



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN ASTROFISICA

CICLO 37

Settore Concorsuale: 02/C1 - ASTRONOMIA, ASTROFISICA, FISICA DELLA TERRA E DEI PIANETI

Settore Scientifico Disciplinare: FIS/05 - ASTRONOMIA E ASTROFISICA

A Comprehensive Galaxy and AGN Mock Catalogue for the Next Generation of Surveys

Presentata da: **Xavier López López**

Coordinatore Dottorato
Andrea Miglio

Supervisore
Micol Bolzonella

Co-supervisore
Lucia Pozzetti

Pero el negro secreto de la noche
y el secreto del agua
¿son misterios tan sólo para el ojo
de la conciencia humana?
¿La niebla del misterio no estremece
al árbol, al insecto y la montaña?
¿El terror de las sombras no lo sienten
las piedras y las plantas?
¿Es sonido tan sólo esta voz mía?
¿Y el casto manantial no dice nada?

Federico García Lorca, 1919

Abstract

Current and future large-sky surveys will produce unprecedented amounts of photometric and spectroscopic data for billions of sources, opening a new era for the development of extragalactic astronomy. To fully exploit the scientific potential of these surveys, realistic simulations of astrophysical sources such as galaxies and Active Galactic Nuclei (AGN) have become essential and invaluable tools. These simulations are needed at all stages of the survey, from the design of the telescopes and the observation plans to the interpretation of the final results.

In this context, we introduce a new, flexible, and efficient computational model designed to simulate mock catalogues of galaxies and AGN. This model aims to accurately replicate the statistical properties and observational features of these astrophysical sources, providing a robust tool for testing survey strategies and refining data analysis pipelines and techniques.

To achieve these requirements, we followed a completely empirical approach: the main constraints that shape the physical and observed properties of the sources within the model come from observed scaling relations. Specifically, we started with simulated dark matter (DM) haloes from an N-body simulation, to preserve the link with the cosmic web, and we populated them with galaxies and AGN using abundance matching techniques. To achieve this we used several observational inputs, such as 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.).

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 $z \sim 7$ and in the ultraviolet up to $z \sim 5$, optical/near-infrared colour-colour diagrams, and narrow emission line diagnostic diagrams.

Finally, we demonstrate several applications of this catalogue, mainly using *Euclid* as a case example. We computed several forecasts of *Euclid* observations, and we tested a pipeline for retrieving photometric redshifts on simulated AGN with *Euclid*-like photometry and a spectroscopic analysis pipeline on *Euclid*-like AGN spectral energy distributions. Finally, we show an application of this model for the preparation of an X-ray observational campaign.

Contents

1	Introduction	1
1.1	Cosmological context: DM to galaxies	2
1.2	Active Galactic Nuclei	5
1.2.1	Accretion onto SMBH	6
1.2.2	The AGN zoo and unification models	12
1.2.3	Identification of AGN	17
1.2.4	Co-evolution of AGN and galaxies	23
1.3	Next generation of surveys	24
1.3.1	The Euclid Surveys	24
1.4	Simulations of galaxies and AGN	29
1.4.1	Empirical models	35
1.5	Aim of this thesis	38
2	Building the catalogue	40
2.1	Mock Galaxy catalogue: MAMBO	41
2.2	Building the empirical AGN catalogue	45
2.2.1	Fraction of AGN	45
2.2.2	X-ray luminosity	50
2.2.3	Obscuration model	52
2.2.4	AGN emitted spectra	57
3	Validating the catalogue	75
3.1	AGN fraction	76
3.2	X-ray luminosity	77
3.3	Narrow emission lines	79
3.4	AGN colours and UVLF	83
3.5	Clustering	87

4 Applications of MAMBO	93
4.1 General forecasts	94
4.1.1 Number densities	94
4.1.2 Diagnostic diagrams	96
4.2 Euclid-like AGN spectra	99
4.3 Flagship 2 + MAMBO AGN	119
4.4 AGN photo-z in Euclid	120
4.5 XMM observations of a EDF: Fornax	137
5 Summary	145
A Rejected methodologies	157
B Computational time	160

Acronyms

- **AGN** : Active Galactic Nuclei
- **AM** : Abundance Matching
- **BH** : Black Hole
- **CANDELS** : Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey
- **CDFS** : Chandra Deep Field South
- **COSMOS** : Cosmic Evolution Survey
- **Dec** : Declination
- **DM** : Dark Matter
- **EDS** : Euclid Deep Survey
- **EDFF** : Euclid Deep Field Fornax
- **EW** : Equivalent Width
- **EWS** : Euclid Wide Survey
- **FUV** : Far Ultraviolet
- **FWHM** : Full Width at Half Maximum
- **IMF** : Initial Mass Function
- **IR** : Infrared
- **ISM** : Inter-Stellar Medium
- **JWST** : James Webb Space Telescope
- **LSST** : Vera Rubin Large Synoptic Survey Telescope
- **MOONS** : Multi-Object Optical and Near-Infrared Spectrograph
- **NIR** : Near Infrared

- **PSF** : Point Spread Function
- **RA** : Right Ascension
- **SAM** : Semi-Analytic Model
- **SARDF** : Specific Accretion Rate Distribution Function
- **SDSS** : Sloan Digital Sky Survey
- **SED** : Spectral Energy Distribution
- **SHAM** : Sub-Halo Abundance Matching
- **SHMR** : Stellar-to-Halo Mass Relation
- **SMBH** : Super-Massive Black Hole
- **SMF** : Stellar Mass Function
- **SNR**: Signal-to-noise ratio
- **UV** : Ultraviolet
- **VIPERS** : VIMOS Public Extragalactic Survey
- **VLT** : Very Large Telescope
- **XMM-Newton** : X-ray Multi-Mirror Mission

Chapter 1: Introduction

The main aim of this PhD thesis is to develop a new methodology to efficiently simulate empirical catalogues of Active Galactic Nuclei (AGN). During this first Chapter, we will explain why this is a relevant question to address, and what are the key ingredients involved in tackling it. To lay the groundwork, we will address the following questions: What is an AGN? What defines an empirical simulation? And, most importantly, why is it relevant to simulate AGN in this type of simulations?

Before we dive into these topics, it is essential to understand the cosmological principles that govern not only the large-scale structure of the universe, but also the fate of matter and energy within it. In Sec. 1.1 we will briefly introduce this cosmological context, and explain how the physical properties of galaxies and AGN are influenced by it.

Afterwards, in Sec. 1.2 we will provide a comprehensive overview of AGN. We will start by introducing the main physical mechanism fuelling their powerful emission, i.e. accretion onto a supermassive black hole (SMBH). Then we will explore the wide variety of observational features across the full electromagnetic spectrum that AGN exhibit, and how we can use this to select AGN from observations. Finally, we will briefly discuss the interplay between AGN and their host galaxies.

In Sec. 1.3 we will introduce some of the current and upcoming large-sky photometric and spectroscopic surveys for the next decade. These surveys present promising opportunities to tackle some of the key open questions in AGN and galaxy formation and evolution.

Finally, in Sec. 1.4 we will present why simulations are an indispensable tool in modern astrophysical research, with a special focus on the type of simulation relevant for this work: empirical models. We will see how and why large-sky surveys rely on this type of simulations for their design and for the full exploitation of their scientific results.

1.1 Cosmological context: DM to galaxies

Studying the processes that drive AGN formation and evolution requires simultaneously studying those that shape their host galaxies. At the same time, galaxy evolution is deeply influenced by the broader cosmological context. Understanding the universe's evolution, therefore, is an essential prerequisite to any galaxy and AGN studies.

The standard cosmological model: Λ CDM

Our current understanding of cosmology is based on two main pillars. On one side, Einstein's theory of General Relativity, which describes how the structure of space-time is shaped by the universe's mass and energy content and distribution. On the other side, the cosmological principle, which states that the universe is homogeneous and isotropic on large scales.

Alongside this, several ground-breaking observational discoveries over the past century have defined our cosmological model. The discovery of the cosmic microwave background (CMB; [Penzias & Wilson 1965](#)) and its subsequent extensive study through the years have been especially crucial in shaping our current standard cosmological model, the Λ CDM (Λ Cold Dark Matter) model.

According to the Λ CDM model, we live in a flat (Euclidean) universe, which is accelerating due to the effect of a component known as dark energy (parametrised in the model by the cosmological constant, Λ). Dark energy constitutes about 70% of the total matter-energy content of the universe. The remaining is formed by about 25% of dark matter (DM) and only less than 5% of baryonic matter, that constitutes the visible universe (stars, gas, and galaxies). Observational evidence for the accelerated expansion of the universe comes, among other sources, from distance measurements of Type Ia supernovae up to $z \sim 1.5 - 2$, ([Riess et al., 1998](#); [Perlmutter et al., 1999](#)). The expansion rate of the universe is parametrised by the Hubble parameter, H , which is a function of cosmic time. The present-day value of this parameter is called the Hubble constant, H_0 .

Another fundamental parameter related with the expansion of the universe is redshift, defined as $1 + z = \lambda_{\text{observed}}/\lambda_{\text{emitted}}$, which quantifies the shift between the

emitted and observed wavelength of any electromagnetic radiation. Assuming a given cosmological model, which relates z with the universe's expansion, allows to use z as a measurement of cosmic distances.

Dark matter accounts for about 85% of the entire mass content in the Universe. Despite this, its physical nature is still a matter of debate. According to the standard Λ CDM model, DM is cold (non-relativistic) and collisionless. Even though DM does not interact electromagnetically with baryonic matter, its existence is supported by several observational findings that demonstrate its gravitational effects. Some of the most relevant evidences include the flattening of galactic rotation curves at large radii (Roberts & Rots, 1973; Rubin et al., 1978, 1980), the high dynamic mass in clusters of galaxies, and gravitational lensing of background galaxies (Zwicky, 1933).

The relative contribution of each one of the matter and energy components of the universe (dark energy, dark matter and baryonic matter) is parametrised by their time-dependent densities $\Omega(z)$. The most precise measurements of the present-day values of these densities, as well as of the Hubble constant are provided by the Planck Collaboration et al. (2016) as:

$$\left\{ \begin{array}{l} H_0 = 67.8 \pm 0.9 \text{ km s}^{-1}\text{Mpc}^{-1}, \\ \Omega_m = 0.315 \pm 0.013, \\ \Omega_{DM} = 0.2666 \pm 0.0022, \\ \Omega_B = 0.0494 \pm 0.0002, \\ \Omega_\Lambda = 0.685 \pm 0.012. \end{array} \right. \quad (1)$$

Galaxy formation in the Λ CDM model

As we just saw, DM dominates the matter content of the universe, accounting for about 85% of its total. Therefore, the birth and growth of large-scale structures is dominated by the gravitational potential of DM.

According to the standard cosmological model, when a perturbation of the density field exceeds a critical overdensity relative to the background density of the universe, the region within that excess density stops expanding and begins collapsing, forming gravitationally self-bound structures known as haloes (see e.g. Navarro et al., 1996). Once formed, these haloes grow over time by merging with and accreting smaller systems, resulting in progressively larger and more massive haloes. This model is known

as the hierarchical (or bottom-up) structure formation scenario. One consequence of this process is the creation of subhaloes, which are gravitationally bound structures smaller than the main halo they belong to, orbiting within the gravitational potential of their main halo. Baryonic matter follows the gravitational potential wells created by DM, and consequently, galaxies form at the centres of these haloes and subhaloes (White & Rees, 1978), with a stellar mass \mathcal{M} that is proportional to the DM mass of the halo (more on this on Sec. 1.4.1).

However, the formation and evolution of galaxies are not only governed by gravity, as it is for dark matter, but also by a vast range of diverse and complex physical processes. Some of these processes include: gas cooling, which is a necessary condition in order for the Giant Molecular Clouds of gas to collapse and form stars and structures; ionisation produced by the strong radiation fields of the newly born stars, which deeply influences the thermal and dynamical evolution of galaxies; and mechanical and radiative feedback, produced mainly by massive stars, supernovae and Active Galactic Nuclei (Sect. 1.2), which can significantly suppress star formation. The combination of all of these processes determines the fate of baryons that form the visible matter from the galaxies we observe today.

Understanding these processes is fundamental, since most of the information we have about the universe is obtained from observations of electromagnetic signals that originate from the baryonic component of galaxies. However, in the following, we will focus mainly on a specific type of galaxies, Active Galactic Nuclei. For a detailed description of the key processes that shape galaxy formation and evolution we refer the reader to the following reviews and books: Dayal & Ferrara (2018), Naab & Ostriker (2017), Mo et al. (2010) and Cimatti et al. (2019).

1.2 Active Galactic Nuclei

”Wir fanden ein Licht ganz hell und klar,
dass heller als die Sonne war;
als wir es sahen,
verging um uns die Nacht.

Wir haben es von allen Seiten betrachtet,
auf Farbe und Benehmen des Lichtes geachtet,
doch die Gründe für des Lichtes Kraft
bleiben rätselhaft!”

Witthüser & Westrupp, 1972

The central region of certain galaxies emits large amounts of radiation across nearly the entire electromagnetic spectrum, from radio waves to gamma rays. These regions, known as Active Galactic Nuclei (AGN), can exhibit extremely high bolometric¹ luminosities – up to $L_{\text{bol}} \sim 10^{47} \text{ erg s}^{-1}$ – originating from a compact and often spatially unresolved region. For decades, the most widely accepted explanation for the physical origin of this radiation has been the accretion of matter onto a supermassive black hole (SMBH; [Salpeter 1964](#); [Zel’dovich 1964](#); [Lynden-Bell 1969](#)): release of gravitational potential energy into heat and radiation, with a much higher mass-to-light conversion efficiency than nuclear fusion², could actually explain the observed luminosities with accretion rates of only a few solar masses per year (see Eq. 4 and the accompanying discussion in Sect 1.2.1).

This accretion flow is commonly assumed to take place through a thin accretion disc (Sect 1.2.1). Because of its angular momentum, the infalling gas naturally tends to accumulate in a plane. Viscosity is the key element of this disc: it allows the gas to transfer outwards its net angular momentum so that it can spiral into the central BH (and eventually be advected). In this process, a considerable fraction of the gas gravitational energy is converted into radiation (especially in the optical and UV range), while the rest is converted into kinetic energy or heat.

¹That is, integrated across the full electromagnetic spectrum.

²An efficiency of $\sim 10\%$ is commonly assumed for AGN, whereas the most favourable case of nuclear fusion, which is hydrogen burning, has an efficiency of 0.7% in the conversion of rest mass to energy.

The term AGN is often used to refer to a wide range of astronomical objects that exhibit a vast variety of observational features (Sect 1.2.2). Depending on their emission line widths and strengths and their multi-wavelength continuum emission, AGN can be classified as Seyferts, quasars, or LINERs, among other classes. Despite this observed diversity, unification models suggest that these different objects can be explained by changes in a few key factors, such as the presence – and orientation relative to the observer – of a dusty torus surrounding the central SMBH and the presence – or not – of jets. These models propose that most AGN share the same basic structure, with differences in appearance arising primarily from the angle at which we observe them. These observational properties of AGN have led to different methods to identify and classify AGN in different wavelengths (Sect 1.2.3).

Finally, several studies suggest that SMBHs are present in the nuclei of nearly all massive galaxies (Sect 1.2.4) and that their mass is closely linked to the properties of the galaxy’s bulge, suggesting that the evolution SMBHs and that of their host galaxies is codependent. During their active phases, SMBHs grow through gas accretion and influence star formation via feedback. Understanding these processes is crucial to understanding galaxy formation and evolution.

1.2.1 Accretion onto SMBH

Since the discovery of the first quasars³ in the 1960s, accretion onto a SMBH is believed to be the main source of energy powering these objects (e.g., Zel’dovich, 1964). This assumption, still widely accepted to this day, was mainly supported by two observational features of quasars: the high redshifts at which they were observed, which implied intrinsic high luminosities, and the rapid variability they exhibited, which implied that the underlying physical processes going on must occur at small scales. Current unification models assume that the same mechanism also powers other less luminous classes of AGN, like Seyferts or LINERs. In this section, we will cover the key aspects of the theory of accretion onto SMBHs.

Assuming a central accreting spherical body of mass M and radius R , the gravitational potential energy released by accretion of a mass m (falling free from infinity)

³We refer to quasars instead of AGN in this context because these were the first type of AGN to be discovered, due to their brightness. For a more detailed definition of quasars and other subclasses of AGN see Sect. 1.2.2.

onto its surface is

$$\Delta E_{\text{acc}} = \frac{GMm}{R}, \quad (2)$$

where G is the gravitational constant. If instead of for a single particle of mass m , accretion happens at a rate \dot{m} , and all mechanical energy is eventually radiated, the resultant luminosity is given by

$$L_{\text{acc}} = \frac{GM\dot{m}}{R}, \quad (3)$$

where we see that for a given central object with a given M/R ratio, the accretion luminosity depends only on the infall accretion rate, \dot{m} . Unlike other accreting objects like white dwarfs or neutron stars, black holes do not have a hard surface at a given radius R . Instead, there is a region at $r < r_s$ ⁴ where matter could fall in without losing its mechanical energy, and therefore not contribute to the total radiated energy. This uncertainty is usually parametrised by the introduction of a dimensionless quantity η , the efficiency. Therefore, for the relevant case of study in this work – accretion onto a SMBH – Eq. 3 becomes

$$L_{\text{acc}} = \frac{2\eta GM_{\text{BH}}\dot{m}}{r_s} = \eta \dot{m} c^2 \approx 5.7 \times 10^{46} \frac{\eta \dot{m}}{M_{\odot} \text{ yr}^{-1}} \text{ erg s}^{-1}, \quad (4)$$

where M_{BH} is the mass of the central black hole. We see from this equation that η measures the efficiency of the conversion of the rest mass of the accreting material into radiation. Making a good estimation of η is a key ingredient of modelling accretion onto BHs, for which detailed calculations involving general relativity are needed, and the result depends on the exact geometry of the accretion flow. Generally, a value of $\eta \approx 0.1$ is usually assumed as the standard. This equation also shows why accretion is a good candidate for explaining the observed luminosities in powerful AGN: assuming $\eta \approx 0.1$, accretion rates of a few tens of solar masses per year can yield intrinsic luminosities of $L_{\text{acc}} > 10^{47} \text{ erg s}^{-1}$.

Eddington Limit

In this section, we will derive an approximate theoretical upper limit for \dot{m} , the so-called Eddington accretion rate (\dot{m}_{Edd}). For a more detailed description, we refer

⁴ $r_s = 2GM_{\text{BH}}/c^2$ is the Schwarzschild radius of a non-rotating BH.

the reader to Chapter 1 of [Frank et al. \(2002\)](#). Some hypotheses are made for this derivation: *a)* the accretion flow onto the SMBH is steady (i.e., constant \dot{m}) and spherically symmetric, *b)* the accreting material is mainly fully ionised hydrogen and *c)* all accretion energy is converted into radiation. Under these circumstances, accretion can be halted if the accretion rate is high enough. Radiation emitted by the plasma which is closer to the central BH exerts an outward force on the infalling plasma at larger radii. This is due to Thomson scattering from that radiation on the free electrons⁵. If L_{acc} is the luminosity due to accretion (the only source of luminosity in the case of accretion onto BHs) and $\sigma_T = 6.7 \times 10^{-25} \text{ cm}^2$ the Thomson cross-section for electrons. The outward radial force exerted on a single electron at a distance r from the centre of the BH is given by

$$F_r = \frac{\sigma_T}{c} \frac{L_{\text{acc}}}{4\pi r^2}, \quad (5)$$

where c is the speed of light. As the radiation pushes the electrons out, they drag the protons because of their mutual electromagnetic attraction. On the other hand, the gravitational force pushes these electron-proton pairs to the centre with a strength

$$F_g = \frac{GM_{\text{BH}}(m_p + m_e)}{r^2} \approx \frac{GM_{\text{BH}}m_p}{r^2}. \quad (6)$$

The condition for radiation to be strong enough to stop accretion is that $F_r \geq F_g$, which implies

$$L_{\text{acc}} \geq L_{\text{Edd}} \equiv \frac{4\pi G m_p c}{\sigma_T} M_{\text{BH}} \approx 1.26 \times 10^{38} \left(\frac{M_{\text{BH}}}{M_{\odot}} \right) \text{ erg s}^{-1}. \quad (7)$$

This equation defines the maximum luminosity of the central source, the Eddington luminosity, such that for $L > L_{\text{Edd}}$ the radiation pressure would exceed the inward gravitational attraction and accretion would stop. From equation 4 we have $\dot{m} = L/\eta c^2$, which means that the existence of a limiting luminosity also implies a limit on the steady accretion rate, the Eddington accretion rate

$$\dot{m}_{\text{Edd}} = \frac{L_{\text{Edd}}}{\eta c^2} \approx 3.7 \times 10^{-8} \left(\frac{M_{\text{BH}}}{M_{\odot}} \right) M_{\odot} \text{ yr}^{-1}, \quad (8)$$

⁵Protons are not taken into account for this radiation pressure, since the scattering cross-section for protons is a factor $(m_e/m_p)^2$ smaller than that of electrons, with $(m_e/m_p) \approx 5 \times 10^{-4}$

where, for obtaining the numerical value on the right side of the equation, we have used $\eta = 0.06$. For a typical AGN harbouring a SMBH of mass $M \sim 10^8 M_\odot$, from equations 7 and 8 we have $L_{\text{Edd}} \sim 10^{46} \text{ erg s}^{-1}$ and $\dot{m}_{\text{Edd}} \sim 4 M_\odot \text{ yr}^{-1}$. Thus, powerful quasars like those seen at very high redshift can be sustained with just a few solar masses per year.

We note that because of the assumptions made for its derivation, we can only expect Eqs. 7 and 8 to be a first-order approximation. More realistic calculations depart from these values when the hypotheses are no longer valid, for instance with metallicity different to zero, non-spherical symmetry (e.g. accretion disc, see below), variability (very common in AGN) or partly neutral gas. Despite this, it is common when discussing about AGN to express the observed quantities in Eddington units (L/L_{Edd} and $\dot{m}/\dot{m}_{\text{Edd}}$).

Accretion Discs

In a realistic situation, the infalling gas will have a non-zero angular momentum component and will, therefore, orbit around the BH. The natural tendency of the accreting material is to concentrate in the equatorial plane, forming a thin accretion disc.

[Shakura & Sunyaev \(1973\)](#) developed the classical theory of geometrically thin and optically thick accretion discs. The basic idea is that, because of viscosity, the angular momentum of the gas is transferred outward, allowing it to spiral further into the centre. During this process, a considerable fraction of the gravitational energy in the gas is converted into electromagnetic radiation, with a relatively high efficiency that depends on the BH spin. General relativity predicts the existence of an innermost stable circular orbit (ISCO) around a black hole, such that matter at radii below this ISCO cannot maintain its orbital motion and falls directly into the BH event horizon without radiating its remaining energy. Therefore, a larger ISCO implies a smaller efficiency since less gravitational energy is extracted before the gas is advected. For a non-rotating (Schwarzschild) BH, general relativity predicts a value of $6r_g$ ⁶ for the ISCO which in turn implies $\eta = 0.057$, while for a Kerr BH rotating at its maximum spin the ISCO approaches r_g , and $\eta = 0.42$. A value of $\eta \approx 0.1$ is usually assumed as

⁶ r_g is the so-called gravitational radius, $r_g = GM/c^2$

the standard. A more detailed explanation of this matter can be found in Chapter 3 of [Netzer \(2013\)](#).

The emitted spectrum of such a geometrically thin and optically thick accretion disc can be approximately calculated making a very simple assumption, namely, that the disc is composed of rings of infinitesimal radial thickness, each one emitting as a black body with different temperature $T(R)$. The integrated spectrum resulting from adding up all the individual contributions has the shape shown in Figure 1.1, and is characterised by a temperature

$$T_* = \left(\frac{3GM_{\text{BH}}\dot{m}}{8\pi r_{\text{in}}^3 \sigma} \right)^{1/4} \approx 2 \times 10^5 \left(\frac{M_{\text{BH}}}{10^8 M_{\odot}} \frac{\dot{m}}{0.1 \dot{m}_{\text{Edd}}} \frac{(6 r_g)^3}{r_{\text{in}}^3} \right)^{1/4} \text{ K}, \quad (9)$$

where r_{in} is the innermost radius of the disc.

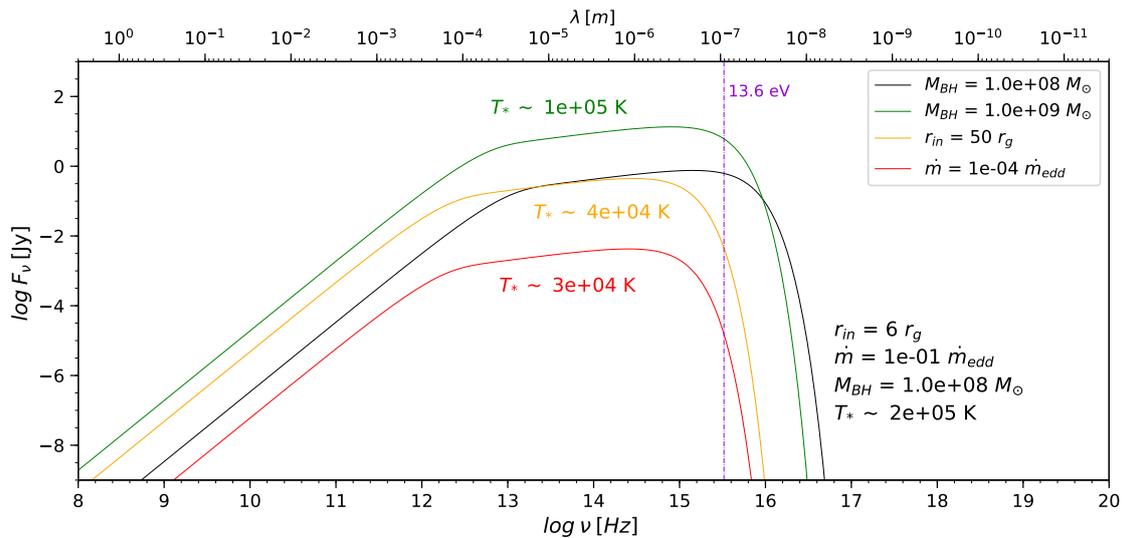


Figure 1.1: Emitted spectrum of a classical [Shakura & Sunyaev \(1973\)](#) thin accretion disc. The black line represents the emission from the disc of a standard quasar, reaching the ISCO, with r_{in} , \dot{m} and M_{BH} as indicated in the figure. The other lines represent changes in one of these parameters, keeping the others fixed. The green line shows the effect of increasing the mass of black hole one order of magnitude. The yellow line corresponds to a truncated disc, and the red line to a disc with a very low accretion rate. The vertical purple line shows the ionisation potential of hydrogen. Note how both the truncated disc and the one with low accretion rate provide far less ionising photons than the standard disc. For all discs $\eta = 0.06$ and $r_{\text{out}} = 10^4 r_{\text{in}}$ was assumed.

The order of magnitude of T_* tells us where to expect the peak emission of the AD. For a typical disc in a luminous quasar, the peak emission of the disc falls in the UV/optical range (black and green lines in Fig. 1.1). This feature is observed in real quasars and is usually known as the ‘big blue bump’. In contrast, for a truncated accretion disc (i.e. $r_{\text{in}} < \text{ISCO}$), or one with a highly sub-Eddington accretion rate ($\dot{m} \lesssim 10^{-3} \dot{m}_{\text{Edd}}$), this peak is expected to be shifted towards lower wavelengths (yellow and red lines in Fig. 1.1).

Specific accretion rate

Performing a direct measure of the real mass accretion rate onto a SMBH is effectively impossible with current technology. Instead, the *specific accretion rate* is a quantity derived from the ratio of two directly measurable properties, namely the X-ray luminosity and the host-galaxy stellar mass ($\lambda_{\text{SAR}} \propto L_{\text{X}}/\mathcal{M}$)⁷, and 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 relative to the stellar mass of the host galaxy. If a bolometric correction ($L_{\text{bol}} = k_{\text{bol}} L_{\text{X}}$) and a $M_{\text{BH}}/\mathcal{M}$ scaling relation are assumed, the specific accretion rate can also be regarded as a proxy for the Eddington ratio of the SMBH, $\lambda_{\text{Edd}} = L_{\text{bol}}/M_{\text{BH}}$ (Bongiorno et al., 2016; Aird et al., 2018).

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

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

where k_{bol} is the bolometric correction factor, assumed to have a constant value of $k_{\text{bol}} = 25$. This definition also assumes a constant scaling relation $M_{\text{BH}} = 0.002\mathcal{M}$ (Marconi & Hunt 2003, assuming also $\mathcal{M} \approx M_{\text{bulge}}$). Under these assumptions, $\lambda_{\text{sBHAR}} = \lambda_{\text{Edd}}$, and therefore an AGN accreting at 1% of the Eddington limit would have $\lambda_{\text{sBHAR}} \sim 10^{-2}$. It is possible to convert between the two definitions of the

⁷Throughout this thesis, we use L_{X} and F_{X} to refer, respectively, to the intrinsic luminosity and observed flux in the hard ([2 – 10keV]) X-ray band.

specific accretion rate as $\lambda_{\text{SAR}} \approx 10^{34} \lambda_{\text{SBHAR}}$. Therefore, with the same assumptions, the 1% of the Eddington limit corresponds to $\lambda_{\text{SAR}} = 10^{32} \text{ erg s}^{-1} M_{\odot}^{-1}$. 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.

1.2.2 The AGN zoo and unification models

In the next sections, we will explore how AGN have been historically classified into various subclasses depending on their diverse observational properties. This classification is relevant since most of these classes are still widely used nowadays. After, we will see how this diversity is interpreted as variations of a few physical parameters within a unified framework.

Observational Properties of AGN

AGN present a rich phenomenology that has motivated different historical classifications, including a large variety of types and subtypes associated with a complex diversity of observational characteristics (see e.g. [Heckman & Best, 2014](#)). These classifications follow different criteria, such as the width of the emission lines in the optical spectra, the relative strength of the radio band with regard to the optical, or the total bolometric luminosity. A short overview of this classification is presented here to introduce the subject of this work; for a much deeper view of the various classifications, we refer to [Padovani et al. \(2017\)](#).

In his pioneering work, [Seyfert \(1943\)](#) discovered that bright nuclei found in a number of nearby galaxies exhibited particularly strong emission lines in their optical spectra, corresponding to highly-excited transitions. Galaxies with these attributes are now called Seyfert galaxies. Their spectral energy distributions (SEDs), in the most general case, are characterized by both broad (FWHM⁸ $\sim 1000 - 5000 \text{ km s}^{-1}$) and narrow (FWHM $\sim 200 - 1000 \text{ km s}^{-1}$) emission lines. The broad lines arise from allowed atomic transitions (e.g. $\text{H}\alpha$), while the narrow ones arise from both allowed and forbidden transitions (e.g. $[\text{O II}] \lambda 3727$). Nuclei where these two types of lines are detected are classified as Seyfert I, while those exhibiting only narrow lines are called Seyfert II ([Khachikian & Weedman, 1971](#)). This separation between type 1

⁸Full width at half maximum.

and 2 is extended to other types of AGN. We show in Fig. 1.2 a comparison between typical optical type 1 and 2 AGN spectra, where the different emission line properties are clearly visible.

A crucial step in the understanding of AGN came twenty years later (Schmidt, 1963), with the discovery of quasars. Radio sources with point-like optical counterparts were identified, showing strong, broad emission lines at high redshift. However, because of the observed differences between these *quasi-stellar radio sources* and Seyfert galaxies, it still took some years until these were interpreted as two different manifestations of the same phenomena. The radio-quiet counterparts of quasars have also been identified, and they are known as quasi-stellar objects (QSOs), although the terms quasar and QSO are used indistinctly in the literature. Most quasars are type 1 AGN. Radio-loud galaxies of intermediate luminosity are distinguished between broad-line radio galaxies (BLRGs) and narrow-line radio galaxies (NLRGs) for types 1 and 2, respectively. Today, radio-loudness is associated with synchrotron radiation coming from relativistic jets or a compact radio core.

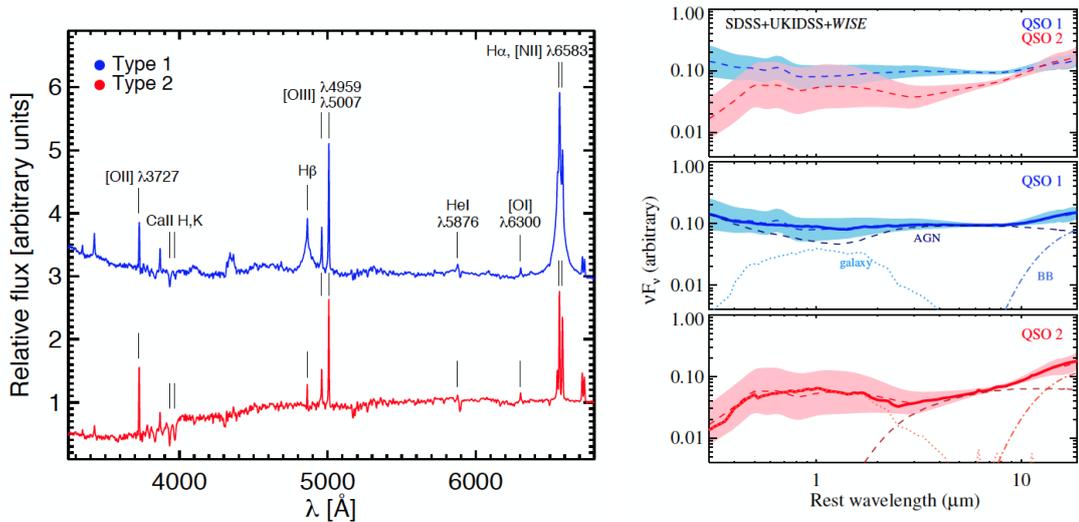


Figure 1.2: Comparison between type 1 (blue) and type 2 (red) optical and MIR spectra. *Left panel:* Composite optical SDSS spectra of type 1 and 2 AGN. As visible from the figure, type 1 exhibit both broad and narrow lines, while type 2 display only narrow lines. *Right panel:* Composite optical to MIR SEDs of type 1 and 2 quasars. In the middle and lower panel, the host galaxy and AGN contributions to the total SED are modelled for each AGN. Credits: [Hickox & Alexander \(2018\)](#).

While quasars represent the brightest type of AGN, most AGN are found in a much less luminous state, often even outshined by its host galaxy. Strong emission

lines corresponding to low ionisation potentials are commonly detected in these faint nuclei, which are therefore often called low ionisation nuclear emission regions, or LINERs (Heckman, 1980). Typical strong lines in the spectrum of LINERs are, e.g. [N II] λ 6548, 6584, [O I] λ 6300 and [S II] λ 6717, 6731. Another common definition for faint AGN is the class of low luminosity AGN (LLAGN), which tend to be dominated by its host galaxy emission in the optical emission, but are detectable with high angular resolution (typically sub-arcsecond) observations.

Finally, a major observational breakthrough took place in the last years, with the first direct images of accretion onto the event horizon of the SMBH in the nucleus of the M87 galaxy (Event Horizon Telescope Collaboration et al., 2019). This accomplishment provided unequivocal proof of the existence of SMBHs, and the accretion flows surrounding them.

Unification Models

Unification models aim to explain the wide variety of observational properties described in Section 1.2.2 as a function of a relatively small number of parameters, also taking into account the possible observational biases associated with the selection techniques (Padovani et al., 2017).

The unified models reviewed in Antonucci (1993) and Urry & Padovani (1995) share a common view of the AGN inner structure, where all AGN have the following main components: *i*) a supermassive black hole at its centre, with a typical mass of 10^6 – $10^9 M_{\odot}$; *ii*) a strong accretion flow, typically assumed to happen through a geometrically thin disc (Section 1.2.1); *iii*) a torus of cold gas and dust surrounding the accretion disc, as well as smaller gas clouds moving at different speeds, which form the broad line region (BLR) and narrow line region (NLR).

These two different regions are used to explain the two families of lines (broad and narrow) found in AGN. The BLR is assumed to be a high-density region close to the BH, formed by fast-moving gas clouds accelerated by the BH with typical Doppler broadenings of $\Delta v_{\text{FWHM}} \sim 1$ – $5 \times 10^3 \text{ km s}^{-1}$. This model also explains why broad lines corresponding to forbidden transitions are not observed: the high density of the BLR causes them to be collisionally suppressed, instead of being de-excited radiatively. On the other hand, the NLR is less dense, farther away from the BH,

and moving at lower speeds ($\Delta v_{\text{FWHM}} \sim 5 \times 10^2 \text{ km s}^{-1}$) in comparison to the BLR, which explains the emission of forbidden and allowed narrow lines.

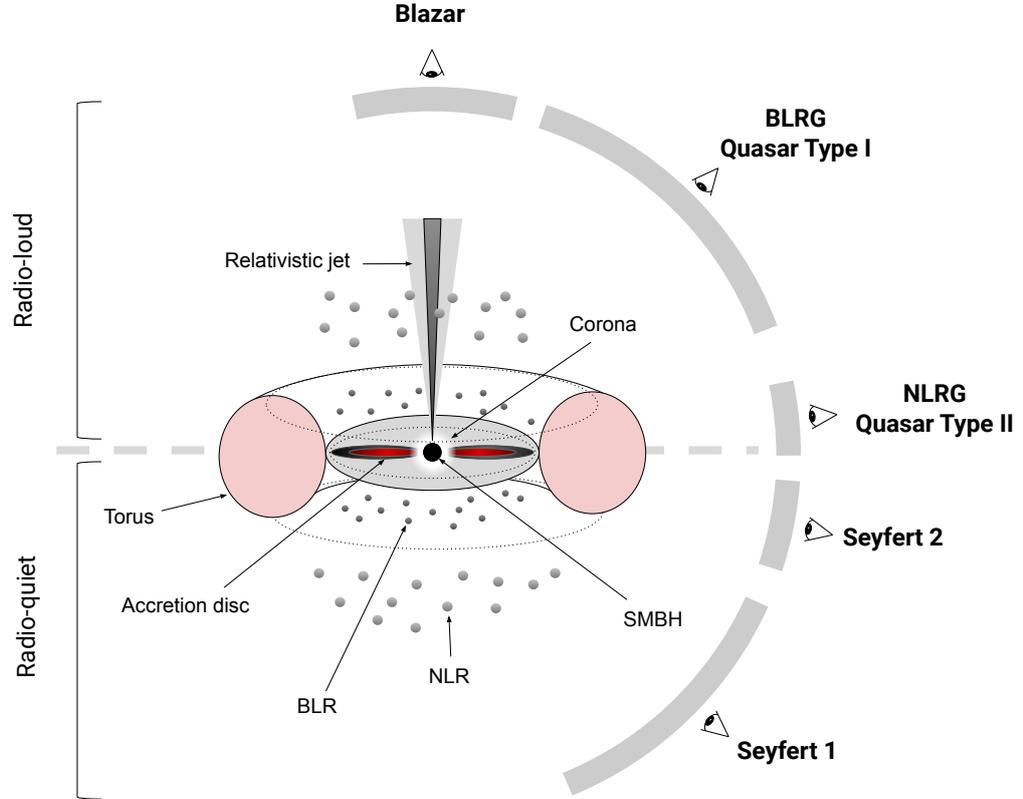


Figure 1.3: Schematic representation of a typical AGN according to unification models. The main components of the model are visible in the figure: a supermassive black hole at the centre, surrounded by an accretion disc and a dusty torus. Gas clouds, moving at different speeds and at different distances from the SMBH form the BLR and the NLR. X-ray emission originates from a region near to the BH known as the corona. Some AGN can also possess relativistic jets of collimated charged particles (upper half of figure). Depending to the viewing angle of the observer, and the presence or not of a jet, the AGN will be observed as different classes, such as Seyfert 1 or 2, quasar 1 or 2, or blazar. Credits: [Lopez \(2024\)](#), adapted from [Beckmann & Shrader \(2012\)](#).

This torus is a key component of unification models, as it is used to explain the observed dichotomy between type 1 and 2 AGN; when AGN are observed nearly edge-on, the torus shields the nucleus and the BLR and we observe a type 2 AGN. On the other hand, on type 1 AGN the orientation of the torus allows us to directly detect the BLR. For this reason, type 1 and type 2 AGN are often referred to as

unobscured and obscured AGN, repetitively.

Besides, the torus fulfils two other important roles: it acts as a source of gas and molecular material that feed the accretion flow, and it collimates radiation from the NLR, which is observed with the shape of a cone in many galaxies. The shape of the torus determines the opening angle of this cone. Finally, reemission of the accretion disc radiation by the dusty torus is used to explain the IR emission of AGN.

Relativistic jets suppose another important feature of unification models, which is, however, not included in all of them. These are collimated outflows of charged particles, which are accelerated at relativistic speeds along the axis perpendicular to the disc. Their size ranges from a few parsecs up to several hundred kilo-parsecs, exceeding the size of their host galaxy. The non-thermal emission related to the jets is associated with radio-loud AGN and is also used to explain the power-law shape of the X-ray flux. The formation mechanism of these jets is not well understood, but they are thought to involve magnetic fields and the rotational energy of the black hole (Blandford & Znajek, 1977; Lu et al., 2023). Besides, it is unclear if jets are a common component of all AGN. Jets also define a new type of AGN, blazars, which are AGN in which the jet direction is closely aligned with our line of sight.

Finally, most current unification models also include a hot corona composed of relativistic electrons surrounding the accretion disc (Haardt & Maraschi, 1991; Fabian et al., 1989). Although the precise mechanisms behind the heating of the corona are not yet fully understood, it is believed to result from a combination of magnetic reconnection, turbulence, and Inverse Compton processes. The simplest corona models consist of hot ($T \sim 10^8 - 10^9$ K), optically thin gas situated above or within the inner regions of the accretion disc.

Each of these individual components is reflected on a characteristic signature on the global AGN emission across the full electromagnetic spectrum, as shown in Fig. 1.4. The accretion disc is the main radiation source, and its emission peaks in the optical/UV wavelengths. Part of this radiation is absorbed and reemitted by the dirty torus in the NIR to FIR. The hot corona and the jets are the main sources of X-ray emission, while jets are also used to explain the radio emission. The figure does not include the emission lines produced in the BLR and the NLR, another fundamental piece of unification models to understand the observed overall SED of AGN.

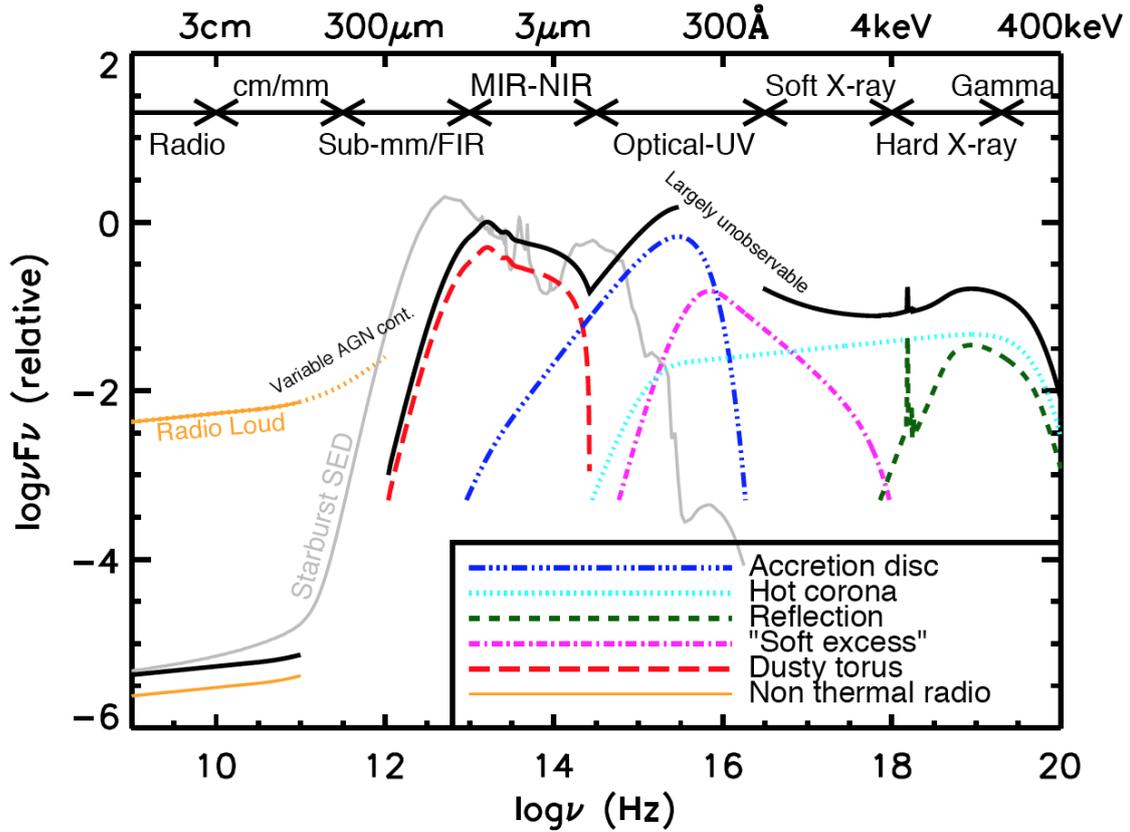


Figure 1.4: Schematic representation of the emission produced by the individual components of AGN according to unification models. The black solid line represents the overall SED, while the various coloured lines illustrate each individual component. The two different yellow lines correspond to radio-loud and radio-quiete AGN. Credits: [Harrison \(2014\)](#)

1.2.3 Identification of AGN

In this section, we will briefly review some of the methods for identifying and selecting AGN. We will focus on two portions of the electromagnetic spectrum: the X-ray band, due to its high efficiency and completeness at selecting broad populations of AGN with different properties, and the IR/optical/UV bands, since methods based on these bands will be used later in this thesis to test and validate our catalogues.

X-ray selection of AGN

X-ray observations offer one of the most complete and pure methods for identifying AGN. Although X-ray emission typically account for less than 10% of the AGN bolometric luminosity (Duras et al., 2020), using the X-ray band for AGN selection presents significant advantages over other wavelength bands.

First, it provides a selection with a high degree of purity, with very low contamination from other X-ray emitters such as star-forming galaxies or stars. X-ray AGN photons are generated through non-thermal processes near the central black hole (Sec. 1.2.2), and therefore, they serve as an ideal tracer of AGN activity (Mushotzky et al., 1993; Brandt & Hasinger, 2005). This is further confirmed by observations performed on deep X-ray surveys: as shown in Fig. 1.5, the density of X-ray emitters is totally dominated by AGN for $F_{0.5-2\text{keV}} \gtrsim 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$; for the hard band (2 – 8 keV) the contamination from galaxies is negligible for even lower fluxes.

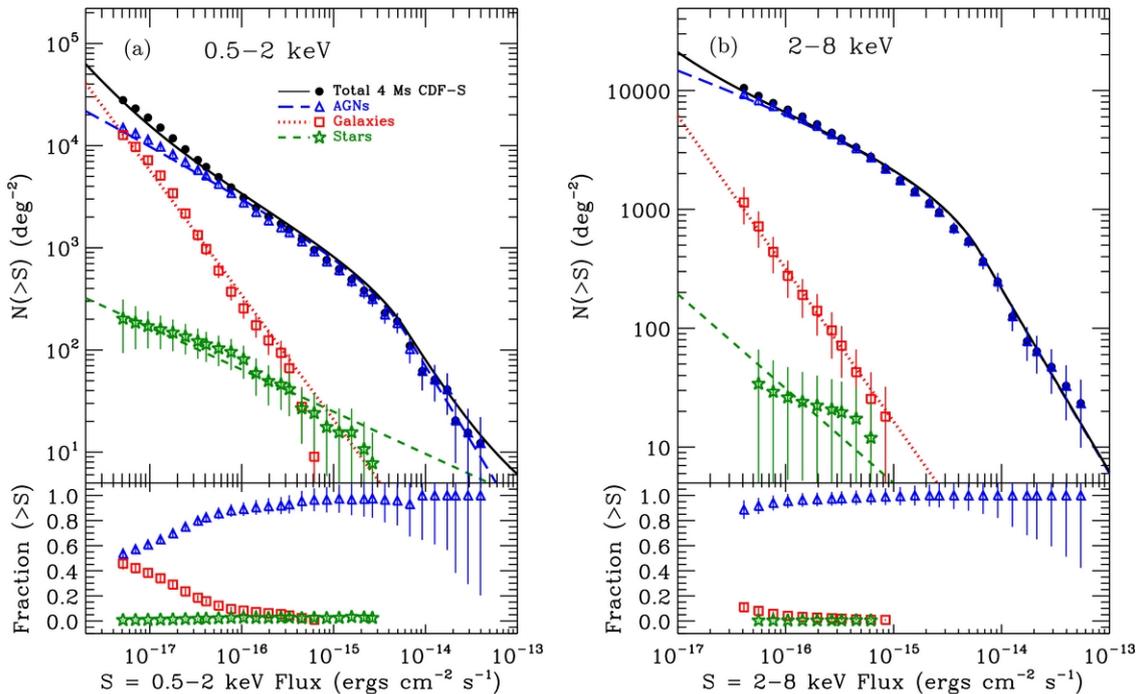


Figure 1.5: Cumulative number counts of sources emitting above a given X-ray flux from the 4 Ms *Chandra* Deep Field-South (CDF-S) survey, the deepest X-ray survey to date. Each coloured line represents the contribution from AGN, star-forming galaxies and stars, while the black line shows the total counts. The left (right) panel shows the results for the soft (hard) X-ray band. Credits: Lehmer et al. (2012).

Besides, observations from the nearby universe (Iwasawa et al., 2009, 2011; Lehmer et al., 2010; Pereira-Santaella et al., 2011), show that the most powerful star-forming galaxies all have X-ray luminosities of $L_{0.5-8\text{keV}} \lesssim 10^{42} \text{ erg s}^{-1}$. Therefore, sources emitting above this threshold can be safely identified as AGN. Other common techniques to sort out AGN from galaxies based on their X-ray emission include examining the steepness of their X-ray power-law spectrum, setting thresholds on flux ratios between X-ray and other wavelengths (such as optical or radio), and analysing the X-ray variability.

On the other hand, the X-ray selection of AGN provides also samples with a high degree of completeness. As mentioned before (Sect. 1.2.2), AGN spend most of their life-time in a low-luminosity state rather than in a quasar state. In fact, LLAGN are the most common population of active nuclei, including about one third of all galaxies in the local Universe (Ho, 2008). X-ray observations provide an effective method to identify these LLAGN, which are often outshined by their host galaxy in other wavelengths.

A complete census of AGN must account non only for the intrinsically faint population, but also for the obscured one. The amount of obscuration in the X-ray spectrum of AGN is quantified with the hydrogen column density (N_{H}), which measures the amount of intervening gas and dust along the line of sight. Based on this parameter, AGN are typically classified in the X-ray as unobscured for values of $N_{\text{H}} < 10^{22} \text{ cm}^{-2}$, and obscured above this threshold. When the hydrogen column density exceeds a threshold given by the inverse of the Thomson cross-section $N_{\text{H}} > \sigma_{\text{T}}^{-1} \approx 1.5 \times 10^{24} \text{ cm}^{-2}$, the absorbing medium becomes effectively opaque to X-ray photons, and AGN are classified as Compton-thick (CTK) AGN (see e.g. Comastri, 2004).

The effect of obscuration depends on the rest-frame energy of the emission: lower-energy photons are more absorbed than those with higher energies. Therefore, X-ray selections (and particularly in the hard band, i.e., $E_{\text{phot}} > 2 \text{ keV}$) are highly effective at identifying obscured AGN. This includes those with significant nuclear obscuration ($22 < \log N_{\text{H}}/\text{cm}^{-2} < 24$) that may be overlooked in other bands like optical or UV (e.g. Buchner et al., 2015; Padovani et al., 2017; Hickox & Alexander, 2018). However, the most heavily obscured CTK sources are still missed by most X-ray surveys, and their exact fraction is still unclear (see e.g. Treister et al., 2004; Buchner et al., 2015; Ricci et al., 2017b; Pouliaxis et al., 2024).

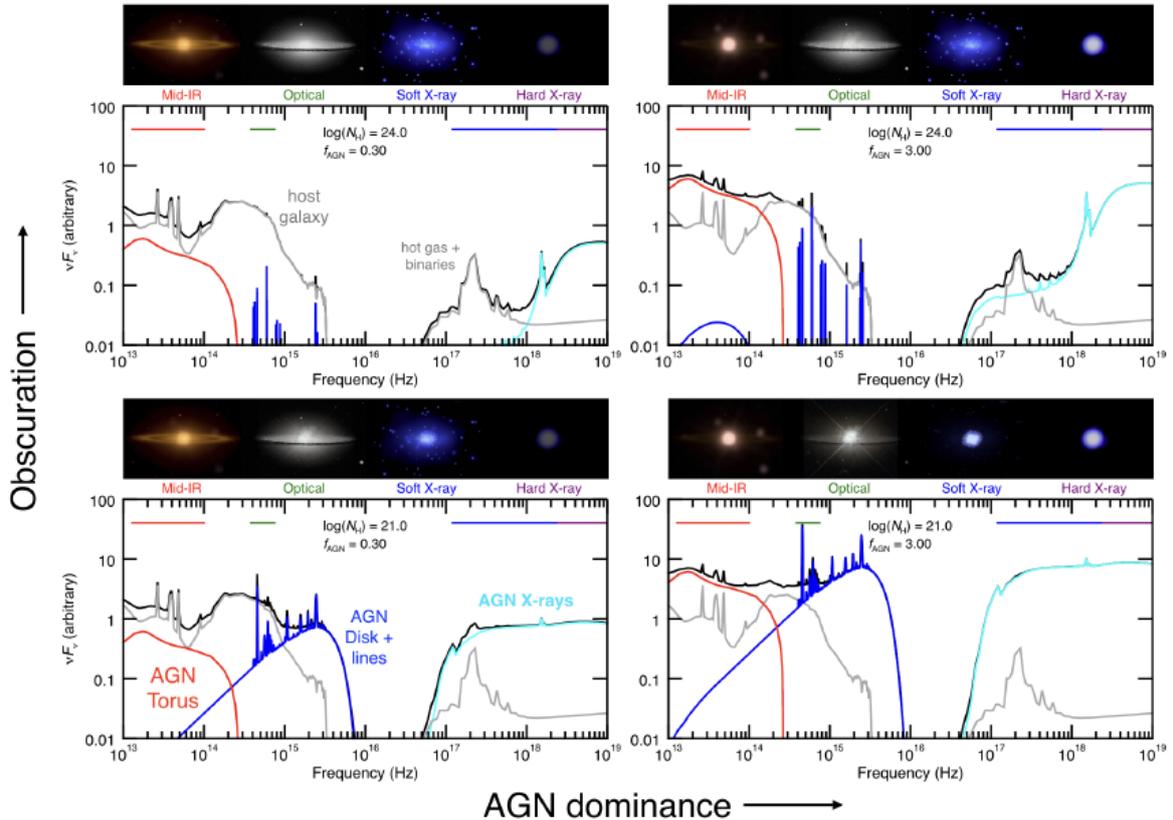


Figure 1.6: Schematic representation of how different levels of obscuration and ratio of AGN to host galaxy contributions influence the overall multi-wavelength spectrum of the source. Credits: [Hickox & Alexander \(2018\)](#).

To illustrate this discussion, we show in Fig. 1.6 how varying levels of obscuration and intrinsic brightness of AGN relative to their host galaxies influence the overall multi-wavelength SED of the source. This figure clearly demonstrates why hard-band X-ray selections of AGN can achieve levels of purity and completeness that are not attainable with other wavelengths. For nuclear-dominated sources (right panels of Fig. 1.6), unobscured AGN can easily be identified at almost any wavelengths. Instead, for obscured sources, the host galaxy totally dominates the optical/UV continuum, making the AGN signature detectable only through its emission lines. In the case of sources dominated by the galaxy emission (left panels), we see a similar situation, with the addition that the AGN continuum is not detectable any more at the Mid-IR. In all cases, however, the AGN emission dominated over its host galaxy at the hard X-ray band.

Ultra-violet to near-infrared

As we have seen in the previous sections, AGN release a considerable fraction of IR, optical and UV emission, with distinctive features that differentiate them from those of non-active galaxies. Therefore, several methods have been developed to select AGN on these bands.

Unobscured (type 1) and bright AGN can be efficiently selected based on their IR to UV broadband colours (see e.g. Richards et al., 2002; Jin et al., 2019; Euclid Collaboration: Bisigello et al., 2024). However, this type of photometric selections are much less efficient at identifying AGN with either significant amounts of obscuration or a big contribution from their host galaxy emission (see Fig. 1.6).

Spectroscopy offers a key opportunity to address this challenge. Indeed, AGN emission is rich in UV-NIR emission lines (see e.g. Fig. 1.2 for examples of optical lines). By studying the FWHM of certain emission lines (e.g. $H\alpha$), it is possible to discriminate between star-forming galaxies and type 2 AGN (often with $\text{FWHM} \lesssim 1000 \text{ km s}^{-1}$, and lower in the case of galaxies) against type 1 AGN, that can show dispersions of several thousands of km s^{-1} (right panel of Fig. 1.7).

Besides, by studying the ratios of the fluxes of different narrow emission lines⁹, often combining forbidden and permitted lines, it is possible to infer information about the physical conditions (density, metallicity, strength of the ionising radiation,...) of different gas phases. This allows for the construction of emission-line diagnostic diagrams, where different line ratios are studied against each other to differentiate between AGN and star-forming emission (Ho et al., 1997; Kewley et al., 2001; Kauffmann et al., 2003; Juneau et al., 2011).

We show in the left panel of Fig. 1.7 an example of such diagram, namely the BPT (Baldwin et al., 1981) diagnostic diagram. The plot is divided into three regions, which differentiate sources depending on what is their main source of ionisation: H II regions (i.e. star-forming regions in galaxies), a type 2 AGN, or a LINER. Finally, we note that while these diagrams offer a powerful tool to identify obscured AGN, they can still miss a considerable fraction of them, for example, in cases of low SMBH mass (Cann et al., 2019) or for AGN with significant host-galaxy contribution. A possible

⁹To minimise effects of dust reddening on the emission-line ratio, the wavelengths of both lines should be as close as possible

workaround for the latter limitation consists on studying spatially resolved emission-line diagnostic diagrams, that can catch AGN signatures missed in the integrated emission (Mezcua & Domínguez Sánchez, 2020, 2024).

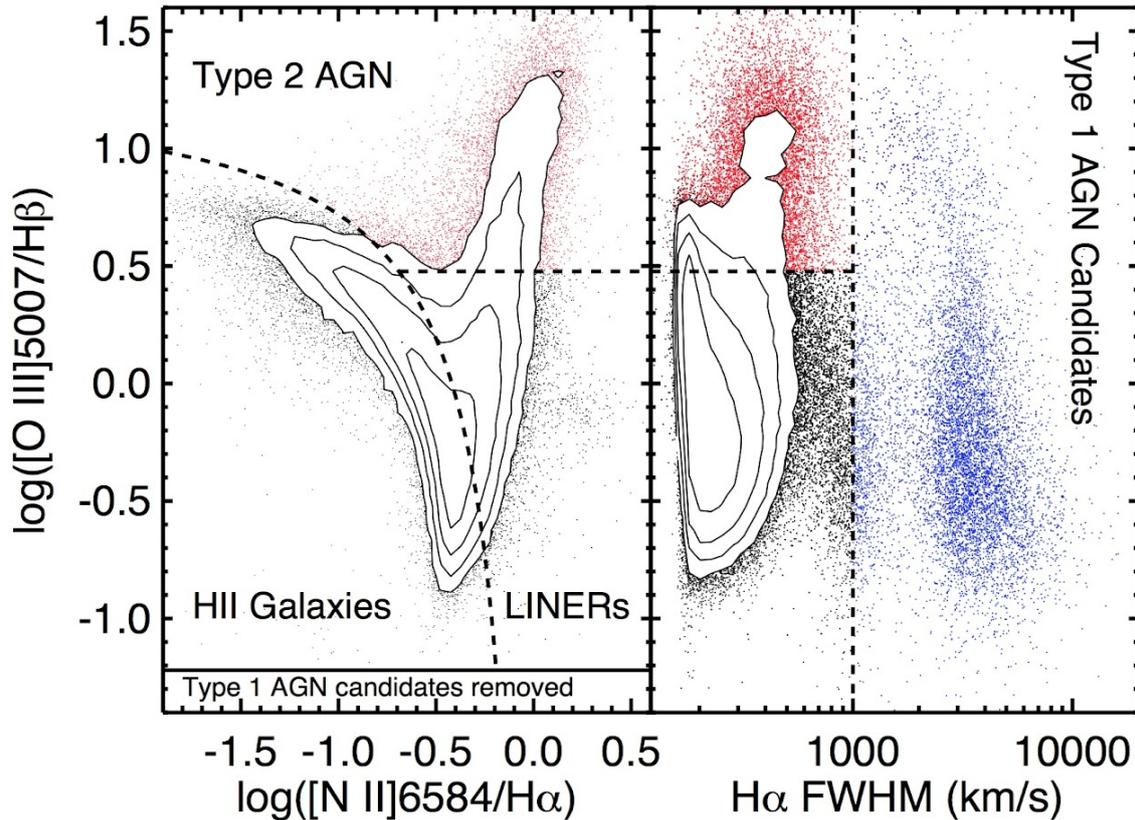


Figure 1.7: Optical emission-line diagnostics used to identify AGN. All data shown comes from the SDSS DR7 database for extragalactic sources with $z < 0.4$. *Left:* BPT diagram for sources with $\text{FWHM}_{\text{H}\alpha} < 1000 \text{ km s}^{-1}$ (to remove the majority of Type 1 AGN). The dashed lines are taken from Ho et al. (1997) and Kauffmann et al. (2003) to identify different source classifications (i.e., type 2 AGN, LINERS and H II regions (in star-forming galaxies)). *Right:* broad $\text{H}\alpha$ emission-line profiles (i.e., $\text{FWHM} \gtrsim 1000 \text{ km s}^{-1}$) may indicate the presence of an AGN broad line region. It is possible for reasonably broad emission lines (i.e., $\text{FWHM} \sim 1000 - 2000 \text{ km s}^{-1}$) to be produced in the NLR due to outflows, and therefore care needs to be taken when classifying AGN with these ‘intermediate’ line widths. Credits: Harrison (2014)

1.2.4 Co-evolution of AGN and galaxies

There are strong observational evidences supporting the presence of supermassive black holes ($M_{\text{BH}} \sim 10^6\text{--}10^9 M_{\odot}$) in the nucleus of essentially every massive galaxy ($\mathcal{M} \gtrsim 10^{10} M_{\odot}$; [Magorrian et al. 1998](#); [Kormendy & Ho 2013](#)). Besides, the mass of this BH is tightly related to several properties of its host galaxy bulge (like velocity dispersion of the stars, stellar mass and luminosity; [Ferrarese & Merritt 2000](#)). This suggests that massive BHs and their host galaxies have grown together during a common evolutionary phase, a scenario supported by observations of star formation at high-redshift galaxies ([Madau & Dickinson, 2014](#)) and the larger fraction of active nuclei among quenched (passive) galaxies ([Heckman & Best, 2014](#)).

Galaxies with very massive but (currently) passive BHs at their nucleus might have gone through active AGN phases in the past. During this phase, their BHs grow through gas accretion ([Rees, 1984](#); [Marconi et al., 2004](#)), while releasing large amounts of energy to the ISM via outflows and radiation, a process known as AGN feedback and that has a great impact on regulating star formation ([Harrison, 2017](#); [Martín-Navarro et al., 2018](#)). Therefore, the study of AGN phenomena is directly connected to the formation and evolution of present-day galaxies, and it is consequently fundamental to study these two populations (AGN and non-active galaxies) simultaneously.

1.3 Next generation of surveys

In the upcoming years, ‘full-sky’ space and ground-based surveys will cover unprecedentedly large sky areas, gathering photometric and spectroscopic data for billions of galaxies and (at least) millions AGN. This perspective offers promising advancements in our understanding of galaxy formation and evolution, cosmological parameters, and the nature of dark energy and dark matter.

Some of these surveys will be (or, in some cases, are being) conducted by space telescopes, such as *Euclid* (Euclid Collaboration: Mellier et al., 2024) and the Roman Space Telescope (Spergel et al., 2015), which will perform both photometric and spectroscopic surveys in the NIR and optical bands, or eROSITA (Merloni et al., 2024), which will survey the X-ray sky with an unprecedented spectral and angular resolution.

In addition, other large-scale surveys will be conducted from ground-based observatories. Some examples of this are: MOONS (Multi-Object Optical and Near-infrared Spectrograph; Cabral et al., 2020) at the Very Large Telescope (VLT) in Chile; UNIONS (Ultraviolet Near-Infrared Optical Northern Survey; Ibata et al., 2017), a collaboration of three wide-field telescopes based in Hawaii, which will complement *Euclid*’s view of the northern sky with ground-based photometry; the Vera C. Rubin Observatory (Rubin; Ivezić et al., 2008; LSST Science Collaboration et al., 2009), currently under construction in Chile, that will carry out the LSST survey, covering the entire available sky every few nights on the UV to NIR *ugrizy* bands.

In the following section, we will focus specifically on the *Euclid* mission, and we will describe in detail its scientific goals and technical specifications. We have chosen this mission as a case study to demonstrate different applications of the empirical model presented in this thesis, such as making forecasts or testing pipelines.

1.3.1 The Euclid Surveys

The European Space Agency’s *Euclid* space telescope, successfully launched in July 2023, is designed to explore the origin of the Universe’s accelerating expansion. By measuring the growth of cosmic structures with unprecedented precision over most of the cosmic time, *Euclid* will investigate the nature of dark matter, dark energy, and

gravity itself. In essence, its surveys aim to reveal how cosmic acceleration influences the expansion history and the 3-dimensional distribution of matter in the Universe. To achieve this, *Euclid* will measure the shapes of over a billion galaxies and determine precise redshifts for tens of millions, enabling studies of weak gravitational lensing and galaxy clustering. The mission is specifically optimised for two key cosmological probes: Weak Gravitational Lensing (WL) and Baryon Acoustic Oscillations (BAO). However, given the unprecedented volume of gathered observational data, the mission is also expected to provide major advancements for galaxy evolution and AGN studies.

Instruments

Euclid is equipped with two on-board instruments; the VISible instrument (VIS, [Euclid Collaboration: Cropper et al., 2024](#)) carries a single broadband optical filter, I_E , covering the approximate wavelength range 550–900 nm (see Fig. 1.8 and Table 1), with a mean image resolution of about $0.23''$ and a field of view of 0.57 deg^2 .

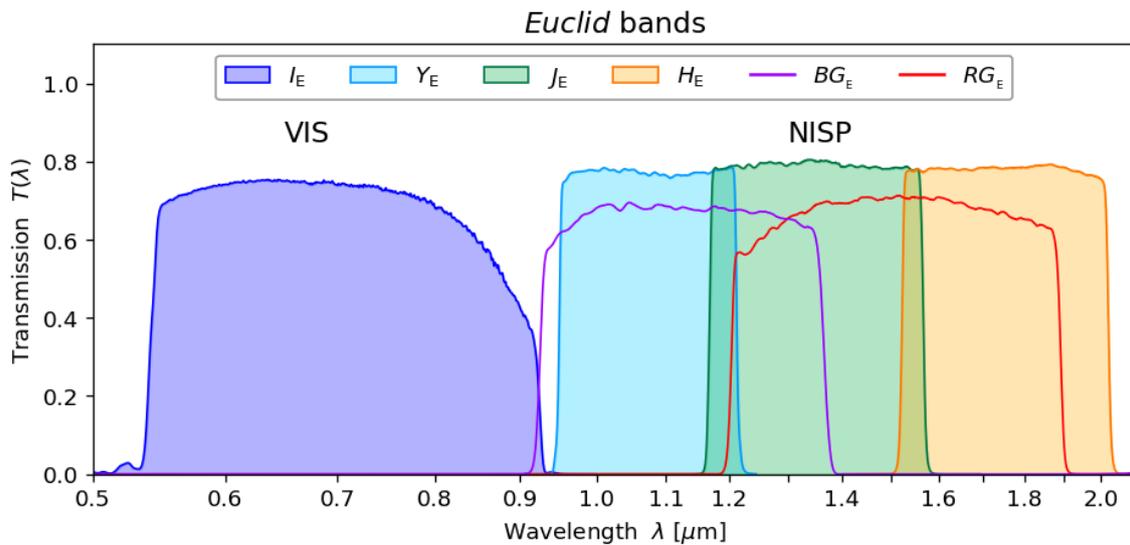


Figure 1.8: Transmission of *Euclid* broadband filters from the VIS (I_E) and NISP (Y_E , J_E and H_E) instruments, and from the blue and red grisms. Source: <https://www.euclid-ec.org>.

The Near-Infrared Spectrometer and Photometer (NISP, [Euclid Collaboration: Jahnke et al., 2024](#)), on the other hand, possesses three NIR photometric broadband filters: Y_E , J_E and H_E (Fig. 1.8 and Table 1). The novelty of *Euclid*, though, is the slitless spectroscopy in the NIR bands that aims at spectroscopically observing

galaxies and AGN in the range $z \sim 1 - 2$, thanks to the detection of the $H\alpha$ emission line. To achieve this, the NISP instrument is equipped with 4 different low-resolution NIR grisms (with a spectral resolution of $R = 380$ for a $0.5''$ diameter source): one blue grism ($0.92 - 1.3 \mu\text{m}$), and 3 identical red grisms ($1.25 - 1.85 \mu\text{m}$, Fig. 1.8). Thanks to these instruments, *Euclid* will produce a legacy dataset with images and photometry of more than a billion galaxies and several million spectra, out to high redshifts ($z \sim 2$), delivering also morphologies, masses and SFRs for all these targets. Such dataset is particularly relevant for galaxy formation and evolution.

The surveys

During its 6 year mission, *Euclid* will perform two main surveys: a Wide Survey (EWS), planned to cover an area of about 14 000 square degrees (about one-third of the total sky), and a Deep Survey (EDS) of 53 deg^2 (Euclid Collaboration: Mellier et al., 2024). The observations for the EWS started on February 2024, and are currently ongoing.

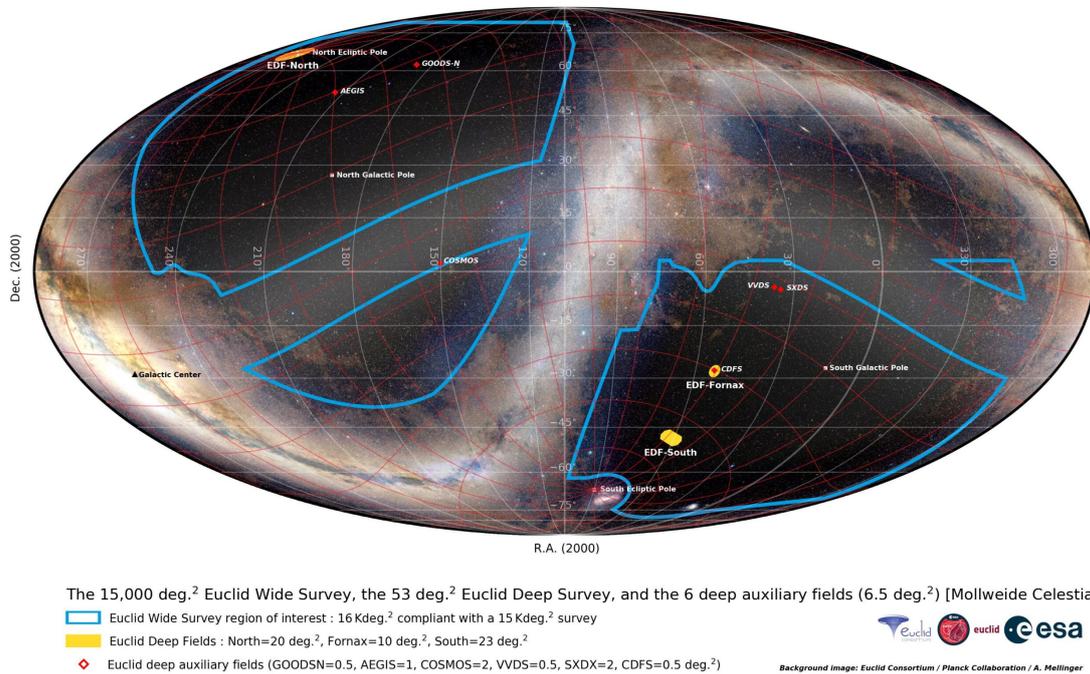


Figure 1.9: All-sky map in a 2D projection showing the main *Euclid* surveys. The Wide Survey is shown in blue, the three *Euclid* Deep Fields are marked in yellow, and the Auxiliary Fields in red. Source: <https://www.euclid-ec.org>

For the EWS, the sky will be observed with only one pass of the red grism, which for the $H\alpha$ line translates into a redshift coverage of $0.9 \leq z \leq 1.8$ (see Fig. 1.10 for the redshift coverage of several emission lines at the EWS and EDS) and a flux limit of $F_{H\alpha} > 2 \times 10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$ for a $\text{SNR} = 3.5$ detection. Instead, for the EDS the redshift coverage of $H\alpha$ spans to $0.4 \leq z \leq 1.8$ with a limiting flux $F_{H\alpha} > 6 \times 10^{-17} \text{erg s}^{-1} \text{cm}^{-2}$. The expected limiting magnitudes in the EWS for point sources detected with a minimum SNR of 10 are about $m_{\text{NISP}} = 23.5$ and $m_{\text{VIS}} = 25.0$ (Table 1), and at least two magnitudes deeper in the EDS.

Besides, *Euclid* will perform observations on seven Auxiliary Fields (six well-known fields, plus an extra field for photometric self-calibration), that will be used to calibrate photometric redshifts and colour gradients (Euclid Collaboration: Scaramella et al., 2022). A sky 2d projection of the expected EWS, EDS, and Auxiliary Fields regions is shown in Figure 1.9.

To achieve its scientific goals, high-precision photometric redshifts (see Sec. 4.4 will be required. To this aim, *Euclid* observations will be complemented with ground-based photometry from the UNIONS, LSST, and DES (Dark Energy Survey; Dark Energy Survey Collaboration et al., 2016) surveys in order to have a complete wavelength coverage between $0.3 \mu\text{m}$ and $1.8 \mu\text{m}$.

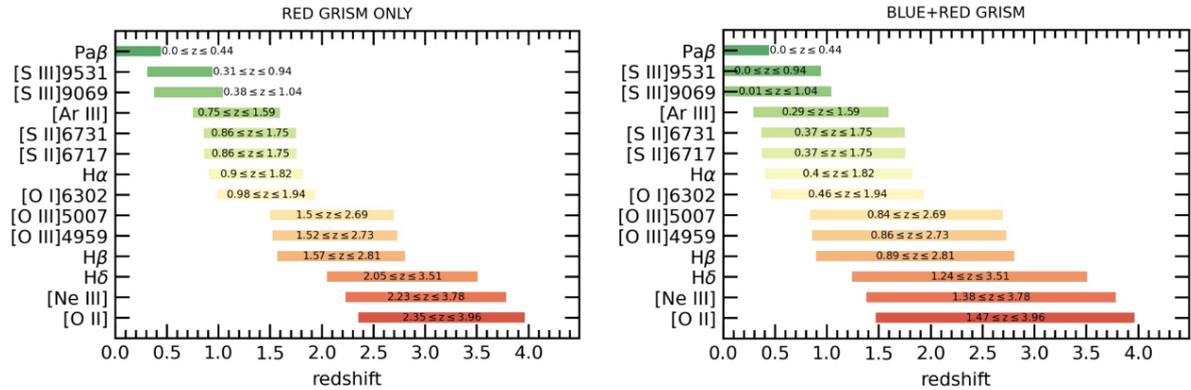


Figure 1.10: Redshift coverage of several emission lines as observed with *Euclid*'s red grism only (as for the EWS) and with the red and blue grisms (EDS).

Table 1: *Euclid* filters, together with other filters used in this work. Columns one to three: reference, minimum and maximum wavelength of each filter, in Å. Source: SVO Filter Profile Service (Rodrigo et al., 2012). Fourth column: expected limiting magnitude for a point-source observed with a 2'' diameter aperture (proxy for extended sources) for a minimum SNR = 10, at the EWS.

Telescope/Survey	Filter	λ_{ref}	λ_{min}	λ_{max}	$m_{\text{lim}} @10\sigma$, EWS
<i>Euclid</i>	I_{E}	7103	5300	9318	25.0
<i>Euclid</i>	Y_{E}	10 785	9382	12 435	23.5
<i>Euclid</i>	J_{E}	13 621	11 523	15 952	23.5
<i>Euclid</i>	H_{E}	17 649	14 972	20 568	23.5
LSST	u	3680	3206	4081	23.6 [†]
LSST	g	4782	3876	5665	24.5 [†]
LSST	r	6218	5377	7055	24.1 [†]
LSST	i	7532	6766	8325	24.2 [†]
LSST	z	8685	8035	9375	23.4 [†]
GALEX	FUV	1535	1340	1809	–
GALEX	NUV	2301	1693	3008	–

Notes: For the EWS we assumed the conditions of the first data-release of the northern sky (NDR1). The corresponding SNR = 5 limiting magnitudes can be found by adding 0.75 magnitudes to the reported one, while for the EDS (assuming 40 passes) one should add 2 magnitudes.

[†] We note that for the NDR1, the *ugriz* photometry will be provided by the UNIONS collaboration, and the SNR limits reported here correspond to these filters. While the λ coverage is reported for the LSST filters, we note that the differences are minimal.

1.4 Simulations of galaxies and AGN

Over the last decades, simulations of astrophysical phenomena have become an essential tool across nearly all areas of astrophysics. Simulations act as ‘virtual laboratories’, allowing researchers to experiment with different parameters, test new theories, and refine models that may be impossible to validate through direct observation or experimentation. In practice, they allow us to explore processes and structures across a wide range of scales, from atomic processes and stellar formation to cosmological models of the entire observable universe.

In the case of AGN and galaxy evolution, different types of simulations have been developed over the years, depending mainly on the scales involved and the aim of the simulation (see [Wechsler & Tinker, 2018](#), for a review). First, we have the models based on codes that apply well-established laws of physics, starting from well-defined initial conditions based on observational data. These initial conditions then evolve over time until arriving to the formation of galaxies and/or large-scale cosmic structures. These simulations can be DM-only simulations or hydrodynamical models (if they involve DM and baryons). A second family of models consists of simulations that apply baryonic physics that model galaxy formation and evolution on top of DM-only simulations. These are known as semi-analytic models (SAMs).

These two classes of models have two main purposes: on one side, by comparing the outputs of the simulations with observations, we can test the validity of the recipes and physical laws imposed a priori; on the other side, assuming that these recipes truly describe the real universe, they can be used to make predictions on future observations, or even give us hints about the fate of the universe. Despite their significant value, these simulations have the main limitation of being computationally very expensive¹⁰, and that they need to use prescriptions to model some key processes of galaxy evolution (like star formation or AGN feedback) that cannot be yet described from fundamental physical laws.

A third class of simulations are the so-called empirical models, which are the type studied in this thesis. The main goal of empirical models is to produce mock catalogues of astrophysical sources that mimic the observations as realistically as possible.

¹⁰For example, the TNG50 hydrodynamical simulation took ~ 130 Mh ($\approx 15\,000$ years) of CPU core time, running on 16 320 supercomputer cores ([Pillepich et al., 2019](#)).

Instead of relying on physical first principles, these models are fully calibrated based on observed empirical relations. Empirical models have the main advantage that they are computationally much more efficient than their analytical counterparts, allowing them to simulate large volumes of space in a relatively short time-scale (see e.g. Appendix B for a review on the computational efficiency of the model presented in this work). This quality makes them particularly useful to provide forecasts for future large surveys. We will describe these models in more detail in Sec.1.4.1.

In practice, these modelling approaches are more of a continuum rather than three very distinctive methodologies. We show in Fig. 1.11 a compilation of different types of models used to simulate DM and galaxy properties. Moving from left to right in this figure, we move from more physically oriented to more observational data-based models.

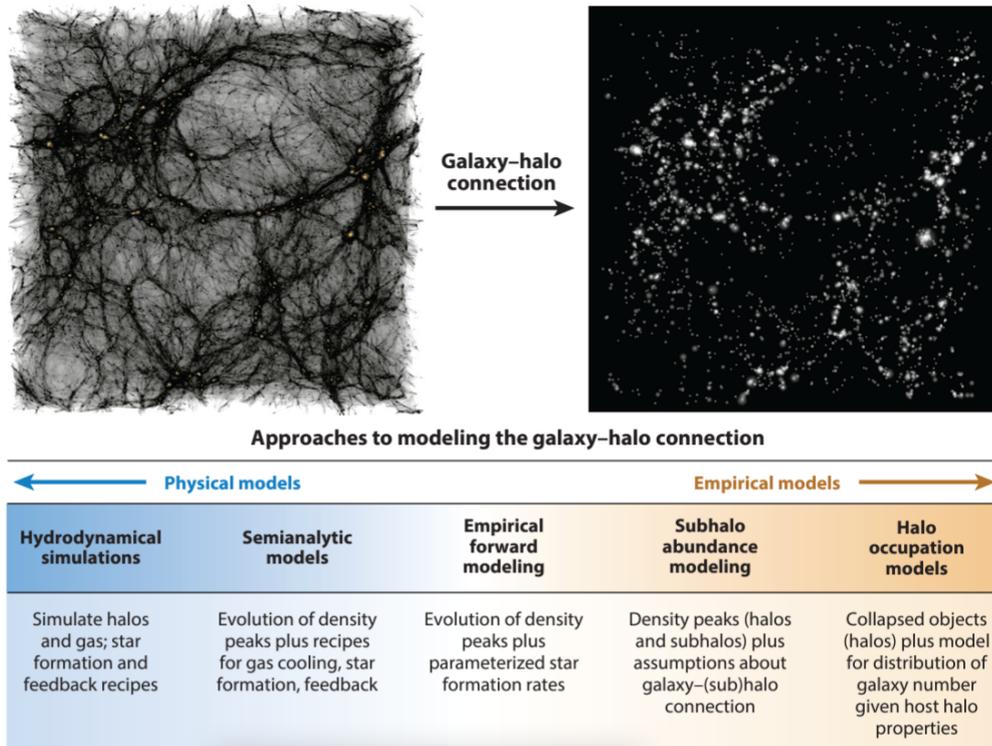


Figure 1.11: Different modelling approaches used to simulate DM and galaxy properties. *Top panel:* DM distribution in a cosmological simulation (left), compared to the galaxy distribution (right) modelled through an abundance matching model (see Sec. 1.4.1). The bottom grid shows the key assumptions of various models for the galaxy-halo connection. From left to right, these models range from more physical to more empirical models. Credits: [Wechsler & Tinker \(2018\)](#).

N-body DM simulations

As mentioned in Sec. 1.1, in the Λ CDM scenario, the growth of large-scale structure in the universe is essentially shaped by the DM distribution and evolution. Dark matter-only simulations model this by solving the gravity equations of general relativity, assuming a universe made only of DM, that is, without considering baryonic physics.

One of the most popular numerical techniques used to achieve this is the N-body method. N-body simulations solve the equations of motion for particles under a gravitational field, typically involving an immense number¹¹ of particles N . The number of particle-particle interactions that need to be computed increases on the order of N^2 , making direct integration of the differential equations prohibitively computationally expensive, especially when simulating large volumes of the universe.

To address this issue, numerical integration is typically carried out on small timesteps, and often using various approximation techniques (see e.g. Kravtsov et al., 1997; Monaco, 2016). Alternatively, one can reduce the number of particles involved (usually at the cost of increasing their individual mass, therefore reducing the simulation’s resolution) or simulate smaller volumes of the universe. Storing the output simulated data provides another technical difficulty, due to its large volumes. Therefore, simulation data are usually only saved for some (pre-fixed) timesteps (the so-called snapshots). These snapshots are typically three-dimensional boxes that contain information on the position and velocities of every particle at a fixed redshift.

While snapshots can be useful for studying the properties of a large number of galaxies at the same redshift, it is difficult to compare them with actual observations. One of the most interesting derived data products of N-body DM simulations is the so-called lightcones, which mimic the cone shape of three-dimensional space when observed from Earth with a given opening angle. Lightcones are constructed by arranging particles belonging to sections of different snapshots at evolving redshift, based on their comoving distances (with respect to the observer) and their position within a defined field of view (see e.g. Blaizot et al., 2005; Giocoli et al., 2014). In practice, this method transforms the comoving coordinates of the particles (X, Y, Z) into observed coordinates (RA, Dec, redshift) by combining information from several snapshots, to create a smooth redshift distribution. Fig. 1.12 offers a graphical

¹¹The *Euclid* Flagship N-body DM simulation (Euclid Collaboration: Castander et al., 2024), the largest N-body simulation performed to date, contains four trillion particles.

schematic representation of this process.

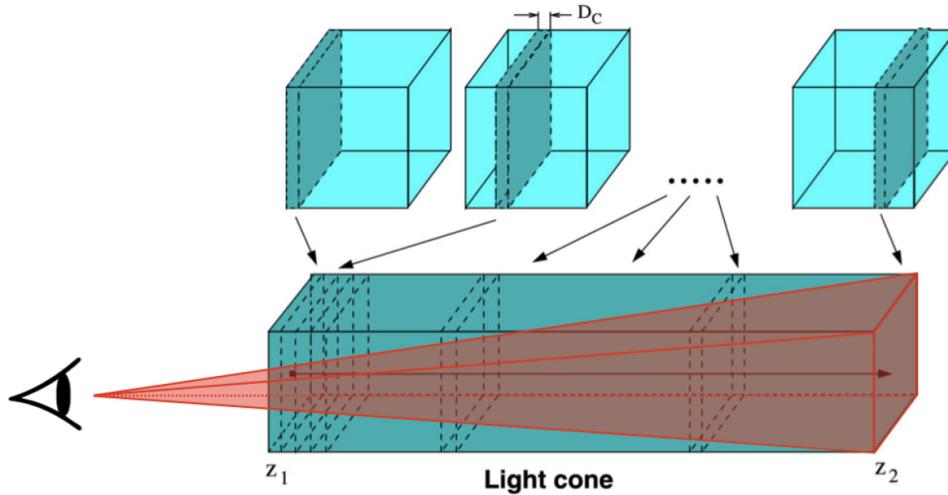


Figure 1.12: Schematic representation of the construction process of DM lightcones by combining sections from several snapshots, each one at a different fixed redshift. Credits: M. Bolzonella, adapted from [Zawada et al. \(2014\)](#).

Finally, to compare simulation results with observations of the real universe, analytical tools are essential for mapping simulation data onto actual cosmic structures. Traditionally, this has been achieved through the use of ‘halo finders’, which identify DM haloes as locally over-dense gravitationally bound systems. Examples of such methods are the friends-of-friends algorithm ([Davis et al., 1985](#)) or more recent and sophisticated methods, such as SUBFIND ([Springel et al., 2001](#)) and ROCKSTAR ([Behroozi et al., 2013](#)), which offer improved accuracy and efficiency in halo identification.

Hydrodynamical simulations

Hydrodynamical simulations suppose an additional level of complexity, modelling the formation of individual DM haloes and galaxies by simultaneously solving the fundamental equations of gravity, hydrodynamics and thermodynamics. These simulations begin with initial conditions grounded in observational cosmology and evolve these conditions across particles or grid cells representing DM, gas, and stars. With this framework, simulations incorporate a wide range of known galaxy formation physics into a cohesive model.

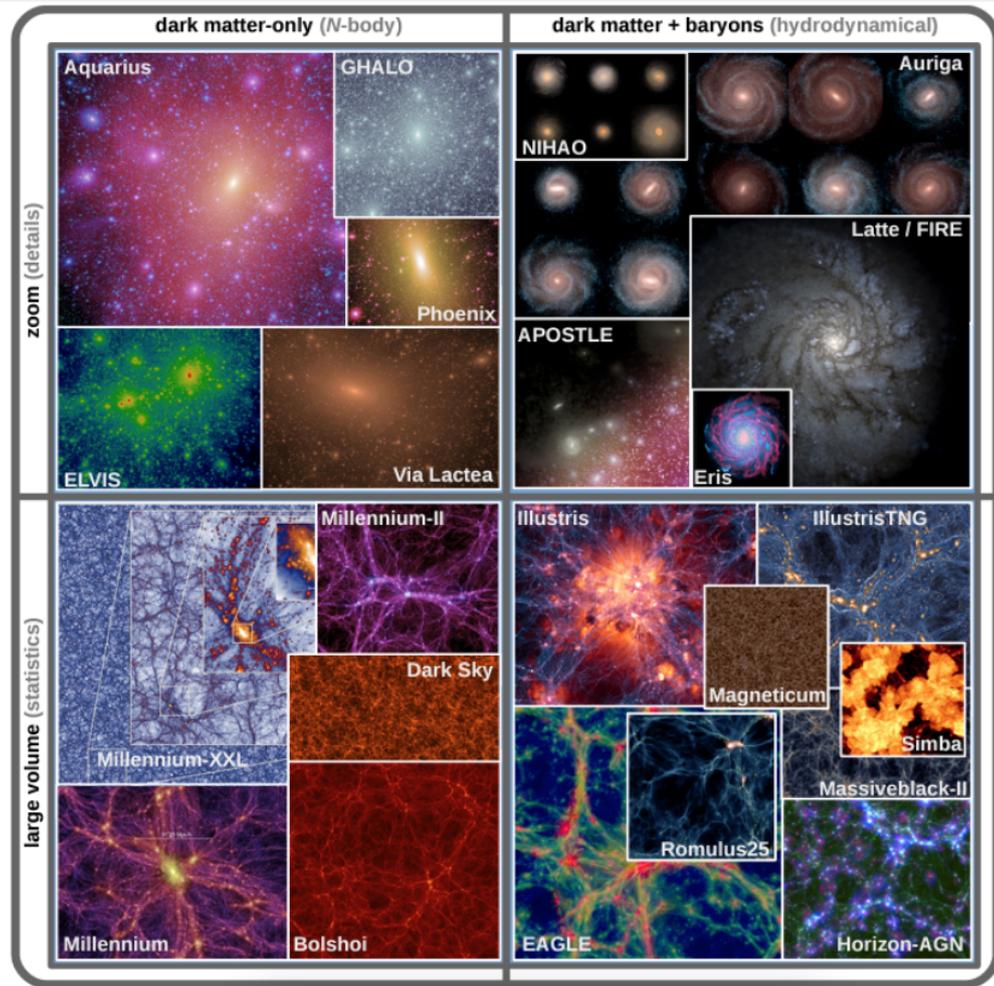


Figure 1.13: Visual representations of the results from some recent large numerical simulations. The left panels show results for DM-only simulations, while the right panel shows hydrodynamical simulations of DM and baryons. The top panels show zoom-in (high-resolution) simulations that allow us to see small-scale details such as structure within galaxies. In contrast, the lower panels focus on large-scale cosmological simulations. Credits: [Vogelsberger et al. \(2020\)](#).

This is an extremely powerful method that can track and predict the interplay between dark matter and baryons and the resulting large-scale structure (depending on the simulation size). Galaxy properties are typically modelled through sub-grid (i.e. scales below the resolution limit) recipes for gas cooling, star formation, and feedback.

However, the enormous computational effort required for hydrodynamic simulations restricts the size of volumes and range of physical parameters that can be explored at high resolution, even more so than for the above described DM-only sim-

ulations. This trade-off often requires choosing between detailed modelling of smaller volumes (zoom-in simulations) or simulating larger scales with reduced resolution (cosmological large-scale simulations).

Recent hydrodynamic simulations successfully reproduce observed properties of galaxy populations. (e.g. EAGLE, [Schaye et al. 2015](#); NIHAO, [Wang et al. 2015](#); Illustris, [Sparre et al. 2015](#); IllustrisTNG, [Nelson et al. 2018](#)). However, the accuracy of detailed predictions remains highly dependent on the underlying implementation of baryonic physics and the assumptions made for complex, not fully understood processes, such as star formation and AGN feedback.

We show in [Fig. 1.13](#) a representation of some recent hydrodynamic simulations, divided into large-volume simulations, that provide statistical samples of galaxies, and zoom-in simulations, that resolve smaller scales in more detail. Furthermore, they are also divided into DM-only (N-body) and DM plus baryons (hydrodynamical) simulations.

Semi-analytic models

Semi-analytic models (SAMs) incorporate baryonic physics (through analytic models) on top of N-body DM-only simulations. These models use simplified prescriptions (that approximate the complex processes involving galaxy formation and evolution), integrating them with DM merger trees. Specific components such as gas accretion, star formation, AGN activity, and feedback mechanisms are included with varying complexity levels depending on the specific aims of the model. Many different semi-analytic models have been developed in the last years (see e.g. [De Lucia & Blaizot, 2007](#); [Henriques et al., 2012](#); [Gonzalez-Perez et al., 2014](#); [Henriques et al., 2015](#); [Fontanot et al., 2024](#)). We refer to [Benson \(2010\)](#) for a review on the topic.

One of the primary advantages of SAMs is their computational efficiency. They significantly reduce the computational demands compared to fully numerical simulations, making it possible to make predictions for larger volumes or simulate galaxies over a larger range of halo mass.

However, both SAMs and hydrodynamic simulations face the limitation that they must approximate physics below their respective resolution limits, which usually correspond to galaxies for SAMs and grid elements for hydrodynamical simulations. Dif-

ferent possible reasonable approximations can lead to significantly different resulting galaxy properties (Lu et al., 2014; Kim et al., 2016)

1.4.1 Empirical models

As mentioned earlier (see Sect. 1.3), the next generation of missions and surveys will gather unprecedentedly large amounts data for billions of galaxies and AGN. In order to efficiently design these surveys, as well as to enhance the scientific return from this new data, it is essential to count on simulations that allow us to produce synthetic data that mimics the observed properties of astrophysical sources. Empirical models offer a very valuable tool to achieve this. Many of these models have been developed in the last years (see e.g. Birrer et al., 2014; Schreiber et al., 2017a; Moster et al., 2018; Georgakakis et al., 2019, 2020; Comparat et al., 2020; Allevato et al., 2021; Girelli, 2021; Bisigello et al., 2021; Euclid Collaboration: Selwood et al., 2024)

Empirical models are calibrated based on observed empirical relations. As opposed to the other types of models described above, they do not rely solely on physical first principles, but instead, the physical constraints come entirely from observations.

These simulations are fundamental to complement the scientific outcome of large surveys. They can be used, for example, to provide forecasts of the outcomes of future observations that can, in turn, be used to optimise the observing strategies. This is especially crucial for space-based observatories, given their limited operative lifetime. Besides, mock catalogues produced with these models can be used to test and optimise data analysis pipelines and to understand selection biases and systematic uncertainties that might arise in the real data. In Chapter 4, we will give examples of all these applications using the empirical model presented in this thesis.

Synthetic data can also be helpful for interpreting the observations; in particular, it can be used to estimate incompleteness and biases and to refine and validate hypotheses based on the simulations.

Empirical methods have several advantages. First, their physical properties, by construction, are in good agreement with observations, which is particularly useful for the applications mentioned above. Second, they are computationally efficient compared to other methods, such as SAMs and hydrodynamical simulations, which allows the simulation of large volumes of the universe. And third, the parametrisation

of the model is very flexible, given that it depends on the input observations and can be easily changed depending on specific needs. This allows empirical models to constrain the unknown aspects of the galaxy with observed data. These last two points (efficiency and flexibility) allow to run several realisations of the model with different parametrisations and explore the differences in the output simulations.

At the same time, the main drawback of this type of models come precisely from the fact that they are less directly connected to fundamental physical principles than hydrodynamic simulations or SAMs, and they can inherit biases from observations. Because of this, they have significantly lower predictive power than them, and they are not as effective for confronting specific physical models against observations (Wechsler & Tinker, 2018).

The stellar-to-halo mass relation

Empirical models often start from N-body DM simulations and assign all the physical properties of galaxies on top of DM haloes. A necessary first step for doing so is to determine what is the typical stellar mass (\mathcal{M}) of a galaxy that resides in the core of a given DM halo of mass M_h . This correlation is known as the stellar-to-halo mass relation (SHMR). We show in Fig. 1.14 the SHMR parametrised by several different works using different techniques, collected in Wechsler & Tinker (2018).

One way to parametrise the SHMR is to make a very simple assumption: the most massive galaxies live in the most massive DM haloes, and vice-versa. This approach can be generalised by assuming that a given galaxy property (like its stellar mass or its luminosity) scales monotonically to a halo property (like its DM mass or circular velocity), and also by taking into account the DM subhalos that live within larger haloes. Therefore, the relation between dark matter halos or subhalos and galaxy properties can be found by performing a one-to-one association from their distributions. This approach is called (sub)halo abundance matching (SHAM), and the only observational input needed to perform it is the stellar mass function (SMF) or luminosity function of galaxies. We note that in practice a certain scatter is added to this one-to-one relation.

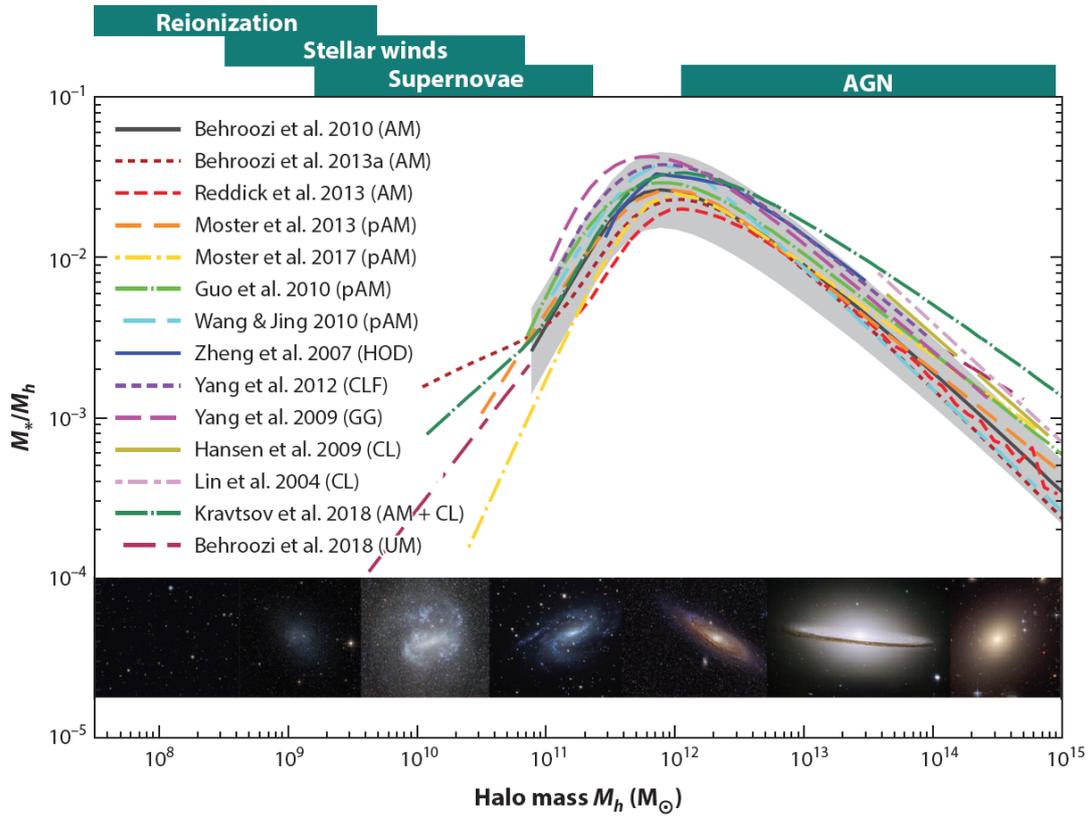


Figure 1.14: Galaxy stellar-to-halo mass ratio of central galaxies at $z = 0$. The figure shows constraints from a number of different methods indicated in the legend, such as abundance matching (AM in the label, see also SHAM in the text). The bottom panel shows example galaxies hosted by halos in the specified mass range. The top of the figure indicates several key physical processes that may be responsible for ejecting or heating gas or suppressing star formation at those mass scales, therefore shaping the SHMR. Credits: [Wechsler & Tinker \(2018\)](#).

1.5 Aim of this thesis

As presented along this Introduction, synthetic data that can accurately reproduce the observed properties of astrophysical sources (i.e., empirical mock catalogues) are a fundamental tool for the scientific exploitation of current and forthcoming large-sky surveys.

Besides, it is essential that these catalogues include the contribution from AGN in order to accurately model a significant fraction of the extragalactic sources. Indeed, at certain cosmic epochs, AGN can account for a significant fraction of the entire galaxy population (up to $\sim 20\%$ for the most massive galaxies, see e.g. [Aird et al. 2018](#)). Moreover, AGN exhibit distinctive photometric and spectroscopic signatures across the full electromagnetic spectrum, such as strong and broad emission lines in the optical and ultraviolet, infrared excess due to dust reemission, hard X-ray emission from the hot corona, and enhanced ultraviolet continuum emission (the so-called big blue bump), among others. Accurately reproducing these features in realistic empirical catalogues is crucial for a comprehensive understanding of galaxy evolution.

In this context, the main aim of this PhD thesis is to develop a new computational model to create galaxy and AGN empirical mock catalogues, that simultaneously reproduce the physical and observed properties of both populations. Specifically, in this thesis we focused on the inclusion of the AGN component into an already existing galaxy mock catalogue (see [Girelli 2021](#) for a detailed description of the galaxy simulations). Previous similar works often model only one of these two populations, focusing either on reproducing galaxy (see e.g. [Birrer et al., 2014](#); [Schreiber et al., 2017a](#); [Moster et al., 2018](#)) or AGN properties (see e.g. [Comparat et al., 2020](#); [Allevalo et al., 2021](#); [Euclid Collaboration: Selwood et al., 2024](#)). Besides, these works often focus on a given wavelength range (e.g. only the X-ray or the optical band). For this work, we aimed to realistically reproduce the physical properties of galaxies and AGN, such as stellar masses, star-formation rates, sizes, or AGN obscuration at optical and X-ray bands. Additionally, the catalogue includes multiwavelength photometric and spectroscopic information for all these sources, such as intrinsic luminosities and observed broadband fluxes in the UV-NIR and X-ray, narrow emission line fluxes, and SEDs in the UV to NIR.

The empirical approach that we have followed to produce this model presents several advantages: not only does it save computation time with respect to other more physics-oriented models (hydrodynamic simulations or semi-analytic models), but it also ensures that results align well with most observed scaling relations. In addition, thanks to its optimised computational efficiency, it can also be quickly adapted to implement new empirical relations following new discoveries. At the same time, with this methodology, we preserve the link with the cosmic web traced by DM haloes. This connection is fundamental for deriving cosmological forecasts and linking the visible properties of galaxies and AGN to the DM distribution.

This thesis is organised as follows:

- In Chapter 2: Building the catalogue, we briefly describe the previously existing empirical galaxy catalogue, MAMBO (Mocks with Abundance Matching in BOlogna), developed by Girelli (2021). We then extensively describe the methodology followed to populate this catalogue with obscured and unobscured AGN, and how we modelled their intrinsic physical properties, their X-ray fluxes, and their UV to NIR broadband emission and spectral energy distribution (SED) and spectral features.
- In Chapter 3: Validating the catalogue, we present several comparisons between our mock catalogue and observations that were not used for its calibration, in order to check the validity of its forecasts and applications.
- In Chapter 4: Applications of MAMBO, we present several application of this catalogue, using the space mission *Euclid* as a case example. We present examples of different forecasts (e.g. AGN number counts on a given band), test the performance of a pipeline for retrieving photometric redshifts, and analyse *Euclid*-like simulated AGN spectra using official *Euclid* pipelines.
- Finally, in 5: Summary, we briefly recap the main methodologies and results presented during the thesis.

Chapter 2: Building the catalogue

In this Chapter we will present the methodology followed to construct the mock empirical catalogues of galaxies and AGN. In Sec. 2.1, we outline the methodology used to create the galaxy mock catalogue. Subsequently, Sec. 2.2 describes the process of populating the galaxy catalogue with AGN.

2.1 Mock Galaxy catalogue: MAMBO

MAMBO (Mocks with Abundance Matching in BOlogna) is a workflow designed to construct an empirical mock catalogue of galaxies which 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. In the rest of this thesis, instead, we will 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. 2.2, is presented in the flowchart of Fig. 2.1.

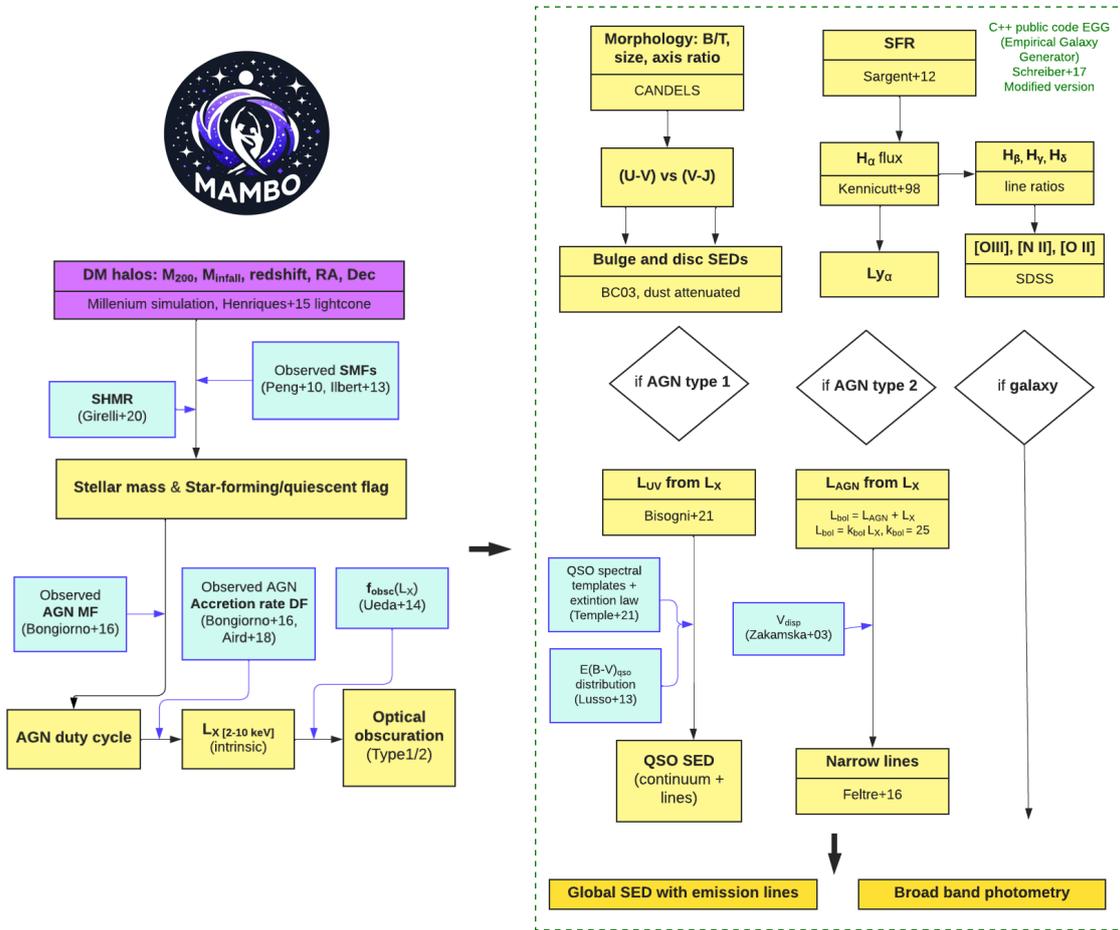


Figure 2.1: Schematic view of the full MAMBO workflow, used to create the galaxy and AGN mock catalogue. Yellow boxes represent the result of each step, while blue boxes show the necessary inputs. Steps within the green dotted box take place within the modified version of the public code EGG.

MAMBO takes in input few quantities from cosmological N-body DM simulations, i.e. the DM halo mass, expressed as M_{200} ¹² for main haloes and M_{infall} ¹³ for subhaloes and orphans¹⁴; the redshift z of each halo and subhalo. The sky coordinates RA and Dec are not used as inputs, but are needed 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 which 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 $M_{\text{halo}} \gtrsim 10^{10.24} M_{\odot} h^{-1}$, 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 $z \sim 0$, the SMF evaluated by [Peng et al. \(2010\)](#) on the SDSS survey and divided into passive and star-forming 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

¹²Mass within the radius where the halo has an overdensity 200 times the critical density of the simulation.

¹³Subhalo 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](#)).

¹⁴Subhaloes that have lost all or part of their mass.

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 star-forming fraction was extrapolated from the results at lower redshifts, arriving at a maximum of $f_{\text{SF}} = 99\%$ 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. (2023) 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. 2.2 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., 2017b). EGG is a C++ code designed to generate an empirical mock catalogue of galaxies with realistic physical properties (such as 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 UV to the submillimeter. The code has been calibrated purely by using empirical relations to produce realistic observable properties. In order to assign a 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 star-forming galaxies. Infrared SEDs are instead derived from a set of libraries (Chary & Elbaz, 2001; Magdis et al., 2012; Schreiber et al., 2017a) aimed at reproducing dust emission, characterised by the values of infrared luminosity of dust and PAHs (polycyclic

aromatic hydrocarbons) at $\lambda = [8, 1000] \mu\text{m}$, the dust temperature, and the ratio of IR to $8 \mu\text{m}$ luminosity (IR8, [Elbaz et al., 2011](#)) for dust and PAH.

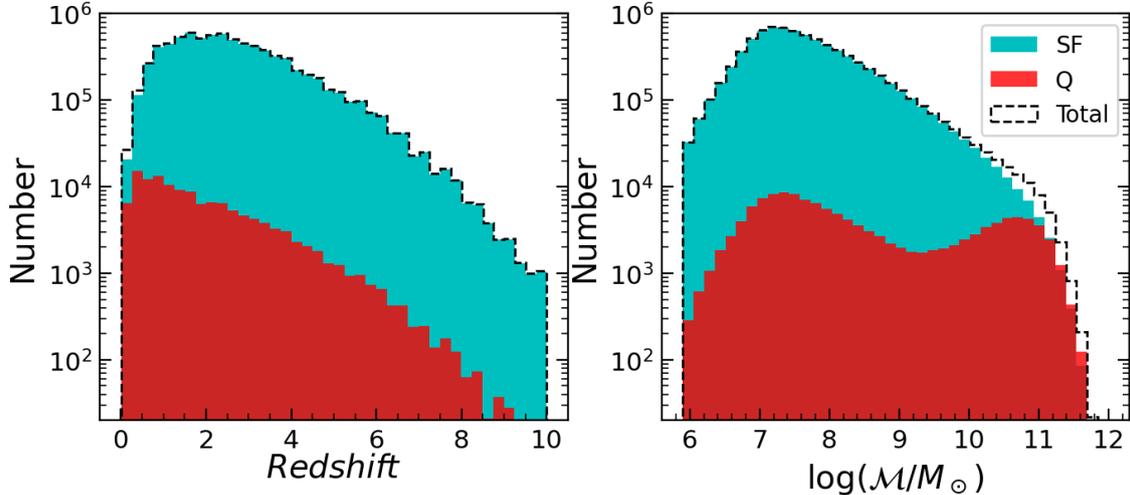


Figure 2.2: Distribution of stellar mass and redshift of all the galaxies in the MAMBO lightcone used in this work, separated into the star forming (blue) and quiescent (red) populations. The dashed black line shows the total (SF+Q) galaxy population.

Afterwards, Gaussian emission lines with a given velocity dispersion¹⁵ is 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 (star-forming or quiescent) of each galaxy, or produce these quantities 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 did to the code are given in Sect. 2.2.4.

It is worth stressing that the full MAMBO workflow can be applied to any DM simulation, and, similarly, the method presented in Sect. 2.2 to populate galaxies with AGN can be applied to any mock galaxy catalogue containing information about galaxy stellar mass, redshift and galaxy type.

¹⁵We modified the original recipe in EGG with a mass-dependent σ_{gas} from [Bezanson et al. \(2018\)](#).

2.2 Building the empirical AGN 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 flag each galaxy from the lightcone as hosting an AGN or not, following a probabilistic method which depends on \mathcal{M} and z (Sect. 2.2.1); *ii*) every object flagged as AGN is assigned an intrinsic X-ray luminosity (Sect. 2.2.2); *iii*) we separate the AGN population into optically unobscured/obscured (type 1/type 2 respectively) by means of their intrinsic X-ray luminosity (Sect. 2.2.3); *iv*) we build 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 SED of typical QSO (that model both the continuum and emission lines; Sect. 2.2.4). A schematic view of this workflow is given in Fig. 2.1.

2.2.1 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 to be hosting an active nucleus, $p(\text{AGN} | \mathcal{M}, z)$, we make 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 $F_X \sim 3 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$. This sample was also selected with a stellar mass limit of $\mathcal{M} > 10^{9.5} M_\odot$ and specific accretion rate $\lambda_{\text{SAR}} > 10^{32} \text{ erg s}^{-1} M_\odot^{-1}$.

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 λ_{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_{H} . This was done considering column density values in the range $20 < \log N_{\text{H}}/\text{cm}^{-2} < 24$, and therefore not including Compton-thick AGN, i.e., heavily obscured objects with $\log N_{\text{H}}/\text{cm}^{-2} > 24$. 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 $f_{\text{CTK}} \sim 35\%$, independent of redshift and accretion luminosity, while Pou-

liasis et al., 2024 found a much lower fraction of $f_{\text{CTK}} \sim 17\%$ 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., 2017b). Therefore, the inclusion of these sources is left to future work.

It is worth noting that the choice of a minimum value of λ_{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. An alternative approach, very common in the literature, is to select sources emitting above a threshold intrinsic X-ray luminosity (generally $L_{\text{X}} > 10^{42} \text{ erg s}^{-1}$, e.g. Brandt & Alexander, 2015). 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. 2.2.2), 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 stellar mass and specific accretion rate, $\Psi(\mathcal{M}, \lambda_{\text{SAR}}, z)$. As a result of this approach, the HGMF cannot be expressed as a simple analytic function, but instead, the authors provide an analytic approximation by performing a least-squares fit to the HGMF with a Schechter (Schechter, 1976) function:

$$\Phi(\mathcal{M}) \text{ d}\mathcal{M} = \Phi^* \left(\frac{\mathcal{M}}{\mathcal{M}^*} \right)^\alpha \exp\left(-\frac{\mathcal{M}}{\mathcal{M}^*}\right) \text{ d}\left(\frac{\mathcal{M}}{\mathcal{M}^*}\right), \quad (11)$$

evaluated at the centre of 3 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 B16.¹⁶

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 visible in Fig. 2.2 (see Appendix A for an example using the original slopes). Therefore, we re-derived these quantities by fitting the $1/V_{\text{max}}$ points shown in B16 with a Schechter function using Eq. 11,

¹⁶B16 used a different definition for the Schechter function, therefore to use the values given in their Table 2 one must remove the term $\text{d}\left(\frac{\mathcal{M}}{\mathcal{M}^*}\right)$ from Eq. 11.

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 didn't modify it.

We then calculate $p(\text{AGN} | \mathcal{M}, z)$ 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 interpolate and extrapolate $p(\text{AGN} | \mathcal{M}, z)$ from the centre of each redshift bin. For this, we assume a minimum fraction $p(\text{AGN} | z = 0) = 0.01$ at all \mathcal{M} when extrapolating at $z < 0.55$, while we maintain a constant fraction when extrapolating at $z > 2$. Afterwards, every galaxy is statistically assigned as hosting an AGN or not with a Bernoulli trial proportional to $p(\text{AGN})$ (i.e., by comparing $p(\text{AGN})$ 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 ($< 120\text{Mpc}$) galaxies from the Palomar sample and classified 14 (6.6%) of them as Seyferts, while only 4 (1.9%) of them have $L_X > 10^{42}$. 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 shown in Fig. 3.2.

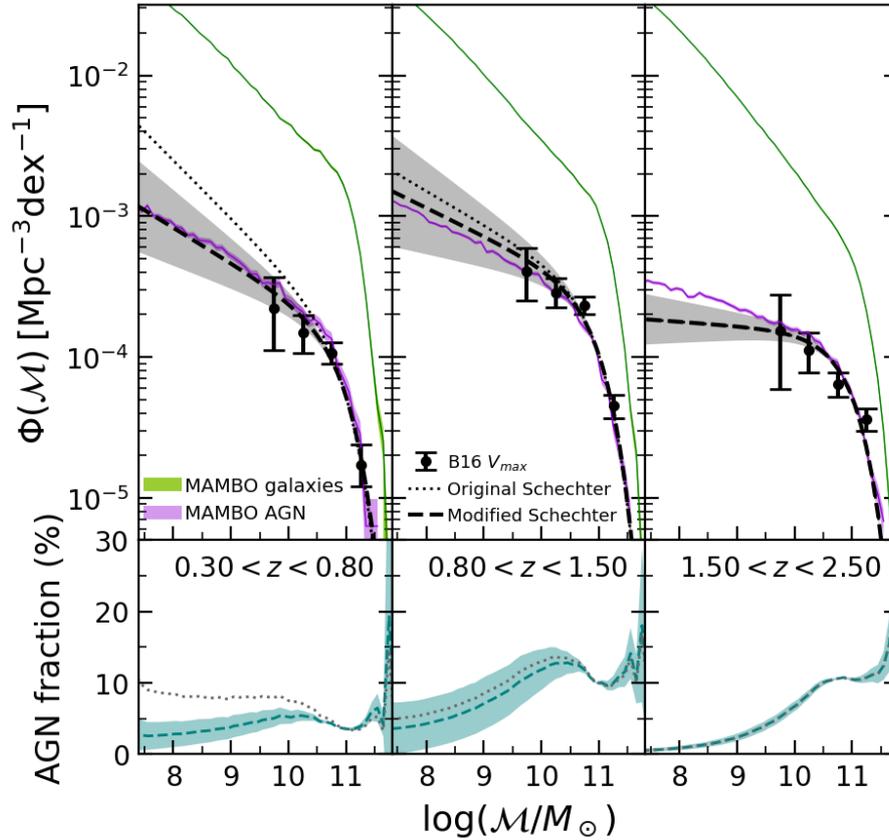


Figure 2.3: *Upper panels:* Stellar mass function of galaxies and AGN. The green and purple shaded areas correspond to the galaxies and AGN from the MAMBO lightcone respectively. We also show the best-fit Schechter function to the host galaxy mass function (HGMF) of AGN from [Bongiorno et al. \(2016\)](#) with a black dotted line, and the modified Schechter fit used in this work with a black dashed line. The shaded grey area shows the uncertainty in α . Black dots with errorbars show the AGN mass function computed using the V_{max} method, derived in B16. The three panels correspond to the three redshift bins studied in B16. *Lower panels:* AGN fraction derived as the ratio of the modified Schechter function and the galaxy mass function from MAMBO (blue dashed line), and the same but using the original Schechter function from B16 (black dotted line).

In the upper panels of Fig. 2.3 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_{max} method, derived in B16, as an additional consistency check to the Schechter model. As 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_{max}

estimate, due to the fact that we kept a constant $p(\text{AGN} | \mathcal{M}, z)$ when extrapolating at $z > 2$ (see Fig. 2.4). In the lower panels of Fig. 2.3 we show the fraction of AGN over the galaxy population as a function of stellar mass and at a given redshift bin, $p(\text{AGN} | \mathcal{M}, \langle z \rangle)$. As expected, this fraction increases with increasing stellar mass of the host galaxy.

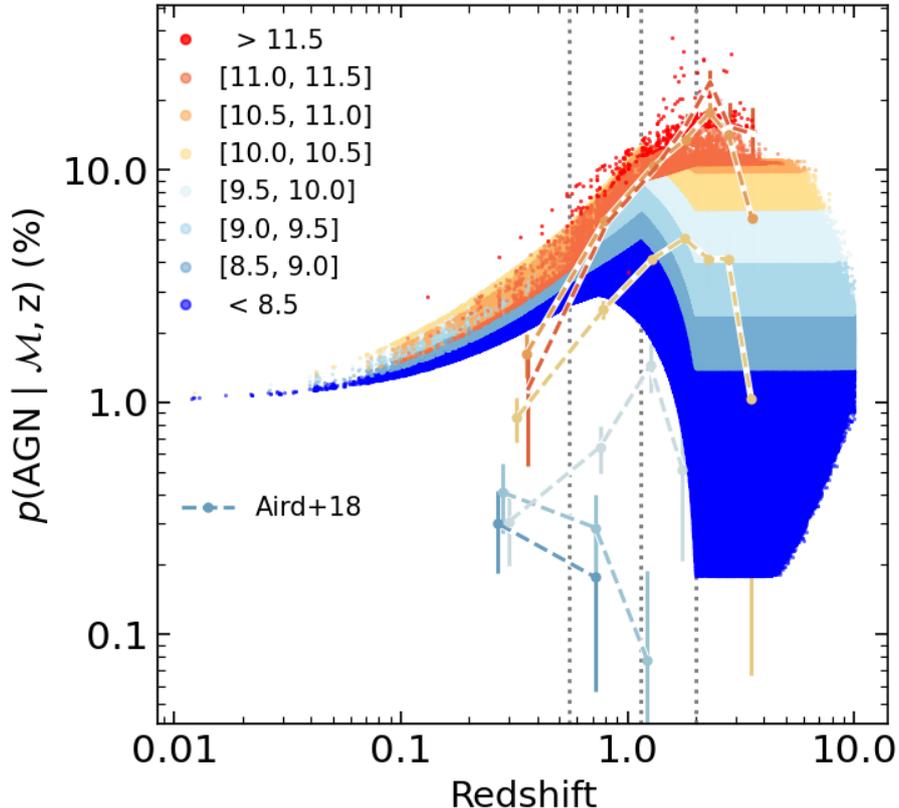


Figure 2.4: Probability of a galaxy hosting an AGN as a function of redshift. Each point corresponds to a galaxy in the MAMBO lightcone, colour-coded in different \mathcal{M} bins indicated in the legend. For comparison, the points with errorbars connected by dashed lines show the AGN fraction derived in A18, also in different mass bins. Vertical dotted lines mark the centre of the three redshift bins studied in B16, namely $z = 0.55, 1.15, 2.0$.

We show in Fig. 2.4 the probability of a given galaxy to be 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 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. 2.4 shows also the fraction of AGN as derived in Aird et al. (2018, hereafter A18), where 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 which tends to be underestimated when comparing it to the observed ones, especially for $z \lesssim 1$ and $z \gtrsim 3$.

2.2.2 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 assign a specific accretion rate. At $0 < z < 2$ we make use of the SARDF derived in B16. As explained in Sect. 2.2.1, B16 derived the SARDF and the HGMF simultaneously as a bivariate distribution function of \mathcal{M} and λ_{SAR} , that is, $\Psi(\mathcal{M}, \lambda_{\text{SAR}}, z)$. 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:

$$\Phi(\lambda_{\text{SAR}}, \mathcal{M}) = \frac{\Phi_{\lambda}^*}{\left(\frac{\lambda_{\text{SAR}}}{\lambda_{\text{SAR}}^*(\mathcal{M})}\right)^{-\gamma_1} + \left(\frac{\lambda_{\text{SAR}}}{\lambda_{\text{SAR}}^*(\mathcal{M})}\right)^{-\gamma_2}}, \quad (12)$$

where the mass dependence is given by $\log \lambda_{\text{SAR}}^*(\mathcal{M}) = \log \lambda_{\text{SAR},0}^* + k_{\lambda}(\log \mathcal{M} - \log \mathcal{M}_0)$, where $\log \lambda_{\text{SAR},0}^* = 33.8$, $\log \mathcal{M}_0 = 11$ and $k_{\lambda} = 0.58$. The best-fit values of the normalisation Φ_{λ}^* 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. 12 we compute the SARDF at 5 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. 2.5). We then divide the AGN in our lightcone into 5 mass bins, each of them centred at one of the above-mentioned \mathcal{M} and all of them of width 0.5 dex in \mathcal{M} , except the lowest mass bin, extending up to the lowest \mathcal{M} in our lightcone. Then, we normalise the λ_{SAR} distribution functions by dividing them by the integrated density of AGN (Mpc^{-3}) in each bin of \mathcal{M} and z , therefore converting these distributions into probability distributions (PDF). We assign to each AGN a value of λ_{SAR} by randomly extracting values from the corresponding PDF at each \mathcal{M} and z bin.

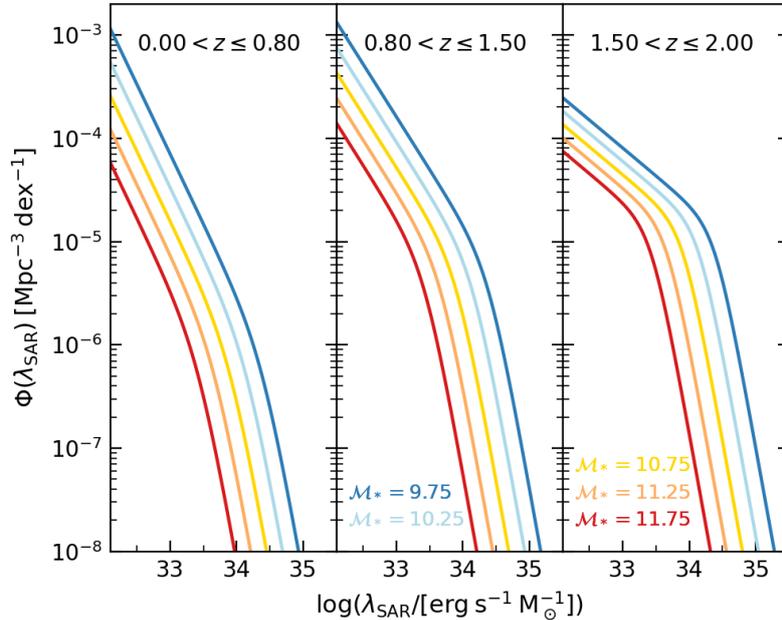


Figure 2.5: Specific accretion rate distribution function (SARDF) computed at different values of \mathcal{M} using Eq. 12, in the three redshift bins from B16.

At higher redshifts ($z > 2$), instead, we make use of the accretion rate distributions from A18, who studied a sample of 1,797 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 $p(\log \lambda_{\text{sBHAR}} | \mathcal{M}, z)$ distributions in the range $-2 < \log \lambda_{\text{sBHAR}} < 1$ and $2 < z < 4$, and use it to assign a value of λ_{sBHAR} to each object previously labeled as AGN at $z > 2$. Since A18 calculated the λ_{sBHAR} distributions separately for star-forming and quiescent galaxies, we use both sets of distributions for the galaxies in our catalogue that are split into these two classes. After assigning every AGN in the lightcone with a specific accretion rate, we convert this into intrinsic X-ray luminosity (in the 2 – 10 keV band).

We also explored the possibility of using the accretion rate distributions from A18 to assign L_X at all redshifts (see Appendix A), but we decided to use the methodology presented in this section since it produces an XLF which is in better agreement with the observed ones.

2.2.3 Obscuration model

The observed emission of AGN can vary significantly depending on the level of obscuration of the nucleus by its surrounding material along the line of sight. Moreover, this effect is wavelength dependent, since emission at different wavelengths interact differently with the obscuring material (see e.g. Fig. 1.6). 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 emission lines (FWHM $> 1000 \text{ km s}^{-1}$) and a very blue continuum emission, while type 2 AGN show only narrow emission lines (with typical FWHMs of a few hundreds km s^{-1}) and their optical continuum emission tends to be dominated by their host galaxy (see Fig. 1.2). 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 $F_X = 2 \times 10^{-15} \text{ erg cm}^{-2} \text{ s}^{-1}$ 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

$$F_{\text{obs}} = A + \frac{1}{\pi} \text{atan} \left(\frac{l_0 - \log L_X}{\sigma_x} \right), \quad (13)$$

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 as type 1 and type 2.

Additionally, the X-ray emission from AGN (especially in the soft band) can be heavily obscured by the 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 $20 < \log N_{\text{H}}/\text{cm}^{-2} < 24$. The choice of this range is consistent with the fact that, as stated in Sect. 2.2.1, Compton-thick AGN ($\log N_{\text{H}}/\text{cm}^{-2} > 24$) are not included in our model. Following Section 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 $\log N_{\text{H}}/\text{cm}^{-2} < 22$, and vice-versa.

It is worth noting that in nature these two classifications, although on average correlate with each other, do not match perfectly, i.e., some objects present optical obscuration but not X-ray obscuration and vice-versa (e.g. Merloni et al., 2014; Marchesi et al., 2016b). Therefore, the classification of AGN into type 1 and 2 depends on the wavelength range of interest. For the rest of this work, unless specifically stated otherwise, we use the terms type 1 and type 2 AGN to refer to their optical/UV classification, regardless of their X-ray obscuration.

In the left panel of Fig. 2.6 we show the fraction of optically obscured AGN as a function of L_{X} for the presented lightcone, and compare it with the calibration used to derive it, that is, Eq. 13. We also show the fraction of objects that are classified as obscured in the X-ray, that is, sources with $\log N_{\text{H}}/\text{cm}^{-2} > 22$. Additionally, the figure shows the fraction of sources that are obscured in both bands, and that of sources obscured in at least one of the two bands.

As mentioned above, the absorption function presented in Ueda et al. (2014) is a function of L_{X} and redshift. Therefore, we show in the right panel of Fig. 2.6 the redshift evolution of the X-ray obscured fraction. For this, we show this fraction as a function of L_{X} for different redshift bins: for the full lightcone ($0 < z < 10$, as shown also in the left panel), and at redshifts bins centred around $z = 0.5$ and $z = 2.5$ respectively.

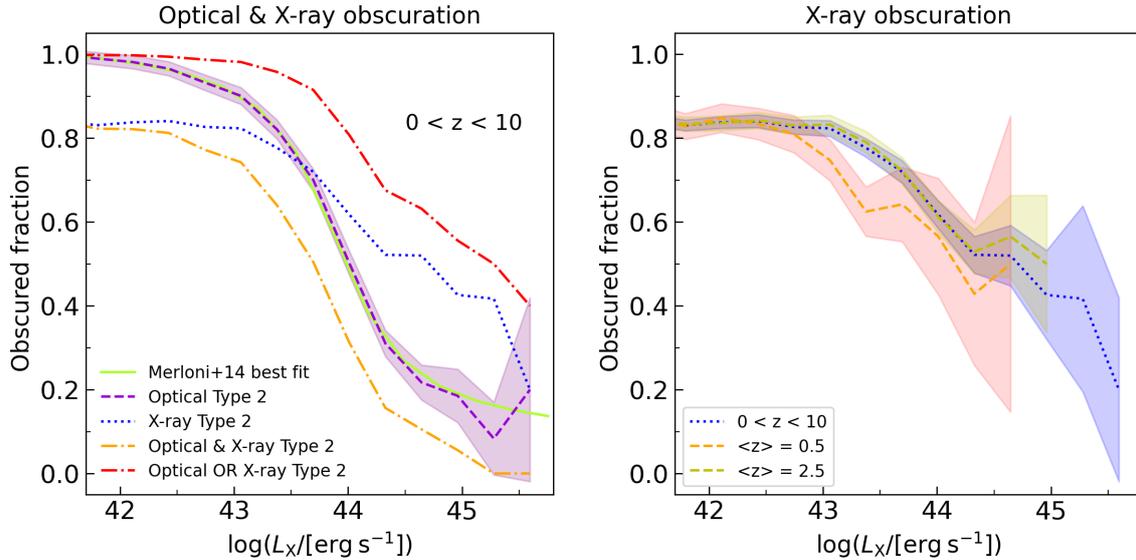


Figure 2.6: Fraction of obscured (type 2) AGN as a function of intrinsic X-ray luminosity. *Left panel:* The green solid line shows the relation used to separate AGN (Eq. 13, originally from Merloni et al., 2014) into optically obscured/unobscured, while the purple dashed line shows the actual fraction of type 2 AGN in our catalogue (with the Poissonian uncertainty shown by the shaded area). The blue dotted line shows the fraction of objects classified as obscured in the X-ray. The other two dot-dashed lines show the fraction of sources that are obscured in both bands, and that of sources obscured in at least one of the two bands. *Right panel:* Fraction of X-ray obscured sources studied in different redshift bins.

Additionally, we show in Fig. 2.7 the histogram of the assigned L_X in different redshift bins for the entire population of AGN, separated into type 1 and type 2 AGN. We note that, as visible from this figure, a large fraction of AGN in the lightcone have luminosities $L_X < 10^{42} \text{ erg s}^{-1}$, that is, below a classical threshold applied to define an X-ray emitter as an AGN. The reason for this is twofold: on one hand, as previously mentioned, we started from a sample of AGN selected above a given accretion rate limit and not above a L_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. It is also visible from this figure that most of these AGN with $L_X < 10^{42} \text{ erg s}^{-1}$ are classified as type 2 AGN in our catalogue, as a consequence of the obscuration model applied (Eq. 13, see also Fig. 2.6).

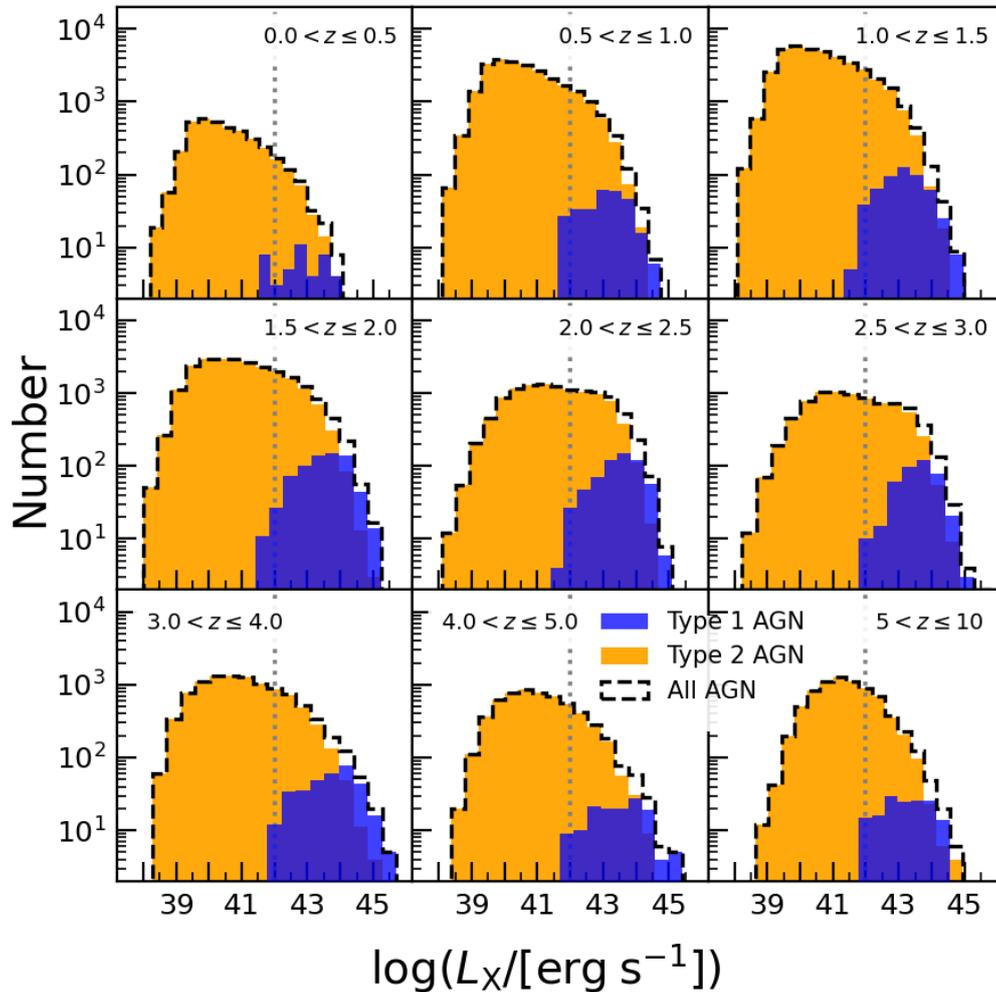


Figure 2.7: Distribution of intrinsic hard-band X-ray luminosity (L_X) of all the AGN in the lightcone, separated into type 1 (blue) and type 2 (orange) AGN, in different redshift bins. We also show with dashed and solid lines the subpopulations of type 1 and 2 in star forming and quiescent galaxies. The vertical dotted line marks $L_X = 10^{42} \text{ erg s}^{-1}$, the threshold typically applied to separate X-ray emission from AGN or from other origins.

In Fig. 2.8 we show the distribution of the AGN in our mock in the $L_X - \mathcal{M}$ plane. As expected by construction, all the AGN have accretion rates above $\lambda_{\text{SBHAR}} > 0.01$ ($\equiv \lambda_{\text{SAR}} > 10^{32} \text{ erg s}^{-1} M_{\odot}^{-1}$), and only a small fraction of them are above the Eddington limit ($\lambda_{\text{SBHAR}} \sim 1$). Additionally, we also show in this figure the distribution of sources from the Chandra COSMOS Legacy Spectral Survey (Marchesi et al., 2016b) in the $L_X - \mathcal{M}$ plane, selected above a minimum X-ray flux $F_X = 1.9 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$, 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. 2.8) 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.

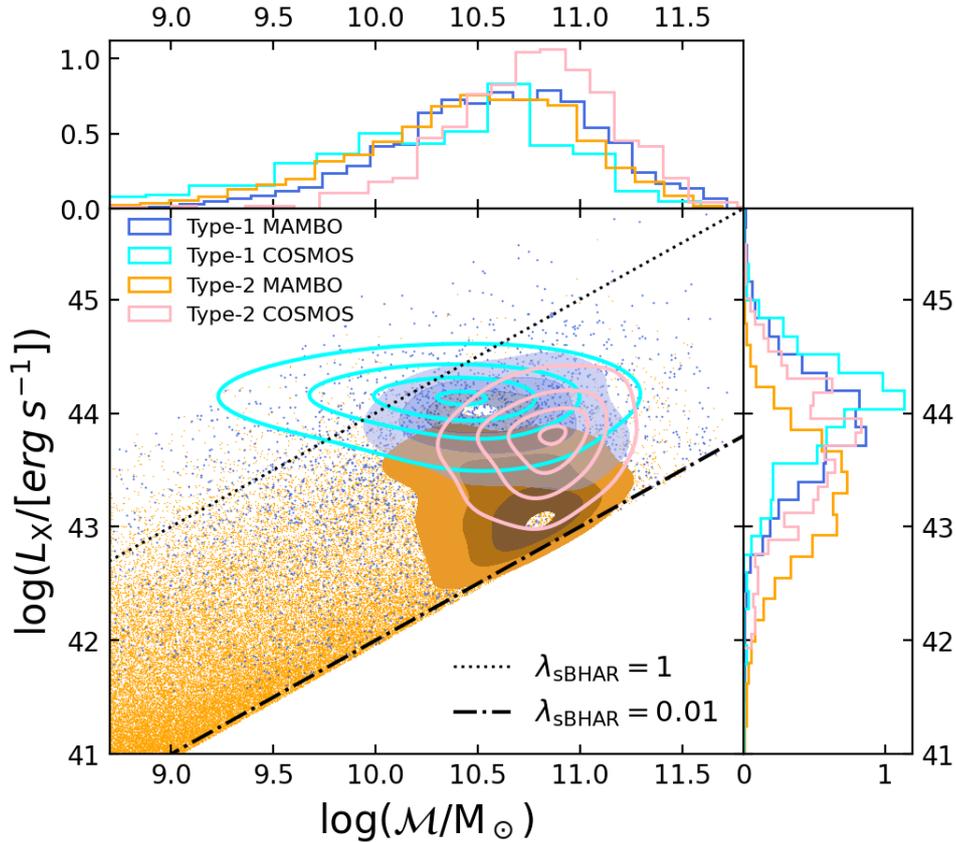


Figure 2.8: AGN from our lightcone scattered in the $L_X - \mathcal{M}$ plane, separated into type 1 (blue) and type 2 (orange) AGN. The points in the central panel show the distribution of all the sources, while the solid contours and the histograms from the upper and right panel correspond to sources selected above $F_X = 1.9 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$. 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 $k_{bol} = 25$ and a constant mass ratio of black hole to host galaxy $M_{BH} \approx 0.002\mathcal{M}$ correspond approximately to 1 and 100 percent of the Eddington limit respectively.

Finally, we show in Fig. 2.9 the spatial distribution of the sources in the lightcone here presented, separated into galaxies, type 1 AGN and type 2 AGN, for a given redshift bin.

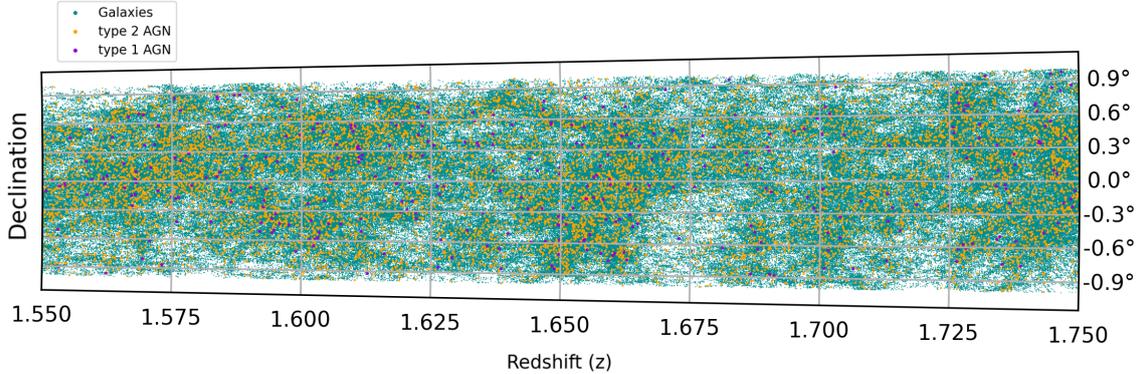


Figure 2.9: Spatial distribution of galaxies (green), type 1 AGN (purple) and type 2 AGN (orange) in the redshift-Dec plane.

2.2.4 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. 2017b](#)) 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 (star-forming 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.

Type 1 AGN SEDs

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, that is, UV to NIR. These synthetic SEDs have been shown to accurately reproduce, to a high degree of accuracy, the observed-frame optical and near-infrared 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 model was calibrated on a large cross-matched sample from the SDSS DR16 quasar catalogue to UKIDSS and *WISE* catalogues, with photometric data covering the *ugrizYJHKW12* bands.

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 ultraviolet lines, such as C IV_{λ1549}. 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. Specifically, this parameter ranges from -2 to +3, which correspond, respectively, to spectra with weak, highly blueshifted lines and to those with strong, symmetric lines. For intermediate values, the emission line properties are modelled by interpolating between these two extreme SEDs. We show in [Fig. 2.10](#) three SEDs generated with different `emline_type`, and fixing other parameters.

The Baldwin effect is parametrised in the model by relating this parameter to the intrinsic absolute magnitude M_i of the quasar by means of the following empirical equation:

$$\text{emline_type} = 0.183 \times (M_i + 27). \quad (14)$$

For this work, we constructed the QSO SED library by creating a set of templates, for which we changed only this parameter between its minimum and maximum values (-2 and +3) in increments of 0.1. Additionally, we modified the code EGG so that, for each type 1 AGN, its SED is rescaled to the L_{UV} (at 2500 Å) that corresponds to that object. We calculated this luminosity according to the observed $L_X - L_{\text{UV}}$ relation, as presented in [Bisogni et al. \(2021\)](#). This relation is parametrised as:

$$\log(L_X) = \gamma \log(L_{\text{UV}}) + \beta, \quad (15)$$

where the proxies for the UV and X-ray emissions correspond to the monochromatic rest-frame 2500 Å and 2 keV luminosities 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 the hard band unabsorbed X-ray luminosity 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:

$$L_E = \frac{(2 - \Gamma)E^{1-\Gamma}}{E_2^{2-\Gamma} - E_1^{2-\Gamma}} L_{E_1-E_2}, \quad (16)$$

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. 15 and 16 we derive the monochromatic luminosity L_{2500} from the full band luminosity L_X , and use it to rescale the SED.

Once we calculate the intrinsic L_{UV} of each QSO SED, we can also determine their intrinsic absolute magnitude M_i , by rescaling the template's M_i with the same factor as L_{UV} (or technically, M_{UV}). This requires the assumption that all templates share the same M_i , meaning that the emission line properties have minimal impact on the total magnitude. We checked this assumption by measuring M_i on all the templates, and we found an average value of $M_i = -25.77 \pm 0.07$. To verify this, we measured M_i for all templates and found an average value of $M_i = -25.77 \pm 0.07$. Given the small standard deviation, we adopted this value as the exact M_i for each template.

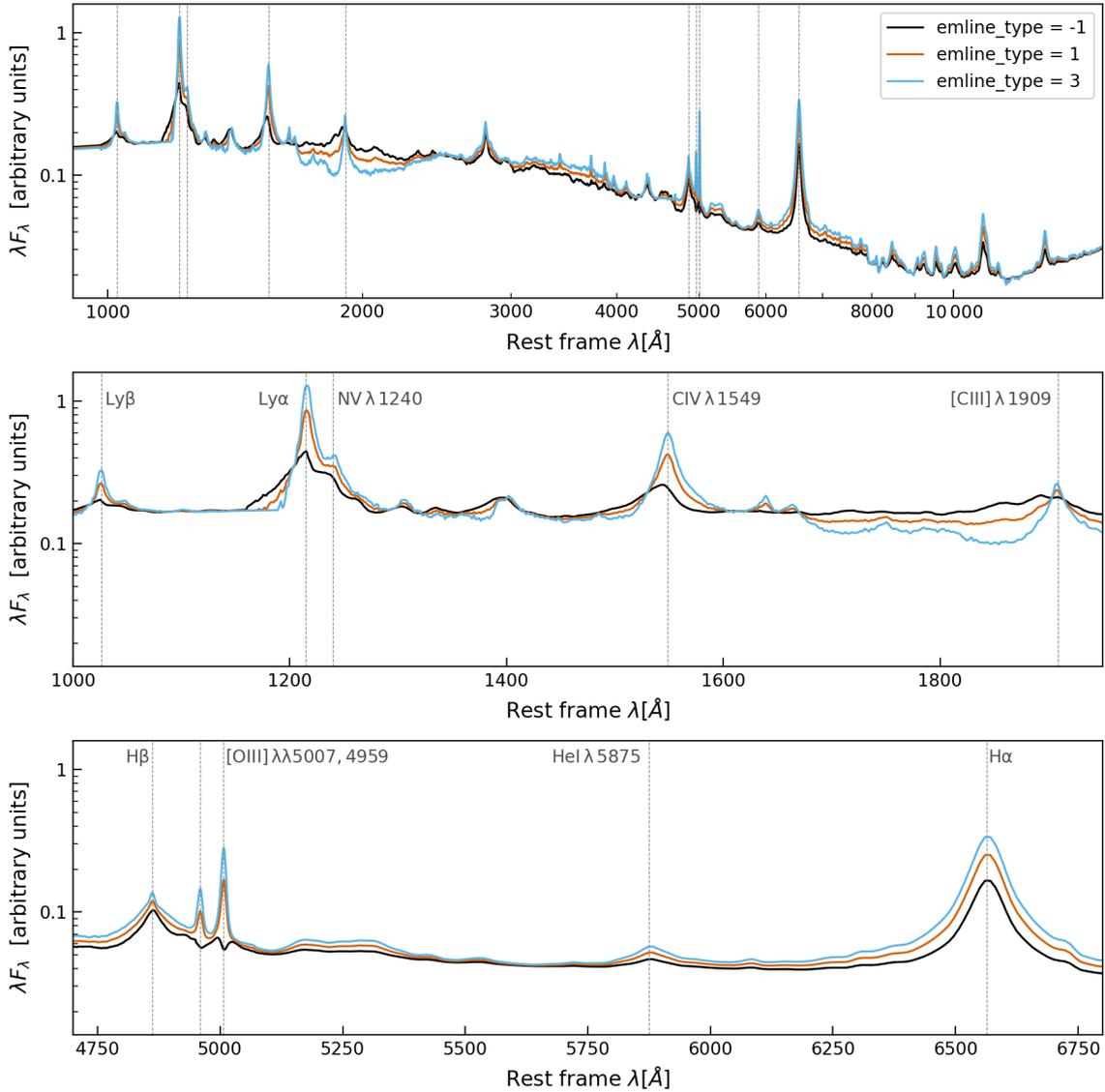


Figure 2.10: Three different templates from the QSO SED library constructed for this work. The templates were produced using the parametric model developed by Temple et al. (2021). As it is visible from the figure, the template with `emline_type` = -1 (black line) presents weak, highly blueshifted lines, while the one with `emline_type` = 3 (blue line) shows stronger and more symmetric lines. The spectrum generated with `emline_type` = 1 (orange line) corresponds to an interpolation between the other two extreme cases. We chose to show these three cases since they represent approximately the extrema and median value of the distribution of `emline_type` of the QSO in the lighcone here presented (see Fig. 2.11). *Upper panel:* full SED from the UV to the NIR, where we marked the lines shown in the other two panels. *Middle panel:* Zoom on the UV region, where we marked the centre of some of the most important lines and their names, as in the next panel. *Lower panel:* Zoom to some optical lines.

After this, for each type 1 AGN, EGG selects an SED from the pre-built library of QSO SEDs following Eq. 14. We show in Fig. 2.11 the distribution of the assigned `emline_type` for the type 1 AGN of the Millennium lightcone presented in this work.

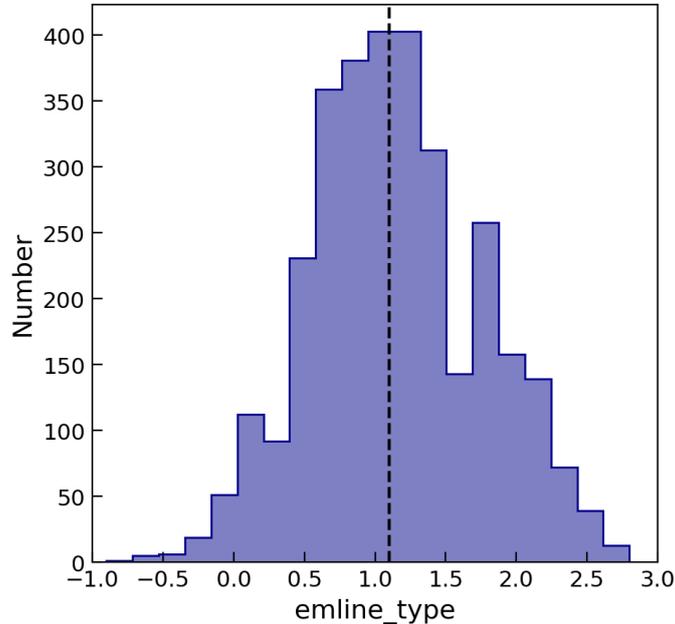


Figure 2.11: Distribution of the assigned `emline_type` for the type 1 AGN of the lightcone presented in this work. The vertical dashed line shows the median value.

In the next step, 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 $\lambda \gtrsim 1700 \text{ \AA}$, while it increases less rapidly for shorter wavelengths. The $E(B-V)$ of each source is chosen randomly from a distribution which we parameterised based on the observational distribution presented in Fig. 2 of Lusso et al. (2013), which peaks at low $E(B-V)$, with a median value $\langle E(B-V) \rangle = 0.03$. The final A_V distribution of the type 1 AGN presented in this work is shown in Fig. 2.12

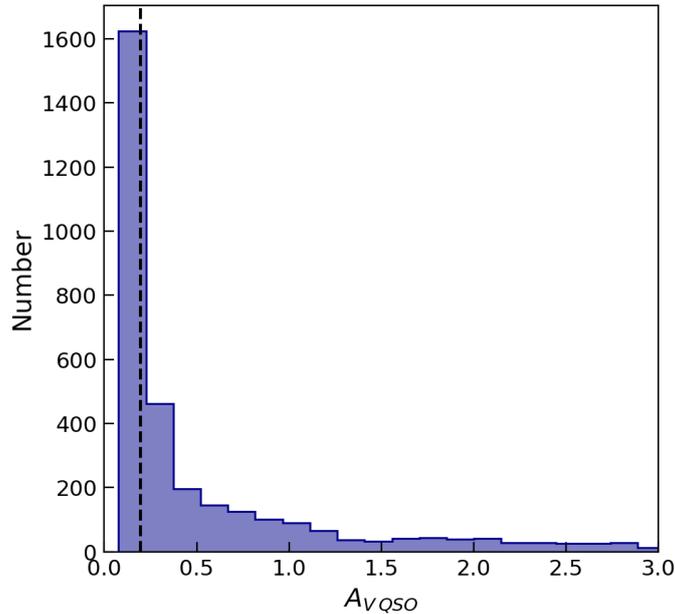


Figure 2.12: Distribution of the assigned A_V for the type 1 AGN of the lightcone presented in this work, which is based on the observed Fig. 2 of [Lusso et al. \(2013\)](#). The vertical dashed line shows the median value.

Finally, the host 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

$$f_{AGN_{band}} = \frac{F_{band}^{AGN}}{F_{band}}, \quad (17)$$

where F_{band}^{AGN} 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, M_{FUV} (which corresponds to the wavelength range where AGN emission typically peaks, and therefore gives a good estimation of the intrinsic dominance of the AGN component), and using the observed NIR magnitude m_H (more

sensitive to the galaxy component, and also more related to a directly observable quantity, since we use an observed magnitude instead of the intrinsic one). Both of these filters are described in Table 1.

We show in the left panel of Fig. 2.13 the distribution of these two quantities for the type 1 AGN of our lightcone. It is visible that the distributions are bimodal, with most objects having values of f_{AGN} near 0 or 1 (especially for the FUV band). The reason for this lies in the fact that the distribution of M_{FUV} for the AGN component is much wider than that of the host galaxy component. As a consequence of this, there is a large fraction of sources for which either the AGN or the host galaxy component totally dominates, and therefore f_{AGN} is near 0 or 1. This is also visible from the right panel of Fig. 2.13.

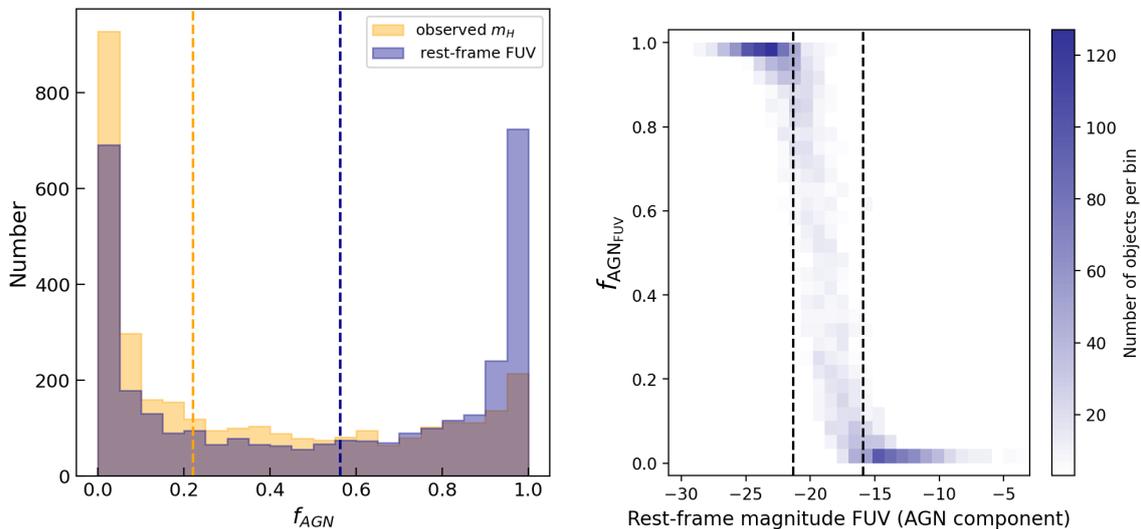


Figure 2.13: *Left panel:* Distribution of f_{AGN} , that is, the contribution of the AGN component with respect to the total emission of the source, in two different bands. The blue histogram shows $f_{\text{AGN}_{\text{FUV}}}$ as defined in Eq. 17, from the rest-frame magnitude GALEX FUV. The orange histogram shows the equivalent quantity using the observed NIR magnitude m_H . Each vertical dashed line shows the corresponding median value of each distribution. *Right panel:* Density plot of f_{AGN} (defined in the FUV band) as a function of rest-frame FUV magnitude (from the AGN component only). The vertical dashed lines show the 5% and 95% percentiles of the M_{FUV} (from the host-galaxy component only) distribution.

In Fig. 2.14, we show some examples of the rest-frame SED of type 1 AGN from the lightcone. We have decided to show three representative examples. The first one is a source for which the AGN component completely dominates the UV and NIR emission and has a clear contribution to the optical emission through continuum and

broad-line emission. The second source shows AGN contribution to the UV emission, while star formation dominates the rest of the spectrum. In both of these sources the AGN component is only mildly attenuated, as it is visible also from their A_V^{QSO} , shown also in Fig. 2.14. Instead, the third source that we show presents a very attenuated QSO component (with $A_V^{QSO} = 1.39$), which nevertheless dominates over the host galaxy emission. Such objects are often known as red QSOs (see e.g. Richards et al. 2003). We note that these very reddened QSOs are rare in our catalogue, as it can be seen from the A_V^{QSO} distribution shown in Fig 2.12.

In Fig. 2.14 we show also the value of the most relevant physical parameter related to the construction of the SED, such as \mathcal{M} , L_X , SFR, A_V and $f_{AGN_{FUV}}$ and f_{AGN_H} .

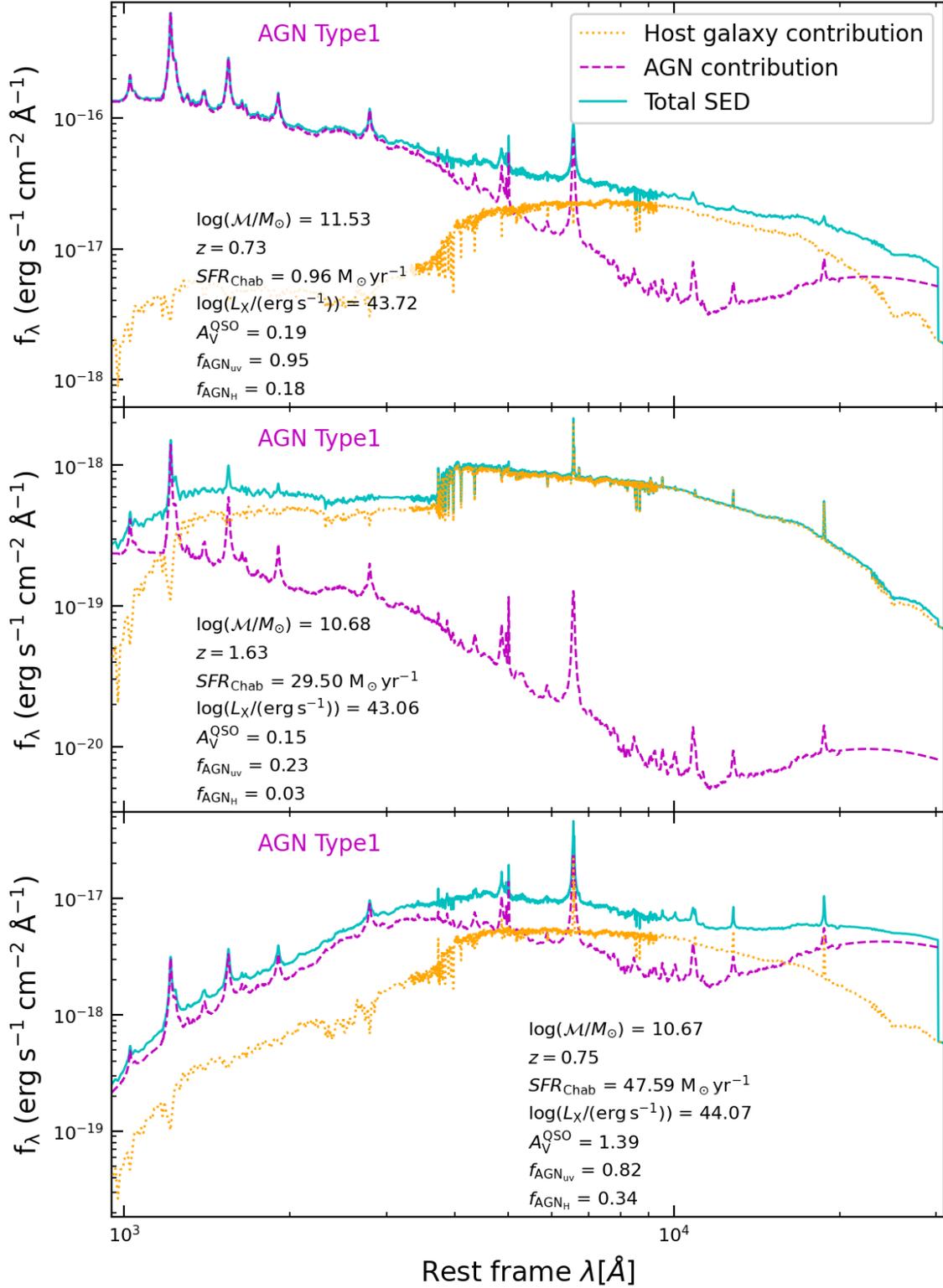


Figure 2.14: Some examples of SEDs of type 1 AGN from our lightcone. We show three representative cases: one source mainly dominated by AGN emission, one dominated by the host-galaxy emission, and a third one dominated by highly attenuated AGN emission. In all cases, the galaxy component (bulge + disc) is shown with an orange dotted line and is based on [Bruzual & Charlot \(2003\)](#) models. The AGN component is shown with a dashed violet line. The cyan line shows the composed SED. Dust absorption is applied in all cases.

As a validation test for this methodology, we show in Fig. 2.15 the relation between the output quantities `emline_type` and the attenuated magnitude M_i . As it is visible from this figure, when considering the emission only from the QSO component of the SED, all of the points have a M_i magnitude which follows the relation shown in Eq. 14, or fainter. This is due to the attenuation that we applied to the QSO component. Instead, when considering the total emission (QSO + host), we see that some sources have M_i magnitudes brighter than the one given by Eq. 14. These are sources for which the host dominates over the AGN component at the photometric band i .

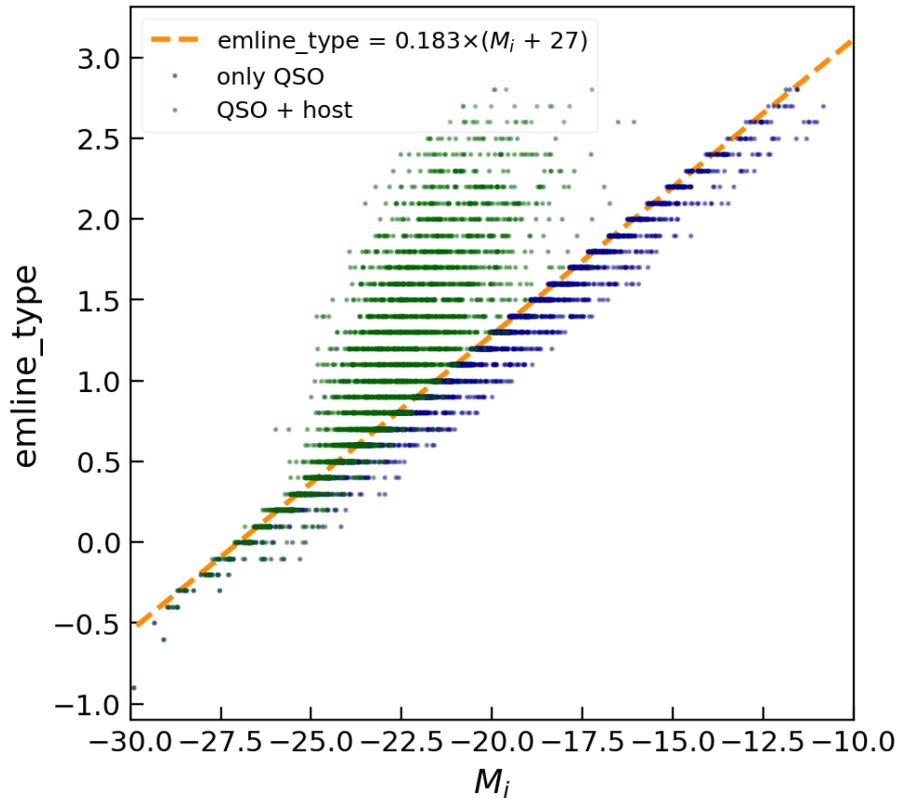


Figure 2.15: Relation between the output quantities `emline_type` and the attenuated magnitude M_i . We show with blue dots the magnitude coming only from the AGN component, and with green dots the total contribution from the AGN and the host galaxy. The dashed orange line shows the relation shown in Eq. 14.

Measuring AGN line properties

The type 1 AGN templates that we used to construct our library do not include information about the emission line properties, like flux, equivalent width (EW) or

FWHM. Therefore, we measured these quantities using the public code PyQSOFit (Python QSO fitting code, Guo et al., 2018; Shen et al., 2019). PyQSOFit is a Python code designed to measure spectral properties of SDSS quasars. The code decomposes the input SED into different components (the host galaxy contribution, the QSO power-law continuum, the Fe II component, and broad and narrow emission lines) and provides a fit for each component.

We measured with PyQSOFit the continuum and spectral properties of all the templates from our QSO library. We show in Fig 2.16 one example of the fit to the SED of one of these templates, namely the one with `emline.type = 1`. In the figure we show the different components fitted by the code. Because the code needs the error of the flux as an additional input, we defined the error spectrum as one percent of the original flux. Overall we see that the code provides a good fit to the input spectrum, specially for the most important broad lines, which are our main target to be fitted.

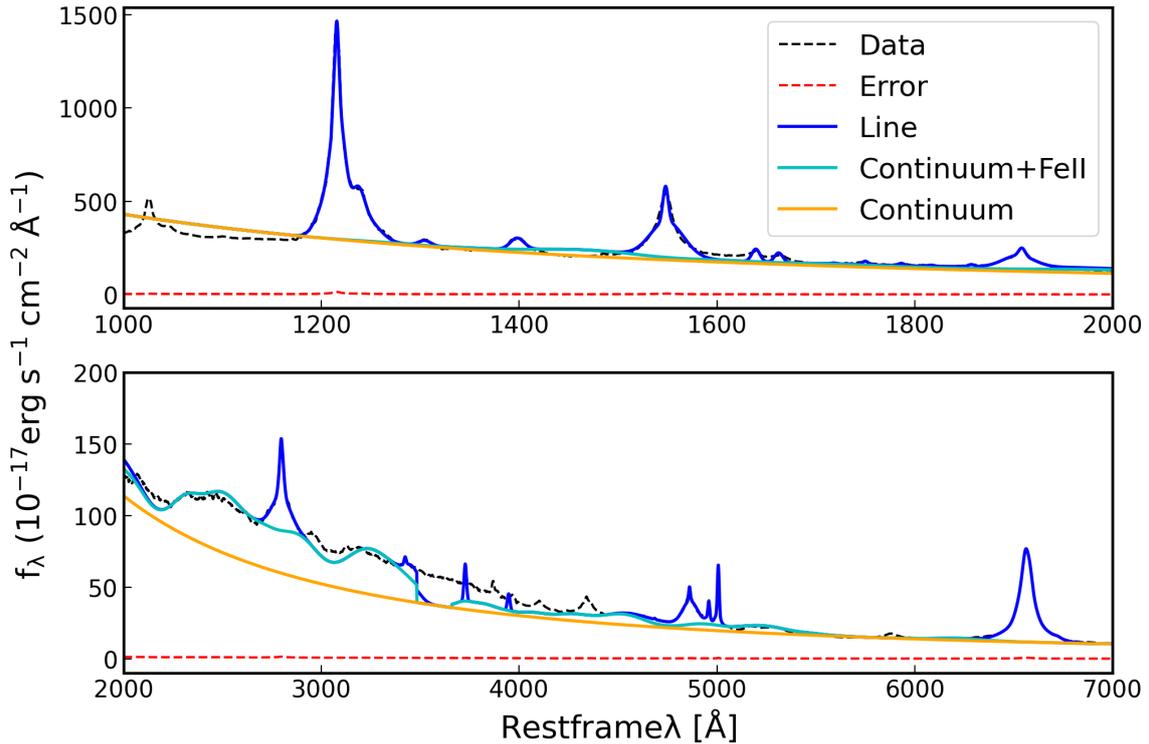


Figure 2.16: Fit to the SED of one QSO from the library presented in sec. 2.2.4, performed with the Python code PyQSOFit. The grey line shows the input flux, and the red line its associated error. The rest of the lines show the different components that the code uses to fit the input spectrum. The yellow line represents the power-law continuum, the cyan line the continuum with Fe II component, and the dark blue line the emission line component.

PyQSOFit fits several line complexes, such as $H\alpha$, $H\beta$, $Mg\ II$, $C\ III]$, $C\ IV$ and $Ly\alpha$ by fitting, in each complex, broad and narrow Gaussian profiles to different emission lines. The code returns in output several properties of the emission lines, such as the FWHM, line dispersion, EW, Peak and line flux. We show in Fig. 2.17 a detailed view of the line-fit to one template from our library (again with `emline_type = 1`, as in Fig 2.16). Specifically, we show the fit to 4 different line complexes. The first panel shows the fit to the $C\ IV$ line, which is fitted together with the broad lines $He\ II$ and $O\ III]$, both of them visible in the plot. The second shows $Mg\ II$, which is fitted by two broad components and one narrow component. The $H\beta$ line is fitted together with $He\ II$ and the $[O\ III]$ doublet, with each line fitted by a narrow and a broad component. And finally, $H\alpha$ is fitted by two broad components and one narrow component, plus narrow components for the $[N\ II]$ and $[S\ II]$ doublets.

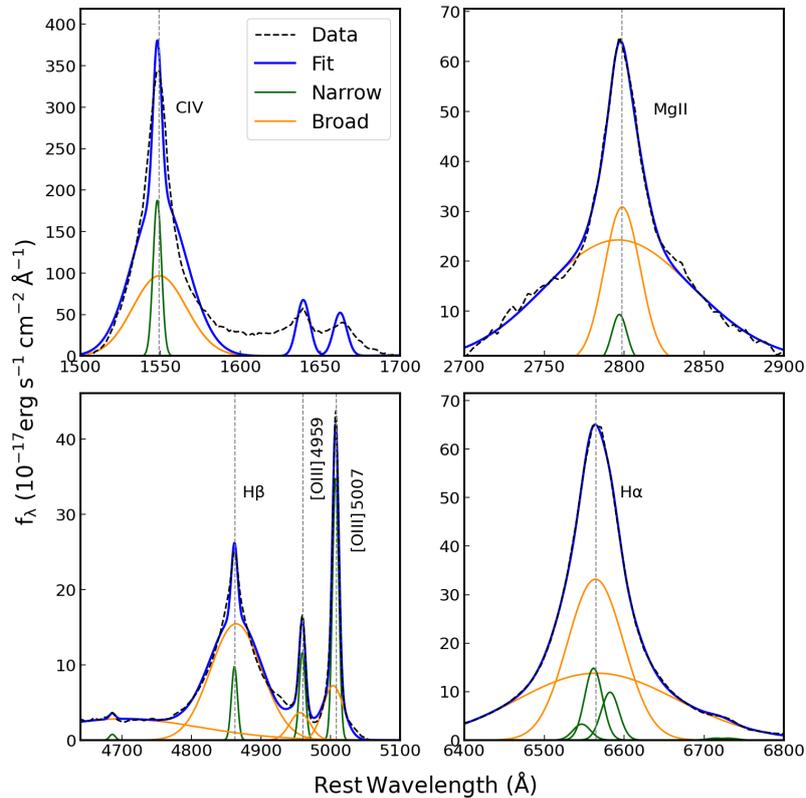


Figure 2.17: Zoom-in to the fit of some emission lines from the SED of one QSO from the library presented in sec. 2.2.4, performed with the Python code PyQSOFit. The black dashed line shows the input flux. The green and orange lines show, respectively, the narrow and broad line fit components. The blue line shows the total fit (continuum + all lines). We marked with vertical dashed lines the centre of some of the most relevant lines.

To give an overview of how the line properties change with `emline_type`, we show in Fig. 2.18 the measured FWHM and line flux for some emission lines as a function of `emline_type`. It is visible from this figure that the library of templates is made in such a way that the $H\alpha$ flux is interpolated linearly between two extreme values. As a sanity check for these values, we plot in Fig. 2.18 (right bottom panel) the ratio of $H\alpha$ to $H\beta$ flux (the so-called Balmer decrement). We observe that the value of this ratio is between 3 and 3.5, which is in agreement with observations of this ratio when performed with the intrinsic fluxes of broad-line AGN (Dong et al., 2007).

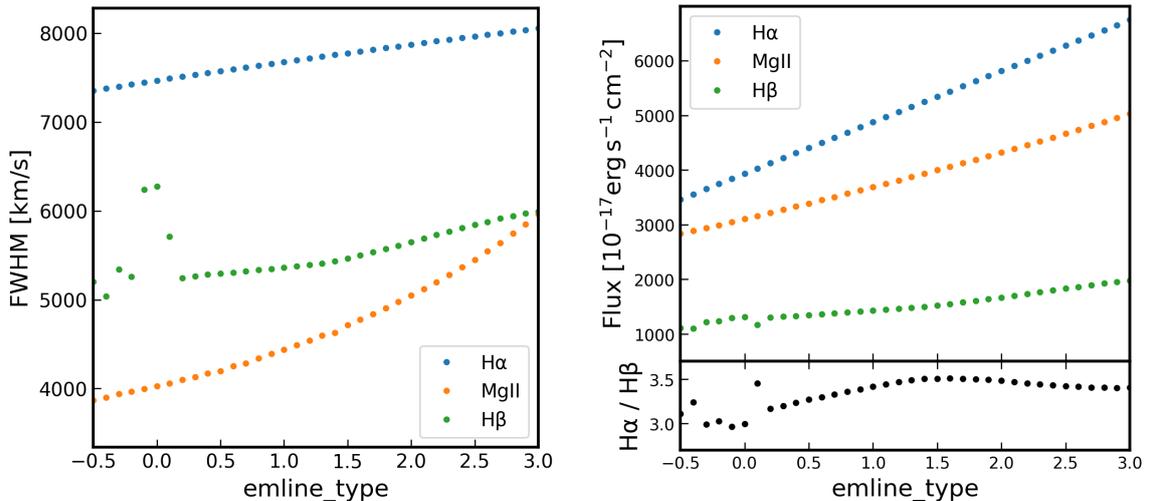


Figure 2.18: Line properties as measured with PyQSOFit on our library of type 1 AGN templates as a function of `emline_type`. The right bottom panel shows the ratio of $H\alpha$ to $H\beta$ fluxes.

Type 2 AGN SEDs

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 which 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 which 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 $0.001 \leq \lambda/\mu\text{m} \leq 0.25$, 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. \(2016\)](#), i.e. -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 ranges¹⁷:

$$\begin{cases} -5 < \log U < -3 & \text{for } 0 < z < 1 \\ -4 < \log U < -2 & \text{for } 1 < z < 2 \\ -3.5 < \log U < -1.5 & \text{for } z > 2, \end{cases} \quad (18)$$

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

¹⁷These values correspond to the volume-averaged ionisation parameter $\langle U \rangle$, as defined in equation 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 $U_S = \frac{4}{9}\langle U \rangle$.

AGN narrow-line ratios is purely linked to that of $\log U$.

Hydrogen number density n_{H} , i.e., volume-averaged hydrogen density of the narrow-line region. Chosen randomly $n_{\text{H}} = 10^3$ or $n_{\text{H}} = 10^4 \text{ cm}^{-3}$, 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 ($L_{\text{bol}} = k_{\text{bol}}L_{\text{X}}$), L_{acc} can be deduced from the following equation:

$$L_{\text{acc}} = L_{\text{X}}(k_{\text{bol}} - 1), \quad (19)$$

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

Another important aspect to take into account is the width of these lines. The emission lines of type 2 AGN are characterised by typical velocity dispersions of a few hundreds 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 $\text{H}\beta$ emission line (for $\text{FWHM}_{\text{H}\beta} \lesssim 1000 \text{ km s}^{-1}$), shown in their Figure 6, as a lognormal distribution centred at 355 km s^{-1} , covering the range $\text{FWHM} \sim [200, 1000] \text{ km s}^{-1}$. We assigned randomly the velocity of dispersion of type 2 AGN in our catalogue following this distribution. The resulting FWHM distribution is shown in Fig. [2.19](#).

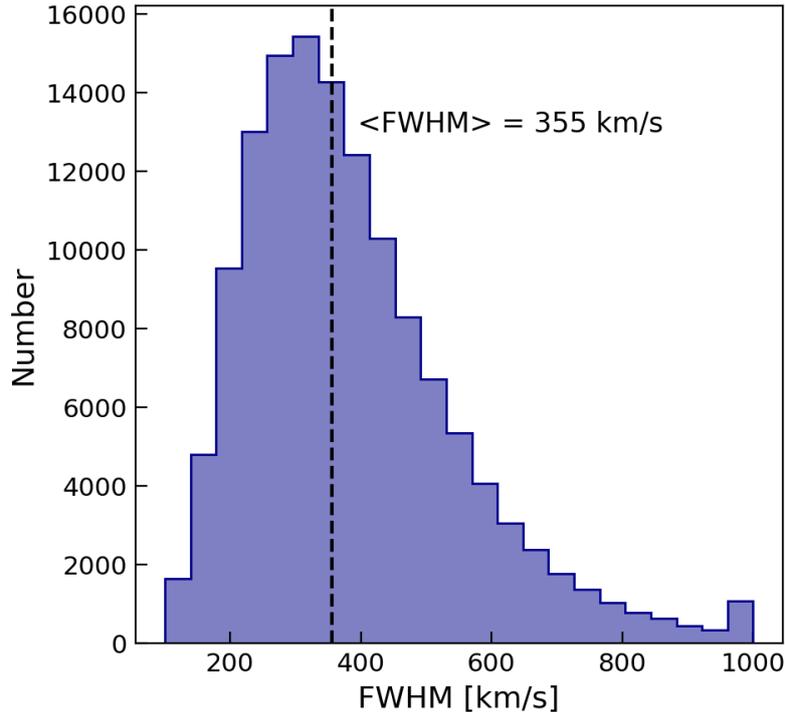


Figure 2.19: Distribution of the FWHM for the type 2 AGN of the lightcone presented in this work. The vertical dashed line shows the median value.

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 AGN component is left to future work.

We show in Fig. 2.20 three examples of type 2 AGN modelled with the method described above. 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 $[\text{O III}]_{\lambda 5007}$ from the AGN component with respect to the total. The three SEDs that we show in this figure have been selected with increasing values of this ratio, from close to 0, to close to 1, and with an intermediate value (~ 0.5). This quantity is printed also in Fig. 2.20 for each AGN, together with other relevant quantities (\mathcal{M} , L_X , SFR, etc.) Finally, the full list of narrow-region lines added to the spectra of type 2 AGN is reported in Tab. 2.

Table 2: Narrow-region lines added to type 2 AGN spectra. Wavelengths are given in vacuum, in units of Å.

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 (H α to H η)	
[O III]	4960.30	Paschen series (Pa α to Pa η)	
[O III]	5008.24		

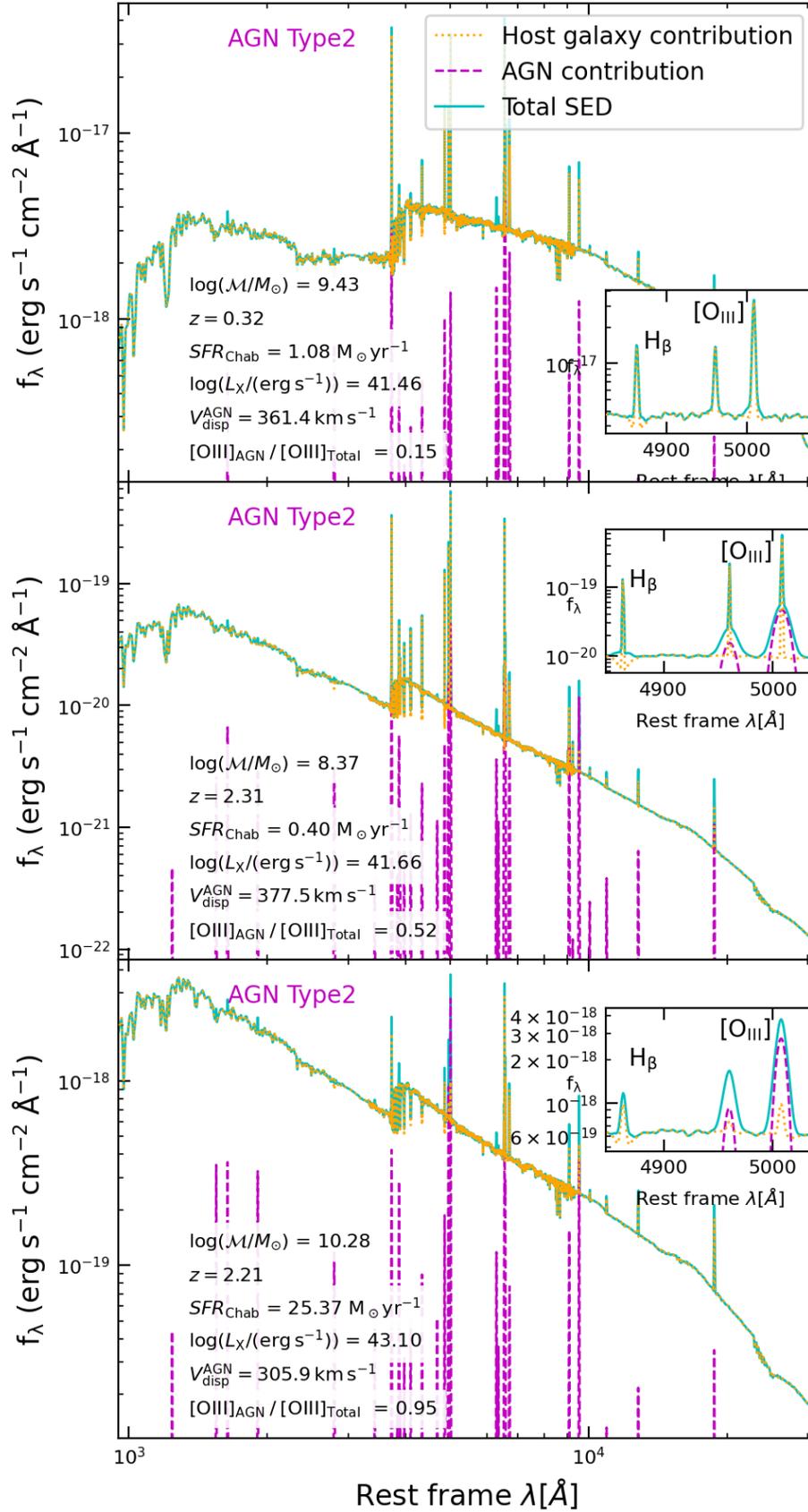


Figure 2.20: Some examples of SEDs of type 2 AGN from our lightcone. We show three representative cases: one source with emission lines mainly dominated by AGN emission, one dominated by the host-galaxy emission, and one intermediate case. In all cases, the galaxy component (bulge + disc) is shown with an orange dotted line and is based on [Bruzual & Charlot \(2003\)](#) models. The AGN component (only emission lines in this case) is shown with a dashed violet line. The cyan line shows the composed SED. Dust absorption is applied in all cases.

Chapter 3: Validating 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 quality and reliability of its predictions.

3.1 AGN fraction

The AGN host galaxy mass function 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. 3.1 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 $\log \lambda_{\text{sBHAR}} < -1.5$ and $\mathcal{M} > 10^{8.5} M_{\odot}$. We see that for all redshift bins the fractions are roughly consistent within their uncertainties.

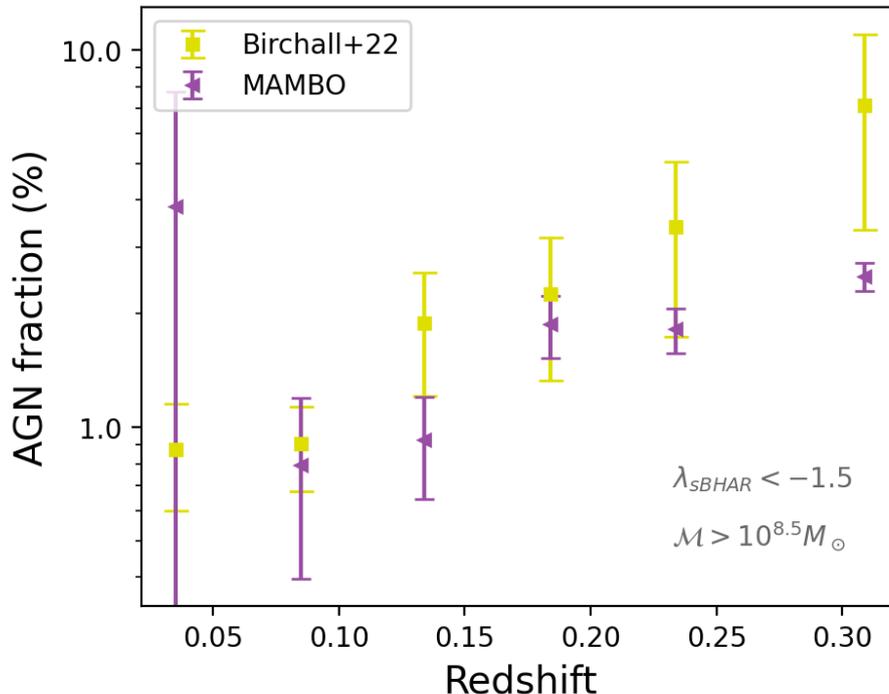


Figure 3.1: Fraction of AGN at $z < 0.33$ from our catalogue (purple triangles), selected with $\log \lambda_{\text{sBHAR}} < -1.5$ to compare them with the fraction reported in Birchall et al. (2022) (yellow squares). The error bars of the MAMBO AGN fraction show the Poissonian uncertainty.

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 $1.4 \pm 0.2\%$, 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.

3.2 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_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 3.2 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. 3.2 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_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. 3.3 we show the cumulative number counts of objects above a given X-ray flux. For this, we estimated the X-ray flux from L_X using

$$F_X = \frac{L_X}{4\pi D_L^2 K(z)}, \quad (20)$$

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

$$K(z) = (1+z)^{\Gamma-2}, \quad (21)$$

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 $N(> F_X)$ from our catalogue in Fig.3.3 to different data from the literature, we see overall a good agreement over a large range of flux until $F_X \gtrsim 2 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2}$, that is, two orders of magnitudes fainter than the sample used to calibrate our methodology.

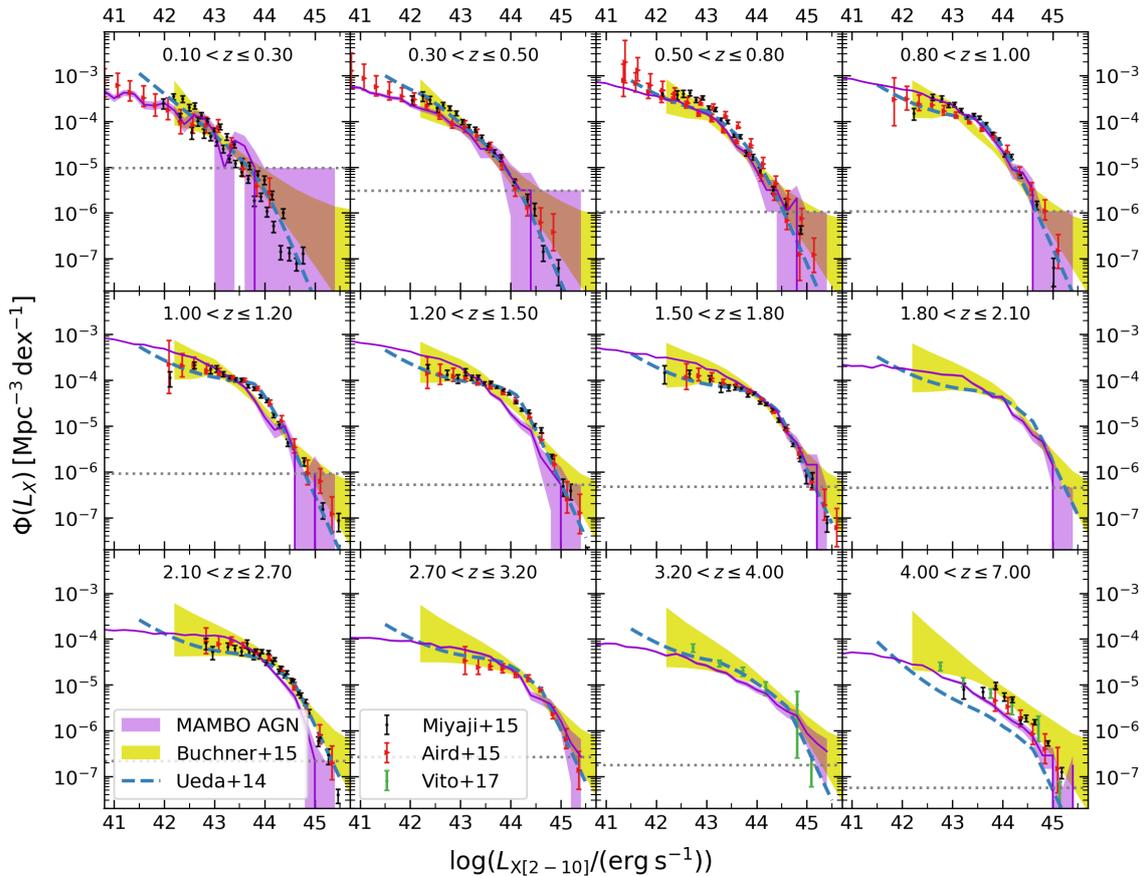


Figure 3.2: Hard X-ray luminosity function of the total population of AGN of our catalogue in different redshift bins, shown by the purple line. The shaded region represents the Poissonian uncertainty and the dotted horizontal line marks the limiting density from our lightcone (corresponding to 1 object $\text{Mpc}^{-3} \text{ dex}^{-1}$). 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).

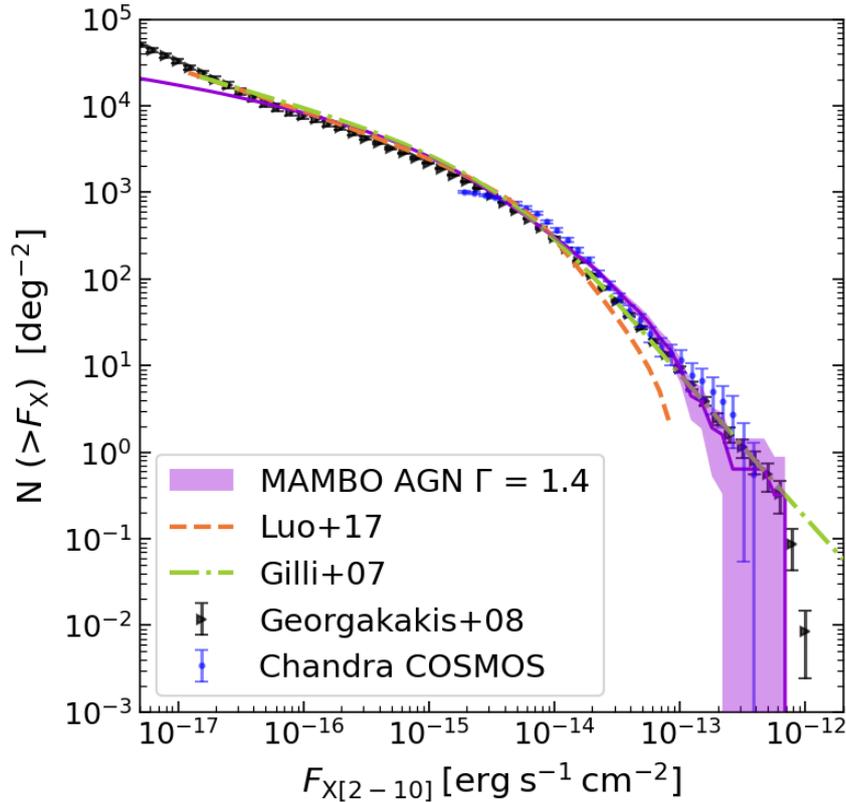


Figure 3.3: Cumulative number counts of X-ray flux in the hard band. The violet shaded area corresponds to the AGN in the MAMBO catalogue, and it shows the uncertainty of our values estimated as the Poissonian error. We show as comparison the number counts reported in different works. References: Luo et al. (2017), Gilli et al. (2007), Georgakakis et al. (2008), Marchesi et al. (2016b).

3.3 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 $[\text{O III}]_{\lambda 5007}/\text{H}\beta$ against $[\text{N II}]_{\lambda 6584}/\text{H}\alpha$. 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. 3.4 the BPT diagram constructed with the star-forming galaxies and the type 2 AGN from our catalogue. All sources have been selected with $z \leq 0.8$, since, as shown by Kewley et al. (2013a), it is not clear if at higher redshifts the BPT diagram can be used 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_X , and therefore this ratio is also correlated to L_X . We also show as comparison the observed line ratios from local galaxies and AGN 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 star-forming 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, $L_X < 10^{42} \text{ erg s}^{-1}$). In fact, different studies have pointed out that the BPT diagram is biased towards more luminous AGN, missing 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 $[\text{O III}]_{\lambda 5007}/\text{H}\beta$ against \mathcal{M} and was calibrated to separated star-forming galaxies from AGN at $0.3 < z < 1$. Fig. 3.5 shows the MEx diagram for the lightcone 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 which 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 $L_X < 10^{42} \text{ erg s}^{-1}$, 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.

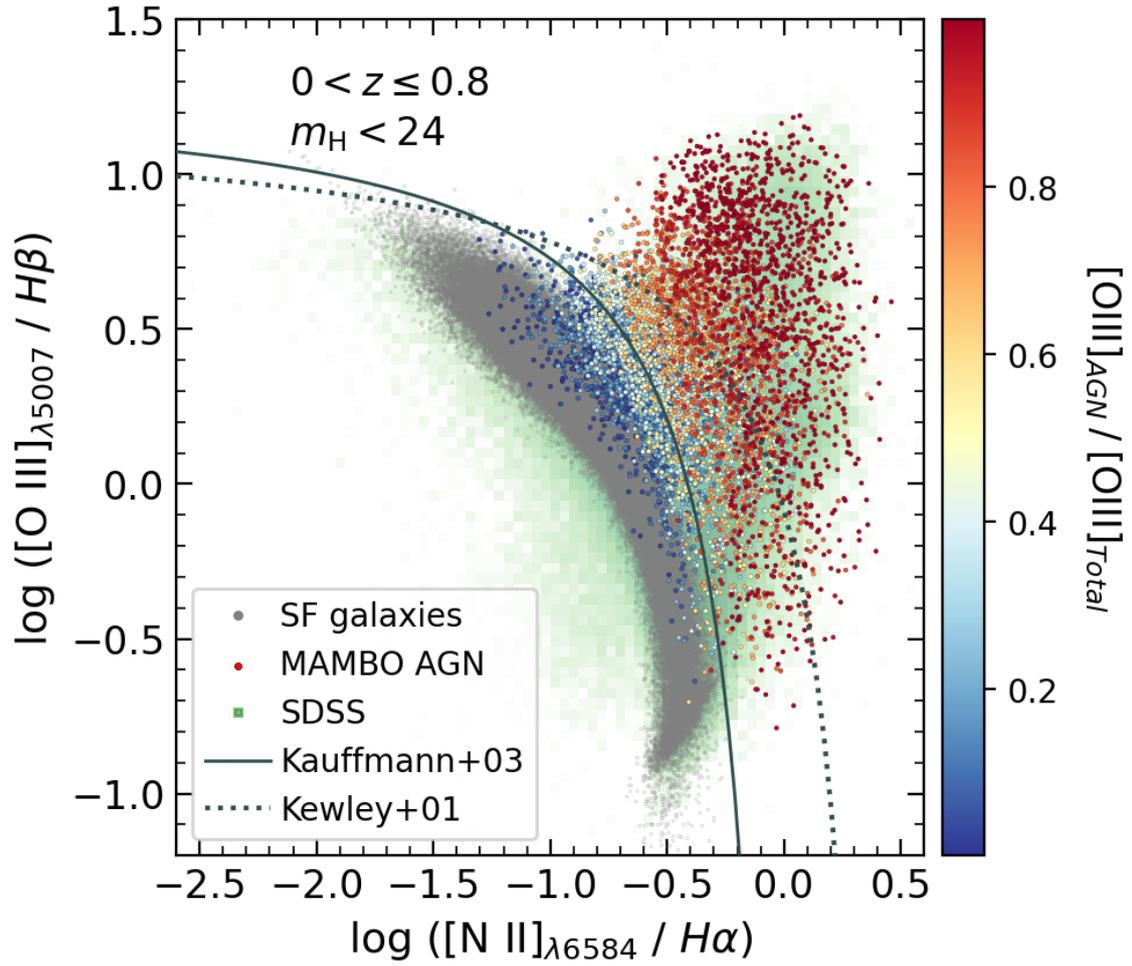


Figure 3.4: BPT diagram for galaxies and type 2 AGN from the MAMBO lightcone selected with $z \leq 0.8$ and $m_{H_E} < 24$. Galaxies are shown in grey, while AGN are colour-coded from blue to red according to the ratio of the emission line flux of [O III] from the AGN component with respect to the total (AGN + host galaxy). Objects from the SDSS DR8 are shown in green in the background, and are traditionally classified as star-forming galaxies if they 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.

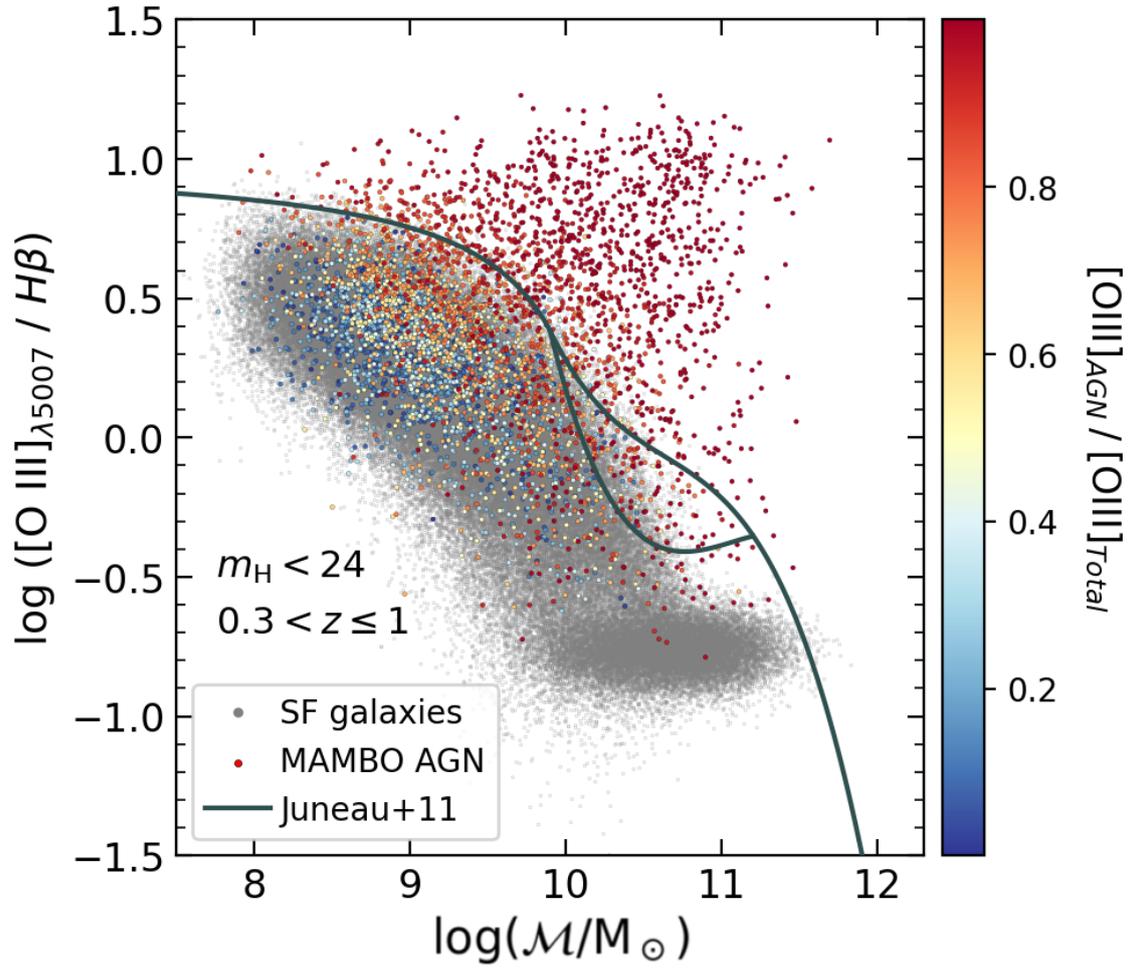


Figure 3.5: Mass-Excitation (MEx) diagram for galaxies and AGN from the MAMBO lightcone selected with $0.3 \leq z \leq 1.0$ and $m_{H_e} < 24$. Galaxies are shown in grey, while AGN are colour-coded from blue to red according to the ratio of the emission line flux of [O III] from the AGN component with 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 star-forming galaxies (below the line) and composite galaxies (between the two lines).

3.4 AGN colours and UVLF

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 $\lambda = 1450 \text{ \AA}$, 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. 3.6 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 doesn't include other systematic effects like selection effects or completeness level, which would increase the uncertainty.

In Fig. 3.6 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 restframe wavelength of 1450 \AA to construct the UVLF. From $z = 3$ to 6 we show also the UVLF from Finkelstein & Bagley (2022), 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 to 1 object/Mpc³/mag) for each redshift bin. On the other hand, the vertical dotted line shows the break magnitude M_* at 1500 \AA 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. (2017a), this magnitude is calculated as:

$$M_* = (1 + z)^{0.206} (-17.793 + z^{0.762}). \quad (22)$$

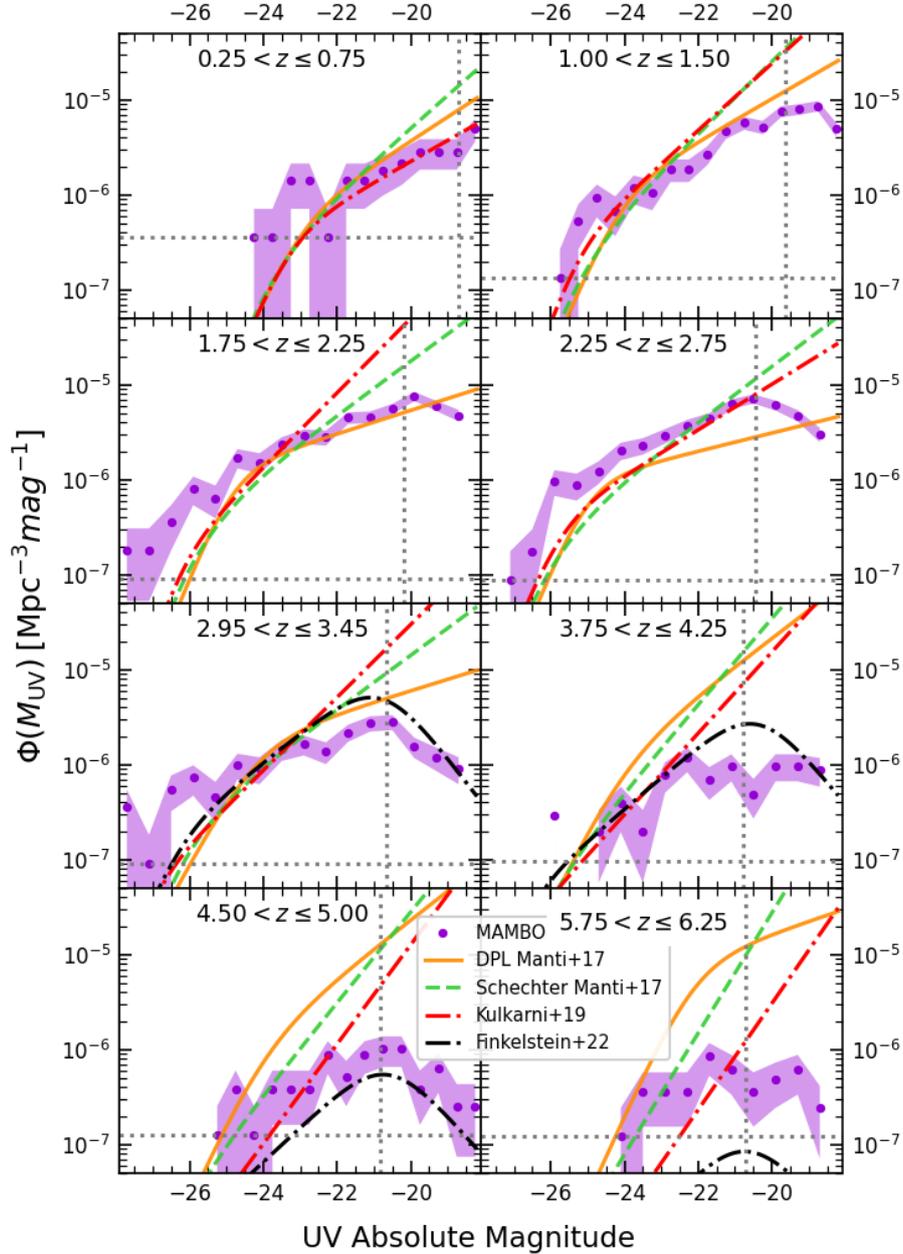


Figure 3.6: UV luminosity function of type 1 AGN from our catalogue (purple points) compared with different literature LFs in different redshift bins. The purple shaded area shows the uncertainty of our values, estimated as the Poissonian error. In the case of MAMBO, the UV magnitude is computed in the GALEX FUV filter, while for the other cases, it refers to M_{1500} . The orange solid and green dashed lines show the parametric LF from Manti et al. (2017), parameterised as a DPL and a Schechter function respectively. The red and black dash-dotted 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.

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 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. \(2017\)](#) or [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 to be completely dominated by the host-galaxy emission at these wavelengths.

In Fig. 3.7 we show the $(i - H)$ vs $(u - z)$ diagram, which was found in [Euclid Collaboration: Bisigello et al. \(2024\)](#) 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 ($f_{\text{AGN}}^{18} > 0.5$) or the galaxy component ($f_{\text{AGN}} < 0.5$). 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 $f_{\text{AGN}} > 0.5$, while both distributions are clearly separated from that of AGN with $f_{\text{AGN}} < 0.5$. Also, the MAMBO AGN dominated by the AGN component are in agreement with the best colour by [Euclid Collaboration: Bisigello et al. \(2024\)](#).

¹⁸In this section we use f_{AGN} as shorthand for $f_{\text{AGN}_{FUV}}$ (Eq. 17).

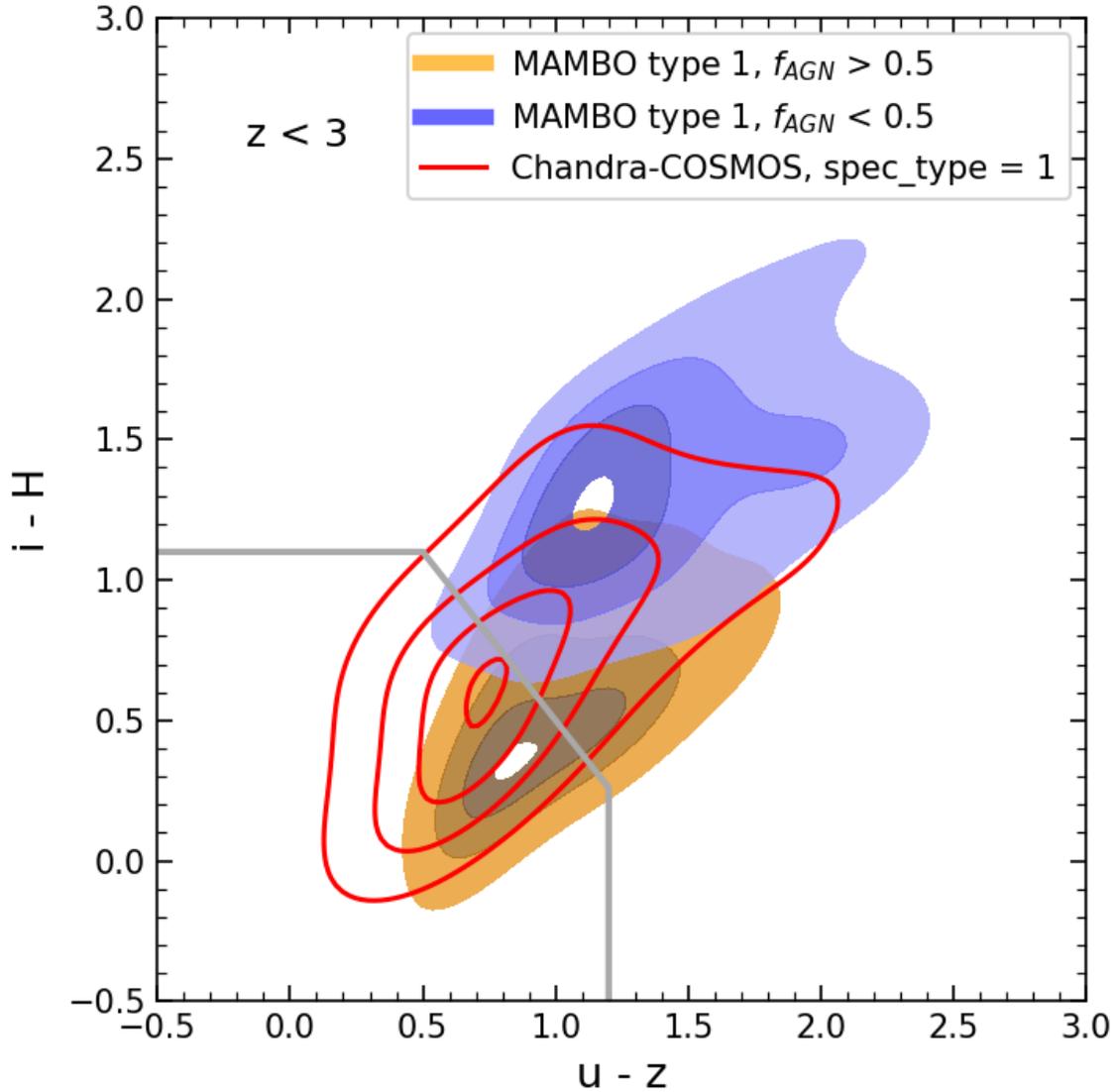


Figure 3.7: Colour-colour diagram to separate galaxies from type 1 AGN as in [Euclid Collaboration: Bisigello et al. \(2024\)](#). Filled contours show the distribution of type 1 AGN from the lightcone here presented, separated into those dominated by the AGN component ($f_{AGN} > 0.5$, orange contours) or the galaxy component ($f_{AGN} < 0.5$, 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: Bisigello et al. \(2024\)](#) to separate type 1 AGN from galaxies and type 2 AGN.

3.5 Clustering

A key quantity our mock must reproduce to make predictions for future surveys is the spatial distribution of galaxies and AGN, that is, their clustering signal. The study of the clustering signal of the MAMBO galaxies has already been presented in [Girelli \(2021\)](#). In this section, we present the results for the MAMBO AGN.

The two-point correlation function $\xi(r)$ is a common metric for quantifying the clustering signal, and it measures the variance in the AGN distribution. It represents the excess probability, compared to a random (unclustered) distribution, of finding a pair of AGN, one within the volume dV_1 and the second within dV_2 , separated by a distance r ([Peebles, 1980](#)):

$$dP = n^2 [1 + \xi(r)] dV_1 dV_2, \quad (23)$$

where n is the mean AGN number density, and therefore, $n^2 dV_1 dV_2$ represents the mean probability of finding an AGN pair on an unclustered sample.

However, the measurement of the pair separation r is affected by redshift space distortions. That is, distances between AGN pairs are inferred from their redshifts, which are affected by the peculiar velocity of each source. The overall impact of peculiar velocities is to increase the radial separations of pairs, thereby shifting the clustering power to larger scales. This results in a generally flatter slope compared to the real-space correlation function ([Gilli et al., 2009](#)). Additionally, uncertainties in redshift measurements produce a similar effect.

To mitigate this effect, we computed the projected two-point correlation function, $\omega(r_p)$, which is insensitive to distortions from redshift space, and be calculated as ([Davis & Peebles, 1983](#)):

$$w_p(r_p) = 2 \int_0^\infty \xi(r_p, \pi) d\pi, \quad (24)$$

where $\xi(r_p, \pi) d\pi$ is the two-point correlation function expressed in terms of distance parallel (π) and perpendicular (r_p) to the observer's line of sight.

In practice, since the integration in Eq. 24 is carried out numerically in bins of π , it is not performed to infinity but rather to a finite value π_{\max} . The value of π_{\max}

should be large enough to capture most of the clustering signal dispersed by peculiar velocities along the line of sight, but not large enough as to introduce excess noise due to the signal by uncorrelated pairs, that would dominate beyond a certain π value.

A standard way of calculating $\xi(r_p, \pi)$ in bins of r_p and π is the [Landy & Szalay \(1993\)](#) estimator, given by

$$\xi(r_p, \pi) = \frac{dd(r_p, \pi) - 2dr(r_p, \pi) + rr(r_p, \pi)}{rr(r_p, \pi)} \quad (25)$$

where dd , dr , and rr correspond to the normalised number of data–data, data–random, and random–random pairs, respectively. Therefore, estimating the projected two-point correlation function requires the construction of a random catalogue that acts as an unclustered distribution of sources.

We performed all calculations described in this section, from the creation of the random catalogue to the computation of $\omega(r_p)$ using [CosmoBolognaLib¹⁹](#) ([Marulli et al., 2016](#)), an open-software Python/C++ library for numerical cosmological calculations.

Samples and random catalogues

We compared the measurements of the projected two-point correlation function on the MAMBO AGN with different results from the literature, namely [Viitanen et al. \(2019\)](#) and [Gilli et al. \(2009\)](#). For each comparison, we created different subsamples of MAMBO AGN, applying appropriate selection criteria to match the characteristics of the observed samples.

In [Viitanen et al. \(2019\)](#), the authors studied the clustering signal of a sample for 1130 X-ray selected AGN in the redshift range $0.1 < z < 2.5$ from the XMM-COSMOS dataset ([Bongiorno et al., 2012](#)), with a limiting flux of $F_X \sim 3 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$, covering a luminosity range of $\log(L_X/\text{erg s}^{-1}) = 42.3 - 45.5$.

The AGN sample studied in [Gilli et al. \(2009\)](#) comprises 538 X-ray selected AGN also in the XMM-COSMOS field, spectroscopically identified with $I_{AB} < 23$ within the redshift range $0.2 < z < 3.0$.

¹⁹<https://github.com/federicomarulli/CosmoBolognaLib>

We selected subsamples of type 1 and 2 AGN from the MAMBO catalogue within each study’s redshift ranges and applied the same limit in F_X . We further selected objects with $\log(L_X/\text{erg s}^{-1}) > 42$.

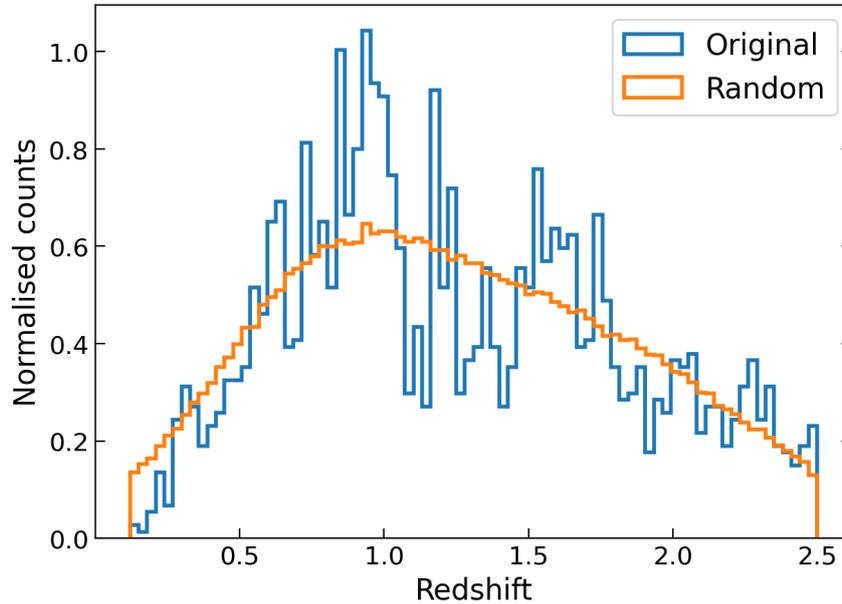


Figure 3.8: Redshift distribution of a sample of MAMBO AGN (blue) and of its associated random catalogue (orange). The random distribution was generated with a Gaussian smoothing kernel with $\sigma_z = 0.3$. We denote with "Original" the distribution as it is in the MAMBO lightcone and "Random" the sample we randomly extracted from the "Original".

To compare with the results from [Gilli et al. \(2009\)](#), we also applied the magnitude limit $I_{AB} < 23$. Besides, that study separately examined the clustering of broad line AGN (BLAGN) or non-BLAGN. Therefore we created two distinct MAMBO AGN subsamples for type 1 and type 2 AGN, respectively.

Next, we constructed random catalogues for each subsample, assigning random redshifts from each smoothed distribution of AGN redshifts. We used a Gaussian smoothing kernel with $\sigma_z = 0.3$ to achieve an optimal balance between avoiding overfitting and ensuring appropriate smoothing of the distribution. Following this, we drew random coordinates for right ascension and declination from a random distribution. As an example, we show in [Fig. 3.8](#) the redshift distribution of the sample we built to compare with [Viitanen et al. \(2019\)](#), together with the redshift distribution of its associated random catalogue. Similarly, in [Fig. 3.9](#) we show the RA and Dec distributions of the same two catalogues.

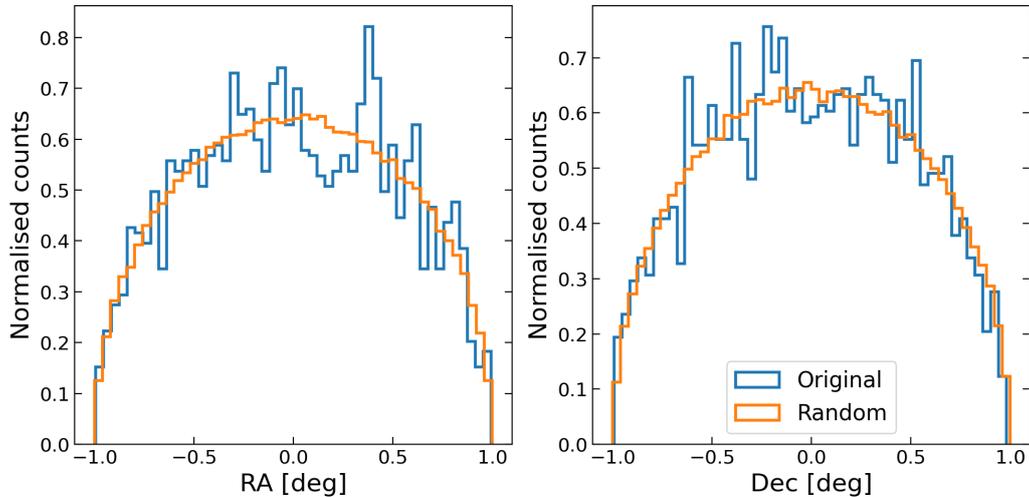


Figure 3.9: RA and Dec distributions of a sample of MAMBO AGN (blue) and of its associated random catalogue (orange).

Projected two-point correlation function

We performed the calculation of the projected two-point correlation function, $\omega(r_p)$, within given ranges in the perpendicular (r_p) and parallel (π) dimensions. These values were chosen to be the same as in each study in order to perform a meaningful comparison with each of them (see e.g. [Storey-Fisher & Hogg 2020](#) for a discussion on the effect that the binning can have on measurements of the two-point correlation function).

For [Viitanen et al. \(2019\)](#), these values correspond to $r_p = 1 - 100 \text{ Mpc } h^{-1}$ in 12 logarithmically spaced bins, and only one bin in the π direction, with $\pi_{\text{max}} = 40 \text{ Mpc } h^{-1}$. Instead, [Gilli et al. \(2009\)](#) measured $\omega(r_p)$ on projected scales $r_p = 0.3 - 40 \text{ Mpc } h^{-1}$ in intervals of $\Delta \log(r_p) = 0.2$, and used also $\pi_{\text{max}} = 40 \text{ Mpc } h^{-1}$.

In [Fig. 3.10](#), we show the results of the projected two-point correlation function from the MAMBO AGN compared to the results from [Viitanen et al. \(2019\)](#). We observe a very good agreement between the AGN clustering signal predicted by the MAMBO catalogue and the observations. While at large scales ($r_p = 40 \text{ Mpc } h^{-1}$) the clustering signal predicted from MAMBO is significantly larger than the model from [Viitanen et al. \(2019\)](#), we note that at that scale, there are no binned points to compare with, and the model from [Viitanen et al. \(2019\)](#) is an extrapolation from the previous points.

In Fig. 3.11, we compare the clustering of MAMBO AGN, separated into type 1 and type 2, to that of the BLAGN and non-BLAGN samples from Gilli et al. (2009). Again, we observe a very good agreement between the MAMBO AGN and the observations, both for the type 1 and type 2 AGN samples. In this figure, we have also extrapolated the best-fit model from Gilli et al. (2009) to scales larger than the ones they could perform the analysis due to the observational limitations.

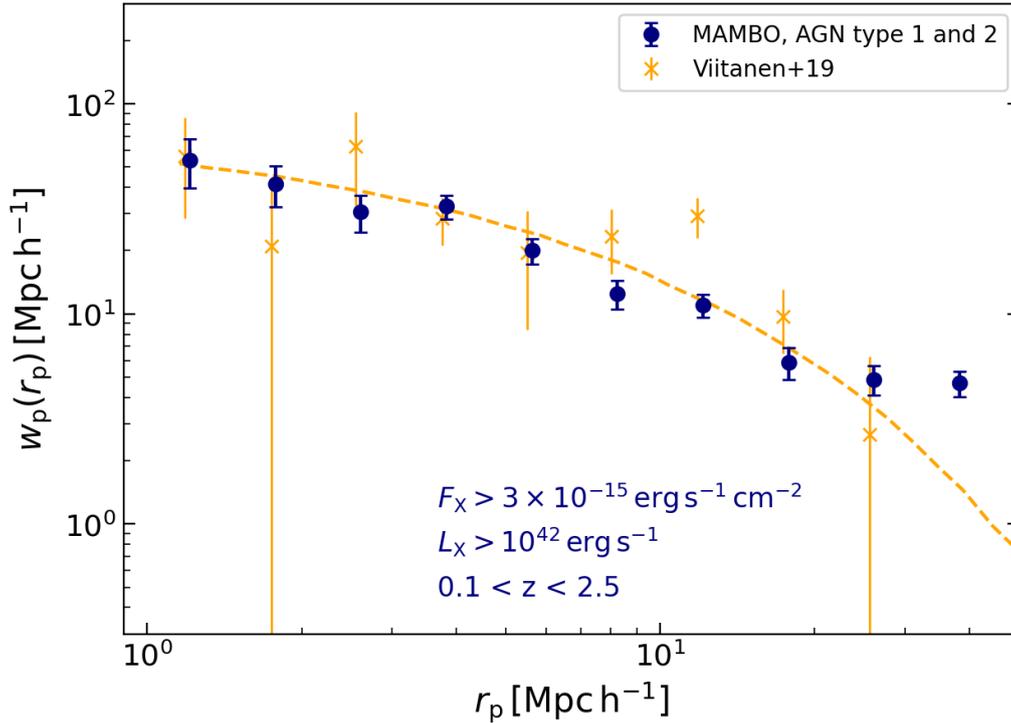


Figure 3.10: Projected two-point correlation function as a function of perpendicular distance r_p . The dark blue points with errorbars show the binned $\omega(r_p)$ calculated from a subsample of MAMBO AGN with similar redshift and flux limits as the ones in Viitanen et al. (2019), indicated in the figure. The dashed orange line shows the best-fit model from Viitanen et al. (2019). Error bars for the MAMBO AGN points show the Poissonian uncertainties.

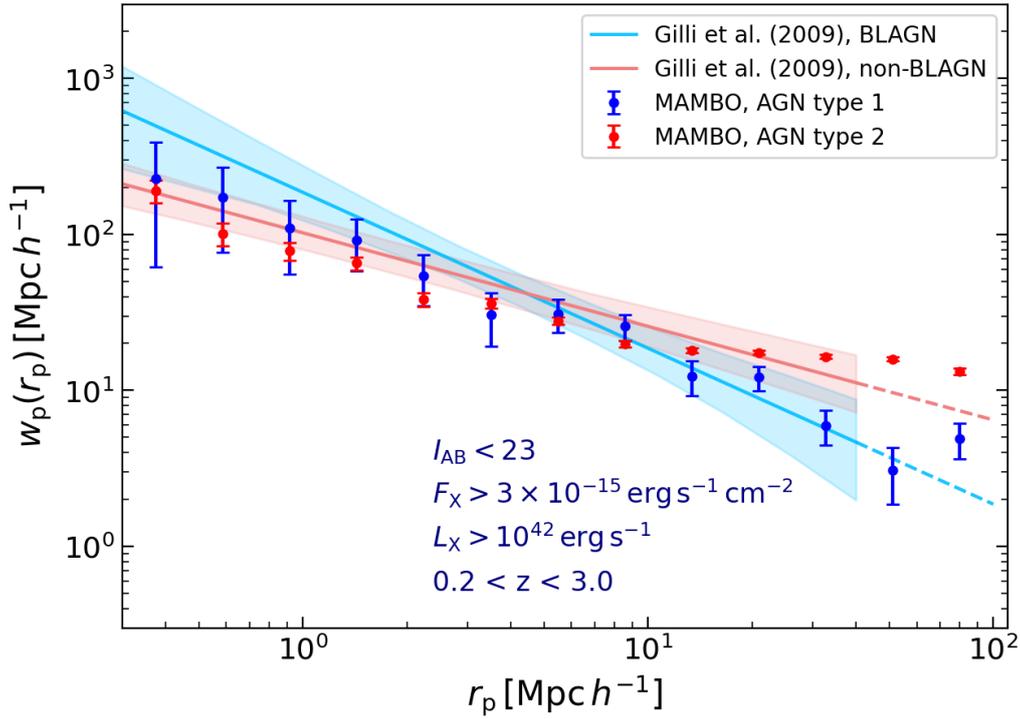


Figure 3.11: Projected two-point correlation function as a function of perpendicular distance r_p . The blue and red points with errorbars show the binned $\omega(r_p)$ calculated from subsamples of MAMBO type 1 and 2 AGN (respectively), with similar redshift, flux, and magnitude limits as the ones in Gilli et al. (2009), indicated in the figure. The solid blue and red lines with shaded areas show the best-fit model from Gilli et al. (2009) to the clustering of the BLAGN and non-BLAGN samples, respectively, with its uncertainty. Error bars for the MAMBO AGN points show the Poissonian uncertainties.

Chapter 4: Applications of MAMBO

In this Chapter we will show several applications of the AGN catalogue presented and validated during the previous Chapters. We will focus on *Euclid* (see Sect. 1.3.1) as a case study to apply the catalogue: we start by presenting general forecasts in Sect. 4.1 of the observations performed by this survey, focusing on number densities of different sources observed in a given photometric band, and narrow-line diagnostic diagrams to differentiate AGN from star-forming galaxies.

We also simulated AGN spectra that emulate the spectroscopic observations from *Euclid*. We analysed this spectra using official *Euclid* pipelines, and studied diagnostic diagrams build with the *Euclid*-like measured fluxes. We present these results in Sect. 4.2.

Then, in Sect. 4.3 we present the implementation of our AGN model applied to a larger DM+galaxy empirical simulation, the Flagship 2. We used this large AGN simulation for two different projects: first, to test the accuracy of an official *Euclid* pipeline at retrieving photometric redshifts of AGN simulated with *Euclid*-like photometry (Sect. 4.4); and second, to aid in the preparation of an incoming X-ray observational campaign (Sect. 4.5).

4.1 General forecasts

4.1.1 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. 4.1 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. 4.1 the m_H distribution of type 1 and 2 AGN (classified into these 2 categories by optical spectroscopy), selected with minimum X-ray flux $F_X = 1.9 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$. 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.

Table 3: Surface density and total numbers of galaxies and AGN (full population, type 1 and type 2). Each row corresponds to a different selection, written in the table and explained in greater detail in the text. All AGN are selected with $L_X > 10^{42} \text{ erg s}^{-1}$. Total numbers assume that the *Euclid* Wide (EWS) and Deep (EDS) surveys will cover areas of $14\,679 \text{ deg}^2$ and 53 deg^2 respectively (Euclid Collaboration: Mellier et al., 2024). Fluxes are given in $\text{erg s}^{-1} \text{ cm}^{-2}$. Densities are in deg^{-2}

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

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 $L_X > 10^{42} \text{ erg s}^{-1}$, to remove possible non-AGN X-ray emitters. We compare these numbers with the ones reported in [Euclid Collaboration: Selwood et al. \(2024\)](#), 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: Selwood et al. \(2024\)](#), the authors report the surface densities of AGN detectable in the H_E band for the *Euclid* Wide and Deep Surveys. For the EWS, the reported surface densities are $4.5 \times 10^3 \text{ deg}^{-2}$, $6.8 \times 10^2 \text{ deg}^{-2}$ and $1.7 \times 10^3 \text{ deg}^{-2}$ for all AGN, type 1 and type 2 respectively. The corresponding numbers for the EDS are $2.4 \times 10^3 \text{ deg}^{-2}$, $8.8 \times 10^2 \text{ deg}^{-2}$ and $3.5 \times 10^3 \text{ deg}^{-2}$. 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 $H\alpha$ emission line flux above the detection limit of each survey (see Sect. 1.3.1), 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.

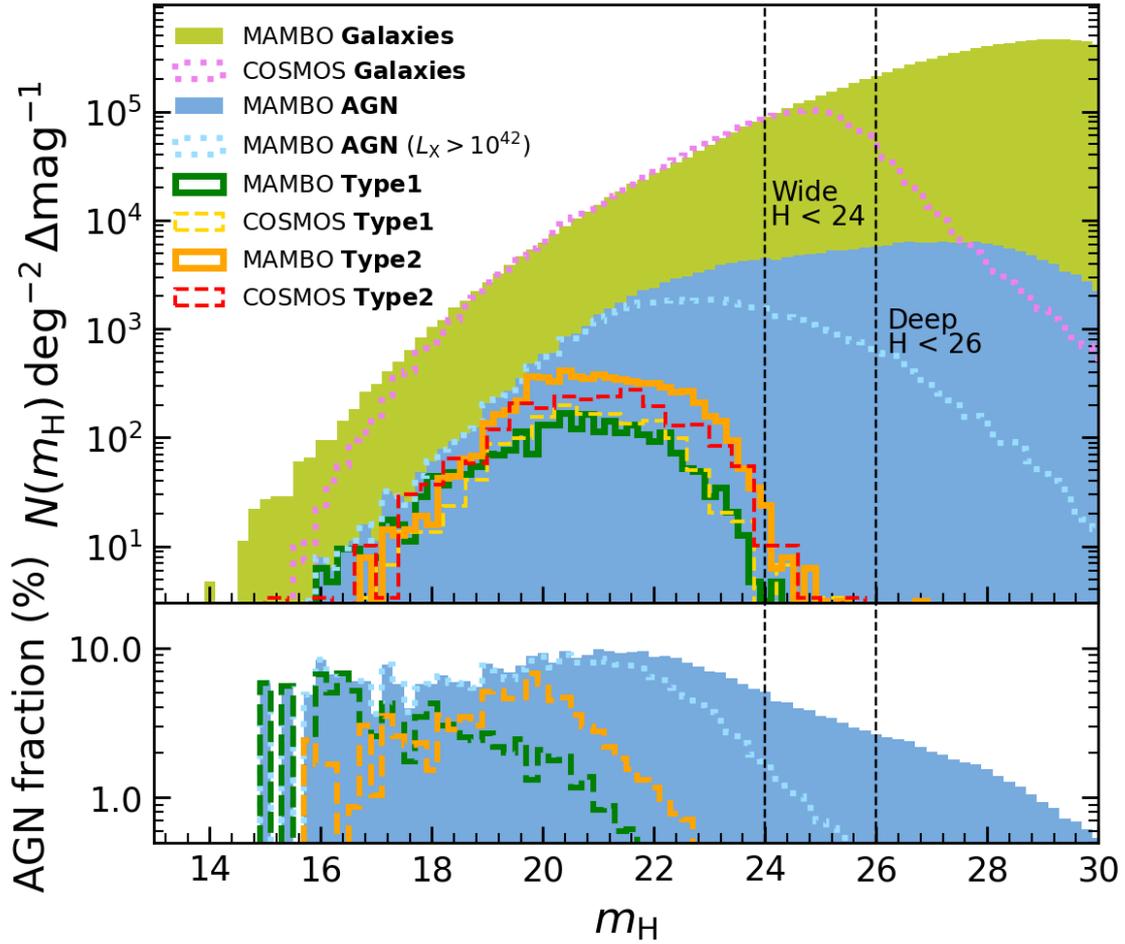


Figure 4.1: Upper panel: Number density counts in the m_H magnitude. Galaxies and AGN from the MAMBO lightcone are represented by green and blue filled histograms respectively. AGN selected with $L_X > 10^{42}$ erg s $^{-1}$ are shown with a cyan dotted line. We also show the density counts for galaxies (pink dotted line) and AGN from the COSMOS catalogue (Laigle et al., 2016; Marchesi et al., 2016b), where AGN have been selected with $F_X > 1.9 \times 10^{-15}$ erg s $^{-1}$ cm $^{-2}$, and separated into type 1 and 2 based on optical spectroscopic criteria (yellow and red dashed lines). The green and orange solid lines show the number density counts of type 1 and 2 AGN from our mock after applying the same cut in F_X as for the COSMOS sample. We show with vertical black dashed lines the limiting magnitude of the EWS and EDS. Lower panel: Fraction of MAMBO AGN as a function of m_H . The different lines represent different subpopulations of AGN, as in the upper panel.

4.1.2 Diagnostic diagrams

In Sect. 3 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., 2013b; 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 H α lines are likely to be blended in real *Euclid* spectra (Euclid Collaboration: Lusso et al., 2024).

We show in Fig. 4.2 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.

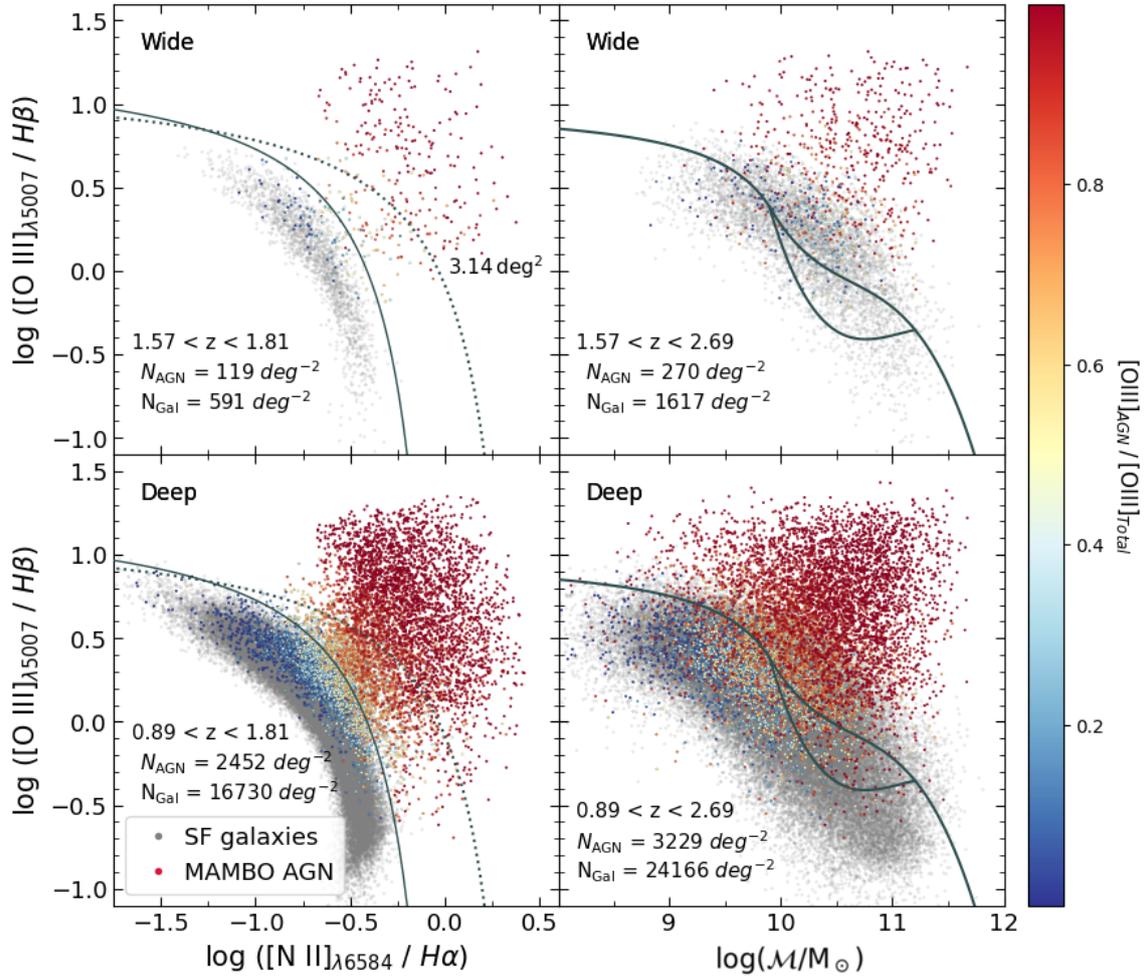


Figure 4.2: Predictions from our lightcone for the BPT (left panels) and MEx diagrams (right panels) as observed by the *Euclid* Wide and Deep surveys. For each panel we have plotted the galaxies (grey dots) and AGN (colour-coded from blue to red) corresponding to the specific redshift range, magnitude limit (in H-band) and $\text{H}\alpha$ 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. 3.4 and 3.5 for further details.

4.2 Euclid-like AGN spectra

In this part of the work, our objective was to generate mock *Euclid*-like AGN spectra to assess the accuracy with which the official *Euclid* pipelines can recover key spectral properties (e.g., line fluxes and FWHM) of AGN. By *Euclid*-like spectra, we refer to simulations that closely resemble the spectra that will be captured by the *Euclid* NISP instrument, incorporating as many instrumental and environmental effects as possible. We simulated the *Euclid*-like spectra using the bypass code FastSpec.

Specifically, for this exercise, we focused on studying the accuracy of the flux measurements of narrow lines in type 2 AGN, focusing on the necessary lines to construct the BPT diagnostic diagram.

Fastspec

FastSpec (Euclid Consortium: Granett et al., in prep.) is a Python-based bypass code designed to efficiently model the instrumental and environmental effects of Euclid NISP spectra. In summary, FastSpec takes in input incident spectra and incorporates a range of factors affecting both the instrument and observations, including the point spread function (PSF) model, flux losses due to the spectral extraction window, and noise contributions from environmental sources such as the zodiacal background and stray light. Additionally, it simulates instrumental noise components like readout noise and dark current. The code also considers the relevant transmission functions, quantum efficiency, effective collecting area, and exposure times for each observation. Additionally, FastSpec enables customisation of the exposure time ratios between the blue and red grisms for simulated EDS spectra.

The advantage of a bypass code simulator like FastSpec lies in its simplicity and speed, which is achieved by incorporating simplified yet accurate representations of instrumental effects. This ensures the high computational efficiency of the code, allowing for the generation of large amounts of *Euclid*-like mock spectra within short timescales. A complete end-to-end simulation of the full process chain, from the mock catalogue through the imaging processing to the final data product (the measured spectral features, in this case), is a time-consuming and complex process. In the case of *Euclid*, this process involves the coordinated activity of several Organisation Units (OUs), namely OU-SIM (OU for SIMulation data), OU-SIR (in charge of data

reduction and spectra extraction) and OU-SPE (in charge of redshift determination and spectral measurements). As shown in Fig. 4.3, FastSpec allows to bypass most of these steps.

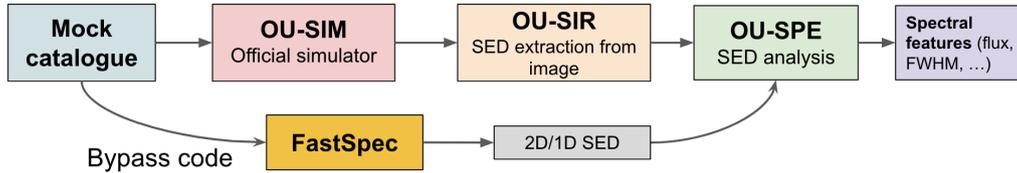


Figure 4.3: Schematic representation of the steps that FastSpec bypasses. OU-SIM, OU-SIR and OU-SPE correspond to different Organisation Units within the Euclid Consortium, described in the text. Its coordinated effort allows, among many other tasks, for the simulation and analysis of mock *Euclid*-like spectra. FastSpec bypasses most of these steps.

However, this approach also has certain significant limitations. First, in slitless spectroscopy, overlapping spectra from nearby objects substantially contribute to the noise in individual spectra. To mitigate this effect, *Euclid* follows a particular observing sequence involving several²⁰ exposures taken at different dispersion angles. FastSpec simplifies this process by arranging galaxies on a grid without interference from neighbouring objects and simulating only the first-order spectra, where over 96% of the object flux is concentrated. Therefore, the simulated *Euclid*-like spectra represent an idealised scenario, free from contamination by overlapping spectra, as if we are handling fully decontaminated spectral products.

Second, FastSpec processes a single incident SED per galaxy, which limits the ability to differentiate between nuclear, bulge and disk components. To accommodate this feature of the simulator, we merged the three components (AGN, bulge and disc) of our MAMBO incident SEDs into a single incident SED. FastSpec then convolves this merged SED with the galaxy’s surface brightness profile and instrumental PSF along the dispersion direction, including background contributions.

However, FastSpec models the galaxy’s 2D surface brightness profile of a galaxy as a composite of only two Sérsic profiles: one for the bulge and the other for the disc. The composite model takes from the MAMBO simulation the radii of each component, preserves the total luminosity, and adjusts the relative contributions of

²⁰Four in the EWS and 40 in the EDS.

the two components based on the bulge-to-total (BT) ratio, inclination, and position angle. As a result, the emission lines are dispersed uniformly across the 2D galaxy spectrum, neglecting spatial variations from inner to outer galactic regions and the spectral contributions from older, centrally-concentrated stellar populations extending into the disc outskirts, where younger stars typically dominate. Similarly, the nuclear emission from the AGN is dispersed along the full galaxy. To try to mitigate this effect, we modelled type 1 AGN as point-like sources.

Finally, FastSpec does not consider other possible observational effects, such as contamination from cosmic rays.

Euclid-like spectra

We applied FastSpec to a sample of type 1 and 2 AGN from the MAMBO catalogue, simulating the limits of both the EWS and the EDS. To reproduce the EDS, we simulated the configuration corresponding to the DEEP-FORNAX field, which will be observed with 15 passes of the red grism and 25 passes of the blue grism. We remind the reader that for the EWS the sky will be observed with only one pass (at four different angles) of the red grism, which is what we simulated for this work as well.

The sample we selected comprises 979 type 1 AGN and 955 type 2 AGN for the EWS limits, and 982 type 1 AGN and 997 type 2 AGN for the EDS. This sample was created selecting type 1 AGN with $f_{\text{AGN}_H} \geq 0.7$ (Eq. 17) and type 2 AGN with $L_X \geq 10^{42} \text{ erg s}^{-1}$. Additionally, we applied for each survey the corresponding limiting $F_{\text{H}\alpha}$ flux and m_H magnitude (see Sec. 1.3.1).

In the next figures, we show some examples of these simulated spectra. We show in Fig. 4.4 and Fig. 4.5 the spectra of type 1 AGN simulated for the EWS and the EDS, respectively, and in Fig. 4.6 and Fig. 4.7 the equivalent for type 2 AGN. In all cases, we show examples of sources with varying H_E band magnitude, from bright sources to sources near the limiting magnitude of the EWS ($m_H = 24$) and EDS ($m_H = 26$). We also show the incident spectra (i.e., before FastSpec) for comparison.

In the upper panel of Fig. 4.4 and Fig. 4.5, it is evident that the mock *Euclid*-like SED is systematically fainter than the intrinsic (incident) one. This is due to the fact that FastSpec uses a fixed spectral extraction window of 5 px (= 1.5"), regardless of

the apparent size of the source. Therefore, if a significant portion of the source’s flux extends beyond this radius, the resulting flux loss can be quite considerable. This effect is clearly more pronounced in sources with larger angular size (and therefore generally closer and apparently brighter sources), as it is visible by looking at the other panels in these two figures. In real observations, it is expected that the so-called optimal extraction will be performed, where the extraction window will be adjusted based on measurements of each source’s size and Sérsic index.

Looking at the two middle panels of Fig. 4.4, we see that for bright type 1 AGN ($m_H \lesssim 22.5$), at least some spectral features of the intrinsic SEDs are clearly visible in the *Euclid*-like SED ones. For spectra with a low SNR where at least one line can be safely identified and, therefore, redshift can be determined, stacking of multiple spectra will be performed in order to increase the signal level (Quai et al. in prep.).

On the other hand, as we approach the limiting magnitude of each survey, the noise starts dominating over the signal of the emission lines, making it impossible to recognize any spectral features. To be able to apply stacking techniques to these highly noisy spectra, it is crucial to know their redshift confidently. For this, it is expected that future data releases of *Euclid* data will include a redshift-reliability flag, derived through machine-learning methods trained on previous data releases.

For type 2 AGN (Figs. 4.6 and 4.7), we see that in all cases, the emission line fluxes at the *Euclid*-like SEDs are lower than in the incident spectra. Additionally, due to the spectral resolution of the NISP instrument ($R = 380$), nearby lines such as $H\alpha$ and $[N\ II]$ appear blended in the *Euclid*-like spectra even for the brightest sources.

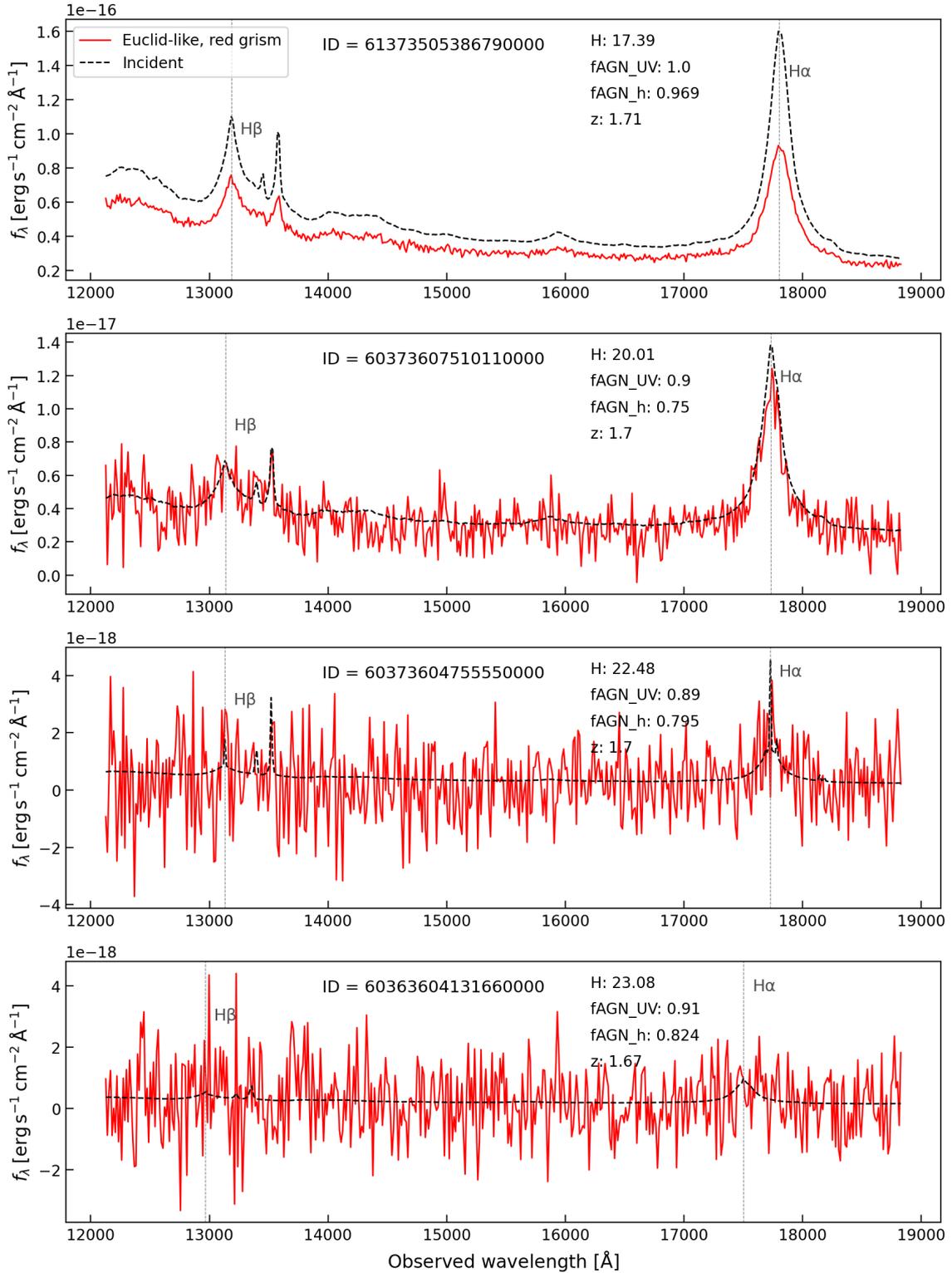


Figure 4.4: Mock *Euclid*-like spectra of type 1 AGN at the EWS limits simulated with FastSpec. The red line shows the *Euclid*-like spectra, while the black dotted line shows the incident one.

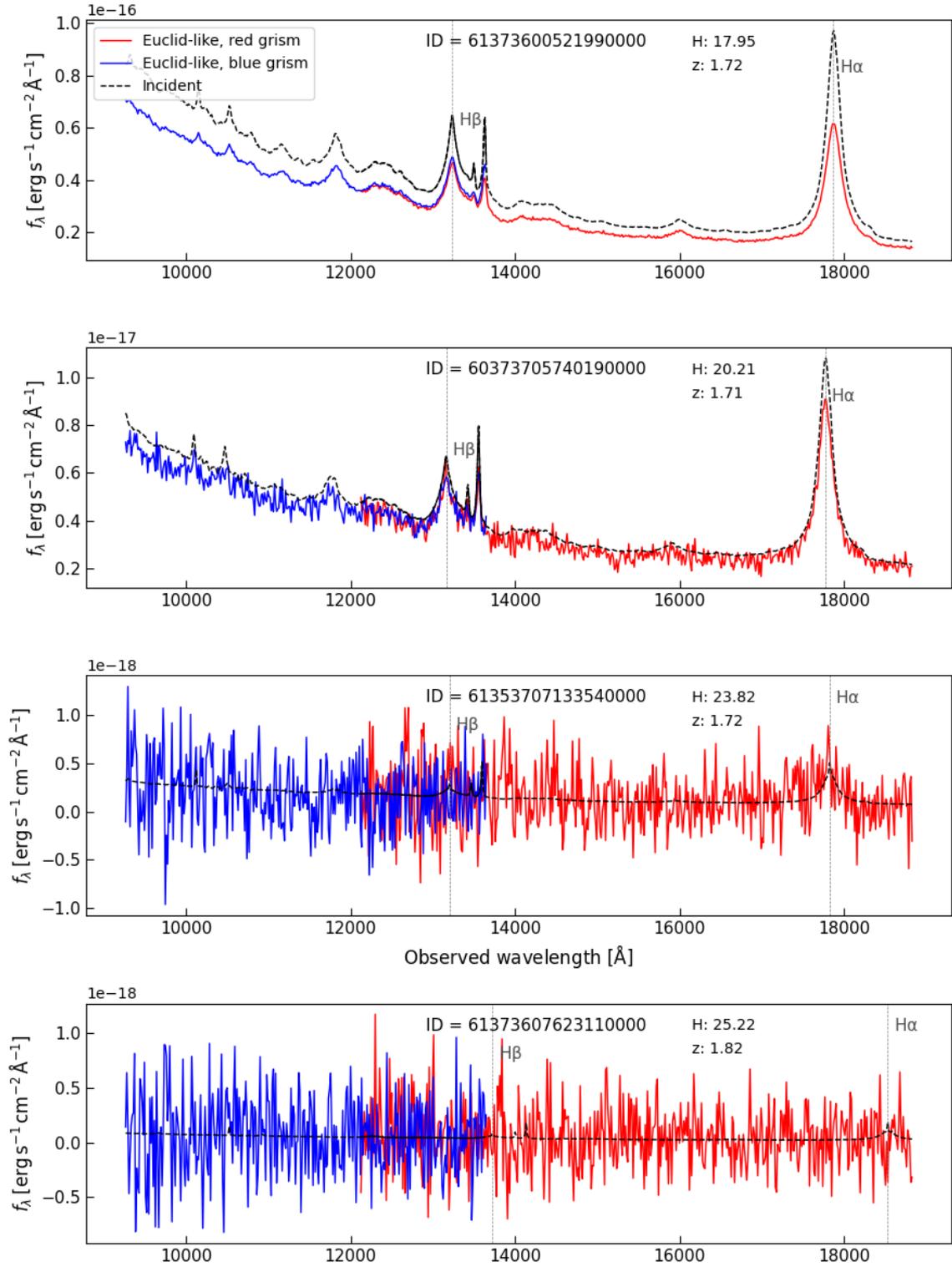


Figure 4.5: Mock *Euclid*-like spectra of type 1 AGN at the EDS limits simulated with FastSpec. The red and blue lines show the *Euclid*-like spectra, while the black dotted line shows the incident one

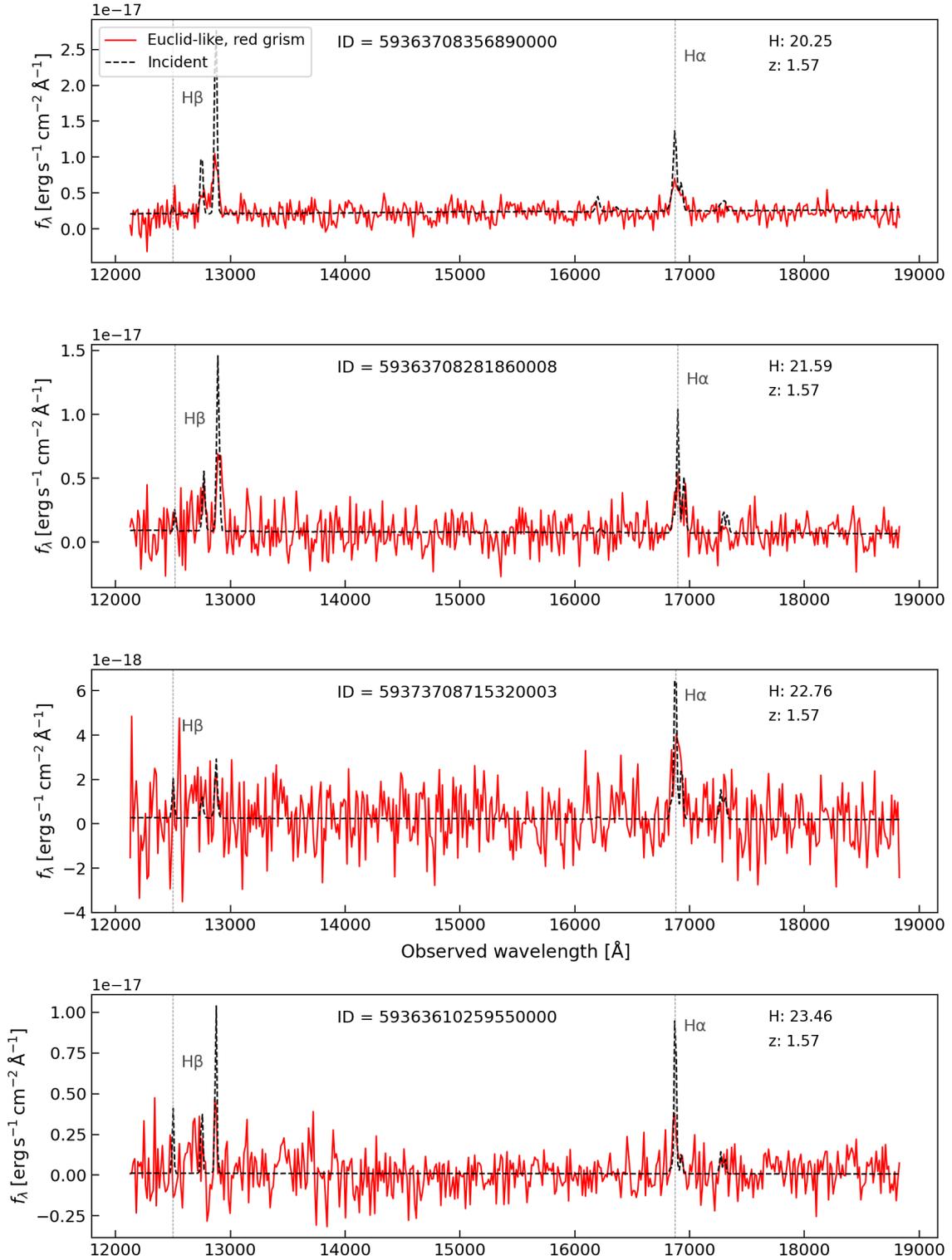


Figure 4.6: Mock *Euclid*-like spectra of type 2 AGN at the EWS limits simulated with FastSpec. The red line shows the *Euclid*-like spectra, while the black dotted line shows the incident one.

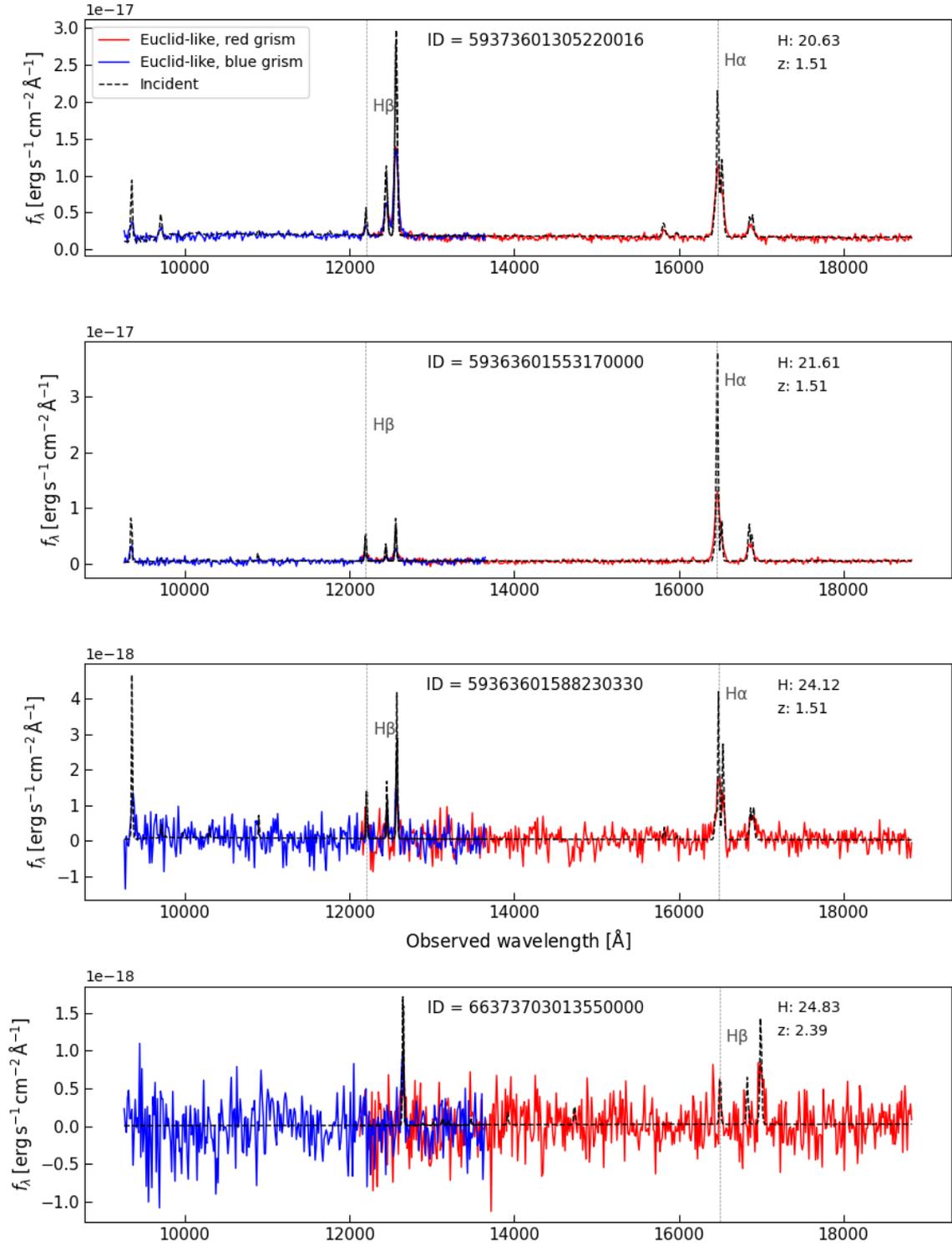


Figure 4.7: Mock *Euclid*-like spectra of type 2 AGN at the EDS limits simulated with FastSpec. The red and blue lines show the *Euclid*-like spectra, while the black dotted line shows the incident one

We asked OU-SPE to analyse the incident and the *Euclid*-like SEDs produced with FastSpec. For this, they use two different methods. On one side, they perform direct integration of the lines, which is model-independent and provides a measurement of the total flux under one line, or under the combination of several lines if they are blended (like $H\alpha$ and $[N II]$). On the other side, they model the emission lines with a multi-Gaussian fit, which provides deblended fluxes for each line, as well as measurements of EW, FWHM, SNR and position (with errors). The results presented below have been derived using the Gaussian fitting model.

In Figs 4.8 and 4.9, we compare the intrinsic flux of the lines of type 2 AGN (that is, the one from the MAMBO catalogue) with the one measured by OU-SPE, for the EDS and the EWS respectively. We show the flux measurements both on the incident and on the *Euclid*-like spectra.

In both figures, we observe that the fluxes of $H\alpha$ and $H\beta$ measured from OU-SPE tend to be underestimated with respect to the intrinsic one by approximately 0.1 dex, also when measured on the incident spectra. While we are still investigating the cause of this issue, a major contribution to this offset might come from the fact that the pipeline to measure line fluxes does not take into account absorption caused by intervening gas clouds in the line of sight. This absorption, on the other hand, is included in the MAMBO spectra through the [Bruzual & Charlot \(2003\)](#) galaxy templates. This hypothesis is supported by the fact that the bias is significantly lower for the $[N II]$ and $[O III]$ lines, for which the intrinsic absorption is absent.

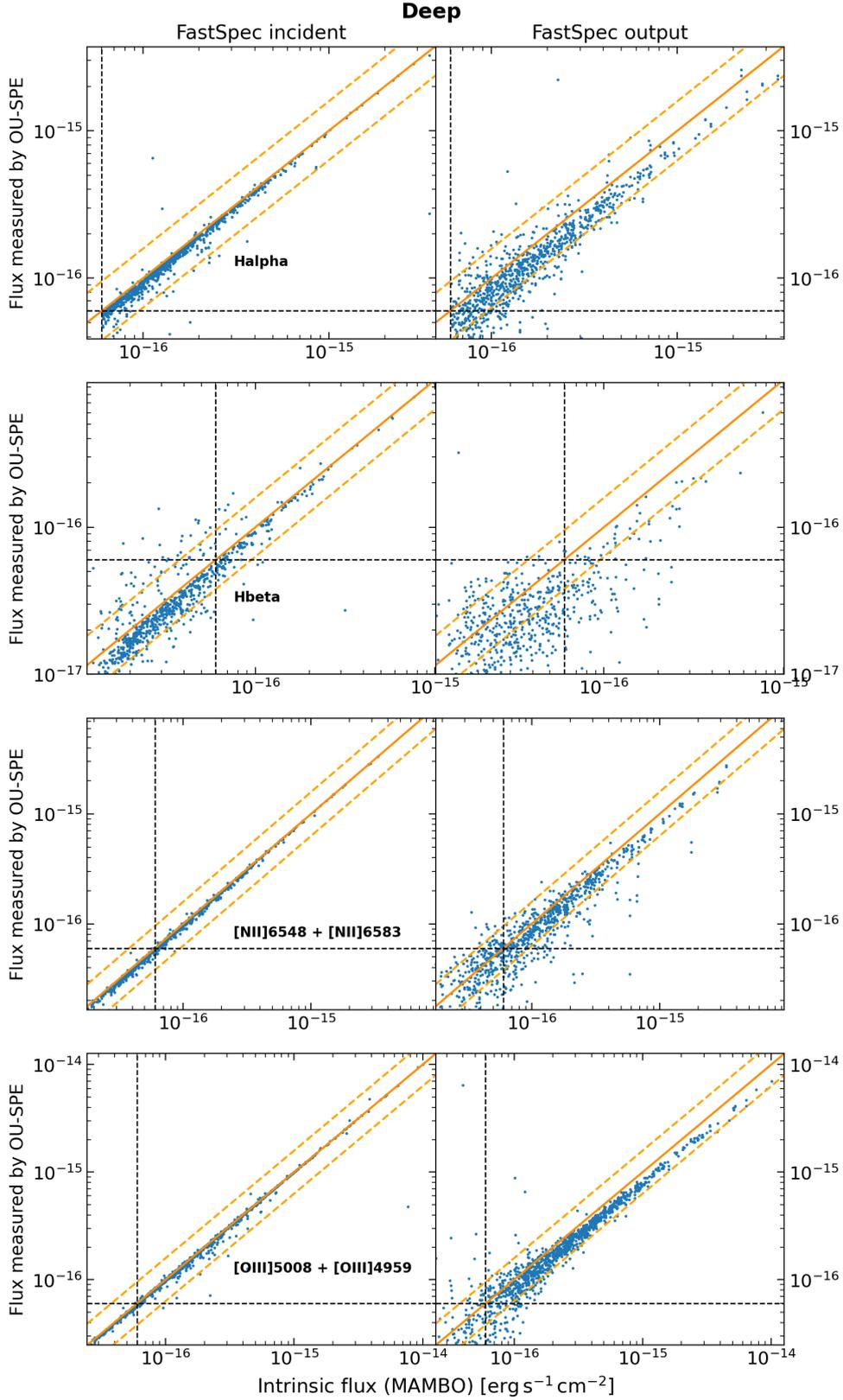


Figure 4.8: Emission line flux of type 2 AGN as measured from OU-SEP against the intrinsic flux from the MAMBO catalogue. The left columns show the flux measured on the incident spectra while the right columns show the flux measured on the *Euclid*-like spectra produced with FastSpec and assuming EDS observing conditions. The solid orange line represents $y = x$, and the dashed orange lines mark $y = x \pm 0.2$ dex. We mark with dashed lines the 3.5σ limiting flux of the EDS, that is, $F = 6 \times 10^{-17} \text{erg s}^{-1} \text{cm}^{-2}$.

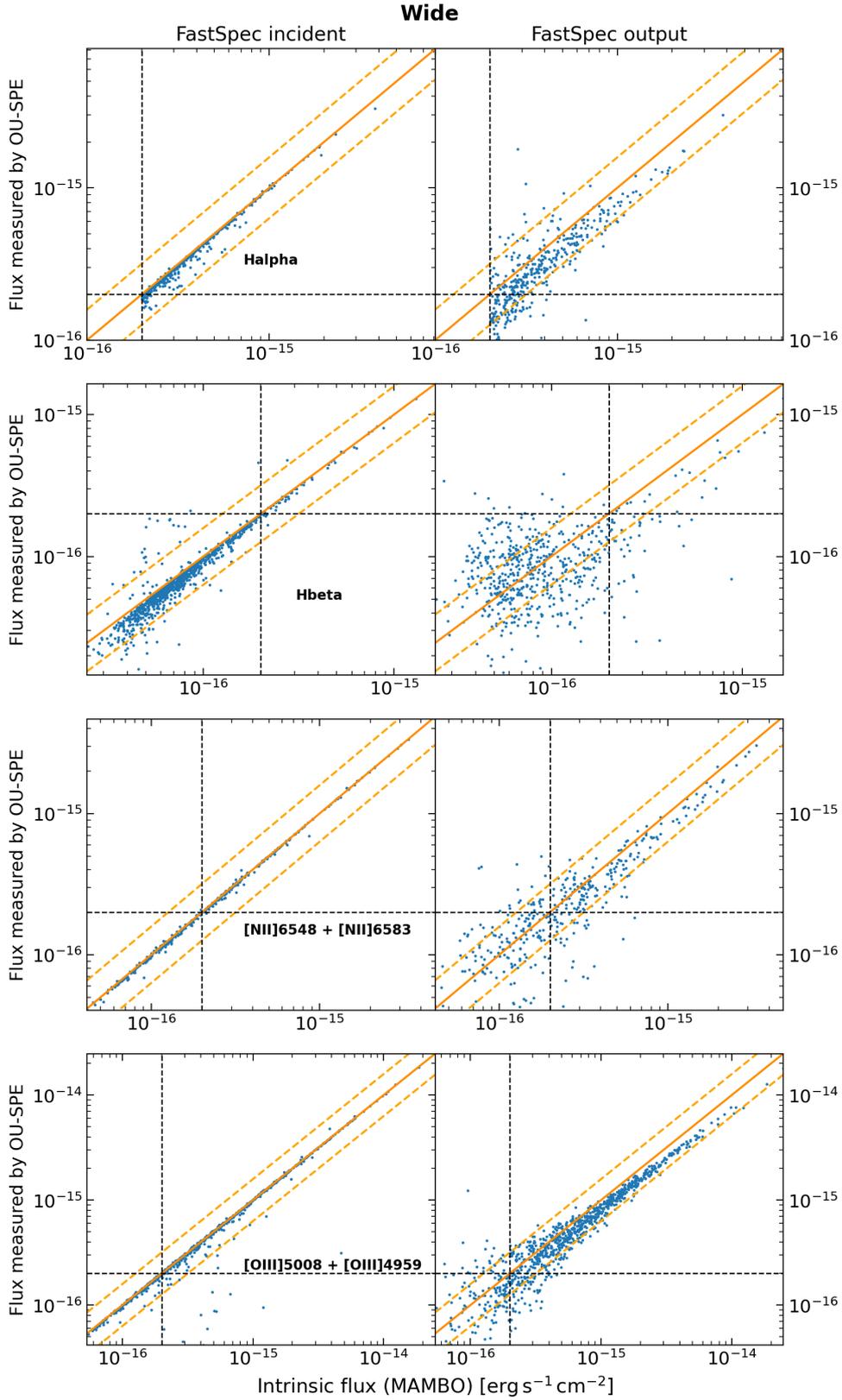


Figure 4.9: Emission line flux of type 2 AGN as measured from OU-SEP against the intrinsic flux from the MAMBO catalogue. The left columns show the flux measured on the incident spectra while the right columns show the flux measured on the *Euclid*-like spectra produced with FastSpec and assuming EWS observing conditions. The solid orange line represents $y = x$, and the dashed orange lines mark $y = x \pm 0.2$ dex. We mark with dashed lines the 3.5σ limiting flux of the EWS, that is, $F = 2 \times 10^{-16} \text{erg s}^{-1} \text{cm}^{-2}$.

In Fig. 4.11 we show the SNR of different emission lines as measured by OU-SPE on the *Euclid*-like spectra as a function of the line flux, for the EDS and EWS simulations. For the Gaussian fit model, the SNR is calculated with the following equation:

$$\text{SNR} = \frac{F_{2\sigma}}{\text{mean_err} \times \text{disp} \times \sqrt{N_{\text{pix}}}}, \quad (26)$$

where $F_{2\sigma}$ is the line flux measured within 2σ of the Gaussian fit, `mean_err` is the mean error (see Fig. 4.10 for an example error spectrum) of the flux within that 2σ window, N_{pix} is the size of the window in pixels, and `disp` is the dispersion element (sometimes also called resolution element), which corresponds to 13.4 \AA for the NISP instrument. In the cases where several lines are blended, such as $\text{H}\alpha$ with the $[\text{N II}]$ doublet, $F_{2\sigma}$ represents the total flux, that is, the sum of the fluxes of all the blended lines. We note that in order to run FastSpec, the incident spectra must have an associated error, which we assumed to be equal to 1% of the incident flux.

Looking at Fig. 4.11, we observe that for all the studied lines, the SNR of the *Euclid*-like spectra emission lines at the limiting flux of each survey, as measured by OU-SPE, is equal to or larger than 3.5, which agrees with the expected performance of the NISP instrument (Sect. 1.3.1). This serves both as a validation of FastSpec and of the OU-SPE spectral analysis pipeline.

Finally, we would like to point out that these spectra have been distributed within the Euclid Consortium to be used at different projects. For example they have been used to train machine-learning algorithms to differentiate SF galaxies from AGN based on their spectra.

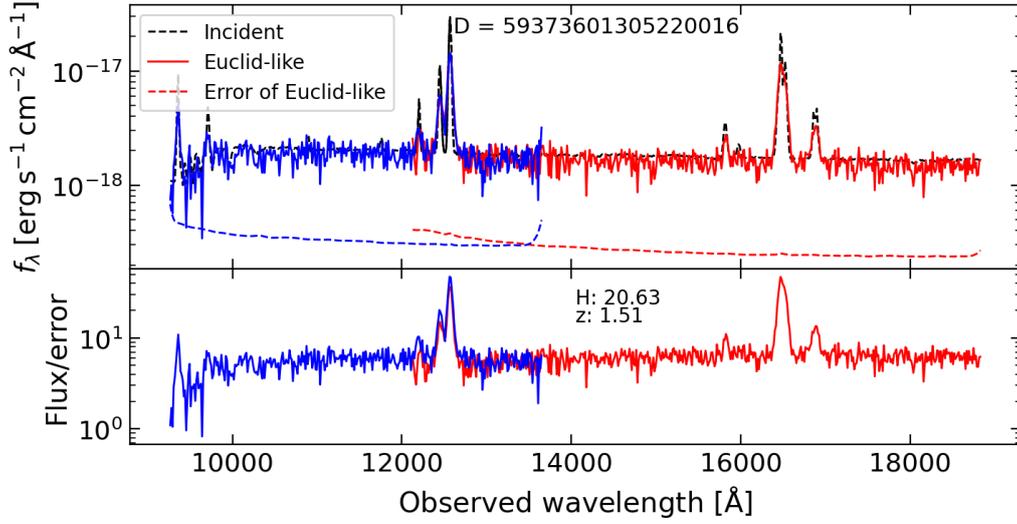


Figure 4.10: Mock *Euclid*-like spectra of a type 2 AGN at the EDS limits simulated with FastSpec. The red and blue lines show the *Euclid*-like spectra, while the black dotted line shows the incident one. The dashed blue and red lines show the error spectrum produced by FastSpec. In the lower panel, we show the ratio of the *Euclid*-like flux and its error spectrum. We do not show in this figure the error of the incident flux, which is 1% of the flux.

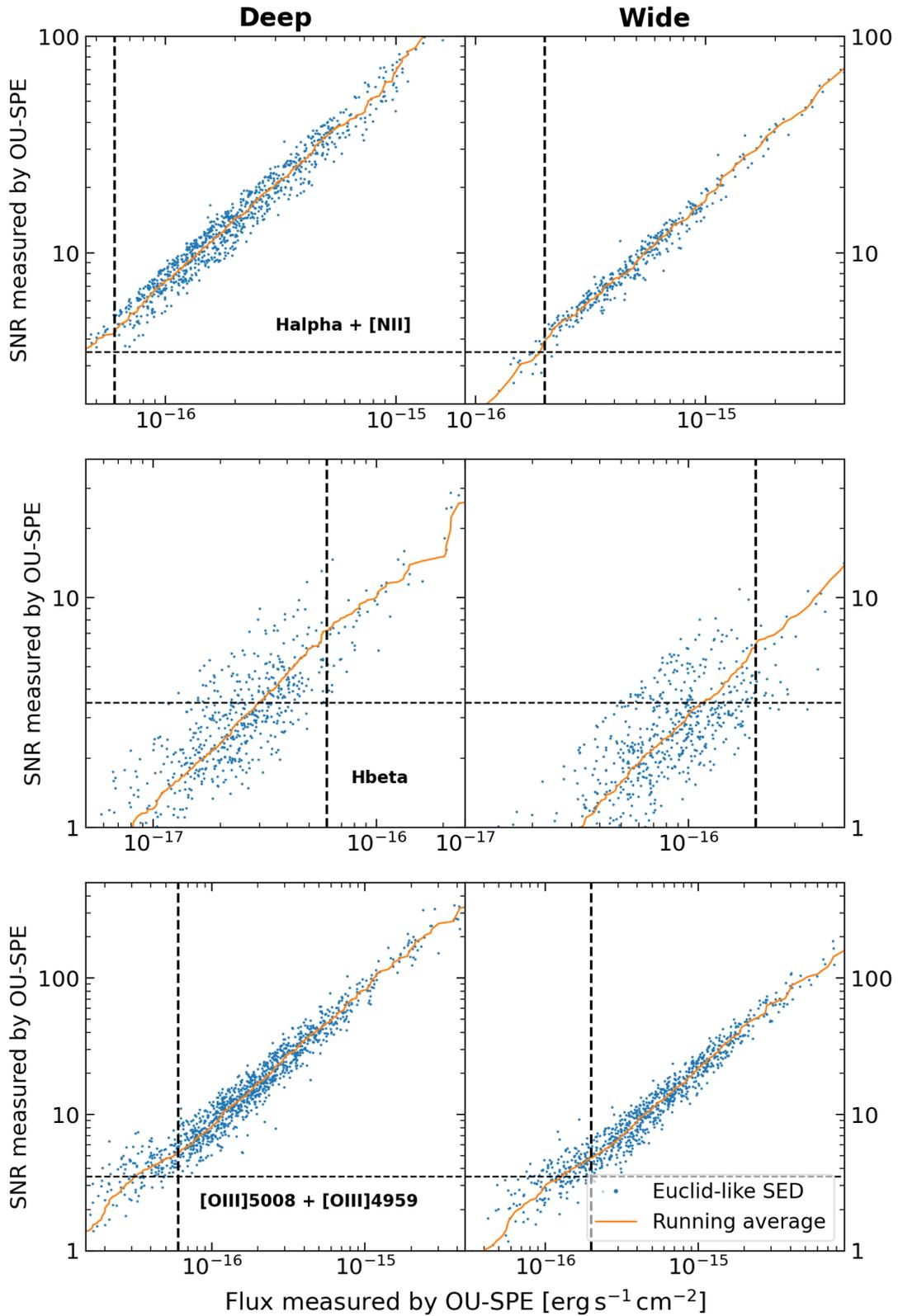


Figure 4.11: Signal-to-noise ratio of different lines as a function of line flux. We show the results only for the SNR measured on the *Euclid*-like spectra. The left column corresponds to spectra generated assuming EDS conditions, while the right column shows the equivalent for the EWS. Vertical dashed lines mark the limiting flux of each survey, while horizontal dash lines mark the threshold $SNR = 3.5$. The orange line shows the running average of the points.

Narrow-line diagnostic diagrams

In this section, we study the instrumental and pipeline-related effects that can have a significant impact when constructing narrow-line diagnostic diagrams with *Euclid* spectroscopic data. Specifically, we focused on studying the BPT diagram (Baldwin et al., 1981). All the BPT diagrams showed in this section have been constructed using narrow lines of type 2 AGN, that is, they do not include star-forming galaxies nor type 1 AGN.

We show in Fig. 4.12 the BPT diagram for the redshift range $1.57 < z < 1.81$. Within this range, the four relevant emission lines will be observed with the NISP instrument’s red grism, making it possible to study this diagram at the EWS limits.

In the top panel, we constructed the BPT diagram using the intrinsic flux of the lines, that is, the one from the MAMBO catalogue. We selected sources based solely on their redshift, which left us with a sample of 350 AGN. As expected, sources whose lines are dominated by the AGN component are properly selected in this diagram.

The middle panel shows the same diagram, but using the fluxes measured by OUSPE on the incident FastSpec spectra. In this case, we selected sources by applying a criterion that required all four relevant emission lines to have SNR values greater than 3.5. We see that even in this idealised case, in which the fluxes are measured on the intrinsic spectra, not all the sample (but 84% of it) is represented in the BPT diagram. This number can be understood as an estimate of how many sources would be missed in this diagram because of the limitations of the *Euclid* spectral analysis pipeline when measuring AGN lines. As is visible from Fig 4.11, the flux measurement over the $H\beta$ line is the main cause for missing such a large fraction of sources in this diagram, since for this line, a significant fraction of the sources fail to meet the SNR requirements. Indeed, if we relax the selection criterion so that all the lines except for $H\beta$ are above the SNR threshold, 99.1% of the sources are represented in this diagram.

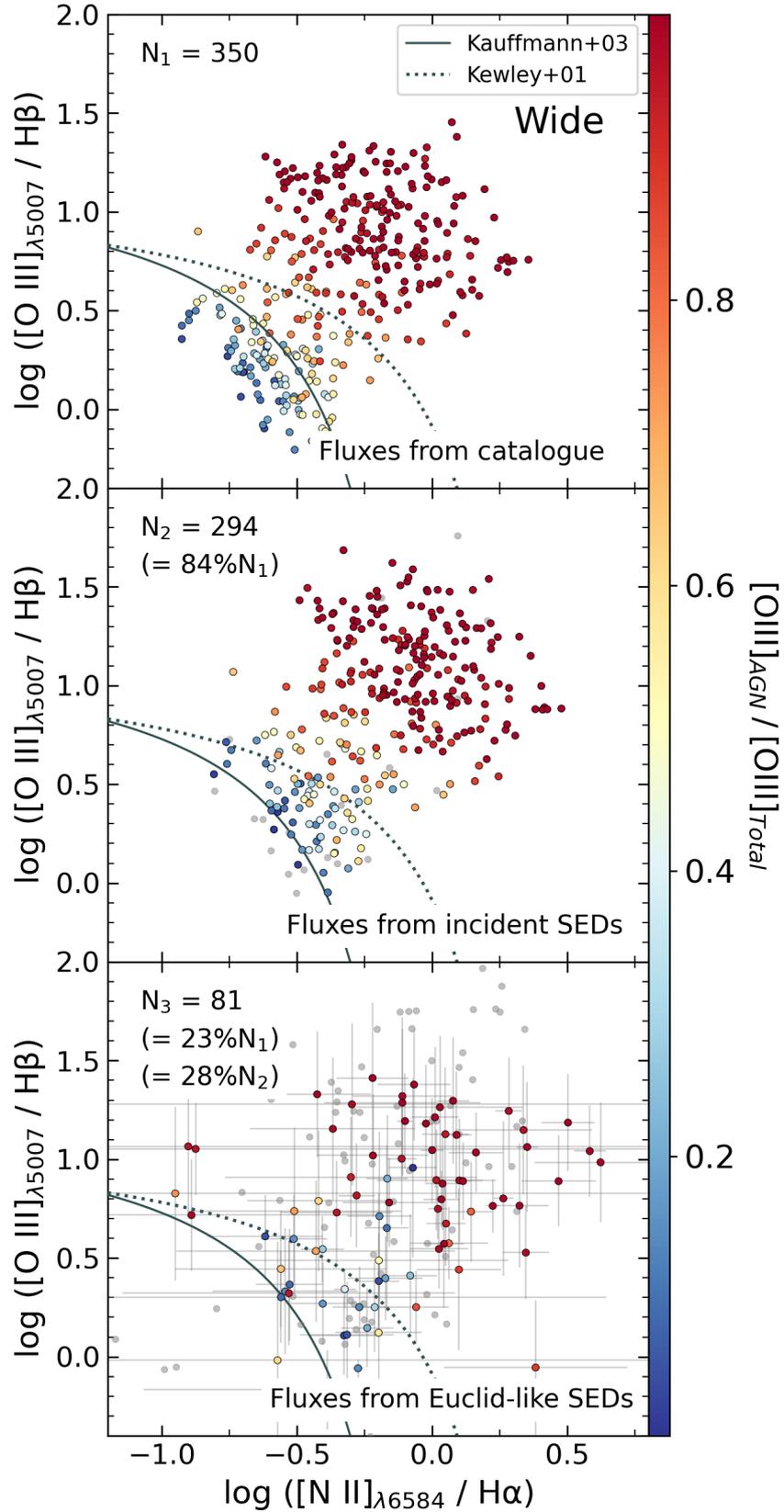


Figure 4.12: BPT diagram at the EWS limits using fluxes from different origins. We show in each panel the number of AGN corresponding to each selection. The AGN are colour-coded from blue to red according to the ratio of the intrinsic emission line flux of [O III] from the AGN component with respect to the total (AGN + host galaxy). We show with grey circles AGN that do not meet the SNR threshold described in the text.

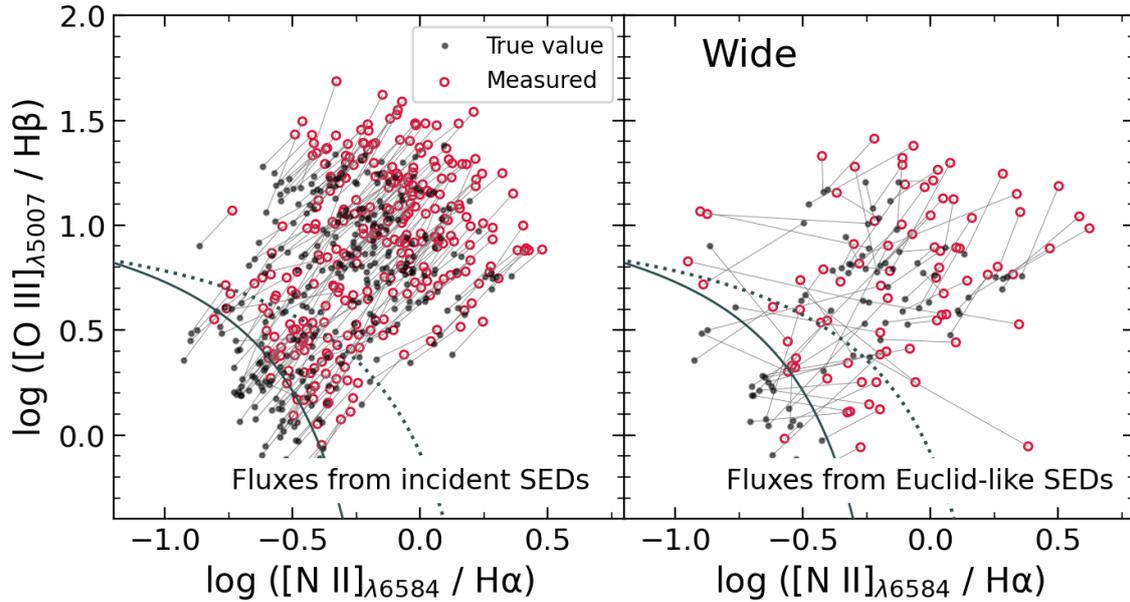


Figure 4.13: Offset of position in the BPT diagram with respect to using intrinsic (MAMBO) fluxes (black dots), when using fluxes measured either on FastSpec incident (left panel, red circles) or output (right panel) spectra. Assuming EWS conditions.

On the other hand, since the pipeline tends to underestimate the flux of the hydrogen lines by a bigger factor than for the [N II] and [O III] lines (as shown in Fig. 4.9), the AGN in this diagram exhibit artificially enhanced metallicity with respect to the intrinsic values. This shift is more visible in Fig 4.13, where we show the shift of the position in the BPT diagram for the measured fluxes with respect to the intrinsic ones. In Fig 4.13, it is visible that on average, this effect can lead to the misclassification of SF galaxies as AGN.

In the third panel, we present the same diagram but using the line fluxes measured by OU-SPE on the *Euclid*-like spectra. As earlier, we selected only sources with $\text{SNR} > 3.5$ measurements on all four lines. As a result, we see that, in this case, only 23% of the original sample is represented in the final BPT diagram. Again, if we relax our selection criteria not to consider the SNR of $H\beta$, the number of sources increases to 147, that is, half of the original sample.

In Fig. 4.14 we present the results for the spectra generated assuming EDS limits. In this case, the redshift range for which it will be possible to study the BPT diagram corresponds to $0.89 < z < 1.81$, and therefore all the samples in this figure are selected within this redshift. This results in an initial sample of 698 AGN.

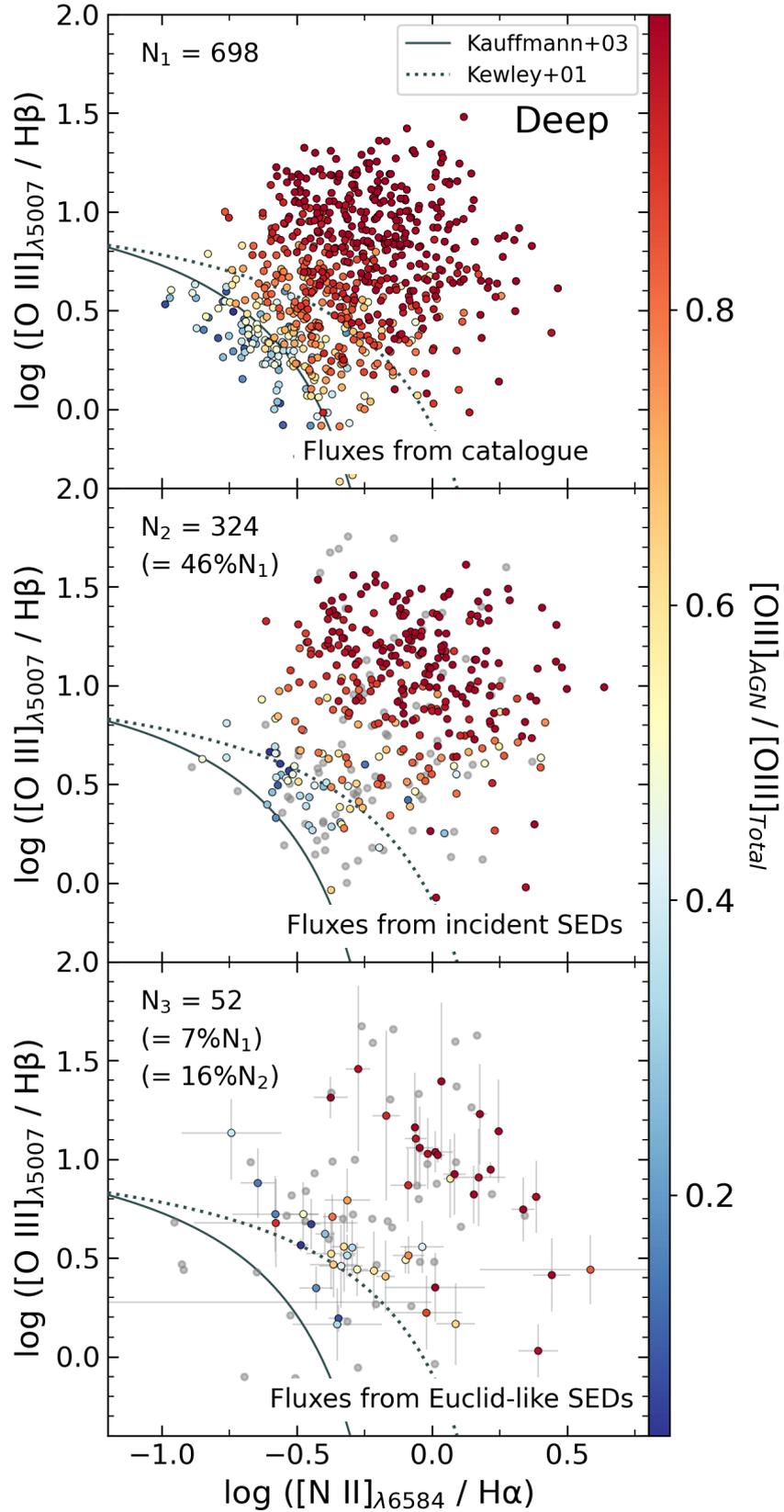


Figure 4.14: BPT diagram at the EDS limits using fluxes from different origins. We show in each panel the number of AGN corresponding to each selection. The AGN are colour-coded from blue to red according to the ratio of the intrinsic emission line flux of [O III] from the AGN component with respect to the total (AGN + host galaxy). We show with grey circles AGN that do not meet the SNR threshold described in the text.

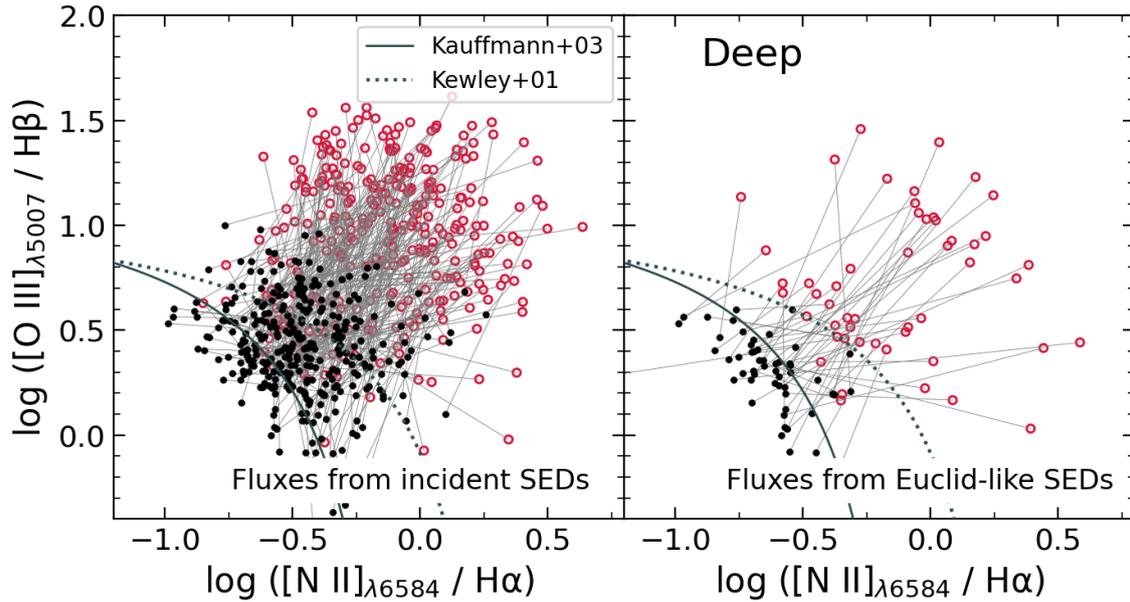


Figure 4.15: Offset of position in the BPT diagram with respect to using intrinsic (MAMBO) fluxes (black dots), when using fluxes measured either on FastSpec incident (left panel, red circles) or output (right panel) spectra. Assuming EDS conditions.

In the EDS, each source will be observed using the red and the blue grism. Therefore, some lines could be resolved by one of the two grisms, by both of them, or by none (see Fig. 4.7). Hence, to construct the BPT diagram with the FastSpec incident and output spectra, we selected sources with $\text{SNR} > 3.5$ for the four lines, using the line flux from the grism with the higher SNR for each line. In other words, we discarded sources where the SNR requirement was not met by at least one of the lines at both grisms simultaneously.

As it is visible in the middle panel of Fig. 4.14, in this case only 46% of the incident spectra have line measurements with $\text{SNR} > 3.5$. However, this fraction increases to 99.6% if we impose the SNR threshold in all the lines except for $\text{H}\beta$. Besides, as it was the case for the EWS, the line ratios appear artificially high on this diagram (see also left panel of Fig. 4.15). Finally, from the bottom panel of this figure, we see that when using the fluxes measured on the *Euclid*-like spectra at EDS limits, only 7% of the original sample fulfils the SNR requirements, while this number increases to 89% if we do not impose any SNR limit on $\text{H}\beta$.

The exact cause of this significant effect of the $\text{H}\beta$ line measurement on the final selected sample is not clear to us and requires further investigation. Given that the

effect manifests in the incident spectra already, it is improbable that FastSpec plays a significant role in this issue. Part of the cause may rely on the methodology used in constructing the MAMBO mock type 2 AGN spectra: the $H\beta$ line has two stellar contributions (one in emission and one in absorption, which dominates in some cases over the former) plus the nuclear (AGN) contribution. The combination of these three components can potentially lead to cases where the final shape of the line departs significantly from a Gaussian shape. Besides, as mentioned previously, the OU-SPE spectral analysis pipeline does not take into account the absorption component, which can be significant for $H\beta$. This analysis should be repeated starting with a different set of mock spectra, to determine if the main cause of this issue lies in the characteristics of our simulated spectra, or if it is related with the spectroscopic analysis pipeline.

One important last remark is that this analysis was performed assuming that we knew with perfect accuracy the true redshift of every source. In a realistic case, the redshift would be measured by OU-SPE on the real *Euclid* spectra as a previous step to the rest of the spectral analysis. On the other hand, using the true redshift as an additional parameter allows a better estimation of the emission-line properties. Therefore, in that sense, the results here presented suppose a best-case scenario.

In conclusion, this exercise provides valuable insights into the fraction of AGN that could be missed or misclassified in a BPT diagram using *Euclid* data. However, there are some important limitations to consider. First, the study needs to be repeated with a larger sample size to ensure statistical significance. Additionally, the real *Euclid* spectra will be even more complex than those simulated with MAMBO and FastSpec. Despite these caveats, the results offer a useful first glimpse into the potential challenges of AGN spectral identification with *Euclid* data. Currently, similar studies focusing solely on the spectral analysis of non-active galaxies are being performed within the Euclid Consortium (e.g. Cassata et al. in prep, Mancini et al. in prep.).

4.3 Flagship 2 + MAMBO AGN

For the next part of the study, we required a larger sample of AGN to improve the statistical significance. Therefore, we applied the methodology outlined in this thesis to 30.17 deg² of the Flagship 2 simulation (FS2).

Flagship 2 ([Euclid Collaboration: Castander et al., 2024](#)) is the official galaxy mock of the *Euclid* mission, comprising billions of galaxies over more than ten billion years of cosmic time. The Flagship N -body DM simulation has a box size of $3600 h^{-1}\text{Mpc}$ on a side with $16\,000^3$ particles, leading to a particle mass of $m_p = 10^9 h^{-1} M_\odot$, which allows to resolve $10^{11} h^{-1} M_\odot$ DM haloes, which host the faintest galaxies Euclid will observe. The halo catalogue was computed directly on the DM lightcone particle data using the ROCKSTAR halo finder ([Behroozi et al., 2013](#)).

Dark matter haloes were populated with galaxies using a prescription ([Carretero et al., 2015](#)) that includes HOD and AM techniques, together with observed scale relations between galaxy properties. Haloes were populated with central and satellite galaxies by computing the number of satellites in each halo and assigning the galaxy luminosities. Galaxy clustering measurements were used to determine the parameters and the relations implemented. We refer to ([Euclid Collaboration: Castander et al., 2024](#)) for a more detailed description of the FS2 simulation.

Using the methodology presented in this thesis (see Sec. 2.2), we populated with AGN a sample of galaxies selected from a 30.17 deg² FS2 lightcone assuming EDS magnitude limits. The lightcone spans to $z = 3$. For the rest of this work, we refer to this sample as the FS2+MAMBO catalogue. Similar to when we constructed the Millennium galaxy and AGN lightcone, the only parameters from FS2 needed to build the galaxy+AGN catalogue are the galaxy stellar mass, redshift, and, optionally, the sky coordinates of each galaxy.

4.4 AGN photo-z in Euclid

In this part of the work, we aim to assess the accuracy of recovering photometric redshifts from a sample of galaxies and AGN using an official *Euclid* algorithm. For this, we simulated *Euclid*-like magnitude measurements (with photometric noise added to the intrinsic magnitudes) and run a template-fitting code (Phosphoros, see below) on them. We evaluated the reliability of the results for different subsamples of galaxies and AGN, selected based on different constraints in physical parameters and photometric SNR.

A similar analysis, using a version of MAMBO prior to the inclusion of AGN, has been performed in [Euclid Collaboration: Enia et al. \(2024\)](#) to forecast the reliability of different algorithms at recovering redshifts and physical parameters (stellar masses and star-formation rates) from *Euclid* photometry.

Photometric redshifts

Retrieving reliable redshifts for the billions of sources that *Euclid*, *Roman*, Rubin, and other large surveys will observe is a crucial prerequisite for achieving their scientific objectives. In the case of *Euclid*, accurate and precise redshifts are a fundamental quantity to study the large-scale structure of the universe by mapping the distribution of galaxies and dark matter across cosmic time and, therefore, fulfil its main cosmological goals.

On the other hand, reliable redshifts are also fundamental to conducting galaxy and AGN evolution studies. Redshift estimates, combined with spectral energy distributions observations, are often used to infer the intrinsic physical properties of galaxies and AGN (such as their stellar mass, intrinsic luminosity, dust content, star-formation rate, etc.). Significant inaccuracies in redshift measurements can lead to substantial errors in these properties.

Generally, the most precise estimates of galaxy redshifts are determined through spectroscopic measurements. The galaxy’s spectrum is observed and recorded, and specific spectral lines – typically emission lines, although absorption lines can be used in the case of quiescent galaxies – are identified and measured. The redshift is calculated by comparing the observed wavelengths of these lines with their known

rest-frame values.

In some situations, though, measuring spectroscopic redshifts is not feasible for several possible reasons. If the number of sources is too high, or if they are too faint to perform reliable spectroscopy, redshifts can also be estimated from broadband photometry alone, using broad features such as the Lyman and Balmer breaks, strong emission and absorption lines, or the overall shape of the continuum of the spectrum. In this case, they are called photometric redshifts (photo- z s, see [Salvato et al. 2019](#) and [Newman & Gruen 2022](#) for reviews on the topic). Photo- z s can be measured much faster than their spectroscopic counterparts (spec- z s), and they can be derived simultaneously for all sources identified in an imaging survey. Additionally, photo- z s are often used to identify rare objects (such as high-redshift galaxies or AGN), for which it can be interesting to perform follow-up spectroscopic analysis. The downside of photo- z s in their precision, typically a factor of 10–100 times lower than spec- z s obtained with a low-resolution spectrograph ([Ilbert et al., 2009](#)).

These characteristics of photo- z s make them particularly advantageous in the case of future large surveys. *Euclid*, for example, will obtain high-resolution and high signal-to-noise images for billions of sources, for which a photometric redshifts will be derived. Spectroscopic redshifts, instead, will only be derived for the brightest sources, with emission lines brighter than 2×10^{-16} erg s $^{-1}$ cm $^{-2}$ (mainly using the H α line, which will be detected by *Euclid*'s red grism in the redshift range $0.9 < z < 1.8$). Besides, these sources will be preselected above a brightness threshold. For example, for the first quick data release (Q1), only the spectra of sources brighter than $H_E = 22.5$ will be extracted.

Phosphoros

One of the main methods to derive photo- z s is by means of template-fitting algorithms. These methods use theoretical or empirical precomputed models (templates) of SEDs across a range of redshifts and physical properties to estimate the photo- z by comparing observed photometric data in given filters with these templates. Among these codes, Phosphoros²¹ ([Paltani et al., in prep.](#)) is a fully Bayesian template-fitting code which was developed for use within the Euclid photometric redshift pipeline, and which is expected to be supported throughout the mission's lifetime. The code builds

²¹<https://phosphoros.readthedocs.io/en/latest/install/index.html>

upon previous well known template-fitting algorithms such as Le Phare (Arnouts et al., 1999; Ilbert et al., 2006), HyperZ (Bolzonella et al., 2000), or EAZY (Brammer et al., 2008), adapting its recipes to be used within the Euclid photometric pipeline. It is written and maintained by the Euclid Swiss Science Data Centre.

The method consists of two steps. During the first step, which has to be performed only once, Phosphoros constructs a grid of model photometry, which consists on the integrated flux at each selected band, covering all possible model parameters. These parameters are the redshift z , the rest-frame SED template, the IGM attenuation, and the intrinsic interstellar dust absorption, modelled by two free parameters: the colour excess E_{B-V} and the reddening curve $k(\lambda)$.

All the results presented in this section have been obtained using one of the two following sets of templates to construct the model photometry grid. On the one hand, we used a library of 31 pure galaxy SEDs, including spiral, elliptical and young blue star-forming galaxies. This same set of templates have been previously used to derived photo- z s on the COSMOS survey (Ilbert et al., 2013; Laigle et al., 2016).

On the other hand, we used a library of AGN templates which includes a mixture of SEDs from single objects (totally dominated by either stellar or nuclear emission, and covering different classes of AGN, such as Seyfert 1 and 2, QSO 1 and 2, etc.) and from hybrids (objects with different contributions of host and nuclear components). This library is a compilation of several empirical and non-empirical templates from different works and is extensively described in Appendix B.1 of Salvato et al. (2022). We show in Fig. 4.16 some examples of SEDs from this library.

In both cases, we derived photo- z s from these templates assuming possible solutions within $0 < z < 6$ (in steps of 0.01) and E_{B-V} values from 0 to 0.5 in steps of 0.1. For the galaxy templates three attenuation laws were considered: SMC (Prevot et al., 1984), Calzetti (Calzetti et al., 2000), and a modified Calzetti’s law including a ”bump” at 2175 Å (Fitzpatrick & Massa, 1986). Instead, for the AGN templates, only the SMC extinction law was considered.

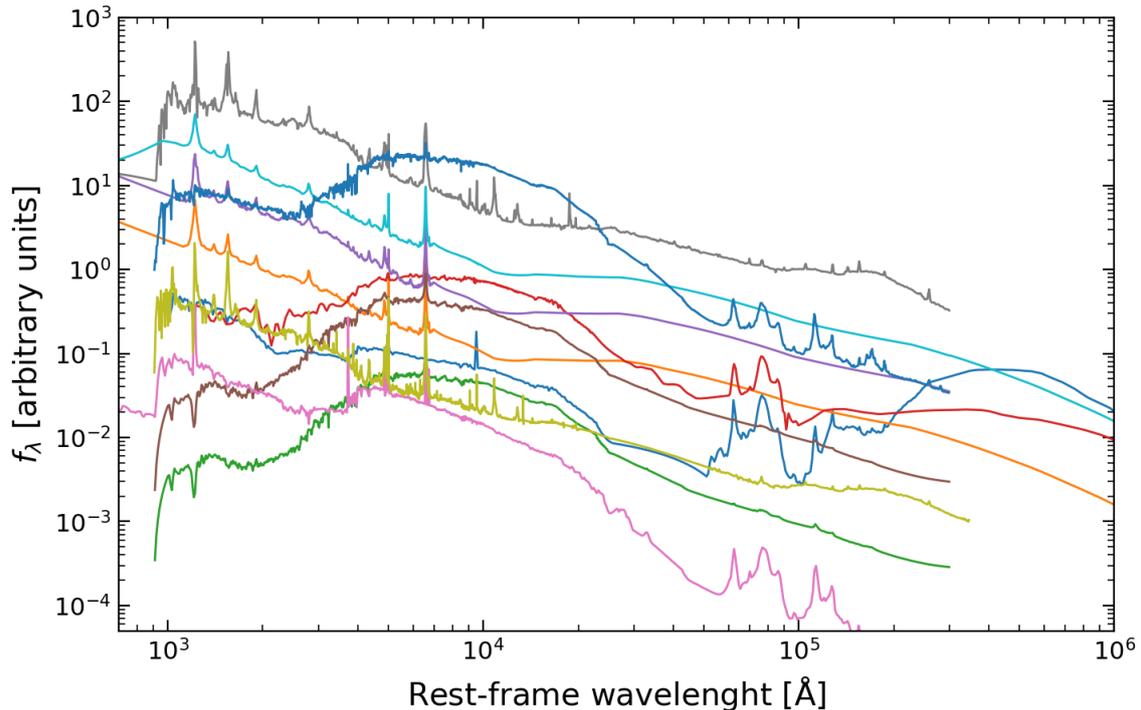


Figure 4.16: Some examples of SEDs from the AGN template library described in [Salvato et al. \(2022\)](#) and used in this work. All templates have been arbitrarily rescaled for a better visualisation.

During the second step of the method, Phosphoros computes, for each source of the input catalogue, the likelihood \mathcal{L} distribution that the observed photometry is described by a given model m . If we name f_{obs}^i the observed flux at the filter i , with associated uncertainty σ^i , and f_m^i the model flux in the same band, the likelihood can be calculated as

$$\ln(\mathcal{L}) = -\frac{\chi^2}{2} = -\frac{1}{2} \sum_i \left(\frac{f_{obs}^i - \alpha f_m^i}{\sigma^i} \right)^2, \quad (27)$$

where χ^2 is a measure of the distance between the observed and the model flux values and α is a normalisation factor which acts as an additional free parameter of the model.

Additionally, Bayesian inference allows the inclusion of additional information on some model parameters, acquired during previous observations (e.g. the redshift distribution of a given type of galaxy or AGN, their luminosity function, the relation between galaxy mass and star formation rate, etc). In practice, this extra empirical

information is included through the use of priors (denoted as \mathcal{P}), which can improve significantly the accuracy of the inferred photo- z (see e.g. [Salvato et al., 2019](#)).

Following Bayes’s theorem, the posterior probability distribution $p(m | F, \mathcal{P})$, i.e., the probability that the observed photometry F is described by a model m given the prior information \mathcal{P} , is given by

$$p(m | F, \mathcal{P}) \propto \mathcal{L}(F | m)\mathcal{P}(m). \quad (28)$$

Unless stated otherwise, all the results presented in this section have been obtained assuming a volume prior in Phosphoros. The volume prior takes into account the fact that the volume sampled by a survey increases with redshift. Therefore, this prior increases the likelihood of sources being at higher redshifts and disfavours low-redshift solutions, where the volume sampled is smaller. In Phosphoros, the volume prior is proportional to the redshift-dependent differential comoving volume, i.e. $\mathcal{P}(z) \propto dV_C(z)$.

Phosphoros returns two main outputs: the best-model solution, which is found by minimising χ^2 , and the full redshift probability density function, $p(z | F, \mathcal{P})$, calculated as explained above.

Sample and methodology

Starting from the FS2+MAMBO catalogue (see Sec 4.3), we selected sources brighter than $m_H = 24.25$, that is, the $\text{SNR} = 5$ limiting magnitude of the EWS (Table 1). We randomly selected subsamples, resulting in a final dataset of 56 059 galaxies (random sample), 43 928 type 2 AGN (10% of the total) 26 035 type 1 AGN (100% of the total). These numbers reduce to 32 601 galaxies, 29 795 type 2 AGN and 24 389 Type 1 AGN if we select sources brighter than $m_H = 23.5$ (representing the approximate value of $\text{SNR}=10$ in the EWS). We show in Fig. 4.17 the distributions of stellar mass, redshift and perturbed magnitude (see below) for the $\text{SNR} > 5$ samples.

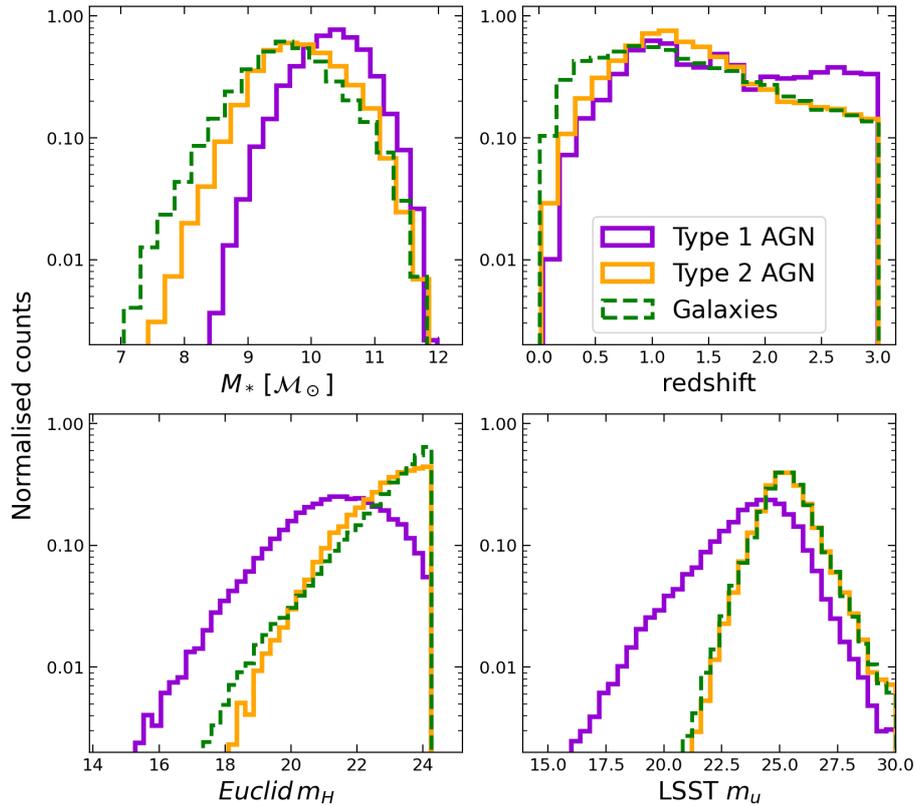


Figure 4.17: Normalised distributions of stellar mass, redshift, and H_E and u -band magnitudes of the sample of galaxies and AGN (with $\text{SNR} > 5$) used as an input for Phosphoros.

By default, the broadband fluxes and magnitudes in the MAMBO catalogue do not include any noise effect and do not have uncertainties associated with them. For this exercise, though, we wanted to simulate *Euclid*-like photometric measurements with their associated uncertainties.

We modelled the magnitude error (Δm) by assuming that it does not depend on the sky background. If a given limiting signal-to-noise ratio ($\text{SNR}_{\text{lim}} = f_{\text{lim}}/\sigma_{\text{lim}}$) is reached at a given limiting magnitude ($m_{\text{lim}} \propto -2.5 \log_{10} f_{\text{lim}}$), we can calculate the flux error at this SNR as:

$$\sigma_{\text{lim}} = \frac{10^{(-m_{\text{lim}}/2.5)}}{\text{SNR}_{\text{lim}}}. \quad (29)$$

We also implemented a systematic error of $\Delta m_{\text{sys}} = 0.005$ magnitudes (see e.g. [Graham et al., 2020](#), Sect. 2.2.2). The total noise is the quadratic sum of all the noise contributions, that is,

$$\sigma_f = \sqrt{\sigma_{\text{lim}}^2 + \sigma_{\text{sys}}^2}. \quad (30)$$

Specifically, we simulated observed magnitudes and uncertainties assuming the conditions of the first data-release of the northern hemisphere part of *Euclid*'s wide survey (NDR1). The limiting *Euclid* magnitudes at $\text{SNR} = 10$ for the NDR1 can be found in Table 1. This table also shows the limiting magnitude of the ground-based photometry that will complement *Euclid*'s NDR1, provided by the UNIONS (Ultraviolet Near-Infrared Optical Northern Survey; [Ibata et al., 2017](#)) collaboration (see footnote of Table 1), and that we also use in this exercise to retrieve photo-zs.

We show in the upper panel of Fig. 4.18 the relation between the perturbed magnitude, calculated as explained above, and the intrinsic magnitude of each source (for the subsample of galaxies), that is, the one from the MAMBO catalogue. In the figure, we marked also the limiting magnitudes corresponding to $\text{SNR} = 10$ and 5. We see that the perturbed magnitude closely aligns with the intrinsic one for $\text{SNR} > 10$, and that it significantly deviates from it for $\text{SNR} \leq 5$.

In the lower panel of Fig. 4.18 we show the relation between the perturbed magnitude and its associated error, where we marked with horizontal lines the magnitude error corresponding to each SNR_{lim} , calculated as in Eq. 29.

To quantify the performance of Phosphoros at retrieving photo-zs from these samples, we used standard metrics, namely, the normalised median absolute deviation (NMAD), the outlier fraction, and the bias. The NMAD is a robust estimator of the scatter in a distribution, and is defined as:

$$\text{NMAD} = 1.48 \times \text{median} \times \frac{|z_{\text{phot}} - z_{\text{true}}|}{1 + z_{\text{true}}} - b, \quad (31)$$

where b is the model bias (see below), z_{phot} is the photometric redshift retrieved by Phosphoros, and z_{true} is true redshift of the source, as from the MAMBO catalogue. The outlier fraction f_{out} is defined as the fraction of catastrophic outliers ([Hildebrandt et al., 2010](#)) over a certain threshold. Specifically, we define it as the fractions of sources for which

$$\frac{|z_{\text{phot}} - z_{\text{true}}|}{1 + z_{\text{true}}} > 0.15. \quad (32)$$

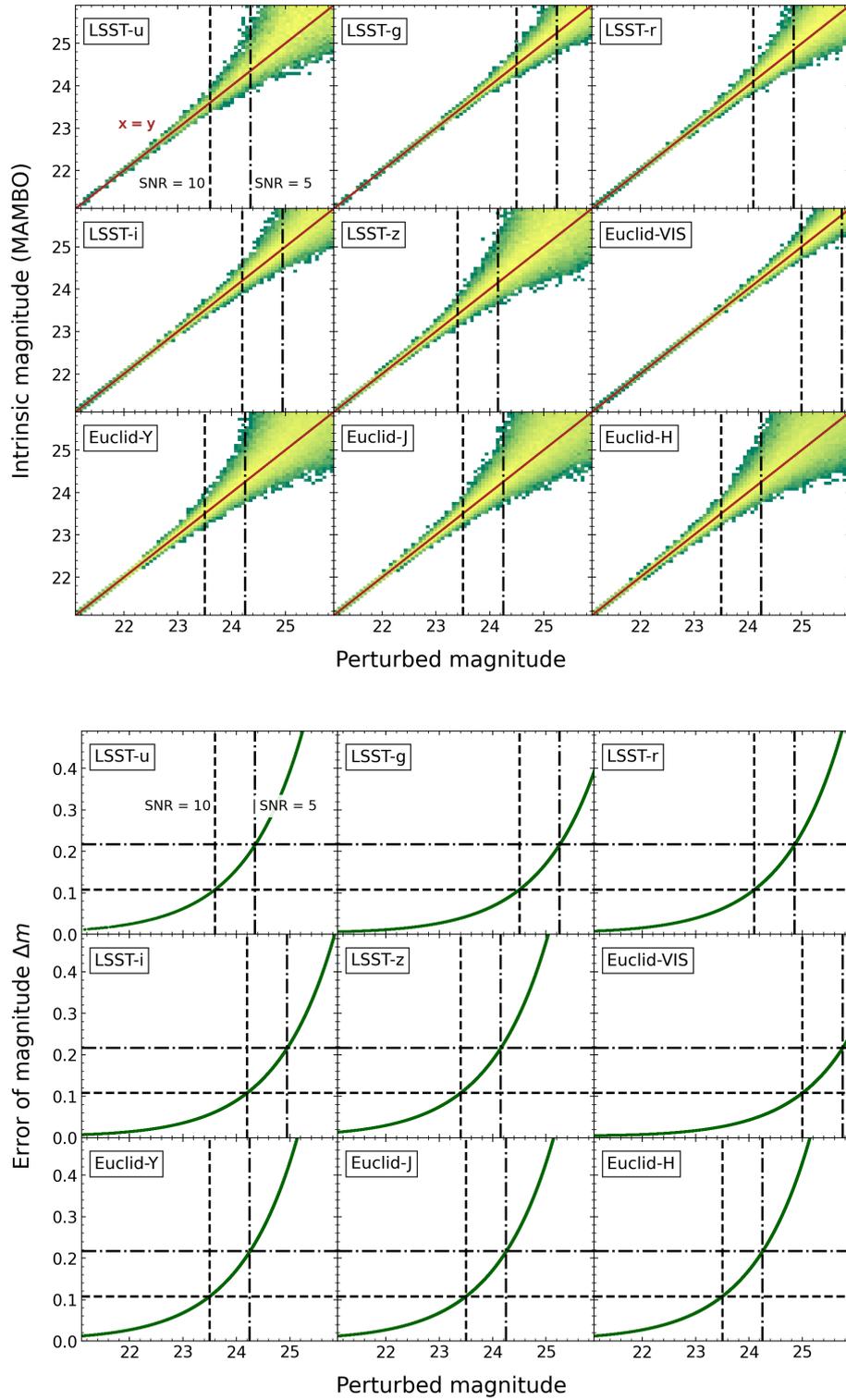


Figure 4.18: Perturbed magnitude against intrinsic magnitude (as from the MAMBO catalogue, upper panel) and against its associated uncertainty (lower panel). The vertical dashed and dash-dotted lines show, respectively, the limiting magnitudes, for each photometric band, at which the SNR = 10 and 5 are reached. The equivalent horizontal lines show the magnitude error which corresponds to each SNR limit, calculated as $\Delta m = 2.5 \ln(10)^{-1} \text{SNR}^{-1}$.

Finally, the model’s bias b measures the systematic offset between the estimated values and the true or reference values, and is defined as:

$$b = \text{median} \left(\frac{z_{\text{phot}} - z_{\text{true}}}{1 + z_{\text{true}}} \right). \quad (33)$$

In all three cases, the closer to zero, the better the predicted values resemble the test ones. Of all three, only the bias can take either positive or negative values. In the case of *Euclid*, achieving its scientific objectives requires measuring photo- z s with a standard deviation of less than $0.05(1 + z)$ and a catastrophic failure rate below 10% (Euclid Collaboration: Mellier et al., 2024).

Results

In the next series of figures, we compare z_{phot} with z_{true} for different subsamples of galaxies and AGN, selected with different SNR limits (either 10 or 5). The number of objects in each sample (N_{obj}), SNR limit, and the quality assessment statistics (NMAD, f_{out} and b) are also shown within each figure. These results have been obtained using the pure galaxy SED library as model templates.

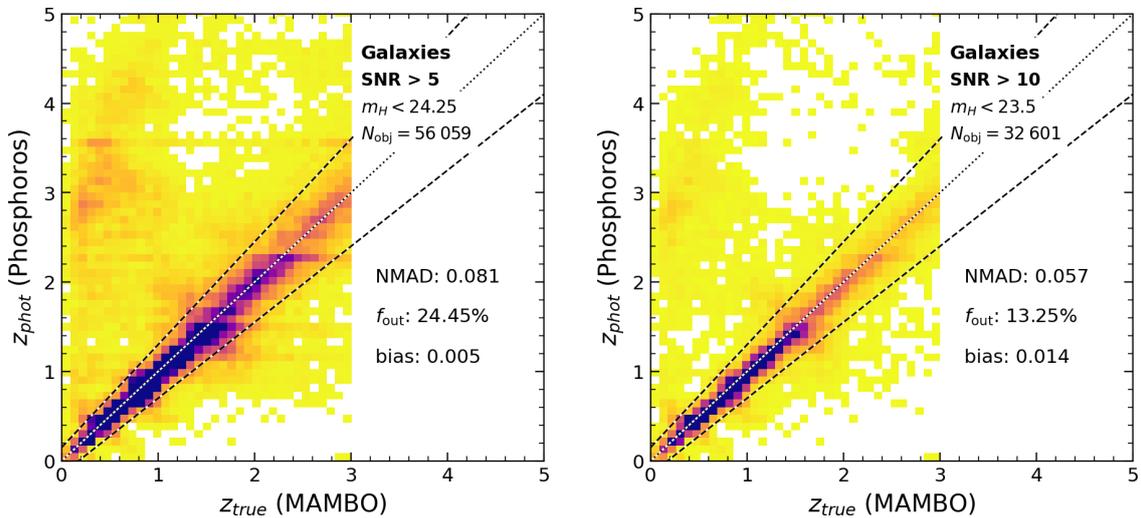


Figure 4.19: Comparison of photometric redshifts from Phosphoros with true redshifts for the galaxy sample. The colour-coding reflects the number of sources per bin. The dotted line shows the one-to-one relation ($z_{\text{phot}} = z_{\text{true}}$), while the dashed lines show the thresholds for the outlier definition. We run Phosphoros with pure galaxy SEDs as model templates.

In Fig. 4.19, we show the results for the galaxy sample. We see that for the subsample with $\text{SNR} > 5$, there is a total of 24.45% of outliers, which decreases to 13.25% for $\text{SNR} > 10$ sources. For type 2 AGN (Fig. 4.20), we observe similar results in terms of the global statistics, with a slightly lower fraction of outliers.

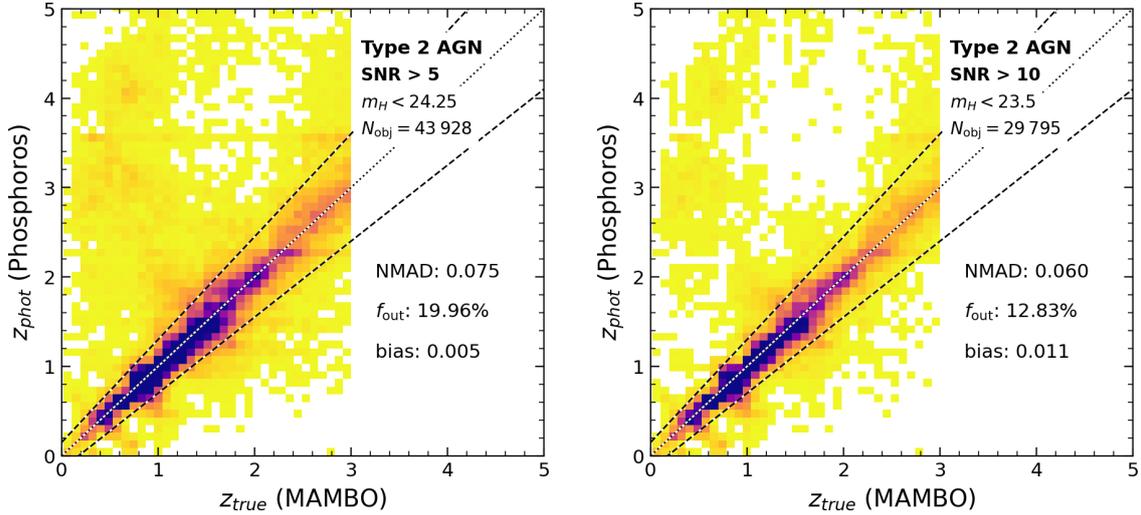


Figure 4.20: Similar to Fig. 4.19, but showing results for type 2 AGN.

While one could expect worse results for the AGN than for the galaxies (since we have used pure galaxy SEDs as model templates), it is worth reminding the reader that the type 2 AGN in the MAMBO catalogue are modelled by adding narrow emission-lines to the SED of normal galaxies. These narrow lines have a minor contribution to the broadband photometry²² of the object. On the other hand, galaxies are not randomly populated with AGN, but with a given probability that depends on redshift and stellar mass. Therefore, type 2 AGN present different distributions of e.g. \mathcal{M} and observed magnitudes (see Fig. 4.17) which can explain the observed difference in performance at recovering true redshifts.

For type 1 AGN (Fig. 4.21), we observe a remarkably high fraction of outliers (47.18%) , which does not improve when considering only $\text{SNR} > 10$ sources. This type of result is not surprising since type 1 AGN exhibit a distinctive continuum emission that differs significantly from the galactic continuum of the model templates.

²²See e.g. Amorín et al. (2015) for a discussion on the effect of emission lines when estimating physical parameters using template-fitting algorithms.

Moreover, the broad and strong lines in type 1 AGN can substantially contribute to the observed broadband magnitudes.

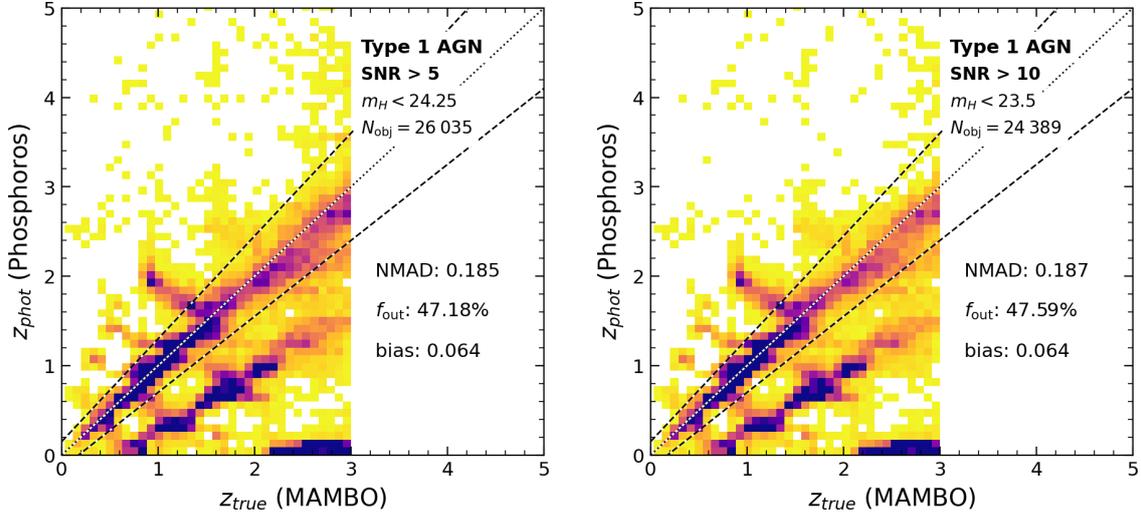


Figure 4.21: Similar to Fig. 4.19, but showing results for type 1 AGN.

To better understand which sources are responsible for the poor performance of the photo- z estimates, we created various subsamples of type 1 AGN, categorized by the contribution of the AGN component to the total emission. For simplicity, in the next figures and for the rest of this section, we show results only for subsamples selected with $\text{SNR} > 5$. The corresponding results for the $\text{SNR} > 10$ samples can be found in Table 4.

In Fig. 4.22, we show the results of the photo- z for four different subsamples of type 1 AGN, selected either above or below a threshold of 0.5 in the quantities $f_{\text{AGN}_{FUV}}$ and f_{AGN_H} (see Eq. 17). We see that for the sample where the AGN contribution dominates in the observed H band wavelength range ($f_{\text{AGN}_H} > 0.5$), the fraction of outliers reaches 95.99%.

Instead, the subsamples dominated by host-galaxy contribution show a similar or even better performance than the pure galaxy samples at the same SNR limit. For example, for $f_{\text{AGN}_{FUV}} < 0.5$, the fraction of outliers is considerably lower than for the galaxies (16.66% instead of 24.45%), even though the bias is about five times larger. As was the case for the type 2 AGN, this increase in performance can be explained by selection effects, which are much stronger in the case of type 1 AGN (see Fig. 4.17). That is, these AGN subsamples correspond to objects characterised by galaxy-like

SEDs, but that are much brighter on average than the pure galaxy sample. This also explains the fact that the results for the $\text{SNR} > 10$ samples do not show significant improvement (see Table 4).

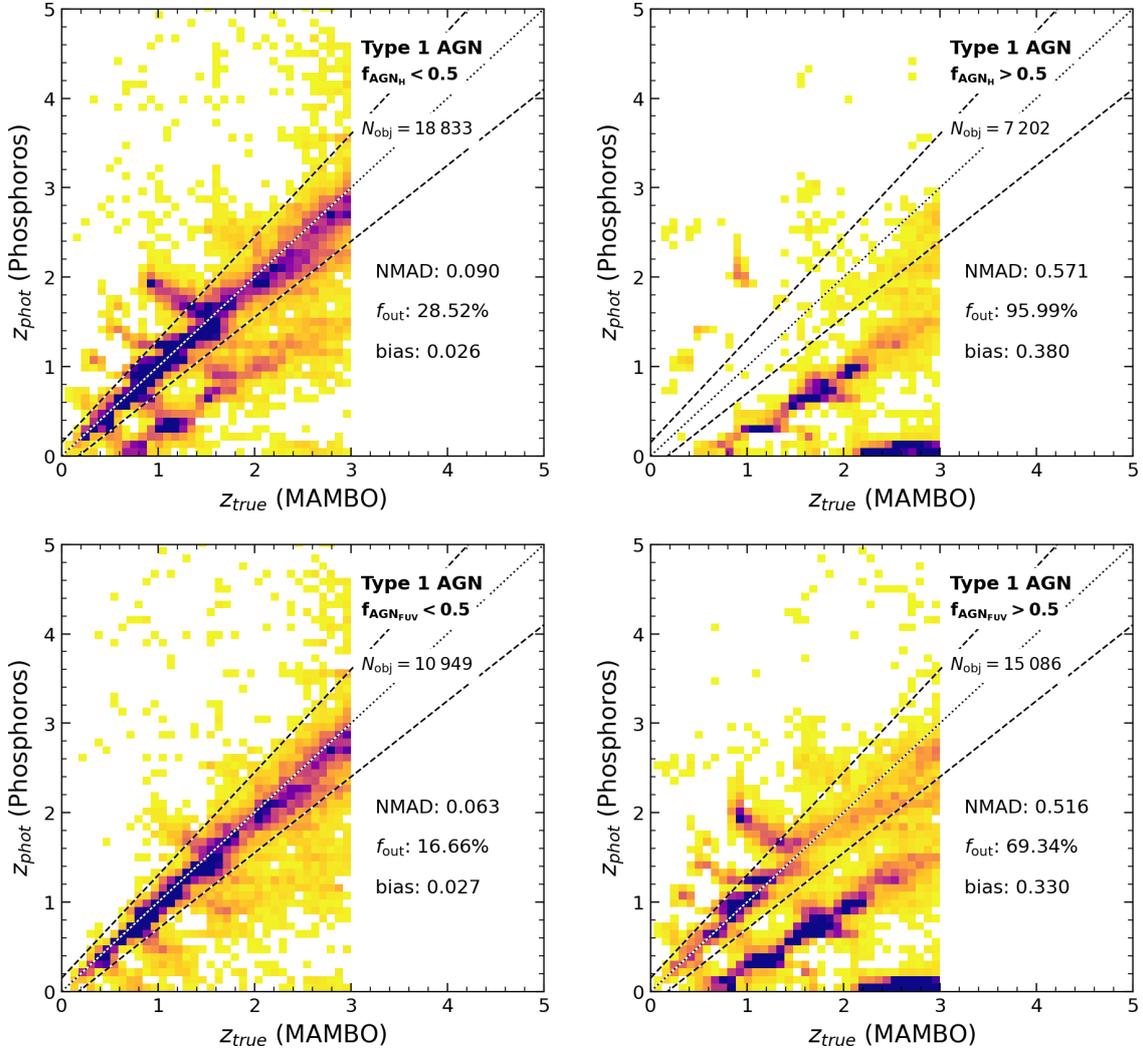


Figure 4.22: Comparison between the photometric and true redshift of type 1 AGN selected above or below a threshold of 0.5 in f_{AGN_H} (upper panels) and $f_{\text{AGN}_{FUV}}$ (lower panels). We run Phosphoros with pure galaxy SEDs as model templates. All figures correspond to samples selected with $\text{SNR} > 5$. See the caption of Fig. 4.19 for further details.

While it may seem obvious that fitting AGN-dominated sources with galaxy-dominated templates would lead to poor results, and therefore one should use AGN templates instead, the reality is more complex. In practice, it is not always possible to obtain a reliable classification of how much a source is dominated by nuclear versus stellar emission before performing the photo- z computation. Therefore, this exercise provides valuable insights into the extent of potential bias when proper classification is not performed. As we have seen, these biases can be substantially high, reaching a fraction of outliers as high as 95.99% when the AGN dominates in at least one of the bands used for the photo- z estimation.

In view of these results, we performed the photo- z computation again for the AGN-dominated subsamples (i.e., $f_{\text{AGN}} > 0.5$), this time using in Phosphoros a template library that incorporates AGN-dominated sources (see Fig. 4.16 and the accompanying text). We show the results in Fig. 4.23. By comparing these two figures to the equivalent ones using galaxy templates (right panels of Fig. 4.22), we see a clear improvement in performance, especially notable for the $f_{\text{AGN}_H} > 0.5$ sample.

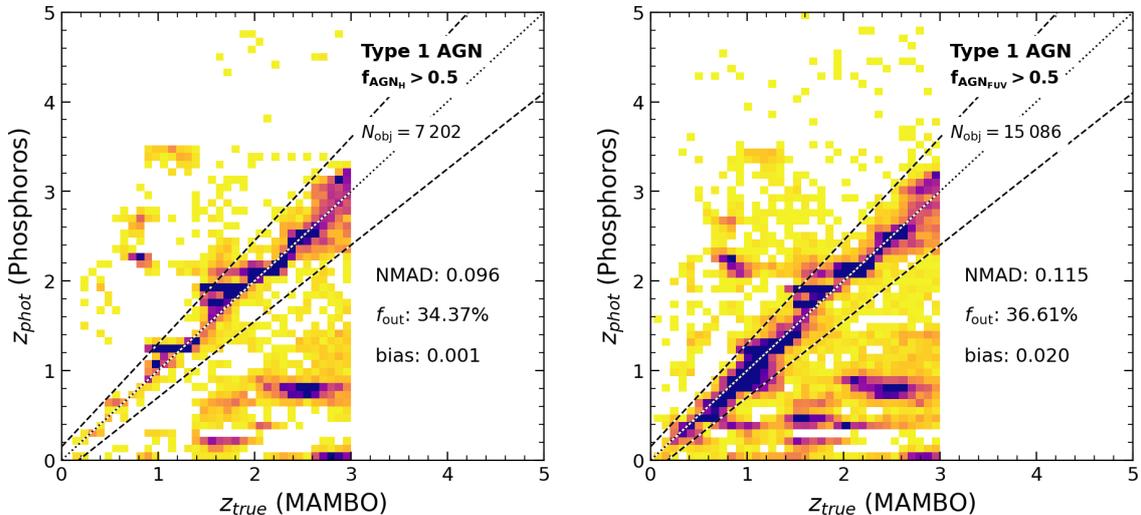


Figure 4.23: Photometric redshift performance for the AGN-dominated subsamples using AGN templates as models in Phosphoros.

While we observe an improvement in the results with respect to when we used galaxy templates, these results can be further improved by applying a suitable prior. As shown in Salvato et al. (2011), the inclusion of a luminosity prior can significantly increase the performance of the photo- z computation. Therefore, following

the flowchart shown in Figure 8 of Salvato et al. (2011) (and using $f_{\text{AGN}} > 0.5$ as a proxy for defining a point-like source), we imposed as a prior that the absolute magnitude M_B of these AGN-dominated sources must be $-30 < M_B < -20$. We remind the reader that the results previously shown in this section have been achieved, instead, assuming a volume prior $\mathcal{P}(z) \propto dV_C(z)$.

We show in Fig. 4.24 the intrinsic M_B distribution of type 1 AGN with different f_{AGN} selections. We see that the $M_B < -20$ limit required by the luminosity prior is actually well representing the $f_{\text{AGN}_H} > 0.5$ subsample, while it is also close for the $f_{\text{AGN}_{FUV}} > 0.5$ one. This is an extra indication that this is a suitable prior for these specific samples.

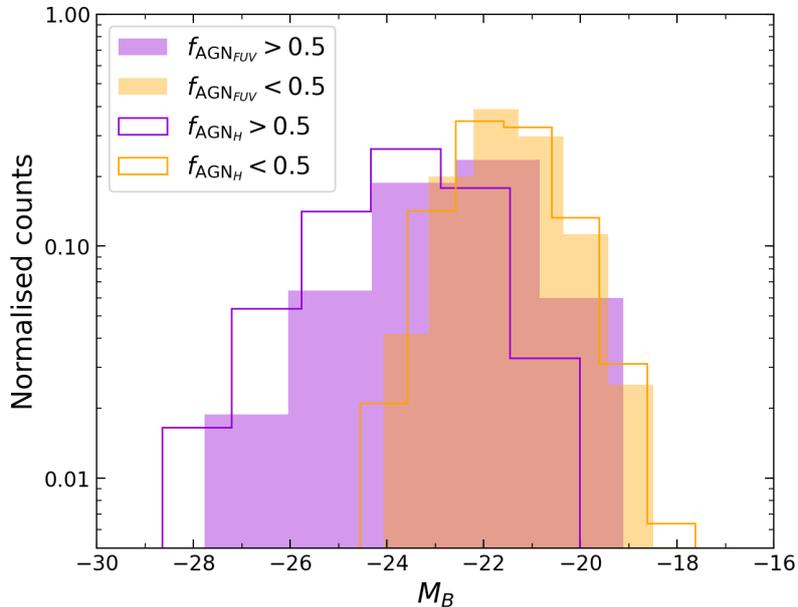


Figure 4.24: Intrinsic absolute magnitude M_B of the type 1 AGN sample with different f_{AGN} selections.

We show in Fig. 4.25 the photo- z results using AGN templates and the luminosity prior. We see that these results are significantly better with respect to using the volume prior (Fig. 4.23), showing a reduction of about 10% in the fraction of outliers. In fact, we see that in terms of the global statistics, these results are very similar to the ones obtained for the galaxy sample using galaxy templates (Fig. 4.19).

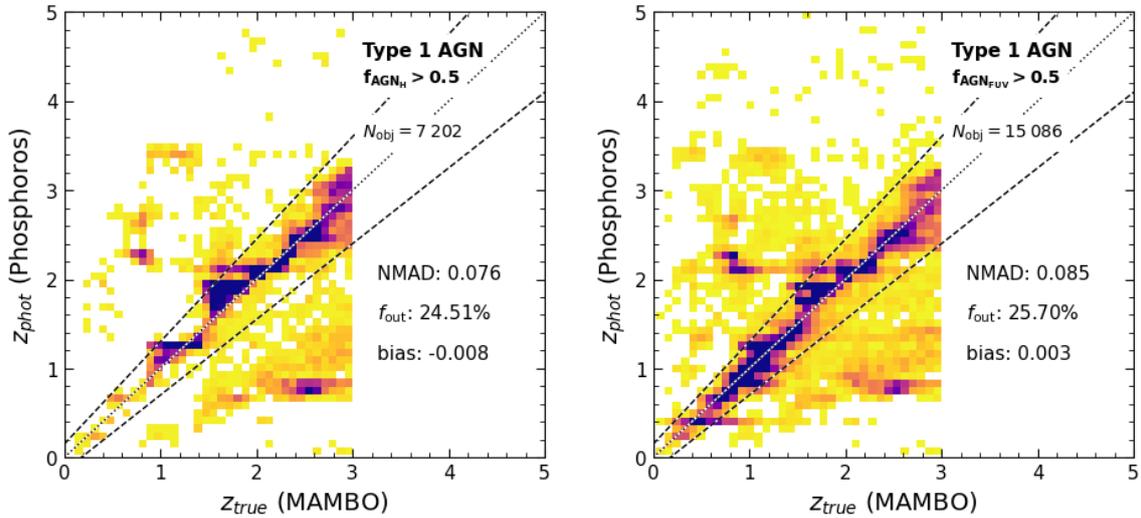


Figure 4.25: Photometric redshift performance for the AGN-dominated subsamples, using AGN templates as models in Phosphoros and a luminosity prior as the one presented in [Salvato et al. \(2011\)](#).

Conclusions

In this section, we have assessed the performance of the Phosphoros photo- z algorithm using a range of galaxy and AGN subsamples extracted from the FS2+MAMBO catalogue. We incorporated realistic noise and observational conditions corresponding to the NDR1 to simulate photometric measurements under different SNR conditions.

The results show that Phosphoros performs well for galaxies and type 2 AGN when using pure galaxy templates but fails to accurately model AGN-dominated subsamples, where the AGN component contributes significantly to the total emission.

We addressed this challenge by re-running the photo- z estimation using a dedicated AGN template library for AGN-dominated sources. This led to a significant improvement, particularly for sources dominated by AGN emission in the H band, where outlier fractions were significantly reduced. Furthermore, we explored the impact of applying a luminosity prior. We showed that this approach yields further improvements, bringing the photo- z results for type 1 AGN to a comparable level of accuracy as for galaxies.

These findings highlight the importance of selecting appropriate templates and priors for different source populations, especially in mixed galaxy-AGN samples where the AGN component can strongly influence the photometric measurements. The results also underscore the necessity of incorporating classification methods to achieve robust photo- z estimates, both for AGN and for galaxies, as noted previously by other authors (see e.g. [Fotopoulou & Paltani, 2018](#)).

Table 4: Summary of the quality-assessment statistics for the photo- z computations shown in this section. Phosphoros runs for galaxies and type 2 AGN were performed using galaxy templates and volume prior. For the type 1 AGN, this information is given in the table.

			NMAD	f_{out} (%)	bias	N_{obj}	
Galaxies		SNR = 5	0.081	24.45	0.005	56 059	
		SNR = 10	0.057	13.25	0.014	32 601	
Type 2 AGN		SNR = 5	0.075	19.96	0.005	43 928	
		SNR = 10	0.060	12.83	0.011	29 795	
Type 1 AGN Galaxy templates	All	SNR = 5	0.185	47.18	0.064	26 035	
		SNR = 10	0.187	47.59	0.064	24 389	
	$f_{\text{AGN}_{FUV}} > 0.5$	SNR = 5	0.516	69.34	0.330	15 086	
		SNR = 10	0.519	69.74	0.335	14 558	
	$f_{\text{AGN}_{FUV}} < 0.5$	SNR = 5	0.063	16.66	0.027	10 949	
		SNR = 10	0.057	14.79	0.027	9831	
	V_{prior}	$f_{\text{AGN}_H} > 0.5$	SNR = 5	0.571	95.99	0.380	7202
			SNR = 10	0.573	96.51	0.381	7030
	$f_{\text{AGN}_H} < 0.5$	SNR = 5	0.090	28.52	0.026	18 833	
		SNR = 10	0.084	27.78	0.025	17 359	
Type 1 AGN AGN templates	$f_{\text{AGN}_{FUV}} > 0.5$	SNR = 5	0.115	36.61	0.020	15 086	
		SNR = 10	0.111	35.87	0.019	14 558	
	V_{prior}	$f_{\text{AGN}_H} > 0.5$	SNR = 5	0.096	34.37	0.001	7202
			SNR = 10	0.094	33.85	0.0002	7030
Type 1 AGN AGN templates	$f_{\text{AGN}_{FUV}} > 0.5$	SNR = 5	0.085	25.70	0.003	15 086	
		SNR = 10	0.083	25.42	0.003	14 558	
	L_{prior}	$f_{\text{AGN}_H} > 0.5$	SNR = 5	0.076	24.50	-0.008	7202
			SNR = 10	0.076	24.50	-0.008	7030

4.5 XMM observations on a Euclid Deep Field: Fornax

The European Space Agency’s X-ray Multi-Mirror Mission (XMM-Newton, or simply XMM), launched in December 1999, is a space observatory designed to explore the high-energy universe. It is equipped with three advanced X-ray telescopes that offer an exceptional combination of a large effective area and an unprecedented field of view of 30 arcminutes, and an Optical Monitor, the first such instrument ever included in an X-ray observatory. Its large collecting area and ability to conduct long, uninterrupted exposures make XMM highly sensitive to faint X-ray sources, which makes it ideal for performing wide-field sky surveys and in-depth studies of extended objects.

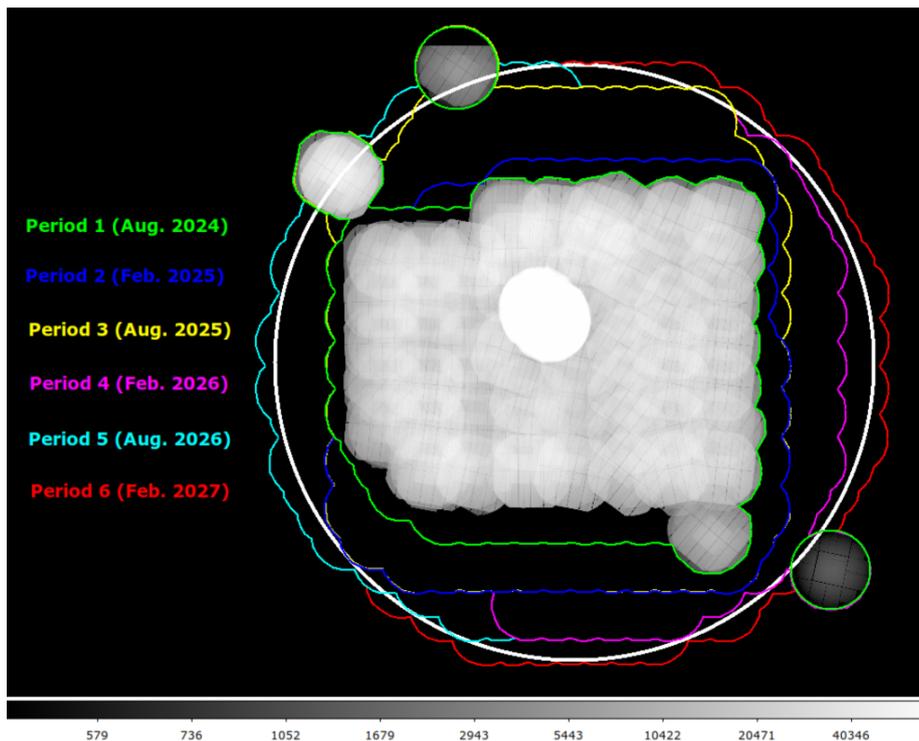


Figure 4.26: Observing plan for the XMM Multi Years Heritage programme Fornax (as of May 2024). The white area shows the already existing XMM observations. The grey scale indicates the cumulative exposure time in seconds.

In December 2023, the multi-year XMM Heritage programme Fornax²³, was awarded 3.6 Ms of XMM observing time. The aim of the program is to uniformly map the 10 deg² of the Euclid Deep Field Fornax (EDFF) at 40 ks depth. This region is particularly interesting since it benefits from a uniquely deep and uniform multi- λ coverage, ranging from X-ray to radio wavelengths, and will be the subject of extensive spectroscopic follow-up using facilities located in the Northern and Southern hemispheres. The observations of the Fornax programme started in August 2024 and are planned to reach final depth by February 2027 (see Fig. 4.26).

The main scientific goals of the project are the study of AGN and galaxy clusters and their co-evolution. In particular, it is expected to bring critical insights into the physics of the intra-cluster medium (ICM) and its interaction with galactic nucleus activity, especially for the low-mass cluster population at $1 < z < 2$. Besides, the Fornax project will play a crucial role in characterising the Euclid cluster selection function, a key factor for the upcoming cluster cosmological analysis. Conversely, the project will also examine potential selection biases in X-ray surveys.

On the other side, the project will also enable the study of rare AGN and clusters in a variety of environments. The identification of obscured AGN up to high redshifts will help constrain the BH accretion rate density as a function of cosmic time. Other scientific goals include studying the role of AGN feedback mechanisms and its link with star formation at high redshift ($z > 3$) and studying the scaling relations of AGN and clusters of galaxies.

Finally, Fornax will also deliver a complete, validated multi-wavelength legacy dataset over the 10 deg² of the EDFF, supporting a wide range of scientific investigations and providing benchmark calibration for AGN and cluster science across the 14 000 deg² of the Euclid Wide Survey.

AGN modelling for Fornax

To fully exploit the available dataset in the optical/NIR and X-ray over the coming years, the project relies on dedicated simulations. Specifically, in the X-ray band, we must model two source populations: clusters of galaxies (extended sources) and

²³<https://fornax.cosmostat.org/> (PI: M. Pierre, co-PIs: M. Bolzonella, B. Maughan, and S. Paltani)

AGN (point-like), which are the two main X-ray sources observed at high galactic latitudes. Of these two, AGN are about 50-100 times more numerous than clusters at the X-ray depths covered by FornaX. An accurate modelling of the AGN population is therefore required not only to achieve the AGN scientific goals, but is also essential for cluster studies: most of the uncertainties in the X-ray cluster selection function come from the difficulty to discriminate between faint extended and point-like sources, in a regime dominated by photon noise.

The FS2 simulation (see Sec. 4.3) is currently employed for deriving optical and X-ray properties of these two populations. In FS2, optical clusters are simulated by assigning properties to galaxies based on empirical relations such as the observed luminosity function and colour distribution (Euclid Collaboration: Castander et al., 2024). The X-ray properties of these clusters are assigned following a set of scaling relations which relate the gas temperature T and the X-ray luminosity in the rest-frame soft band $L_{0.5-2\text{keV}}$ of a cluster with its mass (Cerardi et al., 2024). Finally, the X-ray surface brightness for clusters is modelled by radially distributing the flux following a spherically symmetric β model (Cavaliere & Fusco-Femiano, 1976):

$$F_X = F_0 \left(1 + r^2/r_c^2\right)^{0.5-3\beta}, \quad (34)$$

where F_0 is the flux normalisation, and r_c radius of the cluster core. The angular projection of the latter is known as the EXT parameter and will be used below.

On the other hand, to model the AGN population, we followed a similar approach as the one described in Sec. 4.3. However, in this case, we are interested in a new quantity, namely the observed (i.e. obscured) X-ray flux in the rest-frame soft band $F_{0.5-2\text{keV}}$. We calculated this flux in two different ways, in both cases starting from the intrinsic luminosity $L_{2-10\text{keV}}$. Assuming that the X-ray emission of AGN is dominated by a power-law component, characterized by a photon index Γ , we can calculate the intrinsic soft band luminosity as:

$$L_{0.5-2\text{keV}} = \frac{E_2^{2-\Gamma} - E_1^{2-\Gamma}}{E_4^{2-\Gamma} - E_3^{2-\Gamma}} L_{2-10\text{keV}}, \quad (35)$$

with $E_1 = 0.5$ keV, $E_2 = E_3 = 2$ keV and $E_4 = 10$ keV.

For the first method, we assumed that the X-ray spectra of every source can be described with a power-law with a photon index $\Gamma = 1.4$. In reality, AGN are known

to exhibit steeper X-ray spectra (i.e. higher values of photon index) with typical values ranging in $\Gamma \sim 1.6 - 2.2$ (e.g. [Nandra & Pounds, 1994](#); [Marchesi et al., 2016b](#)). However the value $\Gamma = 1.4$ corresponds to the slope of the cosmic X-ray background (e.g. [Marshall et al., 1980](#); [Hasinger et al., 1998](#); [Gendreau et al., 1995](#)), and therefore represents the full population of both obscured and unobscured objects. Therefore, we calculated the soft-band flux taking into account a Γ -dependent K-correction as:

$$F_{0.5-2\text{keV}} = \frac{L_{0.5-2\text{keV}}}{4\pi D_L^2 (1+z)^{(\Gamma-2)}}, \quad (36)$$

where D_L is the luminosity distance of each source, and $\Gamma = 1.4$ in Eqs. [35](#) and [36](#). As can be seen in [Fig. 4.27](#) (orange line), this method yields soft-band fluxes which are in very good agreement with the observed $\text{Log}N$ - $\text{Log}S$ distribution (that is, the number source density as a function of flux) for $F_{0.5-2\text{keV}} > 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$, while we observe an overprediction in the number of sources at lower fluxes.

For the second method, we applied an individualised obscuration to each source. For this, we used the tool PIMMS²⁴. For a given value of intrinsic absorption column density N_H , redshift z , and intrinsic power-law photon index Γ , PIMMS allows to compute the obscured flux of an X-ray source. We created a grid of PIMMS models at varying z and N_H covering the full space parameter of our mock catalogue ($z = 0 - 3$, $20 < \log N_H/\text{cm}^{-2} < 24$), with fixed intrinsic photon index $\Gamma = 1.9$. The N_H in our simulation was modelled following [Ueda et al. \(2014\)](#), as in [Sec. 2.2.3](#). We computed the obscured flux using Eqs. [35](#) and [36](#), assuming an intrinsic $\Gamma = 1.9$, and applying the obscuration from the corresponding PIMMS model.

As visible in [Fig. 4.27](#), both of these methods yield very similar results at for $F_{0.5-2\text{keV}} > 10^{-14} \text{ erg s}^{-1} \text{ cm}^{-2}$, while at lower fluxes the second method predicts slightly lower number counts than the first. Nevertheless, both methods reproduce with enough accuracy the observed number counts.

The AGN and cluster models described in this section will be published in M. Pierre et al. (in prep.).²⁵

²⁴<https://ixpe.msfc.nasa.gov/cgi-aft/w3pimms/w3pimms.pl>

²⁵”The XMM-Euclid Fornax Deep-Survey – I. Simulations: Scientific motivations and XMM modelling”

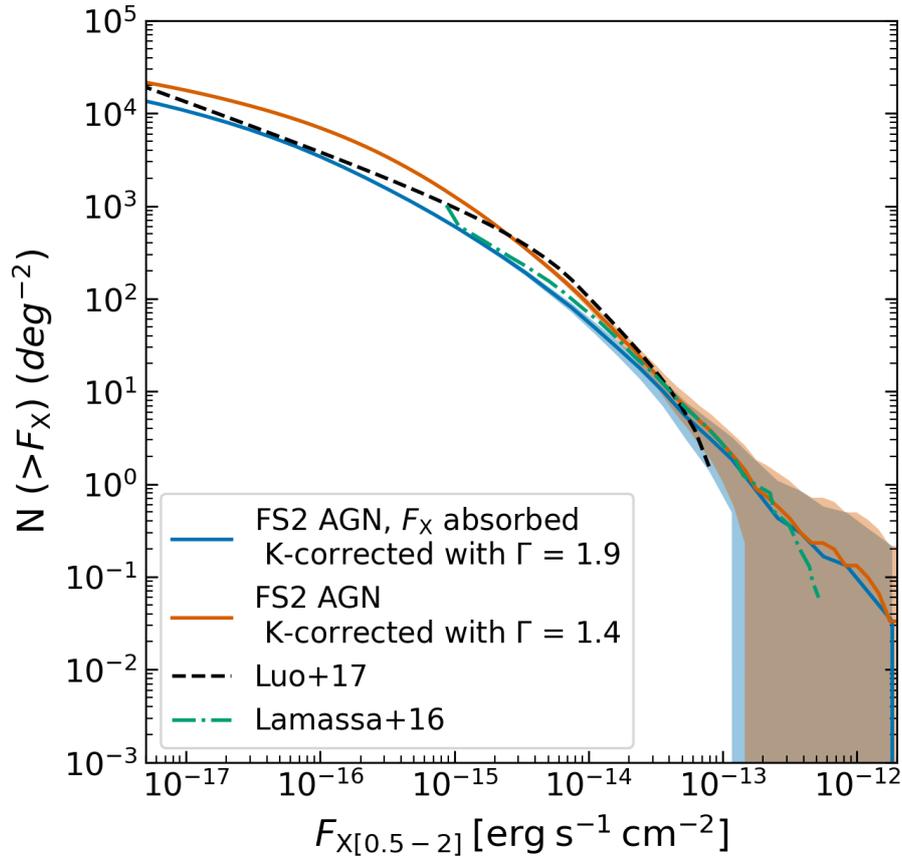


Figure 4.27: Cumulative number source density in the soft X-ray band. The orange line corresponds to the case where the flux is computed assuming a photon index $\Gamma = 1.4$ for all sources, which is the model adopted for the official Fornax simulations. The blue line shows the case where the flux absorption is modelled with PIMMS. For comparison, we show the observed cumulative number densities reported in different works that studied wide [LaMassa et al. \(2016\)](#) and deep ([Luo et al., 2017](#)) X-ray surveys.

XMM simulated images

In order to illustrate the effects of exposure time and the inclusion of the AGN population on the X-ray simulated observations, we show in [Fig. 4.28](#) a simulated image of one observation tile at different depths. The image is a combination of the MOS1, MOS2 and pn (the three instruments on board of XMM-Newton) individual images, and includes all background components and instrumental effects.

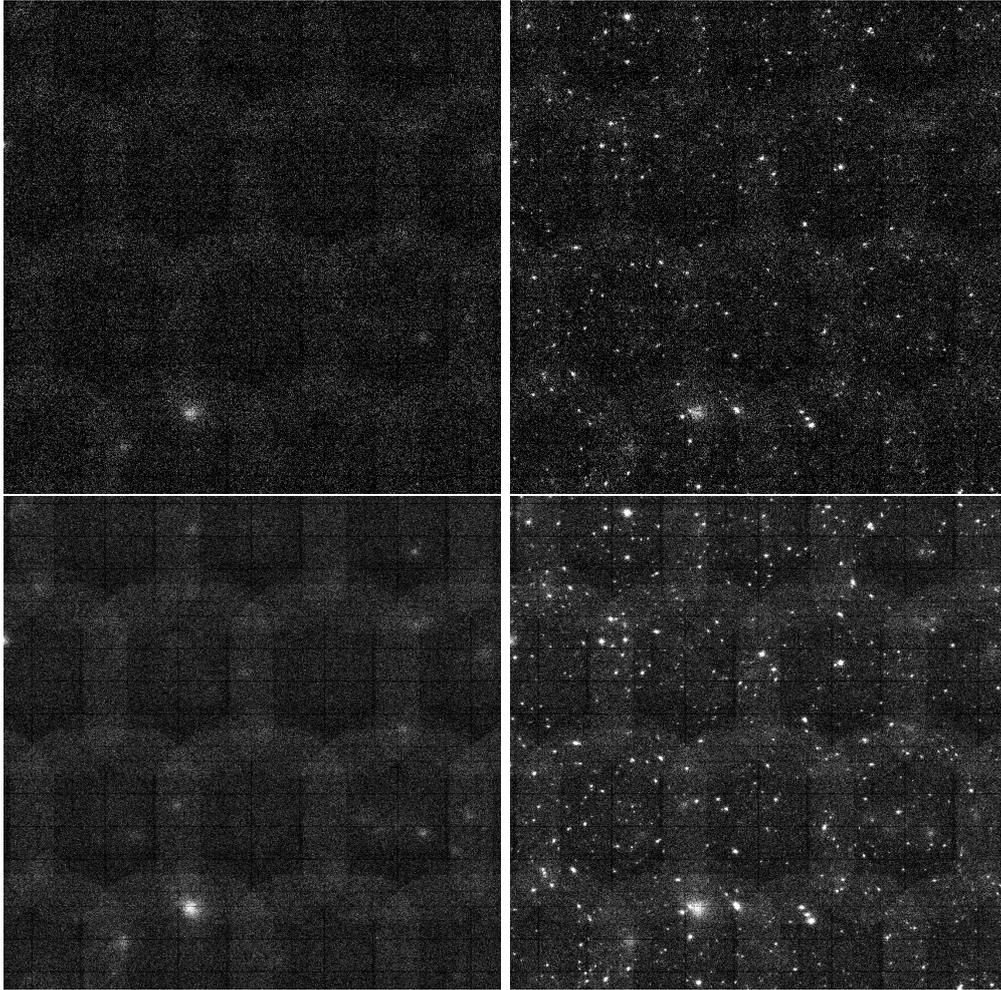


Figure 4.28: Combined MOS1+MOS2+pn image of the central tile, including all background components and instrumental effects. *Left panels:* DM cluster halo only. *Right panels :* final image including the cluster and AGN populations. The upper and lower rows show the 10 ks and 40 ks depths, respectively. The display low-high cuts are different in each case in order to give a visual impression of the gain in SNR when switching from 10 to 40ks. Credits: M. Pierre et al. (in prep.).

The left panels on this figure show the DM cluster halo only, while on the right panels, galaxy clusters and AGN have been included. The top and bottom rows represent 10 ks and 40 ks exposures, respectively. The increase in SNR with exposure time is visible from this image.

Cluster selection and AGN contamination

As mentioned earlier, one of the main purposes of modelling the AGN population for the Fornax project is to study how they can act as contaminants in the cluster selection process. Understanding this interference is crucial for improving the accuracy of identifying and characterizing galaxy clusters.

To construct the cluster catalogue, every source on the XMM count image is examined by a maximum-likelihood analysis, which compares the likelihood of various source models: extended, point-like, unresolved double point-source, and extended source contaminated by a central point-source. The extended source model consists of a $\beta = 2/3$ profile, and it is characterised by only two free parameters: the angular radius of the source core `EXT`, and the normalisation F_0 (see Eq. 34).

For this work, two types of cluster samples are defined, corresponding to different contamination levels by misclassified AGN point sources (Pacaud et al., 2006). The C1 class is defined with a contamination rate of approximately 10% (i.e., a sample purity of approximately 90%) and the C2 class with 50% contamination.

Galaxy clusters are selected on the basis of their extended emission, by considering appropriate thresholds in the `EXT` parameter and the likelihood of this measurement, `EXT_LH`. In particular, for the 10 ks simulation, we define the C1 class as `EXT > 15''` and the C2 class as $8'' < \text{EXT} < 15''$, while we require `EXT_LH > 7` for both classes²⁶.

We note that to define these thresholds, the physical AGN modelling is required. This represents a clear advantage with respect to previous works, where the thresholds were set by randomly positioning clusters and AGN.

We show in Fig. 4.29 the position of all the sources from one sky-realisation of the 10 ks simulation on the `EXT – EXT_LH` plane. By applying the selection criteria described above, 181 C1 and 54 C2 true clusters are identified. In total, 441 clusters are detected, but most of them are too faint (`EXT < 8''`) or too contaminated by AGN emission to be classified as C1 or C2.

²⁶The exact value of these thresholds, in order to fulfil the C1 and C2 sample requirements, is still under study. Therefore, the values reported here are preliminary.

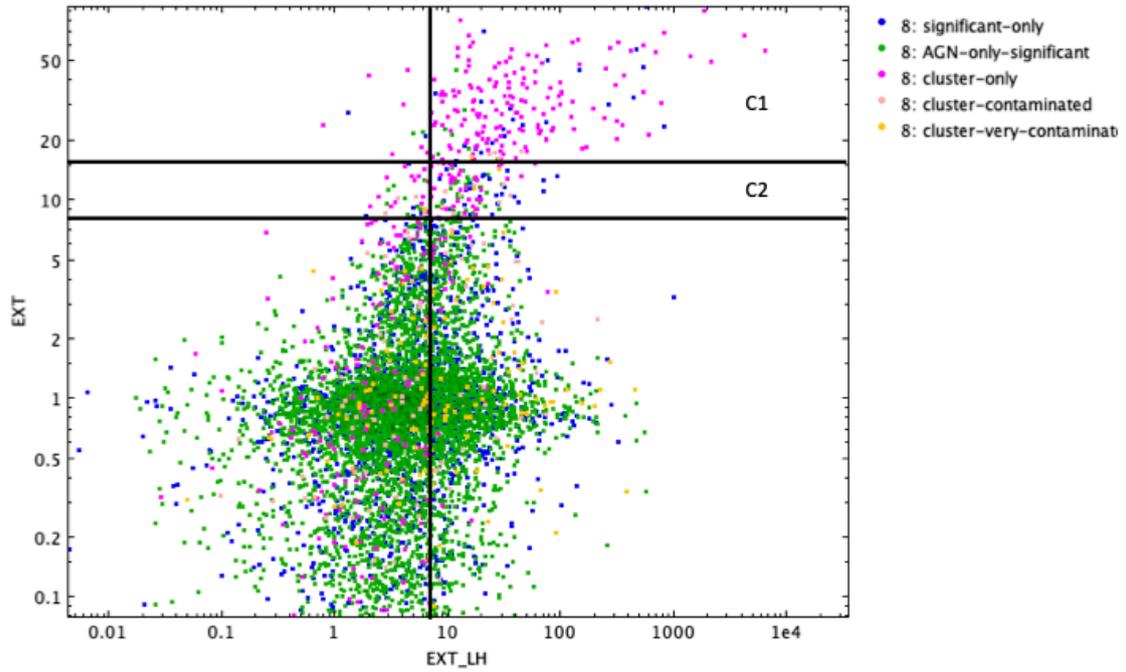


Figure 4.29: The EXT – EXT_LH plane allows for the selection of two samples of clusters: C1 (90% purity) and C2 (50% purity). The bulk of the AGN population (green dots) occupies a clearly different region of the space. This representation also shows clusters that are too faint to be unambiguously recognised as extended sources ($EXT < 8$), and clusters that are contaminated by an AGN. The flags ‘contaminated’ and ‘very contaminated’ stand for clusters whose $\text{flux(AGN)}/\text{flux(cluster)}$ ratio is > 0.25 and > 0.75 respectively. Credits: M. Pierre et al. (in prep.).

Chapter 5: Summary

In this thesis 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 thesis, we focused on the inclusion of AGN into this model 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) and assigned a proper SED to each of them. To do 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 work 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 several applications of this catalogue for future large surveys, using *Euclid* as an example. First, we provided forecasts to *Euclid* observations, such as 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 $5.5 \times 10^3 \text{ deg}^{-2}$ 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 $\text{H}\alpha$ emission with flux $F_{\text{H}\alpha} > 2 \times 10^{-16}$ (6×10^{-17}) $\text{erg s}^{-1} \text{cm}^{-2}$ in the EWS (EDS), finding good agreement with other published forecasts. We also give examples of the *Euclid* view of narrow-line diagnostic diagrams, which are used to separate local AGN from SF galaxies. We simulated *Euclid*-like AGN spectra, and analysed them using official *Euclid* pipelines. While the results we obtained require further investigation, our analysis provides an estimate of the biases and incompleteness that we can expect to find in real data. We also tested a pipeline for retrieving photometric redshifts using *Euclid*-like galaxy and AGN photometry, and demonstrated the techniques necessary to achieve efficient results for AGN. Finally, we applied the AGN modelling developed in this thesis to enhance the simulations required for the preparation of an incoming X-ray observational program.

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. Besides, the code and some examples of derived galaxy and AGN mock catalogues are publicly available at <https://github.com/xalolo/MAMBO>.

We plan to update this workflow in the near future in order to tackle certain open issues. 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). Similarly, we plan to use a more updated bolometric correction (Duras et al., 2020) that correlates with L_X , instead of assuming a fixed value as we did in this work. We note that these two points are related, since our choice of a fixed bolometric correction was imposed to ensure consistency with the accretion rate distributions used as input. 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. Besides, we plan to apply this method starting from a bigger DM lightcone. Finally, we also plan to apply this catalogue for the exploitation of the first *Euclid* data release, that was very recently (November 2024) published within the consortium.

Acknowledgements

I want to start by thanking my supervisors, Lucia and Micol, who from (before) the very beginning of my Ph.D., have helped me with every aspect of it, from the scientific part to the bureaucratic, and with all the difficulties that come with starting a new life in a foreign country. They have always been there for me, making things easy, and have adapted to my personal pace and scientific interests. I consider them great scientists and human beings, and I'm glad to have worked side by side with them.

I have been lucky enough to do my Ph.D. in a very vivid environment, with a very big community of Ph.D. students (currently $\gtrsim 50$) and postdocs working at different topics. Besides, I have been lucky enough that many of my Ph.D. colleagues are amazing experts in AGN, galaxy evolution, programming, statistics, scientific writing, LaTeX, plotting fancy figures, and some of them also great chess rivals. To all of them I am very grateful for sharing their knowledge and their time with me. To name just a few, I want to thank Ivan, Dhruvi, Katia, Alessandro, Giovanni, Farida, Laura, Alberto, Massimiliano, Nicola, Salvatore, Amir, Stefano...

I also want to express my gratitude to Mara Salvato for hosting me at the MPE for some months, and to all the people I met there and that taught me a lot about AGN and mock catalogues in very short time, specially Johan Comparat and Andrea Merloni.

Many conversations and email exchanges with many other researchers have also deeply shaped this project. Among others I want to thank for this to Akke Viitanen, Viola Allevato, Laura Bisigello, Gianni Zamorani, Anna Feltre, Angela Bongiorno, Salvatore Quai, Roberto Gilli, Federico Marulli, Michele Moresco, Matthew Temple.

I want to thank also to the two referees, Elias Koulouridis and Fabio La Franca, who read the first version of this thesis, and whose comments have greatly improved the scientific quality and the style of this document.

Finally, I want to thank my first mentor Juan Antonio, with whom I learned a lot about AGN, physics in general, and scientific writing.

Y por supuesto a mi familia, ya que sin vosotros nada de esto sería posible, muchísimas gracias por vuestro apoyo incondicional en todo lo que hago.

References

- Aird J., et al., 2012, [ApJ](#), **746**, 90
- Aird J., Coil A. L., Georgakakis A., Nandra K., Barro G., Pérez-González P. G., 2015, [Monthly Notices of the Royal Astronomical Society](#), 451, 1892
- Aird J., Coil A. L., Georgakakis A., 2018, [MNRAS](#), **474**, 1225
- Alberts S., et al., 2023, [arXiv e-prints](#), p. [arXiv:2312.12207](#)
- Allevato V., Shankar F., Marsden C., Rasufov U., Viitanen A., Georgakakis A., Ferrara A., Finoguenov A., 2021, [ApJ](#), **916**, 34
- Amorín R., et al., 2015, [A&A](#), **578**, A105
- Antonucci R., 1993, [ARA&A](#), **31**, 473
- Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Gialongo E., 1999, [MNRAS](#), **310**, 540
- Baldwin J. A., 1977, [ApJ](#), **214**, 679
- Baldwin J. A., Phillips M. M., Terlevich R., 1981, [PASP](#), **93**, 5
- Beckmann V., Shrader C. R., 2012, *Active Galactic Nuclei*
- Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, [ApJ](#), **762**, 109
- Benson A. J., 2010, [Phys. Rep.](#), **495**, 33
- Bezanson R., et al., 2018, [ApJ](#), **868**, L36
- Binette L., et al., 2024, [A&A](#), **684**, A53
- Birchall K. L., Watson M. G., Aird J., Starling R. L. C., 2022, [MNRAS](#), **510**, 4556
- Birrer S., Lilly S., Amara A., Paranjape A., Refregier A., 2014, [ApJ](#), **793**, 12
- Bisigello L., et al., 2021, [A&A](#), **651**, A52
- Bisogni S., Lusso E., Civano F., Nardini E., Risaliti G., Elvis M., Fabbiano G., 2021, [A&A](#), **655**, A109
- Blaizot J., Wadadekar Y., Guiderdoni B., Colombi S. T., Bertin E., Bouchet F. R., Devriendt J. E. G., Hatton S., 2005, [Monthly Notices of the Royal Astronomical Society](#), 360, 159
- Blandford R. D., Znajek R. L., 1977, [Monthly Notices of the Royal Astronomical Society](#), 179, 433
- Bolzonella M., Miralles J. M., Pelló R., 2000, [A&A](#), **363**, 476
- Bongiorno A., et al., 2012, [MNRAS](#), **427**, 3103

- Bongiorno A., et al., 2016, *A&A*, **588**, A78
- Brammer G. B., van Dokkum P. G., Coppi P., 2008, *ApJ*, **686**, 1503
- Brandt W. N., Alexander D. M., 2015, *A&A Rev.*, **23**, 1
- Brandt W. N., Hasinger G., 2005, *ARA&A*, **43**, 827
- Brusa M., et al., 2010, *ApJ*, **716**, 348
- Bruzual G., Charlot S., 2003, *MNRAS*, **344**, 1000
- Buchner J., et al., 2015, *ApJ*, **802**, 89
- Cabral A., et al., 2020, in Evans C. J., Bryant J. J., Motohara K., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 11447, Ground-based and Airborne Instrumentation for Astronomy VIII. p. 114478D, [doi:10.1117/12.2561412](https://doi.org/10.1117/12.2561412)
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, *ApJ*, **533**, 682
- Cann J. M., Satyapal S., Abel N. P., Blecha L., Mushotzky R. F., Reynolds C. S., Secret N. J., 2019, *ApJ*, **870**, L2
- Cappelluti N., et al., 2009, *A&A*, 497, 635
- Carnall A. C., et al., 2023, *MNRAS*, **520**, 3974
- Carnall A. C., et al., 2024, *arXiv e-prints*, p. [arXiv:2405.02242](https://arxiv.org/abs/2405.02242)
- Carretero J., Castander F. J., Gaztañaga E., Croce M., Fosalba P., 2015, *MNRAS*, **447**, 646
- Cavaliere A., Fusco-Femiano R., 1976, *A&A*, **49**, 137
- Cerardi N., Pierre M., Valageas P., Garrel C., Pacaud F., 2024, *A&A*, **682**, A138
- Chary R., Elbaz D., 2001, *ApJ*, **556**, 562
- Cimatti A., Fraternali F., Nipoti C., 2019, Introduction to Galaxy Formation and Evolution: From Primordial Gas to Present-Day Galaxies
- Comastri A., 2004, in Barger A. J., ed., Astrophysics and Space Science Library Vol. 308, Supermassive Black Holes in the Distant Universe. p. 245 ([arXiv:astro-ph/0403693](https://arxiv.org/abs/astro-ph/0403693)), [doi:10.1007/978-1-4020-2471-9_8](https://doi.org/10.1007/978-1-4020-2471-9_8)
- Comparat J., et al., 2020, *The Open Journal of Astrophysics*, **3**, 13
- Conroy C., Wechsler R. H., Kravtsov A. V., 2006, *ApJ*, **647**, 201
- Dark Energy Survey Collaboration et al., 2016, *MNRAS*, **460**, 1270
- Davidzon I., et al., 2017, *A&A*, **605**, A70
- Davis M., Peebles P. J. E., 1983, *ApJ*, **267**, 465

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371

Dayal P., Ferrara A., 2018, *Phys. Rep.*, 780, 1

De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2

Dong X., Wang T., Wang J., Yuan W., Zhou H., Dai H., Zhang K., 2007, *Monthly Notices of the Royal Astronomical Society*, 383, 581

Drory N., et al., 2009, *ApJ*, 707, 1595

Duras F., et al., 2020, *A&A*, 636, A73

Elbaz D., et al., 2011, *A&A*, 533, A119

Euclid Collaboration: Bisigello L., et al., 2024, *arXiv e-prints*, p. [arXiv:2409.00175](#)

Euclid Collaboration: Castander F. J., et al., 2024, *arXiv e-prints*, p. [arXiv:2405.13495](#)

Euclid Collaboration: Cropper M., et al., 2024, *arXiv e-prints*, p. [arXiv:2405.13492](#)

Euclid Collaboration: Enia A., et al., 2024, *arXiv e-prints*, p. [arXiv:2407.07940](#)

Euclid Collaboration: Jahnke K., et al., 2024, *arXiv e-prints*, p. [arXiv:2405.13493](#)

Euclid Collaboration: Lusso E., et al., 2024, *A&A*, 685, A108

Euclid Collaboration: Mellier Y., Abdurro'uf Acevedo Barroso J., Achúcarro A., et al., 2024, *A&A*, submitted, p. [arXiv:2405.13491](#)

Euclid Collaboration: Scaramella R., et al., 2022, *A&A*, 662, A112

Euclid Collaboration: Selwood M., et al., 2024, *arXiv e-prints*, p. [arXiv:2405.18126](#)

Event Horizon Telescope Collaboration et al., 2019, *ApJ*, 875, L1

Fabian A. C., Rees M. J., Stella L., White N. E., 1989, *MNRAS*, 238, 729

Feltre A., Charlot S., Gutkin J., 2016, *MNRAS*, 456, 3354

Ferland G. J., et al., 2017, *Rev. Mexicana Astron. Astrofis.*, 53, 385

Ferrarese L., Merritt D., 2000, *ApJ*, 539, L9

Finkelstein S. L., Bagley M. B., 2022, *ApJ*, 938, 25

Fitzpatrick E. L., Massa D., 1986, *ApJ*, 307, 286

Fontanot F., De Lucia G., Xie L., Hirschmann M., Baugh C., Helly J. C., 2024, *arXiv e-prints*, p. [arXiv:2409.02194](#)

Fotopoulou S., Paltani S., 2018, *A&A*, 619, A14

Frank J., King A., Raine D., 2002, *Accretion Power in Astrophysics*, 3 edn. Cambridge University Press, [doi:10.1017/CBO9781139164245](#)

Gendreau K. C., et al., 1995, *PASJ*, 47, L5

Georgakakis A., Nandra K., Laird E. S., Aird J., Trichas M., 2008, *MNRAS*, 388,

1205

- Georgakakis A., et al., 2014, *MNRAS*, **440**, 339
- Georgakakis A., Comparat J., Merloni A., Ciesla L., Aird J., Finoguenov A., 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
- Gilli R., et al., 2009, *A&A*, **494**, 33
- Giocoli C., Meneghetti M., Metcalf R. B., Etori S., Moscardini L., 2014, *MNRAS*, **440**, 1899
- Girelli G., 2021, PhD thesis, Alma Mater Studiorum Università di Bologna, Italy, <http://amsdottorato.unibo.it/9820/>
- Girelli G., Bolzonella, Micol Cimatti, Andrea 2019, *A&A*, **632**, A80
- Girelli G., Pozzetti L., Bolzonella M., Giocoli C., Marulli F., Baldi M., 2020, *A&A*, **634**, A135
- Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, *MNRAS*, **439**, 264
- Graham M. L., et al., 2020, *AJ*, **159**, 258
- Grogin N. A., et al., 2011, *ApJS*, **197**, 35
- Groves B. A., Heckman T. M., Kauffmann G., 2006, *MNRAS*, **371**, 1559
- Guo H., Shen Y., Wang S., 2018, PyQSOFit: Python code to fit the spectrum of quasars, Astrophysics Source Code Library, record ascl:1809.008
- Gutkin J., Charlot S., Bruzual G., 2016, *Monthly Notices of the Royal Astronomical Society*, **462**, 1757
- Haardt F., Maraschi L., 1991, *ApJ*, **380**, L51
- Harrison C. M. C., 2014, PhD thesis, Durham University, UK
- Harrison C. M., 2017, Impact of supermassive black hole growth on star formation ([arXiv:1703.06889](https://arxiv.org/abs/1703.06889))
- Hasinger G., Burg R., Giacconi R., Schmidt M., Trumper J., Zamorani G., 1998, *A&A*, **329**, 482
- Hasinger G., et al., 2007, *ApJS*, **172**, 29
- Heckman T. M., 1980, *A&A*, **500**, 187
- Heckman T. M., Best P. N., 2014, *ARA&A*, **52**, 589
- Henriques B. M. B., White S. D. M., Lemson G., Thomas P. A., Guo Q., Marleau

G., Overzier R. A., 2012, [Monthly Notices of the Royal Astronomical Society](#), 421, 2904

Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, [MNRAS](#), 451, 2663

Hickox R. C., Alexander D. M., 2018, [ARA&A](#), 56, 625

Hildebrandt H., et al., 2010, [A&A](#), 523, A31

Hirschmann M., Charlot S., Feltre A., Naab T., Choi E., Ostriker J. P., Somerville R. S., 2017, [Monthly Notices of the Royal Astronomical Society](#), 472, 2468

Hirschmann M., Charlot S., Feltre A., Naab T., Somerville R. S., Choi E., 2019, [MNRAS](#), 487, 333

Ho L. C., 2008, [ARA&A](#), 46, 475

Ho L. C., Filippenko A. V., Sargent W. L. W., 1997, [ApJS](#), 112, 315

Ibata R. A., et al., 2017, [ApJ](#), 848, 128

Ilbert O., et al., 2006, [A&A](#), 457, 841

Ilbert O., et al., 2009, [ApJ](#), 690, 1236

Ilbert O., et al., 2010, [ApJ](#), 709, 644

Ilbert O., et al., 2013, [A&A](#), 556, A55

Ivezic Z., et al., 2008, [Serbian Astronomical Journal](#), 176, 1

Iwasawa K., Sanders D. B., Evans A. S., Mazzarella J. M., Armus L., Surace J. A., 2009, [ApJ](#), 695, L103

Iwasawa K., et al., 2011, [A&A](#), 529, A106

Jin X., Zhang Y., Zhang J., Zhao Y., Wu X.-b., Fan D., 2019, [MNRAS](#), 485, 4539

Juneau S., Dickinson M., Alexander D. M., Salim S., 2011, [ApJ](#), 736, 104

Kauffmann G., 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., Maier C., Yabe K., Ohta K., Akiyama M., Dopita M. A., Yuan T., 2013a, [ApJ](#), 774, L10

Kewley L. J., Dopita M. A., Leitherer C., Davé R., Yuan T., Allen M., Groves B., Sutherland R., 2013b, [ApJ](#), 774, 100

Khachikian E. E., Weedman D. W., 1971, [Astrofizika](#), 7, 389

Kim J.-h., et al., 2016, [ApJ](#), 833, 202

Kormendy J., Ho L. C., 2013, [ARA&A](#), 51, 511

Kravtsov A. V., Klypin A. A., Khokhlov A. M., 1997, *ApJS*, **111**, 73

Kulkarni G., Worseck G., Hennawi J. F., 2019, *MNRAS*, **488**, 1035

LSST Science Collaboration et al., 2009, *arXiv e-prints*, p. [arXiv:0912.0201](#)

LaMassa S. M., et al., 2016, *ApJ*, **817**, 172

Laigle C., et al., 2016, *ApJS*, **224**, 24

Landy S. D., Szalay A. S., 1993, *ApJ*, **412**, 64

Latimer L. J., Reines A. E., Bogdan A., Kraft R., 2021, *ApJ*, **922**, L40

Lehmer B. D., Alexander D. M., Bauer F. E., Brandt W. N., Goulding A. D., Jenkins L. P., Ptak A., Roberts T. P., 2010, *ApJ*, **724**, 559

Lehmer B. D., et al., 2012, *ApJ*, **752**, 46

Looser T. J., et al., 2023, *arXiv e-prints*, p. [arXiv:2302.14155](#)

Lopez I. E., 2024, PhD thesis, alma, <http://amsdottorato.unibo.it/11583/>

Lu Y., et al., 2014, *ApJ*, **795**, 123

Lu R.-S., et al., 2023, *Nature*, **616**, 686

Luo B., et al., 2017, *ApJS*, **228**, 2

Lusso E., et al., 2013, *ApJ*, **777**, 86

Lynden-Bell D., 1969, *Nature*, **223**, 690

Madau P., Dickinson M., 2014, *ARA&A*, **52**, 415

Magdis G. E., et al., 2012, *ApJ*, **760**, 6

Magorrian J., et al., 1998, *AJ*, **115**, 2285

Manti S., Gallerani S., Ferrara A., Greig B., Feruglio C., 2017, *MNRAS*, **466**, 1160

Marchesi S., et al., 2016a, *ApJ*, **817**, 34

Marchesi S., et al., 2016b, *ApJ*, **830**, 100

Marconi A., Hunt L. K., 2003, *ApJ*, **589**, L21

Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, *MNRAS*, **351**, 169

Marshall F. E., Boldt E. A., Holt S. S., Miller R. B., Mushotzky R. F., Rose L. A., Rothschild R. E., Serlemitsos P. J., 1980, *ApJ*, **235**, 4

Martín-Navarro I., Brodie J. P., Romanowsky A. J., Ruiz-Lara T., van de Ven G., 2018, *Nature*, **553**, 307–309

Marulli F., Veropalumbo A., Moresco M., 2016, *Astronomy and Computing*, **14**, 35

McCracken H. J., et al., 2012, *A&A*, **544**, A156

Menzel M. L., et al., 2016, *MNRAS*, **457**, 110

Merlin E., et al., 2017, [Monthly Notices of the Royal Astronomical Society](#), 473, 2098

Merloni A., et al., 2014, [MNRAS](#), 437, 3550

Merloni A., et al., 2024, [A&A](#), 682, A34

Mezcua M., Domínguez Sánchez H., 2020, [ApJ](#), 898, L30

Mezcua M., Domínguez Sánchez H., 2024, [MNRAS](#), 528, 5252

Miyaji T., et al., 2015, [ApJ](#), 804, 104

Mo H., van den Bosch F. C., White S., 2010, *Galaxy Formation and Evolution*

Monaco P., 2016, [Galaxies](#), 4, 53

Moster B. P., Somerville R. S., Maulbetsch C., van den Bosch F. C., Macciò A. V., Naab T., Oser L., 2010, [ApJ](#), 710, 903

Moster B. P., Naab T., White S. D. M., 2018, [MNRAS](#), 477, 1822

Mushotzky R. F., Done C., Pounds K. A., 1993, [ARA&A](#), 31, 717

Naab T., Ostriker J. P., 2017, [ARA&A](#), 55, 59

Nandra K., Pounds K. A., 1994, [MNRAS](#), 268, 405

Navarro J. F., Frenk C. S., White S. D. M., 1996, [ApJ](#), 462, 563

Nelson D., et al., 2018, [MNRAS](#), 475, 624

Netzer H., 2013, *The Physics and Evolution of Active Galactic Nuclei*. Cambridge University Press, [doi:10.1017/CBO9781139109291](https://doi.org/10.1017/CBO9781139109291)

Newman J. A., Gruen D., 2022, [Annual Review of Astronomy and Astrophysics](#), 60, 363

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*

Pacaud F., et al., 2006, [MNRAS](#), 372, 578

Padovani P., et al., 2017, [A&A Rev.](#), 25, 2

Parsa S., Dunlop J. S., McLure R. J., Mortlock A., 2016, [Monthly Notices of the Royal Astronomical Society](#), 456, 3194

Peebles P. J. E., 1980, *The large-scale structure of the universe*

Peluso G., Radovich M., Moretti A., Mingozzi M., Vulcani B., Poggianti B. M., Marasco A., Gullieuszik M., 2023, [ApJ](#), 958, 147

Peng Y.-j., et al., 2010, [ApJ](#), 721, 193

Penzias A. A., Wilson R. W., 1965, [ApJ](#), 142, 419

Pereira-Santaella M., et al., 2011, *A&A*, 535, A93
Perlmutter S., et al., 1999, *ApJ*, 517, 565
Pillepich A., et al., 2019, *MNRAS*, 490, 3196
Planck Collaboration et al., 2016, *A&A*, 594, A1
Pouliasis E., et al., 2024, *A&A*, 685, A97
Prevot M. L., Lequeux J., Maurice E., Prevot L., Rocca-Volmerange B., 1984, *A&A*, 132, 389
Rees M. J., 1984, *ARA&A*, 22, 471
Ricci F., Marchesi S., Shankar F., La Franca F., Civano F., 2017a, *MNRAS*, 465, 1915
Ricci C., et al., 2017b, *Nature*, 549, 488
Richards G. T., et al., 2002, *AJ*, 123, 2945
Richards G. T., et al., 2003, *AJ*, 126, 1131
Riess A. G., et al., 1998, *AJ*, 116, 1009
Roberts M. S., Rots A. H., 1973, *A&A*, 26, 483
Rodrigo C., Solano E., Bayo A., 2012, SVO Filter Profile Service Version 1.0, IVOA Working Draft 15 October 2012, [doi:10.5479/ADS/bib/2012ivoa.rept.1015R](https://doi.org/10.5479/ADS/bib/2012ivoa.rept.1015R)
Rubin V. C., Ford W. K. J., Thonnard N., 1978, *ApJ*, 225, L107
Rubin V. C., Ford W. K. J., Thonnard N., 1980, *ApJ*, 238, 471
Salpeter E. E., 1964, *ApJ*, 140, 796
Salvato M., et al., 2011, *ApJ*, 742, 61
Salvato M., Ilbert O., Hoyle B., 2019, *Nature Astronomy*, 3, 212
Salvato M., et al., 2022, *A&A*, 661, A3
Schawinski K., et al., 2010, *ApJ*, 711, 284
Schaye J., et al., 2015, *MNRAS*, 446, 521
Schechter P., 1976, *ApJ*, 203, 297
Schmidt M., 1963, *Nature*, 197, 1040
Schreiber C., et al., 2017a, *A&A*, 602, A96
Schreiber C., et al., 2017b, *A&A*, 602, A96
Seyfert C. K., 1943, *ApJ*, 97, 28
Shakura N. I., Sunyaev R. A., 1973, *A&A*, 500, 33
Shen Y., et al., 2019, *ApJS*, 241, 34
Sparre M., et al., 2015, *MNRAS*, 447, 3548

- Spergel D., et al., 2015, [arXiv e-prints](#), p. [arXiv:1503.03757](#)
- Spinoglio L., Fernández-Ontiveros J. A., 2021, in Pović M., Marziani P., Masegosa J., Netzer H., Negu S. H., Tessema S. B., eds, IAU Symposium Vol. 356, Nuclear Activity in Galaxies Across Cosmic Time. pp 29–43 ([arXiv:1911.12176](#)), [doi:10.1017/S1743921320002549](#)
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, [MNRAS](#), **328**, 726
- Springel V., et al., 2005, [Nature](#), **435**, 629
- Storey-Fisher K., Hogg D., 2020, in American Astronomical Society Meeting Abstracts #235. p. 266.05
- Temple M. J., Hewett P. C., Banerji M., 2021, [MNRAS](#), **508**, 737
- Treister E., et al., 2004, [ApJ](#), **616**, 123
- Ueda Y., Akiyama M., Hasinger G., Miyaji T., Watson M. G., 2014, [ApJ](#), **786**, 104
- Urry C. M., Padovani P., 1995, [PASP](#), **107**, 803
- Viitanen A., Alleato V., Finoguenov A., Bongiorno A., Cappelluti N., Gilli R., Miyaji T., Salvato M., 2019, [A&A](#), **629**, A14
- Vito F., et al., 2017, [Monthly Notices of the Royal Astronomical Society](#), 473, 2378
- Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, [Nature Reviews Physics](#), **2**, 42
- Wang L., Dutton A. A., Stinson G. S., Macciò A. V., Penzo C., Kang X., Keller B. W., Wadsley J., 2015, [MNRAS](#), **454**, 83
- Weaver J. R., et al., 2023, [A&A](#), **677**, A184
- Wechsler R. H., Tinker J. L., 2018, [ARA&A](#), **56**, 435
- White S. D. M., Rees M. J., 1978, [MNRAS](#), **183**, 341
- Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, [ApJ](#), **691**, 1879
- Williams D. R. A., et al., 2022, [MNRAS](#), **510**, 4909
- Xie L., et al., 2024, [ApJ](#), **966**, L2
- York D. G., et al., 2000, [AJ](#), **120**, 1579
- Zawada K., Semelin B., Vonlanthen P., Baek S., Revaz Y., 2014, [MNRAS](#), **439**, 1615
- Zel'dovich Y. B., 1964, Soviet Physics Doklady, **9**, 195
- Zou F., Yu Z., Brandt W. N., Tak H., Yang G., Ni Q., 2024, [ApJ](#), **964**, 183
- Zwicky F., 1933, Helvetica Physica Acta, **6**, 110

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.

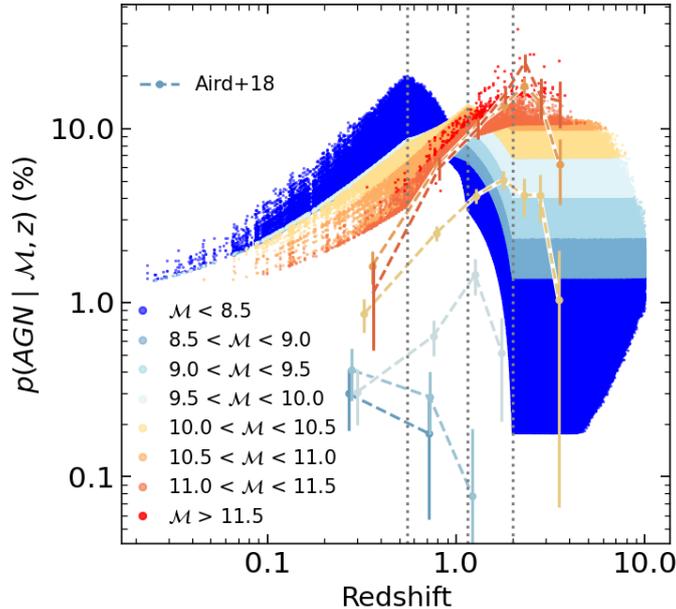


Figure A.1: Probability of a galaxy to be an AGN as in Fig. 2.4 but using the original Schechter fit from B16 to assign the AGN fraction at $z < 1.15$. The biggest difference with respect to the distribution shown in Fig. 2.4 is in the high density of low-mass AGN at low redshift.

In Sect. 2.2.1 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. 2.2.1 and 2.2.2 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 $p(\log \lambda_{\text{sBHAR}} | \mathcal{M}, z)$ distributions from A18 at all \mathcal{M} and z and statistically sample random values from those distributions in order to assign λ_{sBHAR} to every galaxy in our catalogue, and later L_X using Eq. 10. After this, AGN can be selected as objects above a given threshold in either L_X or λ_{sBHAR} .

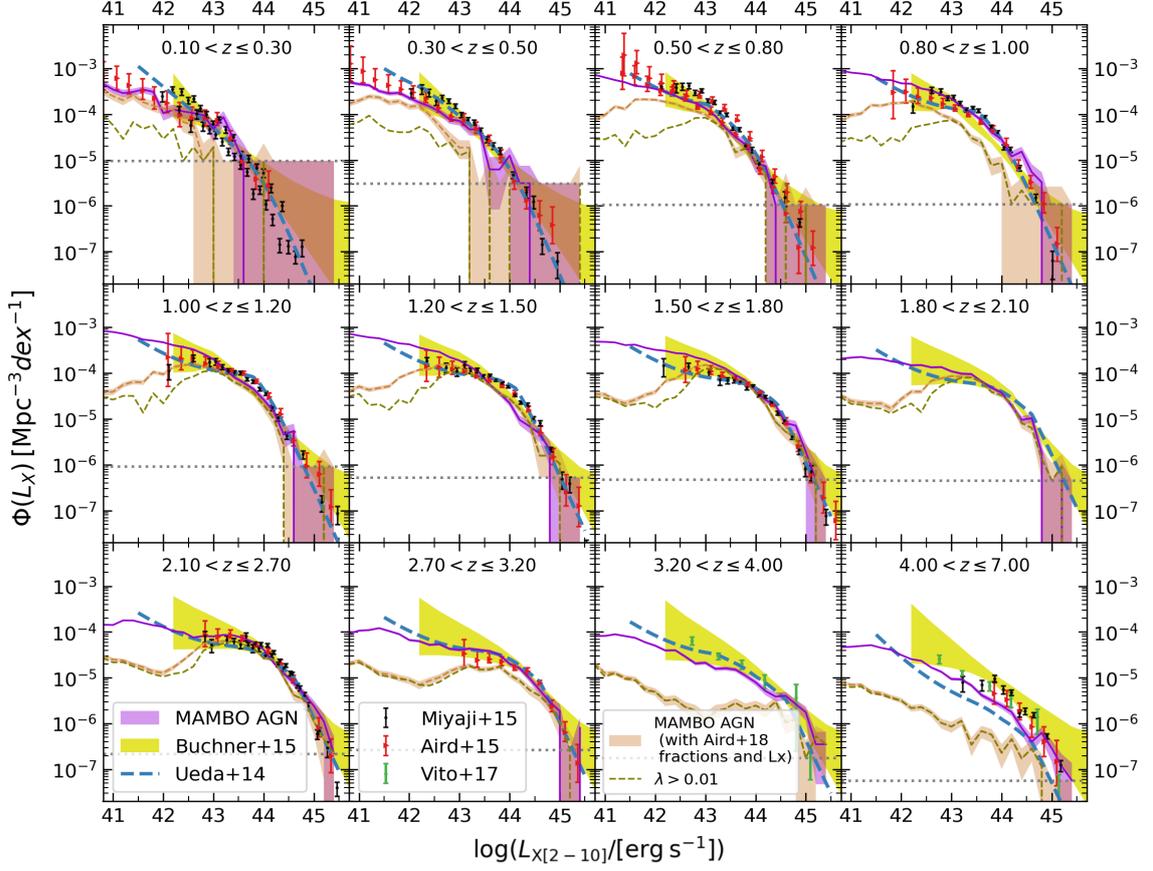


Figure A.2: Hard X-ray luminosity function, using the $p(\log \lambda_{\text{SBHAR}} | \mathcal{M}, z)$ distributions from A18 to infer the AGN fraction and L_X distribution at all \mathcal{M} and z . The dashed brown line corresponds to all the X-ray emitters, while the dashed green line corresponds to the objects selected with $\lambda_{\text{SBHAR}} > 0.01$. For comparison, the solid purple line shows the XLF from our lightcone using the methodology adopted in this paper, as in Fig. 3.2. See the caption of Fig. 3.2 for further details and references to the observed XLFs.

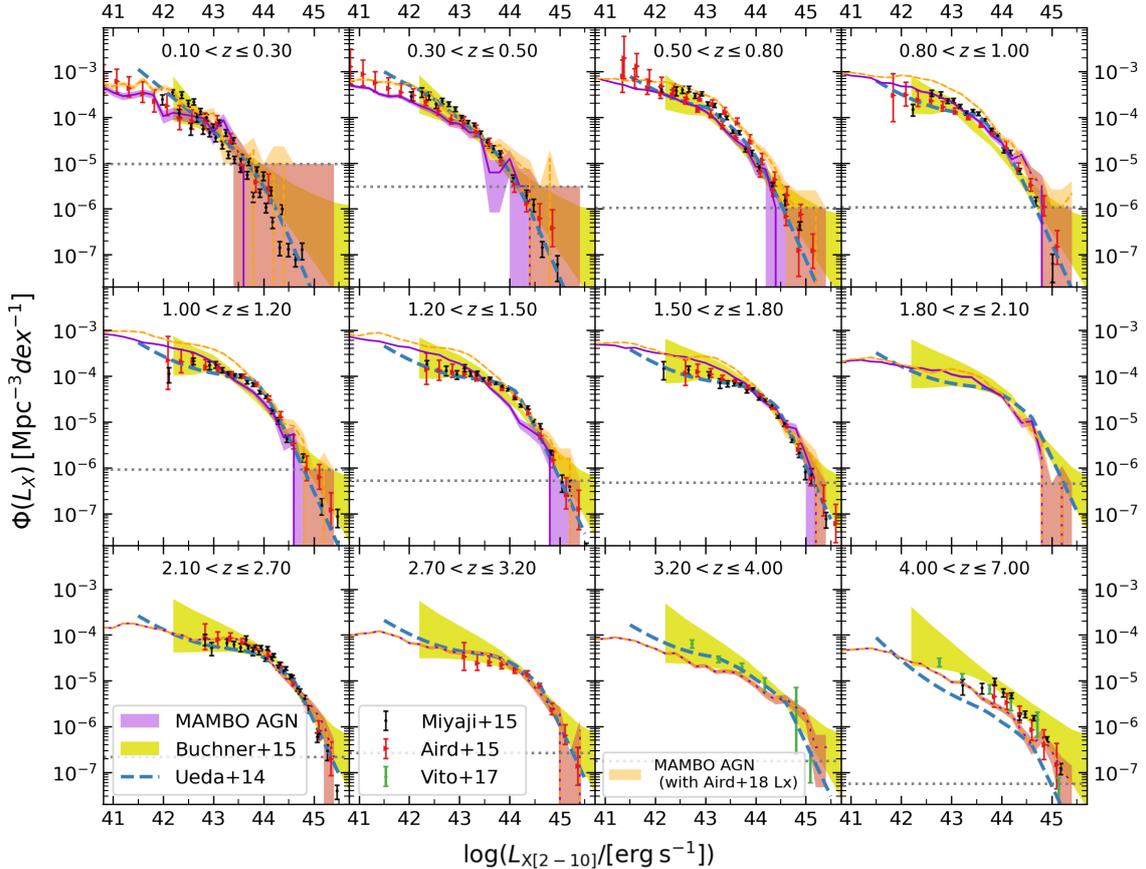


Figure A.3: Hard X-ray luminosity function, using the AGN fraction inferred from B16 (as in Sect. 2.2.1), but the $p(\log \lambda_{\text{sBHAR}} | \mathcal{M}, z)$ distributions from A18 to infer L_X at all z (dashed orange line). For comparison, the solid purple line shows the XLF from our lightcone using the methodology adopted in this paper, as in Fig. 3.2. See the caption of Fig. 3.2 for further details and references to the observed XLFs.

We show in Fig. A.2 the X-ray luminosity function resulting from this methodology, both for AGN selected with $\lambda_{\text{sBHAR}} > 0.01$ (for consistency with the definition adopted in A18 and in this work), and also for all the X-ray emitters (without any constrain in λ_{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 $p(\log \lambda_{\text{sBHAR}} | \mathcal{M}, z)$ distributions from A18 are defined up to $z \leq 4$, we also explored the possibility of using them to infer L_X at all z , starting from the AGN fraction inferred from B16 (as in Sect. 2.2.1). 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. 2.2.2.

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.