B16[6].

The original slope (α) from the Schechter function in the first two redshift bins would produce an unrealistic overestimate of low-mass AGN when extrapolating this model to $\mathcal{M} < 10^{9.5} M_{\odot}$, which is the stellar mass limit in the sample of B16, while our catalogue spans to lower stellar masses, as is visible in Fig. 2 (see Appendix A for an example using the original slopes). Therefore, we re-derived these quantities by fitting the $1/V_{\rm max}$ points shown in B16 with a Schechter function using Eq. (2), obtaining $\alpha = -0.25 \pm 0.09$ for $0.3 < z < 0.8$, and $\alpha = -0.19 \pm 0.11$ for $0.8 < z < 1.5$. For the highest redshift bin, our fit was compatible with the value shown in Table 2 of B16; therefore, we did not modify it.

We then calculated $p\,(\mathrm{AGN} \mid \mathcal{M}, z)$ as the ratio between the SMF of the galaxies of the MAMBO lightcone and the HGMF from B16. Because the MAMBO lightcone covers a wider range of redshifts ($0 < z < 10$), we interpolated and extrapolated $p\,(\mathrm{AGN} \mid \mathcal{M}, z)$ from the centre of each redshift bin. For this, we assumed a minimum fraction, $p\,(\mathrm{AGN} \mid z = 0) = 0.01$, at all $\mathcal{M}$ when extrapolating at $z < 0.55$, while we maintained a constant fraction when extrapolating at $z > 2$. Afterwards, every galaxy was statistically assigned as either a host of an AGN or not a host, with a Bernoulli trial proportional to p(AGN) (i.e. by comparing p(AGN) to a random number extracted from a uniform probability distribution from 0 to 1).

Although the choice of a minimal fraction of 1% at $z \sim 0$ for all masses is a rough approximation, recent studies of the local Universe motivate this assumption. For example, Birchall et al. (2022) studied 917 X-ray selected AGN (found as XMM counterparts of 25 949 SDSS galaxies) with $z \leq 0.33$, which corresponds to a global AGN fraction of 3.5%. Instead, when selecting AGN by means of their accretion rate they found a fraction of about 1%, constant over stellar masses of $8 < \log(\mathcal{M}/M_{\odot}) < 12$, and increasing from 1% to about 10% with redshift. In Williams et al. (2022) the authors studied 213 *Chandra* X-ray counterparts of 280 nearby (<120 Mpc) galaxies from the Palomar sample and classified 14 (6.6%) of them as Seyferts, while only four (1.9%) of them have $L_{\rm X} > 10^{42}$. Similarly, Osorio-Clavijo et al. (2023) studied 138 *Chandra* X-ray counterparts of the CALIFA sample, with a wide range of stellar masses and $z < 0.1$ and found an AGN fraction of 5%. At $z > 2$, we chose to have a constant AGN fraction: despite being a rough approximation, with this choice we are able to reproduce the observed X-ray luminosity functions up to $z = 7$, as is shown in Sect. 4.2.

[6] B16 used a different definition for the Schechter function, and therefore to use the values given in their Table 2 one must remove the term $\mathrm{d}\left(\frac{\mathcal{M}}{\mathcal{M}^{\star}}\right)$ from Eq. (2).

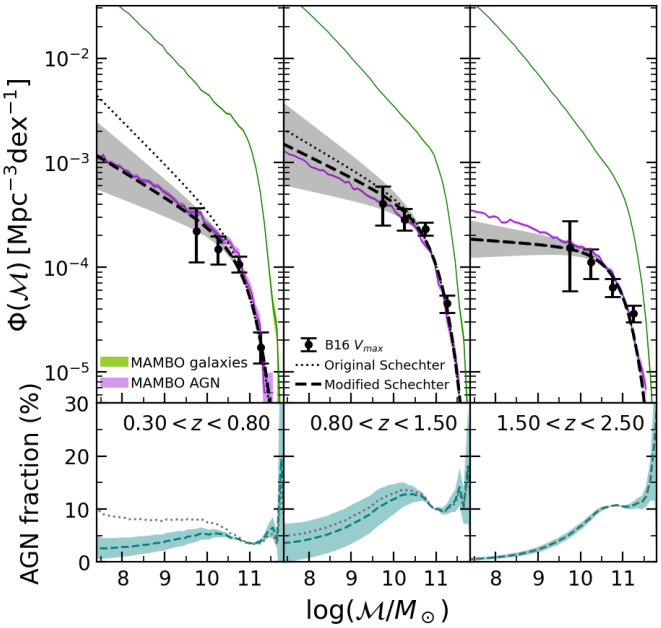

Fig. 3. Stellar mass functions of galaxies and AGN in the three redshift bins studied in B16. The green and purple shaded areas correspond to the galaxies and AGN from the MAMBO lightcone, respectively. We also show the best-fit Schechter function to the HGMF of AGN from Bongiorno et al. (2016) with a dotted black line, and the modified Schechter fit used in this work with a dashed black line. The shaded grey area shows the uncertainty in α. Black dots with error bars show the AGN mass function computed using the $V_{\rm max}$ method, derived in B16. The lower pannels show the AGN fraction derived as the ratio of the modified Schechter function and the galaxy mass function from MAMBO (dashed blue line), and the same but using the original Schechter function from B16 (dotted black line).

In the upper panels of Fig. 3, we show the best-fit Schechter function of the AGN HGMF (both the original from B16 and our modified fit) together with the SMF of the galaxies of the MAMBO lightcone and the AGN mass function of our catalogue, derived with the methodology described above. We also show the AGN HGMF computed using the $V_{\rm max}$ method, derived in B16, as an additional consistency check to the Schechter model. As was expected by construction, the AGN mass functions of MAMBO reproduce those of B16. An exception is the highest redshift bin, where the low-mass end of the HGMF of MAMBO is higher, but still compatible with the $V_{\rm max}$ estimate, due to the fact that we kept a constant $p\,(\mathrm{AGN} \mid \mathcal{M}, z)$ when extrapolating at $z > 2$ (see Fig. 4). In the lower panels of Fig. 3 we show the fraction of AGN over the galaxy population as a function of stellar mass and at a given redshift bin, $p\,(\mathrm{AGN} \mid \mathcal{M}, \langle z \rangle)$. As was expected, this fraction increases with increasing stellar mass of the host galaxy.

We show in Fig. 4 the probability of a given galaxy hosting an AGN as a function of redshift and in different mass bins for the full lightcone, after performing the interpolations and extrapolations described above. Each dot in this figure corresponds to a different galaxy, which is flagged as an AGN (or not) in a statistical way depending on this probability. This probability, at all masses, increases with redshift, reaching a maximum around $z \simeq 1$ and decreasing at higher redshifts. For comparison, Fig. 4

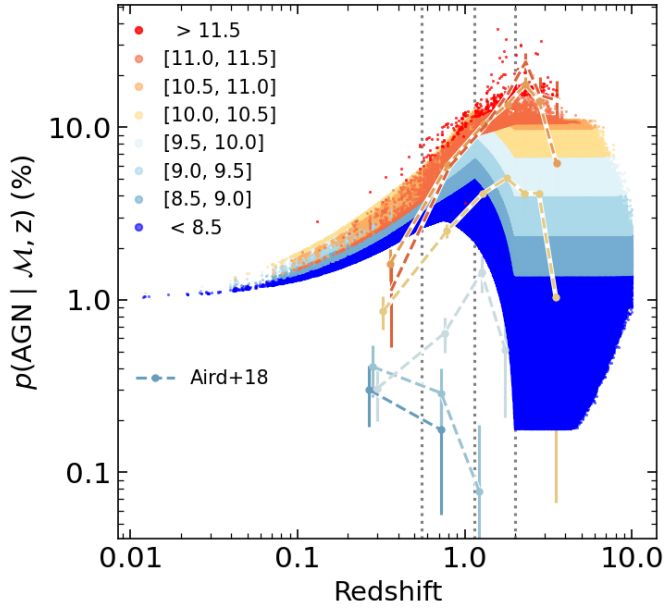

Fig. 4. Probability of a galaxy hosting an AGN as a function of redshift. Each point corresponds to a galaxy in the MAMBO lightcone, colour-coded in different $\mathcal{M}$ bins indicated in the legend. For comparison, the points with error bars connected by dashed lines show the AGN fraction derived in A18, also in different mass bins. Vertical dotted lines mark the centre of the three redshift bins studied in B16; namely, $z = 0.55, 1.15, 2.0$.

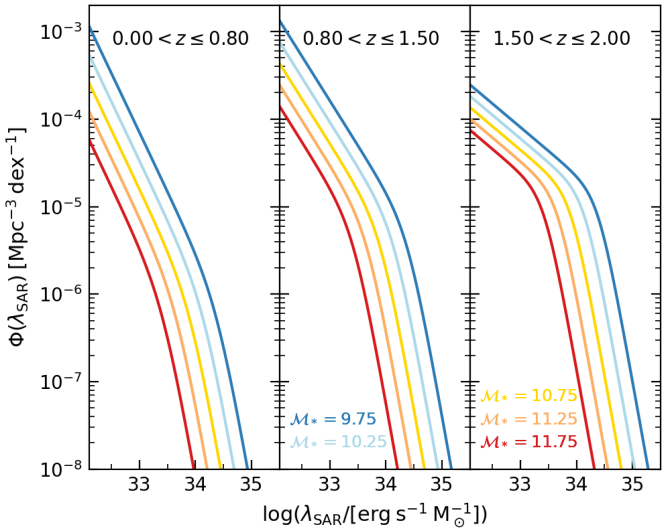

Fig. 5. Specific accretion rate distribution function (SARDF) computed at different values of $\mathcal{M}$ using Eq. (3), in the three redshift bins from B16.

also shows the fraction of AGN derived in Aird et al. (2018, hereafter A18), in which they derived the duty cycle of a sample of NIR-selected galaxies matched with X-ray data. We note that the fractions differ significantly, being higher in this work, in particular at $\mathcal{M} < 10^{10}\,M_\odot$. We show in Appendix A that using the AGN fractions from A18 as an input in our workflow produces an X-ray luminosity function that tends to be underestimated when comparing it to the observed ones, especially for $z \lesssim 1$ and $z \gtrsim 3$.

3.3. X-ray luminosity

In order to assign to each object flagged as AGN an intrinsic luminosity in the hard ([2–10 keV]) X-ray band, we first assigned a specific accretion rate. At $0 < z < 2$ we make use of the SARDF derived in B16. As is explained in Sect. 3.2, B16 derived the SARDF and the HGMF simultaneously as a bivariate distribution function of $\mathcal{M}$ and $\lambda_{\rm SAR}$; that is, $\Psi(\mathcal{M}, \lambda_{\rm SAR}, z)$. As was the case with the HGMF, this implies that the SARDF cannot be expressed as a simple analytic function of $\mathcal{M}$ and z. Instead, we used an analytic approximation of the SARDF (evaluated at the centre of three redshift bins) described as a double power-law (DPL) of the form

$$\Phi(\lambda_{\rm SAR}, \mathcal{M}) = \frac{\Phi^\star_\lambda}{\left(\frac{\lambda_{\rm SAR}}{\lambda^\star_{\rm SAR}(\mathcal{M})}\right)^{-\gamma_1} + \left(\frac{\lambda_{\rm SAR}}{\lambda^\star_{\rm SAR}(\mathcal{M})}\right)^{-\gamma_2}}, \quad (3)$$

where the mass dependence is given by $\log \lambda^\star_{\rm SAR}(\mathcal{M}) = \log \lambda^\star_{\rm SAR,0} + k_\lambda(\log \mathcal{M} - \log \mathcal{M}_0)$, where $\log \lambda^\star_{\rm SAR,0} = 33.8$, $\log \mathcal{M}_0 = 11$, and $k_\lambda = 0.58$. The best-fit values of the normalisation, $\Phi^\star_\lambda$, and slopes, γ_1, γ_2, of the DPL evaluated at the centre of the three redshift bins are given in Table 3 of B16.

Using Eq. (3), we computed the SARDF at five values of $\mathcal{M}$, from $\log(\mathcal{M}/M_\odot) = 9.75$ to $\log(\mathcal{M}/M_\odot) = 11.75$ in steps of 0.5 dex in $\mathcal{M}$ (Fig. 5). We then divided the AGN in our lightcone into five mass bins, each of them centred at one of the above-mentioned $\mathcal{M}$ and all of them with a width of 0.5 dex in $\mathcal{M}$, except for the lowest-mass bin, extending up to the lowest $\mathcal{M}$ in our lightcone. Then, we normalised the $\lambda_{\rm SAR}$ distribution functions by dividing them by the integrated AGN density ($\rm Mpc^{-3}$) in each bin of $\mathcal{M}$ and z, and therefore converting these distributions into probability distributions. We assigned to each AGN a value of $\lambda_{\rm SAR}$ by randomly extracting values from the corresponding probability distribution at each $\mathcal{M}$ and z bin.

At higher redshifts ($z > 2$), we instead made use of the accretion rate distributions from A18, who studied a sample of 1797 X-ray *Chandra* counterparts of 126 971 NIR-selected galaxies up to $z \sim 4$ and with stellar and $8.5 < \log(\mathcal{M}/M_\odot) < 11.5$. Using a methodology similar to that used for the B16 accretion rate distributions previously described, we constructed the cumulative distribution function relative to each $p(\log \lambda_{\rm sBHAR} \mid \mathcal{M}, z)$ distributions in the range of $-2 < \log \lambda_{\rm sBHAR} < 1$ and $2 < z < 4$, and used it to assign a value of $\lambda_{\rm sBHAR}$ to each object previously labelled as an AGN at $z > 2$. Since A18 calculated the $\lambda_{\rm sBHAR}$ distributions separately for SF and quiescent galaxies, we used both sets of distributions for the galaxies in our catalogue that are split into these two classes. We also explored the possibility of using the accretion rate distributions from A18 to assign $L_{\rm X}$ at all redshifts, but we decided to use the methodology presented in this section since it produces an XLF that is in better agreement with the observed ones (see Appendix A).

After assigning every AGN in the lightcone a specific accretion rate, we converted this into the intrinsic X-ray luminosity (in the 2–10 keV band). We show in Fig. 6 the histogram of the assigned $L_{\rm X}$ in different redshift bins for the entire population of AGN. We note that, as is visible in Fig. 6, a large fraction of AGN have luminosities of $L_{\rm X} < 10^{42}\,\rm erg\,s^{-1}$; that is, below a classical threshold applied to define an X-ray emitter as an AGN. The reason for this is twofold: on the one hand, as was previously mentioned, we started from a sample of AGN selected above a given accretion rate limit and not above a $L_{\rm X}$ limit. On the other hand, the galaxy mock probes low stellar masses ($\mathcal{M} \sim 10^7\,M_\odot$), which translates into low X-ray luminosities.

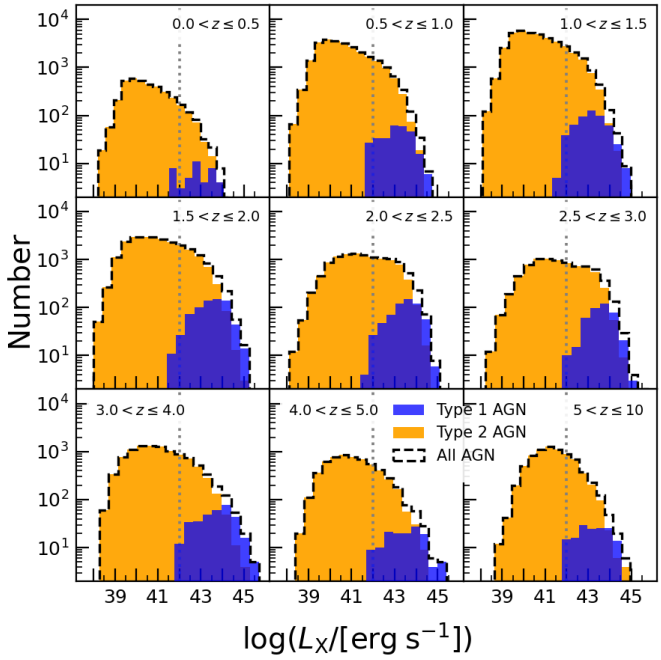

Fig. 6. Distribution of intrinsic hard-band X-ray luminosity (L_X) of all the AGN in the lightcone, separated into type 1 (blue) and type 2 (orange) AGN, in different redshift bins. We also show with dashed and solid lines the subpopulations of type 1 and 2 in SF and quiescent galaxies. The vertical dotted line marks $L_X = 10^{42}$ erg s^{-1}, the threshold typically applied to separate X-ray emission from AGN or from other origins.

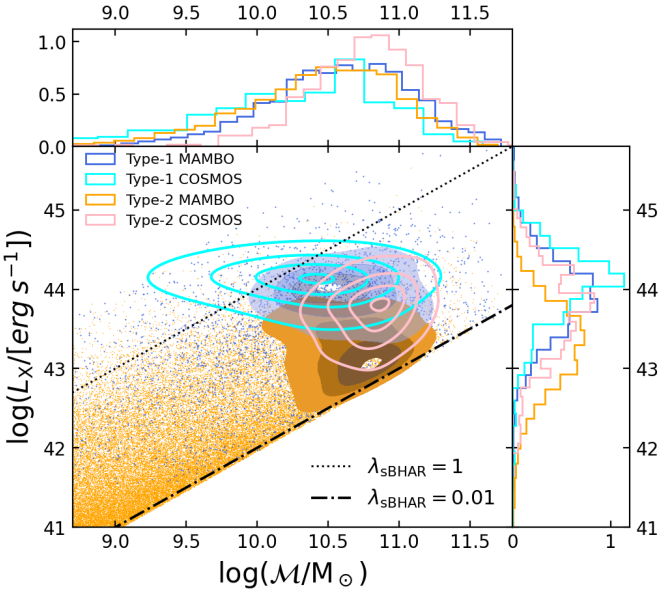

Fig. 7. AGN from our lightcone scattered in the L_X–$\mathcal{M}$ plane, separated into type 1 (blue) and type 2 (orange) AGN. The points in the central panel show the distribution of all the sources, while the solid contours and the histograms from the upper and right panel correspond to sources selected above $F_X = 1.9 \times 10^{-15}$ erg s^{-1} cm^{-2}. We show also the distribution of sources from the Chandra COSMOS Legacy Spectral Survey (Marchesi et al. 2016b), with the same cut in F_X, and separated into type 1 (cyan) and type 2 (pink) AGN. The contour levels represent iso-density lines, corresponding to the 50th, 75th, 90th, and 99th percentiles of the distribution. The dash-dotted and dotted black lines mark the locus where $\lambda_{sBHAR} = 0.01$ and 1, respectively, which assuming a mean bolometric correction of $k_{bol} = 25$ and a constant mass ratio of black hole to host galaxy of $M_{BH} \approx 0.002\mathcal{M}$ correspond approximately to 1 and 100% of the Eddington limit, respectively.

In Fig. 7, we show the distribution of the AGN in our mock in the L_X–$\mathcal{M}$ plane. As was expected by construction, all the AGN have accretion rates above $\lambda_{sBHAR} > -0.01$ ($\equiv \lambda_{SAR} > 10^{32}$ erg s^{-1}M$_\odot^{-1}$), and only a small fraction of them are above the Eddington limit of $\lambda_{sBHAR} = 1$.

3.4. Obscuration model

The observed spectra of AGN in different bands can vary significantly depending on the level of obscuration of the nucleus by its surrounding material along the line of sight. In this context, AGN are generally separated into two main sub-classes, namely type 1 and type 2 AGN (unobscured and obscured, respectively). In the UV/optical/NIR bands, type 1 AGN are characterised by the presence of broad optical emission lines (full width at half-maximum $\gtrsim$1000 km s^{-1}) and a SED with a bluer continuum, while type 2 AGN typically do not present these characteristics. In reality, AGN display a much wider observational variety, which allows us to classify them in many other sub-classes (e.g. Antonucci 1993; Urry & Padovani 1995; Spinoglio & Fernández-Ontiveros 2021), but such classification is beyond the scope of this work.

In order to separate the AGN of our catalogue into optically obscured and unobscured, we make use of the results presented in Merloni et al. (2014). In this study, the authors studied a sample of 1310 AGN selected from the XMM-COSMOS point-like source catalogue (Hasinger et al. 2007; Cappelluti et al. 2009) with a limiting rest-frame X-ray flux of $F_X = 2 \times 10^{-15}$ erg cm^{-2} s^{-1} in the redshift range $0.3 < z < 3.5$. For this sample, the authors studied the luminosity dependence of the (optically) obscured fraction of AGN, and found a relation, almost redshift independent, of the form

$$F_{obs} = A + \frac{1}{\pi} \operatorname{atan}\left(\frac{l_0 - \log L_X}{\sigma_x}\right), \quad (4)$$

where F_{obs} is the fraction of type 2 AGN with respect to the total population, L_X is the intrinsic X-ray luminosity in the hard band, and the best-fitting parameters are $A = 0.56$, $l_0 = 43.89$ and $\sigma_x = 0.46$. For this work, we assumed this relation to be valid for the full redshift range $0 < z < 10$ without any redshift evolution. Using this relation, we assign to every AGN a probability of being optically obscured depending on their L_X, and statistically separate every AGN in type 1 and type 2.

Additionally, the X-ray emission from AGN (especially in the soft band) can be heavily obscured by gas and dust surrounding the SMBH. To quantify this effect, we assign to each AGN a value of absorption column density (N_H), following the absorption function presented in Ueda et al. (2014), within the range $20 < \log N_H/\mathrm{cm}^{-2} < 24$. The choice of this range is consistent with the fact that, as stated in Sect. 3.2, Compton-thick AGN ($\log N_H/\mathrm{cm}^{-2} > 24$) are not included in our model. Following Sect. 3 of Ueda et al. (2014), we assign N_H in a probabilistic way as a function of L_X and redshift. This allows us to classify the AGN in our catalogue as X-ray type 1 and type 2, where type 1 X-ray AGN are defined as an object with $\log N_H/\mathrm{cm}^{-2} < 22$, and vice-versa.

It is worth noting that in nature these two classifications, although they on average correlate with each other, do not match perfectly; that is, some objects present optical obscuration but

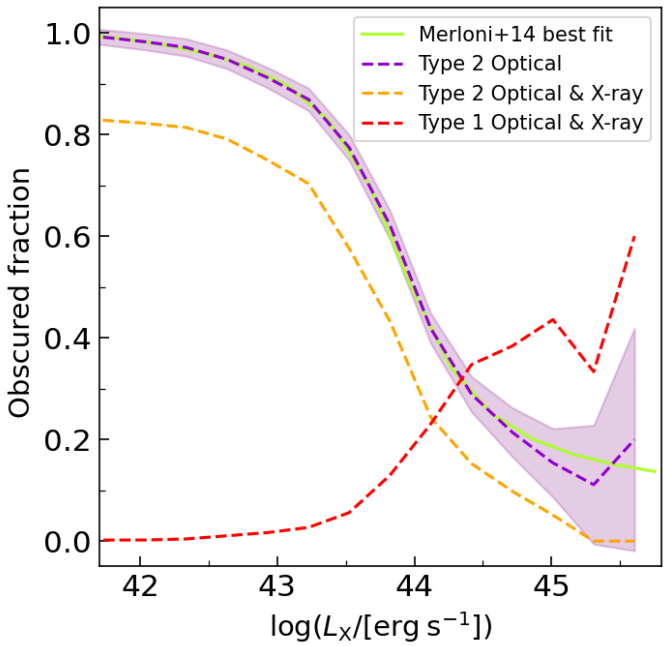

Fig. 8. Fraction of optically obscured (type 2) AGN as a function of intrinsic X-ray luminosity. The solid green line shows the relation used to separate AGN (Eq. (4), originally from Merloni et al. 2014) into optically obscured/unobscured, while the dashed purple line shows the actual fraction of type 2 AGN in our catalogue (with the Poissonian uncertainty shown by the shaded area). The other two dashed lines give information about the X-ray obscuration (where type 1 X-ray AGN are defined as objects with $\log N_{\rm H}/{\rm cm}^{-2} < 22$ and vice-versa): the dashed orange line shows the fraction of objects classified as type 2 both in optical and the X-ray, while the dashed red line represents the fraction of all the AGN that are type 1 in both bands at the same time.

not X-ray obscuration and vice versa (e.g. Marchesi et al. 2016b; Merloni et al. 2014). Because the aim of this work is to produce a catalogue to be used in optical/NIR surveys (e.g. *Euclid*, MOONS), in the following sections we focus on type 1 AGN defined as optically unobscured sources, regardless of their X-ray obscuration.

In Fig. 6, we show the $L_{\rm X}$ distribution separated into type 1 and type 2 AGN. It is visible from this figure that most of the AGN with $L_{\rm X} < 10^{42}$ erg s^{-1} are classified as type 2 AGN in our catalogue.

In Fig. 7, we show also the distribution of sources from the Chandra COSMOS Legacy Spectral Survey (Marchesi et al. 2016b) in the $L_{\rm X}$–$\mathcal{M}$ plane, selected above a minimum X-ray flux $F_{\rm X} = 1.9 \times 10^{-15}$ erg s^{-1} cm^{-2}, which is the flux limit reported in Marchesi et al. (2016b). We applied the same flux cut to the AGN from our catalogue (contours in Fig. 7) in order to compare with the COSMOS data. With this flux cut, our AGN and the COSMOS AGN are limited to $z \lesssim 3.5$. We observe a reasonable agreement between the MAMBO and the COSMOS distributions, although MAMBO type 1 AGN seem to cover a narrower range in stellar mass than the COSMOS ones, and MAMBO type 2 AGN have on average lower X-ray luminosities than their COSMOS counterparts.

In Fig. 8, we show the fraction of optically obscured AGN as a function of $L_{\rm X}$ for the presented lightcone, and compare it with the calibration used to derive it; that is, Eq. (4). We also show the fraction of objects that are obscured both in optical and X-ray, and those that are unobscured in both bands.

3.5. AGN emitted spectra

After every galaxy in the catalogue has been classified as either hosting an AGN or not, and AGN have been characterised in terms of their X-ray luminosity and optical obscuration, we employ the publicly available code EGG (Empirical Galaxy Generator, Schreiber et al. 2017) to assign the rest of physical properties and observables (e.g. bulge/disc ratio, dust attenuation, etc.). Additionally, EGG allows us to generate the photometry of every object in any desired band, as well as the complete SED from the UV to the submillimeter.

In EGG, each galaxy is represented as a two-component system composed of a disc and a bulge, each of which is associated with a distinct SED, selected from a predefined lookup table. The choice of the galaxy SED is based on specific recipes that are tied to three main galaxy properties: its total stellar mass, its redshift, and its type (SF or quiescent).

We modified the original code by adding a third component to account for the AGN emission. For type 1 AGN (optically unobscured), this component includes the continuum and narrow and broad line emission typical of QSO, while for type 2 it accounts only for narrow-line emission (i.e. the galaxy stellar continuum is assumed to dominate at all wavelengths). For normal galaxies, this component remains effectively null.

3.5.1. Type 1 spectral energy distributions

In order to construct the SEDs of type 1 AGN in the lightcone, we made use of the parametric model developed by Temple et al. (2021). This model allows for the generation of synthetic quasar SEDs over the rest-frame wavelength range 912 Å to 3 µm. These synthetic SEDs have been shown to accurately reproduce, to a high degree of accuracy, the observed-frame optical and NIR colours of large samples of quasars, over redshifts $0.2 \leq z \leq 7$ and absolute magnitude in the i-band $-29 < M_i < -22$.

The observed variety in emission line properties is included in the model through the interpolation between two emission-line templates, which correspond to the observed limits of very strong and very weak emission in terms of the equivalent width of high-ionisation UV lines, such as C IV. Furthermore, observations from quasar spectra show that the equivalent width of strong emission lines is anti-correlated with the intrinsic luminosity of the source. This phenomenon is known as the Baldwin effect (Baldwin 1977). In order to reproduce this phenomenology, the model from Temple et al. (2021) incorporates a single parameter (`emline_type`) that allows for the generation of spectra with different emission line properties. This parameter is related to the absolute magnitude M_i of the quasar by means of the following empirical equation:

$$\texttt{emline_type} = 0.183 \times (M_i + 27). \quad (5)$$

For this work, we constructed the QSO SED library by changing only this parameter between its minimum and maximum values (–2 to +3), ranging from spectra with weak, highly blueshifted lines to those with strong, symmetric lines.

For each type 1 AGN, we modified the code EGG to select a SED from the pre-built library of QSO SEDs following Eq. (5). The SED is rescaled to the $L_{\rm UV}$ (at 2500 Å) that corresponds to that object according to the observed $L_{\rm X}$–$L_{\rm UV}$, as is presented in Bisogni et al. (2021). This relation is parametrised as

$$\log(L_{\rm X}) = \gamma \log(L_{\rm UV}) + \beta, \quad (6)$$

where the proxies for the UV and X-ray emissions correspond to the monochromatic rest-frame 2500 Å and 2 keV luminosities,

Table 1. Filters used in this work.

Filter	λ_{ref}	λ_{min}	λ_{max}
I_E	7103	5300	9318
H_E	17 649	14 971	20 568
u	3680	3206	4081
z	8685	8035	9375
GALEX FUV	1535	1340	1809

Notes. All wavelengths are given in Å. Source: SVO Filter Profile Service (Rodrigo et al. 2012).

respectively. The authors in Bisogni et al. (2021) find an average value for the slope of the relation of $\gamma = 0.58 \pm 0.06$ up to $z \approx 4.5$ and a dispersion of $\delta = 0.24$ dex, which we used also for this work. For this, we calculated the monochromatic luminosity at 2 keV from L_X following the general relation between the total luminosity in the band ($L_{E_1-E_2}$) and at a monochromatic energy (L_E); that is,

$$L_E = \frac{(2-\Gamma)E^{1-\Gamma}}{E_2^{2-\Gamma} - E_1^{2-\Gamma}} L_{E_1-E_2}, \quad (7)$$

where in this case $E = E_1 = 2$ keV, $E_2 = 10$ keV, and we used a fixed value for $\Gamma = 1.8$. Therefore, following Eqs. (6) and (7) we derive the monochromatic luminosity L_{2500} from the full band luminosity L_X, and use it to rescale the SED.

We apply the reddening by dust attenuation to the QSO SED using an empirically derived extinction curve presented in Temple et al. (2021). This extinction curve is similar to that of the Small Magellanic Cloud for $\lambda \gtrapprox 1700$ Å, while it increases less rapidly for shorter wavelengths. The $E(B-V)$ of each source is chosen randomly from the distribution presented in Fig. 2 of Lusso et al. (2013), which peaks at low $E(B-V)$, with a median value $\langle E(B-V) \rangle = 0.03$.

Finally, the galaxy SED (bulge + disc components) is added to that of the QSO in order to create the composite spectrum. To quantify how much each type 1 AGN is dominated by the host-galaxy stellar light or by the nuclear emission, we derived the quantity f_{AGN}, which is defined as

$$f_{AGN_{band}} = \frac{F^{AGN}_{band}}{F_{band}}, \quad (8)$$

where F^{AGN}_{band} is the flux in a given band from the AGN component only, and F_{band} is the equivalent for the total spectrum (i.e. disc + bulge + AGN). With this definition, f_{AGN} ranges from 0 to 1, where an object with $f_{AGN} = 0\,(1)$ would be completely dominated by the galactic (AGN) component, and $f_{AGN} = 0.5$ corresponds to the limit where both the galactic and the nuclear components contribute equally to that specific band. We derived this quantity in two bands: using the rest-frame magnitude from the GALEX FUV filter (which corresponds to the wavelength range where AGN emission typically peaks), and using the observed magnitude m_H (see Table 1).

In Fig. 9, we show an example of rest-frame SED for both a type 1 and a type 2 AGN from the lightcone. We show also in the figure the value of the most relevant physical parameter related to the construction of the SED, such as $\mathcal{M}$, L_X, SFR and, only for the type 1 AGN, $f_{AGN_{UV}}$ and f_{AGN_H}.

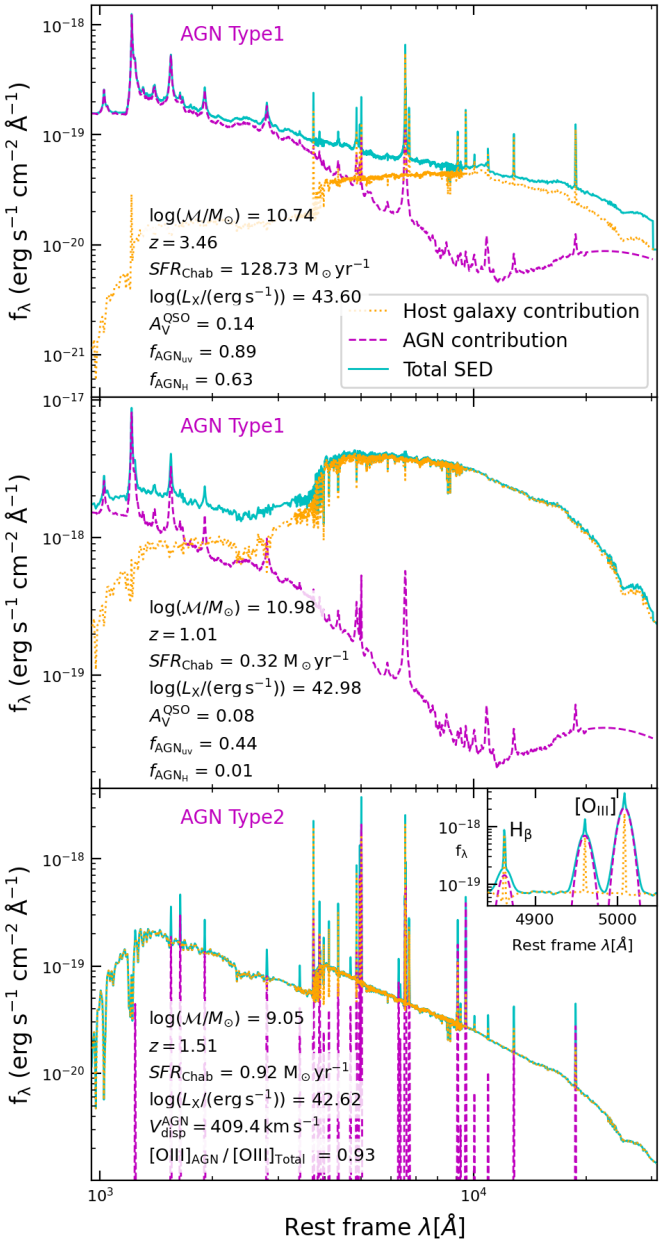

Fig. 9. Some examples of SEDs of AGN from our lightcone. We show three representative cases: in the upper panel, we show a type 1 AGN with dominant QSO component in the rest-frame UV band, while in the middle panel, we show a type 1 AGN with similar QSO and host-galaxy contribution for the same band. The lower panel shows a type 2 AGN, with a zoom-in subpanel showing the H_β and the [O III] $\lambda\lambda$ 5007, 4959 doublet emission lines. The galaxy component (bulge + disc) is shown with a dotted orange line, and is based on Bruzual & Charlot (2003) models. The AGN component, which consists on a full SED template from Temple et al. (2021) for type 1 and narrow emission lines for type 2 is shown with a dashed violet line. The cyan line shows the composed SED. Dust absorption is applied in all cases.

3.5.2. Type 2 spectral energy distributions

For objects flagged as type 2 AGN, typical narrow emission lines from AGN are added to the SED of the bulge component of the host galaxy. The emission lines are modelled with a Gaussian profile, with a total luminosity that was computed using photoionisation models made by Feltre et al. (2016). These models were generated using a standard photoionisation code (CLOUDY,

Ferland et al. 2017), and in them, the luminosity of the lines depends on a series of input parameters that describe the physical properties of the region where the narrow lines of AGN are emitted (the narrow-line region, or NLR), and which we explain in detail below:

Gas metallicity: $\log(O/H) + 12$. The gas-phase oxygen abundance is chosen randomly in the range $8.7 < \log(O/H) + 12 < 9.3$, where we adopted the value $\log(O/H)_\odot + 12 = 8.71$ for the (gas-phase) solar metallicity, as in Feltre et al. (2016) and Gutkin et al. (2016) for a corresponding dust-to metal mass ratio $\xi_d = 0.3$ (see below). The fact that we are using only models with solar or super-solar metallicities for the NLR of type 2 AGN is motivated by the fact that these are the models that best sample the region covered by local Seyfert galaxies in standard emission-line diagnostic diagrams (Feltre et al. 2016). These values are also consistent with those found in local AGN from observations (e.g. Peluso et al. 2023).

Ionising spectrum. The ionising spectrum used in the CLOUDY models to represent the accretion disc of the AGN has the shape $S_\nu \propto \nu^\alpha$ for the wavelength range $0.001 \leq \lambda/\mu m \leq 0.25$, where the power-law index α is an adjustable parameter. We used only models with either $\alpha = -1.4$ or $\alpha = -1.7$, randomly chosen, which sample the centre of the range modelled by Feltre et al. (2016); that is, -1.2 to -2.

Ionisation parameter: $\log U$. The ionisation parameter is defined as the dimensionless ratio of the number density of H-ionising photons to that of hydrogen. Using a combination of photoionisation models and high-resolution cosmological zoom-in simulations of galaxies, Hirschmann et al. (2017) found that, at fixed stellar mass, U is one of the main physical parameters driving the cosmic evolution of optical-line ratios. To reproduce this effect in our catalogue, the ionisation parameter is selected randomly (evolving with redshift) within the following ranges[7]:

$$\begin{cases} -5 < \log U < -3 & \text{for} \quad 0 < z < 1 \\ -4 < \log U < -2 & \text{for} \quad 1 < z < 2 \\ -3.5 < \log U < -1.5 & \text{for} \quad z > 2, \end{cases} \tag{9}$$

where these ranges have been derived from Hirschmann et al. (2017) (their Fig. 6, central panel). We note that this is the only NLR parameter in our simulation for which we assumed an evolution with redshift, and therefore the redshift evolution of AGN narrow-line ratios is purely linked to that of $\log U$.

Hydrogen number density, n_H: the volume-averaged hydrogen density of the narrow-line region. Chosen randomly $n_H = 10^3$ or $n_H = 10^4\,cm^{-3}$, which are typical gas densities estimated from optical line-doublet analyses of NLRs (see e.g. Osterbrock & Ferland 2006; Binette et al. 2024).

Dust-to-metal mass ratio, ξ_d, which accounts for the depletion of metals onto dust grains in the ionised gas. We used only models with $\xi_d = 0.3$, which implies assuming that 30 per cent by mass of all heavy elements are in the solid phase (Feltre et al. 2016; Gutkin et al. 2016).

In these models, the intensity of the lines is scaled to the accretion luminosity, L_{acc}, of the AGN; that is, the luminosity due to the accretion onto the central black hole. Assuming that the bolometric luminosity coming from the AGN, L_{bol}, is the sum of L_{acc} and L_X, and that L_{bol} can be retrieved from the X-ray luminosity with a bolometric correction ($L_{bol} = k_{bol} L_X$), L_{acc} can be deduced from the following equation:

$$L_{acc} = L_X(k_{bol} - 1), \tag{10}$$

where again we chose $k_{bol} = 25$ for consistency with the rest of the work.

The emission lines of type 2 AGN are characterised by velocity dispersions that are higher than those of SF galaxies (with FWHM lower than a few hundred $km\,s^{-1}$), but lower than those of their type 1 counterparts (with FWHM $\gtrsim 1000\,km\,s^{-1}$). To model this, we used the results from Menzel et al. (2016), who studied the spectroscopic properties of a sample 2578 X-ray selected AGN in the redshift range $z = [0.02, 5.0]$. We modelled the FWHM distribution of the Hβ emission line (for $FWHM_{H\beta} \lesssim 1000\,km\,s^{-1}$) shown in their Fig. 6 as a log-normal distribution centred at $355\,km\,s^{-1}$, covering the range FWHM $\sim [200, 1000]\,km\,s^{-1}$. We assigned randomly the velocity of dispersion of type 2 AGN in our catalogue following this distribution.

Finally, it is worth noting that we did not add an AGN component to the continuum of the host galaxies of type 2 AGN. While at UV and optical wavelengths this approach can be a good approximation for galaxies with strongly attenuated AGN emission, the AGN contribution is expected to dominate in the mid and far IR, even for the most attenuated sources, due to the emission from the dusty torus surrounding the accretion disc. The inclusion of such an AGN component is left to future work.

The full list of narrow-region lines added to the spectra of type 2 AGN is reported in Table 2. We show in Fig. 9 an example of type 2 AGN with these lines. In a similar fashion as we did for type 1 AGN, we quantified the dominance of the type 2 AGN component with respect to the total (AGN + host galaxy) using the ratio of the emission line flux of $[O\,III]_{\lambda 5007}$ from the AGN component with respect to the total. This quantity is shown also in Fig. 9.

Table 2. Narrow-region lines added to type 2 AGN spectra.

Line	Wavelength	Line	Wavelength
Ne V	1242.80	[O I]	6302.05
C IV	1549.86	[O I]	6365.54
He II	1640.42	[N II]	6549.86
C III]	1907.71	[N II]	6585.27
Mg II	2795.53	[S II]	6718.32
Mg II	2802.71	[S II]	6732.71
[Ne V]	3426.85	[S III]	6312.06
[O II]	3728.49	[S III]	9071.10
[Ne III]	3870.16	[S III]	9533.20
He II	4686.01	Balmer series (Hα to Hη)	
[O III]	4960.30	Paschen series (Paα to Paη)	
[O III]	5008.24		

Notes. Wavelengths are given in vacuum, in units of Å.

4. Validation of the catalogue

In this section we compare the physical properties of the AGN in our catalogue with data that were not used for its calibration, in order to test the validity of its predictions.

[7] These values correspond to the volume-averaged ionisation parameter $\langle U \rangle$, as defined in Eq. (1) of Hirschmann et al. (2017). Instead, Feltre et al. (2016) used a different definition, namely the ionisation parameter at the Strömgren radius (U_S). The conversion between these two quantities is $U_S = \frac{4}{9}\langle U \rangle$.

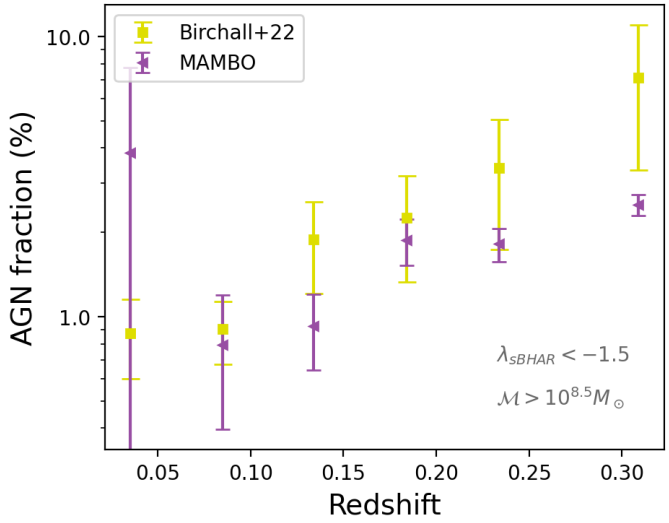

Fig. 10. Fraction of AGN at $z < 0.33$ from our catalogue (purple triangles), selected with $\log \lambda_{\rm sBHAR} < 1.5$ to compare them with the fraction reported in Birchall et al. (2022) (yellow squares). The error bars of the MAMBO AGN fraction show the Poissonian uncertainty.

4.1. AGN fraction

The AGN HGMF that we used to calibrate the fraction of AGN over the total galaxy population at $z < 2$ (Bongiorno et al. 2016) is defined for $z > 0.30$ and $\mathcal{M} > 10^{9.5}\,M_\odot$. To test the validity of the extrapolations we did at lower redshifts and stellar masses, we compare our results with different recent works. For example, in Fig. 10 we compare the fraction of AGN at $z < 0.33$ in different redshift bins with the results from Birchall et al. (2022), who studied 917 X-ray counterparts of SDSS galaxies. To make our sample as similar as possible to that of Birchall et al. (2022), we further selected the MAMBO AGN with $\log \lambda_{\rm sBHAR} < 1.5$ and $\mathcal{M} > 10^{8.5}\,M_\odot$. We see that for all redshift bins the fractions are roughly consistent within their uncertainities.

Regarding low-mass galaxies, Latimer et al. (2021) studied a sample of 495 dwarf local galaxies ($\mathcal{M} \leq 3 \times 10^9\,M_\odot$, $z \leq 0.15$) observed by eRosita, and found an upper limit of 1.8% for the AGN fraction. In the same redshift and mass bin our catalogue yields a fraction of 1.4 ± 0.2%, which is roughly compatible with these results. On the other hand, Mezcua & Domínguez Sánchez (2024) found a much higher AGN fraction (about 20%) when studying a sample of dwarf local galaxies from the MaNGA survey using integral-field spectroscopy. We note that the exact AGN fraction at these low stellar mass and redshift regimes is still a matter of debate, and therefore the predictions from our catalogue should be taken with caution here.

4.2. X-ray luminosity

In order to validate the X-ray properties of our catalogue we used two main observables, the X-ray luminosity function and the $F_{\rm X}$ cumulative number counts. These comparisons are relevant since they give hints about the purity and completeness of our catalogue in comparison to other X-ray selected catalogues.

Figure 11 shows the hard X-ray luminosity function (XLF) of our mock catalogue compared with several observational luminosity functions from the literature. We note that the XLF by Miyaji et al. (2015) was used in B16 as an extra observational constraint to determine their HGMF and SARDF, and therefore our catalogue should reproduce it by construction. In Fig. 11 the XLF from Buchner et al. (2015) is scaled by a factor 0.65 in order to remove the contribution of Compton-Thick (CTK) sources, which were not considered in the work of B16, and therefore are not represented in our catalogue. This factor was chosen because Buchner et al. (2015) found a constant fraction of CTK objects over the total AGN population of about 35% independent of z and $L_{\rm X}$. We observe in general a good agreement between the XLF from our mock and the observed ones at all redshift bins.

As a further check, in Fig. 12 we show the cumulative number counts of objects above a given X-ray flux. For this, we estimated the X-ray flux from $L_{\rm X}$ using

$$F_{\rm X} = \frac{L_{\rm X}}{4\pi D_L^2 K(z)}, \tag{11}$$

where D_L is the luminosity distance and $K(z)$ is a K-correction of the form

$$K(z) = (1+z)^{\Gamma-2}, \tag{12}$$

where Γ is the slope of the X-ray spectrum. For the K-correction we assumed a photon index of $\Gamma = 1.4$, which corresponds to the slope of the cosmic X-ray background, and therefore should represent the full population of both obscured and unobscured objects. By comparing the $N(> F_{\rm X})$ from our catalogue in Fig.12 to different data from the literature, we see overall a good agreement over a large range of flux until $F_{\rm X} \gtrapprox 2 \times 10^{-17}$ erg s^{-1} cm^{-2}; that is, two orders of magnitudes fainter than the sample used to calibrate our methodology.

4.3. Narrow emission lines

Diagnostic diagrams are a frequently utilised tool for identifying AGN. By comparing the ratios of various emission lines, we can gain insight into whether star formation, AGN or a composite of both processes dominate in the spectra of a given galaxy. In this section we employ various optical nebular line diagnostics in order to validate the AGN on our catalogue.

One of the diagrams used for this purpose is the BPT diagram (Baldwin et al. 1981), which plots the ratio of the optical lines [O III]$_{\lambda 5007}$/Hβ against [N II]$_{\lambda 6584}$/Hα. In this diagram, the Kauffmann et al. (2003) and Kewley et al. (2001) lines are commonly used to separate between objects dominated by AGN emission (that fall on the top-right of the diagram), those dominated by star-formation (in the bottom-left) and those that have composite spectra (in the central region).

We show in Fig. 13 an example of such a diagram for the objects of our catalogue, selected with $z \leq 0.8$, since, as was shown by Kewley et al. (2013b), this would be the maximum redshift up to which it is safe to use the BPT diagram to separate SF galaxies from AGN. For clarity of visualisation, we show only galaxies and AGN selected with observed magnitude $m_H < 24$. The AGN in this figure are colour-coded according to the ratio of the emission line flux of [O III] from the AGN component with respect to the total (AGN + host galaxy). We remind the reader that in our mock, the emission line flux in type 2 AGN is directly proportional to $L_{\rm X}$, and therefore this ratio is also correlated to $L_{\rm X}$. We also show as comparison the observed line ratios from local galaxies from the SDSS catalogue. We see that, in general, the simulated AGN from our mock fall in the regions that correspond to AGN-dominated or composite objects. While there are some that fall in the region of SF objects, the great majority of them are AGN with dominant [O III] emission from the host galaxy (and most of them have low X-ray luminosities, $L_{\rm X} < 10^{42}$ erg s^{-1}). In fact, different studies have pointed out that the BPT diagram is biased towards more luminous AGN, missing

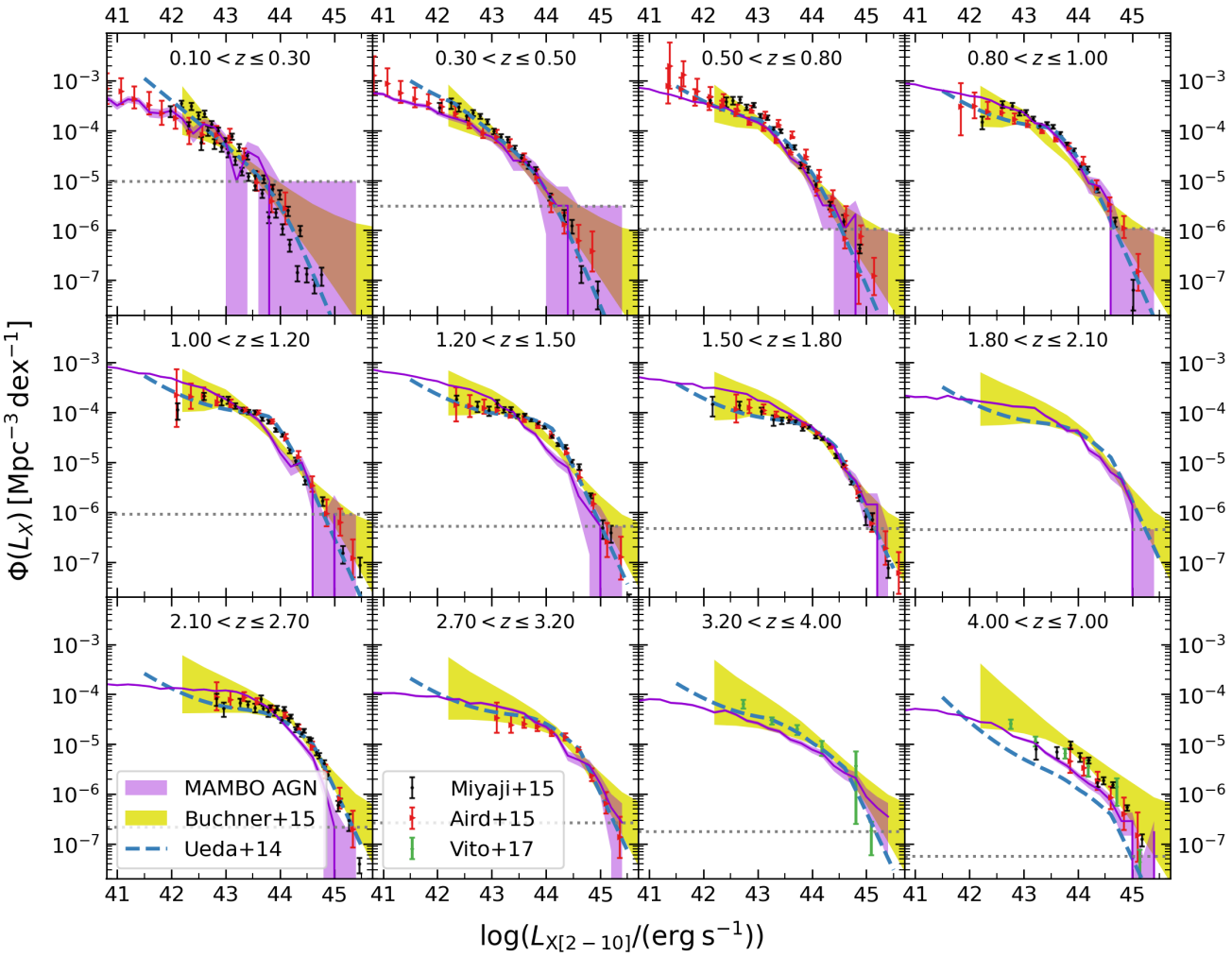

Fig. 11. Hard X-ray luminosity function of the total population of AGN of our catalogue in different redshift bins, shown by the pruple line. The shaded region represents the Poissonian uncertainty and the dotted horizontal line marks the limiting density from our lightcone (corresponding to 1 object $\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-1}$). For comparison, we show several observed XLFs: Ueda et al. (2014), Buchner et al. (2015), Miyaji et al. (2015), Aird et al. (2015) and Vito et al. (2017).

objects with low X-ray (Birchall et al. 2022) or optical luminosity (Schawinski et al. 2010).

An alternative emission line diagnostic to classify AGN is the mass–excitation (MEx) diagram (Juneau et al. 2011), which plots $[\mathrm{O\,III}]_{\lambda5007}/\mathrm{H}\beta$ against $\mathcal{M}$ and was calibrated to separated SF galaxies from AGN at $0.3 < z < 1$. Figure 14 shows the MEx diagram for the lighcone presented in this work, where both galaxies and AGN have been selected with $0.3 \leq z \leq 1.0$ and $m_{H_E} < 24$. The limit in magnitude allows for a better comparison with the work from Juneau et al. (2011), as it removes the low stellar mass tail of our catalogue. We see that all of the AGN that fall in the AGN locus of the diagram have line emission dominated by the AGN component (at least for the [O III] line). However, there are some AGN-dominated sources that are classified as MEx-SF galaxies. Similarly to the previous discussion regarding the BPT diagram, the majority of them have $L_X < 10^{42}\,\mathrm{erg\,s^{-1}}$, while the MEx diagram was validated using AGN selected above this threshold, and therefore, it is difficult to draw clear conclusions from this sub-sample.

4.4. AGN colours and the ultraviolet luminosity function

An important validation for the catalogue if we intend to reproduce the observed AGN population is the luminosity function. In this section we study the redshift evolution of the AGN UV luminosity function (UVLF) at rest-frame wavelength $\lambda = 1450$ Å, where the majority of UV rest-frame data on AGN is gathered and where type 1 AGN typically present a peak in their SED (the so-called "big blue bump").

In Fig. 15, we show the UVLF of the type 1 AGN from the lightcone presented in this work, in different redshift bins, up to $z = 6$. We have used the absolute rest-frame magnitude from the GALEX FUV filter (see Table 1) as a proxy for M_{1450}. We note that the uncertainty shown is only Poissonian, and therefore constitutes a lower boundary since it does not include other systematic effects like selection effects or completeness level, which would increase the uncertainty.

In Fig. 15, we also compare our UVLF with different works from the literature. Specifically, we show at all redshift bins the QSO UVLF from Manti et al. (2017), who parameterised the LF both as a DPL and a Schechter function using a collection of state-of-the-art measurements from z = 0.5 to z = 6.5, and also from Kulkarni et al. (2019), who used a sample of more than 80 000 colour-selected AGN from redshift $z = 0$ to 7.5 to parameterise the UVLF as DPL evolving with redshift. Both of these works use the absolute monochromatic AB magnitude at a rest-frame wavelength of 1450 Å to construct the UVLF. From $z = 3$ to 6 we show also the UVLF from Finkelstein & Bagley (2022),

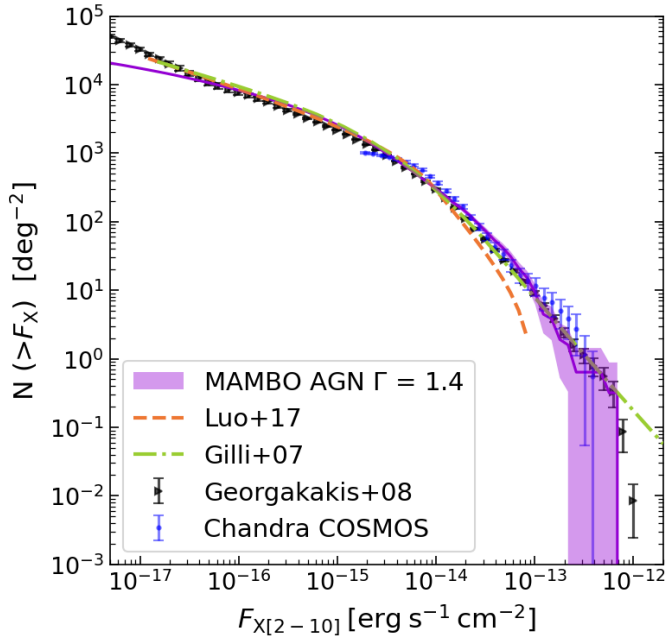

Fig. 12. Cumulative number counts of X-ray flux in the hard band. The violet shaded area corresponds to the AGN in the MAMBO catalogue, and it shows the uncertainty of our values estimated as the Poissonian error. We show as comparison the number counts reported in different works. References: Luo et al. (2017), Gilli et al. (2007), Georgakakis et al. (2008), Marchesi et al. (2016b).

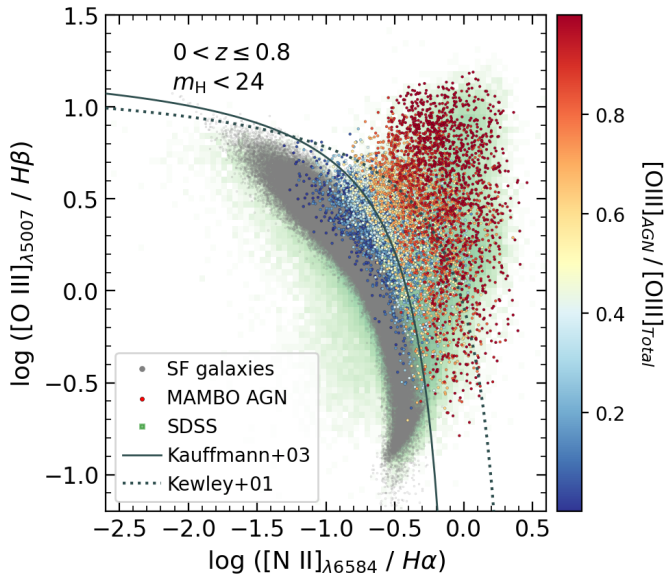

Fig. 13. BPT diagram for galaxies and AGN from the MAMBO lightcone selected with $z \leq 0.8$ and $m_{H_E} < 24$. Galaxies are shown in grey, while AGN are colour-coded from blue to red according to the ratio of the emission line flux of [O III] from the AGN component with respect to the total (AGN + host galaxy). Objects from the SDSS DR8 are shown in green in the background, and are traditionally classified as SF galaxies if the fall at the left of the Kauffmann et al. (2003), as AGN if they fall at the right of the Kewley et al. (2001) line, or as composite if they fall in between these two lines.

who studied jointly the UVLF of galaxies and QSO and parameterised each population individually with a modified DPL, in order to account for the drop in the faint end of the LF. Additionally, in the figure we show with horizontal dotted line marks the limiting density from our lightcone (corresponding

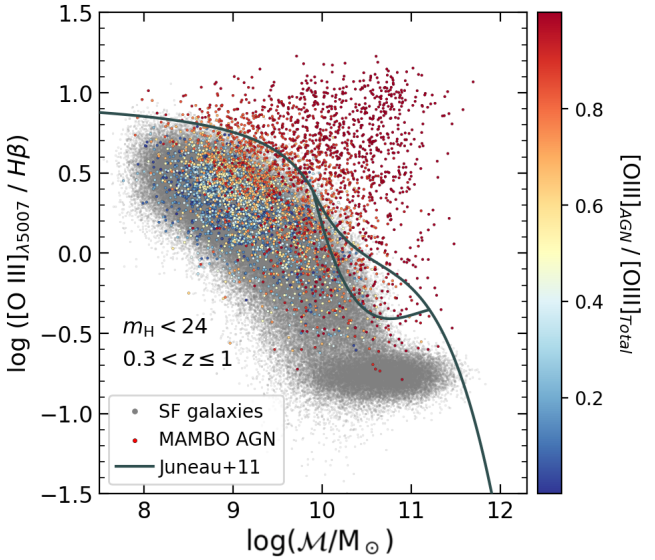

Fig. 14. Mass–excitation (MEx) diagram for galaxies and AGN from the MAMBO lightcone selected with $0.3 \leq z \leq 1.0$ and $m_{H_E} < 24$. Galaxies are shown in grey, while AGN are colour-coded from blue to red according to the ratio of the emission line flux of [O III] from the AGN component with respect to the total (AGN + host galaxy). The solid black lines show the empirical division found in Juneau et al. (2011) to separate AGN (above the line) from SF galaxies (below the line) and composite galaxies (between the two lines).

to 1 object/Mpc^3/mag) for each redshift bin. On the other hand, the vertical dotted line shows the break magnitude, M_*, at 1500 Å of the galaxy UVLF; that is, the magnitude where the galaxy contribution to the ionising background could be relevant, and indeed the galaxy number density higher than the UV/optically selected QSO one. Following Parsa et al. (2016), Ricci et al. (2017b), this magnitude is calculated as

$$M_* = (1+z)^{0.206}\left(-17.793 + z^{0.762}\right). \quad (13)$$

By comparing the UVLF constructed from our catalogue with the ones derived directly from observations, we observe a general agreement up to $z \lesssim 5$, for magnitudes brighter than the break magnitude M_*. At fainter magnitudes ($M_{UV} > M_*$) we observe a drop in our QSO UVLF, as was expected from the definition of M_* (see above). We note also that at these faint magnitudes there is a big discrepancy also among the observed QSO UVLFs. For example, from $z = 3$ to 5, the faint end of our UVLF is orders of magnitude lower than that of Manti et al. (2017) of Kulkarni et al. (2019), but agrees quite well with the only-QSO LF of Finkelstein & Bagley (2022).

On the other side, our mock produces an overprediction of the bright end of the UVLF. This can be partially due to the fact we assigned the AGN/galaxy fraction starting from X-ray selected catalogues, which tend to be more complete than UV/optical selections of AGN.

Colour–colour diagrams that use UV to mid-infrared colours can be used to select AGN from a galaxy and AGN sample. Additionally, these selections have the advantage that they can be quickly applied to very large data sets without spectroscopic information. In this section we use such diagrams to validate the colour properties of the AGN in our mock. For this, we checked different colour–colour diagrams that are known to separate AGN from galaxies. For UV/optical bands, these diagrams are able to select mainly luminous type 1 AGN, since type 2 tend

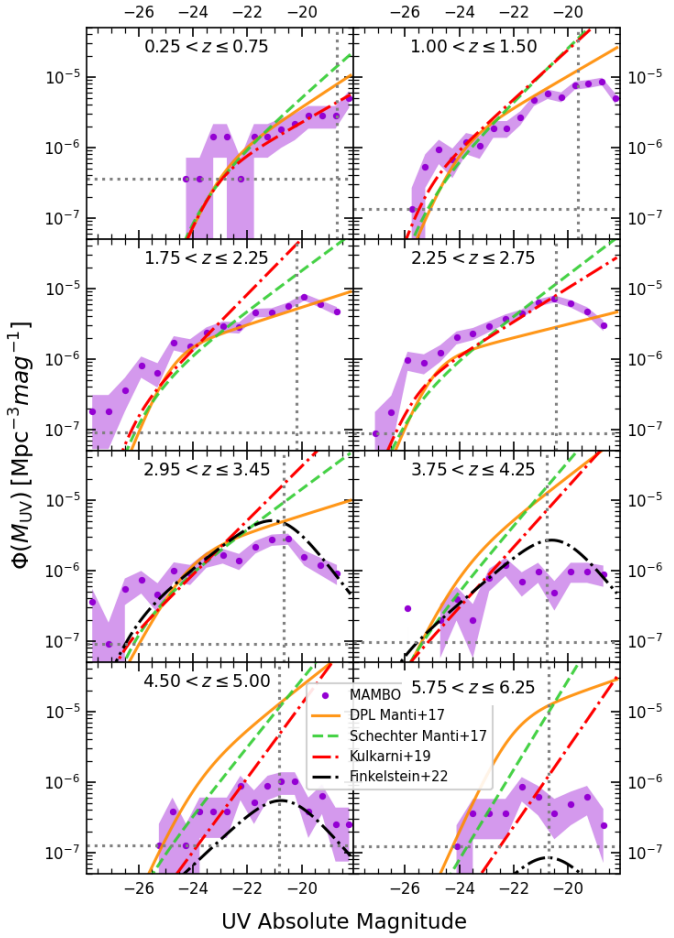

Fig. 15. UV luminosity function of type 1 AGN from our catalogue (purple points) compared with different literature LFs in different redshift bins. The purple shaded area shows the uncertainty of our values, estimated as the Poissonian error. In the case of MAMBO, the UV magnitude is computed in the GALEX FUV filter, while for the other cases, it refers to M_{1500}. The solid orange and dashed green lines show the parametric LF from Manti et al. (2017), parameterised as a DPL and a Schechter function, respectively. The dash-dotted red and black lines show the DPL parametric LFs from Kulkarni et al. (2019) and Finkelstein & Bagley (2022), respectively. The vertical dotted line shows the break magnitude at which the galaxy contribution should start dominating the UVLF (Parsa et al. 2016), while the horizontal dashed line marks the limiting density from our lightcone (corresponding to 1 object/Mpc^3/mag) for each redshift bin.

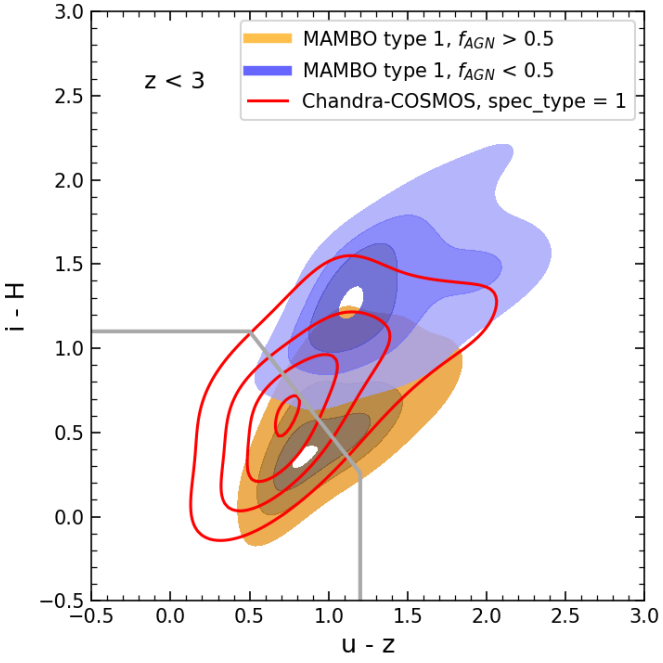

Fig. 16. Colour–colour diagram to separate galaxies from type 1 AGN as in Euclid Collaboration (2024a). Filled contours show the distribution of type 1 AGN from the lightcone here presented, separated into those dominated by the AGN component ($f_{\rm AGN} > 0.5$, orange contours) or the galaxy component ($f_{\rm AGN} < 0.5$, blue contours). The distribution of spectroscopically confirmed type 1 AGN from the Chandra-COSMOS catalogue Marchesi et al. (2016a) is shown with non-filled red contours. In all cases, the samples have been cut at $z < 3$. The contour levels represent iso-density lines, corresponding to the 50th, 75th, 90th, and 99th percentiles of the distribution. The grey line shows the best selection criteria found in Euclid Collaboration (2024a) to separate type 1 AGN from galaxies and type 2 AGN.

to be completely dominated by the host-galaxy emission at these wavelengths.

In Fig. 16, we show the $(i - H)$ vs. $(u - z)$ diagram, which was found in Euclid Collaboration (2024a) to be the best colour selection to separate type 1 AGN from galaxies using *Euclid* and *Rubin*/LSST filters. These filters are described in Table 1. The filled contours show the distribution of type 1 AGN from the lightcone here presented, separated into those dominated by the AGN component in the rest-frame FUV ($f_{\rm AGN}$[8] > 0.5) or the galaxy component ($f_{\rm AGN} < 0.5$). We also show the distribution of spectroscopically confirmed Type 1 AGN from the Chandra-COSMOS catalogue by Marchesi et al. (2016a) as a comparison. In all cases, the distributions have been cut at $z < 3$. We observe a good agreement between the Chandra-COSMOS AGN and those with $f_{\rm AGN} > 0.5$, while both distributions are clearly separated from that of AGN with $f_{\rm AGN} < 0.5$. Also, the MAMBO AGN dominated by the AGN component are in agreement with the best colour by Euclid Collaboration (2024a).

5. Uses for future surveys

This section aims to present some examples of how this catalogue can be used to make predictions that shall aid in the preparation of future large surveys (*Euclid*, *Rubin*/LSST, Moons, etc.). We focus on two examples, namely the expected number densities of galaxies and AGN in a given photometric band, and the spectroscopic selection of AGN through diagnostic diagrams. For this purpose, we choose the *Euclid* mission as a case study.

5.1. Caveats

First of all, we would like to warn the reader of a series of caveats to be taken into account when using this catalogue to perform this kind of scientific analysis. Most of them have already been discussed in the text in different sections, but we gather them here for clarity:

The DM lightcone out of which the galaxy and AGN catalogue is constructed contains DM haloes up to $z = 10$. However, the relations we used to construct the galaxy and AGN catalogue are calibrated at lower redshifts. For example, the galaxy SMF is based on observations up to $z = 7.5$, the AGN HGMF up to

[8] In this section we use $f_{\rm AGN}$ as shorthand for $f_{{\rm AGN}_{FUV}}$ (Eq. (8)).

z = 2.5, and the accretion rate distribution up to z = 4. Therefore, the extrapolations we did at higher redshifts are to be taken with precaution.

The spectra of type 2 AGN are constructed by simply adding narrow AGN lines to the continuum of the host galaxy, without adding any nuclear (non-stellar) continuum contribution. This represents a strong assumption that might not hold true for all sources. Furthermore, we separated the AGN in our catalogue into two only categories: type 1 and 2. In reality, however, we know that such binary classification does not represent the full population of AGN, and AGN are sometimes classified in intermediate classes (1.5, 1.9, etc.).

Likewise, we did not include the IR component from the torus, which usually dominates the continuum emission of type 1 and 2 AGN for wavelengths redder than 2 µm.

Throughout this paper we assumed a constant bolometric correction $k_{\rm bol}$ = 25 ($L_{\rm bol} = k_{\rm bol} L_{\rm X}$) to ensure self-consistency. However, the bolometric correction is known to correlate with $L_{\rm X}$, and can have a wide range of values covering orders of magnitude (Duras et al. 2020).

Finally, we did not include stars in our mock. Some stars, such as cataclysmic variables, have multi-wavelength features that can lead to their misclassification as QSO. Cataclysmic variables appear as point-like sources with QSO-like colours, and often exhibit X-ray emission. Therefore, our mock does not account for contamination from such stellar objects, which is instead present in observations.

We plan to keep developing our pipeline in order to tackle most of these effects.

5.2. The Euclid Surveys

The European Space Agency *Euclid* space telescope (Euclid Collaboration 2024g), successfully launched in July 2023, is equipped with two on-board instruments. The VISible instrument (VIS, Euclid Collaboration 2024c) carries a single broad-band optical filter, $I_{\rm E}$. The Near-Infrared Spectrometer and Photometer (NISP, Euclid Collaboration 2024e), on the other hand, possesses three NIR photometric filters: $Y_{\rm E}$, $J_{\rm E}$, and $H_{\rm E}$ (Table 1). The analysis performed in the following subsections is restricted to the $H_{\rm E}$ filter. Additionally, the NISP instrument is equipped with four different low resolution NIR grisms (with a spectral resolution of R = 380 for a 0.5 arcsecond diameter source): one blue grism (0.92–1.3 µm), and three identical red grisms (1.25–1.85 µm).

During its 6-year mission, *Euclid* will perform two main surveys: a Wide Survey (EWS), planned to cover an area of 14 679 square degrees on the sky, and a Deep Survey (EDS) of 53 deg^2 (Euclid Collaboration 2024g). The limiting magnitude for point sources detected with a minimum S/N of 5 is m_H = 24 in the EWS (Euclid Collaboration: Scaramella et al. 2022), and at least two magnitudes deeper in the EDS (m_H = 26). During the EWS, the sky will be observed with only one pass of the red grism, which for the Hα line translates into a redshift coverage of $0.9 \leq z \leq 1.8$ and a flux limit of $F_{\rm H\alpha} > 2 \times 10^{-16}$erg s^{-1} cm^{-2}. Instead, for the EDS the redshift coverage of Hα spans to $0.4 \leq z \leq 1.8$ with a limiting flux $F_{\rm H\alpha} > 6 \times 10^{-17}$erg s^{-1} cm^{-2}.

5.3. Number densities

One of the most important predictions that can be done with this type of catalogue is the number densities (number of objects per squared degree) of galaxies and AGN in a given magnitude bin. We show in Fig. 17 the number density in the *Euclid* H-band for both galaxies and AGN, showing with vertical lines the limiting m_H for the *Euclid* Wide and Deep surveys. We show also for comparison the number density of objects in the COSMOS catalogue. We observe a good agreement between the galaxies from our catalogue and the ones from COSMOS 2015 (Laigle et al. 2016). For the AGN sample, we used again the Chandra COSMOS Legacy Spectral Survey catalogue (Marchesi et al. 2016b). We show in Fig. 17 the m_H distribution of type 1 and 2 AGN (classified into these 2 categories by optical spectroscopy), selected with minimum X-ray flux $F_{\rm X} = 1.9 \times 10^{-15}$ erg s^{-1} cm^{-2}. After applying the same flux limit to type 1 and 2 MAMBO AGN, we observe a good agreement between the respective type 1 and 2 populations in MAMBO and COSMOS.

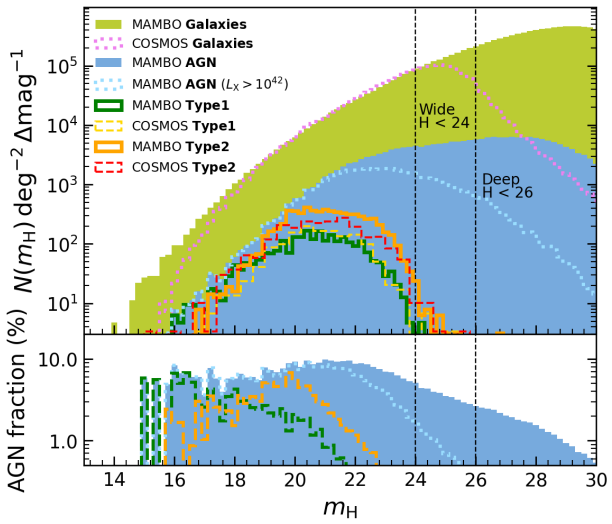

Fig. 17. Number density counts in the m_H magnitude. Galaxies and AGN from the MAMBO lightcone are represented by filled green and blue histograms, respectively. AGN selected with $L_{\rm X} > 10^{42}$ erg s^{-1} are shown with a dotted cyan line. We also show the density counts for galaxies (dotted pink line) and AGN from the COSMOS catalogue (Laigle et al. 2016; Marchesi et al. 2016b), where AGN have been selected with $F_{\rm X} > 1.9 \times 10^{-15}$ erg s^{-1} cm^{-2}, and separated into type 1 and 2 based on optical spectroscopic criteria (dashed yellow and red lines). The solid green and orange lines show the number density counts of type 1 and 2 AGN from our mock after applying the same cut in $F_{\rm X}$ as for the COSMOS sample. We show with vertical dashed black lines the limiting magnitude of the EWS and EDS. The lower panel show the fraction of MAMBO AGN as a function of m_H. The different lines represent different subpopulations of AGN, as in the upper panel.

Additionally, we show in Table 3 the expected surface densities and total (integrated) numbers of galaxies and AGN with different selections, in the EWS and the EDS. First, we performed a photometric selection of sources detected above the limiting magnitude of each survey. Furthermore, the numbers shown in Table 3 consider only AGN selected with intrinsic hard band luminosity $L_{\rm X} > 10^{42}$ erg s^{-1}, to remove possible non-AGN X-ray emitters. We compare these numbers with the ones reported in Euclid Collaboration (2024h), who performed a similar analysis to the one presented in this work, with the aim of forecasting the expected surface densities of AGN in the *Euclid* Surveys. In Tables 4 and 5 of Euclid Collaboration (2024h), the authors report the surface densities of AGN detectable in the $H_{\rm E}$ band for the *Euclid* Wide and Deep Surveys. For the EWS, the reported surface densities are 4.5×10^3 deg^{-2}, 6.8×10^2 deg^{-2} and 1.7×10^3 deg^{-2} for all AGN, type 1 and type 2, respectively.

Table 3. Surface density and total numbers of galaxies and AGN (full population, type 1 and type 2).

	Galaxies		Type 1 AGN		Type 2 AGN		All AGN
	Surface density deg^{-2}	Total number	Surface density deg^{-2}	Total number	Surface density deg^{-2}	Total number	Surface density deg^{-2}
EWS $m_H < 24$	1.4×10^5	2.1×10^9	8.3×10^2	1.2×10^7	5.5×10^3	8.0×10^7	6.3×10^3
EDS $m_H < 26$	4.0×10^5	2.0×10^7	9.4×10^2	5.0×10^4	7.4×10^3	3.9×10^5	8.4×10^3
EWS $F_{\mathrm{H}\alpha} > 2 \times 10^{-16}$ $0.9 \leq z \leq 1.8$	2.9×10^3	4.4×10^7	–	–	6.1×10^2	8.9×10^6	6.1×10^2
EDS $F_{\mathrm{H}\alpha} > 6 \times 10^{-17}$ $0.4 \leq z \leq 1.8$	3.0×10^4	1.5×10^6	–	–	2.6×10^3	1.4×10^5	2.6×10^3

Notes. Each row corresponds to a different selection, written in the table and explained in greater detail in the text. All AGN are selected with $L_X > 10^{42}$ erg s^{-1}. Total numbers assume that the *Euclid* Wide (EWS) and Deep (EDS) surveys will cover areas of 14 679 deg^2 and 53 deg^2, respectively (Euclid Collaboration 2024g). Fluxes are given in erg s^{-1} cm^{-2}.

The corresponding numbers for the EDS are 2.4×10^3 deg^{-2}, 8.8×10^2 deg^{-2} and 3.5×10^3 deg^{-2}. We observe in general a good agreement between these numbers and the ones reported in Table 3, especially for the type 1 AGN, while the numbers differ more for the type 2.

In Table 3 we also show the surface densities of sources with Hα emission line flux above the detection limit of each survey (see Sect. 5.2), and only considering sources within the redshift window at which this line will be observed at each survey. For this part, we only considered narrow-line emitters; that is, galaxies and type 2 AGN, since the values of the emission line fluxes for type 1 AGN are not included in the current version of the catalogue.

5.4. Diagnostic diagrams

In Sect. 4, we used two emission line diagnostic diagrams, namely the BPT and the MEx, as validation tools for our lightcone, by studying them at the redshift ranges at which these diagrams have been calibrated. In this section, we studied these same diagrams applying the redshift, magnitude and flux limits corresponding to the *Euclid* Wide and Deep surveys.

Two aspects should be noted in this regard: first, both these diagrams have been calibrated at low redshift ($z \lesssim 1$), while at higher redshifts the physical properties (metallicity, density, ionising radiation, etc.) of the regions where the lines are emitted are expected to change with respect to local conditions. For example, as studied by different works, the position of AGN in the BPT strongly depends on the gas metallicity, and below $Z \sim 0.5\,Z_\odot$, AGN start to populate the SF region side of the diagram (Groves et al. 2006; Kewley et al. 2013a; Hirschmann et al. 2019). Therefore, these diagrams must be taken more cautiously when used at these redshifts with real data. Secondly, a precise study of the number of AGN that *Euclid* could select with these diagrams (assuming they work at high z) would involve using *Euclid*-like spectra with realistic noise and resolution, and official *Euclid* pipelines for the extraction of the line fluxes, which is beyond the scope of this study. Besides, given the spectral resolution of *Euclid*, the [N II] and Hα lines are likely to be blended in real *Euclid* spectra (Euclid Collaboration 2024f).

We show in Fig. 18 the BPT and MEx diagrams for the Wide and Deep *Euclid* surveys, applying in each case the corresponding selection in emission line flux and magnitude limit, and redshift range where all relevant lines will be observed. The number density of objects for each case is also given in the figure.

6. Summary

We developed an empirical workflow to generate mock catalogues of galaxies and AGN starting from a DM-only simulation. Following Girelli (2021), we populated the DM haloes with galaxies by means of a stellar-to-halo mass relation, developed using a subhalo abundance matching technique on observed SMFs. Galaxies were also separated into quiescent or SF ones, following the relative ratio of the blue and red populations in observed SMFs.

In this paper, we further populated galaxies with AGN following observed host galaxy AGN mass functions at different redshifts and AGN accretion rate distribution functions, which were derived starting from X-ray samples of AGN at $z < 4$ (Bongiorno et al. 2016; Aird et al. 2018). Following Merloni et al. (2014), we separated AGN into optically unobscured (type 1) or obscured (type 2) ones and assigned a proper SED to each of them. For this, for type 1 AGN we used the parametric SED model by Temple et al. (2021), which accounts for the continuum emission of type 1 AGN, as well as their broad and narrow emission lines. For type 2, instead, we added narrow lines generated using photoionisation models (Feltre et al. 2016) to the host galaxy stellar continuum.

We tested this workflow by applying it to a 3.14 deg^2 DM Millennium lightcone up to $z = 10$. The result is a mock catalogue of galaxies and AGN with realistic physical properties and observables (such as broadband rest-frame and observed magnitudes and spectral features), complete at least up to a magnitude of $m_H \sim 28$ and down to a stellar mass of $\mathcal{M} \sim 10^{7.5}\,M_\odot$. We obtained good agreement between our mock data and state-of-the-art observations such as published L_X luminosity functions (e.g. Buchner et al. 2015; Aird et al. 2015), number counts in different NIR to optical bands (in this paper we showed results only on the H band for simplicity), colour–colour diagrams (using u, z, i, and H bands), and emission line diagnostic diagrams (BPT and MEx).

Finally, we demonstrated how this catalogue can be used to make forecasts for future large surveys, using *Euclid* as an example. We computed the expected surface densities of type 1 and 2 AGN detectable with a given *Euclid* broad filter. We show the results for the H_E band, forecasting that *Euclid* will observe about 8.3×10^2 and 5.5×10^3 deg^{-2} type 1 and 2 AGN, respectively, selected with $m_H < 24$, and 6.1×10^2 (2.6×10^3) deg^{-2} type 2 AGN with narrow-line H$_\alpha$ emission with flux $F_{\mathrm{H}\alpha} > 2 \times 10^{-16}$ (6×10^{-17}) erg s^{-1} cm^{-2} in the EWS (EDS), finding good agreement with other published forecasts. We also give

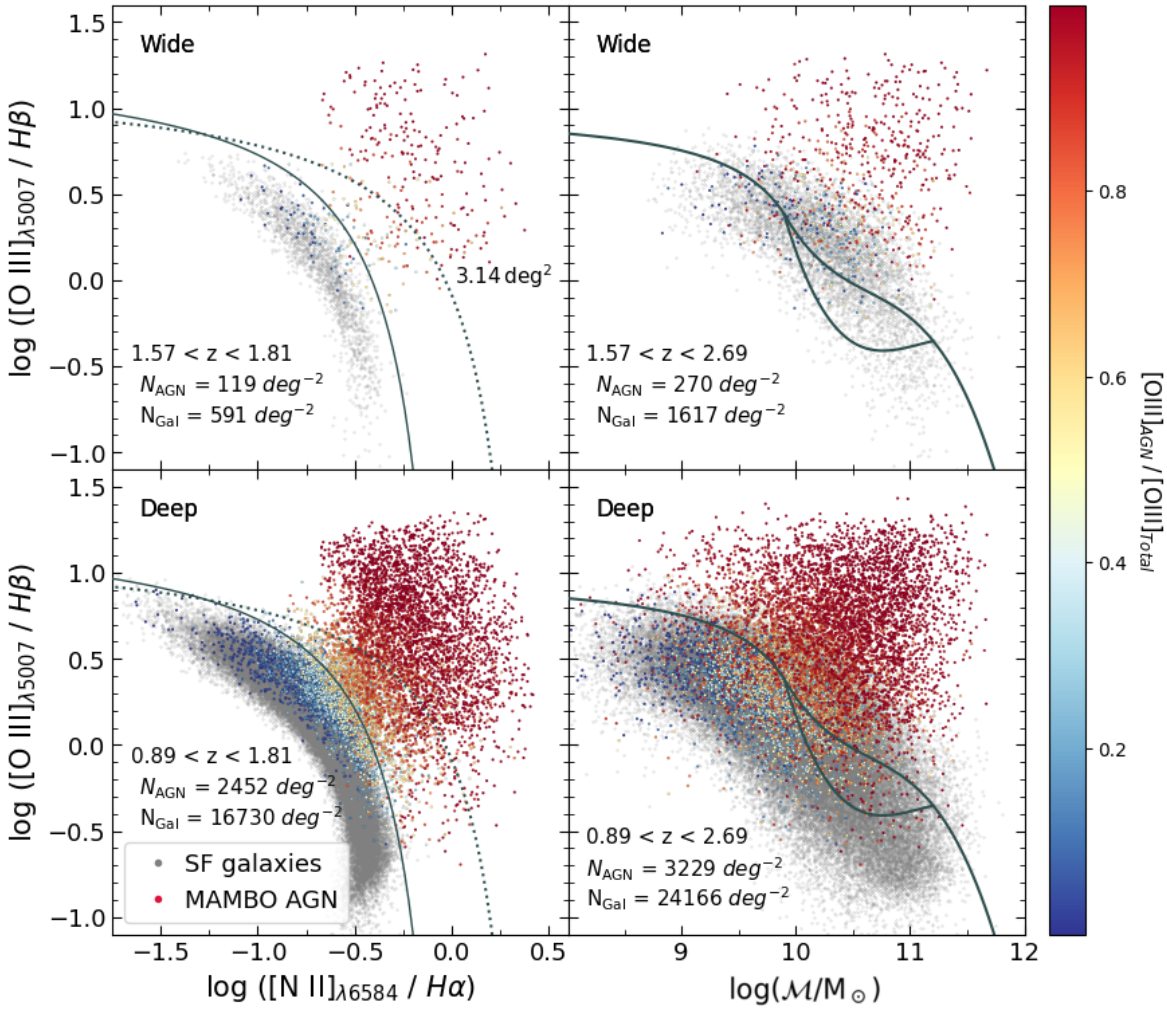

Fig. 18. Predictions from our lightcone for the BPT (left panels) and MEx diagrams (right panels) as observed by the *Euclid* Wide and Deep surveys. For each panel we have plotted the galaxies (grey dots) and AGN (colour-coded from blue to red) corresponding to the specific redshift range, magnitude limit (in H-band) and Hα line flux limit of each survey. The surface density of galaxies and AGN in each diagram are given in the figure. See captions of Figs. 13 and 14 for further details.

examples of the *Euclid* view of narrow-line diagnostic diagrams, which are used to separate local AGN from SF galaxies.

The full workflow is designed to be as computationally efficient as possible so that it can be run on a personal computer. In Appendix B, we give more details on the execution time of the main steps of the method.

We plan to update this workflow in the near future in order to tackle the open issues described in Sect. 5.1. For example, we plan to revise the AGN fraction and the assignment of the X-ray luminosity using more up-to-date accretion rate distributions (e.g. Zou et al. 2024). Also, we are working on the inclusion of an obscured AGN continuum for type 2 sources, as well as the emission from a dusty torus, which dominates the IR emission of type 1 and 2 AGN. Additionally, we plan to investigate the clustering properties of the AGN in our mock catalogues and compare them with observational data as an additional validation step. Although similar studies have been conducted for MAMBO galaxies (Girelli 2021), we leave the corresponding analysis for AGN to future work. Finally, we plan to apply this method starting from a bigger DM lightcone.

Data availability

The catalogue used for the analysis performed in this paper, together with the necessary code to produce it, are available through the GitHub repository https://github.com/xalolo/MAMBO/tree/main.

Acknowledgements. We thank the referee for the careful reading and helpful suggestions that helped to improve the manuscript. We acknowledge the helpful conversations and input from Johan Comparat, Andrea Merloni, Matthew Selwood, James Aird, Matthew J. Temple, Roberto Gilli, Marco Mignoli. XLL acknowledges the "IAC International Scholarships Program". LP and MB acknowledge ASI and Italian Ministry grant "Premiale MITIC 2017". MB acknowledges the support from INAF Minigrant 2023. AF acknowledges the support from the project "VLT-MOONS" CRAM 1.05.03.07, INAF Large Grant 2022 "The metal circle: a new sharp view of the baryon cycle up to Cosmic Dawn with the latest generation IFU facilities" and INAF Large Grant 2022 "Dual and binary SMBH in the multi-messenger era". VA acknowledges the support from the INAF Large Grant "AGN & *Euclid*: a close entanglement" 01.05.23.01.14. IEL acknowledges support from the Cassini Fellowship program at INAF-OAS and the European Union's Horizon 2020 research and innovation program under Marie Sklodowska-Curie grant agreement No. 860744 "Big Data Applications for Black Hole Evolution Studies" (BiD4BESt).

References

Aird, J., Coil, A. L., Moustakas, J., et al. 2012, ApJ, 746, 90
Aird, J., Coil, A. L., Georgakakis, A., et al. 2015, MNRAS, 451, 1892
Aird, J., Coil, A. L., & Georgakakis, A. 2018, MNRAS, 474, 1225
Alberts, S., Williams, C. C., Helton, J. M., et al. 2023, arXiv e-prints, [arXiv:2312.12207]
Allevato, V., Shankar, F., Marsden, C., et al. 2021, ApJ, 916, 34
Antonucci, R. 1993, ARA&A, 31, 473
Baldwin, J. A. 1977, ApJ, 214, 679
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Bezanson, R., van der Wel, A., Straatman, C., et al. 2018, ApJ, 868, L36
Binette, L., Zovaro, H. R. M., Villar Martín, M., et al. 2024, A&A, 684, A53
Birchall, K. L., Watson, M. G., & Aird, J. 2020, MNRAS, 492, 2268
Birchall, K. L., Watson, M. G., Aird, J., & Starling, R. L. C. 2022, MNRAS, 510, 4556
Bisigello, L., Gruppioni, C., Feltre, A., et al. 2021, A&A, 651, A52
Bisogni, S., Lusso, E., Civano, F., et al. 2021, A&A, 655, A109
Blaizot, J., Wadadekar, Y., Guiderdoni, B., et al. 2005, MNRAS, 360, 159
Bongiorno, A., Merloni, A., Brusa, M., et al. 2012, MNRAS, 427, 3103
Bongiorno, A., Schulze, A., Merloni, A., et al. 2016, A&A, 588, A78
Brandt, W. N., & Alexander, D. M. 2015, A&A Rev., 23, 1
Brandt, W. N., & Hasinger, G. 2005, ARA&A, 43, 827
Brusa, M., Civano, F., Comastri, A., et al. 2010, ApJ, 716, 348
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Buchner, J., Georgakakis, A., Nandra, K., et al. 2015, ApJ, 802, 89
Cabral, A., Abreu, M., Coelho, J. M. P., et al. 2020, SPIE Conf. Ser., 11447, 114478D
Cappelluti, N., Brusa, M., Hasinger, G., et al. 2009, A&A, 497, 635
Carnall, A. C., McLeod, D. J., McLure, R. J., et al. 2023, MNRAS, 520, 3974
Carnall, A. C., Cullen, F., McLure, R. J., et al. 2024, MNRAS, 534, 325
Chabrier, G. 2003, PASP, 115, 763
Chary, R., & Elbaz, D. 2001, ApJ, 556, 562
Comparat, J., Eckert, D., Finoguenov, A., et al. 2020, Open J. Astrophys., 3, 13
Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, ApJ, 647, 201
Davidzon, I., Ilbert, O., Laigle, C., et al. 2017, A&A, 605, A70
Drory, N., Bundy, K., Leauthaud, A., et al. 2009, ApJ, 707, 1595
Duras, F., Bongiorno, A., Ricci, F., et al. 2020, A&A, 636, A73
Elbaz, D., Dickinson, M., Hwang, H. S., et al. 2011, A&A, 533, A119
Euclid Collaboration (Scaramella, R., et al.) 2022, A&A, 662, A112
Euclid Collaboration (Bisigello, L., et al.) 2024a, A&A, in press, https://doi.org/10.1051/0004-6361/202450446
Euclid Collaboration (Castander, F. J., et al.) 2024b, A&A, submitted [arXiv:2405.13495]
Euclid Collaboration (Cropper, M., et al.) 2024c, A&A, submitted [arXiv:2405.13492]
Euclid Collaboration (Enia, A., et al.) 2024d, A&A, in press, https://doi.org/10.1051/0004-6361/202451425
Euclid Collaboration (Jahnke, K., et al.) 2024e, A&A, in press, https://doi.org/10.1051/0004-6361/202450786
Euclid Collaboration (Lusso, E., et al.) 2024f, A&A, 685, A108
Euclid Collaboration (Mellier, Y., et al.) 2024g, A&A, in press, https://doi.org/10.1051/0004-6361/202450810
Euclid Collaboration (Selwood, M., et al.) 2024h, A&A, submitted [arXiv:2405.18126]
Feltre, A., Charlot, S., & Gutkin, J. 2016, MNRAS, 456, 3354
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, Rev. Mexicana Astron. Astrofis., 53, 385
Finkelstein, S. L., & Bagley, M. B. 2022, ApJ, 938, 25
Gebhardt, K., Bender, R., Bower, G., et al. 2000, ApJ, 539, L13
Georgakakis, A., Nandra, K., Laird, E. S., Aird, J., & Trichas, M. 2008, MNRAS, 388, 1205
Georgakakis, A., Pérez-González, P. G., Fanidakis, N., et al. 2014, MNRAS, 440, 339
Georgakakis, A., Comparat, J., Merloni, A., et al. 2019, MNRAS, 487, 275
Georgakakis, A., Ruiz, A., & LaMassa, S. M. 2020, MNRAS, 499, 710
Gilli, R., Comastri, A., & Hasinger, G. 2007, A&A, 463, 79
Girelli, G. 2021, PhD thesis, Alma Mater Studiorum Università di Bologna, Italy, http://amsdottorato.unibo.it/9820/
Girelli, G., Bolzonella, M., & Cimatti, A. 2019, A&A, 632, A80
Girelli, G., Pozzetti, L., Bolzonella, M., et al. 2020, A&A, 634, A135
Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35
Groves, B. A., Heckman, T. M., & Kauffmann, G. 2006, MNRAS, 371, 1559
Gu, Y., Yang, X., Han, J., et al. 2024, MNRAS, 529, 4015
Gutkin, J., Charlot, S., & Bruzual, G. 2016, MNRAS, 462, 1757
Hadzhiyska, B., Yuan, S., Blake, C., et al. 2023, MNRAS, 525, 4367
Hasinger, G., Cappelluti, N., Brunner, H., et al. 2007, ApJS, 172, 29
Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. 2015, MNRAS, 451, 2663
Hirschmann, M., Charlot, S., Feltre, A., et al. 2017, MNRAS, 472, 2468
Hirschmann, M., Charlot, S., Feltre, A., et al. 2019, MNRAS, 487, 333
Ilbert, O., Salvato, M., Le Floc'h, E., et al. 2010, ApJ, 709, 644
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55
Ishiyama, T., Prada, F., Klypin, A. A., et al. 2021, MNRAS, 506, 4210
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Juneau, S., Dickinson, M., Alexander, D. M., & Salim, S. 2011, ApJ, 736, 104
Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, MNRAS, 346, 1055
Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121
Kewley, L. J., Dopita, M. A., Leitherer, C., et al. 2013a, ApJ, 774, 100
Kewley, L. J., Maier, C., Yabe, K., et al. 2013b, ApJ, 774, L10
Kormendy, J., & Ho, L. C. 2013, ARA&A, 51, 511
Korytov, D., Hearin, A., Kovacs, E., et al. 2019, ApJS, 245, 26
Kovacs, E., Mao, Y.-Y., Aguena, M., et al. 2022, Open J. Astrophys., 5, 1
Kulkarni, G., Worseck, G., & Hennawi, J. F. 2019, MNRAS, 488, 1035
Laigle, C., McCracken, H. J., Ilbert, O., et al. 2016, ApJS, 224, 24
Latimer, L. J., Reines, A. E., Bogdan, A., & Kraft, R. 2021, ApJ, 922, L40
Looser, T. J., D'Eugenio, F., Maiolino, R., et al. 2024, Nature, 8010, 53
Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2017, ApJS, 228, 2
Lusso, E., Hennawi, J. F., Comastri, A., et al. 2013, ApJ, 777, 86
Magdis, G. E., Daddi, E., Béthermin, M., et al. 2012, ApJ, 760, 6
Manti, S., Gallerani, S., Ferrara, A., Greig, B., & Feruglio, C. 2017, MNRAS, 466, 1160
Marchesi, S., Civano, F., Elvis, M., et al. 2016a, ApJ, 817, 34
Marchesi, S., Lanzuisi, G., Civano, F., et al. 2016b, ApJ, 830, 100
Marconi, A., & Hunt, L. K. 2003, ApJ, 589, L21
McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156
Menzel, M. L., Merloni, A., Georgakakis, A., et al. 2016, MNRAS, 457, 110
Merlin, E., Fontana, A., Castellano, M., et al. 2017, MNRAS, 473, 2098
Merloni, A., Bongiorno, A., Brusa, M., et al. 2014, MNRAS, 437, 3550
Merloni, A., Lamer, G., Liu, T., et al. 2024, A&A, 682, A34
Merritt, D., & Ferrarese, L. 2001, ApJ, 547, 140
Mezcua, M., & Domínguez Sánchez, H. 2024, MNRAS, 528, 5252
Miyaji, T., Hasinger, G., Salvato, M., et al. 2015, ApJ, 804, 104
Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, ApJ, 710, 903
Mushotzky, R. F., Done, C., & Pounds, K. A. 1993, ARA&A, 31, 717
Osorio-Clavijo, N., Gonzalez-Martín, O., Sánchez, S. F., Guainazzi, M., & Cruz-González, I. 2023, MNRAS, 522, 5788
Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of gaseous nebulae and active galactic nuclei (Sausalito, CA: University Science Books)
Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, A&A Rev., 25, 2
Parsa, S., Dunlop, J. S., McLure, R. J., & Mortlock, A. 2016, MNRAS, 456, 3194
Peluso, G., Radovich, M., Moretti, A., et al. 2023, ApJ, 958, 147
Peng, Y.-j., Lilly, S. J., Kovač, K., et al. 2010, ApJ, 721, 193
Pouliasis, E., Ruiz, A., Georgantopoulos, I., et al. 2024, A&A, 685, A97
Ricci, C., Trakhtenbrot, B., Koss, M. J., et al. 2017a, Nature, 549, 488
Ricci, F., Marchesi, S., Shankar, F., La Franca, F., & Civano, F. 2017b, MNRAS, 465, 1915
Richstone, D., Ajhar, E. A., Bender, R., et al. 1998, Nature, 385, A14
Rodrigo, C., Solano, E., & Bayo, A. 2012, SVO Filter Profile Service Version 1.0, IVOA Working Draft 15 October 2012
Schawinski, K., Urry, C. M., Virani, S., et al. 2010, ApJ, 711, 284
Schechter, P. 1976, ApJ, 203, 297
Schreiber, C., Elbaz, D., Pannella, M., et al. 2017, A&A, 602, A96
Spinoglio, L., & Fernández-Ontiveros, J. A. 2021, in Nuclear Activity in Galaxies Across Cosmic Time, 356, eds. M. Pović, P. Marziani, J. Masegosa, et al., 356, 29
Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629
Temple, M. J., Hewett, P. C., & Banerji, M. 2021, MNRAS, 508, 737
Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, ApJ, 786, 104
Urry, C. M., & Padovani, P. 1995, PASP, 107, 803
Vito, F., Brandt, W. N., Yang, G., et al. 2017, MNRAS, 473, 2378
Weaver, J. R., Davidzon, I., Toft, S., et al. 2023, A&A, 677, A184
Wechsler, R. H., & Tinker, J. L. 2018, ARA&A, 56, 435
Wechsler, R. H., DeRose, J., Busha, M. T., et al. 2022, ApJ, 931, 145
Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879
Williams, D. R. A., Pahari, M., Baldi, R. D., et al. 2022, MNRAS, 510, 4909
Xie, L., De Lucia, G., Fontanot, F., et al. 2024, ApJ, 966, L2
York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579
Zou, F., Yu, Z., Brandt, W. N., et al. 2024, ApJ, 964, 183

Appendix A: Rejected methodologies

In this appendix we show different tests that we did before arriving at the final workflow presented in this paper and that motivates some of the choices presented above.

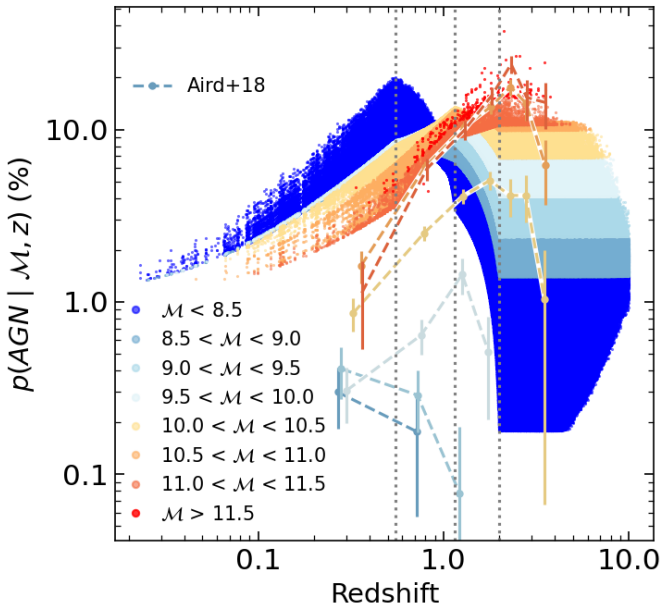

Fig. A.1. Probability of a galaxy to be an AGN as in Fig. 4 but using the original Schechter fit from B16 to assign the AGN fraction at $z < 1.15$. The biggest difference with respect to the distribution shown in Fig. 4 is in the high density of low-mass AGN at low redshift.

In Sect. 3.2 we noted that, when using the HGMF from Bongiorno et al. (2016), we modified the slope of the Schechter fit at the two lowest redshift bins. This decision was motivated because the original Schechter fit from B16, when extrapolated to $\mathcal{M} < 10^{9.5}$, predicts a large fraction of low-mass AGN. This can be seen in Fig. A.1

In principle, one could avoid doing the steps presented in Sects. 3.2 and 3.3 and instead derive the AGN fraction and the X-ray distribution in one step, starting from an observed accretion rate distribution. For this, we used the $p\,(\log \lambda_{\rm sBHAR} \mid \mathcal{M}, z)$ distributions from A18 at all $\mathcal{M}$ and z and statistically sample random values from those distributions in order to assign $\lambda_{\rm sBHAR}$ to every galaxy in our catalogue, and later $L_{\rm X}$ using Eq. (1). After this, AGN can be selected as objects above a given threshold in either $L_{\rm X}$ or $\lambda_{\rm sBHAR}$.

We show in Fig. A.2 the X-ray luminosity function resulting from this methodology, both for AGN selected with $\lambda_{\rm sBHAR} > 0.01$ (for consistency with the definition adopted in A18 and in this work), and also for all the X-ray emitters (without any constrain in $\lambda_{\rm sBHAR}$). It is visible from this figure that in both cases, the AGN population coming from this methodology is underestimated when comparing it to the observed ones, especially for $z \lesssim 1$ and $z \gtrsim 3$.

Since the $p\,(\log \lambda_{\rm sBHAR} \mid \mathcal{M}, z)$ distributions from A18 are defined up to $z \leq 4$, we also explored the possibility of using them to infer $L_{\rm X}$ at all z, starting from the AGN fraction inferred from B16 (as in Sect. 3.2). As it can be seen in Fig. A.3, the results from this methodology tend to overestimate the XLF with respect to observed ones, especially in the range $0.8 \lesssim z \lesssim 1.8$, and therefore we decided to use the approach explained in Sect. 3.3.

Appendix B: Computational time

The most computationally expensive step of the workflow described in this paper is running the modified version of the C++ code EGG. In this step, physical properties and observables are assigned to every galaxy and AGN from the catalogue. The execution time scales mainly with two parameters; namely, the number of sources and the number of filters at which one wishes to compute the rest-frame and observed magnitudes. Running EGG on a 2011 iMac with 12 GB of RAM memory and a 2.8 GHz Quad-Core Intel Core i7 processor, with 10^6 sources and 10 filters both for rest-frame and observed magnitudes took a total time of 36 min 51 s; therefore, around 2.2 ms per source. In all our tests, the running time scales almost linearly with the number of sources.

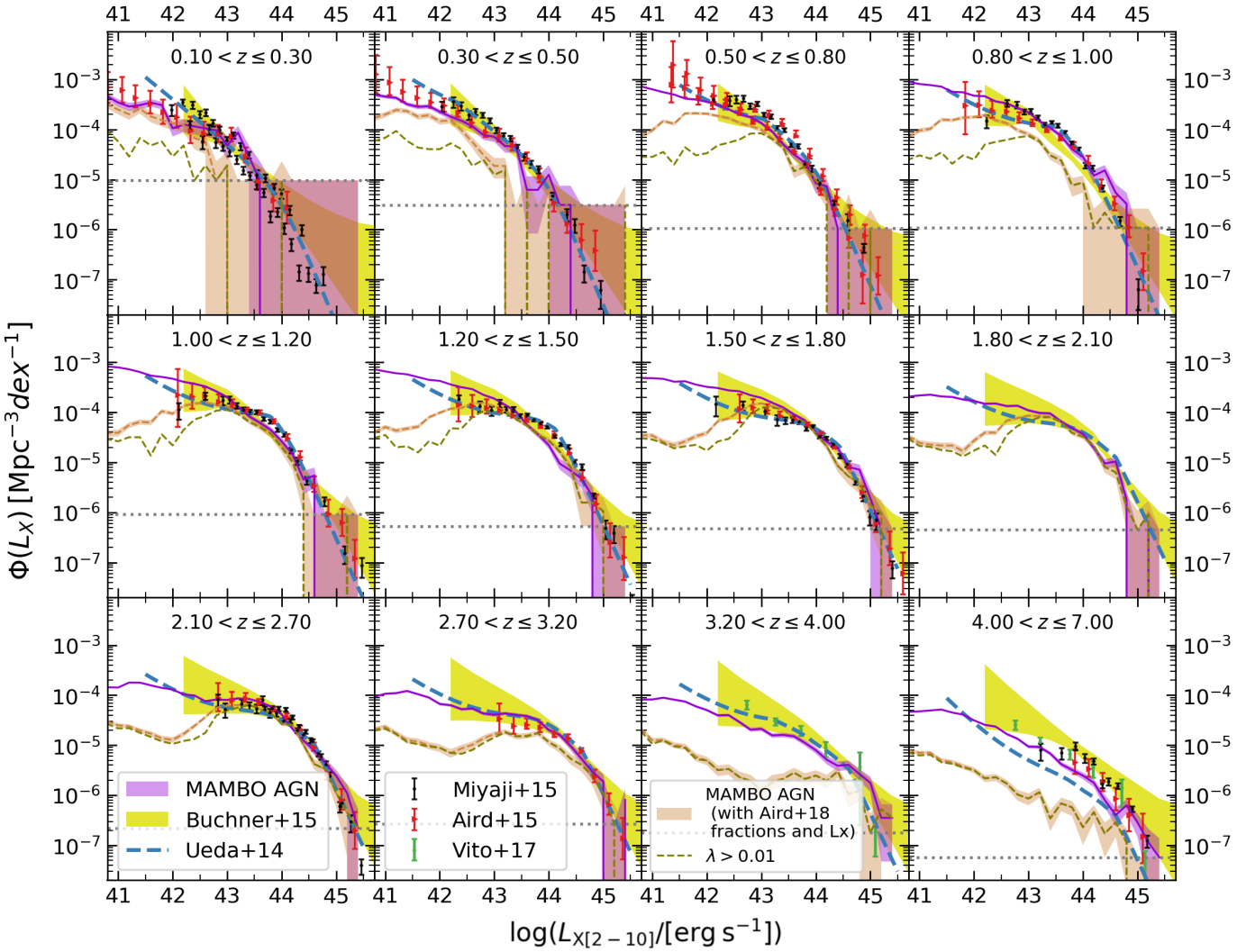

Fig. A.2. Hard X-ray luminosity function, using the $p(\log \lambda_{\rm sBHAR} \mid \mathcal{M}, z)$ distributions from A18 to infer the AGN fraction and $L_{\rm X}$ distribution at all $\mathcal{M}$ and z. The dashed brown line corresponds to all the X-ray emitters, while the dashed green line corresponds to the objects selected with $\lambda_{\rm sBHAR} > 0.01$. For comparison, the solid purple line shows the XLF from our lightcone using the methodology adopted in this paper, as in Fig.11. See the caption of Fig. 11 for further details and references to the observed XLFs.

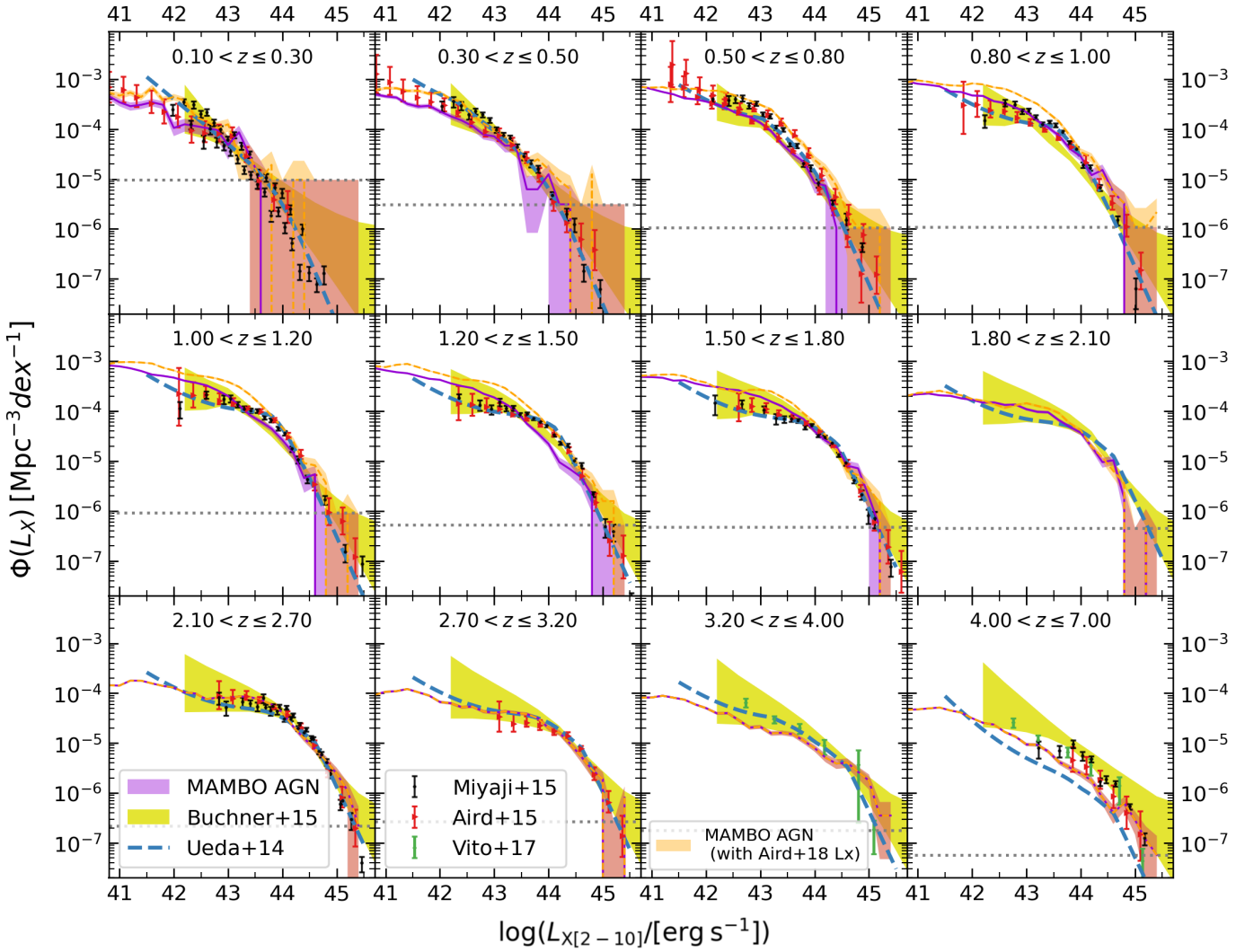

Fig. A.3. Hard X-ray luminosity function, using the AGN fraction inferred from B16 (as in Sect. 3.2), but the $p\left(\log \lambda_{\rm sBHAR} \mid \mathcal{M}, z\right)$ distributions from A18 to infer $L_{\rm X}$ at all z (dashed orange line). For comparison, the solid purple line shows the XLF from our lightcone using the methodology adopted in this paper, as in Fig.11. See the caption of Fig. 11 for further details and references to the observed XLFs.