TESIS DE LA UNIVERSIDAD DE ZARAGOZA
2022
71

Daniele Pasquale Spinoso

The high-redshift formation and evolution of Super-Massive Black Holes through semi-analytic models and photometric data

Director/es
Bonoli, Silvia

http://zaguan.unizar.es/collection/Tesis

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

© Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

THE HIGH-REDSHIFT FORMATION AND EVOLUTION OF SUPER-MASSIVE BLACK HOLES THROUGH SEMI-ANALYTIC MODELS AND PHOTOMETRIC DATA

Autor

Daniele Pasquale Spinoso

Director/es

Bonoli, Silvia

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Física

2022

Repositorio de la Universidad de Zaragoza – Zaguan http://zaguan.unizar.es

UNIVERSIDAD DE ZARAGOZA

Universidad Zaragoza

FACULTAD DE CIENCIAS

ESCUELA DE DOCTORADO

The high-redshift formation and evolution of Super-Massive Black Holes through semi-analytic models and photometric data

Ph.D. THESIS

Author
Daniele Spinoso

Supervisor
Dr. Silvia Bonoli

Tutor
Prof. Gloria Luzón Marco

CEFCA
Centro de Estudios de Física del Cosmos de Aragón

Donostia International Physics Center

Facultad de Ciencias
Universidad Zaragoza

1542

Escuela
de Doctorado
Universidad Zaragoza

In a sky where no sun is shining,
in a sea where no wave is moving,
in a land where nobody is walking,
there's a light where nobody stands.

(Aucan - Black Rainbow ; 2011 ©Africantape/Ruminance)

ACKNOWLEDGMENTS

The effort thanks to which the content of this thesis work was put together has been made significantly easier by the continuous guidance, support and sustain of a multitude of people, to whom I express my enormous gratitude.

I would like to start from my supervisor, Silvia Bonoli, who patiently followed my oscillations between the observational and theoretical topics of this thesis during the last years. Her continuous and insightful inputs, and her willingness to discuss results or ideas, taking advantage of every free minute available, have constantly pushed and driven the improvement of my analysis as well as my scientific growth as a student. In addition, a significant part of my work would not have been possible without the supervision and guidance of Álvaro Orsi and Kerttu Viironen, which guided me through the first analysis of preliminary `J-PLUS` data, motivating the search for elusive results with inexhaustible passion and enthusiasm. A special thanks goes to Carlos López Sanjuán, which took the burden of helping me to finalize the observational results of this work, and to Alessandro Ederoclite for making clear that observational proposals should be submitted *before* their deadline (sic transit gloria mundi). My sincere gratitude goes also to Massimo Dotti, whose insightful and focused guidance during my master thesis ultimately gave me the possibility to opt for a Ph.D. and pursue my interests in Astronomy. I would also like to acknowledge the assistance of Naoki Yoshida and Tilman Hartwig which made my working visit in Tokyo extremely fruitful and stimulating. Finally, I would like to acknowledge the help and support of my tutor Gloria Luzón, as well as of Theopisti Dafni. Their professional and considerate presence has always been a secure reference at the University of Zaragoza.

As for the human cost of my Ph.D., I was only able to carry it thanks to the presence of a wide ensemble of people at my side. All of them deserve my extreme gratitude. Arianna, whose steady, strenuous and selfless support always showed me the transient and ephemeral nature of anxiety and discouragement. Her external view on "*that weird stuff you do*" greatly helped me to put in perspective and overcome the challenges and choices posed by my Ph.D. along these years. My brother Riccardo, with his crystalline excitement for good news and effortless encouragement, my parents Caterina and Mimmo with their silent support, extreme enthusiasm and life-long motivation to pursue knowledge and improve myself, along with all the extended ensemble of aunts, uncles and cousins, which steadily followed and motivated my trajectory. All the *becarios* of CEFCA, with whom I had the privilege to share the cold, dark and door-less "*pasillo de los becarios*": Giovanni, Gonzalo, Jonas, Lorenzo, Matteo, Rafa and Siddhartha. Long lasting friendships were forged through the countless series of events we shared along these years, which motivated my efforts and ultimately built-up to an ensemble of bright memories defining the years in Teruel. A very special acknowledgment and my full gratitude goes to David, my closest companion among the becarios. Without his presence, his continuous input, constructive criticism and example, my Ph.D. journey would have just not been possible. Stretching outside the boundaries of Astronomy, my gratitude extends to the extremely close friends which have sustained my efforts from afar. Their constant presence and support have strongly defeated the limits imposed by the distance, offering a solid reference to which turning has always been immensely pleasant. In sparse order, but all with the same intensity, my sincere thanks go to Jack, Jan, Tasso, Madda, Edo, Calo, Nitti, Milo, Davide, Simo, Abba, Alice, Melo, Tommo, Mala, Teo, Tom, Marcio, Irene and Tommy Zana. For all the truly countless times in which your human contact has supported my efforts during the last years.

Finally, I would like to acknowledge the members of Victoria Station Disorder, SMP and Circolo Arci L'impegno. Unfortunately, in this case "*non sarà mai più come prima*" is an outstanding truth; nevertheless the memories of those truly uplifting experiences have filled the emptiness of nostalgia along my path along my Ph.D. years.

The high-redshift formation and evolution of Super-Massive Black Holes through semi-analytic models and photometric data

Abstract

The formation and early evolution of Supermassive Black Holes (SMBHs) in the high redshift Universe ($z > 6$) represents one of the most enigmatic open issues in modern Astrophysics. Indeed, a growing wealth of observational evidences points towards the existence of billion-solar-masses BHs within just $\sim$1Gyr from the Big-Bang, powering extremely luminous quasars (QSOs). Despite intense theoretical efforts in the last decade, current models still struggle to fit the formation of these extreme objects in such a short cosmological time. Furthermore, it is not yet clear how these high-z SMBHs are related to the more common lower-redshift population of SMBHs which are thought to be ubiquitously hosted in the nuclei of massive galaxies. Explaining the formation and growth of a global SMBH population in its cosmological context presents extreme theoretical difficulties which arise from the need to orchestrate the small-scale physics of gas cooling, star-formation and its radiative and chemical feedback on primordial gas, with the large-scale structure formation and evolution. On the numerical side, the formation of SMBHs is generally studied by exploiting high-resolution, small-scale simulations which can effectively capture the small-scale physics involved in this process. Nevertheless, the application of these results to wide, cosmological simulated boxes is computationally prohibitive, therefore the generalization of these results to cosmological scales is still uncertain. On the observational side, current and next-future experiments are providing improved constraints on the cosmological evolution of the SMBH population across cosmic history. In particular, extensive spectroscopic surveys and wide-area narrow-bands photometric surveys offer a complementary view on the population of luminous active galactic nuclei (AGN) and QSOs, allowing to constrain the theoretical models for the formation and growth of SMBHs. This thesis presents a novel approach which tackles this complex phenomenon through a combination of numerical methods and observational techniques. More in detail, we embed a comprehensive model for the formation and growth of SMBHs in the L-Galaxies semi-analytical model (SAM). We then apply our prescriptions to the merger trees of the Millennium-II N-Body simulation, which offers an optimal compromise between mass-resolution and simulated volume. This allows to study the occurrence of SMBHs formation via all the currently-envisioned physical processes, as well as to follow self-consistently the evolution of SMBHs within their cosmological context. Therefore, this represents one of the first attempts to model self-consistently the evolution of a cosmological SMBH population, emerging only from high-z formation processes. We complement this theoretical approach with an observational study of the AGN/QSOs Lyman-alpha luminosity function (LF) at $2 \lesssim z \lesssim 3.3$ by exploiting the wide-area, multi-band photometric survey J-PLUS. Thanks to the wide area sampled by this survey with narrow-band photometry, we are able to extend for the first time the constraints on the very bright-end of the Lyman-alpha luminosity function over a previously un-sampled region of luminosity and density. This sets a constraint to the far-UV luminosity distribution of accreting SMBHs at the epoch of the cosmological-peak AGN activity. The combination of our numerical and observational results suggest that high-z ($z > 6$) processes leading to the formation of SMBHs can actively account both for the formation of extremely massive objects and QSOs after $\sim$1 Gyr from the Big-Bang, as well as the cosmological build-up of the global SMBH population observed at more moderate redshifts ($2 \lesssim z \lesssim 3.3$). Furthermore, by employing a combination of numerical and observational techniques, our study is suitable for being generalized to larger simulated volumes (on the theoretical side) and future wide-area surveys of the high-redshift Universe (on the observational side).

The high-redshift formation and evolution of Super-Massive Black Holes through semi-analytic models and photometric data

Resumen

La formación y evolución de los Agujeros Negros Supermasivos (SMBHs) en el Universo temprano ($z>6$) representa una de las cuestiones abiertas más enigmáticas de la Astrofísica moderna. De hecho, una creciente cantidad de evidencias observacionales apunta a la existencia de BHs de mil millones de masas solares a tan sólo 1Gyr del BigBang, impulsando cuásares extremadamente luminosos (QSOs). A pesar de los intensos esfuerzos teóricos de la última década, los modelos actuales siguen teniendo dificultades para encajar la formación de estos objetos extremos en un tiempo cosmológico tan corto. Además, aún no está claro cómo se relacionan estos SMBH de alto z con la población más común de SMBH de desplazamiento al rojo inferior, que se cree que está alojada de forma ubicua en los núcleos de las galaxias masivas. Explicar la formación y el crecimiento de una población global de SMBH en su contexto cosmológico presenta dificultades teóricas extremas que surgen de la necesidad de orquestar la física a pequeña escala del enfriamiento del gas, la formación de estrellas y su retroalimentación radiativa y química sobre el gas primordial, con la formación y evolución de la estructura a gran escala. Desde el punto de vista numérico, la formación de los SMBH se estudia generalmente mediante simulaciones de alta resolución a pequeña escala que pueden captar eficazmente la física a pequeña escala implicada en este proceso. Sin embargo, la aplicación de estos resultados a cajas simuladas amplias y cosmológicas es computacionalmente prohibitiva, por lo que la generalización de estos resultados a escalas cosmológicas es aún incierta. Desde el punto de vista de la observación, los experimentos actuales y futuros están proporcionando mejores restricciones sobre la evolución cosmológica de la población de SMBH a lo largo de la historia cósmica. En particular, los estudios espectroscópicos extensos y los estudios fotométricos de banda estrecha de área amplia ofrecen una visión complementaria de la población de núcleos galácticos activos luminosos (AGN) y QSO, lo que permite restringir los modelos teóricos para la formación y el crecimiento de los SMBH. Esta tesis presenta un enfoque novedoso que aborda este complejo fenómeno mediante una combinación de métodos numéricos y técnicas observacionales. Más detalladamente, incrustamos un modelo completo para la formación y el crecimiento de los SMBHs en el modelo semi-analítico (SAM) de las L-Galaxias. A continuación, aplicamos nuestras prescripciones a los árboles de fusión de la simulación Millennium-II N-Body, que ofrece un compromiso óptimo entre la resolución de masa y el volumen simulado. Esto permite estudiar la ocurrencia de la formación de SMBHs a través de todos los procesos físicos actualmente previstos, así como seguir de forma autoconsistente la evolución de los SMBHs dentro de su contexto cosmológico. Por lo tanto, esto representa uno de los primeros intentos de modelar de forma autoconsistente la evolución de una población cosmológica de SMBHs, emergiendo sólo de procesos de formación de alto z. Complementamos este enfoque teórico con un estudio observacional de la función de luminosidad Lyman-alfa (LF) de los AGN/QSOs a $2 \lesssim z \lesssim 3.3$ explotando el estudio fotométrico multibanda de área amplia J-PLUS. Gracias a la amplia área muestreada por este sondeo con fotometría de banda estrecha, somos capaces de ampliar por primera vez las restricciones en el extremo muy brillante de la función de luminosidad Lyman-alfa sobre una región de luminosidad y densidad no muestreada previamente. Esto establece una restricción a la distribución de la luminosidad en el ultravioleta lejano de los SMBH en acreción en la época del pico de actividad cosmológica de los AGN. La combinación de nuestros resultados numéricos y observacionales sugiere que los procesos de alto z ($z>6$) que conducen a la formación de SMBHs pueden explicar activamente tanto la formación de objetos extremadamente masivos y QSOs después de 1 Gyr desde el BigBang, como la acumulación cosmológica de la población global de SMBHs observada a desplazamientos al rojo más moderados ($2 \lesssim z \lesssim 3.3$). Además, al emplear una combinación de técnicas numéricas y observacionales, nuestro estudio es adecuado para ser generalizado a volúmenes simulados más grandes (en el lado teórico) y a futuros estudios de área amplia del Universo a alto desplazamiento al rojo (en el lado observacional).

Index

Thesis outline

The Ph.D. work presented in this thesis is based on two main, complementary branches which rely on both theoretical and observational approaches to tackle the study of formation and evolution of super-massive Black Holes. Therefore, the methods and results presented in this manuscript are rooted on a wide range of topics and methodologies. To provide a structured presentation of all the concepts, definitions and current knowledge which support this work, we separate the technical aspects of our introduction from the theoretical ones. The latter are presented in chapter §1, while §2 presents an introduction to the numerical and observational techniques we use. Nevertheless, for the sake of discussion, some of the concepts which are thoroughly introduced and detailed in §2 are used in §1 by making use of preliminary, quick definitions where needed.

Following these introduction sections, we present the main results of the two research lines pursued by our work. In particular, chapter §3 details the model for the formation of SMBHs we developed and the results it produced, chapter §4 presents the observational determination of the luminosity function of $2 \lesssim z \lesssim 3.3$ AGN evaluated at their rest-frame ultra-violet wavelengths. In chapter §6 we discuss our results and fit them into an organic picture, analyzing their outcomes, applications and possible future developments. Finally, we also provide both an initial abstract and final remarks to summarize the major results and novelty of the presented work.

Units and main physical quantities

Throughout this work we make use of the `cgs` units system, following common conventions in Astronomy and Astrophysics. According to this system, physical distances are measured in centimeters (cm), mass is measured in grams (g) and time in seconds (s). With this choice, energy is measured in erg, namely: $1\,erg = 1\,cm^2\,g\,s^{-2}$ In addition, we use Kelvin units (K) to measure temperatures, whenever needed. Since the `cgs` system can be uncomfortable for astronomical quantities, we couple `cgs` units to `solar` units. Note that when using the latter, we always refer to their values in `cgs` units. According to these choices, the numerical values of common physical quantities are as follows:

Speed of light	$c = 2.9979 \times 10^{10}\,cm\,s^{-1}$
Gravity constant	$G = 6.6726 \times 10^{-8}\,cm^3\,g^{-1}\,s^{-2}$
Boltzmann's constant	$k = 1.3807 \times 10^{-16}\,cm^2\,g\,s^{-2}\,\mathrm{K}^{-1} = erg\,\mathrm{K}^{-1}$
Planck's constant	$h = 6.6261 \times 10^{-27}\,erg\,s$
Proton mass	$m_p = 1.6726 \times 10^{-24}\,g$
Hydrogen atom mass	$m_\mathrm{H} = 1.6733 \times 10^{-24}\,g$
Thomson cross section	$\sigma_\mathrm{T} = 6.6524 \times 10^{-25}\,cm^2$
Astronomical Unit	$\mathrm{UA} = 1.4959 \times 10^{13}\,cm$
Parsec	$\mathrm{pc} = 3.0857 \times 10^{18}\,cm$
Year	$\mathrm{yr} = 3.154 \times 10^{7}\,s$
Solar mass	$\mathrm{M}_\odot = 1.9885 \times 10^{33}\,g$
Solar Luminosity	$\mathrm{L}_\odot = 3.828 \times 10^{33}\,erg\,s^{-1}$

Table 1: Summary of units and main physical quantities used in this work

Chapter 1

Introduction

The last decades might be considered as one of the most fervent periods of progress in our understanding of astronomical phenomena and cosmological observables in modern history. A growing wealth of experimental results, observational constraints and theoretical descriptions is being accumulated over a wide range of fields in both astronomy and cosmology. The study of the formation, growth and cosmological evolution of compact objects is one of the topics which have undergone a tremendous impulse, driven by fundamental observations and results. Among many, the latter include the observation of extremely luminous point-like sources which are thought to be associated to extremely-massive, compact objects already in place when the Universe was less than 1 Gyr old (see e.g., Fan et al., 2006b; Mortlock et al., 2011; Venemans et al., 2013; Wu et al., 2015; Mazzucchelli et al., 2017; Bañados et al., 2018, 2021; Wang et al., 2021), the detection of gravitational waves (GWs, see e.g., Abbott et al., 2016), the association between a GW event and its electromagnetic counterpart (see Abbott et al., 2017), the estimate of the mass of the SgrA$*$ compact source in the Milky-Way (MW) center (e.g., Ghez et al., 2008) and finally the direct observation of the M87-galaxy center (e.g., Event Horizon Telescope Collaboration et al., 2019). These milestones cumulatively hints towards the existence of compact objects with extreme masses, i.e. in excess of millions of solar masses ($M_\odot$), residing at the centers of massive galaxies and co-evolving with them across the whole cosmic history (see e.g. Heckman & Best, 2014, for a review). These compact objects are thought to be the super-massive counterparts of stellar black holes (BHs) and are therefore referred-to as super-massive black holes (SMBHs). Their formation, evolution and observation represent the main topics of the current thesis work.

Open questions and thesis rationale

While the origin of relatively small BHs is fairly-well understood thanks to stellar evolution models, the formation of SMBHs is currently enshrouded in uncertainty. Indeed, extreme challenges are required to theoretically model and observationally constrain it. On the other hand, the cosmological history of SMBHs as well as their co-evolution with their hosting galaxies has been widely studied (Kormendy & Ho, 2013, e.g.,). This allowed to reach reasonable consensus on many key aspects, such as the connection between the accretion of matter onto SMBHs and the intense activity observed at different wavelengths from the central regions of some galaxies (AGN, from *Active Galactic Nuclei*, e.g. Shakura & Sunyaev, 1976; Soltan, 1982). In particular, a significant progress in the study of galaxy-SMBHs co-evolution has been the discovery of relationships between observables (i.e. measurable properties) of SMBHs and galaxies, such as the mass of SMBHs and the one of their host galaxies, as well as between the SMBHs masses and the velocity dispersion of stars within their host galactic centres (see e.g. Ferrarese & Merritt, 2000; Gebhardt et al., 2000; Gültekin et al., 2009b; Shankar, 2013; Fiore et al., 2017). Nevertheless, while the constraints on these relationships are being improved in time, several open question still remain about the global population of SMBHs. For instance, it is expected that SMBHs should merge with each other, but a direct observational confirmation of this phenomenon has yet to be obtained. Furthermore, it is theoretically predicted that the merger of compact objects might impress a "kick-velocity" to the remnant object, ejecting it from the host galaxy in extreme cases. It has therefore been suggested that a population of "wandering" SMBHs might exist in the surroundings of massive galaxies (e.g., Volonteri,

2007; Volonteri & Madau, 2008; Volonteri et al., 2010; Tremmel et al., 2018; Izquierdo-Villalba et al., 2020). Finally, despite stellar evolution and mass-growth models can explain to some extent the presence of SMBHs, the actual origin of the population of SMBHs hosted at the centers of massive galaxies is substantially unknown.

This thesis addresses a part of these open questions by focusing on the study of formation and evolution of SMBHs in a cosmological context. The approach on which it is based join both observational and numerical methods to link the theoretical knowledge about SMBHs formation to the observational evidences describing their evolution. In particular, this work embeds the current models for the formation of SMBHs in their cosmological context and study the production of an initial SMBHs population as well as its cosmological evolution, both numerically and observationally. The goal of this analysis is to understand if and how current SMBHs formation models are able to explain the build-up in time of the large SMBHs populations currently observed. The final outcome of this study is a comprehensive study which connects the high-z formation of SMBHs to their subsequent evolution, with a number of possible future applications both on the theoretical and observational fields of astrophysics.

1.1 Current paradigms

Since astronomical observations were extended beyond optical wavelengths, breakthrough discoveries have confirmed fundamental theoretical predictions and have driven key advancements in the understanding of the whole Universe. Typical examples of this are the serendipitous detection of the *Cosmic Microwave Background* (CMB) through measurements at radio-wavelengths (Penzias & Wilson, 1965), or first direct evidences and constraints on the presence of Dark Matter (DM) from X-ray observations[1] (Clowe et al., 2004; Markevitch et al., 2004). Analogously, infra-red (IR) observations helped to detect and study the emission of high-z galaxies and QSOs, thus helping to build a coherent interpretation of galaxy formation and evolution throughout the history of our Universe. In this section we contextualize our work by introducing its theoretical basis. In particular, §1.1.1 presents the current picture for the growth of structures in the Universe while §1.1.2 describes the evolution of galaxies within them. Finally, §1.1.3 introduces the current knowledge about SMBHs residing at the center of most massive galaxies.

1.1.1 The growth of structures in the ΛCDM model

The astronomical panorama observable by naked-eye seems to suggests that visible matter is distributed into discrete objects separated by large "dark" spaces. Furthermore, these objects are not uniformly distributed over the night sky, but appear denser towards a specific direction, i.e. the projection of the disk-like structure of our Galaxy over the night-sky plane (namely, the "Milky Way" or MW). Nevertheless, thanks to large astronomical surveys, nowadays we now know that the Universe extends much farther than the MW and that its non-homogeneity is only manifest on very-small astronomical distances. This observation is formalized by the so-called *cosmological principle*, which can be expressed as: "the Universe is homogeneous and isotropic over large-enough spatial scales". Indeed, when considering distances of $\gtrsim 100\,\mathrm{Mpc}$[2], observable matter in the Universe shows a relatively uniform distribution, without over-densities or specific patterns towards any preferential direction (Wu et al., 1999).

The fundamental importance of the cosmological principle is to allow a global description of the Universe in the form of a homogeneous and isotropic *fluid* comprising different components, each with specific physical properties (e.g. Ryden, 2016). In particular, according to the currently more favoured cosmological models, our Universe includes five main components: radiation, neutrinos, ordinary matter[3],

[1] Here we only refer to X-ray detection of DM, since its presence was already required by both theoretical arguments and optical observations (see e.g. Ostriker & Peebles, 1973; Bosma, 1978; Fall & Efstathiou, 1980; Carignan & Freeman, 1985).

[2] A Mega-parsec (Mpc) corresponds to 10^6 parsecs (pc), which is a common unit of measure in astronomy and cosmology. Parsecs have been defined historically as the distance of an object (e.g. a star) showing an angular apparent-motion of 1 arcsecond with respect to farther stars. This apparent motion is called parallax, while 1 arcsecond is exactly the 1/3600th part of an angular degree. In the standard international metric system units, one parsecs can be expressed as: $1\,\mathrm{pc} = 3.086 \times 10^{16}\,\mathrm{m}$.

[3] Ordinary matter is generically referred-to as "baryonic" matter due to the strong dominance of protons and neutrons (baryons) over electrons (fermions) in terms of total mass contribution.

DM and Dark Energy (or *cosmological constant* Λ). Each of these contributes to the total energy content[4] of the Universe, with a degree that has been measured with increasing precision in the last decades, especially after satellite experiments such as `COBE` (Bennett et al., 1996), `WMAP` (Hinshaw et al., 2009) and `Planck` (Planck Collaboration et al., 2020) provided their observational contraints on the CMB signal. Indeed, these favoured cosmological models characterized by the prevalence of Λ and non-relativistic ("cold") DM, hence leading to the definition of the Λ-Cold Dark Matter (Λ-CDM) paradigm, according to which Λ and CDM respectively accounting for the $\sim$26% and $\sim$69% of the total energy-density balance (Planck Collaboration et al., 2020).

The ΛCDM model for the expanding Universe

Dark Matter and Λ can be associated respectively to two fundamental observed properties of our Universe, respectively: the presence of gravitational forces and the accelerated expansion of its space-time fabric. These two can be thought as competing factors, with gravity producing the contraction of matter in space and time, and the expansion fostering structures to drift apart.

The observational evidence of the accelerated expansion of the Universe was established by studying the relation between the intrinsic and measured emission of SuperNovae Type-Ia (SNe-Ia), i.e. a specific class of astronomical sources (e.g. Riess et al., 1998; Perlmutter et al., 1999). This relation, indeed, carries fundamental information about the sources distance d, which can be compared to the predictions drawn by a given *cosmological model*, i.e. a theoretical description of the Universe geometry and evolution. The starting point of these analysis can be traced back to the identification of the proportionality between d and the sources redshift z, i.e. the shift towards long wavelengths of their emitted light. By interpreting z as a doppler shift (i.e. $z = v/c$), Hubble (1929) enunciated the *Hubble's law*:

$$v = \mathrm{H}_0 \cdot d\,, \tag{1.1}$$

where H_0 is a constant (the *Hubble's constant*) which describes the rate of recession between two points in space at a given time t_0. Hubble's law can be generalized to any time t by introducing the *hubble's parameter* $H(t)$ which describes the evolution of the Universe expansion rate through time.

The formation of structures in a scenario where space-time expansion and gravitational forces compete is currently explained by applying the principles of General Relativity (GR, Einstein, 1915) to the Universe as a whole. In particular, by solving the Einstein's field equations of GR for an expanding space-time in spherical coordinates, it is possible to obtain the exact solution known as the Friedman-Lemaître-Robertson-Walker (FLRW) metric:

$$dS^2 = c^2\,dt^2 - a^2(t) \cdot \left[\frac{dr^2}{1 - \mathrm{K}\,r^2} + r^2 d\Omega^2\right], \tag{1.2}$$

where c is the speed of light, while the factor K accounts for the geometric curvature of space-time ($\mathrm{K} = -1$, 0 and 1 respectively for negative, null or positive curvature).

The quantity $a(t)$ is a dimensionless *scale factor* which encodes the time-dependent expansion of space, hence allowing to define the *comoving* coordinates, i.e.: $[r\ ;\ \Omega = r(d\theta + sin\theta\, d\phi)]$ as the system of reference in which the distance between two points in space does not depend on time. As an example, the physical distance $R(t)$ and the comoving one (r) are related through: $R(t) = a(t) \cdot r$, which leads to reformulate the Hubble's law as:

$$v = \frac{d\,R(t)}{dt} = \frac{da}{dt} \cdot r = \frac{da}{dt} \cdot \frac{R(t)}{a(t)} = \left[\frac{da/dt}{a}\right] \cdot R(t)\,, \tag{1.3}$$

and therefore to obtain a formal definition of the Hubble's parameter:

$$H(t) = \frac{\dot{a}}{a}\,. \tag{1.4}$$

[4] Thanks to the Einstein's energy-mass equivalence (Einstein, 1905), in this section we use the expressions "energy density" and "mass density" as synonims.

By definition, the value of $a(t)$ at the present time t_0 is set to be $a(t_0) = 1$, so that physical and comoving coordinates coincide in this particular instant.

Applying the FLRW metric to the Einstein's field equations leads to the fundamental Friedman's equation, which relates the evolution of $H(t)$ to the total content of matter-energy in the Universe:

$$H^2(t) = \frac{8\pi G}{3}\rho - \frac{\mathrm{K}\,c^2}{a^2} \tag{1.5}$$

where $\rho = \varepsilon/c^2$ is the total density of matter-energy. In a Universe containing multiple components, the latter can be obtained as a sum of different contributions: $\rho = \sum_i \rho_i$. By defining the *critical density* as the value $\rho_c(t)$ for which $\mathrm{K} = 0$, the equation above reduces to:

$$\rho_c(t) = \frac{3\,H^2(t)}{8\pi G} \quad ; \quad \rho_c(t=0) = \rho_c^0 = \frac{3\,H_0^2}{8\pi G}\,, \tag{1.6}$$

and hence the generic case in which $\mathrm{K} \neq 0$ can be rewritten as:

$$\frac{H^2(t)}{H_0^2} = \frac{\sum_i \rho_i}{\rho_c^0} - \frac{\mathrm{K}c^2}{a^2\,H_0^2} = \sum_i \Omega_i - \frac{K c^2}{a^2\,H_0^2}\,, \tag{1.7}$$

with the quantities:

$$\Omega_i = \frac{8\pi G}{3\,H_0^2}\cdot\rho_i(t)\,. \tag{1.8}$$

being the *density parameters* of each component in the Universe. The curvature term, on the contrary, is associated to the geometric properties of space-time instead of a mass-energy density. As a consequence, its contribution is usually determined indirectly by evaluating eq. 1.7 at t_0, i.e. when $H(t) = H_0$ and $a(t) = 1$:

$$1 - \sum_i \Omega_i = 1 - \Omega_{tot} = -\frac{\mathrm{K}c^2}{H_0^2} = \Omega_k\,, \tag{1.9}$$

so that the Friedman's equation finally becomes;

$$\frac{H^2(t)}{H_0^2} = \sum_i \Omega_i - \frac{\Omega_k}{a^2}\,. \tag{1.10}$$

The density evolution ρ_i of each component can be obtained by assuming a generic equation of state in the form $P_i = \omega\,\rho_i\,c^2$ (with ω being a parameter which identifies each component) and applying the energy-conservation equation for the case of an adiabatically-expanding fluid:

$$dU = -P_i\,dV\,, \tag{1.11}$$

where V is the volume of the expanding fluid and U its total internal energy, so that: $dU = d(\rho_i\,c^2\cdot V)$. Therefore, by considering time derivatives of both definitions of dU:

$$-P_i\dot{V} = -P_i\frac{dV}{dt} = \frac{dU}{dt} = \frac{d(\rho_i\,c^2\,V)}{dt} = \dot{\rho}\,c^2\,V + \rho\,c^2\,\dot{V}\,, \tag{1.12}$$

and using $P_i = \omega\,\rho_i\,c^2$, these equations can be solved for $\rho_i(t)$:

$$\rho_i(t) = \rho_i(t_0)\cdot a(t)^{-3(1+\omega)} \tag{1.13}$$

Both baryonic matter and DM can be considered pressure-less components, so that $\omega_m = 0$ (with m identifying the ensembre of both kinds of matter). On the other hand, radiation is a relativistic fluid for which the equation $P_{\mathrm{rad}} = \frac{1}{3}\rho c^2$ holds, hence $\omega_{\mathrm{rad}} = 1/3$. Finally, the mass-energy density of the cosmological constant must not depend on time, by definition, therefore: $\omega_\Lambda = -1$. By using these values in the density parameters definition for each component (eq. 1.8) we obtain expressions in the form: $\Omega_i = \Omega_{i;0}\,a^\beta$, with the exponent β varying for each component and $\Omega_{i;0}$ being the density parameter of

the i_{th} component evaluated at t_0. Substituting into eq. 1.10, we finally obtain the explicit dependence of the Friedman equation on the scale parameter:

$$\frac{H^2(t)}{H_0^2} = \left[\frac{\Omega_{\text{rad};0}}{a^4} + \frac{\Omega_{m;0}}{a^3} + \Omega_\Lambda - \frac{\Omega_k}{a^2}\right] . \tag{1.14}$$

We underline that $\Omega_{m;0}$ includes both the contribution of baryonic matter and DM, while we neglected the contribution of neutrinos to keep a synthetic approach. Equation 1.14 compactly describes the evolution of $H(t)$ for a series of Universe models, each described by a set of cosmological parameters Ω_i. As already commented, the current ΛCDM paradigm predicts $\Omega_{\text{rad};0} \sim 0$, $\Omega_{m;0} \sim 0.315$, $\Omega_\Lambda = 0.685$, $\Omega_k = 0$ and $H_0 \sim 67.4\ \text{km s}^{-1}\,\text{Mpc}^{-1}$ (Planck Collaboration et al., 2020). Usually, the value of $H(t)$ or H_0 are provided in units of h, which is defined as: $H_0 = 100\,\text{h}$. Therefore, h only represents a parametrization of H_0 which disentangles the numerical factors in cosmological equations from the exact value of H_0[5].

Formation of structures from density perturbations

Despite being currently object of some tension regarding the value of H_0 (e.g. Freedman, 2017), the general picture we delineated was recently supported by observations of the CMB (Hinshaw et al., 2009; Planck Collaboration et al., 2020). These also assessed the the overall validity of the cosmological principle, by showing a remarkably uniform temperature of the CMB, at least down to very small anisotropies ($\Delta T_{\text{CMB}} / \langle T_{\text{CMB}} \rangle \approx 10^{-5}$). These are currently interpreted as an effect of local density perturbations $\delta(x, y, z)$ in the 3D distribution of matter-density[6] $\rho(x, y, z)$, which grew in time to form the currently-observed galaxies and cluster of galaxies.

The overall growth of perturbations δ can be studied by assuming a *linear regime*, in which δ is small with respect to the average density of the Universe $\overline{\rho}$, namely:

$$\delta(\mathbf{x}) = \frac{|\delta(\mathbf{x}) - \overline{\rho}\,|}{\overline{\rho}} \ll 1 , \tag{1.15}$$

where $\mathbf{x}$ is a generic 3D position in space. In other words, the overall field of matter-density can be described as a constant, average density with a spatially varying perturbation function:

$$\rho(\mathbf{x}) = \overline{\rho} + \delta(\mathbf{x}) . \tag{1.16}$$

We can consider a volume-element of this field, characterized by local density δ and peculiar velocity $\mathbf{u} = \mathbf{v}' - v$, where $\mathbf{u}$ and $\mathbf{v}'$ respectively are the element velocity in comoving and physical coordinates, while $v = H \cdot R(t)$ is the *hubble flow* defined in eq. 1.3. The motion of this volume-element is regulated by the Euler fluid equations, expressed in comoving coordinates:

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla[(1+\delta)\mathbf{u}] = 0 , \tag{1.17}$$

$$\frac{\partial \mathbf{u}}{\partial t} + H\mathbf{u} + \frac{1}{a}(\mathbf{u}\nabla)\mathbf{u} = -\frac{1}{a}\nabla\phi , \tag{1.18}$$

where ϕ is the gravitational potential in which the volume-element moves. ϕ can be obtained by applying the Poisson equation to the perturbed field $\rho(\mathbf{x})$, from which it follows:

$$\nabla^2\phi = 4\pi G\,\overline{\rho}\,a^2\delta(\mathbf{x}) . \tag{1.19}$$

The joint solution of equations 1.17, 1.18 and 1.19 describes the time-evolution of density perturbations. Under simplifying assumptions, namely $\delta(\mathbf{x}) \ll 1$ and $\mathbf{v}' \sim 0$, it is possible to linearize it:

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G\,\overline{\rho}\,\delta . \tag{1.20}$$

[5] The h factor must not be confused with the Planck's constant h (see the thesis outline, above).

[6] When studying the growth of structures in the Universe it is generally assumed that DM drives the evolution of the overall matter content, given its prevalence over ordinary, baryonic matter. As a consequence, the term "matter" refers exclusively to DM in the following discussion.

The coefficients $H(t)$ and $\overline{\rho}(t)$ of this equation depend on the cosmological model. For instance, the geometrically flat, matter-only model proposed by Einstein-deSitter (EdS, Einstein & de Sitter, 1932) produces two different solutions: a *growing* (δ^+) and a *decaying* (δ^-) mode. In the first case, perturbations grow in time at a rate $\delta^+(t) \propto t^{2/3}$, while in the latter case they are damped as $\delta^-(t) \propto t^{-1}$. It can be shown that the scale parameter is related to time as $a(t) \propto t^{2/3}$, so that $\delta^+(t) \propto a(t)$ in the EdS model (e.g. Liddle & Lyth, 2000; Longair, 2008). Growing and decaying modes can be also derived analytically for different cosmological models, but in general their expression shows a complex form, i.e.:

$$\delta^+ = \frac{5\,\Omega_{\mathrm{tot};0}}{2}\left(\frac{1}{a}\frac{da}{dt}\right)\int_0^a \frac{da}{\dot{a}^3}\,, \tag{1.21}$$

where it is evident that the growth-rate of $\delta(t)$ depends on the details of the $a(t)$ function. Finally, the above equation can be used to define the growth-factor of perturbations as: $D = \delta^+(a)/\delta^+(a_{\,\mathrm{ref}})$, where $a_{\,\mathrm{ref}}$ is an arbitrary scale-factor value of reference (i.e. a reference time in the Universe history).

Virialization of DM halos

The growth of density perturbations can be effectively described by resorting to linearized equations[7] until $\delta(\mathbf{x}) \ll 1$. When this condition is not verified, the mathematical description of structures formation becomes highly non-linear. Only few specific cases can be addressed analytically, in particular by requiring that the collapsing structure satisfies specific symmetries. An example of this is the *spherical collapse* which describes the condensation of spherically-symmetric structures with uniform (i.e. *top-hat*) density distribution within a given radius $\mathrm{R} \ll H(t)/c$.

This process can be described as the motion of a spherical element subject at the same time to gravity and to the Hubble flow produced by Λ:

$$\frac{d^2\mathrm{R}}{dt^2} = -G\frac{M(<\mathrm{R})}{R^2} + H_0^2\Omega_\Lambda\, R\,. \tag{1.22}$$

In other words, this equation represents the decelerated radial motion of a test-particle with initial velocity $v = H_0^2\Omega_\Lambda\, R$. Indeed, the density perturbation will initially expand with the Hubble flow while growing at a given rate determined by the energy content of the Universe (as in eq. 1.21). When $\delta \sim 1$, the perturbation reaches its maximum radius, i.e. the *turn-around* radius R_{ta}, detaches from the Hubble flow and begins collapsing. Through rigorous calculations it is possible to show that the density-contrast at the turn-around moment is given by:

$$\delta_{ta} = \frac{1.686}{D(z)} \tag{1.23}$$

with $D(z=0) = 1$ by convention.

If the spherical symmetry is not broken during the whole collapse, the structure compresses down to a single point. Nevertheless, realistic numerical computations show that volume elements of the collapsing structure might acquire angular momentum during the collapse, which ultimately provides a form of kinetic energy to the collapsing structure. This allows the system to reach *virial equilibrium* (see e.g. Collins, 1978; Davis et al., 2011), i.e. a configuration characterized by a specific energy balance between total kinetic energy K and total internal energy U, namely:

$$2K + U = 0 \tag{1.24}$$

Structures in virial equilibrium are characterized by a diffuse mass distribution and are generally referred to as *virialized halos* (or simply *halos*). In an EdS Universe, virialization is usually reached when the final over-density, relative to the critical density evaluated at the turn-around time, is given by:

$$\frac{\delta_{\mathrm{vir}} - \rho_{c\,;\,ta}}{\rho_{c\,;\,ta}} = \Delta_c \sim 18\pi^2 \sim 178 \tag{1.25}$$

[7] This approach is also known as *linear theory of structures growth*.

In a ΛCDM Universe, this value is modified through the fitting formula of Bryan & Norman (1998):

$$\Delta_c = 18\pi^2 + 82\,(\Omega_m^z - 1) - 39\,(\Omega_m^z - 1)^2 \tag{1.26}$$

where the quantity:

$$\Omega_m^z = \frac{\Omega_{m;0}(1+z_{ta})^3}{\Omega_{m;0}(1+z_{ta})^3 + \Omega_\Lambda + \Omega_{k;0}(1+z_{ta})^2} \tag{1.27}$$

is evaluated at the turn-around redshift z_{ta}. From these equations it is possible to derive fundamental physical properties of virialized DM halos, such as their *virial radius* $\mathrm{R_{vir}}$ and their *virial temperature* $\mathrm{T_{vir}}$. All these quantities depend on the *virial mass* $\mathrm{M_{vir}}$, namely the mass enclosed within a sphere of radius $\mathrm{R_{vir}}$ (see Barkana & Loeb, 2001):

$$\mathrm{R_{vir}} = 0.784 \left(\frac{\mathrm{M_{vir}}}{10^8\,\mathrm{M_\odot\,h^{-1}}}\right)^{1/3} \left(\frac{\Omega_m}{\Omega_m^z}\frac{\Delta_\mathrm{c}}{18\pi^2}\right)^{1/6} \left(\frac{1+z}{10}\right)^{-1} \mathrm{Kpc\,h^{-1}}\,, \tag{1.28}$$

$$\mathrm{T_{vir}} = 1.98\times10^4 \left(\frac{\mu}{0.6}\right) \left(\frac{\mathrm{M_{vir}}}{10^8\,\mathrm{M_\odot\,h^{-1}}}\right)^{2/3} \left(\frac{\Omega_{m;0}}{\Omega_\mathrm{m}^z}\frac{\Delta_\mathrm{c}}{18\pi^2}\right)^{1/3} \left(\frac{1+z}{10}\right) \mathrm{K}\,. \tag{1.29}$$

The large scale structure of the Universe

Equation 1.28 shows that, at fixed $\mathrm{M_{vir}}$ and cosmological parameters, $\mathrm{R_{vir}}$ diminishes with increasing z. This is an effect of the so-called *hierarchical growth* of DM structures. Indeed, within the ΛCDM scenario, smaller perturbations form earlier in time than massive ones, growing by the smooth accretion of matter and mergers with other small structures, throughout their evolution. Therefore, initially small and isolated density perturbations gradually join together in increasingly large structures, formed of massive halos and smaller satellites orbiting around them. This process eventually drives the formation of *groups*, *clusters* or massive *super-clusters* at the expense of under-dense regions of the universe (i.e. *cosmic voids*). Clusters and super-clusters are connected to each others through elongated formations known as *filaments*. These elements collecively form the sponge-like distribution of matter known as the *large-scale structure* (LSS) of the Universe.

This picture describes a bottom-up growth of structures, by which groups and clusters form late in time through the hierarchical merging of halos which originated from low-mass density perturbations at high-z. This is in open contrast with the observed *downsizing* of galaxies (e.g. Neistein et al., 2006; Fontanot et al., 2009; Baugh, 2013; Pacifici et al., 2016), according to which massive stellar systems such as elliptical galaxies assembled their stellar-mass earlier in time than low-mass, dwarf galaxies. Indeed, it may look reasonable to expect that the evolution of baryonic-matter would parallel the one of DM, given the strong gravitational prevalence of the latter over the former. We provide further details about this apparent issue in §1.1.2.

1.1.2 Primordial infall of gas and galaxy formation

As soon as DM halos are virialized, baryons are captured within their gravitational-potential wells and acquire both internal and kinetic energy, setteling into a rotationally-supported disk (e.g. Fall & Efstathiou, 1980). This process is usually referred-to as *primordial baryonic infall* and is provides the initial reservoir for early star formation (SF) processes within DM halos.

Differently to DM, baryonic matter responds to electromagnetic interactions, which ultimately produce *thermo-dynamic pressure* and contrast gravitational collapse. On the other hand, electromagnetic interactions allow to dissipate, in the form of radiative emission, part of the energy acquired during the primordial infall. The balance between *heating* due to gravitational contraction and *cooling* triggered by radiative processes in the collapsing structures, ultimately determines the fate of baryonic matter within virialized DM halos (e.g. White & Rees, 1978). In particular, the details of cooling depend on the physics and chemistry of pristine gas at the atomic level, i.e. on the *micro-physics* of gas. To keep the current discussion simple, here we address the overall stability of gas clouds and its effect on the formation of stars and galaxies. Nevertheless, some details about the micro-physics of primordial gas deeply impact the possible formation of the first SMBHs, hence we comment further on this topic in §1.2.1.

The Jeans stability criterion

The stability of a spherical mass distribution under the competing effects of gravity and hydrodynamic pressure has been addressed by Jeans (1902, 1928). The rationale of its analysis is similar to the one showed for the collapsing of DM structures, except for the facts that the baryonic cloud can develop hydrodynamical pressure and it is not subject to the Hubble flow, since it is contained in a virialized DM halo which already detached from it. Therefore, the cloud stability can be studied by applying a different form of the Euler equations, namely:

$$\frac{\partial \rho}{\partial t} + \nabla(\rho\,\mathbf{u}) = 0\,, \tag{1.30}$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla(\mathbf{u}) = -\frac{1}{\rho}\nabla(\mathrm{P}) - \nabla(\phi)\,, \tag{1.31}$$

$$\nabla^2\phi = 4\pi G\,\rho\,. \tag{1.32}$$

where ρ is the 3D density of the baryonic cloud, $\mathbf{u}$ is the velocity of a volume element and P is the gas pressure. The solution to this system of equations is a second-order, partial-derivatives, non-homogeneous differential equation which generally describes an oscillatory phenomenon, i.e. a density wave. This, in turn can exhibit two different behaviors, i.e. an *oscillatory* and an *exponentially-growing* mode. Indeed, it is reasonable to imagine that very small perturbations cannot exhibit enough "gravitational pull" in order to win the balancing effect of pressure and trigger their own gravitational collapse. Therefore, they will oscillate between gravitational collapse and pressure-driven expansion. On the other hand, big-enough perturbations will likely grow and collapse within some characteristic time interval.

This qualitative argument hints at the presence of a typical scale-length which separates the two behaviors. This is called *Jeans length* (λ_J) and, by means of the cloud density, it can be easily associated to a typical mass-scale, i.e. the *Jeans mass*, defined as the mass of a sphere with diameter $2R = \lambda_J$ (see e.g. Jeans, 1902; Stahler & Palla, 2004):

$$\lambda_J = \left(\frac{\pi c_s^2}{G\,\overline{\rho}}\right)^{1/2} \quad ; \quad M_J = \frac{4}{3}\,\pi\,\overline{\rho}\left(\frac{\lambda_J}{2}\right)^3 \approx 2.5\times 10^5 \left(\frac{aT}{\mu}\right)^{3/2} \Omega_m^{-1/2}\,, \tag{1.33}$$

where $c_s = \gamma \cdot k_B T/m_p\mu$ is the sound-speed within the cloud with adiabatic index γ, temperature T and molecular mass $m_p\mu$, while a is the cosmological scale factor (Madau & Kuhlen, 2003). Therefore, according to the local physical conditions within the baryonic cloud, clumps of $M > M_J$ might collapse within the DM halo. Inside these clumps, the gas can further cool via radiative processes, so that fragmentation into smaller clumps is favoured.

Basic picture of star formation

Under realistic conditions, the gravitational compression of gas clouds is hardly symmetric. Therefore, despite offering profound insights, the theoretical scenario delineated until this point has a limited application. Therefore, numerical approaches are usually pursued to study the process of stars and galaxies formation. Simulations show that realistic baryonic collapse at high-z generally leads to the formation of dense, rotationally-supported, disk-like structures in which the local physical conditions regulate the eventual fragmentation of gas into clumps (e.g. Mo et al., 1998).

This process is thought to proceed until clumps of very high densities are formed, for which radiative processes might become inefficient due to the gas opacity. Under these conditions, the excess energy cannot be radiated away from the gas cloud which therefore heats under the effect of gravitational compression. The endpoint of this scenario is reached when the central part of the cloud (i.e. its *core*) reaches the temperature and density needed for nuclear reactions to set-in. These complex, exothermic phenomena convert the primordial hydrogen into helium and heavier elements, providing the necessary source of internal energy to support pressure forces and balance the cloud self-gravity. This state is known as *hydrostatic equilibrium* and conventionally marks the transition from the early proto-star phase to the actual formation of a stable star (e.g. Omukai, 2001a; Omukai & Palla, 2001; Stahler & Palla, 2004; Bodenheimer, 2011).

The star-formation (SF) process described above is thought to happen ubiquitously within turbulent, disk-like baryonic structures inside DM halos, hence leading to the formation of *galaxies*, i.e. bound structures composed of stars and gas. Within the latter, stars are produced at a given *star-formation rate* (SFR) in time from the diffuse gas in the inter-stellar medium (ISM), and evolve until their nuclear cores of hydrogen are consumed. This marks the end of their stable evolution and their subsequent fate depends primarily on their total mass M_s (e.g. Prialnik, 2009). Indeed, isolated stars with $M_s \lesssim 8M_\odot$ are thought to end their life as white dwarfs with an associated planetary nebula, i.e. systems where the compact stellar core expels the outer stellar envelope in the ISM and remains "naked". On the contrary, massive stars with $M_s \gtrsim 10M_\odot$ have a more violent, explosive end which can eventually lead to the complete disruption of the evolved star (e.g. Shapiro & Teukolsky, 1986).

These latter scenarios are usually associated with extremely energetic and luminous phenomena called SuperNovae (SNe). SNe cover a fundamental role in the evolution of the whole Universe. Indeed, their immediate effect is to inject both energy and heavy elements in the ISM, through a combination of mechanical, chemical and radiative effects globally called SNe *feedback* (e.g. Leitherer et al., 1992; Leitherer, 1994; Dale, 2015). This process is believed to have profound impact on the evolution of galaxies by e.g. regulating their SFR and chemical enrichment (e.g. Scannapieco et al., 2002; Bertone et al., 2007; Wise & Abel, 2008; Hayward & Hopkins, 2017). In addition, SNe are believed to mark the formation of extremely compact remnants such as neutron stars (NS) and BHs (Prialnik, 2009), while SNe-Ia can be used to precisely measure very large distances, due to the link between their intrinsic emitted power and the time-evolution of their observed light (as discussed in §1.1.1).

Basic picture of galaxy evolution

Once galaxies are formed their gaseous and stellar content evolve through a series of dynamical, mechanical, thermodynamic and chemical processes. Examples of these baryonic processes are the continuous exchange of gas with the surrounding environment (either by accretion of diffuse gas or by the expulsion of chemical-enriched one), or the formation of stellar disks and dynamical instabilities within them (e.g. Longair, 2008). These processes involving a single galaxy and its diffuse surroundings are usually labelled as *secular* processes, in order to differentiate them from *external* processes, such as *mergers* (see below). Secular processes can be thought as slow and continuous evolutionary steps through which galactic systems build-up their stellar mass and structural properties (see Kormendy, 2013).

Galaxy mergers, on the other hand, are the end-point of gravitational interactions between different galactic systems which were formed in separate halos and later-on brought together by their mutual gravitational attraction. They generally represent fast episodes within the overall evolutionary history of galaxies, as opposed to secular processes. Mergers usually proceed through a series of subsequent dynamical interactions such as close-passages (fly-by) and tidal disturbances which can modify the structure of the interacting systems, trigger gas compression (and subsequent star formation) or even expel diffuse gas and isolated stars into the *intra-galactic medium* (IGM). Depending on the mass-ratio m_r of the interacting systems, galaxy mergers are generally divided into minor ($m_r \lesssim 0.3$) and major ($m_r \gtrsim 0.3$), with the latter class usually leading to a overall structural disruption of the interacting galaxies which coalesce into a single, elliptical (or triaxial) remnant (Kormendy, 2013).

The complex combination of secular and external processes shapes the properties of galaxies throughout their evolution, producing a plethora of morphological types and properties (see e.g. Kormendy, 2013). For historical reasons, this variate population is generally divided into two broad classes, namely: disk-like, star forming galaxies (i.e. *late-type*) and spheroid-like, quenched systems (i.e. *early-type*). The first ones are characterized by a circular, ordered motion with a well-defined rotation curve. This can be measured through the differential doppler shift of their emitted light, at different positions along the rotation plane. On the contrary, the second class does not show a preferential axis of rotation, but rather a thermal-like dispersion of star velocities. For each of these two morphological classes, a number of correlations between structural properties has been studied in the past, with the most famous examples being the Faber-Jackson (Faber & Jackson, 1976) and the Tully-Fisher relations (Tully & Fisher, 1977). These connect the dynamical properties of galaxies (respectively: the dispersion

velocity of early-types and the rotational velocity of late-types) to the intrinsic luminosity[8] of the galactic system.

The morphological distinction of galaxies broadly holds when considering the relationship between their stellar mass (M_*) and SFR over large populations (González Delgado et al., 2016). Indeed, late-types appear to show a tight linear correlation between these two properties, know as the SFR *main sequence* (e.g. Brinchmann et al., 2004; Noeske et al., 2007; Vilella-Rojo et al., 2021), while early-types cluster in a so-called *red sequence*. Nevertheless, it must be stressed that the picture according to which galaxies can be divided into two broad classes is likely too simple to capture all the diversity of realistic galaxy populations. As an example, it has been noted that the central, spheroidal *bulge* of some massive disk-like galaxies show many similarities with early-type systems (e.g. Kormendy, 2016). To conclude, massive early-type galaxies at low-z generally show low SFRs and old stellar populations (e.g. Kormendy, 2013). This led to the conclusions that elliptical galaxies might have formed before low-mass, actively star forming galaxies, in apparent contrast with the hierarchical growth of DM structures. This observation is generally called *downsizing* and it is currently understood within hierarchical models of structures formation as an effect of the different evolutionary paths followed by DM and their baryonic counterparts (see e.g. White & Frenk, 1991; Bower et al., 2006; Neistein et al., 2006; Fontanot et al., 2009).

1.1.3 Compact objects in galaxy evolution and cosmology

As already mentioned, massive stars with $M_s \gtrsim 8-10\,M\odot$ which terminate their life catastrophically produce extremely compact objects known as NS and BHs. These originate from the extreme compression of their stellar cores, which exhausted the fuel for all the possible exothermic nuclear reactions. Indeed, these central parts of massive stars suffer extreme gravitational stress, without being able to produce the required energy to balance the star self-gravity through pressure forces. Therefore, their gravitational collapse can proceed unimpeded and compress their matter up to extreme densities, i.e. in excess of: $\rho_{\,\mathrm{core}} \gtrsim 10^9 g\,cm^{-3}$ (e.g. Shapiro & Teukolsky, 1986; Prialnik, 2009). The only known outcomes of this condition are either the formation of a NS or a BH (e.g. Oppenheimer & Volkoff, 1939).

Compact objects are able to produce intense gravitational fields in their vicinity. This idea can be formalized by the concept of *escape velocity*[9] from an object with mass $M_\bullet$ and radius $R_\bullet$. According to Newtonian mechanics, this is given by:

$$v_{esc} = \sqrt{\frac{2GM_\bullet}{R_\bullet}} \tag{1.34}$$

where G is the gravitational constant. This qualitatively shows that massive objects compressed into small radii require high velocities in order to escape from their gravitational fields. Taking this argument to the limit, if an object shows such a high compactness in order to have $v_{esc} = c$, nothing can escape from it, not even light.

Such extreme mass configurations cannot be described by Newtonian laws of gravitation, but rather a GR treatment is necessary (e.g. Shapiro & Teukolsky, 1986). As an example, the distance at which $v_{esc} = c$ is usually called *Schwarzschild radius* (R_s), as it was derived by Schwarzschild (1916) through solving the Einstein's field equations around a spherical, non-rotating mass distribution in polar coordinates. Within the Schwarzschild metric, R_s defines a spherical surface which causally disconnects the volume within it from the outside Universe, since any information which should cross this surface outwards, must have a speed higher than c. This is referred-to as *event-horizon*, to underline that any space-time event inside this surface cannot be observed from outside it, being causally disconnected from observers beyond it (see Rindler, 1956; Finkelstein, 1958).

The existence of extremely compact astrophysical objects such as NS and BHs was generally limited to theoretical considerations, until the discovery of *Pulsating Radio Sources* signals (*pulsars*) by Jocelyn Bell (Hewish et al., 1968; Pilkington et al., 1968), the observation of the first quasi-stellar source (QSO, Schmidt, 1968) and the determination of the presence of a BH in the X-ray binary system Cygnus-X (Bolton, 1972; Webster & Murdin, 1972). In particular, the intrinsic power of the QSO source, source

[8] The Luminosity (L) of an astronomical source is defined as the intrinsic emitted power of that source, that is: the energy emitted per unit of time. In astronomical contexts, it is common to express quantities in `cgs` units, so that $[L] = [\mathrm{erg\ s^{-1}}]$.

[9] The escape velocity is defined as the speed which is necessary for a test particle to reach, with a final null velocity, an infinite distance from an astronomical object.

combined with its high redshift (for the epoch of its discovery) clearly pointed towards a non-stellar origin of its electromagnetic emission. This observational issue was overcome with the development of models for the mass-accretion onto massive BHs, which produced a theoretical basis to explain the high observed luminosity of QSOs. In particular, the viscous, thin-disk model of Shakura & Sunyaev (1976) allowed to explain the conversion of matter into energy with much higher efficiencies than the estimated values for nuclear reactions within stellar cores (see §1.3.3). This, in turn, provided a natural explanation for the observed QSO luminosity.

The inclusion of BHs and SMBHs as possible astronomical sources fostered the interpretation of apparently luminous AGN as powered by SMBHs in the process of accreting matter through an accretion disk (or, in general, an accretion *flow*). This speculation lead Soltan (1982) to formulate the conjecture according to which the observed constraints on the AGN luminosity function, together with the limits on the density of mass currently compacted into SMBHs, suggest that these latter objects acquired most of their mass through episodes of luminous and radiatively efficient mass-growth. We further comment the mass-growth of SMBHs and its associated radiative emission in §1.3.

Nowadays, the presence of SMBHs in the nuclei of galaxies is considered to be ubiquitous, at least at the massive end of the galaxy-mass distribution (see e.g. Miller et al., 2015). Furthermore, SMBHs are thought to co-evolve with their host galaxies by regulating their SFR with a series of AGN-feedback mechanisms (e.g. Boyle & Terlevich, 1998; Wyithe & Loeb, 2003; Croton et al., 2006; Fabian, 2012; Heckman & Best, 2014; Volonteri et al., 2015b; Mezcua et al., 2019). Although this complex phenomenon is still debated, several arguments support the scenario of galaxy-SMBHs co-evolution, such as: the observed correlations (or *scaling relations*) between the SMBH mass $M_\bullet$ and the total stellar mass (M_*) or bulge mass (M_{bulge}) of the host-galaxy (see e.g. Davis et al., 2018), as well as between $M_\bullet$ and the velocity-dispersion (σ) of stars within the host-galaxy bulge (e.g. Ferrarese & Merritt, 2000; Gebhardt et al., 2000; Heckman et al., 2004; Wyithe & Loeb, 2005; Wyithe, 2006; Gültekin et al., 2009b; Kormendy & Ho, 2013).

The presence of SMBHs can appear as a natural consequence of the production of compact objects by high-z stellar evolution, together with prolonged mass-growth on these objects. Indeed, according to this hypothetic scenario, it would be possible to fit within the Universe history the growth of initial $M_\bullet \sim 1-10 M_\odot$ into massive $M_\bullet \gtrsim 10^7 M_\odot$ objects within few Gyr (assuming an episodic Eddington-limited growth with *radiative efficiency* $\epsilon_r = 0.1$, see §1.3.2 for details). Nevertheless, the recent observation of extremely luminous QSOs (possibly associated to very massive BHs) when the Universe was less than 1 Gyr old severely challenges this picture (e.g. Latif & Ferrara, 2016; Valiante et al., 2017). Since this issue represents one of the central topics of the current work, we discuss it in more detail in §1.2.

1.2 The formation of Super-Massive Black Holes

The observation of extremely luminous QSOs at $z \gtrsim 6$ (e.g. Mortlock et al., 2011; Venemans et al., 2013; Wu et al., 2015) represents one of the most fascinating open questions in modern astrophysics, as it directly challenges our current understanding of the formation of super-massive black-holes (SMBHs, e.g. Latif & Ferrara, 2016; Valiante et al., 2017; Inayoshi et al., 2020, for recent reviews). Indeed, powerful QSOs are likely powered by the accretion of matter onto $M_\bullet \gtrsim 10^9 M_\odot$ SMBHs, which must have assembled in less than $\lesssim 10^9$ yr, as suggested by their extreme redshift (e.g. Mazzucchelli et al., 2017). In the last deacades, the interest of the scientific community towards this issue has been increasing, fuelled by the ongoing discovery of extremely luminous and high-z QSOs (see Bañados et al., 2016, 2018; Fan et al., 2019; Wang et al., 2021). Although several possible pathways have been proposed, current theoretical models still struggle to identify the initial *seeds* of these massive compact objects and to fit their mass-growth within such short time-scales from the Big Bang.

Given the theoretical and observational challenges it poses, the issue of early SMBH formation and growth has gained the status of "problem" (i.e. the *BH-seeding problem*). Indeed, the evolution of the first generation of metal-free (PopIII) stars is thought to produce relatively common seeds of $M_{seed} \sim 10^2 M_\odot$ at $z \gtrsim 20$ (e.g. Schneider et al., 2002; Yoshida et al., 2003; Bromm & Larson, 2004; Yoshida et al., 2007), nevertheless these would require stringent conditions on their growth-rates to promptly evolve into SMBHs (e.g. Pacucci et al., 2015; Pezzulli et al., 2017; Regan et al., 2019). On the other side,

models which assume a more massive origin of SMBHs avoid strict growth constraints but generally require peculiar physical conditions for the formation of BH seeds. Consequently, their occurrence is yet poorly constrained over cosmological contexts, hence it is not clear whether their predicted abundance can explain the observed number density of high-z QSOs (e.g. Bromm & Loeb, 2003; Omukai et al., 2008; Latif et al., 2015; Agarwal et al., 2016; Latif et al., 2018). Finally, in order to relax the time-constraints introduced by baryonic physics, exotic scenarios envisioning the formation of *primordial* BH-seeds (PBHs) during the first, inflationary phases of the Universe expansion have been recently proposed (e.g. Bean & Magueijo, 2002; Bernal et al., 2018; Inomata et al., 2021). In this work we focus on the baryonic channels that might lead to the formation of SMBHs, hence we do not discuss the formation of PBHs.

On top of this already complex issue, it is not clear whether the first phases of BH-seed growth at extreme redshifts ($z \gtrsim 10$) can be directly observed in order to shed light on the origin of SMBHs. Indeed, the first phases of BH formation and growth might easily be obscured by the gaseous medium in which they take place (see e.g. Davies et al., 2019). Indirect, statistical constraints about high-z BH seeding processes might alternatively come from low-z observations. Among these, GW-detections (e.g. Sesana et al., 2011) and the eventual observation of low-mass SMBHs in dwarf galaxies (e.g. Mezcua, 2021) might help to constrain the overall efficiency of light, intermediate or massive seeding scenarios. This section is devoted to introducing the physical conditions which are expected to foster high-z BH formation processes (§1.2.1) and to detail the formation scenarios of SMBH seeds based on baryonic physics (§1.2.2).

1.2.1 Physical conditions for high-z BH formation

Generally speaking, the occurrence of different scenarios critically depends on the small-scale physical conditions of high-z, primordial gas clouds which are expected to nurture the formation of SMBH seeds. More in detail, gas cooling and fragmentation, star-formation processes and stellar feedback concur to discriminate among different BH seeding channels. These phenomena are thought to happen in the first virialized halos at high-z, where primordial gas infall has already settled into DM potential wells, as delineated in §1.1.1. This neutral gas exhibits pristine chemical composition, namely: $\sim$75% H, $\sim$24% $\mathrm{H_e}$ and trace amounts of heavier elements such as Lithium or Beryllium (Ryden, 2016; Planck Collaboration et al., 2020). Starting from these conditions at $z \gtrsim 30$, the formation of SMBHs by $z \gtrsim 7$ requires that: i) a given amount of matter is compressed into a compact object on short time-scales, and ii) a greater amount of mass is accreted onto this compact object. All the theoretical solutions we discuss here to the first part of this issue are based on the micro-physics of primordial gas, independently of the final mass they envision for the compact object.

Cooling of primordial gas and PopIII star formation

The process by which a distribution of gas loses its internal energy by radiative emission is refered-to as *gas cooling*, and it is generally followed by the fragmentation of gas into *clumps* of different masses. At high-z, cooling and fragmentation of pristine gas are thought to regulate the formation of the first generations of metal-free stars (i.e. Population III, or PopIII stars, e.g. Schaerer, 2002b; Schneider & Omukai, 2010; Schneider et al., 2012; Latif & Schleicher, 2015). These processes critically depends on the gas temperature, density, *metallicity* (Z, i.e. the ratio of concentration between *metals*[10] and hydrogen), as well as on its dust content (e.g. Bromm et al., 2001; Schneider et al., 2002; Santoro & Shull, 2006; Clark et al., 2008; Smith et al., 2008). Indeed, due to their atomic structure, metals can generally lose their excess kinetic energy through the emission of photons by electrons in excited states, during collisional interactions (e.g. Rybicki & Lightman, 1986). This process is generally called *metal*-cooling and it has a low efficiency at gas temperatures below $\mathrm{T_{gas}} \lesssim 10^4$ K (e.g. Stahler & Palla, 2004; Bodenheimer, 2011). Nevertheless, as soon as the gas temperature allows it, metal cooling is extremely efficient in lowering the temperature of gas, reducing its Jeans mass and hence in diminishing the typical mass-scale on which fragments can form (see the discussion in §1.1.2). Consequently, gas clouds are induced to fragment into low-mass clumps which can furthermore contract and form stars (e.g. Greif et al., 2012). Several works have pointed out that sufficiently-high metallicity increases the efficiency of metal cooling and induces

[10] Conventionally, in most astronomical and astrophysical contexts, all elements heavier than helium are collectively referred-to as "metals".

vigorous gas fragmentation (e.g. Bromm et al., 2001; Smith et al., 2008, 2009a). Generally, this limit is set to $Z \sim 10^{-4} - 10^{-3}\,Z_{\odot}$ (where $Z_{\odot} = 0.0134$ is the solar metallicity Asplund et al., 2009), and it is used to distinguish between the regimes of metal-free and metal-enriched star formation (e.g. Maio et al., 2010; Valiante et al., 2016). Nevertheless, significant differences are found in the value for this critical meallicity when considering realistic models which include cooling by dust and fine-structure metal lines or heating by an external radiation fields (e.g. Schneider et al., 2002, 2006; Santoro & Shull, 2006; Tsuribe & Omukai, 2008).

On the other hand, in the absence of metals, the cooling of gas can proceed via i) the interaction between free-electrons and H atoms, ii) the recombination of free protons and electrons (see e.g. Dijkstra, 2017) or iii) the roto-vibrational transitions of H_2 molecules (e.g. Abel et al., 2000; Bromm, 2013). More in detail, the first two processes efficiently subtract internal energy from the gas through radiative emission associated to atomic-hydrogen transitions. The production of this *cooling* radiation depends on the gas density and temperature, being significantly efficient only at $T_{gas} \gtrsim 10^4\,K$ (e.g. Abel & Haiman, 2000; Oh & Haiman, 2002). Consequently, pristine (i.e. $Z \approx 0$) high-z DM halos with $T_{vir} \gtrsim 10^4 K$ (see Eq. 1.29) are called *atomic cooling* halos.

At $T_{gas} \lesssim 10^4\,K$ and pristine conditions, gas cooling is driven by H_2 molecular transitions (e.g. Abel et al., 2000; Abel & Haiman, 2000; Greif & Bromm, 2006; Greif et al., 2008; Bromm, 2013). For this reason, the H_2 content of pristine gas clouds is considered crucial for the formation of PopIII stars (e.g. Omukai, 2000; Omukai & Palla, 2001; Bromm & Larson, 2004; Tsuribe & Omukai, 2008; Latif & Schleicher, 2015). Equation 1.29 shows that, at $z \gtrsim 15$, the temperature conditions for efficient H_2 cooling are reached by DM halos with $10^5 \lesssim M_{vir}/M_{\odot} \lesssim 10^7$. Therefore at these early epochs PopIII star-formation is thought to happen in low-mass H_2-cooling *mini-halos* (Abel et al., 2000; Abel & Haiman, 2000; Schaerer, 2002a; Schneider, 2006; Greif et al., 2011, 2013; Hartwig et al., 2015a; Haemmerlé et al., 2020). At lower redshift, PopIII SF is hindered by the intervening metal-pollution of the IGM due to the explosions of the first SNe, as discussed above. This ultimately leads to a rapid damping of PopIII SFR in the Universe (e.g. Smith et al., 2009b; Maio et al., 2010; Wise et al., 2012).

Delaying PopIII stars formation - Keeping primordial gas H_2-free

In order to produce seeds more massive than the typical mass of PopIII remnants, it is crucial to prevent the cooling and fragmentation of gas within high-z mini-halos. Indeed, if their fragmentation into small clumps is hindered, high-z pristine clouds can reach relatively high masses and densities before the virial temperature of their DM hosts reaches the limit of $T \gtrsim 10^4\,K$ and triggers efficient atomic cooling (e.g. Abel & Haiman, 2000; Oh & Haiman, 2002).

As discussed above, the gas fragmentation into small clumps can be efficiently contrasted by raising T_{gas}. For the case of pristine mini-halos collapsing at high-z, this is equivalent to diminish the efficiency of H_2 cooling. For this reason, several mechanisms have been proposed in the literature in order to delay or prevent gas fragmentation by keeping gas temperature high or by reducing its H_2 content. These theoretical scenarios invoke either dynamical heating (e.g. Yoshida et al., 2003; Fernandez et al., 2014; Inayoshi et al., 2015; Chon et al., 2016; Inayoshi et al., 2018), baryon streaming velocities (e.g. Tanaka et al., 2013; Latif et al., 2014b; Tanaka & Li, 2014; Popa et al., 2016) or the presence of radiation fields, able to photo-dissociate H_2 molecules (Omukai, 2001b; Wise & Abel, 2007; Dijkstra et al., 2008; Omukai et al., 2008; Latif et al., 2011; Wolcott-Green & Haiman, 2011; Agarwal et al., 2012; Latif et al., 2014a; Regan & Downes, 2018; Visbal et al., 2020). In particular, the last solution is one of the most explored in the literature, since photo-dissociating fluxes can be explained as the effect of radiative feedback from young stars (e.g. Haardt & Madau, 1996, 2012; Wise et al., 2014; Barrow et al., 2017; Kim et al., 2017b), forming in the close vicinity of the pristine cloud. This led several works to conclude that the conditions for the photo-dissociation of H_2 can be preferentially found in *synchronized halo pairs*, where the radiative feedback of one, luminous neighbor sterilizes a collapsing cloud against the presence of H_2 (e.g. Dijkstra et al., 2008; Visbal et al., 2014b; Agarwal et al., 2017; Regan et al., 2017). Within this theoretical framework, the presence of Lyman-Werner (LW) radiation fields is usually required, being the LW band a specific interval of UV frequencies responsible for the dissociation of H_2 molecules (i.e. $h\nu = [11.2 - 13.6]$ eV). In high-z studies, LW flux (i.e. J_{LW}) is generally measured in units of

$J_{21} = 10^{-21}\,\mathrm{erg\,s^{-1}\,cm^{-2}\,Hz^{-1}\,sr^{-1}}$.

The local values of J_{LW} necessary for the full dissociation of H_2 have recently been the object of intense debate. A common approach is to consider as completely sterilized against H_2 all the structures illuminated by a local $J_{LW} > J_{crit}$, where J_{crit} is a given critical threshold whose value can be derived from a variety of different assumptions. For instance, values of the order of $J_{crit} \sim 10^3\,J_{21}$ are found by requiring that the H_2 content is completely eliminated from $M_{cloud} \sim 10^5 M_\odot$ gas clouds (e.g. Omukai et al., 2008; Shang et al., 2010; Latif et al., 2014a). Nevertheless, orders-of-magnitude differences with respect to $J_{crit} \sim 10^3\,J_{21}$ can be found when considering the self-shielding of H_2 clouds against UV radiation (e.g. Wolcott-Green et al., 2011; Hartwig et al., 2015b; Wolcott-Green & Haiman, 2019), Lyα trapping (Wolcott-Green et al., 2021) or the effect of turbulence and rotation (e.g. Latif et al., 2013; Latif & Volonteri, 2015). Currently there is no consensus about the J_{crit} values; rather, recent works have enriched the discussion by suggesting to abandon the simplified requirement of a single J_{crit} value (Wolcott-Green et al., 2017). This is motivated by the complex dependencies of the local J_{LW} onto the spectral properties of the emitting source, its SFR and its distance (e.g. Agarwal et al., 2019).

Overall, the efficiency of heating mechanisms is correlated to the typical mass of the gas clumps which can form. In limiting cases, the fragmentation is inhibited for the whole gas content of a collapsing DM halo, so that eventually massive, compact objects can form as soon as cooling sets in. Nevertheless, it is reasonable to envision intermediate cases in which clumps form only in peculiar regions of collapsing gas clouds, where the conditions for gas fragmentation are satisfied locally (e.g. Schneider et al., 2006). In particular, past works have studied the case in which mildly-enriched gas clouds contract and cool until a critical density is reached, above which vigorous fragmentation is triggered within a dense, compact region of the collapsing cloud (e.g. Clark et al., 2008; O'Shea & Norman, 2008; Devecchi & Volonteri, 2009). Finally, the overall picture discussed in this section is complicated by the presence of dust or trace amounts of metals, which can significantly modify the behavior of low-metallicity gas and trigger its cooling even at $Z < 10^{-4}\,Z_\odot$ (e.g. Schneider et al., 2006; Omukai et al., 2008; Schneider et al., 2012). Furthermore, additional processes such as the self-shielding of H_2-enriched regions of gas clouds within mini-halos further complicate the evolution of high-z collapsing clouds (e.g. Wolcott-Green et al., 2011; Hartwig et al., 2015b; Wolcott-Green & Haiman, 2019).

To summarize, the complex interaction between all the above-mentioned physical properties of high-z gas ultimately determines the fate of the collapsing gas and its ability to produce BH-seeds of different masses (e.g. §1.2.2 and Valiante et al., 2017, for a recent review). As discussed, the leading factors in this analysis are the gas temperature, its metallicity and eventual heating mechanisms, such as the presence of an external LW flux (see e.g. Inayoshi et al., 2020). Therefore, at least under a physically-motivated, simplified approach, the formation of the first stars and BHs of different masses, within either mini-halos or atomic-cooling halos at high-z, can be described in terms of local metallicity and LW flux (e.g. Valiante et al., 2017; Maio et al., 2019; Sassano et al., 2021; Lupi et al., 2021b).

1.2.2 The different channels for SMBH seeds formation

Several possible pathways have been proposed to solve the SMBH-seeding problem. Most of these theoretical scenarios rely on baryonic processes which are thought to happen in low-metallicity, $z \gtrsim 8$ gas clouds. These models are usually divided into three broad classes, depending on the final seed mass (M_{seed}) they predict. In order of increasing M_{seed}, they are labelled as *light* ($M_{seed} < 10^3\,M_\odot$), *intermediate* ($10^3 \lesssim M_{seed}/M_\odot \lesssim 10^4$) and *heavy* ($M_{seed} \gtrsim 10^5\,M_\odot$) scenarios; we separately address each of them in the following sections.

We underline that recent theoretical research has also focused on the evenutal non-baryonic origin of SMBHs in order to overcome the caveats of either light, intermediate or heavy seeding scenarios. In particular, the main circulating idea is that primordial BHs (PBHs) of small mass ($1 \lesssim M_{seed}/M_\odot \lesssim 10^2$, e.g. De Luca et al., 2020; Bernal et al., 2018) might act as SMBH seeds. PBHs are thought to be formed *before* the nucleosynthesis, i.e. during the inflationary phases of the Universe's history (see e.g. Bean & Magueijo, 2002; Bernal et al., 2018; Inomata et al., 2021; Liu et al., 2021). Since the BH-formation model discussed in this thesis (see §3) is based on a semi-analytic code which tracks the evolution of baryonic matter within condensed DM halos (see §2.1), we only concentrate on processes which follow in time the

primordial nucleosynthesis. Therefore, we do not treat the possible origin of SMBHs from PBHs.

Light-seeding scenarios

As discussed in §1.2.1, the formation and evolution of PopIII stars is thought to happen in low-mass H_2-cooling mini-halos. Physical conditions within these structures ultimately determine the typical mass of PopIII stars and of their compact remnants (e.g. Madau & Rees, 2001; Schaerer, 2002b; Schneider et al., 2006; Omukai et al., 2008; Haemmerlé et al., 2018). This formation scenario easily fits in the paradigm of structures formation, as it emerges naturally from the expected evolution of PopIII stars at high-z. The latter, indeed, are expected to leave behind compact BH seeds within specific mass ranges, namely $M_{seed} \sim 40 - 140\,M_\odot$ and $> 260\,M_\odot$ (e.g. Madau & Rees, 2001; Volonteri et al., 2003; Volonteri, 2010; Valiante et al., 2016; Haemmerlé et al., 2020).

Due to their relatively-low starting mass right after their formation, light seeds should continuously grow at extreme rates in order to reach $\sim 10^{8-9} M_\odot$ by $z \sim 7$ (i.e. even beyond the Eddington limit introduced in §1.3.3, see e.g. Volonteri & Rees, 2005; Natarajan, 2011; Pezzulli et al., 2017; Inayoshi et al., 2020; Haemmerlé et al., 2021). We note that this growth constrain does not exclude that SMBHs at more moderate redshifts (i.e. $z \lesssim 4$) might have indeed a light origin, due to the significantly larger time available for their growth. Therefore, PopIII remnants might offer a relatively easy solution to the formation of seeds for the global population of SMBHs observed at moderately-low redshifts but it is still unclear if this formation scenario can explain the presence of $M_\bullet \sim 10^{8-9} M_\odot$ objects at $z \gtrsim 8$.

Finally, as discussed in §1.2.1, in order to model the formation of PopIII stars and of their compact remnants it is necessary to follow the cooling and fragmentation of pristine gas clouds within mini-halos. This directly translates into a stringent requirement on the mass-resolution needed to perform self-consistent numerical simulations of light-seeds formation (e.g. Regan et al., 2015; Smith et al., 2018).

Intermediate-mass scenarios

In order to relax the critical time constraints on the mass-growth of light seeds, several works considered the case in which SMBH seeds might form with intermediate masses of $M_{seed} \sim 10^{3-4} M_\odot$ (e.g. Omukai et al., 2008; Devecchi & Volonteri, 2009; Das et al., 2021). As introduced in §1.2.1, these models are based on specific conditions which delay the onset of ordinary, high-z SF mechanisms until a point in which gas cooling and fragmentation is effective only in the densest regions of a collapsing gas cloud (e.g. O'Shea & Norman, 2008; Clark et al., 2008; Devecchi & Volonteri, 2009). This generally leads to the rapid formation of intermediate-mass BHs (IMBHs) within these dense regions, which would host favourable conditions for their rapid mass-growth (e.g. Portegies Zwart & McMillan, 2002; Rasio et al., 2004; Lupi et al., 2014; Sassano et al., 2021). The appeal of intermediate-mass scenarios is to envision the ability to reach $\gtrsim 10^7\,M_\odot$ on a short time-scale without requiring extremely peculiar physical conditions for the formation environment of BH seeds (see e.g. Regan & Haehnelt, 2009; Sakurai et al., 2017; Reinoso et al., 2018).

According to current theoretical developments, IMBH seeds might form after the runaway dynamical interactions of stars within a dense, nuclear stellar cluster (e.g. Ebisuzaki et al., 2001; Portegies Zwart & McMillan, 2002; Rasio et al., 2004; Lupi et al., 2014; Katz et al., 2015). These stellar clusters are thought to form as the result of intense episodes of early SF in the nuclear regions of halos where H_2-driven gas fragmentation is delayed. Usually, intermediate-seeding scenarios require either mild illumination from J_{LW} (e.g. Schneider et al., 2006; Omukai et al., 2008) or constraints on the strength of primordial infall which can be sustained by the collapsing gas disk without fragmenting (see the discussion in Devecchi & Volonteri, 2009). Once these conditions are met, dense nuclear stellar clusters are believed to form and evolve through dynamical interactions between stars. By assuming that these can lead to the instability of the stellar cluster towards runaway stellar mergers (RSM scenario hereafter, see Portegies Zwart & McMillan, 2002; Rasio et al., 2004), a very massive star (VMS) might form at the center of the structure and later-on evolve into a single, compact IMBH seed of $M_{seed} \sim 10^{3-4}\,M_\odot$ (see e.g. Devecchi & Volonteri, 2009; Mapelli, 2016; Stone et al., 2017; Woods et al., 2020; Das et al., 2021). Finally, this object is thought to efficiently grow thanks to abundant fueling provided by its dense, surrounding environment (e.g. Lupi et al., 2014).

Heavy-seeding scenarios

The formation of heavy seeds with $M_{seed} > 10^{4-5} M_\odot$ already at $z \gtrsim 10$ has been recently explored as an easier alternative to reach SMBHs of $\sim 10^{8-9} M_\odot$ by $z \sim 7$ (e.g. Bromm & Loeb, 2003; Volonteri & Bellovary, 2012; Latif et al., 2018; Mayer & Bonoli, 2019). Several possible channels might lead to these extremely massive objects, based on different physical processes (e.g. Bromm & Loeb, 2003; Mayer et al., 2010; Latif et al., 2014c; Mayer et al., 2015; Kroupa et al., 2020; Escala, 2021).

A great amount of theoretical effort has been undertaken to explore pathways in which the cooling of massive gas clouds is strongly delayed. Indeed, if gas cooling can be hindered until atomic-cooling sets in (at $T_{vir} \gtrsim 10^4$ K and $M_{vir} \sim 10^{7-8} M_\odot$, see Eq. 1.29 and §1.2.1), massive gas clouds contained in atomic-cooling halos might directly collapse into a $\sim 10^5 M_\odot$ BH through gravitational and dynamical instabilities (e.g. Becerra et al., 2018). As discussed above (§1.2.1), in order to reach the atomic-cooling limit without previously cooling, the gas must be pristine, i.e. without any metals nor dust, and it must not be subject to H_2-cooling. Pristine chemical composition can be ensured by the absence of internal SF processes in the collapsing cloud and by requiring that eventual chemically-enriched winds produced in its vicinity by star-forming halos do not pollute its gas content (e.g. Agarwal et al., 2017). On the other hand, the suppression of its H_2 content can be attained with any of the heating mechanisms discussed in §1.2.1.

Provided that these peculiar conditions are met (e.g. Shang et al., 2010; Visbal et al., 2014a), current models envision the formation of a very massive *direct-collapse* BH (DCBH) with initial mass $M_{seed} \sim 10^5 M_\odot$ (e.g. Latif et al., 2013; Schleicher et al., 2013). Provided that further gas mass can be efficiently funneled towards the newly-formed DCBH, these can grow up to $M_\bullet \sim 10^{8-9} M_\odot$ without requiring stringent conditions on their mass-growth rate (e.g. Volonteri et al., 2008; Latif et al., 2018; Regan et al., 2019, but see also the discussion in Chon & Omukai (2020) about the inefficient growth of DCBHs).

Further scenarios leading to heavy-seeds formation have been suggested as alternatives to the above possibilities. These models usually envision the formation of extremely-massive BH seeds (up to $M_{seed} \sim 10^{8-9} M_\odot$) through gas-rich galaxy mergers (e.g. Mayer et al., 2010, 2015) or the collapse of extreme stellar clusters (e.g. Kroupa et al., 2020; Escala, 2021). The model of Mayer et al. (2010), in particular, envisions the formation of *merger-induced* DCBHs (i.e. miDCBH hereafter) due to vigorous multi-scale inflows of gas towards the center of a galaxy-merger remnant. Indeed, numerical simulations have shown that the result of gravitational interactions between massive, gas-rich galaxies at $z \gtrsim 4$ can harbour the conditions to produce a strong concentration of gas in their centers on very short time-scales. Despite being compressed below the spatial-resolution scale of the simulations, this mass concentration is thought to evolve towards the formation of a very-massive star (e.g. Schleicher et al., 2013; Haemmerlé et al., 2019; Regan et al., 2020), which would then directly collapse into a massive BH. Furthermore, the intense, multi-scale inflow generated during the merger, would then feed this compact objects at sustained rates, hence rapidly leading it to the required $M_\bullet \gtrsim 10^8 M_\odot$ at high-z (e.g. Mayer et al., 2015; Mayer & Bonoli, 2019; Haemmerlé et al., 2021).

To summarize, heavy-seeding scenarios offer possible solutions to the issue of rapid mass-growth of the first BHs, allowing to not impose stringent limits on this process or to avoid resorting to super-eddington accretion (as in e.g. Wyithe & Loeb, 2012; Volonteri et al., 2015c; Inayoshi et al., 2016; Pezzulli et al., 2017; Sugimura et al., 2017; Regan et al., 2019). On the other hand, these scenarios envision the production of rather exotic objects, leading to extremely-massive compact remnants (e.g. Becerra et al., 2018). Indeed, heavy-seeding channels usually require very peculiar conditions whose actual occurrence in the high-z Universe is highly uncertain, especially within wide cosmological contexts (e.g. Devecchi et al., 2012; Agarwal et al., 2014; Bonoli et al., 2014; Habouzit et al., 2016; Di Matteo et al., 2017; Lupi et al., 2021a).

1.3 The mass growth of Super-Massive Black Holes

As discussed in the previous sections, observations across wide resdhift intervals and wavelength ranges strongly hint towards the existence of SMBHs at the center of all massive galaxies. Most of this evidence has been accumulated thanks to the presence of luminous AGN throughout the last decades. This section summarizes the current models for the accretion of matter onto SMBHs and the AGN radiative emission.

1.3.1 Feeding SMBHs: inflows and outflows of gas regulate SMBH growth

The mass-accretion onto SMBHs at the center of galaxies requires that a sufficient amount of matter is able to reach the immediate vicinity of these massive, compact objects. The fuelling of gas towards the latter is not straightforward, especially if considering the huge difference between galactic-scales ($\sim 1-10$ Kpc) and the typical, sub-parsec dimension of accretion disks (~ 0.1 pc, see e.g. Alexander & Hickox, 2012; Abramowicz & Fragile, 2013). In particular, gas fuelling must be driven by processes which are able to efficiently modify the angular momentum of matter on such wide spatial scales. In the past, significant effort has been put forward in order to explain the fuelling of AGN activity in galaxies. The most common solutions rely on gravitational torques produced in galactic interactions Capelo et al. (2015); Anglés-Alcázar et al. (2017b); Capelo & Dotti (2017); Steinborn et al. (2018), shocks or disk instabilities (e.g. Fanali et al., 2015; Kim & Elmegreen, 2017) and galactic stellar bars (e.g. Shlosman et al., 1989; Hopkins & Quataert, 2010; Anglés-Alcázar et al., 2021; Li et al., 2017; Jung et al., 2018), as also recently supported by observations (e.g. Ellison et al., 2011; Mezcua et al., 2015; Audibert et al., 2019). In addition to these multi-scale processes, significant AGN fuelling can be fostered by the action of strong stellar winds in the vicinity of the central SMBH (see e.g. Ressler et al., 2018; Palouš et al., 2020, but see also Anglés-Alcázar et al. (2017a) for an opposite effect of stellar feedback).

As discussed, accretion of matter onto SMBHs also produces intense energetic output from the active regions, either in the form of strong radiation fields or momentum-carrying winds and jets (e.g. Ghisellini, 2013; Mezcua et al., 2019; Sala et al., 2021). This complex *feedback* from the central accreting BH can have strong effects onto both the flow of gas towards the central SMBH and the overall properties of the host galaxy. Indeed, AGN feedback is believed to be one of the main phenomena by which galaxies and SMBHs co-evolve (see e.g. Silk & Rees, 1998; Ferrarese & Merritt, 2000; Di Matteo et al., 2005; Bower et al., 2006). Due to its crucial importance into shaping the properties of galaxies, AGN feedback have been largely studied both observationally and theoretically (see e.g. Bower et al., 2006; Croton et al., 2006; Merloni & Heinz, 2008; Fabian, 2012; Baron et al., 2018; Mezcua et al., 2019). These analysis helped to determine how AGN feedback shapes in time the scaling relations the central SMBHs mass and the stellar properties of their hosts (see §1.1.3 and e.g. Silk & Rees, 1998; Di Matteo et al., 2005; Wyithe & Loeb, 2005; Marulli et al., 2008; Gültekin et al., 2009a; Shankar, 2013).

1.3.2 Accretion of matter onto Black Holes

Once they are formed, SMBHs exert intense gravitational attraction on their surroundings, eventually driving matter in the form of stars and gas to fall within their event horizon (§1.1.3). This mass-accretion process is believed to be the most efficient ways to convert matter into energy (Shapiro & Teukolsky, 1986). Several models have been proposed to describe the infall of matter in the gravitational potential produced by BHs. Currently, two of these models are mainly used in numerical simulations which track BH mass-accretion, namely: the Bondi-Hoyle-Lyttleton model (Hoyle & Lyttleton, 1941; Lyttleton & Hoyle, 1940; Bondi, 1952) and the Shakura-Sunyaev model (Shakura & Sunyaev, 1976). The first one describes BHs mass-growth within idealized, spherically symmetric conditions, assuming a relative motion with velocity v between the BH and the surrounding matter distribution, as discussed at the end of the current section. Given its relative simplicity, it can be used to obtain an estimate for the accretion-rate onto BHs in numerical models accounting for the mass growth of BHs (e.g. Valiante et al., 2011; Bonoli et al., 2016). On the other hand, the Shakura-Sunyaev model presents a more complex treatment of BH mass-growth, where the infalling matter is organized in a rotationally-supported disk. Due to its structural and physical properties, this physical model is able to provide an explanation for the radiative emission of matter infalling into the gravitational field of a compact object, as detailed below.

The Shakura-Sunyaev accretion model

In most realistic cases, matter infalling into a gravitational potential acquires angular momentum, due to non-radial components of its motion. This might prevent the radial infall of matter, since without a mechanism to lose angular momentum, the matter can organize into a rotationally-supported disk and orbit around the central object. The Shakura-Sunyaev model answers to this issue by envisioning a model

in which the *viscosity* of the disk material is responsible for the transport of angular momentum between different annuli of the rotationally-supported disk. This transport ultimately induces an overall flow of matter towards the central BH, which results in the final infall of mass from the *accretion disk*.

Most importantly, the Shakura-Sunyaev model allows to predict the release of energy from the accretion disk, hence providing a theoretical justification to explain the extreme luminosity inferred for AGN sources (e.g. Jung et al., 2018; Inayoshi et al., 2019). The basis of this solution is to assume that the infalling matter settles into a *geometrically thin* and *optically thick* configuration which is able to transport outwards the angular momentum of matter, inducing its infall towards the central BH. According to this model, the internal energy accumulated by the disk matter during the flow towards the central object is mainly released through radiative processes by the innermost regions of the accretion disk (see Frank et al., 2002, for a detailed treatment).

To obtain the mass-accretion rate associated to the thin-disk of the Shakura-Sunyaev model, we start from the Euler equations in polar coordinates for a fluid element of density ρ, velocity $\mathbf{v}$ and pressure P:

$$\frac{\partial \rho}{\partial t} + \nabla(\mathbf{v}\,\rho) = 0\,, \tag{1.35}$$

$$\frac{\partial v}{\partial t} + \mathbf{v}\cdot\nabla(\mathbf{v}) = -\frac{\nabla(P)}{\rho} - \nabla(\phi_\bullet) + \frac{\nabla(\sigma)}{\rho}\,, \tag{1.36}$$

$$\nabla\phi_\bullet = -\frac{G\mathrm{M}_\bullet}{r^2}\,, \tag{1.37}$$

where $\phi_\bullet$ is the gravitational field produced by the central BH, while the term $\nabla(\sigma)/\rho$ in equation 1.36 accounts for the disk viscosity through the viscous-strain tensor $\sigma = \rho\alpha\nabla\mathbf{v}$, being α the scalar viscosity of the disk. This term is ultimately responsible for the outwards angular-momentum transport in the disk (e.g. Shakura & Sunyaev, 1976; Frank et al., 2002). By solving this system of equations it is possible to obtain the accretion rate for a Shakura-Sunyaev thin-disk:

$$\dot{\mathrm{M}}_{\mathrm{disk}} = \frac{3\pi\alpha\Sigma}{\left(1-\sqrt{\frac{\mathrm{R_{ISCO}}}{\mathrm{R}}}\right)} \tag{1.38}$$

where the quantity Σ is the surface-density of the accretion disk, R is the 2D radial distance from the BH and finally $\mathrm{R_{ISCO}}$ is the radius of the *innermost stable circular orbit* around the central BHs and it marks the inner boundary of the disk.

Finally, it is possible that the energy accumulated by the infalling material cannot be efficiently dissipated by the innermost part of the disk. In this case, the inner structure of the accretion flow might "inflate" and develop a vertical structure due to which the thin-disk approximation is no longer valid. When this happens, the disk model must be modified to account for a different type of *advection dominated* accretion flow (ADAF Rees et al., 1982).

The Bondi-Hoyle-Lyttleton accretion rate

The model presented by Bondi-Hoyle-Lyttleton describes the steady, spherically-symmetric inflow of matter onto a BH of mass $\mathrm{M}_\bullet$, moving with velocity v through a uniform mass distribution with density ρ. It is possible to show that by adapting the Euler-equations to this specific case (i.e. Eq. 1.30, 1.31) and with an adequate choice of boundary conditions (see e.g. Frank et al., 2002), this models provide the following expression for the mass-accretion rate:

$$\dot{\mathrm{M}}_\bullet = \frac{4\pi\, G^2\, M_\bullet^2\, \rho}{(c_s^2 + v^2)^{3/2}}\,, \tag{1.39}$$

where c_s is the sound-speed of the infalling gas.

1.3.3 The radiative emission associated to SMBH growth

By exploiting the mass-accretion rate of the Shakura-Sunyaev model it is possible to compute the power emitted by an annulus of the accretion disk, at a given distance R from the BH. Indeed, by applying the conservation of energy for a viscous fluid:

$$\frac{dE}{dt}\,\Delta r = -\frac{3G\mathrm{M}_\bullet\,\dot{\mathrm{M}}_{\mathrm{disk}}}{2\mathrm{R}^2}\cdot\left(1-\sqrt{\frac{\mathrm{R}_{\mathrm{ISCO}}}{\mathrm{R}}}\right)\Delta r\,, \tag{1.40}$$

where Δr is the radial width of a disk annulus. By integrating this over the whole disk radius we obtain the total energy emitted by a Shakura-Sunyaev accretion disk (e.g. Frank et al., 2002):

$$\mathrm{L}_{\mathrm{disk}}(\mathrm{R}>\mathrm{R}_{\mathrm{isco}}) = \frac{\mathrm{G}\,\mathrm{M}_\bullet\,\dot{\mathrm{M}}_{\mathrm{disk}}}{2\mathrm{R}_{\mathrm{ISCO}}}\,. \tag{1.41}$$

This result is of extreme importance for explaining the strong luminosity of AGN. Indeed, according to the equivalence of matter and energy, the disk luminosity can be thought as a process in which the disk material is converted into energy and radiated with a given radiative efficiency ϵ_r, namely: $\mathrm{L}_{\mathrm{disk}} = \epsilon_{\mathrm{r}}\,\dot{\mathrm{M}}_{\mathrm{disk}}\,c^2$. By joining this with the above equation and substituting typical values for SMBH masses and radii (namely: $\mathrm{M}_\bullet \sim 10^8\,\mathrm{M}_\odot$ and $\mathrm{R}_\bullet \sim 10^{14}$ cm) we find $\epsilon_r \sim 0.42$, which is roughly one order of magnitude higher than the typical mass-to-energy conversion efficiency of nuclear reactions taking place in stellar cores. Despite different theoretical and observational arguments reach to lower values, i.e. $\epsilon_r \approx 0.1$ (e.g. Soltan, 1982; Marconi et al., 2004), the thin-accretion disk model still strongly suggests that mass-accretion onto SMBHs is the energy source responsible for the bright emission of AGN (see e.g. Abramowicz & Fragile, 2013, for a review).

Finally, thanks to the Shakura-Sunyaev model, it is possible to build a simple but descriptive model for the spectrum of the radiation emitted by an accretion disk. Indeed, due to its infall in the gravitational potential of the central BH, it can be shown that the accretion disk develops a radial temperature-gradient. By assuming that different annuli behave as black-bodies at different temperatures, it is therefore possible to determine the spectrum emitted by the superposition of all the annuli emissions (e.g. Ghisellini, 2013). According to this scenario, the emitted spectrum of accretion disks, as a function of the emitted frequency ν is proportional to ν^2 in the limit of low frequencies (i.e. $h\nu \ll \mathrm{k_B\,T}$), it evolves as $\nu^{1/3}$ at $h\nu \approx \mathrm{k_B\,T}$ and finally decays exponentially as $\exp[-(h\nu)/(\mathrm{k_B T})]$ in the high-frequency regime. This shows that spectra emitted by thin-disks are expected to show a *peaked* emission. The frequency at which this peak falls depends on the mass of the central objects. For typical SMBH masses, the peak of emission is generally found at X-ray/UV frequencies. These heuristic considerations can be supported by rigorous numerical computations and observations, which confirm that high fluxes of ionizing photons are effectively produced and observed in the emission of AGN sources (e.g. Frank et al., 2002; Abramowicz & Fragile, 2013).

Eddington limit

It must be stressed that mass-accretion onto compact objects, as well as the energy production due to nuclear reactions in stars, cannot produce any arbitrarily high emitted powers. Indeed, within intense radiation fields, the cross-section for the interaction of baryons and photons is generally high, therefore the effects of momentum exchange between radiation and matter can dominate the dynamic of baryons. This concept is formalized by the *radiation pressure*, i.e. the macroscopic effect by which intense radiation fields might produce and sustain pressure forces upon baryonic matter.

This idea was exploited by Eddington (1917) in order to derive the maximum luminosity which a spherically symmetric source can produce before the radiation pressure would balance the system self-gravity. According to this argument, the theoretical limit (i.e. the *Eddington luminosity*) is found by setting equal the radiation-pressure gradient to the gravitational one:

$$\mathrm{L}_{\mathrm{Edd}} = \frac{4\pi\,G\,\mathrm{M}\,m_P\,c}{\sigma_{\mathrm{T}}}\,, \tag{1.42}$$

where: M is the mass of the central object, $\mathrm{m_P}$ is the proton mass, c is the speed of light and finally σ_{T}

is the Thomson-scattering cross-section. By substituting the values of all the physical constants in the above equation we obtain:

$$\mathrm{L_{Edd}} = 1.26 \times 10^{38} \left(\frac{\mathrm{M}}{\mathrm{M_\odot}}\right) \mathrm{erg\,s^{-1}}\,. \tag{1.43}$$

Exponential mass growth of BHs

The theory of matter-accretion onto compact objects through thin-disks allow to determine a model for the time-evolution of $\mathrm{M_\bullet}$. This is based on the idea that, at any point during the process of mass accretion, BHs emit radiation at a given fraction $\mathrm{f_{Edd}} < 1$ of the Eddington limit, namely the *Eddington ratio*:

$$\mathrm{L_{bol}(t)} = \mathrm{f_{Edd}(t)} \cdot \mathrm{L_{Edd}(t)}\,, \tag{1.44}$$

where $\mathrm{L_{bol}} = \mathrm{L_{disk}}$ is the total *bolometric* luminosity of the accretion disk, i.e. the integral of the disk power across all emitted frequencies. We can rewrite eq. 1.42 as:

$$\mathrm{L_{Edd}(t)} = \frac{4\pi\, G\, \mathrm{M_\bullet(t)}\, m_P\, c^2}{\sigma_\mathrm{T}\, \mathrm{c}} = \frac{\mathrm{M_\bullet(t)}\, c^2}{\tau_\mathrm{Edd}}\,, \tag{1.45}$$

by defining the *Eddington time* $\tau_\mathrm{Edd} = 0.451\,\mathrm{Gyr}$. By joining these two equations, we obtain:

$$\mathrm{L_{bol}(t)} = \mathrm{f_{Edd}(t)} \cdot \frac{\mathrm{M_\bullet(t)}\, c^2}{\tau_\mathrm{Edd}}\,, \tag{1.46}$$

which can be used to derive a link between the mass-accretion rate of the central BH, by exploiting the thin-disk accretion scenario. Indeed, assuming that the inflowing matter is converted into *any*[11] form of energy with an efficiency η, then the quantity (1-η) measures the fraction of disk-matter that is actually available to be physically accreted onto the BH. On the other hand, we already defined the radiative efficiency ϵ, which measures the fraction of matter that is transformed into *radiative* energy. These two different concepts are formalized by the following equations:

$$\dot{\mathrm{M}}_\bullet(\mathrm{t}) = (1-\eta) \cdot \dot{\mathrm{M}}_\mathrm{disk}(\mathrm{t})\,, \tag{1.47}$$

$$\mathrm{L_{bol}(t)} = \epsilon_\mathrm{r}\, \dot{\mathrm{M}}_\mathrm{disk}(\mathrm{t})\, c^2\,. \tag{1.48}$$

Finally, by solving together equations 1.46, 1.47 and 1.48 we obtain the differential equation which describes the BH mass-growth evolution in time, namely:

$$\dot{\mathrm{M}}_\bullet(\mathrm{t}) = \frac{(1-\eta)}{\epsilon_\mathrm{r}}\, \mathrm{f_{Edd}}\, \frac{\mathrm{M_\bullet(t)}}{\tau_\mathrm{Edd}}\,, \tag{1.49}$$

which has a simple, exponential solution:

$$\mathrm{M_\bullet(t)} = \mathrm{M_\bullet(t_0)} \cdot \exp\left[\frac{(1-\eta)}{\epsilon_\mathrm{r}}\, \frac{\mathrm{f_{Edd}}}{\tau_\mathrm{Edd}}\, (\mathrm{t}-\mathrm{t_0})\right] \tag{1.50}$$

where $\mathrm{t_0}$ is an arbitrary reference time for the beginning of the BH mass-growth.

1.3.4 The basic structure of AGN

The Shakura-Sunyaev model provided strong impulse to the comprehension and interpretation of AGN emission. Nevertheless, the spectral energy distribution (SED) obtained through spectroscopic observations of these objects showed several properties which were not directly explained by the uniform, stationary picture traced by Shakura & Sunyaev (1976). These generally include broad and narrow line emission, multi-wavelength components and "bumps" throughout the whole electromagnetic spectrum as well as pronounced time-variability. Furthermore, it immediately appeared clear that AGN are not uniform spectral classes but rather exhibit a vast range of properties.

[11] The accretion efficiency η relates to the global conversion of disk-matter into all the possible forms of energy, i.e.: radiative, kinetic, magnetic, thermal (etc...) during the infall within the BH gravitational field.

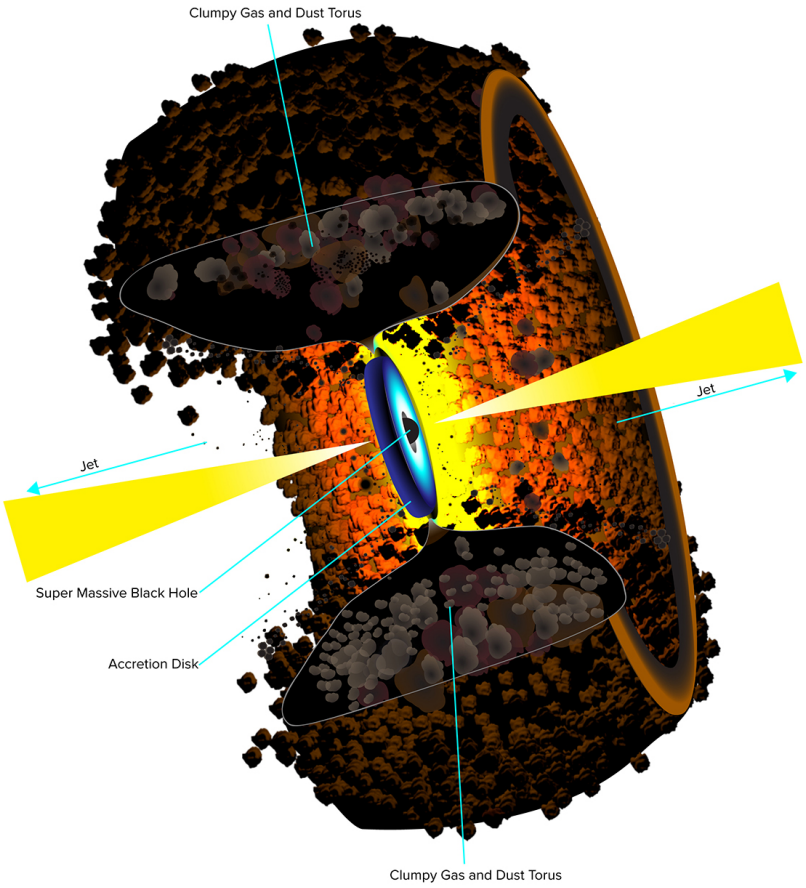

Figure 1.1: Artistic representation of the current model of the regions which surround an accreting SMBH. The depicted scenario is the result of the interpretation of several spectral features of AGN, which gradually integrated the thin-disk model of Shakura & Sunyaev (1976) with additional components in order to explain the observed spectral properties of AGN. *Credit: Bill Saxton, NRAO/AUI/NSF.*

With the progress of data collection, the different spectroscopic properties of AGN have been fit into an organic picture which describes the immediate surroundings of the BH-disk system (see e.g. Antonucci, 1993; Urry & Padovani, 1995; Haas et al., 2004; Beckmann & Shrader, 2012b; Netzer, 2015). This unified scheme can be seen as a modification of the Shakura & Sunyaev (1976) model in which a thin-disk is still considered as the main source of power, with the addition of surrounding structures to account for different spectroscopic features. a significant impulse to the development of this picture was provided by the discovery of the gravitationally lensed system known as the Einstein Cross (Huchra et al., 1985) whose micro-lensing (e.g. Irwin et al., 1989; Yonehara et al., 1998) has allowed to gain fundamental insights about the small-scale structure of the AGN accretion disks (e.g., Wyithe et al., 2000, 2002).

Figure 1.1 presents an artistic representation of the current unified picture of AGN. The massive BH and thin accretion-disk occupy the very central part of this system. They are believed to be enshrouded in a hot *corona* of plasma (not shown in Fig. 1.1), which is responsible for most of the high-energy, ionizing, X-ray emission observed in AGN spectra. This internal section of the system is surrounded by a region of clumpy gas and cloudy structures, which might be organized in a toroidal geometry (i.e. the *obscuring torus*, see figures 1.1 and 1.2, and e.g. Rees et al., 1982; Toba et al., 2013). This cloudy region is thought to be the origin of both the IR emission (e.g., Wyithe et al., 2002) and the line-emission observed in AGN spectra, due to the absorption and re-emission of the disk and corona ionizing fluxes by the orbiting clouds. Generally, this region is divided in two sub-regions, namely the *broad-lines region* (BLR, see e.g. Wyithe & Loeb, 2002b) and the *narrow-lines region* (NLR), named after the typical line-profiles they emit. In particular, the line broadening is attributed to the doppler shift of the emitting material, so that broader lines would correspond to faster rotational velocities and therefore would be produced closer to the central BH than narrow-lines. For this reason, the broadening of AGN lines carry information

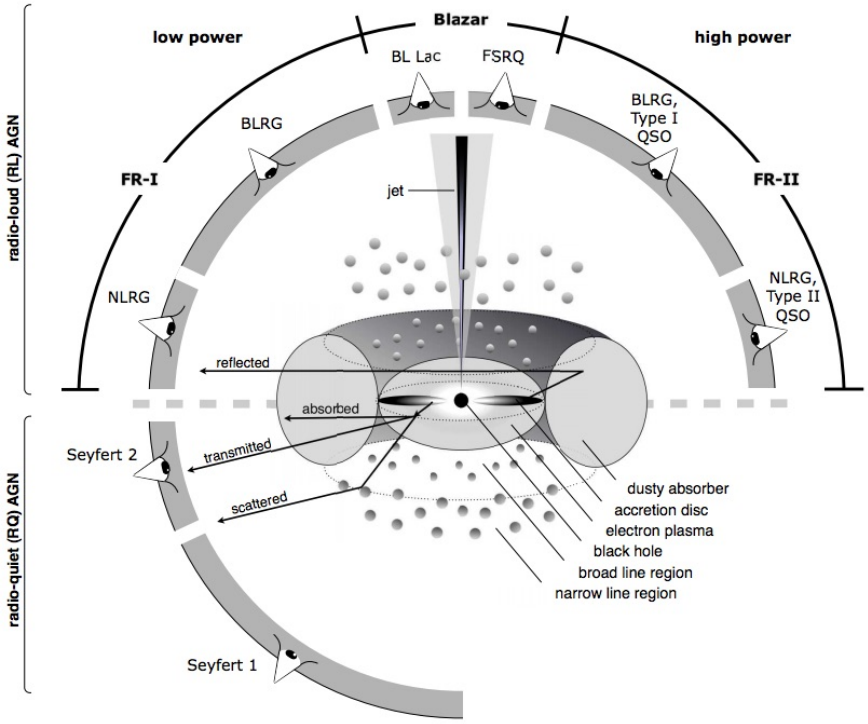

Figure 1.2: Classification scheme of AGN based on observation angle and radio-loudness (i.e. the presence of strong radio-flux in the source observed spectrum). This shows how the same central emitting structure can produce a plethora of different spectroscopic classes by simply allowing to access to the different emissions of its various sub-structures (figure from Beckmann & Shrader, 2012a).

about the mass of the central object, allowing to estimate it observationally (see e.g. Kelly et al., 2009; Marziani et al., 2011; Onoue et al., 2020).

Finally, the presence of two collimated *jets* of relativistic matter and radiation can be emitted perpendicularly to the accretion-disk rotation plane (although indications of jet misalignment have been reported, e.g. Sartori et al., 2016; Singal, 2016). These jets are observed extending up to significant distances from galaxies, and their interaction with the ionized IGM produces giant radio-lobes frequently observed around AGN. Thanks to this model, the various classes of AGN can be explained as different views of the same basic emitting system, observed under different inclination angles or modified in some of its additional components (e.g. by the presence or absence of the jet, or the obscuring torus). Figure 1.2 (from Beckmann & Shrader, 2012a) schematically present the different classifications of AGN as a function of viewing angle, emitted power and spectral properties.

1.4 Lyman-α emission from astronomical sources

The primordial composition of baryonic matter is constrained to be $\sim 75\%$ of hydrogen, $\sim 24\%$ helium and $\sim 1\%$ heavier elements by both observational measurements and theoretical arguments on primordial nucleosynthesis (see e.g. Ryden, 2016; Planck Collaboration et al., 2020). Consequently, it is reasonable to expect that a large amount of processes and observables in our Universe are related to the physics of hydrogen. For instance, as discussed in §1.2.1, gas cooling and star formation are largely based on the physics of hydrogen plasmas. Similarly, the theory of stellar evolution successfully explains the energy output of stars as the effect of nuclear reactions which convert hydrogen into helium, within stellar cores (e.g. Prialnik, 2009). Finally, on larger scales, the imprint of hydrogen absorption into the spectra of high-z QSOs (i.e. the so-called Lyα forest Rauch, 1998; Weinberg et al., 2003; Becker et al., 2015a), suggests that the IGM is largely composed of hydrogen. Consequently, the physiscs of emission, absorption and transfer of electromagnetic radiation produced by hydrogen is a fundamental topic in astronomy, and it is deeply-rooted in many astrophysical and cosmological research fields.

Thanks to quantum-mechanical arguments, it is possible to show that, within most astronomical contexts, the interaction between radiation fields and baryonic matter can efficiently produce strong fluxes of Lyman-*alpha* photons (Dijkstra, 2017, for a complete review). These are associated to the hydrogen transition $2p \longrightarrow 1s$, i.e. the first transition in the Lyman series of hydrogen atom, with a rest-frame wavelength of $\lambda_0 = 1215.67$Å. Indeed, due to the high probability of hydrogen atom transitions leading to the emission of Lyα photons, the Lyα line is by far the most prominent feature observed within the astronomical sources emission (e.g. Hainline et al., 2011; Dijkstra, 2017).

More in detail, Lyα is ubiquitously associated to the reprocessing of ionizing radiation, either by the photo-ionization and recombination of hydrogen atoms or by interaction of the latter with free electrons (i.e. atomic-cooling or *Lyα-cooling* of hydrogen gas, see §1.2.1). Furthermore, the Universe is believed to be globally ionized at $z \lesssim 6$ (see e.g. Malhotra & Rhoads, 2004; Fan et al., 2006a; Kashikawa et al., 2006; Ouchi et al., 2010; Matthee et al., 2015), therefore Lyα photons produced by the first forming galaxies are free to travel for cosmological distances without being absorbed. Finally, due to atmospheric absorption, the far-UV nature of the Lyα line makes it only detectable via space-borne telescopes when emitted by local, $z \sim 0$ sources. On the contrary, when redshifted up to $z \gtrsim 2$, Lyα photons enter into the optical band, hence they become observable by ground facilities. All the above properties make Lyα one of the favourite and most informative tools to study astronomical phenomena such as star-formation, galaxy formation, AGN activity, and the re-ionization of the hydrogen-rich IGM at high-z.

1.4.1 Lyman-α emitting sources: Star-Forming galaxies or AGN

The seminal work of Partridge & Peebles (1967) first introduced the idea of observing young galaxies through their Lyα emission. Indeed, these should appear bright at rest-frame Lyα wavelength, due to the reprocessing of the strong UV fluxes produced by young, newly-formed stars. Since then, a large amount of studies focused on Lyα emitters (LAEs), namely a class of sources emitting strong, rest-frame Lyα fluxes without other significant emission-line features (e.g., Cassata et al., 2011; Nakajima et al., 2018).

LAEs have been targeted at a wide variety of redshifts (e.g. Guaita et al., 2015; Hayes, 2015; Le Fèvre et al., 2015; Sobral et al., 2018b), hence providing an extremely valuable tool to study fundamental processes across cosmological scales and times (see e.g. Cassata et al., 2011; Matthee et al., 2015; Santos et al., 2020). Indeed, the examples and applications of works focused on LAEs are countless, either on the astronomical, cosmological and theoretical sides. For instance, these objects allowed to infer the evolution of the escape fraction of ionizing photons from high-z galaxies (e.g. Gronke & Dijkstra, 2014; Bouwens et al., 2016; Matthee et al., 2016), thus helping to shed light on the conditions of their ISM (e.g. Blanc et al., 2011; Christensen et al., 2012; Sobral et al., 2017, 2018b; Sobral & Matthee, 2019). In addition, the steep drop of LAEs number density observed at $z \gtrsim 6$ has been used to constrain the progress of cosmic reionization (see e.g. Malhotra & Rhoads, 2004; Kashikawa et al., 2006; Ouchi et al., 2010; Clément et al., 2012; Becker et al., 2015b; Cassata et al., 2015; Dijkstra, 2016; Santos et al., 2016; Drake et al., 2017; Konno et al., 2018), thus also contributing to the understanding of physical properties of the IGM on large scales. Furthermore, the distribution of LAEs in the surroundings of bright QSOs at high-z has been exploited to study the environment of these powerful sources (e.g. Steidel & Hamilton, 1992; Cantalupo et al., 2012; Fumagalli et al., 2017; Kikuta et al., 2017), as well as the build-up of large galaxy clusters and the progress of SF within these dense environments (see e.g. Møller & Warren, 1998; Stiavelli et al., 2001; Bădescu et al., 2017; Cai et al., 2018, 2019). Massive, high-z structures emitting diffuse Lyα radiation (namely, Lyα *blobs*, or LABs) have been also observed at high-z, providing extremely valuable information about the assembly of proto-clusters and filaments (e.g. Steidel et al., 2000; Møller & Fynbo, 2001; Monaco et al., 2005; Yang et al., 2009; Prescott et al., 2012; Saito et al., 2015; Cai et al., 2017a,b; Marques-Chaves et al., 2019). Finally, low-z observation of LAEs have provided extremely useful insights, as these systems have been interpreted as the analogues of high-z proto-galaxies. This allowed to infer the properties and evolution of their high-z counterparts (see e.g. Nilsson & Møller, 2009; Guaita et al., 2015; Hayes, 2015).

Common properties of high-z LAEs, such as the absence of strong emission or absorption line features (apart from the prominent Lyα), together with the faint continuum and their generally compact morphology has driven the interpretation of these objects as young ($\lesssim 50$ Myr) and low-mass

($M_* < 10^{10}\,M_\odot$) galaxies (e.g., Wilkins et al., 2011; Amorín et al., 2017; Hao et al., 2018; Santos et al., 2020; Matthee et al., 2021) with small rest-frame UV half-light radii ($R \lesssim 1-2$ Kpc, see e.g., Møller & Warren, 1998; Lai et al., 2008; Bond et al., 2012; Guaita et al., 2015; Kobayashi et al., 2016; Ribeiro et al., 2016; Bouwens et al., 2017a; Paulino-Afonso et al., 2018) which are actively star-forming ($SFR \sim 1-100\,M_\odot/yr$) and dust-poor (dust attenuation $A_V < 0.2$, see e.g., Gawiser et al., 2006, 2007; Guaita et al., 2011; Nilsson et al., 2011; Bouwens et al., 2017b; Arrabal Haro et al., 2020). Therefore, isolated and grouped high-z LAEs are interpreted as the progenitors of present-day galaxies and clusters, respectively (e.g., Matsuda et al., 2004, 2005; Venemans et al., 2005; Overzier et al., 2008; Guaita et al., 2010; Mei et al., 2015; Fumagalli et al., 2017; Khostovan et al., 2019). For this reason, an increasing number of works has focused on the determination of the properties of LAEs populations across different redshifts (e.g. van Breukelen et al., 2005; Ouchi et al., 2008; Sobral et al., 2009a; Konno et al., 2016; Santos et al., 2016; Sobral et al., 2017, 2018b; Zhang et al., 2021).

On the other hand, also AGN emit strong ionizing fluxes (see e.g. §1.3.3, and Haiman & Rees, 2001; Bolton et al., 2011; Dijkstra, 2017) and, therefore, they are expected to show a prominent Lyα line due to the reprocessing of this hard flux by neutral hydrogen in their surroundings (e.g. Wyithe et al., 2005a; Cantalupo et al., 2012; Borisova et al., 2016a,b). Differently from LAEs, AGN usually show multiple, additional lines, such as: `SiIV`, `CIV`, `CIII]` and `MgII` (see e.g. Vanden Berk et al., 2001; Telfer et al., 2002; Selsing et al., 2016; Hernán-Caballero et al., 2016). Therefore, at least in principle, it is easy to disentangle the classes of star-forming LAEs (SF LAEs) from Lyα -emitting AGN on the basis of their emitted radiation (e.g. Hainline et al., 2011).

It is worth noticing that the Lyα emission of AGN has a significantly different origin than the one of SF LAEs. Indeed, the latter is produced by *nebular emission* of gas clouds surrounding SF regions, which are illuminated by the radiative outputs of young stars (e.g. Shibuya et al., 2014). As a consequence, within SF LAEs the regions which reprocess radiation into Lyα photons are generally close to the actual sources of ionizing photons, therefore Lyα emission of SF LAEs carries direct information about ongoing SF processes (see e.g. Dijkstra, 2017). On the other hand, AGN emit Lyα due to the reprocessing of the X-ray and/or UV output of accretion disks by the more distant NLR and BLR cloudy regions (see §1.3.3 and Fig. 1.1). The latter can be found at large distances from the regions where ionizing photons are emitted ($\sim$ 1 pc for $M_\bullet \sim 10^9 M_\odot$ SMBH Negrete et al., 2014; Marziani et al., 2015). Furthermore, bright AGN can even illuminate their cosmological environment by fluorescence processes at even larger distances ($\sim 1-10$ Kpc, e.g. Haiman & Rees, 2001; Cantalupo et al., 2012; Borisova et al., 2016a). For these reasons, the Lyα emission of AGN has been used to infer properties of AGN environments (e.g. Fumagalli et al., 2017; Cai et al., 2019) or the reionization state of the IGM along their line-of-sight (e.g. Becker et al., 2015a; Greig & Mesinger, 2017; Croft et al., 2018).

Nevertheless, due to their strong Lyα line, AGN are expected to take part to the population of Lyα -emitting sources, and to be especially frequent among the most bright ones (Konno et al., 2016; Sobral et al., 2018b). Therefore, determining the relative abundance of both LAEs and AGN in the regime where these two classes overlap can help to constrain the interplay between SF and nuclear activity processes along the history of galaxies (e.g. Calhau et al., 2017, 2020). Recent compelling hints point towards identifying as AGN the majority of high-z Lyα-emitting sources showing high Lyα luminosity (i.e. $L_{Ly\alpha} > 2 \times 10^{43}\,erg\,s^{-1}$) (see e.g., Nilsson et al., 2011; Konno et al., 2016; Matthee et al., 2017b; Sobral et al., 2018a,b; Calhau et al., 2020). The work of Sobral et al. (2018a), in particular, pointed out the co-existence of two different classes of luminous $z \sim 2-3$ LAEs at roughly $L_{Ly\alpha} \gtrsim 3 \times 10^{43}\,erg\,s^{-1}$: on one side dust-free, highly star-forming galaxies (LAEs) and, on the other hand AGN. In addition, a significant fraction (at least $\gtrsim 20\%$) of bright LAEs selected by Matthee et al. (2017b) and Sobral et al. (2018b) respectively on the Boötes and COSMOS fields (with areas of $\sim 0.7\,deg^2$ and $\sim 2\,deg^2$), show X-Ray counterparts, which strongly points towards confirming them as AGN. Finally, Calhau et al. (2020) shows how the fraction of AGN within a sample of $z > 2$ Lyα-emitting candidates approaches $\sim 100\%$ at $L_{Ly\alpha} \gtrsim 10^{43.5}\,erg\,s^{-1}$. For generality, in our following discussion we consider all the sources showing a prominent Lyα as Lyα-emitting objects (see §4), independently of the star-forming or nuclear origin of their Lyα emission.

1.4.2 Morphology of Lyα emitters

With the progress of technical capabilities and observational techniques, it has become rapidly clear that Lyα emission, either from SF LAEs or AGN, can exhibit complex morphology in the surrounding of sources (e.g. Steidel et al., 2000). This is particularly true for high-z Lyα-emitting objects, for which the surrounding IGM can contain a significant fraction of neutral hydrogen (see e.g. Haiman & Rees, 2001; Wyithe et al., 2005a; Rauch et al., 2008; North et al., 2012; Saito et al., 2015; Leclercq et al., 2017; Matthee et al., 2020). Indeed, due to the resonant scatter of Lyα photons by H atoms, SF LAEs can be surrounded by faint Lyα-emitting halos and then appear more extended at Lyα wavelengths than at other wavelengths (e.g., Møller & Warren, 1998; Fynbo et al., 2001, 2003; Nilsson et al., 2009b; Finkelstein et al., 2011; Guaita et al., 2015; Wisotzki et al., 2016; Shibuya et al., 2019, but see also Bond et al. 2010, 2012 and Feldmeier et al. 2013). At the same time, the intense ionizing radiation emitted by bright QSOs at high-z can effectively interact with the IGM around these sources an be re-emitted by fluorescence at rest-frame Lyα wavelengths (e.g. Borisova et al., 2016b).

For these reasons, the importance of extended Lyα emission is to probe directly the circum-galactic medium (CGM) in the close-proximity of star-forming structures or AGN hosts, hence providing a view of their cosmological environments. This motivated a significant observational effort in order to either constrain the faint Lyα halos surrounding SF LAEs or to find and characterize rare, bright LABs at high-z (see e.g. Saito et al., 2006; Yang et al., 2009; Saito et al., 2015; Shibuya et al., 2018b).

Extended Lyα halos are usually characterized by very low surface brightness, therefore it is generally necessary to devote long observational programs to reliably observe them (see Bacon et al., 2015; Bădescu et al., 2017; Drake et al., 2017; Leclercq et al., 2017; Erb et al., 2018). This also applies to the peculiar class of high-z Lyα-emitting LABs, which usually show rest-frame very extended ($d \gtrsim 20-30$ kpc) and bright ($L_{Ly\alpha} > 10^{43}\,erg\,s^{-1}$) Ly$\alpha$ emission (see e.g., Matsuda et al., 2004; Bridge et al., 2013; Ao et al., 2015; Cai et al., 2017b; Cantalupo et al., 2019; Lusso et al., 2019).

On the other hand, without considering LABs or the eventual Lyα halo of SF LAEs, high-z rest-frame Lyα emission usually appears point-like sources, hence being morphologically indistinguishable from QSOs (point-like by definition). This potentially complicates the classification of sources within large catalogs of Lyα-emitting objects. We comment further on this topic in §4, when discussing our observational results.

1.5 The Luminosity Function of AGN and galaxies

As discussed in §1.3.3, the mass-accretion rate of active SMBHs can be linked to the luminosity emitted by their accretion disks, provided an assumption on ϵ and η (equations 1.47 and 1.48). Therefore, the distribution of AGN bolometric luminosity observed at a given z carries information about the mass-growth of active SMBHs at that time. This allows to reconstruct the history of SMBHs mass assembly in a statistical sense throughout their cosmological evolution (see e.g. Ueda et al., 2003; Hopkins et al., 2007; Ueda et al., 2014; Aird et al., 2015; Buchner et al., 2015; Shen et al., 2020). Analogously, the emission of galaxies at specific wavelengths can be linked to fundamental processes at play within them. One of the most important examples in this regards is the ensemble of Kennicutt relations, which link the galaxies SFR to different luminosity tracers (see Kennicutt, 1998). When these are coupled to the Schmidt relation (see Schmidt, 1959), the galaxies SFR can be connected to their gas content, thus allowing to track the progress of star formation and the consumption of gas within these systems throughout cosmic history. These kind of analysis are usually based on the distribution of the sources number density as a function of their luminosity, per unit of luminosity bin (dL), namely: their *luminosity function* (LF) (see e.g. Schmidt, 1968; Schmidt & Green, 1983; Warren et al., 1994; Miyaji et al., 2000; Shioya et al., 2008; Sedgwick et al., 2011; Gunawardhana et al., 2013; Bouwens et al., 2015b; Shen et al., 2020):

$$\Phi(L; z) = \frac{N(L; z)}{dV\, dL}\,, \tag{1.51}$$

where $N(L; z)$ is the number of sources observed at a given luminosity and redshift and dV is the volume over which the observations are performed. The next sections are devoted to introduce this fundamental

statistical tool and its computation.

1.5.1 The AGN luminosity function

Historically, AGN have been observed at different frequencies, allowing to build their LFs from X-ray to radio (see e.g. Fontanot et al., 2007; Croom et al., 2009; Bongiorno et al., 2010; Singal et al., 2011; Bonchi et al., 2013; Ross et al., 2013; Ueda et al., 2014; Aird et al., 2015; Lacy et al., 2015; Singal et al., 2016; Kulkarni et al., 2019; Zhang et al., 2021). This allows to construct a comprehensive picture of SMBHs activity (since different physical processes typically emit at specific frequencies, see e.g. Rybicki & Lightman, 1986; Ghisellini, 2013). For instance, IR observations can target the emission reprocessed by the obscuring material surrounding the active regions (e.g. Matute et al., 2006; Toba et al., 2013; Merloni et al., 2014; Singal et al., 2016), therefore providing a view of AGN obscuration. On the other hand, X-ray surveys can complement IR studies by providing a perspective on the global AGN population, being less biased towards obscuration (see e.g. Han et al., 2012; Ueda et al., 2014; Aird et al., 2015; Buchner et al., 2015; Miyaji et al., 2015). Studies about the AGN LF shed light on fundamental phenomena related to accreting SMBHs such as the impact of AGN feedback on SMBH growth (e.g. Hopkins et al., 2007) or the *BH downsizing* (Hasinger et al., 2005; Hirschmann et al., 2012). The latter, in particular, is an evolutionary picture according to which low-luminosity AGN appear to be more common at lower-z, as opposed to luminous AGN (e.g. Ueda et al., 2014; Miyaji et al., 2015). Thus LF studies allowed to establish that low-mass SMBHs typically assemble their mass later in time, in analogy with the galaxy downsizing.

On the other hand, the measurement of AGN luminosity over limited frequency intervals can be linked to the L_{bol} through relationships called *bolometric corrections* (see Marconi et al., 2004; Duras et al., 2020). In this way, the bolometric LF (bLF) can be built by compiling multi-frequency AGN observations (see e.g. Marconi et al., 2004; Hopkins et al., 2007; Shen et al., 2020). This is of great importance for theoretical models describing SMBH growth throughout cosmic times, which generally rely on the formalism described in §1.3.3 to compute and predict L_{bol}. Consequently, the bLF inferred from multi-frequency observations is a fundamental constraint for SMBH evolution models (see e.g. Bower et al., 2006; Croton et al., 2006; Wyithe & Loeb, 2003; Wyithe & Padmanabhan, 2006; Degraf et al., 2010; Volonteri et al., 2015a, 2016). In particular, the latter usually exploits the comparison with the measured bLF to test fundamental hypothesis on their physical prescriptions, such as the modelling of AGN fuelling (e.g. Hopkins & Hernquist, 2006), the evolution of SMBHs spin (e.g., Volonteri et al., 2013), the interplay between SMBH growth and galaxy mergers (e.g. Kauffmann & Haehnelt, 2000; Wyithe & Loeb, 2002a; Marulli et al., 2008; Hopkins & Quataert, 2010; Capelo et al., 2015; Capelo & Dotti, 2017; Steinborn et al., 2018; Izquierdo-Villalba et al., 2020), the build-up of SMBHs mass through alternate phases of activity and starvation (i.e. the so called *duty-cicle* Inayoshi et al., 2019; Delvecchio et al., 2020) or the consumption rate of accretion disks (e.g. Marulli et al., 2008; Volonteri et al., 2013; Izquierdo-Villalba et al., 2020).

1.5.2 The galaxies luminosity function

Analogously to the AGN case, also the LF of galaxies has been studied at different frequencies and redshifts, to understand galaxy evolution across the Universe history, exploiting both observational and numerical approaches (e.g. Reddy et al., 2008; Gronke et al., 2015b; Liu et al., 2016). For instance, (rest-frame) optical and UV LFs of galaxies greatly contributed to the understanding of SF processes at high-z and their evolution down to $z = 0$, as well as to constrain the progress of reionization (see e.g. Geach et al., 2008; Clément et al., 2012; Loveday et al., 2012; Gunawardhana et al., 2013; Bouwens et al., 2015b,a; Drake et al., 2017; Stefanon et al., 2017; Konno et al., 2018; Oesch et al., 2018; Santos et al., 2021). In a similar way, the galaxies emission measured at specific atomic-transitions wavelengths has been widely used for tracing the presence of ionized regions hosting recent SF within galaxies (see e.g. Bayliss et al., 2011). In particular, the H_α hydrogen transition[12] allows to build the H_α LF, which is one of the most important tools for statistical studies of galaxy evolution (see e.g. Shioya et al., 2008; Hayes

[12] H_α is the first transition of the Balmer hydrogen series, corresponding to $3p \longrightarrow 2s$.

et al., 2010; Sobral et al., 2013; Gunawardhana et al., 2015; Stroe & Sobral, 2015; Vilella-Rojo et al., 2021).

In addition to being a fundamental observable itself, the galaxies LF plays a crucial role also in relation to the AGN LF. Indeed, AGN and galaxies often emits at similar wavelengths, therefore complicating the study of their relative abundance. This issue is particularly evident both at UV-optical or at IR frequencies, where the AGN emission and that produced by stellar populations or ISM dust may overlap (see e.g. Babbedge et al., 2006; Sedgwick et al., 2011). Therefore, knowing the relative contributions of galaxies and AGN to the LF measured at a given frequency is crucial to disentangle the roles of these two source classes (see e.g. Fu et al., 2010; Wu et al., 2011). For instance, the relative abundance of galaxies and AGN can be used to study the role of AGN feedback into galaxies evolution at different redshifts (see e.g. van Velzen et al., 2015; Calhau et al., 2017) or to constrain the abundance of bright-SF galaxies at high-z (e.g. Konno et al., 2016; Stroe et al., 2017b; Sobral et al., 2018b; Calhau et al., 2020).

1.5.3 Binned estimates - Number counts and their corrections

As introduced above, the simplest form of a LF is the binned, volume-weighted histogram of measured luminosity, for a given sample of sources. This idea was formalized by Schmidt (1968) in what is now called the $1/\mathrm{V_{max}}$ method. Indeed, by defining $\mathrm{V_{i;\,max}}$ as the maximum volume within which a given source i can be observed (e.g. its emission is still received within the sampled wavelength range, or the source is still bright enough to be detected), then the LF can be defined as:

$$\Phi(\mathrm{L}, z) = \frac{4\pi}{\Omega_{\mathrm{obs}}} \sum_{\mathrm{i}} \frac{1}{\mathrm{V_{i;\,max}}\ \mathrm{dL}}\,, \tag{1.52}$$

where Ω_{obs} is the solid-angle subtended by the observations, 4π is the total solid angle, dL is the luminosity bin and the sum extends on all the i observed sources (see also Gunawardhana et al., 2013).

This definition accounts for the contribution of each *detected* source which is included in the sample used to compute the LF. Nevertheless, this sample can be affected by observational and/or selection biases. Namely, some sources that should have been detected and included in it might have been missed[13], thus leading to a lack of *completeness* of the catalog. This effect is increasingly evident at faint luminosity, since bright sources are detected more easily than dim ones (see §2.2.4 for further details). Furthermore, some of the sources included in the sample on the basis of their observed properties might be undesired *interlopers*, hence diminishing the sample *purity*. A typical example of the latter case is the inclusion of stars within astronomical samples of QSOs.

To correct for these possible biases, the weighted histogram in Eq. 1.52 are usually multiplied by additional weights which statistically account for the amount of either lost sources (i.e. completeness correction $\mathcal{C}_i$) or undesired interlopers (i.e. purity correction $\mathcal{P}_i$):

$$\Phi(\mathrm{L}, z) = \frac{4\pi}{\Omega_{\mathrm{obs}}} \sum_{\mathrm{i}} \frac{1}{\mathrm{V_{i;\,max}}\ \mathrm{dL}} \cdot \frac{\mathcal{P}_i}{\mathcal{C}_i}\,. \tag{1.53}$$

Both $\mathcal{C}_i$ and $\mathcal{P}_i$ are usually measured as a function of luminosity, but can be generalized to multi-variate functions in case complex biases must be taken into account (as in e.g. Gunawardhana et al., 2015). It is worth noticing that $\mathcal{C}_i$ and $\mathcal{P}_i$ are not the only factors which introduce uncertainties onto $\Phi(\mathrm{L}; z)$. Indeed, statistical errors on luminosity can shift sources among adjacent bins and affect the values of $\Phi(\mathrm{L}; z)$. Even assuming that these statistical errors equally affects luminous and faint sources, if the abundance of the latter in adjacent bins is not equal, then this bias is not symmetric and might produce a "leakage" of sources from the more populated bins towards the less populated ones. This statistical phenomenon is known as *Eddington bias* since it was postulated by Eddington (1913). In most real cases, LFs are characterized by more abundant sources at their faint-end, therefore the Eddington bias can effectively shift faint sources towards bright bins, leading to over-estimating the LFs bright end.

[13] In §2.2.4 we better detail the possible reasons leading to the loss of sources in astronomical observations.

1.5.4 Parametric estimates - Functional forms

In order to better interpret the binned estimates of LFs, several functional forms have been proposed to describe analytically the LFs shapes. In particular, when both luminosity and $\Phi(\mathrm{L}, z)$ are expressed in logarithmic units, LFs tend to appear as linear functions at the faint end (with a given *faint-end slope*) and rapidly decaying at the bright end. To describe this behavior, two main models have been put forward: a relatively simple double power-law and the Schechter function (Schechter, 1976). The first one describes a combination of two straight lines (in logarithmic units) and generally takes the form:

$$\Phi(\mathrm{L}) = \Phi_* \left[\left(\frac{\mathrm{L}}{\mathrm{L}_*} \right)^{\gamma_1} + \left(\frac{\mathrm{L}}{\mathrm{L}_*} \right)^{\gamma_2} \right]^{-1} , \tag{1.54}$$

which is fully determined by the four parameters Φ_* (global normalization) L_* (break luminosity) γ_1 and γ_2 (respectively: faint-end and bright-end slopes). These latter parameters control the actual shape of the DPL model, while Φ_* and L_* respectively control its vertical and horizontal shift. On the other hand, the Schecther function combines a power law at the faint-end and an exponential decay at the bright-end:

$$\Phi(\mathrm{L}) = \Phi_* \cdot \left(\frac{\mathrm{L}}{\mathrm{L}_*} \right)^{-\alpha} \exp\left(-\frac{\mathrm{L}}{\mathrm{L}_*} \right) , \tag{1.55}$$

and it is described by three parameters, namely Φ_* (global normalization), α (faint-end slope) and L_* (characteristic luminosity).

The common assumption in the literature is to describe the AGN LF using a double power law (DPL) model (e.g. Boyle et al., 1988; Ueda et al., 2003; Hopkins et al., 2007; Willott et al., 2010; Ross et al., 2013; Vito et al., 2014; Aird et al., 2015; Palanque-Delabrouille et al., 2016; Shen et al., 2020) and the galaxy LF with a Schechter function (e.g. Ouchi et al., 2008, 2010; Bouwens et al., 2015a; Matthee et al., 2017b; Sobral et al., 2018b; Viironen et al., 2018; Vilella-Rojo et al., 2021), independently on the frequencies at which either of these are evaluated, although different possibilities have been explored for both source classes (e.g. Warren et al., 1994; Schulze & Wisotzki, 2010; Manti et al., 2017; Stroe et al., 2017b; McGreer et al., 2018; Zhang et al., 2021). These parametrizations are fundamental for studies based on the sources LF, since their parameters carry physical significance and therefore efficiently describe the statistical properties of the underlying populations of astronomical objects. For instance, the steepness of the faint-end slope (in both DPL or Schechter cases) determines the relative importance of faint sources over bright ones. Analogously, the characteristic (or break) luminosity provides an indication of the turning-point between power-law regimes of the LF and rapidly decaying ones.

1.5.5 Evolution of the AGN luminosity function

The parametric forms of LFs allow the intuitive comparison between determinations obtained at different redshifts, therefore helping the understanding of the LF evolution throughout cosmic history (e.g. Ueda et al., 2003). The latter can also be parametrized, hence leading to the production of evolutionary models for the LF. In the specific case of the AGN LF, several models have been proposed to parametrize the LF evolution in time. Among these, the pure-luminosity evolution (PLE) or pure-density evolution (PDE) fix in time the LF shape (i.e. γ_1 and γ_2), and model the LF evolution only through rigid shifts of L_* or Φ_* (e.g. Boyle et al., 1988; Ueda et al., 2003). It has been shown that more complex model can better describe the AGN LF evolution, such as the luminosity-dependend density evolution (LDDE) which fixes every DPL parameter except for Φ_*, allowing it to vary as a function of L and z. Finally, also the luminosity *and* density evolution (LADE) or flexible double power-law (FDPL) evolutionary models have been used to extract information about the evolution of the underlying AGN populations (e.g. Schmidt & Green, 1983; Ueda et al., 2003; Hasinger et al., 2005; Hopkins et al., 2007; Ueda et al., 2014; Aird et al., 2015; Miyaji et al., 2015; Kim & Im, 2021).

These models have pointed out that AGN are an extremely variate population which undergoes significant evolution throughout cosmic history. In particular, according to the above-mentioned BH downsizing the peak of activity for bright and faint AGN appear shifted in time, namely at $z \sim 1$ and $z \sim 2-3$ respectively for faint and bright AGN (e.g. Ueda et al., 2014; Miyaji et al., 2015, but see also

Ross et al. (2013)). At the same time, the shape-distortions of AGN LF across cosmic time have been attributed to AGN feedback, which might suppress the growth of massive SMBHs formed early in the Universe history and therefore steepening the faint-end slope at lower redshifts (e.g. Hopkins et al., 2007). Finally, the presence of significant obscuration, whose impact depends on both luminosity and redshift, has been recently studied via multi-frequency determinations of the AGN LF and physical models (e.g. Aird et al., 2015; Buchner et al., 2015, 2021).

Due to its fundamental connections with the physical processes driving the build-up of SMBHs populations, the AGN LF and its evolution are crucial observables to which numerical models and simulations of SMBH activity must adapt. In this regards, several models for SMBH evolution developed in the last decades are generally able to reproduce the observed AGN LF by means of different physical prescriptions (e.g. Kauffmann & Haehnelt, 2000; Marulli et al., 2008; Degraf et al., 2010; Sijacki et al., 2015; Rosas-Guevara et al., 2016; Steinborn et al., 2018; Marshall et al., 2020a; Izquierdo-Villalba et al., 2020). On the other hand, these models are either applied to large simulated volumes (and usually they exploit simplified prescriptions for high-z BH seeding processes), or they describe SMBH mass-growth in idealized contexts, whose extension to cosmological environments is not straightforward (e.g. Anglés-Alcázar et al., 2017a; Valiante et al., 2016). Therefore, a physical model which is able to self-consistently follow the evolution of SMBHs from their formation at $z \gtrsim 7$ down to $z \sim 0$, over a cosmological context, can be thought as a missing link between these two approaches. The theoretical part of this thesis work builds on this idea and presents a comprehensive model of SMBH formation and evolution in a cosmological context. As detailed in Chapter §3, the predictions of this model for the bLF of active SMBHs were instrumental for developing and calibrating its physical prescriptions.

1.5.6 The Lyα luminosity function

As anticipated in §1.5.1, observed LFs can be constructed from measurements performed over broad wavelength intervals (e.g. IR, UV or X-ray) or by targeting the specific rest-frame wavelength of a particular emission-line[14] (see e.g. Hao et al., 2005; Bongiorno et al., 2010; Stroe et al., 2017b; Spinoso et al., 2020; Zhang et al., 2021). The latter usually probe the particular physical phenomena responsible for their production, as in the above-mentioned case of H_α line probing SF processes (see e.g. Shioya et al., 2008; Sobral et al., 2016; Gunawardhana et al., 2013; Vilella-Rojo et al., 2021). Analogously, also Lyα emission can be used to track star-formation and SMBH activity in the Universe, as introduced in §1.4 (see also Dijkstra, 2017). Consequently, the Lyα LF is a widely exploited tool to study the progress of galaxy formation and the evolution of cosmological structures from $z \sim 6$ (e.g. Stiavelli et al., 2001; Ouchi et al., 2005; Overzier et al., 2008; Shibuya et al., 2014; Ouchi et al., 2018; Zahedy et al., 2019; Santos et al., 2020) down to $z \sim 0$ (e.g. Guaita et al., 2015; Weinberger et al., 2018; Shibuya et al., 2019). Indeed, due to the link between Lyα emission and star formation (e.g. Sobral & Matthee, 2019), the Lyα LF carries fundamental information about the initial build-up of galaxies stellar mass and the evolution in time of their SFRs (see also Greif et al., 2009b; Sobral et al., 2014; Konno et al., 2016; Matthee et al., 2021).

Several recent works have focused on the construction of the Lyα LF at $z \geq 2$ either resorting to photometric or spectroscopic approaches (Gronwall et al., 2007; Ouchi et al., 2008; Blanc et al., 2011; Clément et al., 2012; Konno et al., 2016; Sobral et al., 2017, 2018b; Zhang et al., 2021). Generally, these works were based on deep observations carried over narrow sky regions (up to few squared degrees, as in e.g., Matthee et al., 2014; Cassata et al., 2015; Matthee et al., 2017b; Ono et al., 2018). Their findings describe a Lyα LF which follows a Schechter function (Schechter, 1976) at relatively faint line luminosity (i.e. $L_{Ly\alpha} \lesssim 10^{42.5}$, see e.g., Ouchi et al., 2008; Konno et al., 2016; Sobral et al., 2016; Matthee et al., 2017a), a regime mostly occupied by low-mass star-forming galaxies (e.g., Hu et al., 1998; Kudritzki et al., 2000; Stiavelli et al., 2001; Santos et al., 2004; van Breukelen et al., 2005; Gawiser et al., 2007; Rauch et al., 2008; Guaita et al., 2011).

The bright-end of the Lyα LF is instead populated by rare, bright and SF-bursty Lyα-emitting systems (see the discussion in §1.4.1 and e.g., Matsuda et al., 2011; Bridge et al., 2013; Cai et al., 2017b, 2018; Marques-Chaves et al., 2019, 2020) as well as spatially-extended groups of Lyα-emitting sources

[14] See §2.2 for a detailed treatment of emission lines within sources emission.

(e.g. Cai et al., 2017b; Hartwig et al., 2016; Lusso et al., 2019; Matthee et al., 2020) and AGN (Matthee et al., 2017b; Sobral et al., 2018b; Calhau et al., 2020; Zhang et al., 2021). Consequently, the interest towards the Lyα LF bright-end is motivated by both the search for bright and extreme star-formation powered LAEs (Sobral et al., 2015; Hartwig et al., 2016; Sobral et al., 2016; Shibuya et al., 2018a) and the characterization of high-z AGN/QSOs populations (e.g. Konno et al., 2016; Matthee et al., 2017b; Calhau et al., 2020; Zhang et al., 2021).

Current constraints on the Lyα LF at high luminosity are somewhat poor, given the relatively small cosmological volumes usually probed by past works (e.g., Fujita et al., 2003; Blanc et al., 2011; Herenz et al., 2019). In particular, recent measurements show hints about an excess with respect to the expected exponential Schechter decay, at $L_{Ly\alpha} \gtrsim 10^{43}$ (e.g., Konno et al., 2016; Viironen et al., 2018), possibly due to the contribution of a faint-AGN population (see e.g., Matthee et al., 2017b; Sobral et al., 2018b). This is supported by the tomographic analysis performed by Sobral et al. (2018b) in the `COSMOS` field by using a combination of optical, infrared and X-Ray data. In their work, the major contribution to the LF at $L_{Ly\alpha} \gtrsim 10^{43}$ is provided by X-Ray loud sources (i.e. likely AGN, see also Matthee et al., 2017b; Calhau et al., 2020). This contribution completely vanishes at $z \gtrsim 3.5$ in their data, thus paralleling the peak of AGN activity usually observed at $z \lesssim 3$ (e.g., Hasinger et al., 2005; Ueda et al., 2014; Miyaji et al., 2015). Finally, the constraints on the $z \gtrsim 2$ Lyα LF bright-end are prone to contamination by lower-redshift interlopers. For example, Sobral et al. (2017) and Stroe et al. (2017a) show that a consistent fraction of bright LAE candidates at $z > 2$ are actually AGN at $z \gtrsim 1.5$ emitting `CIV` at $\lambda_{obs} \sim 4000$Å.

Unfortunately, most of these works analyzed narrow sky regions, which probe small fractions of the $z \sim 2$ cosmological volumes, therefore being susceptible to cosmic-variance[15] (e.g., Fujita et al., 2003; Blanc et al., 2011; Herenz et al., 2019). These caveats can be addressed by exploiting wide-area observations over frequency ranges which allow to detect the redshifted Lyα emission of $z \sim 2$ sources. As detailed in chapter §4, the current thesis work fits in this context by performing the first photometric determination of the Lyα Lf over unprecedented regimes of luminosity and number density. We underline that the results presented in §4 were recently confirmed by the spectroscopic analysis of Zhang et al. (2021), which nicely complemented and supported the results presented in this thesis.

[15] The uncertainty on observational results introduced by intrinsic field-to-field variations typical of observations carried out in different, narrow, observed sky regions is usually referred to as *cosmic variance* in astronomy.

Chapter 2

Methods

The study of SMBHs formation, evolution and interplay with their host galaxies is carried out through a wealth of methodologies and techniques, which span between the theoretical, numerical and observational fields of astronomy and astrophysics. The current thesis work approaches both the formation of SMBHs (i.e. the BH seeding problem, see §1.2) and their subsequent evolution by exploiting both numerical and observational techniques. Given the wide range of methods involved in our approach, we devote this chapter to introduce their main concepts and definitions.

2.1 Numerical simulations and analytical models

Despite most of the breaking-ground discoveries and advancements in modern Astronomy are inevitably based on observations, the significant theoretical progress of last decades has been greatly favoured by the introduction and development of numerical simulations of complex astrophysical phenomena. Indeed, thanks to current computational capabilities, it is possible to numerically solve highly non-linear equations which generally describe astrophysical systems under realistic (i.e. non-simplified) assumptions.

Numerical methods in theoretical astrophysics can be divided in two broad classes: numerical simulations and analytical (or semi-analytical) models. Both methods solve the evolutionary equations of the systems they analyze but with radically different approaches. Indeed, numerical simulations solve these non-linear equations directly (i.e. numerically) in subsequent time-steps over a spatial grid of coordinates, while analytic and semi-analytic models employ a set of simplified, analytic versions of the same equations (see §2.1.1 for details). The advantage of numerical simulations is their ability to properly capture and describe the physical phenomena they model, which can usually be tracked and resolved self consistently in time. Nevertheless, this comes at the price of a generally high computational cost, which limits the size of the Universe regions which can be simulated. On the other hand, (semi-)analytic models are generally much less expensive in terms of computational effort, but they need to rely on stronger theoretical assumptions in order to describe physical process through a set of linear, analytic expressions.

Numerical simulations generally tackle the description of astrophysical phenomena by solving their descriptive equations over a set of discrete *particles* which sample either continuous fields (e.g. baryonic gas density distributions) or populations of discrete objects (e.g. single BHs, or clusters of stars or galaxies). As discussed in previous sections, most of the realistic descriptions of astrophysical systems involve the combination of gravitational, hydrodynamic and (eventually) magneto-dynamic equations. In most of the cases, the gravitational part of these systems is separable from hydrodynamics, so that usually simulations follow these two aspects separately (see e.g. Springel, 2016, for a comprehensive review). In general, simulations can follow: only gravity (*N-body simulations*) or involve also hydrodynamics computations (i.e. *hydrodynamic simulations*). Finally, a further degree of classification can be defined according to the type of initial conditions:

– *cosmological simulations* are generally performed over large simulated volumes, starting from high-z initial conditions built on DM density distributions. The latter are obtained from the linear theory

of structures growth (see §1.1.1) and then populated with baryonic gas. Finally, the evolution of DM and baryons is followed self consistently.

– *isolated simulations* usually describe a specific system isolated from its cosmological context. This system is initalized with ad-hoc conditions, which are generally chosen in an optimal way to follow the specific phenomenon under analysis.

This broad topic covers several further aspects, such as the actual definition of particles (either gravitational or hydrodynamic ones) and the efficient resolution of physical equations. Since these topics lay beyond the context of this work, we refer to Springel (2016) for an exhaustive discussion.

2.1.1 Analytical and semi-analytical models

As discussed in the previous section, accurate numerical simulations can be computationally expensive despite recently-developed efficient computational methods (Springel, 2016). This is mainly due to the numerical resolution of both i) the gravitational interaction between all particles in a given volume, ii) the Euler equations for gaseous part of the system. This can be further complicated by the introduction of magnetic, radiative phenomena as well as GR effects.

Analytic and semi-analytic models allow to bypass both these computational issues by resorting to simplified approaches. Indeed for the modelling of baryonic processes (i.e. hydrodynamics and eventual magnetic and/or radiative phenomena), these codes use sets of linearized equations, which can be efficiently solved even on large simulated volumes (see e.g. Cole et al., 2000; Guo et al., 2011; Baugh, 2013; De Lucia et al., 2014; Henriques et al., 2015; Lacey et al., 2016; Lagos et al., 2018; Cora et al., 2018). On the other hand, for the gravitational side, they resort to DM merger trees, i.e. catalogs of DM halos which are linked together in parent-descendant relationships through their time-evolution. Generally, there are two main methods to obtain the input merger trees for (semi-)analytic models:

– *purely-analytic* methods allow to construct the merger histories of DM halos starting from the statistical considerations about halos abundance at a given mass and redshift. This theory is known as the Press-Schechter formalism (PS, Press & Schechter, 1974, generalized by Lacey & Cole (1993)) and it uses linear theory (§1.1.1) to predict the distribution of masses and redshifts of DM halos. According to this approach, analytic merger trees can be built along the history of any number of DM halos In particular, by randomly sampling the halo-masses distribution predicted by PS formalism at each z, different realizations of DM merger trees can be produced in a very fast and efficient way. As a drawback, the PS formalism only predicts a halo-mass distribution, therefore the actual relative positions of halos are usually unknown.

– *N-body* methods require that a DM-only, N-body simulation would produce its self-consistent outputs, which are then analyzed to extract the halo catalogs used as inputs of analytic models. Given the non-analytic origin of these merger trees, models resorting to this approach are usually called *semi-analytical models* (SAM). By exploiting the outputs of their underlying DM simulation, SAMs take into account the actual cosmological environment in which the different DM halos evolve, hence they can model self consistently the relative positions of DM halos and the dynamics of their interactions.

Both analytic and semi-analytic codes model the evolution of baryonic components on top of DM merger trees through a set of linearized, evolutionary equations. On one side, this approach comes at the cost of using approximation for the modelling of physical processes. On the other hand, this allows to run these models on large cosmological boxes, especially if compared to e.g. high-resolution N-body hydrodynamic simulations.

Furthermore, (semi-)analytic codes are usually much faster to execute than numerical simulations. This makes them ideal candidates for a number of applications in computational astrophysics. Indeed, they proved to be extremely powerful and versatile tools for testing physical models within cosmological contexts, as well as to efficiently perform the estimation of their parameters. This is mainly due to their relatively short execution-times, which allow to perform several repetitions of the same initial conditions with different model parameters or entirely distinct numerical prescriptions. In the past, several

physical prescriptions were successfully developed and studied using SAMs. For instance, Kauffmann & Haehnelt (2000); Croton et al. (2006); Bonoli et al. (2009); Croton et al. (2016) and Qin et al. (2017) are some examples of works which connected the growth of BHs to the hierarchical build-up of structures. Always in the context of SMBH formation and evolution, Hopkins & Hernquist (2006) developed a model of AGN fuelling via stochastic accretion of gas, Marulli et al. (2008) developed a model for the time-prolonged mass-accretion onto SMBHs, whhich was complemented by Izquierdo-Villalba et al. (2020) with the introduction of BH-spins evolution, GW emission and BH-merger time delays (see §2.1.2 below). Furthermore, Dotti et al. (2013) and Sesana et al. (2014) linked the evolution of BH spins to the properties of galactic bulges using analytic methods and Croton et al. (2006) introduced a relatively simple yet extremely successful model for AGN feedback in a semi-analytic context. Finally, Volonteri et al. (2008), Valiante et al. (2016) and Sassano et al. (2021) are some examples of works which approached the BH-seeding problem and the evolution of the emerging SMBHs population with analytic methods.

2.1.2 The L-Galaxies semi-analytic model

The numerical part of this thesis presents a model for high-z BH seeding, embedded in a modified version of the Munich Galaxy Formation model (`L-Galaxies`, see e.g. Guo et al., 2011). In particular, our work is based on the version presented by Henriques et al. (2015) and recently modified by Izquierdo-Villalba et al. (2019b, 2020). `L-Galaxies` includes detailed treatment for most of the baryonic physics involved in the evolution of galaxies and SMBHs, and it has been designed to be applied on the merger trees of `Millennium` (`MR`) and `Millennium-II` (`MR-II`) N-body simulations (see, respectively: Springel et al., 2005; Boylan-Kolchin et al., 2009). Indeed, the combination of the large simulated volume of `MR` (500 Mpc h^{-1}) and the high halo-mass resolution of `MR-II` ($M_{vir}^{min} = 3.84 \times 10^7\, M_\odot\, h^{-1}$) allows to track baryonic processes over a wide dynamical range within a self-consistent cosmological context (see discussion below).

By embedding our BH-seeding prescriptions into `L-Galaxies`, we have the opportunity to study this process and the evolution of its SMBH descendants in a self-consistent cosmological context. In particular, we rely on the `MR-II` merger trees, since they resolve and track self-consistently the evolution of high-z atomic-cooling halos (see §1.2.1). Furthermore, by using `L-Galaxies`, our BH-seeding prescriptions can potentially be extended to the wider simulated volume of `MR`, hence allowing to study physically-motivated BH-seeding prescriptions on wider cosmological scales. This is a yet poorly-constrained approach, due to the computational challenges it presents. Indeed, current methods usually resort to either detailed, small-scale simulations whose generalization to wide cosmological volumes is uncertain, or to very simplified approaches about BH-formation which fail to capture the long-term effects of BH-seeding processes on the build-up of global SMBH populations.

`L-Galaxies` is built to exploit DM-only merger trees and track the evolution of their baryonic counterparts using a number of time *sub-steps* between different snapshots of the input merger-trees. `L-Galaxies` follows the evolution of galaxies, from small dwarf satellites, to massive galaxies harbouring at the potential-well bottom of large FoF groups (see §1.1.1). In particular, the `L-Galaxies` version presented in Henriques et al. (2015) was calibrated jointly on `MR` and `MR-II` in order to reproduce crucial observables related to galaxy evolution at $z < 4$, such as the stellar mass-function (SMF) and the fraction of passive satellites over a wide dynamic range (see e.g. Henriques et al., 2015).

This version of the SAM follows self-consistently the evolution of baryons starting from their primordial infall into halos. In particular, `L-Galaxies` computes the amount of infalling matter as *baryonic fraction* f_b of the DM halos resolution-mass M_{vir}^{res} ($f_b = 0.155$ for the `Planck2015` cosmology, Planck Collaboration et al., 2016a,b). This gas is assumed to undergo shock-heating during its gradual accretion onto the hosting DM halo, and it is therefore modelled as an hot atmosphere (i.e. a hot-phase gas component). Following the approach of White & Rees (1978) and (White & Frenk, 1991), the hot gas M_{hGas} is allowed to cool according to a temperature and metallicity-dependent cooling function $\Gamma(T_{hot}; Z_{hot})$ which also accounts for the cooling of pristine gas (but not for H_2 cooling, see Sutherland & Dopita, 1993). Indeed, the Henriques et al. (2015) version of `L-Galaxies` does not model the gradual metal-enrichment of the IGM, so that every time a new halo is resolved in the merger-trees of the underlying N-body simulation, its associated primordial infall is pristine, independently of redshift. Gas cooling can follow two regimes,

depending on the value of the *cooling radius*:

$$r_{cool} = \left(\frac{t_{dyn}\, M_{hot}\, \Gamma(T_{hot}; Z_{hot})}{6\pi\, \mu\, m_H\, T_{vir}\, R_{vir}} \right)^{1/2} , \tag{2.1}$$

where $t_{dyn} = R_{vir}/V_{vir} = 0.1/H(z)$ is the halo *dynamical time*, μ and m_H are the gas molecular weight and the hydrogen-atom mass. The quantities labelled as "hot" are referred to the hot gas component, and finally T_{vir} and R_{vir} are given by equations 1.28 and 1.29. This cooling regime is subject to the condition $r_{cool} < R_{vir}$ and it describes the transfer of gas from the hot phase to the cold phase in a stationary flow at a rate:

$$\dot{M}_{cool} = M_{hot} \frac{r_{cool}}{R_{vir} \cdot t_{dyn}} , \tag{2.2}$$

On the other hand, when $r_{cool} > R_{vir}$, the whole hot-gas mass is interested by efficient and vigorous cooling, by what is defined as *rapid-infall* regime. In this case, the cooling rate is given simply by:

$$\dot{M}_{cool} = \frac{M_{hot}}{t_{dyn}} , \tag{2.3}$$

The mass of cooled gas (M_{cGas}) settles in a rotationally-supported disk whose initial angular momentum $\mathbf{J}_{cGas}$[1] is determined by the specific angular momentum of its host DM halo (i.e. $j = \mathbf{J}_{halo}/M_{vir}$). As discussed in Guo et al. (2011), the $\mathbf{J}_{cGas}$ of each structure evolves through mass exchanges with the hot-phase gas, star formation and galaxy mergers.

The rotationally-supported cold-gas disk harbours star formation processes as soon as its accumulated mass overcomes the critical threshold:

$$M_{cGas\,;\,crit} = M_{cGas;0} \left(\frac{V_{vir}}{200\,Km\,s^{-1}} \right) \left(\frac{R_{gas}}{10\,Kpc} \right) , \tag{2.4}$$

where R_{gas} is the cold-gas disk radius determined from $\mathbf{J}_{cGas}$ (see Henriques et al., 2015, and references therein), while $M_{cGas;0}$ is a free parameter of the model. With this definition, the mass of newly formed stars is:

$$\dot{M}_* = SFR = \alpha_{SF} \frac{M_{cGas} - M_{cGas\,;\,crit}}{t_{dyn}} , \tag{2.5}$$

where α_{SF} is the star-formation efficiency parameter of the model. The typical time interval associated to `L-Galaxies` sub-steps is $\delta t_{step} \gtrsim 100\,Myr$, which is much longer than the typical time spent by massive stars in their main-sequence, i.e. $t_{MS} \lesssim 5\,Myr$ (e.g. Prialnik, 2009). Therefore, `L-Galaxies` makes use of the instantaneous-recycling approximation, according to which a fixed fraction $R = 0.43$ of newly-formed stellar mass is immediately returned to M_{cGas}, hence modelling the recycling of gas mass due to SNe feedback. The latter heats and chemically enriches M_{cGas} by releasing energy and metals into it. The energy output of SNe feedback is coupled to the SFR via the equation:

$$E_{SNe} = \epsilon_{SNe} \frac{1}{2} \dot{M}_* \, \delta t_{step} \, V_{SNe}^2 , \tag{2.6}$$

where ϵ_{SNe} is a free-parameter controlling the overall amplitude of SNe feedback, while $\frac{1}{2}V_{SNe}^2$ is the energy per unit mass injected by SNe into the ISM (with $V_{SNe} = 630\,km\,s^{-1}$, see Henriques et al., 2015, for details).

The quantity E_{SNe} is the total energy available to both re-heat a fraction of M_{cGas} into M_{hGas} and, eventually *eject* a fraction of M_{hGas} outside the R_{vir} of DM halos. This possibility is regulated by an

[1] We use the notation according to which bold-face letters denote vector quantities, so that: $\mathbf{J} = \overrightarrow{J} = (J_x, J_y, J_z)$.

energy and mass balance which obeys to the following equations:

$$\mathrm{E_{reheat}} = \frac{1}{2}\,\dot{\mathrm{M}}_{\mathrm{reheat}}\,\delta \mathrm{t_{step}}\,\mathrm{V_{vir}^2}\,, \tag{2.7}$$

$$\dot{\mathrm{M}}_{\mathrm{reheat}} = \epsilon_{\mathrm{reheat}}\,\dot{\mathrm{M}}_{*}\,, \tag{2.8}$$

$$\epsilon_{\mathrm{reheat}} = \varepsilon \cdot \left[0.5 + \left(\frac{\mathrm{V_{max}}}{\mathrm{V_{reheat}}}\right)^{-\beta}\right]\,, \tag{2.9}$$

where $\mathrm{V_{max}}$ is the maximum circular velocity of the DM halo host, while $\mathrm{V_{reheat}}$ and ε are free parameters. From this system of analytic equations it is possible to derive the amount of mass which can be ejected beyond $\mathrm{R_{vir}}$ by each episode of star formation:

$$\dot{\mathrm{M}}_{\mathrm{ej}} = 2\frac{(\mathrm{E_{SNe}} - \mathrm{E_{reheat}})}{\mathrm{V_{vir}^2}}\,. \tag{2.10}$$

The numerator of the above equation is bound to be positive or null, due to energy conservation, i.e. $(\mathrm{E_{SNe}} - \mathrm{E_{reheat}}) \geq 0$. Consequently, when $\mathrm{E_{SNe}} = \mathrm{E_{reheat}}$ the galaxy SFR is not high enough for the associated SNe feedback to eject mass into the IGM. On the contrary, if $\dot{\mathrm{M}}_{\mathrm{ej}} > 0$, mass-ejecta are stored in a separate baryonic reservoir for future reincorporation of metal-enriched IGM in $\mathrm{M_{hGas}}$, along the evolution of galaxies (see Henriques et al., 2015, for details about gas reincorporation).

In `L-Galaxies`, secular SF processes take place in the disks of galaxies, as this is the only structure that forms directly from the cooling of primordial infall. On the other hand, the formation of galactic bulges is strictly related to both merger interactions (either minor or major) and disk instabilities. More in detail, the remnants of major mergers are assumed to be disk-less galaxies, since the stellar content of both merging galaxies is transferred to the bulge component of the remnant. On the other hand, minor interactions do not destroy the disk of the major interacting galaxy, and the entire $\mathrm{M_*}$ of the satellite is directly transferred to the bulge of the central galaxy. Finally, the prescriptions for disk instabilities and extremely-minor interaction were greatly improved by the work of Izquierdo-Villalba et al. (2019b) in order to reconcile the results of `L-Galaxies` with observed galaxy morphologies, when applying the SAM on both `MR` and `MR-II` merger trees. We refer to this latter work for a detailed discussion of merger interactions in the `L-Galaxies` context.

Mergers also trigger SF-bursts, i.e. star-formation events which are not related to the secular accumulation of cold-gas mass in the galactic disk. The newly-formed mass of stars during SF bursts is modelled through the *collisional formulation* of Somerville et al. (2001):

$$\mathrm{M_*^{burst}} = \alpha_{\mathrm{SF}}^{\mathrm{burst}}\left(\frac{\mathrm{M_{sat}}}{\mathrm{M_{cen}}}\right)^{\beta_{\mathrm{burst}}}\mathrm{M_{cGas}}\,, \tag{2.11}$$

and are added to the disk of the merger remnant. The values $\alpha_{\mathrm{SF}}^{\mathrm{burst}}$ and β_{burst} are free parameters of the model, while $\mathrm{M_{sat}}$ and $\mathrm{M_{cen}}$ are respectively the masses of the satellite and central galaxy in the interaction.

BH physics in the `L-Galaxies` model

The standard, Henriques et al. (2015) version of `L-Galaxies` does not account for BH-seeding processes; rather, an BHs are allowed to form and grow at the center of galaxy-mergers remnants, at any z, independently of environmental and local gas conditions. These BHs can grow in time via accretion of either hot or cold gas, as well as via BH-BH mergers. The latter process happens immediately during the interaction of two galaxies which both carry a central BH. More specifically, the Henriques et al. (2015) version on `L-Galaxies` does not account for delays between galaxies and BHs mergers (see e.g. Volonteri et al., 2020, for a recent numerical simulation accounting for BH-BH mergers delay). Once the remnant BH is formed after a merger, it receives a quantity $\Delta\mathrm{M_{acc}}$ of cold gas, thus mimicking the effect of tidal distortions caused by the galaxies interaction (as in e.g. Capelo & Dotti, 2017). $\Delta\mathrm{M_{acc}}$ is proportional to the baryonic merger-ratio $\mathrm{m_r} = \mathrm{M_{sat}}/\mathrm{M_{cen}}$ of the interacting galaxies and to the $\mathrm{M_{cGas}}$

of the central galaxy:

$$\Delta M_{acc} = f_{BH} \cdot m_r \cdot \frac{M_{cGas}}{\left(1 + \frac{V_{BH}}{V_{vir}}\right)^2} . \tag{2.12}$$

Both f_{BH} and V_{BH} in this equation are adjustable parameters which control the efficiency of gas fuelling towards the BH. In the context of `L-Galaxies`, this growth-mode is also referred-to as *quasar mode* due to the typically high accretion-rates exhibited during this growth channel. Nevertheless, ΔM_{acc} is instantaneously added to the central BH mass (i.e. in the same time sub-step of the galaxy merger), without a model for the radiative emission of L_{bol} due to the accretion process.

On the other hand, gas accretion from M_{hGas} is not related to mergers and can happen at any time during galaxies evolution. This growth-mode is called *radio mode* and it generally shows much lower accretion rates than the quasar-mode. Hot-gas accretion plays a crucial role in the modelled co-evolution of SMBHs and their hosts, being directly related to AGN feedback (see e.g. Croton et al., 2006). The mass accretion rate for the radio-mode growth is given by:

$$\dot{M}_\bullet = \frac{\Delta M_{acc}}{\delta t_{step}} = k_{AGN} \left(\frac{M_{hGas}}{10^{11}\, M_\odot}\right) \left(\frac{M_\bullet}{10^8\, M_\odot}\right) , \tag{2.13}$$

where k_{AGN} is a free parameter. We note that the normalization of $M_\bullet$ in the latter equation implies that this accretion mode is efficient for DM halos hosting massive central BHs. Finally, this channel of growth injects energy into the cold gas with a rate given by:

$$\dot{E}_{AGN} = \eta \dot{M}_\bullet c^2 , \tag{2.14}$$

where $\eta = 0.1$ and c is the speed of light. This energy injection effectively slows the cooling of gas in massive galaxies, reducing their ability to form stars. Finally, due to the strict correlation between the evolution of SMBHs and that of their host galaxy, `L-Galaxies` naturally produces the observed correlations between the properties of BHs and galaxies (Henriques et al., 2015).

The above prescriptions for BH growth and evolution were significantly modified and improved by Izquierdo-Villalba et al. (2020). In particular, the new prescriptions for SMBHs growth and evolution include:

- the treatment of time-delays between galaxy mergers and the merger of their central BHs,
- a model for the time-prolonged consumption of an accretion disk around central BHs, as opposed to the instantaneous accretion of ΔM_{acc} from cold gas,
- the possibility that cold-gas mass-accretion may happen also during disk instabilites, as an additional mass-growth channel to galaxy mergers,
- a complete treatment of the BH spin a and its evolution through mergers and gas accretion,
- the evolution of the mass-accretion efficiency η (see §1.3.3) as a function of BH spin,
- the emission of GWs as a consequence of BH-BH mergers,
- and finally the production of kick-velocities as a consequence of GW emission during BH-BH mergers, which may produce the ejection of the central SMBH from its host. The trajectory and evolution of these ejected BHs can then be tracked in time in order to study the population of *wandering* BHs around galaxies.

We refer to Izquierdo-Villalba et al. (2020) for the thorough description and analysis of these upgrades to `L-Galaxies`. Here we just concentrate on the newly-included BH-growth prescriptions, which allow to model the evolution of L_{bol} in time, hence to describe the gradual build-up of the SMBH population properties through a series of active and quiet evolutionary phases.

According to the prescriptions of Izquierdo-Villalba et al. (2020), both mergers and disk instabilities funnel a given amount of matter for the mass-growth of central BHs in galaxies. In particular, galaxy mergers provide:

$$\Delta M_{acc}^{merg}(z) = (1+z)^{\alpha_{BH}} \cdot \Delta M_{acc} , \tag{2.15}$$

where ΔM_{acc} is given by Eq. 2.12 while α_{BH} is a parameter controlling the redshift dependence of ΔM_{acc}^{merg}. Similarly, DIs provide an amount of matter to the central BH:

$$\Delta M_{acc}^{DIs}(z) = f_{BH}^{DIs} \cdot (1+z)^{\alpha_{BH}} \cdot \frac{\Delta M_{*}^{DIs}}{\left(1+\frac{V_{BH}}{V_{vir}}\right)^2}\,, \tag{2.16}$$

where ΔM_{*}^{DIs} is the amount of stellar mass transferred from the galactic disk to the bulge following a DI. In the above equation, f_{BH}^{DIs} is a normalization parameter which controls the amount of mass channeled to the central BH by DIs (see Izquierdo-Villalba et al., 2020, for details).

We underline that these equations only provide the *net amount* of mass to be transferred to the central BH (due to either mergers or DIs), rather than an *accretion rate* onto the BH. In other words, it is necessary to know at which rate BHs consume ΔM_{acc} in order to model their prolonged mass-growth in time. As shown by Eq. 1.50, at fixed mass-accretion and radiative efficiencies (respectively η and ϵ), the BH growth rate is controlled by the Eddington ratio f_{Edd}. Therefore, Izquierdo-Villalba et al. (2020) introduced a 2-phase model for f_{Edd}, following the work of e.g. Marulli et al. (2008). According to this prescription, each amount of mass funneled towards the central BH by any merger or DI is accumulated into an *accretion reservoir* M_{res} which fuels the actual mass-growth of BHs. As soon as some M_{res} mass is available, central BHs experience a first, rapid growth-phase which consumes a given fraction γ of M_{res} (where γ is an adjustable free-parameter set to 0.7, see Izquierdo-Villalba et al., 2020). This phase proceeds at the maximum possible rate, namely $f_{Edd} = 1$ (see §1.3.3) and it is followed by a quieter growth-phase in which f_{Edd} decays in time as:

$$f_{Edd} = \frac{1}{\left[1+\left(\frac{t}{t_Q}\right)^{1/2}\right]^{2/\beta}}\,, \tag{2.17}$$

where $t_Q = t_0\,\varepsilon^{\beta}/(\beta \ln 10)$, with $\varepsilon = 0.3$, $\beta = 0.4$ and $t_0 = 1.26 \times 10^8\,\mathrm{yr}$, as in Hopkins et al. (2006b). This phase mimicks a low-power AGN growth, as opposed to the first high-power QSO phase (see 1.1 for a graphic reference). Finally, the prescription of Izquierdo-Villalba et al. (2020) includes the modelling of an ADAF phase of growth (see §1.3.2), which is activated when f_{Edd} falls below the critical value $f_{Edd}^{crit} = 0.03$. When this happens the radiative efficiency ϵ is connected to f_{Edd} instead of being fixed to the mass accretion efficiency η, namely:

$$\epsilon = \eta \cdot \frac{f_{Edd}}{f_{Edd}^{crit}} \quad \text{if} \quad f_{Edd} < f_{Edd}^{crit} = 0.03\,, \tag{2.18}$$

as in Merloni & Heinz (2008) and Volonteri et al. (2013). These prescriptions for SMBHs growth were applied and calibrated on the `MR` merger trees by Izquierdo-Villalba et al. (2020), in order to reproduce the main observables of SMBH cosmological evolution. The results of the current work presented in chapter 3 are based on the same prescriptions, with the only exception of being applied to the `MR-II` merger trees. In chapter 5 we discuss the possible future modifications to this mass-growth model for SMBHs.

2.2 Astronomical observations: Photometry vs. Spectroscopy

This section is devoted to introduce the definitions and methods commonly used to measure the emission of astronomical sources. This sets the basis for the discussion of §4, where we address the observational study of the luminosity function of high-z QSOs with strong rest-frame UV emission. We first define astronomical quantities in §2.2.1 and the main current observational techniques in §2.2.2 and §2.2.3. We finally introduce the specific method on which our observational results are based in §2.2.5.

2.2.1 Definition of astronomical quantities

Here we introduce the main concepts and quantities connected to astronomical observations, in order to support our following discussion. The luminosity of observed sources (see §1.1.2 for its definition) is related to their emitted *flux*, i.e. the luminosity which crosses a unit of area at a given distance d from the source. Assuming spherical symmetry, the two quantities are connected by:

$$\mathrm{F} = \frac{\mathrm{L}}{4\pi\,\mathrm{d}^2}\,. \tag{2.19}$$

in `cgs` units, flux is measured as $\mathrm{erg\,cm^{-2}\,s^{-1}}$. As for the bolometric luminosity, also in this case the above definition includes all the electromagnetic frequencies. When referred only to a specific interval of frequency, these quantities are generally called *monochromatic*[2] or *flux density* or simply, *spectrum* of a source. Since wavelength λ and frequency ν of an electromagnetic wave are linked by $\nu = c/\lambda$, it is possible to define two equivalent flux densities, namely: f_ν or f_λ, respectively in ν-units ($\mathrm{erg\,cm^{-2}\,s^{-1}\,Hz^{-1}}$) and λ-units ($\mathrm{erg\,cm^{-2}\,s^{-1}\,\AA^{-1}}$). We note that in `cgs` units, frequencies are measured in Hertz (Hz) while wavelengths in Ångstroms (where $1\,\AA = 10^{-8}\,\mathrm{cm}$). These two units conventions are connected by:

$$\mathrm{f}_\nu\,\mathrm{d}\nu = \mathrm{f}_\lambda\,\mathrm{d}\lambda \quad ; \quad \frac{\mathrm{d}\nu}{\mathrm{d}\lambda} = -\frac{c}{\lambda^2}\,. \tag{2.20}$$

we stress that, in the following, all the flux-densities indicated by f_λ are measured in λ-units and they are a function of wavelength, i.e. $\mathrm{f}_\lambda = \mathrm{f}(\lambda)$. Capital F, on the other hand, will denote *wavelength-integrated* flux in units of $\mathrm{erg\,cm^{-2}\,s^{-1}}$.

Flux density allows to define the magnitude m of an source, i.e. a widely-used unit of measure for the flux of astronomical sources. Magnitudes are generally defined as a logarithmic[3] measure of f_ν and several different definitions have been proposed in the past. We make use of the common AB magnitudes (Oke, 1974), i.e.:

$$m_{\mathrm{AB}} = -2.5\,\mathrm{Log}\,(\mathrm{f}_\nu) + 48.6\,. \tag{2.21}$$

The last term in the above equation is the so-called *photometrical zero point* and it is a multiplicative factor (in flux units) that calibrates the photon counts received by a source into a measure of energy (see Oke, 1974).

Spectra of astronomical sources

The flux density f_λ of astronomical sources usually show *continuum* regions where the source emission appears smooth and without any peculiar feature. The sources continuum is generally due to global physical processes involving a significant fraction of the source baryonic matter, which then emits over a very wide wavelength range. Typical examples of these processes are: the thermal emission of stars, AGN accretion-disks (see §1.3.3), or bremsstrahlung of energetic electrons within AGN jets (see Rybicki & Lightman, 1986). On the other hand, some sources show *emission* (*absorption*) *lines* which usually manifest as prominent excesses (depressions) of f_λ within a well-defined wavelength range. These are generally due to the presence of a specific atomic species in an excited or ionized state, which emits (absorbs) photons at a characteristic frequency. For the sake of simplicity, when using the term *line* in the following discussion, we only refer to emission lines. It can be shown that lines are never intrinsically emitted at a single wavelength. Rather, several effects such as doppler broadening can "smear" them across a given wavelength interval. Their light distribution within this wavelength interval is usually called *line profile* and generally it shows a *peak* at a given wavelength λ_{peak}. Except in the most complex cases, lines can be well-approximated by either a single gaussian function or a combination of gaussian functions (see e.g. Calderone et al., 2017).

Spectral lines carry a huge amount of information, as they are directly related to source properties such as the abundance of specific atomic species, the temperature and ionization of the emitting medium

[2] Normally instead of using the *monochromatic* adjective, flux and luminosity referred to a particular frequency interval are named together with that interval, e.g.: *UV flux*, *X-ray luminosity* or *ionizing* flux.

[3] We follow the notation according to which Log X and ln X stand respectively for the base-10 and natural logarithm of the quantity X.

or its dynamic state. For this reason, several measurable quantities related to spectral lines have been defined in the past. The most immediate of them is the *integrated line flux* F_{EL}. This is the integral of the source spectrum across the wavelength range covered by the line profile, once the eventual contribution of the continuum has been removed from the same wavelength interval. In other words, F_{EL} is the integral of the line profile *above* the continuum level. Another fundamental observable of emission lines is their *full-width at half-maximum* (FWHM), which is simply the width of the line profile at the half of its maximum height with respect to the continuum level. Line widths can also be measured via the *equivalent width*, a concept which we define below to avoid confusion. Finally, sources emission is inevitably subject to cosmological redshift, so that the observed wavelength of any given spectral feature is modified as:

$$\lambda_{obs} = \lambda_{em} \cdot (1 + z) \tag{2.22}$$

where λ_{obs} and λ_{em} respectively are the observed and emitted wavelength. This implies that the whole spectrum of high-z sources is "stretched" by a factor $(1 + z)$.

The equivalent width of emission lines

In order to evaluate the relative importance of lines with respect to the continuum, it is common to use their *equivalent width* EW, which relates F_{EL} to the continuum flux-density evaluated at λ_{peak}. This quantity can be defined both for absorption and emission lines, as shown below in Eq. 2.23. Visually, the EW of a line can be thought as the width $\Delta\lambda$ which the line would cover if its profile was a flat, step-function with the same height as the continuum. Consequently, EWs are measured with the same units as wavelength intervals (i.e. Ångstroms in our case).

The EW is mathematically defined by the equation:

$$EW \equiv \int_{\lambda_{min}}^{\lambda_{max}} \left| 1 - \frac{f_\lambda}{f_\lambda^{cont}} \right| d\lambda = \int_{\lambda_{min}}^{\lambda_{max}} \left| 1 - \frac{f_\lambda^{cont} + f_\lambda^{EL}}{f_\lambda^{cont}} \right| d\lambda \,, \tag{2.23}$$

where λ_{min} and λ_{max} encompass the whole line profile. As anticipated, the absolute value in the above integral makes the definition of EWs identical for both absorption (i.e. when $f_\lambda < f_\lambda^{cont}$, within wavelength range covered by the line feature) and emission lines (i.e. $f_\lambda > f_\lambda^{cont}$).

The second equality holds because we consider the total spectrum of a source f_λ as composed by the sum of line and continuum contributions (respectively f_λ^{EL} and f_λ^{cont}). By further assuming that f_λ^{cont} is constant between λ_{min} and λ_{max} we have:

$$EW = \frac{1}{f_{\lambda_{EL}}^{cont}} \left(\int_{\lambda_{min}}^{\lambda_{max}} f_\lambda^{EL} \, d\lambda \right) = \frac{F^{EL}}{f_{\lambda_{peak}}^{cont}} \,, \tag{2.24}$$

where $f_{\lambda_{peak}}^{cont} = f_\lambda^{cont}(\lambda_{peak})$. Just as other spectral quantities related to wavelengths, also the EW is affected by cosmological redshift, so that:

$$EW_{obs} = EW_0 \cdot (1 + z) \,. \tag{2.25}$$

2.2.2 Spectroscopic observations

Sources spectra are obtained by splitting the light collected by a telescope into its different wavelengths contributions. This is achieved through either a diffracting (prism), grating or *grism* element (the latter being the combination of prism and a grating elements). Photons of different wavelengths are then measured by a photon-counting device, namely a CCD. The ability of a given instrument to separate close wavelengths is defined as *resolving power* $R = \lambda/\Delta\lambda$, where $\Delta\lambda$ is the minimum wavelength-difference distinguishable at a wavelength λ. Depending on R, spectroscopy allows to disentangle the various spectral features emitted by the observed source and to delineate the profile of its emission or absorption lines. These often include important physical information, such as doppler broadening (which is connected to the medium temperature and dynamical state), skewed-profiles (which may be indicative of bulk motions in the emitting matter source) or multiple-peaks (which might denote absorption features or specific

ionized states of the emitting material).

By splitting photons according to their wavelength, spectrographs also artificially dim the source luminosity, hence usually requiring very sensitive measuring devices or long observing time-intervals (or *exposure times* t_{exp}). For this reasons, spectroscopic observations are usually time-consuming and therefore need a preliminary, careful selection of the observations targets (e.g. Myers et al., 2015; Clerc et al., 2016; Dwelly et al., 2017; MacLeod et al., 2018). Spectroscopic observations are usually performed by isolating these targets and individually collecting their light, in order to avoid the mixing of spectral features of different objects within the same spectrum. Historically, this was obtained by using thin *slits* which selected only a specific object. Current techniques invlove the use of optical fibers which are placed at the exact position of the target on the optical plane, and therefore collect light exclusively from the desired object.

2.2.3 Photometric observations

Photometric measurements are usually performed through *filters* who *integrate* f_λ over specific wavelength intervals (or *pass-bands*), instead of splitting the photons of different wavelengths received by a source. Filters are defined by their *transmission curves*[4] $T^x_\lambda = T^x(\lambda)$, which describe the efficiency with which photons are transmitted through them, towards the measuring device. The latter are generally large arrays of photon-counting *pixels*, designed to measure entire areas of the sky. This method is significantly different than targeting specific sources, as in spectroscopy, and it is ideal to "blindly" look for sky sources down to a given *detection limit*.

By integrating the sources emission over a given *exposure-time* $t_{[exp]}$, photometric observations can significantly increase the possibility of detecting a source with respect to spectroscopic methods (at fixed t_{exp}). On the other hand, photometric observations tend to lose the details of f_λ. Indeed, the flux of a source measured in a given filter "x" is defined as the weighted *average* of f_λ in the pass-band, where the weight is the filter T^x_λ. For CCDs, this definition is given by (e.g., Tokunaga & Vacca, 2005):

$$\langle f^x_\lambda \rangle = \frac{\int f_\lambda \, T^x_\lambda \, \lambda \, d\lambda}{\int T^x_\lambda \, \lambda \, d\lambda} = \frac{\int (f^{cont}_\lambda + f^{EL}_\lambda) \, T^x_\lambda \, \lambda \, d\lambda}{\int T^x_\lambda \, \lambda \, d\lambda} \ , \tag{2.26}$$

where in the second equality we considered again the case in which f_λ is the superposition of continuum and line emission (as for the definition of EW). This confirms that, when measured through photometric filters, line emission features get mixed with the continuum within the single quantity $\langle f^x_\lambda \rangle$, namely (from eq. 2.26):

$$\langle f^x_\lambda \rangle = \frac{\int f^{cont}_\lambda \, T^x_\lambda \, \lambda \, d\lambda}{\int T^x_\lambda \, \lambda \, d\lambda} + \frac{\int f^{EL}_\lambda \, T^x_\lambda \, \lambda \, d\lambda}{\int T^x_\lambda \, \lambda \, d\lambda} = \langle f^{cont}_\lambda \rangle + \langle F_{EL} \rangle \ , \tag{2.27}$$

Therefore F_{EL} cannot be directly measured from photometry due to the convolution of f_λ with T^x_λ. To mitigate this issue, filters with a *narrow-band* (NB) can be used, so that the line-emission contribution to $\langle f^x_\lambda \rangle$ is largely dominant over the continuum one. This *narrow-band photmetry* technique is commonly-used to look for line features over specific wavelengths intervals and contrast them against the source continuum. On the other hand, line features are easily averaged-out by using *broad-band* (BB) filters, which then probe more efficiently the sources continuum. For this reason, combinations of BBs and NBs have been extensively used to blindly search for faint objects with specific emission (or absorption) lines (e.g. Venemans et al., 2005; Guaita et al., 2010; Sobral et al., 2016; Vilella-Rojo et al., 2021).

The pivot wavelength of photometric filters

The pivot wavelength of a filter (λ_p) is a measure of the average wavelength position of a photometric pass-band, similar to the central (i.e. average) wavelength of the pass-band. In particular, λ_p is the weighted average of all the wavelengths included within the filter pass-band, where the weight is the filter

[4] We define T_λ as the *measured* transmission curve of a filter, i.e. including the quantum efficiency of the measuring device, the atmospheric transmission and the effect of telescope optics.

transmission curve. Tokunaga & Vacca (2005) provided a mathematical definition of λ_p, namely:

$$\lambda_p^2 = \frac{\int \lambda \, T_\lambda^x \, d\lambda}{\int T_\lambda^x \, \frac{d\lambda}{\lambda}} \, . \tag{2.28}$$

2.2.4 Astronomical surveys

Systematic astronomical observations performed over determined sky areas and/or targeting specific classes of sources are usually referred-to as *surveys* (e.g. Scoville et al., 2007; Moles et al., 2008; Matsuda et al., 2011; Best et al., 2013; Le Fèvre et al., 2015; Cai et al., 2016; Matthee et al., 2017b; Ouchi et al., 2018; Yao et al., 2019; Cenarro et al., 2019). These can be either photometric, spectroscopic or offer a mixture of both approaches (e.g. Lawrence et al., 2007; Lilly et al., 2007; Warren et al., 2007; Geach et al., 2008; Sobral et al., 2009b; Adams et al., 2011; Matthee et al., 2014; Le Fèvre et al., 2015; Yan et al., 2016; Blanton et al., 2017; Abolfathi et al., 2018; Bonoli et al., 2020), targeting different parts of the electromagnetic spectrum, from gamma-rays down to radio wavelengths (e.g. Wright et al., 2010; Bundy et al., 2015; Boller et al., 2016; Gaia Collaboration et al., 2018). Several astronomical surveys have fostered ground-breaking advancements in the astronomical community. Some of the most influencing examples are the *Sloan Digital Sky Survey* (`SDSS`, e.g. Stoughton et al., 2002; Ahn et al., 2012; Dawson et al., 2013, 2016; Blanton et al., 2017; Abolfathi et al., 2018; Aguado et al., 2019), or the `COSMOS` survey (e.g. Scoville et al., 2007; Capak et al., 2007; Hasinger et al., 2007; Koekemoer et al., 2007; Lilly et al., 2007; Taniguchi et al., 2007). The first one is a wide-area, multi broad-band *and* spectroscopic optical survey which provided observations in five BB filters (u, g, r, i and z, see Gunn et al., 1998; Doi et al., 2010) over roughly $\sim 15000\,\mathrm{deg}^2$, and optical spectra (see Smee et al., 2013) for 3×10^7 objects (e.g. Ahumada et al., 2020). On the other hand, `COSMOS` is a multi-wavelength, narrow-area survey which provided observations at all wavelengths from X-ray to radio of a small $2\,\mathrm{deg}^2$ area of the sky, performed with several different instruments. Due to these properties, data from both surveys were vastly analyzed and taken as reference for a huge wealth of works focusing on various topics, from cosmology to galaxy formation and evolution.

Astronomical surveys provide the means to explore sky regions or target specific classes of sources with a consistent and coherent method, usually referred-to as *survey strategy*. The latter ensures that observations are carried out uniformly onto the selected sky-regions or targets, thus simplifying the statistical analysis of the collected data. Usually, spectroscopic surveys focus on a given number of pre-selected targets (e.g. Richards et al., 2002; Ross et al., 2012), while photometric surveys concentrate on observing a given sky region up to a given *depth*. The latter is defined on the basis of the *signal-to-noise ratio* (SNR, i.e. the ratio between the observed signal and its associated noise) of the collected data. In particular, especially in the case of photometric surveys, the measured SNR is a direct proxy of the ability to discern an observed source from the background data noise. Following this argument, the survey depth (or *detection limit*) is defined as the measured flux (or magnitude) at which the average SNR of the data reaches a fixed SNR-threshold. This implies that surveys planned to reach a given depth are expected to detect all the sources brighter than their detection limit. These kinds of observing programs are defined as *flux limited* surveys, while if they do not recur to target pre-selection surveys are also dubbed *blind* (see e.g. Stroe et al., 2017a).

Since bright sources are detected more efficiently than faint ones, it is reasonable to expect that the latter will be un-detected more easily within flux-limited survey data than bright ones. Therefore, bright sources might be "artificially" more numerous than faint ones due to this observational bias. This phenomenon is also known as *Malmquist bias* (see Malmquist, 1922, 1925; Butkevich et al., 2005), and it affects sources samples obtained through flux-limited surveys.

Detection limits and completeness

The actual SNR of observational data is influenced by a complex series of factors which range from the technical features of the measuring device, to the specific observing conditions under which data is collected, up to the details of the data-reduction procedures. These factors might influence the actual depth of an observed data-set, so that faint sources might remain un-detected within the survey data. It

is reasonable to imagine that this issue is larger for faint sources, i.e. those closer to the survey detection limit, rather than for bright, well resolved objects. The idea of accounting for the possibility of "missing sources" is formalized by the concept of *detection completeness*. The latter is measured as the ratio of the actually-measured number of objects with respect to their expected number:

$$\mathcal{C} = \frac{\#\ \text{of}\ \text{"true"}\ \text{detections}}{\#\ \text{of}\ \text{expected}\ \text{detections}}\,. \tag{2.29}$$

The detection completeness is usually defined as a function of the observed sources flux (or magnitude) and it is generally difficult to estimate, since the exact amount of expected detections is unknown in most of astronomical cases. One common way for assessing it relies onto the simulation of observing conditions and data-reduction procedures. Indeed, by producing mock data with a known amount of artificial sources, it is possible to measure the rate at which these are recovered as a function of their observed flux. Detection limits and completeness are two fundamental parameters which characterize the quality of the data collected by a given survey and guide the comparison between different surveys.

The Javalambre photometric surveys

The observational part of this thesis is based on the data of the *Javalambre Photometric Local Universe Survey* (`J-PLUS`, Cenarro et al., 2019). This is a wide-area, multi-band, optical, photometric survey, carried out in the Observatorio Astrofísico de Javalambre (OAJ, Cenarro et al., 2014). This new facility, dedicated to large-area photometric surveys, is located in central Spain, on the "Pico del Buitre" mountain (1956 m) within the "sierra de Javalambre" mountain complex of the province of Teruel.[5].

`J-PLUS` observations are being carried out by the T80Cam instrument on the JAST/T80 83cm telescope (Javalambre Auxiliary Telescope, Marin-Franch et al., 2015). T80Cam is a panoramic scientific camera equipped with a 9.2k-by-9.2k 10μm pixels, high-efficiency CCD which covers a significant fraction of the T80 telescope field of view (FoV) with a spatial resolution scale of 0.5"/pixel. This camera provides a wide effective FoV for scientific imaging (FoV $\sim 1.96\,\mathrm{deg}^2$, see Cenarro et al., 2019, for technical details). T80Cam is associated to a custom-designed photometric filter set, composed by a total of 12 pass-band filters. These are divided into five BBs, i.e. *u, g, r, i, z'*, with widths between $\sim$800 and $\sim$2000 Å, which replicate the SDSS photmetric filter set (Fukugita et al., 1996; Doi et al., 2010) with the exception of the custom-designed *u* band[6]. The seven NBs have widths of $\sim 150-500$ Å and are placed in key-positions to detect stellar absorption features, namely `[OII]` (λ3727.42) for *J*0378, `Ca` and `H+K` for *J*0395, `H`δ (λ4101.74) for *J*0410, the G-band for *J*0430, `Mg`*b* triplet for *J*0515, `H`α (λ6562.82) for *J*0660 and finally the `Ca` triplet for *J*0861. Figure 2.1 shows the transmission curves of `J-PLUS` filters, computed after accounting for the effect optical elements, CCD quantum efficiency and sky transparency. On the other hand, table 2.1 presents the filters properties together with their nominal depth in $\mathrm{mag_{AB}}$ and $\mathrm{f_\lambda}$ units.

Despite being designed for detecting local, stellar-absorption features, the `J-PLUS` NB filters can also target typically strong lines of extra-galactic spectra such as Lyα_{1216}, `SiIV` $_{1398}$, `CIV` $_{1549}$), `CIII]` $_{1909}$, `[OIII]` $_{2321}$, and `MgII]` $_{2799}$ within specific redshift windows. We stress that the actual ability of observing these extra-galactic features can be affected by the detection limits of `J-PLUS` bands (see table 2.1). In particular, the survey nominal depth is $r = 22$ at signal-to-noise ratio SNR = 3 (as measured in the *r* reference band, as better detailed in chapter §4), which is shallower than the corresponding limit of comparably wide-area optical photometric surveys such as SDSS ($\mathrm{r'} = 23.1$ at SNR = 5, see York et al., 2000). Nevertheless, this relative shortcoming is balanced by the possibility of observing an unprecedented sky-area with photometric NBs, making `J-PLUS` suitable for extensive searches of bright and rare emission-line galaxies (ELGs).

`J-PLUS` data are periodically made public through official data releases (DR). In particular, DR1 and DR2 were the first two releases, respectively made public in July 2018 and March 2020[7]. These consist of both imaging and source catalogs obtained over $\sim 1000\,\mathrm{deg}^2$ and $\sim 2000\,\mathrm{deg}^2$ respectively for DR1 and DR2. As better detailed in chapter 4, the results presented here are based exclusively on the DR1 data.

[5] OAJ website: https://oajweb.cefca.es/home/home
[6] The apex in the name of the *z'* broad-band only serves to differentiate it from the redshift *z*.
[7] `J-PLUS` DRs are publicly available at: http://www.j-plus.es/datareleases/

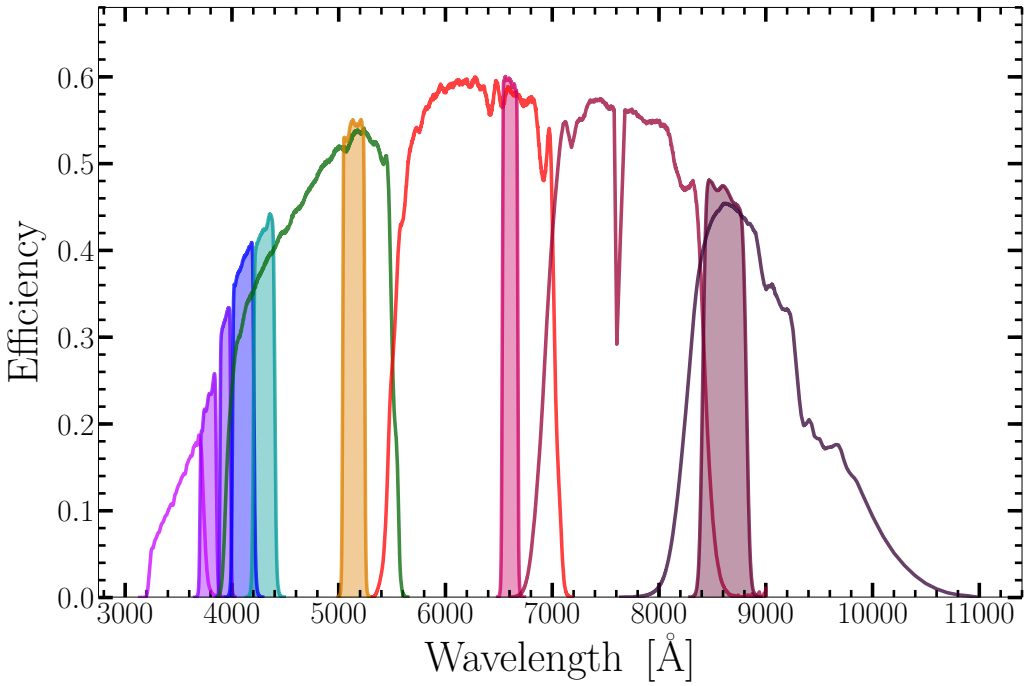

Figure 2.1: Transmission curves of the `J-PLUS` filter set. The 12 optical bands cover the entire optical band and can be divided into 7 narrow-bands (filled colored lines) and 5 broad-bands (colored empty lines). NBs are placed at key stellar features (see discussion in the text) in order to help photometric calibration of the data and pursue key scientific goals related to stellar physics (see e.g. López-Sanjuan et al., 2019a).

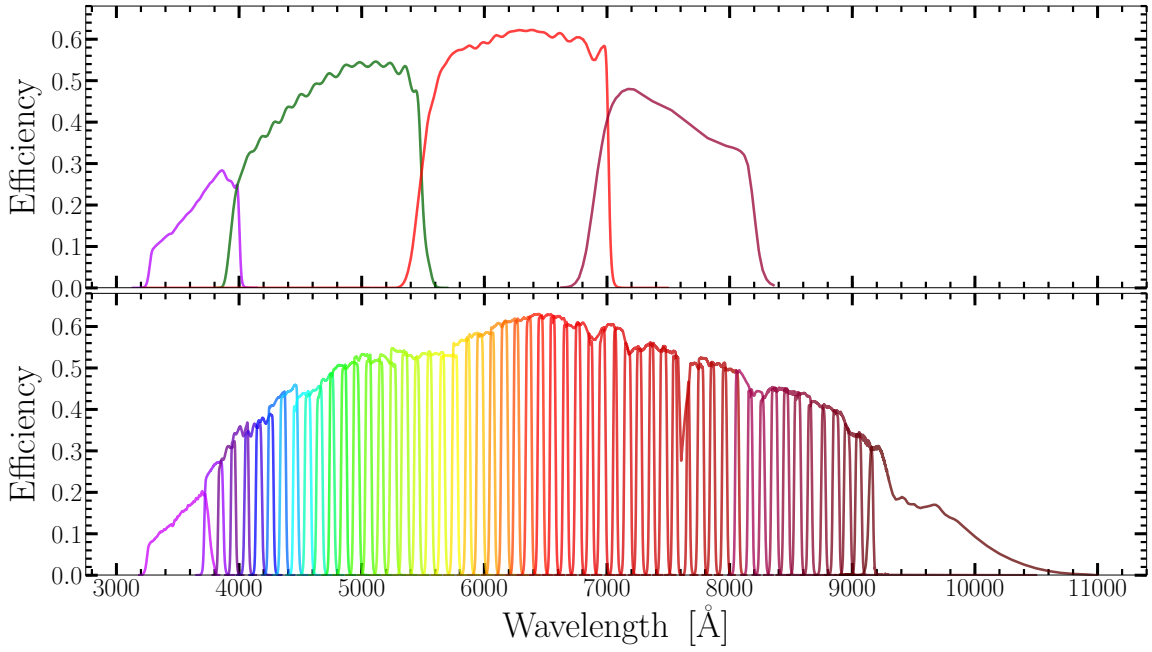

Figure 2.2: Transmission curves of the `J-PAS` filter set, divided in the four BBs (upper panel) and the 56 NBs (bottom panel). The latter span the entire optical bands and exhibit rougly constant widths (with the exception of the two extreme medium-bands at $\lambda\sim3500$ and $\lambda\sim9500$.

Filter	FWHM [Å]	$\mathrm{mag}_{\mathrm{AB}}^{\mathrm{min}}$ (3σ)	$\langle f_\lambda \rangle$ (3σ) [erg cm^{-2} s^{-1} Å^{-1}]
u	363.91	21.17	2.99×10^{-17}
$J0378$	152.74	21.18	2.56×10^{-17}
$J0395$	101.39	21.06	2.63×10^{-17}
$J0410$	201.76	21.28	1.99×10^{-17}
$J0430$	200.80	21.30	1.78×10^{-17}
g	1481.92	22.09	7.05×10^{-18}
$J0515$	207.19	21.35	1.18×10^{-17}
r	1500.20	22.02	4.36×10^{-18}
$J0660$	146.13	21.34	7.27×10^{-18}
i	1483.59	21.54	4.47×10^{-18}
$J0861$	410.50	20.67	7.94×10^{-18}
z	1055.93	20.80	6.54×10^{-18}

Table 2.1: Tabulated FWHMs and 3σ detection limits of J-PLUS filters. For the sake of simplicity, we generally refer to the filters $J0378$, $J0395$, $J0410$, $J0430$, $J0515$, $J0660$ and $J0861$ as J-PLUS NBs, even though some of these filters (e.g. $J0861$) could be defined as *medium bands*.

Despite having provided a big amount of high-quality scientific data, J-PLUS was primarily conceived as a preliminary survey for the main observing program of the OAJ, namely the *Javalambre Physics of the Accelerating Universe Astrophysical Survey* (J-PAS[8]). Indeed, the multi-narrowband nature of J-PLUS was designed to help and guide the calibration of the J-PAS survey data. The latter is designed to perform multi-NB observations over the whole optical band, uniformly covering it with a set of 56 NBs (the majority of which has a roughly constant width of ~ 150 Å), supported by four broad bands (i.e. u, g, r and i, as shown in Fig. 2.2).

The J-PAS survey has recently undergone a preliminary verification phase, where a first round of scientific data has been collected with a pathfinder camera. This phase was conceived to verify the operational behavior of the telescope and optical systems, as well as to practically assess the potential scientific outcome of J-PAS (see Bonoli et al., 2020). The main survey is expected to start in the next future and it will be carried out by the JST/T250 2.5 meters Javalambre Survey Telescope, mounting the JPcam scientific camera (Taylor et al., 2014). The latter is a mosaic camera composed of 14 9.2k-by-9.2k 10μm pixels CCDs, which replicate the J-PLUS ones and cover the extremely large FoV of JST/T250 (namely: $\sim 4\,\mathrm{deg}^2$). Due to the different properties of the J-PAS optical system with respect to the J-PLUS one, JPcam has a better spatial resolution than T80cam, reaching 0.2267"/pixel (to be compared with e.g. the 0.396"/pixel of SDSS, Gunn et al., 1998).

The observing strategy of J-PAS is designed to reach $\sim 8000\,\mathrm{deg}^2$ at a nominal depth of $m_r \sim 23$ for the r reference band and $22 \lesssim \mathrm{mag} \lesssim 22.5$ for narrow-bands (see e.g., Benitez et al., 2014; Bonoli et al., 2020). Nevertheless, the key point of J-PAS is the coverage of the whole surveyed area with each of its 56 NBs. In other words, the emission of each source (or even of each "empty" patch in the sky) observed by J-PAS will be sampled at 56 different points within the optical band, therefore obtaining the analogue of a very-low resolution spectrum for each pixel of the observed FoV. In this sense, J-PAS can be thought as a mid-point between photometric, spectroscopic and IFU approaches, hence potentially providing significant contributions to a wide variety of scientific research fields (see Bonoli et al., 2020, for a detailed discussion). Just to give some examples, J-PAS can efficiently identify the nebular emission of galaxies up to $z \sim 1$, when considering its expected detection limits, or the bright emission-lines of QSOs up to $z \sim 6$ (as recently demonstrated by e.g., Chaves-Montero et al., 2021; Martínez-Solaeche et al., 2021). Given the high number of data-points extracted from the optical spectra of sources, J-PAS data are ideal for SED-fitting procedures, hence allowing to effectively study the properties of stellar populations within galaxies up to $z \sim 1$ (e.g., González Delgado et al., 2021). Furthermore, in the case

[8] J-PAS website: http://www.j-pas.org/

of spatially-resolved galaxies, this analysis can be performed in different pixels (or *spaxels*, i.e. groups of adjacent pixels), therefore allowing for 2D studies of galaxy properties. Finally, J-PAS is expected to provide accurate photometric measurements of redshift (i.e. *photo-z*) up to $z \sim 1$, therefore aiding to the measurement of the BAO scale, the mapping of the large-scale structure and the identification of large samples of galaxy clusters (as discussed in Bonoli et al., 2020).

In the context of this work, the J-PAS survey offers extremely interesting applications, since it is expected to trace the emission of AGN and bright line-emitters up to $z \sim 6$. In addition, its wide-area and relatively good depth ($r^{\rm limit} \sim 24$ and $\mathrm{m}_{\rm NB}^{\rm limit} \sim 23$, see Bonoli et al., 2020) allow to look for high-z, bright and rare sources, while the multiple narrow-bands can provide an efficient mean for distinguishing broad and narrow line features (see §4 for more details about this point). This can be used to study the roles and contributions of both SF galaxies and AGN during cosmic evolution. On one hand, this might provide a natural setting to test the model predictions about formation and evolution of SMBHs discussed in chapter §3. On the other hand, J-PAS properties offer a natural context to expand the results and methods detailed in §4. We discuss these possible applications and future perspectives in much greater detail in our discussions (§5).

2.2.5 Photometric characterization of emission lines: the three-filters method

In the last decades, a considerable effort has been put towards the development of methods to measure emission lines with the only use of photometric bands (see e.g. Venemans et al., 2005; Pascual et al., 2007). These methods are generally approximations of spectroscopic approaches and rely on the combination of multi-wavelength photometric estimates of sources spectra. Generally speaking, the flux of an emission line could be measured with a single NB measurement if the relative contribution of line and continuum emission to the total NB flux, and the source redshift were known. In other words, by knowing the line EW and $\lambda_{\rm peak}$ in the filter band-width. Unfortunately, these quantities are not provided by a single NB measurement without further hypothesis.

In particular, line and continuum emission are degenerate in the single NB measurement, while the exact position of the line peak within the filter bandwith (i.e. the source redshift) is washed-out by the integration of photons over the whole bandwidth. Consequently, it is necessary to make some working assumptions in order to extract spectral information from photometric measurements. Interestingly, an efficient solution to this issue has been developed by Vilella-Rojo et al. (2015) and tested by Logroño-García et al. (2019). Their *three-filters* method (3FM) is designed to extract the flux of a given emission line ($\mathrm{F_{EL}}$) by making use of a set of three filters: one NB and two BB. This parallels well-established methodologies which exploit BB photometry to estimate the sources continuum and then contrast the eventual line emission with a NB measurement (see e.g., Gronwall et al., 2007; Guaita et al., 2010).

Equations of the three-filters method

Here we derive the main equations detailed in Vilella-Rojo et al. (2015), to which we refer for additional details. We stress that the 3FM is designed to extract $\mathrm{F}^{\rm EL}$ by using three photometric measurements (two BBs and one NB). Its theoretical basis are two main hypothesis, i) the emission-line profile can be approximated by a Dirac-delta, and ii) the source continuum is well-traced by a linear function of wavelength over the whole interval covered by the three filters:

$$\mathrm{f}_\lambda^{\rm EL} = \mathrm{F}^{\rm EL} \cdot \delta(\lambda - \lambda_{\rm EL}) \ , \tag{2.30}$$

$$\mathrm{f}_\lambda^{\rm cont} = \mathrm{A}\,\lambda + \mathrm{B} \ , \tag{2.31}$$

where $\delta(\lambda - \lambda_{\rm EL})$ is centered at $\lambda_{\rm EL}$, while A and B are two scalar coefficients. Equation 2.30 implicitly assumes that $\mathrm{F}^{\rm EL}$ is *entirely* included within the NB pass-band. This might be false when part of the emission-line profile lies outside the NB pass-band, for instance: when the line-profile is wider than the NB pass-band, or when $\lambda_{\rm peak}$ lies close to the NB pass-band edge. We further discuss this bias in §4.6.1 and §5.2.3.

By using eq. 2.31 into eq. 2.26 we get:

$$\begin{aligned}\langle \mathrm{f}_\lambda^{\mathrm{x}} \rangle &= \frac{\int (\mathrm{A}\,\lambda + \mathrm{B} + \mathrm{f}_\lambda^{\mathrm{EL}}) \cdot \mathrm{T}_\lambda^{\mathrm{x}}\,\lambda\,\mathrm{d}\lambda}{\int \mathrm{T}_\lambda^{\mathrm{x}}\,\lambda\,\mathrm{d}\lambda} \\ &= \frac{\left[\mathrm{A}\int \lambda^2\,\mathrm{T}_\lambda^{\mathrm{x}}\,\mathrm{d}\lambda \;+\; \mathrm{B}\int \mathrm{T}_\lambda^{\mathrm{x}}\,\lambda\,\mathrm{d}\lambda \;+\; \int \mathrm{f}_\lambda^{\mathrm{EL}}\,\mathrm{T}_\lambda^{\mathrm{x}}\,\lambda\,\mathrm{d}\lambda\right]}{\int \mathrm{T}_\lambda^{\mathrm{x}}\,\lambda\,\mathrm{d}\lambda} \\ &= \frac{\left[\mathrm{A}\int \lambda^2\,\mathrm{T}_\lambda^{\mathrm{x}}\,\mathrm{d}\lambda \;+\; \mathrm{B}\int \mathrm{T}_\lambda^{\mathrm{x}}\,\lambda\,\mathrm{d}\lambda \;+\; \mathrm{F}^{\mathrm{EL}}\,\mathrm{T}_{\lambda_{\mathrm{EL}}}^{\mathrm{x}}\,\lambda_{\mathrm{EL}}\right]}{\int \mathrm{T}_\lambda^{\mathrm{x}}\,\lambda\,\mathrm{d}\lambda}\,,\end{aligned} \tag{2.32}$$

where $\mathrm{T}_{\lambda_{\mathrm{EL}}}^{\mathrm{x}} = \mathrm{T}_\lambda^{\mathrm{x}}(\lambda_{\mathrm{EL}})$, while the last step makes use of Eq. 2.30 in the last term at the numerator and the properties of the Dirac-delta. To simplify the notation we introduce:

$$\alpha_{\mathrm{x}} = \frac{\int \lambda^2\,\mathrm{T}_\lambda^{\mathrm{x}}\,\mathrm{d}\lambda}{\int \mathrm{T}_\lambda^{\mathrm{x}}\,\lambda\,\mathrm{d}\lambda} \quad ; \qquad \beta_{\mathrm{x}} = \frac{\mathrm{T}_{\lambda_{\mathrm{EL}}}^{\mathrm{x}}\,\lambda_{\mathrm{EL}}}{\int \mathrm{T}_\lambda^{\mathrm{x}}\,\lambda\,\mathrm{d}\lambda}\,, \tag{2.33}$$

which only depend on $\mathrm{T}_\lambda^{\mathrm{x}}$ and λ_{EL}. The latter is determined by each source redshift and cannot be measured without a spectroscopic observation, hence λ_{EL} must be assumed a-priori. A reasonable choice for this value is either the central wavelength of a filer or its λ_{p}. We can now re-write 2.32 using 2.33:

$$\langle \mathrm{f}_\lambda^{\mathrm{x}} \rangle = \mathrm{A}\cdot\alpha_{\mathrm{x}} \;+\; \mathrm{B} \;+\; \mathrm{F}^{\mathrm{EL}}\cdot\beta_{\mathrm{x}}\,, \tag{2.34}$$

which is valid for a generic filter. We note that if the targeted emission-line lies outside the pass-band, then $\mathrm{T}_{\lambda_{\mathrm{EL}}}^{\mathrm{x}} = 0$ implying $\beta_{\mathrm{x}} = 0$. To determine A, B and F^{EL}, we apply 2.34 to a set of three filters: a NB, a *line-contaminated* BB (denoted here by LC) and a *line-uncontaminated* BB (denoted by LU):

$$\langle \mathrm{f}_\lambda^{\mathrm{NB}} \rangle = \mathrm{A}\cdot\alpha_{\mathrm{NB}} \;+\; \mathrm{B} \;+\; \mathrm{F}^{\mathrm{EL}}\cdot\beta_{\mathrm{NB}}\,, \tag{2.35}$$

$$\langle \mathrm{f}_\lambda^{\mathrm{LC}} \rangle = \mathrm{A}\cdot\alpha_{\mathrm{LC}} \;+\; \mathrm{B} \;+\; \mathrm{F}^{\mathrm{EL}}\cdot\beta_{\mathrm{LC}}\,, \tag{2.36}$$

$$\langle \mathrm{f}_\lambda^{\mathrm{LU}} \rangle = \mathrm{A}\cdot\alpha_{\mathrm{LU}} \;+\; \mathrm{B}. \tag{2.37}$$

By solving this linear system we finally obtain F^{EL}, A and B:

$$\mathrm{F}^{\mathrm{EL}} = \frac{\langle \mathrm{f}_\lambda^{\mathrm{LC}} \rangle \;-\; \langle \mathrm{f}_\lambda^{\mathrm{LU}} \rangle \;+\; \frac{\alpha_{\mathrm{LU}}-\alpha_{\mathrm{LC}}}{\alpha_{\mathrm{NB}}-\alpha_{\mathrm{LU}}}\cdot\left[\langle \mathrm{f}_\lambda^{\mathrm{NB}} \rangle - \langle \mathrm{f}_\lambda^{\mathrm{LU}} \rangle\right]}{\beta_{\mathrm{LC}} \;+\; \frac{\alpha_{\mathrm{LU}}-\alpha_{\mathrm{LC}}}{\alpha_{\mathrm{NB}}-\alpha_{\mathrm{LU}}}\cdot\beta_{\mathrm{NB}}}\,, \tag{2.38}$$

$$\mathrm{A} = \frac{\langle \mathrm{f}_\lambda^{\mathrm{NB}} \rangle \;-\; \langle \mathrm{f}_\lambda^{\mathrm{LU}} \rangle \;-\; \frac{\beta_{\mathrm{NB}}}{\beta_{\mathrm{LC}}}\cdot\left[\langle \mathrm{f}_\lambda^{\mathrm{LC}} \rangle - \langle \mathrm{f}_\lambda^{\mathrm{LU}} \rangle\right]}{\alpha_{\mathrm{NB}} \;-\; \alpha_{\mathrm{LU}} \;-\; \frac{\beta_{\mathrm{NB}}}{\beta_{\mathrm{LC}}}\cdot(\alpha_{\mathrm{LC}} \;-\; \alpha_{\mathrm{LU}})}\,, \tag{2.39}$$

$$\mathrm{B} = \langle \mathrm{f}_\lambda^{\mathrm{LU}} \rangle - \alpha_{\mathrm{LU}}\cdot\mathrm{A}\,. \tag{2.40}$$

The coefficients A and B can be used to evaluate 2.31 at λ_{p} of the NB and get an estimate of the line-uncontaminated linear continuum in the NB $\left\langle \mathrm{f}_{\lambda\,;\,\mathrm{cont}}^{\mathrm{NB}} \right\rangle$. By introducing the coefficients

$$\alpha_{\mathrm{x}} = \frac{\int \lambda^2\,\mathrm{T}_\lambda^{\mathrm{x}}\,\mathrm{d}\lambda}{\int \mathrm{T}_\lambda^{\mathrm{x}}\,\lambda\,\mathrm{d}\lambda} \quad ; \qquad \beta_{\mathrm{x}} = \frac{\mathrm{T}_\lambda^{\mathrm{x}}(\lambda_{\mathrm{EL}})\;\lambda_{\mathrm{EL}}}{\int \mathrm{T}_\lambda^{\mathrm{x}}\,\lambda\,\mathrm{d}\lambda}\,, \tag{2.41}$$

which only depend on the transmission curve of a given filter "x" (i.e. $\mathrm{T}_\lambda^{\mathrm{x}}$) and on λ_{EL} (i.e. the wavelength of the line-peak in the NB), the method of Vilella-Rojo et al. (2015) can directly estimate F_{EL} as:

$$\mathrm{F}_{\mathrm{EL}}^{\mathrm{3FM}} = \frac{(\mathrm{f}_\lambda^{g} - \mathrm{f}_\lambda^{r}) + \frac{\alpha_r - \alpha_g}{\alpha_{\mathrm{NB}} - \alpha_r}\cdot(\mathrm{f}_{\mathrm{NB}} - \mathrm{f}_\lambda^{r})}{\beta_g + \frac{\alpha_r - \alpha_g}{\alpha_{\mathrm{NB}} - \alpha_r}\cdot\beta_{\mathrm{NB}}}\,. \tag{2.42}$$

As already mentioned, $\mathrm{F}_{\mathrm{EL}}^{\mathrm{3FM}}$ can be biased in case a fraction of the line profile that this methods targets lies outside the NB pass-band. In other words, the part of F_{EL} which lies outside the NB wavelength coverage is simply undetected by the NB photometric measurement. The importance of this bias on F_{EL} depends on both the line profile details (e.g. its width or eventual tails) and the position of its peak

within the NB. In turn, these are determined by a number of complex aspects for different astronomical sources, i.e.: the accretion status for high-z QSOs (e.g., Calhau et al., 2020), the transfer of Lyα photons in the hydrogen-rich ISM and IGM (e.g., Dijkstra, 2017; Gurung-Lopez et al., 2018; Gurung-López et al., 2019, 2020) or the sources metals and dust content (e.g., Christensen et al., 2012). These details can be extracted by high-resolution spectroscopic data but, unfortunately, not from photometry, due to its intrinsic loss of spectral information (see §2.2.3). Therefore the eventual bias on F_{EL}^{3FM} can only be estimated indirectly. We further discuss this topic in §4.6.1 and §5.2.3.

Chapter 3

Black-Hole seeding in the L-Galaxies semi-analytical model

This chapter presents the theoretical side of this thesis, which is based on the implementation of physical recipes for the formation of SMBHs in the L-Galaxies SAM (see §2.1.2). The BH-seeding scenarios detailed in §1.2.2 are embedded into a cosmological context, allowing to explore their effects onto the build-up of SMBHs populations throughout cosmic history. This approach allows to both track high-z SMBH formation within its cosmological context through physically-motivated prescriptions, and to obtain sufficient statistics to study the occurrence of these processes over a wide simulated volume. Furthermore, thanks to the cosmological nature of the L-Galaxies inputs, the evolution of the SMBHs population resulting from high-z seeding processes can be followed down to $z=0$ and analyzed statistically.

This chapter is organized as follows: section 3.1 introduces the rationale of our theoretical work and contextualizes our methods within the current panorama. Section 3.2 presents the modifications we introduced to the version of L-Galaxies described in §2.1.2. Section 3.3 details our physical model for BH-seeding and it explains how we complement L-Galaxies with the outputs of the GQd model (Valiante et al., 2011, 2016, 2021). Section 3.4 details our results and finally section 3.5 presents a discussion of the results of the chapter and its conclusions.

3.1 Introduction

As introduced in §1, the formation of $M_\bullet \gtrsim 10^9\,M_\odot$ SMBHs in just $\sim$1 Gyr from the Big Bang represents one of the most intriguing open questions of modern astrophysics (see e.g., Latif & Ferrara, 2016; Valiante et al., 2017; Inayoshi et al., 2020, for recent reviews). Although several possible pathways have been proposed (see §1.2.2), current theoretical models still struggle to identify the processes by which SMBH seeds formed and grew within such short timescales. For instance, the formation of PopIII stars eventually leads to the formation of light seeds at $z \gtrsim 20$ (e.g., Schneider et al., 2002; Yoshida et al., 2003; Bromm & Larson, 2004; Yoshida et al., 2007). Nevertheless, these would require stringent conditions on their growth-rates to promptly evolve into SMBHs (e.g., Pacucci et al., 2015; Pezzulli et al., 2017; Regan et al., 2019). On the other side, intermediate-mass or heavy seeding scenarios avoid strict growth-rate constraints by envisioning a more massive origin of SMBHs but generally require peculiar physical conditions (e.g., Valiante et al., 2016; Sassano et al., 2021), whose occurrence is yet poorly constrained over cosmological contexts (e.g., Bromm & Loeb, 2003; Omukai et al., 2008; Latif et al., 2015; Agarwal et al., 2016; Latif et al., 2018).

This broad issue has been widely addressed through a multitude of numerical approaches, including hydrodynamical simulations (e.g., Agarwal et al., 2012; Chon et al., 2016; Habouzit et al., 2016; Di Matteo et al., 2017; Habouzit et al., 2017; Ardaneh et al., 2018; DeGraf & Sijacki, 2019; Habouzit et al., 2019) as well as semi-analytic and purely-analytic models (e.g., Devecchi & Volonteri, 2009; Valiante et al., 2016; Pezzulli et al., 2017; Lupi et al., 2021a; Sassano et al., 2021). Hydrodynamical simulations can generally model directly the small-scale processes involved in the formation of BH seeds (see §1.2.1, §2.1

and e.g., Yoshida et al., 2007; Latif et al., 2013; Agarwal et al., 2014; Latif et al., 2016; Chon et al., 2018; Chon & Omukai, 2020). Nevertheless, this limits the volume sizes which can be modelled with current computational capabilities, hence hindering the extrapolation of hydrodynamical results to cosmological volumes. On the other hand, semi-analytic and analytic codes are computationally affordable over very wide cosmological boxes or high mass-resolutions (e.g., Valiante et al., 2016; Cora et al., 2018; Lupi et al., 2021a), despite offering a more limited physical modelling. Consequently, SAMs can help to study how BH-seeding processes relate to the formation of the global SMBH population observed at $z \lesssim 4$, a yet poorly constrained aspect of SMBHs evolution.

Numerical models which track the cosmological build-up of SMBHs generally adopt strong assumptions on their formation scenarios. For instance, a common approach is to assume that BH seeds form with a fixed-mass, throughout the whole history of the Universe (i.e. eventually down to $z \sim 0$) and independently of local metallicity or UV illumination levels (e.g., Dubois et al., 2014a; Schaye et al., 2015; Sijacki et al., 2015; Bonoli et al., 2016; Izquierdo-Villalba et al., 2020). Despite these models are generally able to reconcile the properties of simulated SMBH populations with $z \lesssim 4$ observations (e.g., Volonteri et al., 2013; Dubois et al., 2014b; Habouzit et al., 2017; Volonteri et al., 2020), their simplified description of BH-seeding precludes to study the actual origin of SMBHs. On the other hand, physical models which focus onto the small-scale BH-seeding physics can successfully reproduce the observable properties of high-z QSOs (e.g., Agarwal et al., 2012; Valiante et al., 2014, 2016; Lupi et al., 2019, 2021a; Sassano et al., 2021), but their conclusions are difficult to generalize to cosmological boxes.

As a compromise between these two approaches, our work aims at embedding physically-motivated prescriptions for modelling BH-seeding over a cosmological volume, therefore addressing a "missing link" between the small-scale physics of BH formation and the evolution of large SMBHs populations. As introduced in §2.1.2, we make use of the L-Galaxies semi-analytic code (see e.g., Guo et al., 2011; Henriques et al., 2015). This choice is primarily motivated by the possibility of exploiting the N-body merger trees of the Millennium-II simulation (MR-II, Boylan-Kolchin et al., 2009), which offers an optimal compromise between halo-mass resolution ($M_{vir}^{min} \sim 10^7\,M_\odot\,h^{-1}$) and simulated volume ($L_{box} = 100\,Mpc\,h^{-1}$). Furthermore, differently from purely-analytic codes, L-Galaxies can access to the 3D spatial distribution of structures within its input merger-trees. This allows to track self-consistently the high-z environments where SMBHs form. On the other hand, despite still offering a high mass-resolution, MR-II cannot access to the mass scales of mini-halos hosting PopIII stars formation (see §1.2.1). Therefore, the formation of light seeds cannot be tracked self consistently in L-Galaxies. In order to overcome this limitation, we exploit the results of the GQd analytic model (Valiante et al., 2011, 2014, 2016, 2021) where PopIII evolution and the formation of light seeds is followed self-consistently. In particular, we implement the formation of PopIII remnants in a sub-grid fashion within L-Galaxies, while self-consistently tracking the formation of massive BH seeds. In addition, as detailed in §3.2, the Henriques et al. (2015) version of L-Galaxies was not designed to follow the gradual chemical enrichment of the IGM, nor the diffusion of Lyman-Werner photons in the IGM. As introduced in §1.2, both elements are crucial for any physically-motivated model of SMBH formation. Therefore, we also added a set of prescriptions for treating both phenomena in the L-Galaxies code. In particular, we added a model to trace the average backgrounds of metallicity and LW flux, as well as their spatial variations. With this method, we are able to account at the same time for most of the currently-envisioned seeding processes (namely: light, intermediate and heavy scenarios, see 1.2.2) and embed this comprehensive model for high-z BH formation in a cosmological framework.

We underline that Sassano et al. (2021) recently presented an improved version of GQd where the formation of light, intermediate and heavy seeds is followed self-consistently. We do not use the outputs of this latter GQd version since the combination of MR-II and L-Galaxies allows to track the occurrence of intermediate and heavy seeding scenarios within their cosmological context. Furthermore, our approach is similar to the one recently developed by Lupi et al. (2021a). Nevertheless, this work focused on the role of DCBHs within the progenitors of $z \sim 6$ QSOs, without including the formation of IMBH seeds. Our method lies in between of these two recent works, as it aims to explore the role of all currently-envisioned baryonic channels in the build-up of the global SMBH population, within a cosmological context.

3.2 Tracing the diffusion of metals and UV photons in L-Galaxies

The versions of L-Galaxies presented in Henriques et al. (2015), Izquierdo-Villalba et al. (2019b) and Izquierdo-Villalba et al. (2020) (see §2.1.2) were specifically designed to reproduce the main observables of galaxies and SMBHs populations at $z \lesssim 4$. Therefore, we introduced further features and modifications to these L-Galaxies versions, in order to account for the implementation of BH-seeding processes. These include a model for the diffusion of metals and UV photons in the IGM, which is detailed in the current section. On the other hand, the set of self-consistent physical recipes producing high-z BHs and the prescription to account for the un-resolved formation of PopIII remnants (exploiting the interplay with GQd) constitute the core of our BH-seeding model, therefore we address both of them in §3.3.

As introduced in §1.2, the early radiative and chemical feedback of SF events dramatically impacts the production of BH seeds (see also e.g., Volonteri, 2010; Latif & Ferrara, 2016; Valiante et al., 2017; Inayoshi et al., 2020). Therefore, in order to base our BH-seeding model on physically-motivated and self-consistent local conditions for metallicity and UV background, we introduce the treatment of $\langle \mathrm{Z_{IGM}} \rangle$ and $\mathrm{J_{LW\,;\,bg}}$ (respectively, the average IGM metallicity and the LW background), as well as a model for their spatial variations produced by early SF episodes. Generally speaking, these variations arise from the diffusion of metals and the radiative transfer of photons within the IGM (see e.g., Greif et al., 2009a; Maio et al., 2010; Habouzit et al., 2016; Dale, 2015; Habouzit et al., 2017; Kim et al., 2017b). Nevertheless, we resort to a simpler analytic treatment in order to allow for the fast execution of our model over the MR-II cosmological box. More in detail, we use the Henriques et al. (2015) version of L-Galaxies to compute uniform, averaged backgrounds of $\mathrm{Z_{IGM}}$ and $\mathrm{J_{LW}}$ (§3.2.1), on top of which we model local variations (§3.2.2 and §3.2.3).

3.2.1 Uniform $\mathrm{Z_{IGM}}$ and $\mathrm{J_{LW}}$ backgrounds

As discussed in §2.1.2, the standard version of L-Galaxies does not model the gradual chemical enrichment of the IGM nor the metallicity of the primordial infall. The latter can significantly influence the progress of BH-seeding at high-z therefore, for each newly-initialized structure at $z_{\rm init}$, we link it to the average IGM metallicity: $\langle \mathrm{Z_{IGM}} \rangle (z_{\rm init})$. We obtain $\langle \mathrm{Z_{IGM}} \rangle (z)$ as the ratio between the metals mass ejected into the IGM by all halos in the MR-II box, and the total amount of baryonic mass not yet collapsed into halos, i.e.: $\mathrm{M_{gas}^{IGM}}(z)$. Namely:

$$\langle \mathrm{Z_{IGM}} \rangle (z) = \frac{\sum_i \mathrm{M_{ej}^{met\,;\,\mathit{i}}}(z)}{\mathrm{M_{gas}^{IGM}}(z)} = \frac{\sum_i \mathrm{M_{ej}^{met\,;\,\mathit{i}}}(z)}{\left[\mathrm{M_{box}^{DM}} - \sum_i \mathrm{M_{vir}^{\mathit{i}}}(z)\right] \cdot \mathrm{f_b}} \,, \tag{3.1}$$

where the summations extend over all the structures in the MR-II box. In particular, $\mathrm{M_{ej}^{met\,;\,\mathit{i}}}(z)$ is the mass of metals in the $\mathrm{M_{ej}}$ reservoir of each galaxy i (see §2.1.2), at a given z. $\mathrm{M_{gas}^{IGM}}(z)$ is the total mass of diffuse gas in the IGM, obtained by multiplying the baryon fraction $\mathrm{f_b}$ (see §2.1.2) for the difference between the total DM mass[1] in the whole MR-II box and the amount of DM collapsed into halos (respectively $\mathrm{M_{box}^{DM}}$ and $\sum_i \mathrm{M_{vir}^{\mathit{i}}}$). Figure 3.1 (middle panel) compares the evolution of $\langle \mathrm{Z_{IGM}} \rangle$ in L-Galaxies and GQd. We note that the latter shows a significantly higher level of IGM chemical enrichment. This is due to both its higher mass-resolution and specific implementation of SNe feedback. Furthermore, the GQd run we use is built to reproduce the merger history of a DM halo reaching $\mathrm{M_{vir}} = 10^{13} \mathrm{M_\odot}$ at $z = 2$ (see Valiante et al., 2021), therefore modelling a relatively biased region of the Universe. This produces significantly different IGM properties with respect to those computed over the MR-II volume.

Also UV illumination plays a crucial role in BH-seeding processes (§1.2.1), therefore a model for the diffusion of UV photo-dissociating photons must be added to the Henriques et al. (2015) version of L-Galaxies. We compute a uniform background $\mathrm{J_{LW\,;\,bg}}$ over the MR-II box, to account for the large mean free-paths of LW photons at high-z (up to ~ 100 Mpc, see Ahn et al., 2009). By following Agarwal et al. (2012):

$$\mathrm{J_{LW\,;\,bg}} = \mathrm{f_{esc}} \frac{h\,c}{4\pi \mathrm{m_H}} \eta\, \rho_* \,(1+z)^3 \,, \tag{3.2}$$

[1] We compute $\mathrm{M_{box}^{DM}}$ as $\mathrm{N_p} \cdot \mathrm{m_p}$ where $\mathrm{N_p}$ and $\mathrm{m_p}$ are the particles number and mass of the MR-II simulation, rescaled to the PLANCK15 parameters.

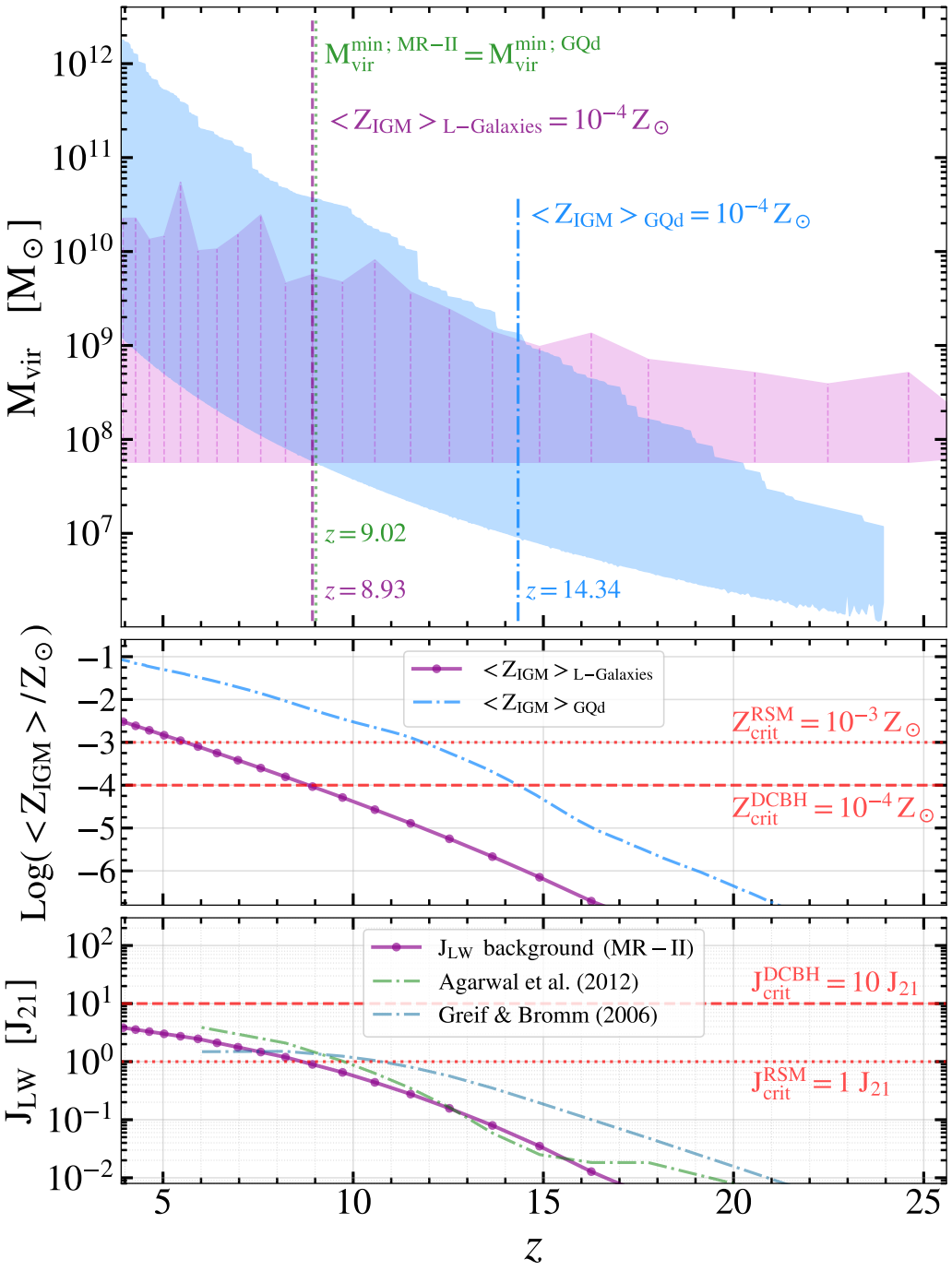

Figure 3.1: *Upper panel*: comparison between the dynamical ranges of newly-resolved `MR-II` halos (pink shaded area) and all the `GQd` halos (cyan shaded area). At $z \gtrsim 9$, `GQd` mass resolution is higher than the `MR-II` one, hence the `GQd` outputs can be used to *graft* `GQd` information into `L-Galaxies` structures (see §3.3.2). On the contrary, at $z \gtrsim 20$ `MR-II` halos are resolved outside the dynamical range of `GQd`, hence we initialize these structure with the most massive `GQd` halos at the corresponding z. *Middle panel*: Evolution of the average IGM metallicity in `L-Galaxies` (purple line and dots), and in `GQd` (cyan dashed-dotted line). While `GQd` concentrates on a relatively biased region where a $M_{vir} \sim 10^{13}\,M_\odot$ halo forms by $z \sim 2$, `L-Galaxies` tracks the evolution of baryonic matter within a cosmological box of $d_{side} = 100\,Mpc\,h^{-1}$. This produces the significant differences between the IGM chemical enrichment obtained from the two models. *Bottom panel*: Background level of J_{LW} over the whole `MR-II` box, in units of J_{21} (purple line and dots; see §3.2.1 for details). Our computation provides values comparable to similar analytic approaches (green and light-blue dashed-dotted lines). We note that $J_{LW\,;\,bg}$ is never strong enough to overcome $J_{crit} = 10\,J_{21}$ at any $z > 4$.

where h is the Planck's constant, c is the speed of light and m_H is the hydrogen atom mass. $\eta = 4 \times 10^3$ is the number of H_2-dissociating photons produced per stellar baryon (following Agarwal et al., 2012, but see also Wise et al. (2014); Maio et al. (2019)), while ρ_* is the mass density of active stars in the whole MR-II box. Finally, $f_{esc} \leq 1$ is the escape fraction of ionizing UV photons. There is no clear consensus on its exact value and dependencies on the properties of the emitting sources (see the discussion in e.g. Agarwal et al., 2012; Dijkstra et al., 2014b). We fix $f_{esc} = 1$ to maximize $J_{LW\,;\,bg}$ although, even with this choice, $J_{LW\,;\,bg} < 10\,J_{crit}^{DCBH}$ at any z (see the bottom panel of Fig. 3.1). On the other hand, f_{esc} might influence the redshift at which $J_{LW\,;\,bg} > J_{crit}^{RSM}$ (see Sections 3.3.1 and 3.3.3 for the definition of J_{crit}^{DCBH} and J_{crit}^{RSM}). We address this point while discussing our results in §3.4.4.

3.2.2 Spatial variations of Z_{IGM}

To model the patchy diffusion of metals in the IGM, we assume that strong SNe feedback episodes power the propagation of metallic winds, i.e.: spherically-symmetric, pressure-driven bubbles around galaxies (e.g., Bertone et al., 2005, 2007; Sharma et al., 2014). Therefore, for each SNe feedback episode which is strong enough to produce M_{ej} (see §2.1.2), we model the time-evolution of a bubble of radius r_s. We follow the analytic approximation of Dijkstra et al. (2014a) which well reproduces more detailed calculations, such as those presented in e.g. Madau et al. (2001); Bertone et al. (2005); Sharma et al. (2014); Kim et al. (2017a); Yadav et al. (2017); Fielding et al. (2017) and Fielding et al. (2018):

$$r_s\,(M, t) = \left(\frac{E_0\,\nu\,M_{*\,;\,new}}{m_p\,n} \right)^{1/5} (t - t_0)^{2/5}\,, \tag{3.3}$$

where t_0 is the time at which SNe are produced, $E_0 = 10^{51}\mathrm{erg} = 1\,\mathrm{Bethe}$ is the energy released by a single SN explosion, ν is the number of SNe per unit-mass of the newly formed stellar mass $M_{*\,;\,new}$, and m_p is the proton mass. We fix ν to the recycled fraction of L-Galaxies, namely $\nu = R = 0.43$ (see §2.1.2). The quantity n at the denominator of Eq. 3.3 is the number density of the IGM gas in which the shell expands. Following Dijkstra et al. (2014a) we set:

$$n = 60\,\Omega_{b;0}\,\frac{\rho_{crit;0}}{m_p}\,(1+z)^3\,. \tag{3.4}$$

The factor 60 in this formula represents the typical density-contrast, with respect to the cosmic average, of the IGM medium in which the SNe ejecta expand (see the discussion in Dijkstra et al., 2014a).

Equation 3.3 shows how $M_{*\,;\,new}$ regulates the speed of the bubble expansion. As a consequence, expanding fronts associated to strong SF events might reach slower fronts eventually launched earlier in time. In this case, we choose to follow only the faster one and update the bubble expansion-rate. Finally, we neglect energy losses during the fronts interaction as well as during the expansion within the IGM. We discuss the impact of these hypothesis on our results in our discussion (§5). Ours is a simplified, yet descriptive, approach which allows us to determine the size of chemically-enriched regions around galaxies. In §3.2.4 we detail how this model is used to actually compute the local Z_{IGM} (i.e. $Z_{IGM\,;\,local}$) at any point of interest in the MR-II box.

3.2.3 Spatial variations of J_{LW}

Due to the action of cosmological redshift, photons are redshifted during their propagation over cosmological distances. Consequently, photons emitted at LW frequencies can exit from the LW band, while more-energetic UV photons can enter in it. At the same time, as soon as they are redshifted to the frequency of an atomic-hydrogen transition, these ionizing photons are efficiently absorbed by neutral hydrogen. This complex process produces *windows* and *barriers* in the transmission function of the neutral IGM at $z \gtrsim 6$. The overall result of this phenomenon is a generally large mean free-path of ionizing UV photons, which can travel for cosmological distances even in a neutral IGM (see e.g., Ahn et al., 2009).

In order to rigorously describe the amount of LW photons received at any location in space and time, a multi-frequency radiative-transfer analysis should be carried out, considering the effects of both neutral

hydrogen absorption and cosmological redshift. Nevertheless, this approach can be computationally expensive, especially over cosmological boxes. Therefore, we follow the equations presented in Ahn et al. (2009); Agarwal et al. (2012) and Dijkstra et al. (2014a) to compute the J_{LW} received by an observer from a single star-forming galaxy (i.e. a LW source):

$$J_{LW}\,(d_L) = \frac{f_{esc}}{4\pi\,\Delta\nu}\,\frac{Q_{LW}\;M_{*\,;\,new}}{4\,\pi\,d_L^2}\,f_{mod}\,. \tag{3.5}$$

Here $\Delta\nu$ is the frequency interval of the LW band (see §1.2.1) while $M_{*\,;\,new}$ is the stellar mass of short-lived, instantly-recycled massive stars (see §2.1.2) at the time of the LW photons emission. We use $M_{*\,;\,new}$ because we assume that only these stars actively produce a significant J_{LW} (see e.g., Schaerer, 2002a; Agarwal et al., 2012; Wise et al., 2014; Maio et al., 2019). This implies that star-forming galaxies in our model exhibit a varying LW emission at each SF episode during their evolution. $d_L = d_c\,(1 + z_{rel})$ is the luminosity distance between observer and LW source, with $z_{rel} = (1 + z_{obs})/(1 + z_s)$ is the LW source redshift computed from the observer perspective, while z_{obs} and z_s are the redshift of observer and LW source with respect to $z = 0$. The quantity Q_{LW} is the rate of energy produced within $\Delta\nu$ by the newly-formed stellar mass $M_{*\,;\,new}$. We fix $Q_{LW} = 10^{44}\,erg\,s^{-1}$ for the PopII/PopI stellar populations modelled by L-Galaxies (see e.g., Greif & Bromm, 2006; Agarwal et al., 2012). Finally, f_{mod} is a factor taking into account both the cosmological redshift and the dimming action of the neutral IGM at high-z. Following Ahn et al. (2009):

$$f_{mod} = \begin{cases} 1.7\,\exp\left[-(\,d_c\,/\,116.26\,\alpha\,)^{\,0.68}\right] - 0.7 & d_c \;\leq\; 97.39\,\alpha\;, \\ 0 & d_c \;>\; 97.39\,\alpha\;, \end{cases} \tag{3.6}$$

where d_c is the comoving distance between source and observer in Mpc, while $d_{hor} = 97.39\alpha$ Mpc is the LW photons horizon, where α is defined as:

$$\alpha = \left(\frac{0.7}{h}\right)\left(\frac{0.27}{\Omega_{m\,;\,0}}\right)^{1/2}\left(\frac{21}{1+z_s}\right)^{1/2}\,. \tag{3.7}$$

In §3.2.4 we detail how we obtain the local J_{LW} at any specific point of interest (i.e. $J_{LW\,;\,local}$).

3.2.4 Implementation of Z_{IGM} and J_{LW} spatial variations

$Z_{IGM\,;\,local}$ and $J_{LW\,;\,local}$ are the result of superposing contributions which may originate from any point of the simulated box. Nevertheless, accounting for all the possible sources of metal winds and LW photons over the whole MR-II volume can be computationally challenging, especially in the case of $J_{LW\,;\,local}$, due to the large LW horizon at high-z (see §3.2.3, and e.g., Ahn et al., 2009). Therefore, we only check $Z_{IGM\,;\,local}$ and $J_{LW\,;\,local}$ at specific positions of interest $\mathbf{x_P}$, namely: the position of newly-resolved halos which may host a RSM or DCBH seed at their initialization redshift z_{init}. For either $Z_{IGM\,;\,local}$ or $J_{LW\,;\,local}$, we follow an inside-out procedure in three separate steps by integrating the contributions of: i) eventual SF episodes inside halos, ii) close neighbors belonging to the same FoF and iii) halos belonging to different FoF groups, hence finally encompassing the eventual contribution of the whole box at $\mathbf{x_P}$. We stress that each of the above steps is only performed if the conditions for RSM or DCBH seeding are still verified after the previous step.

Even with this simplified approach, directly accounting for all the possible sources of J_{LW} or IGM metals for each newly-resolved halo is computationally prohibitive. To overcome this issue we define two catalogs of *sources* which include only the halos which significantly influence their surroundings with metallic winds or LW flux. More in detail, by considering the whole MR-II box and any z, we define as *metal sources* all galaxies with a metallic-bubble extended at least to 10 times their R_{vir}. Analogously, *LW sources* are galaxies which produce at least $J_{LW} = 10\,J_{21}$ at $d_L = R_{vir}$. We stress that these definitions are a trade-off between the inclusion of the highest possible number of sources in their respective catalog and reasonable execution times of our model. These two source-catalogs are stored only once, in a specific preliminary run of L-Galaxies and used in all subsequent runs.

Once the source catalogs are produced, for each newly-resolved halo which does not already host a

BH, we look for metal sources whose metallic-bubble radius r_s is able to reach $\mathbf{x_P}$ at $z_{\rm init}$. In this case, metal sources contribute to $Z_{\rm IGM\,;\,local}$ at $\mathbf{x_P}$ with a fraction f_m of the M_{ej}^{met} and M_{ej} inside their metallic bubble. In particular: $f_m = (rs/R_{vir})^3$, where R_{vir} is the virial radius of the newly-resolved halo. For $J_{\rm LW\,;\,local}$ we reconstruct the past light-cone of the newly-resolved halo and sum the contribution of each LW source in it (as given by Eq. 3.5). Finally, we flag as *metal contributors* or *LW contributors* all the galaxies which provide an actual contribution to $Z_{\rm IGM\,;\,local}$ or $J_{\rm LW\,;\,local}$.

3.3 BH-seeding prescription

We check for the occurrence of favourable conditions for SMBH seeding in every newly-resolved DM halo within the MR-II merger trees. Our prescription includes the formation of BH seeds over a wide mass range, from $M_{seed} \sim 10^2\,M_\odot$ to $M_{seed} \sim 10^5\,M_\odot$, in four different BH-seeds flavours: i) light PopIII remnants, ii) intermediate-mass BHs resulting from the RSM scenario, iii) heavy DCBHs originating from the monolithic compression of pristine gas clouds and finally iv) merger-induced direct-collapse BHs (miDCBHs) produced at the center of gas-rich galaxy mergers. In particular, light PopIII remnants are inherited from the GQd model, as described in §3.3.2, while the specific conditions for RSM and DCBH seeds formed self-consistently in L-Galaxies are detailed in §3.3.3. Ours is a comprehensive and physically-motivated model which accounts for the main SMBH seeding (baryonic) scenarios currently envisioned in high-z collapsing halos (e.g., Latif & Ferrara, 2016; Valiante et al., 2017; Inayoshi et al., 2020). We provide an overview of our BH-seeding prescription in §3.3.1 and summarize it in Table 3.1. In §3.3.2 we detail the interplay between GQd and L-Galaxies models, and finally in §3.3.3 we address the implementation of each self-consistent seeding channel separately.

3.3.1 Overview of the SMBH seeding model

As a first step, we statistically account for the un-resolved formation of light PopIII remnants within MR-II halos by populating L-Galaxies structures with light seeds from GQd. This procedure does not depend on $Z_{\rm IGM\,;\,local}$ nor $J_{\rm LW\,;\,local}$, since these cannot be tracked during the un-resolved evolution of halos. Rather, we match GQd halos to the newly-resolved L-Galaxies ones through their virial mass and redshift, as further detailed in §3.3.2. This method naturally extends the BH occupation of GQd on the MR-II merger trees. Consequently, a fraction of newly-resolved halos in L-Galaxies does not inherit light seeds due to the low BH occupation in GQd halos, especially at $M_{vir} \lesssim 5 \times 10^8 M_\odot\,h^{-1}$ and $z \lesssim 12$ (see Fig. 3.2).

After the initialization with GQd information, all newly-resolved structures without a central BH are candidates for self-consistent BH formation in L-Galaxies. By default, we consider halos hosting GQd PopIII remnants as chemically enriched by early PopIII episodes beyond the limits for DCBH formation (see e.g., Ritter et al., 2012; Katz et al., 2015). Nevertheless, since mild local pollution does not exclude the onset of the RSM scenario (e.g., Omukai et al., 2008; Devecchi & Volonteri, 2009), we still allow the formation of intermediate-mass BHs (IMBHs) into newly resolved L-Galaxies structures hosting light GQd BH seeds.

As better detailed in §3.3.3, we discriminate between different seeding pathways by comparing $Z_{\rm IGM\,;\,local}$ and $J_{\rm LW\,;\,local}$ to different thresholds for different seeding channels, namely: Z_{crit}^{RSM} and J_{crit}^{RSM} as well as Z_{crit}^{DCBH} and J_{crit}^{DCBH} (e.g., Omukai et al., 2008; Volonteri, 2010; Devecchi & Volonteri, 2009; Valiante et al., 2017; Sassano et al., 2021). Finally, we check whether the conditions for the miDCBH channel (Mayer et al., 2010; Bonoli et al., 2014) are verified throughout the cosmological evolution of halos. The activation of the latter channel does not depend on $Z_{\rm IGM\,;\,local}$ nor on $J_{\rm LW\,;\,local}$, hence it can occur at any z.

We stress that the gas metallicity within newly-resolved structures is determined by $\langle Z_{\rm IGM}\rangle$ computed at $z_{\rm init}$ (see §3.2.1). As a consequence, the formation of self-consistent DCBHs and RSM seeds in our model cannot proceed beyond the moments when $\langle Z_{\rm IGM}\rangle(z)$ overcomes respectively Z_{crit}^{DCBH} and Z_{crit}^{RSM} (see Fig. 3.1, middle panel).

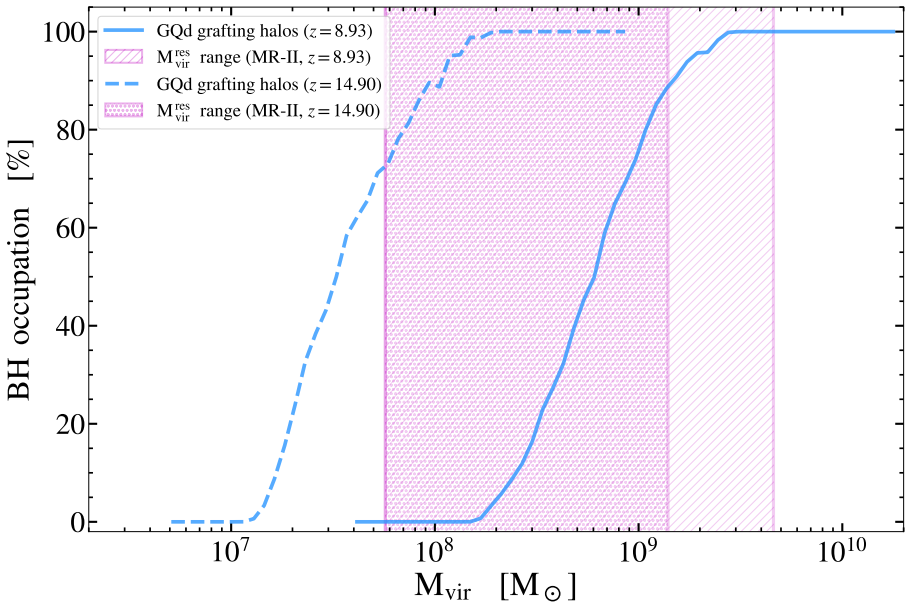

Figure 3.2: BH occupation-fraction as a function of M_{vir} in the GQd halos used for grafting (dashed and solid cyan lines). We only show $z\sim9$ and $z\sim15$ as examples, and highlight the dynamic range of newly-resolved MR-II halos as the pink hatched areas. At $z\gtrsim12$, most of the halos grafted from GQd inherit a BH, while at $z\lesssim12$ an increasingly large fraction of low-mass L-Galaxies structures (i.e. $M_{vir}\lesssim5\times10^8\,M_\odot$) can be initialized without a BH from GQd.

3.3.2 Sub-grid SMBH seeding from GQd

As anticipated, we use the outputs of the GQd model of Valiante et al. (2021) in order to include light seeds in our model; a process which we call *grafting* of GQd information into L-Galaxies halos. In order to initialize the latter at a given z_{init}, we consider all GQd halos between two consecutive MR-II snapshots, namely: $z_{init}\leq z_{GQd}<z_{prev}$ (being z_{GQd} the redshift of GQd outputs). For each newly-resolved halo of MR-II we randomly extract a GQd halo from this sample, within a M_{vir} bin of 0.5 dex, centered on the M_{vir}^{res} of the L-Galaxies halo. To increase the statistics of the GQd sample, we use 10 different realizations of a GQd merger tree. Each of these is an independent run of GQd executed with the same simulation parameters (Valiante et al., 2021), leading to a final halo of $M_{vir}^{final}=10^{13}\,M_\odot$ at $z=2$ through a different reconstruction of its merger history. We refer to this process as the *grafting* of GQd information into L-Galaxies halos.

Our method matches the virial mass distributions of GQd and newly-initialized MR-II halos in the regions where the two dynamical ranges overlap, i.e. at $8\lesssim z\lesssim16$ (Fig. 3.1, upper panel). At $z>16$, the maximum M_{vir}^{res} of L-Galaxies halos is higher than the maximum mass of GQd halos, hence we use the highest-mass bins of the GQd dynamical range. Finally, we interrupt the grafting from GQd as soon as one of the following conditions are verified: i) $\langle Z_{IGM}\rangle>Z_{crit}=10^{-4}Z_\odot$, ii) the (redshift-evolving) mass resolution of GQd overcomes M_{vir}^{min} (Fig. 3.1, upper panel). Indeed, the first condition generally marks the inhibition of PopIII SF (see e.g., Bromm & Larson, 2004; Schneider et al., 2006; Maio et al., 2010; Ritter et al., 2012; Valiante et al., 2016; Pezzulli et al., 2017), hence there is no need to track un-resolved PopIII SF in L-Galaxies, while the second condition implies that GQd can no longer track the un-resolved evolution of MR-II halos, prior to their initialization in L-Galaxies.

Grafted information from GQd

We only initialize the BH mass and seed-type of central BHs in the newly-resolved structures of L-Galaxies with GQd information. We do not account for other baryonic quantities to avoid biases due to the different implementation of physical processes governing baryonic evolution with respect to L-Galaxies (see e.g. §2.1.2, §3.2 Valiante et al., 2011, 2016; Henriques et al., 2015). Moreover, GQd is designed to track the merger tree of a single halo which reaches $M_{vir}=10^{13}\,M_\odot$ at $z\sim2$, hence simulating a relatively biased region of the Universe. Although different realizations of this merger tree enhance

their statistics, the outputs of GQd cannot account for all the different environments simulated in the MR-II cosmological box.

The GQd version presented in Valiante et al. (2016) and Valiante et al. (2021) includes the formation of both light and DCBH seeds. Consequently, we flag the SMBHs inherited from GQd as either descending from *pure-light*, *heavy* or *mixed* seed-types. The first ones can only derive from either isolated or merged PopIII remnants, the second class is the result of isolated or merged pure-heavy types while, finally, SMBHs belonging to the latter class experienced at least one merger between a pure-light and a pure-heavy seed during their evolution in GQd. We choose to ignore the pure-heavy DCBH remnants of GQd since we can model self-consistently their formation in L-Galaxies. On the other hand, we separately account for the contributions of light and heavy channels to the M_{seed} of the mixed GQd BHs we graft in L-Galaxies.

Grafting probability: the G_p parameter

The grafting of GQd BHs does not depend on the $Z_{IGM\,;\,local}$ and $J_{LW\,;\,local}$ values we compute in the MR-II box. Furthermore, GQd models the formation of BHs within the merger tree of a DM halo which reaches $M_{vir} = 10^{12} M_{\odot}$ at $z = 2$. This is only representative of a relatively biased environment among the whole cosmological variety modelled by MR-II. As a consequence, our grafting procedures tend to generalize the occupation of this specific SMBHs population to the whole MR-II box, regardless of the local properties of the environment in which new SMBHs are grafted. As we further discuss in §3.4, this has evident implications onto the properties of our SMBH population at low-z. Therefore, in order to control the abundance of light seeds inherited from GQd we include the possibility to bypass the grafting of GQd BHs even if the conditions for grafting are met. Since most of the structures in L-Galaxies are resolved with $M_{vir}^{res} \sim 2 \times 10^8 M_{\odot}$, we model a grafting probability ($\mathcal{P}_{graft}$) as a linear function of M_{vir}^{res}:

$$\mathcal{P}_{graft}(M_{vir}^{res}) = G_p \cdot \left(\frac{M_{vir}^{res}}{M_P} \right) , \tag{3.8}$$

where G_p is a free parameter ranging from 0 to 1 which controls the magnitude of $\mathcal{P}_{graft}$ at the characteristic mass $M_P = 10^8 M_{\odot}$. We stress that $\mathcal{P}_{graft}$ is saturated to 1 when $M_{vir}^{res} > M_P / G_p$.

3.3.3 Self-consistent SMBH seeding in L-Galaxies

This section briefly details our prescriptions for producing SMBH seeds self consistently in L-Galaxies, as a function of local gas conditions, according to the RSM, DCBH and miDCBH seeding scenarios. We stress that, when forming any of these BH seeds, we always impose mass conservation by requiring that the baryonic content of newly-resolved MR-II halos can provide the required M_{seed}.

The J_{crit} threshold

The exact threshold of J_{crit} required for inhibiting H_2-cooling has been widely discussed in the literature, ranging from $J_{crit} \lesssim 0.01 - 0.1\, J_{21}$ (e.g., Yoshida et al., 2003, 2007; O'Shea & Norman, 2008; Devecchi & Volonteri, 2009; Shang et al., 2010; Regan et al., 2014) up to $\gtrsim 10^4 - 10^5 J_{21}$ when taking into account different factors, such as the properties of LW-emitting sources (e.g., Latif et al., 2015; Agarwal et al., 2016; Regan et al., 2016), the complete dissociation of H_2 and H^- (e.g., Wise & Abel, 2007; Wolcott-Green & Haiman, 2011; Latif et al., 2014a; Latif & Khochfar, 2019), H_2 self-shielding (e.g., Wolcott-Green et al., 2011; Hartwig et al., 2015b; Wolcott-Green & Haiman, 2019) or the eventual rotational-support of cooling clouds (e.g., Latif et al., 2014c; Latif & Volonteri, 2015). Interestingly, recent work suggested new analytic methods which do not employ a fixed J_{crit} threshold (e.g., Wolcott-Green et al., 2017; Agarwal et al., 2019).

Within our simplified, analytical prescription we use the relatively low value $J_{crit}^{DCBH} = 10\, J_{21}$, as a compromise among all the approaches followed in the literature (e.g. see the discussion in Agarwal et al., 2016; Regan & Downes, 2018; Dunn et al., 2018; Inayoshi et al., 2020; Lupi et al., 2021a). Futhermore, the complete dissociation of H_2 and H^- is not strictly required for RSM seeds formation, so that lower levels of $J_{LW\,;\,local}$ might be sufficient to simply delay H_2-cooling (e.g., Omukai, 2001b; Yoshida et al., 2007; O'Shea & Norman, 2008). This likely induced several past work to relax the requirements on

J_{crit} for RSM formation (see e.g., Katz et al., 2015; Kimm et al., 2016; Sakurai et al., 2017; Das et al., 2021), although this condition is necessary to increase the critical density for gas fragmentation (e.g., Omukai et al., 2008; Latif et al., 2014c; Sassano et al., 2021). By following this argument, we choose to fix $J_{crit}^{RSM} = 0.1\, J_{crit}^{DCBH}$, which also allows us to control the abundance of RSM and DCBH seeds by only acting on the J_{crit}^{DCBH} free-parameter.

IMBHs from runaway stellar mergers

RSM seeds are thought to form in the core of star-forming atomic-cooling halos under specific conditions which generally include mild chemical enrichment ($10^{-6} \lesssim Z/Z_{\odot} \lesssim 10^{-3}$, e.g., Chon & Omukai, 2020; Sassano et al., 2021) and the presnece of $J_{LW\,;\,local}$. These conditions ensure that gas cooling is driven by atomic transitions and fragmentation can only occur above a metallicity-dependent density threshold $n_{crit;Z}$ (see §1.2.1 and e.g., Devecchi & Volonteri, 2009; Stone et al., 2017). Consequently, the dense nuclear regions of the halo can host intense star formation, fuelled by strong gas inflows (see e.g., Omukai et al., 2008, for a thorough discussion). The end-product of this scenario is the formation of a nuclear, dense stellar cluster of mass $M_{cl} \sim 10^5\, M_{\odot}$, where efficient stellar collisions eventually lead to the prompt formation of a $M_{seed} \sim 10^{3-4}\, M_{\odot}$ central BH (e.g., Omukai et al., 2008; Sassano et al., 2021).

We add this seeding channel to L-Galaxies by closely following the physical model delineated in Devecchi & Volonteri (2009). We adapt their purely-analytic prescription to our case by using the sub-halos virial masses, radii and spin parameters provided by the MR-II outputs and the values of Z_{local} and $J_{LW\,;\,local}$ computed self-consistently with our model (see §3.2.2 and 3.2.3). By following Devecchi & Volonteri (2009), we require $10^4 \lesssim Tvir/[K] \lesssim 1.8 \times 10^4$ for our RSM candidates, which translates into a condition on their virial mass (see Eq. 1.29). Furthermore, we impose $Z_{local} \leq Z_{crit}^{RSM} = 10^{-3} Z_{\odot}$ and $J_{LW\,;\,local} \geq J_{crit}^{RSM}$ (as discussed in Omukai et al., 2008; Sassano et al., 2021). Under these conditions we apply the procedure detailed in Devecchi & Volonteri (2009) and assume that star formation proceeds in the region of RSM-candidate halos where gas density overcomes $n_{crit;Z}$. This allows to use the halo properties to compute the radius enclosing SF processes and the mass M_{cl} of the resulting stellar cluster. M_{cl} determines the time-scale t_{cc} over which stars are expected to mass-segregate due to dynamic interactions and produce a dense core of massive stars (e.g., Portegies Zwart & McMillan, 2002; Rasio et al., 2004). If t_{cc} is shorter than the typical main-sequence life of massive stars (i.e. $t_{MS} = 5\,Myr$ in our model), then runaway mergers instabilities can effectively produce a central very-massive star (VMS, see e.g., Portegies Zwart et al., 1999). Finally, the latter may evolve into a RSM seed whose mass is determined by M_{cl} and t_{cc} (see Portegies Zwart & McMillan, 2002; Devecchi & Volonteri, 2009; Stone et al., 2017).

Direct-Collapse Black Holes

We model the formation $M_{seed} = 10^5 M_{\odot}$ DCBHs inside newly-resolved atomic-cooling halos with i) $Z_{IGM\,;\,local} < Z_{crit}^{DCBH}$, ii) $J_{LW\,;\,local} \geq J_{crit}^{DCBH}$, iii) a minimum cold-gas mass of $M_{cGas} > M_{seed}$ and iv) $M_* = 0$. Indeed, as soon as stars are formed, the mechanical and chemical feedback of SNe reduces the halo gas content and enriches it with metals and dust, preventing the formation of DCBH seeds (e.g., Ritter et al., 2012; Valiante et al., 2016; Agarwal et al., 2017; Maio et al., 2019; Sassano et al., 2021). We fix the value of $Z_{crit}^{DCBH} = 10^{-4} Z_{\odot}$ by following the approach of e.g. Bromm & Loeb (2003); Agarwal et al. (2012); Dijkstra et al. (2014a); Valiante et al. (2016). Finally, it has been argued that the formation of DCBHs requires a prolonged illumination by H_2 photo-dissociating fluxes, lasting for at least one dynamical time of the pristine collapsing clouds (see e.g., Habouzit et al., 2016; Regan et al., 2017; Lupi et al., 2021a). This is to ensure that SF processes are delayed at least until the gas cloud is able to collapse and form the central massive object, without suffering mass losses due to SNe feedback. Given the relatively large time-step used by L-Galaxies ($10 \lesssim \Delta t \lesssim 60$Myr at $z \gtrsim 8$) we assume that the presence of $J_{LW\,;\,local} > J_{crit}^{DCBH}$ at z_{init} is a sufficient condition for our DCBHs candidates to be illuminated for their whole dynamical time ($\tau_{dyn} \sim 1/\sqrt{G\rho} \sim 10$Myr, for newly-resolved MR-II halos at $z \gtrsim 8$).

Merger-induced Direct-Collapse Black Holes

An alternate route to the formation of massive DCBHs is offered by strong inflows of gas towards the center of high-z gas-rich mergers (see Mayer & Bonoli, 2019, for a recent review). We add this BH seeding channel to our model by closely mirroring the approach of Bonoli et al. (2014), which tested this BH formation scenario on the MR simulation merger trees. This seeding channel does not require conditions on $Z_{IGM\,;\,local}$ or $J_{LW\,;\,local}$. Rather, it is based on the possibility of obtaining strong, multi-scale inflows of gas as a consequence of galaxy mergers. Indeed, numerical simulations have shown that gas-rich interactions can compress $M_{gas} \sim 10^{7-8} M_{\odot}$ within the nuclear region of the merger remnant (see e.g., Mayer et al., 2010; Inayoshi et al., 2015). This process might provide the extreme gas-inflow rates needed to grow a central compact object up to $M_{seed} \sim 10^{4-5} M_{\odot}$ in few 10^5yr (see Mayer et al. (2010) and Bonoli et al. (2014) for detailed discussions).

We require that i) the merger has a minimum mass-ratio of 0.3[2], ii) the merger remnant has a minimum halo mass of $M_{rem}^{min} = 10^9 M_{\odot}$, iii) the two merging galaxies are disk-dominated, hence exhibit a maximum bulge-to-total ratio of $B/T = 0.2$ and finally that iv) the merger remnant does not already host a central BH more massive than $M_{BH\,;\,max} = 5 \times 10^4 M_{\odot}$. With the exception of the condition on M_{rem}^{min}, these requirements are identical to those of Bonoli et al. (2014), hence we refer to this work for their theoretical justification. We underline that M_{rem}^{min} is lower than the one adopted by Bonoli et al. (2014), in order to account for the different dynamic range and merger histories of MR-II with respect to MR. Finally, when all the above conditions are met, we form a massive BH seed with $M_{seed} = 8 \times 10^4 M_{\odot}$.

3.4 Results

Our model allows to track the formation of BH seeds and their evolution as SMBHs in a cosmological context. Thanks to the N-body, cosmological nature of the MR-II merger-trees, we can both analyze the high-z environment in which massive BHs form and trace the statistical properties of their lower-z SMBH descendants. This section details our results on both sides.

3.4.1 Build-up of a multi-flavour SMBHs population

To understand the origin and evolution of SMBHs, we keep track of both their seeding-mass M_{seed} and their formation-channel. With this information, we define four different *seed classes*, namely: pure-light (see §3.3.2), *pure-RSM*, *pure-DCBH* and *miDCBH*. Each of these includes SMBHs which exhibit exclusively one component. For instance, pure-RSM and pure-DCBH descend from SMBH seeds formed self-consistently in L-Galaxies which evolved in isolation or only merged with pure-RSM or pure-DCBH descendants, respectively. On the other hand, mergers can also lead to seed-types mixing. To account for these cases, we define three additional seed-types:

- *RSM*: originating exclusively from the mix between pure-RSM and pure-light descendants,
- *Heavy*: originating either from the grafting of *mixed* seed-type from GQd (see §3.3.2), or the merger between pure-DCBH and pure-light descendants.
- *Hybrid*: this class accounts for all the SMBHs showing at least three different seed-types among their progenitors. Although this seed-type is extremely rare in the MR-II box, we report it here for completeness.

Table 3.1 summarizes the conditions leading to the formation of each of these seed-type classes. Figure 3.3 shows the evolution of the BHs number-density for a run with $G_P = 1$ (see Sect. 3.3.2), split in the different seed-types classes. This highlights the build-up of the multi-flavour seeds-descendants population in time, as well as their gradual mixing. We note that pure-light seeds-descendants (light-blue lines) are numerically dominant over other seed-types by several orders of magnitude, at all z. Heavy seed-types (purple lines) begin to appear in our model at $z \lesssim 13$, either due to the grafting of mixed GQd SMBHs

[2] As in (Bonoli et al., 2014), we compute the merger-ratio by considering both the stellar and cold-gas component of the merging structures.

Seed Type	$Z_{IGM\,;\,local}$	$J_{LW\,;\,local}$	Grafting	Merger
Pure-light	-	-	✓	-
Pure-DCBH	$Z_{IGM\,;\,local} \leq Z_{crit}^{DCBH}$	$J_{LW\,;\,local} \geq J_{crit}^{DCBH}$	-	-
Pure-RSM	$Z_{IGM\,;\,local} \leq Z_{crit}^{DCBH}$ $Z_{crit}^{DCBH} < Z_{IGM\,;\,local} \leq Z_{crit}^{RSM}$	$J_{crit}^{RSM} \leq J_{LW\,;\,local} < J_{crit}^{DCBH}$ $J_{LW\,;\,local} \geq J_{crit}^{RSM}$	- -	- -
miDCBH	-	-	-	✓
	Merging seed types			
RSM	pure-light and pure-RSM		✓	✓
Heavy	pure-light and pure-DCBH light and heavy within GQd evolution		✓	✓
Hybrid	RSM and Heavy and/or miDCBH		✓	✓

Table 3.1: Summary of the conditions for the formation of each seed-type in our model. The table is split horizontally between seed-types having a single-flavour origin (upper half), and seed-types forming through the merger of different seed-types (lower half). The last two columns to the right highlight whether seed-types inherited any component from GQd grafting and if they formed during galaxy mergers. We note that both pure-light and miDCBH seed-types do not require any condition on $Z_{IGM\,;\,local}$and $J_{LW\,;\,local}$ to form, but they exhibit a different origin: pure-light are inherited from GQd while miDCBHs form self-consistently in L-Galaxies mergers. Finally, heavy seed-types can form through the merger of either i) pure-light and pure-DCBHs or ii) the merger between a light and heavy seed in GQd. The latter case accounts for the grafting of *mixed* seed-types from GQd (see Valiante et al., 2016) into L-Galaxies structures (see §3.3.2).

or through the mixing between pure-light seeds and pure-DCBHs. At the same time, pure-RSMs (yellow lines) and pure-DCBHs (red lines) begin to form self-consistently in L-Galaxies. Both these classes always play a minor role in the global SMBH population, with only few of them reaching $z\sim0$ without mixing with other seed-types. We note that pure-DCBHs formation happens relatively late in our model as compared to previous works (e.g. Yue et al., 2014; Valiante et al., 2016; Lupi et al., 2021a; Sassano et al., 2021). This is an effect of the MR-II mass-resolution (see also Valiante et al., 2021, for the effect of mass resolution onto heavy-seeds formation) which only allows T_{vir} to become adequate for atomic-cooling only at $z\lesssim18$ (see Eq. 1.29 and Barkana & Loeb, 2001). Furthermore, as noted in 3.2.1, the value of $J_{LW\,;\,bg}$ is too low at any z to significantly influence the formation of pure-DCBHs (see also Fig. 3.6 and the bottom panel of Fig. 3.1). We note that the late formation of pure-DCBHs poses a challenge to the role which can be played in our model by these "late" pure-DCBHs in the build-up of $z\sim7$ SMBHs. Indeed, in order to reach $M_\bullet \gtrsim 10^8 M_\odot$ at $z\sim7$, the pure-DCBHs forming at $z\sim9$ in L-Galaxies should grow continuously at high rates and low radiative efficiency (i.e. $\epsilon\lesssim0.05$). We further discuss the evolution of pure-DCBHs in §3.4.3 and §3.4.4.

Analogously, the formation of pure-RSM seeds in our model is inefficient at $z\geq9$. Indeed, despite the less stringent limit imposed by J_{crit}^{RSM}, this channel also requires a mild, local chemical enrichment which is difficult to attain in our model at $z\geq9$. On the contrary, as shown in the middle and bottom panel of Fig. 3.1, at $z<9$ both $J_{LW\,;\,bg}$ and $\langle Z_{IGM}\rangle$ become compatible with the formation of pure-RSM seeds in our model, due to the thresholds we impose on $J_{LW\,;\,local}$ and $Z_{IGM\,;\,local}$ for this seeding channel (see Sect. 3.3.3). The combination of these two factors produces the sudden jump in the number of newly-initialized pure-RSM seeds we observe at $z\sim8$ (dashed yellow line in Fig. 3.3). Finally, miDCBHs are extremely rare in our model, since this seeding channel is particularly inefficient in our cosmological box. Indeed, even after adapting the conditions on M_{rem}^{min} (see Sect. 3.3.3), gas-rich mergers with high

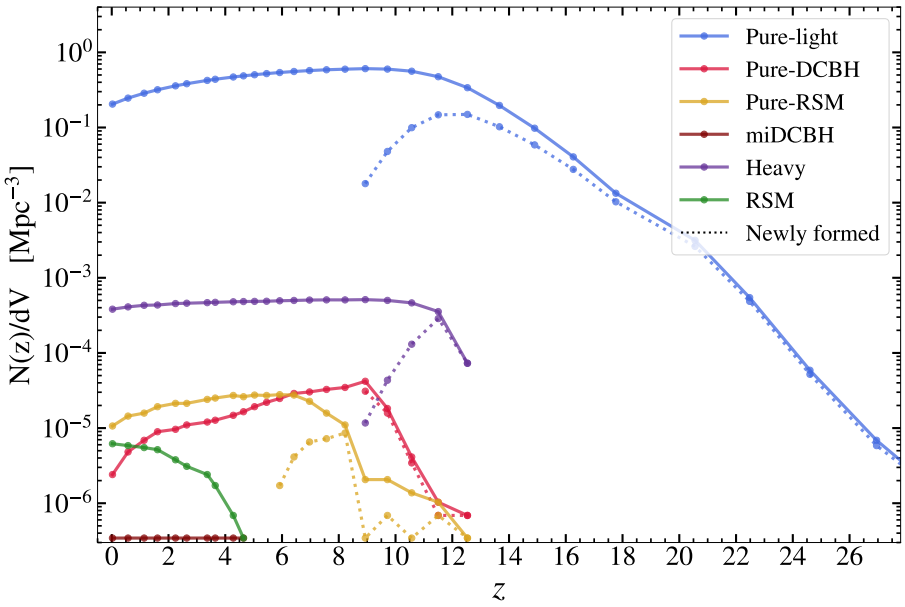

Figure 3.3: Evolution of the BHs number density for our the run of our model with $G_P = 1$ (see §3.3.2), split into different seeds-flavours. We mark the density of newly-formed BHs with dashed lines, to show the progress of BH seeds formation. We note that the latter stops at $z\sim9$ for light and DCBH seeds and proceed down to $z\sim6$ for RSM (see §3.3.1). After this, our SMBH population only evolves through mergers of already-existing SMBHs. It is evident that GQd light seeds (solid light-blue line) numerically dominate over all the other flavours (i.e. solid yellow, red and brown lines) and mixed classes (i.e. purple and green), which only play a marginal role in our version of L-Galaxies. Finally, the fact that $Z_{crit}^{DCBH} < Z_{IGM} < Z_{crit}^{RSM}$ at $6 \lesssim z \lesssim 9$ (see Fig. 3.1, middle panel) explains both the interruption of DCBH seeds formation (dotted red line) and the increase of RSM seeds formation (dotted yellow line).

merger-rates as those required to form miDCBHs are extremely rare in the MR-II simulated volume (see Izquierdo-Villalba et al., 2019b). Although further lowering M_{rem}^{min} might help to increase the efficiency of miDCBH formation, this would prevent the retention of the shocked gas in the merger remnant by its shallow DM potential-well (see e.g. Bonoli et al., 2014). We refrain to explore this possibility, since this is a necessary condition for the formation of strong multi-scale inflow in the Mayer et al. (2010) and Bonoli et al. (2014) models.

We stress that our seeding procedure stops at $z\sim6$ (see §3.3.1), therefore at lower z our population of SMBHs only evolves through mergers of already-existing objects. This explains the turn-over in the number density evolution of pure-light seed-descendants at $z\sim8$ (as well as the one of pure-DCBHs and pure-RSM at their respective z). At later times, only the number density of seed-types with a mixed or merger-driven origin might be able to increase, such as in the case of RSM (solid green line and dots), which emerges at $z<5$ through the mixing of pure-RSM and pure-light seeds-descendants. We note that our results are obtained using $f_{esc} = 1$ in Eq. 3.5. This value is consistent with observations suggesting high f_{esc} values at high-z (although current constrains are highly uncertain, see e.g. Dijkstra et al., 2014b). Consequently, the $J_{LW\,;\,bg}$ values we compute are maximized at any z with respect to this parameter. This implies that acting on f_{esc} would tend to hinder the formation of pure-RSM seeds in our model.

3.4.2 Nurturing environment of BH seeds

As detailed in §3.3.3, we impose a set of conditions on $Z_{IGM\,;\,local}$ and $J_{LW\,;\,local}$ for identifying galaxies that may host BH-seeding processes. Consequently, it is interesting to investigate how these requirements might influence the birthplaces of the various seed flavours.

We note that relatively dense and biased regions at high-z might host actively SF galaxies with a higher probability than less dense and quieter environments, hence allowing to find LW-bright halos in the proximity of a pristine, collapsing cloud. This is reflected in the cumulative number of neighbors as a function of distance from newly-formed SMBH seeds, which is a proxy of the local density of structures. In Fig. 3.4 we show the median distance at which the 1st, 10th, 25th and 40th neighbors are found, for each seed-type class present in the MR-II box at $z\sim9$. For comparison, we add the halos which do not

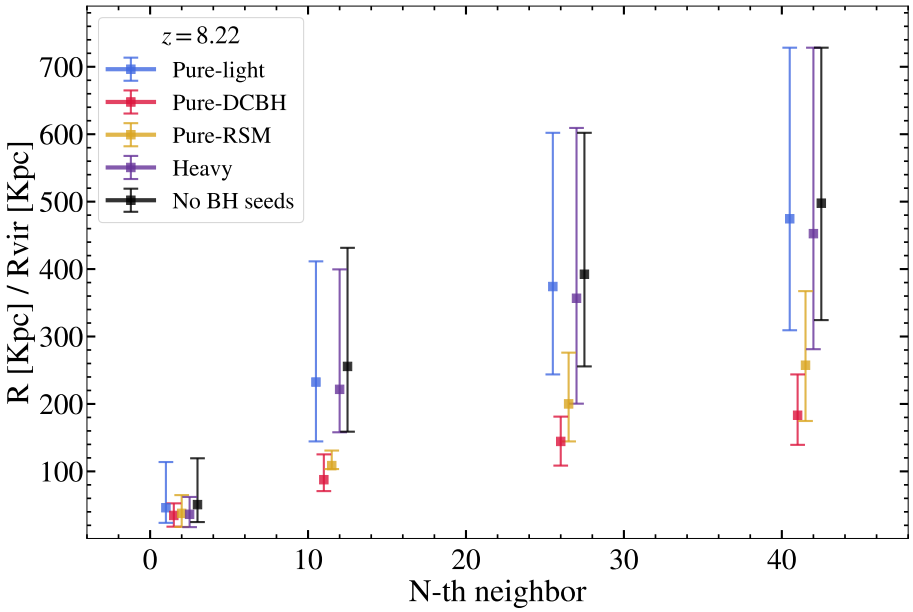

Figure 3.4: Typical distance at which the 1st, 10th, 25th and 40th neighbor is found from newly-formed BH-seeds, at $z \sim 9$. The latter are divided in two different classes: pure-RSMs and DCBHs (respectively orange and red squares). On the other hand, already-evolved pure-light seeds (cyan squares) are inherited from GQd. For comparison, we also show galaxies initialized without a BH-seed (black squares). Points are slightly shifted along the x-axis for a better visualization, while the error bars mark the 16th and 84th percentiles of the distances distribution. This result shows that, on average, pure-DCBHs and pure-RSMs form in more populated regions than other seed-types.

host a central BH (black squares). In this analysis smaller distances imply denser environments, as in the case of newly-formed galaxies hosting pure-DCBHs (red squares) and pure-RSM (yellows squares) with respect to those without BHs or hosting pure-light or heavy seeds descendants (respectively light-blue and purple squares). The latter follow the trend of pure-light descendants because on one hand they are defined by the merger between pure-light and pure-heavy components, and also because pure-light descendants are largely predominant in our model (see Fig. 3.3). We stress that BHs inherited from GQd already underwent unresolved evolution at the moment of their grafting in L-Galaxies, therefore the result in Fig. 3.4 is not representative of their formation environment in GQd. Rather, structures grafted with GQd light seeds trace the average population of newly-initialized halos with $M_{vir}^{init} \gtrsim 10^9\,M_\odot$. This is because our grafting procedure does not depend on environmental conditions (see §3.3.2).

We note that the difference between objects hosting either pure-DCBH (or pure-RSM) descendants and galaxies hosting pure-light (or heavy) seed-types is systematically present at least up to their 40th neighbor. Since the formation of pure-DCBHs and pure-RSMs is dictated by local physical conditions, this suggests that $J_{LW\,;\,local}$ and $Z_{IGM\,;\,local}$ also leave an imprint on the nurturing environment of different seed-types, at least up to few hundreds times the virial radii of their hosts. This is also reflected on the (low significance) systematic shift between the results for pure-DCBHs and pure-RSM seeds, as these channels require different levels of $J_{LW\,;\,local}$ and $Z_{IGM\,;\,local}$.

As a representative example, Fig. 3.5 presents the local environments of a pure-DCBH and a pure-RSM seed (respectively upper and lower row). In each case, we show four different spatial-maps computed at the moment of the BH formation, centered at the position of the BH seed. From left to right: the projected density of virial mass (first panel) and SFR (second panel), the local intensity of LW flux as seen by the newly-formed BH (third panel) and the local level of chemical enrichment (last panel to the right). Each of these only shows a slice of $(x, y, z) = (2, 2, 0.5)$ Mpc, centered on the newly-formed BH. In particular, the LW map represents the 2D projection of the spatial variations of $J_{LW\,;\,local}$, as seen from the newly-formed BH seed, i.e. the contribution to $J_{LW\,;\,local}$ received from its past light-cone. LW sources can contribute to $J_{LW\,;\,local}$ from previous times with respect to the one at which the BH-seed actually forms, hence they might not appear in the M_{vir}, ρ_{SFR} and Z_{IGM} maps. Although being only qualitative, these maps show different degrees of occupation by neighboring structures in the environment of these two different BH-seeds.

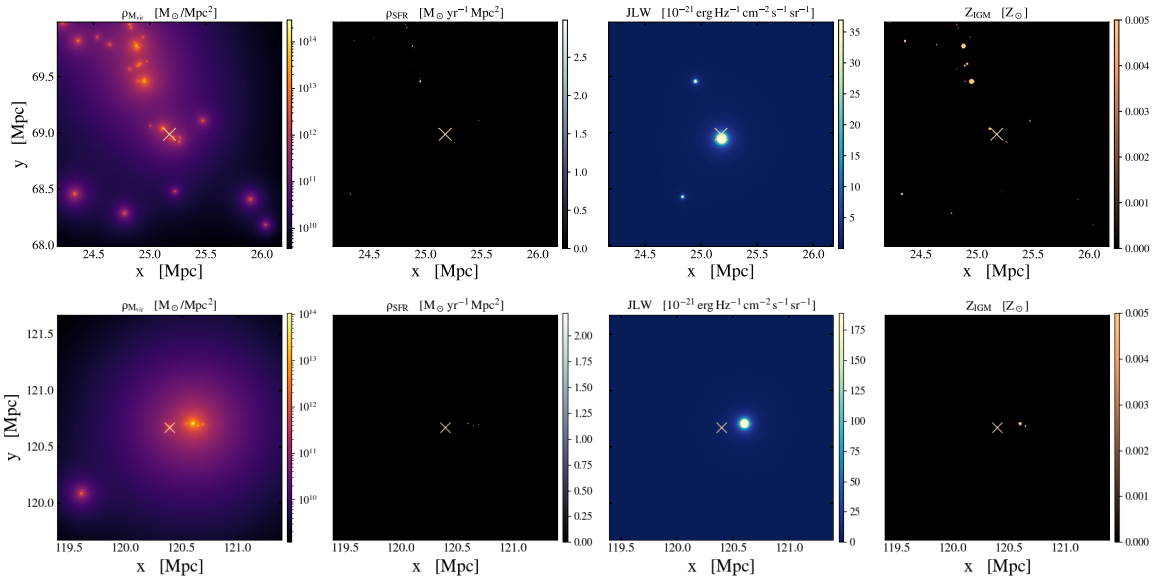

Figure 3.5: Environment of a DCBH (upper row) and a RSM (lower row) taken as representative examples. The position of these BH seeds is marked with a white X symbol at the center of each map. For each case, we show four different spatial-maps obtained by considering all halos within a $(x, y, z) = (2, 2, 0.5)$ Mpc slice, centered on the position of the BH seed. From left to right, we show: the projected-density of M_{vir} and SFR (first and second panel), the local intensity of LW flux, as received by the BH seed and the local level of chemical enrichment. The projected-density maps (namely: first and second panel from the left) are obtained by smoothing the relative quantities on adequate spatial kernels. For M_{vir} we use a standard Navarro, Frenk & White profile, while we smooth the SFR and $Z_{IGM\,;\,local}$ of each galaxy on a simple 2D, circular step-function with a radius equal to the galaxy-disk radius (for the SFR) and to the metallic-shell radius (for $Z_{IGM\,;\,local}$). We then integrate the smoothed quantities along the z coordinate to obtain the maps. Finally, the LW-flux map is obtained by summing the flux received in each point of the map from each source actively contributing to the $J_{LW\,;\,local}$ value.

LW sources contributing to DCBHs formation

Our pure-DCBH seeds form in relatively denser and more active environments, where recent SF is able to provide LW photons without yet polluting the IGM with metals. This is in line with recent works suggesting that DCBHs might form preferably in close halo-pairs (see e.g. Dijkstra et al., 2008; Agarwal et al., 2013; Visbal et al., 2014b; Agarwal et al., 2017; Regan et al., 2017; Agarwal et al., 2019; Lupi et al., 2021a), and supports this picture statistically over a wide, cosmological box. To this purpose, Fig. 3.6 illustrates the comparison between $J_{LW\,;\,local}$ at the sites of DCBHs formation (red dots) and the global $J_{LW\,;\,bg}$ (magenta line and dots, see also §3.2.1), showing that strong, local J_{LW} variations of >1 dex with respect to $J_{LW\,;\,bg}$ are generally responsible for the formation of DCBHs (Shang et al., 2010; Agarwal et al., 2012; Visbal et al., 2014a; Fernandez et al., 2014; Regan et al., 2014; Latif et al., 2015; Latif & Volonteri, 2015; Regan & Downes, 2018; Maio et al., 2019). These local variations are preferentially due to a single, actively star-forming neighbor which is able to influence its surroundings with strong J_{LW} fluxes. We check this scenario quantitatively by computing the median number of LW contributors per dex of *received* J_{LW} at the position of each new pure-DCBH formed in the MR-II cosmological box, independently of its formation redshift. This is shown in the inset panel of Fig. 3.6, where the vertical error bars mark the 16th and 84th percentiles of the J_{LW} distributions in each bin (whose widths are highlighted with horizontal error bars). Statistically, the critical threshold J_{crit} (vertical dashed red line) is surpassed thanks to only one, luminous LW contributor, hence supporting the idea according to which DCBHs might preferentially form in close halo-pairs (e.g. Dijkstra et al., 2008; Visbal et al., 2014b; Yue et al., 2014; Agarwal et al., 2019). Indeed, when considering the whole population of L-Galaxies DCBHs formed in our model (i.e. at any z), we find that 127 DCBHs over 152 ($\sim$84%) receive $J_{LW\,;\,local} > J_{crit}$ thanks to a single, luminous LW contributor. Of the remaining DCBHs, 15 ($\sim$10%) need 2 LW contributors, while only 10 ($\sim$ 6%) need at least 3 contributors.

To provide an illustrative example, in Fig. 3.7 we show the properties of all LW contributors concurring to the $J_{LW\,;\,local}$ measured at the position of the DCBH seed with the highest number of LW contributors in its past light-cone. We clearly note that J_{crit} is surpassed thanks to a single, luminous contributor which

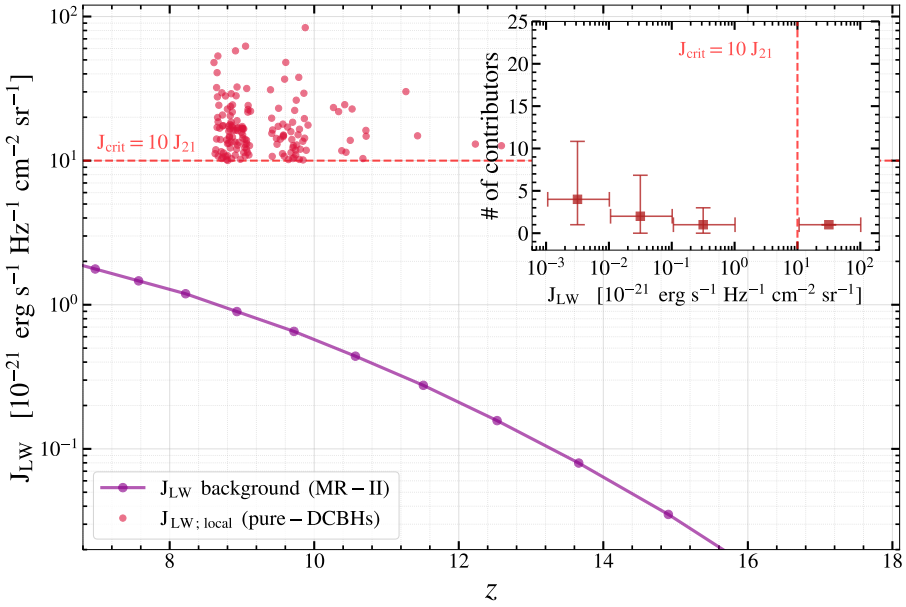

Figure 3.6: *Main panel*: Comparison between the values of $J_{LW\,;\,local}$ received by each newly-formed DCBH (red dots) and the background level of J_{LW} (magenta line and dots, see also §3.2.1). The latter is generally not sufficient to sterilize high-z hydrogen clouds against H_2. *Inset panel*: median number of LW contributors (see §3.2.4 for this definition) per decade of $J_{LW\,;\,local}$ *received* by the newly-formed DCBHs. Statistically, $J_{LW\,;\,local} > J_{crit}$ only thanks to a single luminous neighbor (last square to the right), while the higher number of faint contributors ($\gtrsim$10, on average) only manage to provide faint $J_{LW\,;\,local}$ to the forming DCBH (i.e.: $\sim 10^{-3} J_{21}$).

lies outside the distribution of both $J_{LW\,;\,local}$ vs SFR and $J_{LW\,;\,local}$ vs M_*. Therefore, even if numerous, distant LW-contributors generally do not provide enough illumination to overpass J_{crit} (see e.g., Yue et al., 2014). This picture is supported by the redshift distribution of LW contributors (color-coded crosses in both panels of Fig. 3.7). Indeed, the stronger contributors are also closer in time to the DCBH formation event. This is in line with recent works which focused on the distance between potential LW contributors and DCBH formation sites (e.g., Visbal et al., 2014b; Wolcott-Green et al., 2017; Regan et al., 2017; Agarwal et al., 2019).

3.4.3 SMBH population properties at $z = 0$

As anticipated in §2.1.2, we use the BHs growth prescriptions detailed in Izquierdo-Villalba et al. (2020), which were calibrated on the MR merger trees in order to reproduce the build-up of SMBH masses in time, down to $z \sim 0$, as determined by gas accretion and BH mergers. We stress that, in our model, SMBH seeds can only form until $Z_{IGM\,;\,local}$ and $J_{LW\,;\,local}$ allow the activation of any seeding channel or until the mass resolution of GQd is higher than the MR-II one (see §3.3.1 and §3.3.2). Therefore, a key aspect of our model is that any property of the $z < 6$ SMBH population is a consequence of the self-consistent evolution of BHs formed exclusively at high-z through physically-motivated seeding prescriptions.

Figure 3.8 shows the comparison between the BHMF obtained with two different runs of L-Galaxies performed with the same parameters except for G_P (see §3.3.2). These bracket the efficiency of our light-seeding channel by imposing two different levels of grafting probability, namely: $G_P = 1$ and $G_P = 0.01$ (respectively solid black and purple lines). In both cases, our predictions at $z = 0$ are in good agreement with the determinations of Marconi et al. (2004); Shankar et al. (2004, 2009) and Shankar (2013), especially at $M_\bullet \gtrsim 10^7\,M_\odot$. This shows that the integrated growth of SMBHs down to low-z is well-predicted by the Izquierdo-Villalba et al. (2020) model also on MR-II merger trees. Nevertheless we emphasize that, differently from their work, we only model the formation of BH seeds under the specific conditions required by our seeding model, which are generally satisfied only at $z \gtrsim 6$. To our knowledge, ours is the first semi-analytic code in which a physically-motivated, high-z BH-seeding model, applied to a cosmological context and coupled to refined mass-growth prescriptions, produces a SMBH population consistent with $z \sim 0$ BHMF constraints.

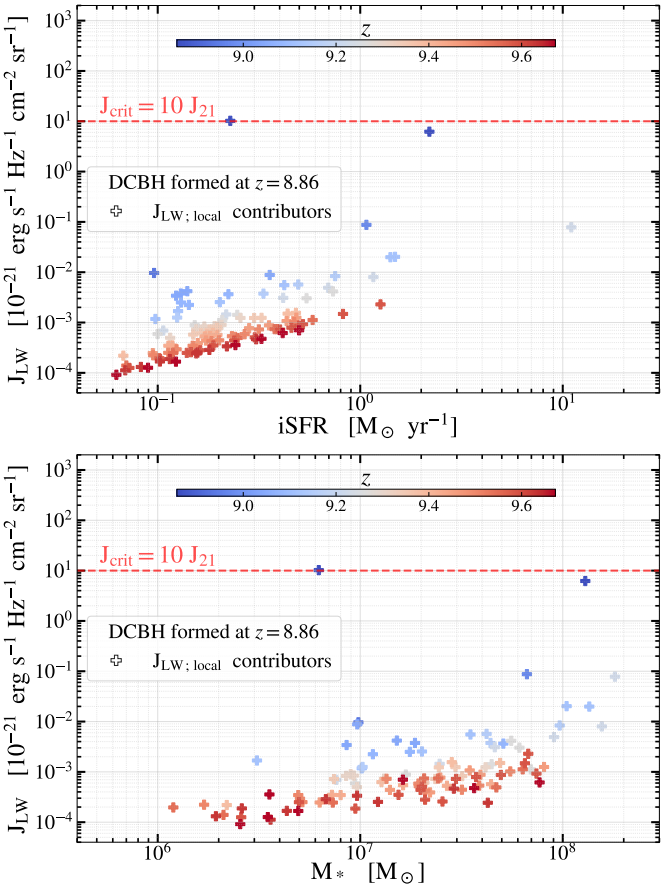

Figure 3.7: Properties of $J_{LW\,;\,local}$ contributors for one DCBH formed in our model, taken as example. In particular, for each contributor, we show the $J_{LW\,;\,local}$ provided to the forming DCBH versus the SFR (*upper panel*) and M_* (*bottom panel*). We underline that $J_{LW\,;\,local}$ is the level of LW flux *received* by the forming DCBH (i.e. at the time and location of its formation), while SFR and M_* are the intrinsic values displayed by each contributor at the time and position of the LW photons emission.

We note that the effect of varying G_P manifests at $M_\bullet \lesssim 10^8\,M_\odot$, hence suggesting that this mass regime can be discriminating towards the efficiency of light-seeds formation at high-z in our model (e.g. Sesana et al., 2007; Volonteri et al., 2008; Valiante et al., 2021). In other words, depending on the specific details of the BH-growth model, the abundance of BH seeds can effectively leave an imprint on the BHMF low-mass end. This suggests that the commonly followed approach of initializing a SMBH seed in every newly-resolved structure, independently of redshift (see e.g., Malbon et al., 2007; Fanidakis et al., 2011; Henriques et al., 2015; Lacey et al., 2016; Qin et al., 2017; Cora et al., 2018; Marshall et al., 2020a; Izquierdo-Villalba et al., 2020) might lead to over-estimations of this low BH-mass regime. Interestingly, the effect of G_P becomes negligible at $M_\bullet \gtrsim 10^9\,M_\odot$, showing that the details of our grafting procedure are washed-out by the cosmological growth of SMBHs only at the highest BH masses. This is especially noteworthy if considering that in our model BHs never form at $z \lesssim 6$ (except for the negligible contribution of miDCBHs). This supports the idea that mass-accretion induced by the hierarchical assembly of their hosts primarily drives the growth and evolution of SMBHs (see e.g., Kauffmann & Haehnelt, 2000; Malbon et al., 2007; Wyithe & Loeb, 2009; Fanidakis et al., 2011); at least for those populating the high-mass end of the $z \sim 0$ BHMF.

In Fig. 3.9 we show the BHMF of our run with $G_P = 1$, split into the contributions provided by different seed flavours (solid colored lines). It is evident that the descendants of `GQd` light seeds are the most numerous types at $z = 0$ and across the whole $M_\bullet$ distribution. On the contrary, heavy and RSM seed-types only provide a minor contribution to the total BHMF at all $M_\bullet > 10^5 M_\odot$. This suggests that, in our model, the mixing of seed-types is not a necessary condition to populate the massive end of the $z = 0$ BHMF, although the relative contribution of heavy types is larger for larger $M_\bullet$ (see Fig.3.9). This finding can be explained by the strong prevalence of light seed-types in our SMBH population, together

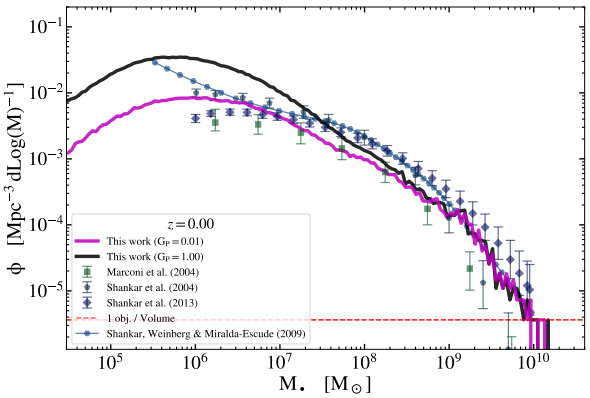

Figure 3.8: Black Hole mass function (BHMF) at $z = 0$ for the same model runs discussed in §3.4.5, namely: $G_P = 1$ and $G_P = 0.01$ (respectively, solid black and purple lines). Both model runs are in good agreement with observational determinations (blue diamonds and pentagons as well as green squares), although the $G_P = 0.01$ run is closer to the data at $M_\bullet \lesssim 10^6 M_\odot$. This suggests that the abundance of GQd light seeds critically impacts the mass distribution of $z \sim 0$ SMBHs.

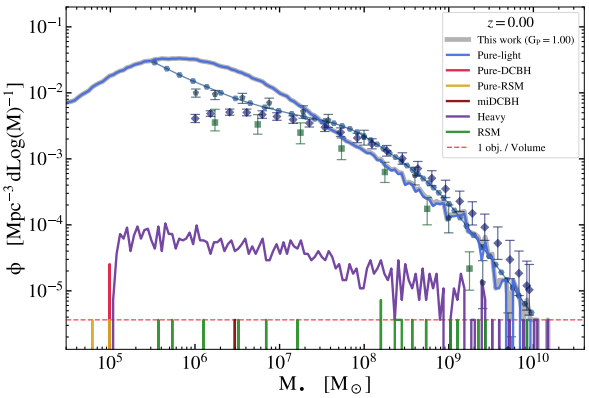

Figure 3.9: BHMF at $z = 0$ split into the contribution of different seed-types descendants (solid colored lines, as shown in the figure legend). GQd light seeds are by far the most abundant class of seeds at $z = 0$, with the remaining seed-flavours providing only a minor contribution. This shows that, in our model, the already-evolved GQd light seeds are generally sufficient to reach $M_\bullet \gtrsim 10^8 M_\odot$ at $z = 0$, although the contribution of heavy seeds is increasingly high at $M_\bullet > 10^5 M_\odot$.

with the model for BH growth we employ (e.g., Izquierdo-Villalba et al., 2020). Indeed, the efficient gas-fuelling of central BHs induced by galaxy mergers and the frequency of the latter processes in the MR-II merger tree (see Izquierdo-Villalba et al., 2019b) allow the light seeds inherited by GQd to grow up to $\sim 10^{10} M_\odot$ by $z \sim 0$, without necessarily requiring the contribution of intermediate or heavy seeds.

This picture contrasts with what generally found by BH formation models (see e.g., Valiante et al., 2016; Lupi et al., 2021b; Sassano et al., 2021; Valiante et al., 2021), especially when focusing on the origin of the first SMBHs at $z > 6$. This tension can be understood in terms of the interplay between L-Galaxies and GQd models. Indeed, the latter generally provides already-evolved light-seeds descendants with masses of $10^2 \lesssim M_\bullet/M_\odot \lesssim 10^5$, as shown by the solid light-blue lines in the bottom-right panels of Fig. 3.10. Despite these already-evolved SMBHs actually descend only from pure-light seeds formed at earlier times in GQd, they act as *effective* intermediate or heavy seeds for newly-initialized L-Galaxies structures. Therefore, they only need few merger-driven accretion episodes in L-Galaxies to reach $M_\bullet \gtrsim 10^7 M_\odot$ already at $z \sim 6$. This is clearly shown in the upper-row panels of Fig. 3.10 which detail the build-up of the BHMF during the transition period in which our grafting procedure ends (i.e. around $z \sim 8$). Indeed, after being inherited from GQd and distributed over the cosmological environments of the MR-II box, pure-light seeds descendants (solid light-blue lines) are able to efficiently grow and cover a relatively wide $M_\bullet$ range already by $z \sim 6$ (upper-left panel of Fig. 3.10). This further shows that the interplay between GQd grafting, the formation of BHs in L-Galaxies and the mass-growth model of SMBHs ultimately produce a prevalence of pure-light seeds descendants at the massive-end of the BHMF predicted by our model. We further comment this point in our discussion (§5).

As discussed by e.g. Volonteri et al. (2008); van Wassenhove et al. (2010), the overall efficiency of BH-seeding processes is expected to manifest on the BH occupation fraction at $z = 0$. We show this prediction in Fig. 3.11 as a function of the stellar mass of SMBH hosts, for the same two runs shown in Fig. 3.8. We compare our results to the observations of Miller et al. (2015), which offer relatively solid constraints at $M_* \gtrsim 10^{10} M_\odot$. Differently to what observed for our BHMF, the $G_P = 0.01$ run worsen the agreement of our model with observations, especially at high stellar masses. On the contrary, the Miller et al. (2015) estimates at $M_* \lesssim 10^9 M_\odot$ are less stringent, thus leaving open the possibility to tune the abundance of SMBHs in dwarf galaxies, within our model. We comment further this latter point in our discussion (§5), since the accurate calibration of our model lies beyond the scope of the current work.

It is interesting to explore whether the different evolution and mass-growth histories of SMBHs belonging to different seed-type classes are reflected also on the M_* of their hosts at $z \sim 0$. In the central panel of Fig. 3.12, we show the relation between M_* and $M_\bullet$ for all galaxies in the MR-II box hosting a central SMBH at $z = 0$ and with $M_{vir} > 3 \times 10^8 M_\odot$ (see §3.4.5). We present the median M_*-to-$M_\bullet$

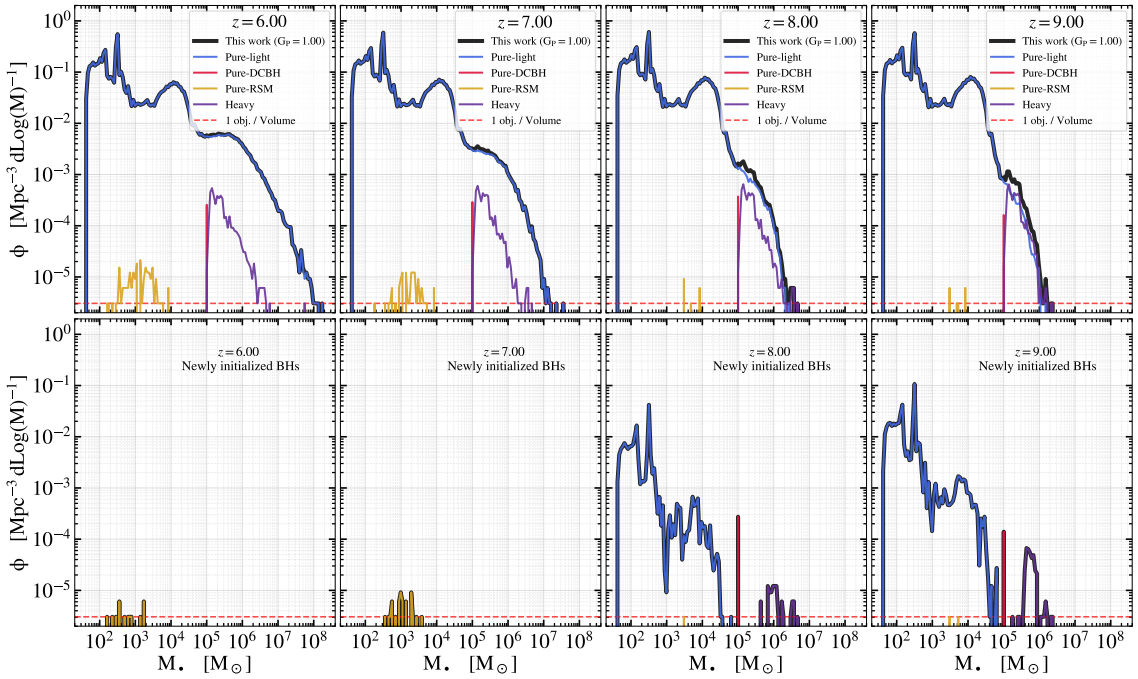

Figure 3.10: *Upper row*: build-up of the BHMF at $6 < z < 9$, encompassing the final epochs of BH seeding in our model. Indeed, both the GQd grafting (two rightmost panels) and the formation of pure-RSM seeds in L-Galaxies (leftmost panels) end by $z \sim 6$. *Lower row*: same as upper row but only for newly-formed (or newly-inherited) BH seeds. These plot shows that GQd provides already-evolved pure-light seeds descendants with masses up to $M_\bullet \lesssim 10^5 M_\odot$. This population inherited by GQd is able to grow via mergers and efficient gas accretion, so that by $z \sim 6$ it already provides a major contribution to the BHMF at all masses, including its high-mass end.

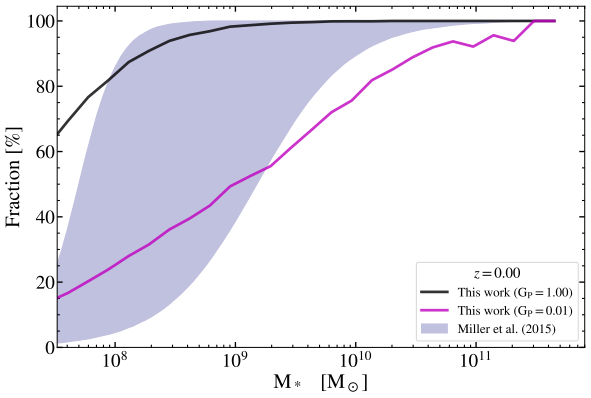

Figure 3.11: Black Hole occupation fraction as a function of stellar mass, for the two runs with $G_P = 1$ (solid black line) and $G_P = 0.01$ (solid magenta line). The relatively stringent determinations of Miller et al. (2015) at $M_* \gtrsim 10^10 M_\odot$ (blue shaded area) favour a high value of our G_P parameter, although the latter over-predicts the abundance of SMBHs in dwarf galaxies ($M_* \lesssim 10^9 M_\odot$) at $z = 0$. The comparison with our $G_P = 0.01$ run shows that by acting on the abundance of SMBHs inherited by GQd we can improve the consistency of our results with observational constraints on different M_* intervals.

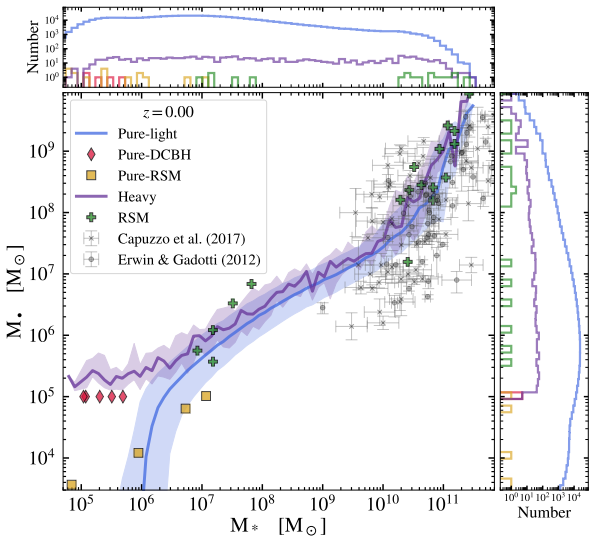

Figure 3.12: *Central panel*: relation between M_* and $M_\bullet$ for all galaxies with $M_{vir} > 3 \times 10^8 M_\odot$, hosting a central SMBH at $z = 0$. We present the median relations for different seed-type classes (solid colored lines) with the corresponding $16^{th} - 84^{th}$ percentiles (shaded areas) or the position of single SMBHs for the least abundant seed-types classes in the MR-II box. *Upper panel*: histogram of M_* split into its different seed-types, hence showing how the latter contribute to different M_* intervals. *Right panel*: same as in the upper panel but for $M_\bullet$. Our results at high masses are comparable with the observational determinations of Erwin & Gadotti (2012) and Capuzzo-Dolcetta & Tosta e Melo (2017).

relation for each seed-type population with their corresponding 16^{th} and 84^{th} percentiles (respectively, colored lines and shaded areas). On the other hand, we show the individual positions of single SMBHs for the less abundant seed-type classes (i.e. red diamonds, yellow squares and green crosses). We note that the high-mass end of our M_*-$M_\bullet$ relation (i.e. at $M_\bullet \gtrsim 10^7$ and $M_* \gtrsim 10^10$) is consistent with the observational constraints of Erwin & Gadotti (2012) and Capuzzo-Dolcetta & Tosta e Melo (2017). At these high masses, our model only predicts the presence of pure-light (solid light-blue line), heavy (solid purple line) and RSM descendants (green crosses). The latter two classes are obtained from the merger of pure-DCBHs or pure-RSM with light seeds, therefore this shows that in our model the highest $M_\bullet$ at $z \sim 0$ are reached as a consequence of the hierarchical build-up of structures. Finally, pure-DCBHs and pure-RSM descendants at $z = 0$ are only hosted in extremely-low mass galaxies ($M_* \lesssim 10^7 M_\odot$). This supports the idea that BHs which are able to retain memory of their formation channel (i.e. which do not experience significant mass growth), end up in dwarf isolated galaxies by $z \sim 0$ (see e.g., van Wassenhove et al., 2010; Mezcua, 2021).

3.4.4 Imprint of the formation environment at low-z

Our model links the formation of massive BHs at high-z to their local environment. Later on, BHs eventually grow into SMBHs, mainly via gas accretion during galaxy mergers. With this in mind, it is interesting to explore whether the BH-formation requirements of different seeding channels induce differences in the properties of SMBH-seeds hosts at different epochs. Furthermore, during galaxy mergers[3] the seed-types of the central SMBHs mix, hence the distribution of seed-flavours evolves throughout cosmic times. Thanks to the N-body nature of the MR-II merger trees, our model allows to track all these features in a cosmological environment over a dynamic range which encompass dwarf satellites and Milky Way-like halos.

As discussed in §3.3.3, a fundamental difference between light, intermediate and heavy seeding scenarios is the different degree of inhibition of SF processes by photo-dissociating fluxes. To analyze whether this reflects onto the global SF history of the hosting structure, in Fig. 3.13 we show the median stellar mass (upper panel) and SFR (lower panel) as a function of redshift, for each seed-type population in our box. We normalize both quantities to the M_{vir} of the BH host to minimize biases due to the wide dynamic range we explore. Furthermore, we split the whole sample of galaxies hosting central BHs in three virial-mass bins, in order to provide a more detailed perspective of the different environments in which SMBH seeds descendants evolve. In particular, from left to right in Fig. 3.13, we show low-mass ($M_* \lesssim 10^9 M_\odot$), dwarf ($10^9 \lesssim M_*/[M_\odot] \lesssim 10^{11}$) and Milky Way-type halos ($M_* \gtrsim 10^{12} M_\odot$).

We note that pure-RSM and pure-DCBH are never hosted in massive halos (rightmost column), which are instead populated either by pure-light seeding descendants (light-blue lines) or seeds descendants with merger-driven origin (namely RSM and heavy channels, respectively yellow and purple lines). The presence of pure-light descendants within high mass halos testifies the efficient hierarchical build-up of massive BHs via subsequent mergers of pure-light seeds. Analogously to what we commented for the high-mass end of the BHMF, this is an effect of the strong prevalence of light-seeds descendants in our model.

The merger-driven origin of the RSM channel is what causes the significant drop of the Sfr-to-M_{vir} ratio (bottom-right panel), which suggests that this class of seed-descendants (in the most massive M_{vir} bin showed in Fig. 3.13) is hosted in systems that get rapidly quenched at $z < 2$. On the other hand, the heavy channel shows a less-pronounced drop of the Sfr-to-M_{vir} ratio, both in the high- and intermediate-M_{vir} bins (right and central columns). Indeed, this class can derive from both heavy-seeds contributions inherited from GQd grafting and pure-DCBHs formed in L-Galaxies. These are two different populations of BHs, hosted in different halos and environments. As shown in §3.4.2, pure-DCBHs tend to form in low-mass satellites of high-z star-forming halos, while heavy seeds descendant inherited from GQd do not form preferentially in a specific environment, by construction. This degeneracy within the heavy seed-types class explains the different results with respect to RSM seed-descendants, and also shows up in the significant presence of heavy seeds-descendants down to the lowest M_{vir} bin (leftmost column of

[3] We stress that in our version of L-Galaxies BHs merge as soon as their host merge, without the time-delay introduced by Izquierdo-Villalba et al. (2020).

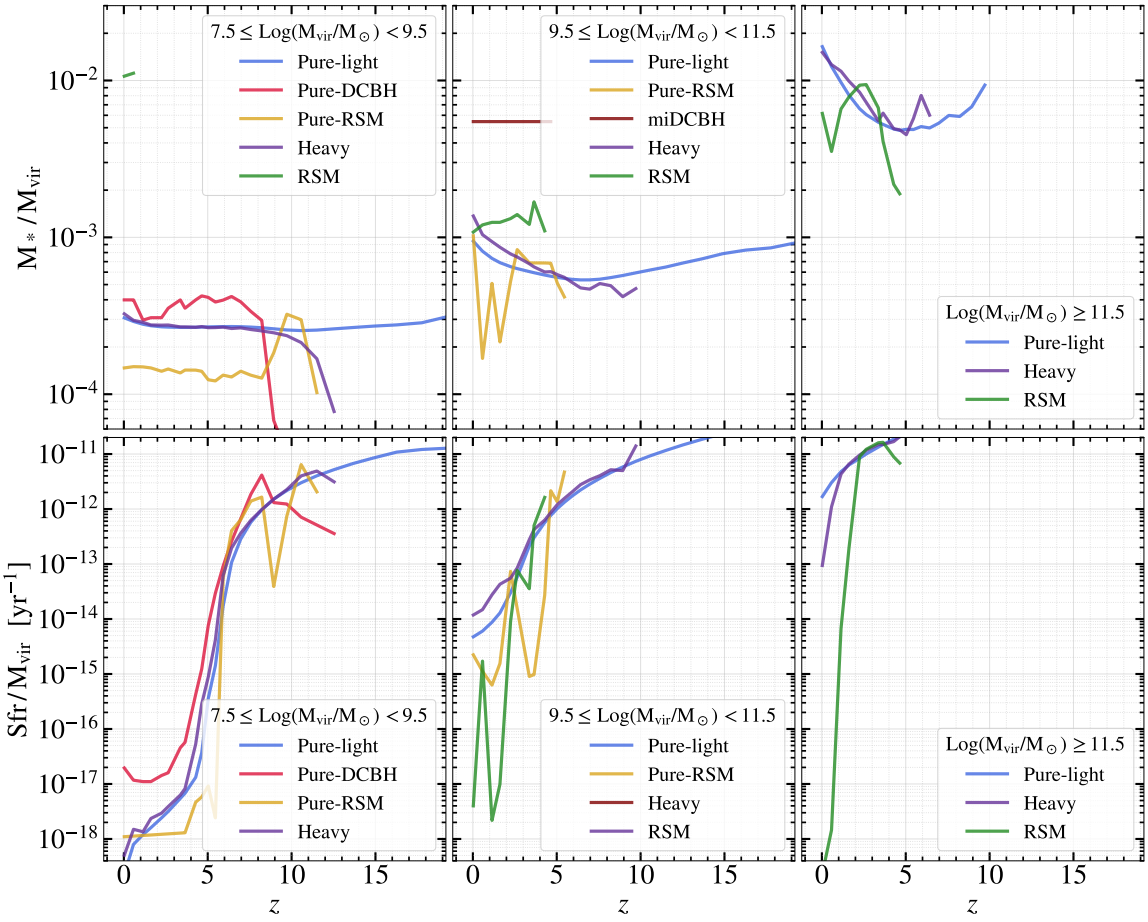

Figure 3.13: Median properties of BH-seeds hosts at different redshifts, split into different seed-types (solid colored lines). *Upper panel*: median M^*-to-M_{vir} ratio for each seed-type, showing the typical build-up history of the M_* of different seed-type classes. *Bottom panel*: median SFR-to-M_{vir} ratio, showing the intensity of SF processes with respect to the halos virial masses. We split the whole population of MR-II structures in three M_{vir} bins, in order to detail the redistribution of different seed-types among low-mass halos ($M_{vir} < 10^9 M_\odot$, left column), dwarf-like halos ($10^9 < M_{vir}/M_\odot < 10^{11}$, central column) and Milky Way-like halos ($M_{vir} \geq 10^{11} M_\odot$, right column).

Fig. 3.13), differently from RSMs.

Furthermore, pure-RSM and pure-DCBH seed descendants which are able to retain memory of their seeding channel (i.e. those which never merge with other seed-types) only populate low-mass halos. In particular, pure-DCBHs are only found at the lowest virial masses (leftmost column) and show higher M_*-to-M_{vir} ratio with respect to pure-RSM. As expected, this difference is also reflected in the Sfr-to-M_{vir} ratio distributions, since pure-DCBHs appear to be among the most star-forming halos at low mass and low-z. Although affected by the low statistics provided by the small number of these seed-types hosts (see also Fig. 3.3), this finding suggests that the formation environment of pure-DCBHs and pure-RSM might leave a significantly different imprint on their hosting galaxy. For instance, although both seeds-descendants hosts evolve in evolution for their entire history, they show significantly different SF histories. We checked that this conclusion is not driven by virial-mass differences between the two classes.

We stress that the grafting of pure-light seeds does not depend on local physical conditions. Therefore the galaxies hosting these seed-types cover a wide dynamic range, which includes a combination of actively SF and quiet galaxies with a wide variety of merging histories. Due to their high number, the probability of a merger involving two pure-light seeds is high. As a consequence, only a minor fraction of these seed-type descendants mixes with pure-RSMs or pure-DCBHs descendants during cosmic history. The combination of these factors implies that the hosts of pure-light seeds are not representative of a determined class of galaxies, with narrow mass and SFR distributions, but rather trace the average evolutionary trend of all MR-II structures.

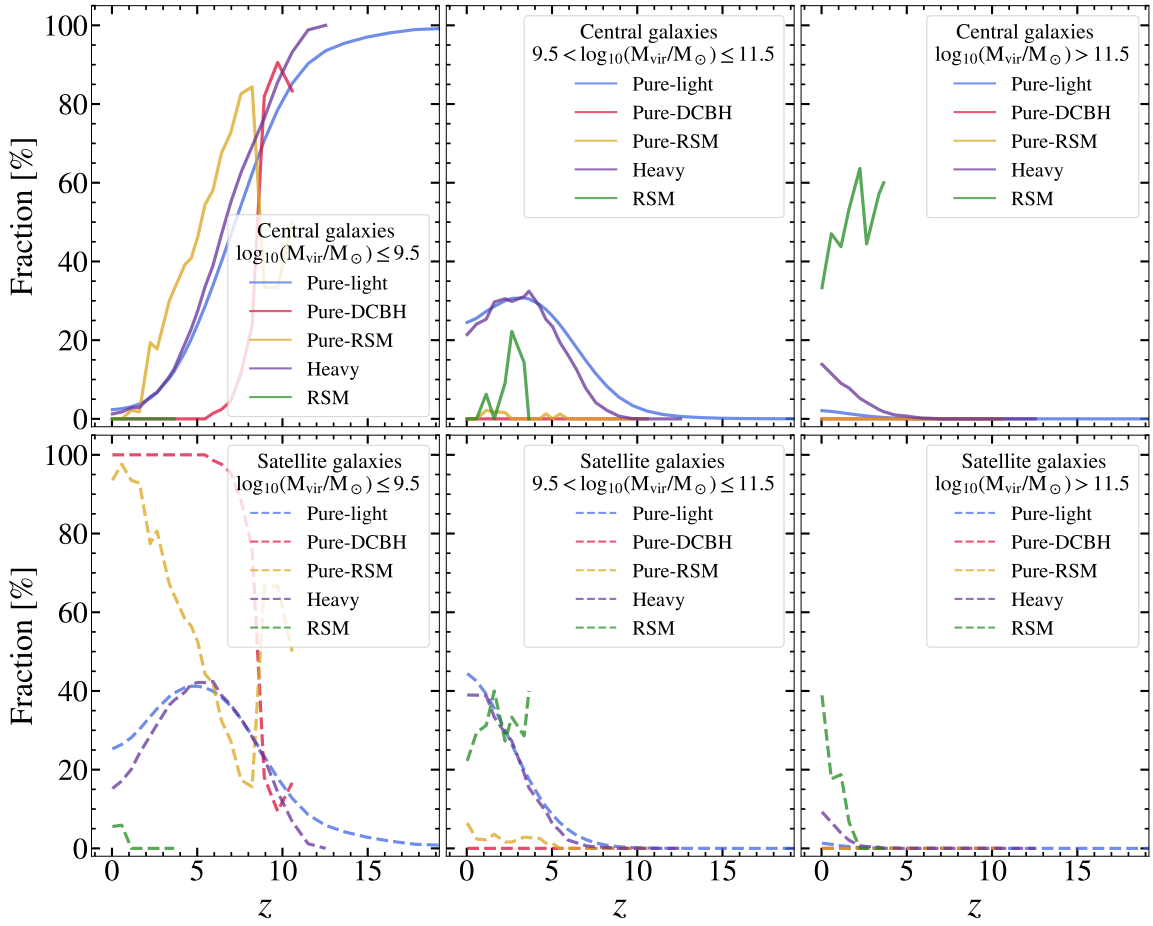

Figure 3.14: Fractions of BHs hosted in central galaxies (upper row) and satellite galaxies (lower row). Different colored lines represent descendants of different seed-type classes, as shown in the plot legend. Fractions add to 100% only if considering each seed-type separately in both upper and lower panels. As in Fig. 3.13, we split the ensemble of MR-II structures in three M_{vir} bins, from low-mass halos (left column) up to Milky Way-like halos (right column). This allows to detail the rearrangement of different seed-types in central and satellite galaxies acreoss the whole MR-II dynamic range.

The imprint of the formation environment of SMBH seeds can also be studied by considering the fraction of BHs which are hosted into central or satellite galaxies at each z, split into different seed-flavours (Fig. 3.14). We underline that, for each seed-type class, the fractions presented in the latter figure add to 100% only when considering the six panels together. In this way, this analysis provides an idea about how the different seed-types mix in time and how they distribute over the MR-II dynamic range. We note that L-Galaxies baryonic structures are initialized as the only member of their FoF, regardless of their mass, therefore (at their initialization) they are hosted by the most massive DM halo of their FoF, by construction. In other words, L-Galaxies structures are always initialized as central galaxies, which explains why the totality of pure-light seeds at $z \gtrsim 15$ is hosted in such structures (solid light-blue line in the upper-left panel of Fig. 3.14). This is also the reason for the high fractions of pure-RSMs and pure-DCBHs hosted in centrals at high-z (respectively: yellow and red lines in the top left panel), despite these seed-types preferentially form in the vicinity of actively SF galaxies (§3.4.2). We underline that the initial fluctuations in the pure-RSM and pure-DCBH hosts evolution are only due to statistical noise.

After the formation of BH seeds, the merger history of their hosts lowers the fraction of seed descendants hosted in central galaxies in the lowest M_{vir} bin (upper left panel of Fig. 3.14). In particular, this fraction rapidly drops to zero in the case of pure-DCBHs remnants. On the other hand, pure-RSM hosts follow a slightly different evolution, being their accretion as satellites delayed in time with respect to pure-DCBHs hosts. Differently from the latter, few hosts of pure-RSM descendants manage to increase their mass over $M_{vir} \sim 3 \times 10^9 M_\odot$ through either the merger with small halos not hosting BHs or the smooth accretion of un-resolved DM (see central column of Fig. 3.14). Nevertheless, the vast majority ($\sim 95\%$) of pure-RSM hosts ends up in low-mass satellites by $z \sim 0$ (bottom left panel). We note that

also the fraction of pure-light and heavy descendants (respectively, light-blue and purple lines) hosted in centrals appears to drop with respect to z in the lowest M_{vir} bin (top-left panel). Nevertheless, this evolution is accompanied by the M_{vir} growth of their hosts, as testified by the increase of pure-light and heavy descendants in the intermediate and massive M_{vir} bins (upper panels of central and rightmost columns). This is not observed for pure-RSMs and pure-DCBHs, which remain hosted by small halos during their whole evolution.

As expected, the hosts of heavy and RSM descendants which experienced mergers with pure-light remnants show a variate evolution. For instance, RSM hosts (solid and dashed green lines) are only found at $M_{vir} \gtrsim 3 \times 10^9 M_\odot$, being split as 60% in centrals and 40% in satellites at $z \sim 4$. Later on, this picture gets gradually modified by the intervening hierarchical assembly of structures, until most of the RSM hosts at $z \sim 0$ are found in massive satellites (bottom right panel). On the contrary, the hosts of heavy-seeds descendants tend to follow the evolution of pure-light descendants hosts, except for their fraction in satellites at the smallest masses (bottom left panel) and their fraction in centrals at the highest masses (top right panel). These two hints, together with the drop of central pure-light hosts at intermediate masses (upper central panel) suggest that the increase of massive central hosts of heavy seed-types is driven by the merger of low-mass satellites hosting heavy descendants (either heavy or pure-DCBHs) with moderately massive centrals hosting pure-light remnants. Overall, both heavy and RSM hosts are predominant at high M_{vir} and $z < 5$, as a consequence of their merger-driven origin.

3.4.5 The mass growth of SMBH seeds

The luminosity function (LF) of accreting BHs is a fundamental statistical tool to constrain and describe the growth of SMBHs through cosmological time (Wyithe & Loeb, 2003; Wyithe & Padmanabhan, 2006; Hopkins et al., 2007; Croom et al., 2009; Ross et al., 2012; Aird et al., 2015; Kulkarni et al., 2019; Shen et al., 2020). Thanks to the prescriptions for prolonged gas-accretion onto SMBHs of Izquierdo-Villalba et al. (2020), we can compute the bolometric luminosity (L_{bol}) of accreting BHs at each z and determine whether our BH-formation model is able to explain the observed properties of active SMBHs. Furthermore, as introduced in §3.3.2, we can modulate our BH-seeding prescription in order to study the role of high-z BH formation on the lower-z properties of active BHs.

For the present work, we neglect the effects of BH-spin evolution, BH-BH merger-delay and GW-induced recoil (after BH-BH mergers) analyzed in Izquierdo-Villalba et al. (2020). Indeed, we plan to dedicate a future analysis to the study of the complex interaction between these prescriptions, the implementation of our BH-formation model and the cosmological evolution of SMBHs in L-Galaxies (see the discussion in §5.1.4). Here we limit to present the main predictions of our BH-formation and growth model, namely the build-up in time of the $M_\bullet$ of our global SMBH population through episodes of nuclear activity described by their L_{bol} function (bLF, see §1.5.1), X-ray LF (XLF) as well as their BH accretion rate (BHAR).

Growth and activity of BH-seeding remnants

Figure 3.15 shows the bLF at four representative redshifts: $z \sim 0, 1, 2$ and $z \sim 4$, spanning the cosmological evolution of SMBHs after the epoch of BH-seeding. We use the same runs as in §3.4.1 in order to bracket the effect of G_P. The horizontal, dotted red line in each panel of Fig. 3.15 shows the number density obtained by the presence of one, single object in the whole MR-II box. This limitation only allows to sample the faint-end of the bLF, namely at $L_{bol} \lesssim 10^{46}\,erg\,s^{-1}$, at any z. Therefore, the MR-II box does not allow to sample the number densities at which $z \sim 6 - 7$ ultra-luminous QSOs are expected (e.g., Valiante et al., 2017, for a recent review). Consequently, the MR-II merger trees cannot be used in combination to our model in order to study the possible origin of these enigmatic objects. Nevertheless, our approach offers the possibility to concentrate on more common and abundant AGN which constitute the bulk of the $z < 6$ population.

At $z = 4$ and $L_{bol} \lesssim 10^{46}\,erg\,s^{-1}$, our model predicts a very high number-density of AGN with respect to recent observations (Hopkins et al., 2007; Shen et al., 2020, respectively blue dots and red pentagons). Similar discrepancies at high-z were also highlighted by recent works addressing BH-formation in a cosmological context (see e.g., DeGraf & Sijacki, 2019; Marshall et al., 2020a; Bhowmick et al., 2021).

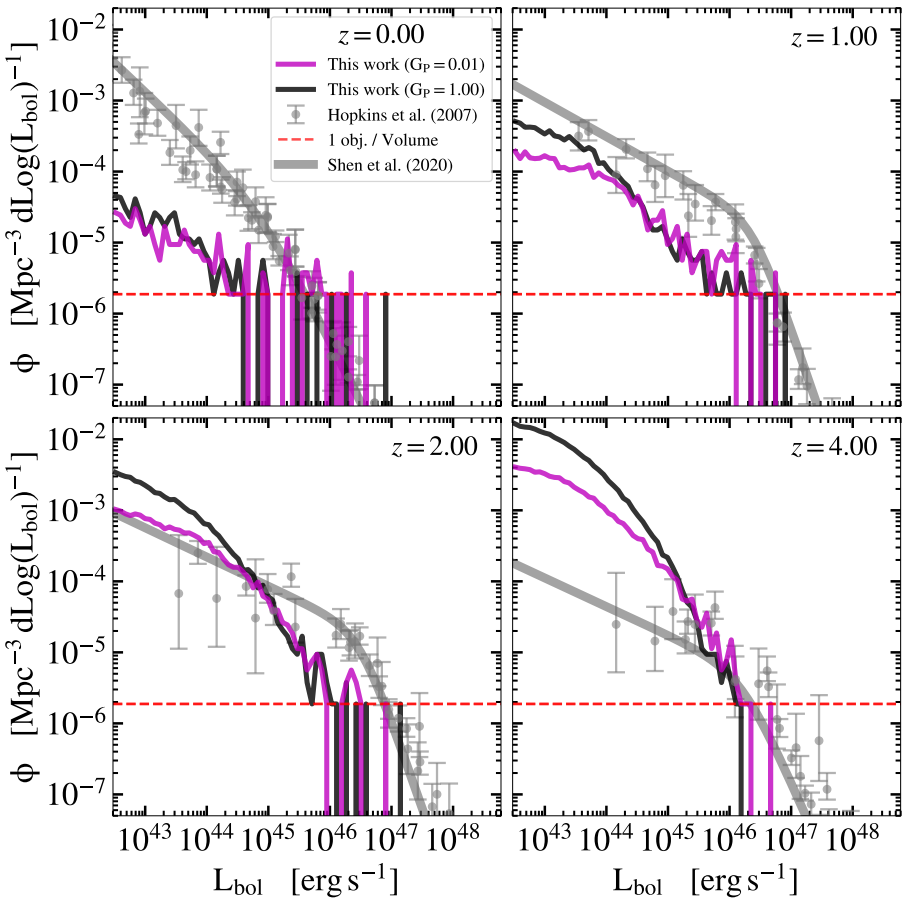

Figure 3.15: L_{bol} function at $z=0, 1, 2$ and $z=4$ (respectively from top-left to bottom-right). We show two runs obtained with different grafting probabilities (see §3.3.2), namely: $G_P=1$ (solid black line) and $G_P=0.01$ (solid purple line). The abundance of GQd light seeds in our model crucially determines the properties of the global SMBHs population at $z<4$. The dashed, horizontal red line shows the number density Φ_{limit} of one, single object in the whole MR-II box. Our bLF measurement starts to be affected by poissonian noise at roughly $\Phi \sim 10\,\Phi_{limit}$, hindering the prediction capacity of our model at $L_{bol} \gtrsim 10^{45}$ erg s^{-1}. Finally, we show the comparison with observational determinations (grey dots and solid grey line), as discussed it in the text.

In our case, this bLF feature is mainly driven by relatively low-mass SMBHs (i.e. $M_\bullet \lesssim 10^7 M_\odot$) formed as GQd light seeds, which accrete at the Eddington limit ($\lambda_{Edd}=1$) during their first mass-growth phases (see Izquierdo-Villalba et al., 2020). The comparison between the $G_P=1$ and $G_P=0.01$ runs shows that the abundance of GQd light-seeds descendants impacts the shape of the bLF at $L_{bol} \lesssim 10^{44}$ erg s^{-1}, $L_{bol} \lesssim 3\times10^{44}$ erg s^{-1} and $L_{bol} \lesssim 10^{45}$ erg s^{-1} respectively at $z=1$, $z=2$ and $z=4$. On the other hand, these differences completely vanish at $z=0$, where our bLFs severely differ from current constraints.

We note that a similar behavior is shown by our XLF, obtained from converting the L_{bol} predicted by our model to X-ray luminosity with the bolometric corrections of Marconi et al. (2004). In particular, the faint-end of our XLF is over-estimated at $z>2$ and under-estimated at $z<2$ by our model, as shown by the solid black line in Fig. 3.16. A very similar result was recently highlighted by Marshall et al. (2020a) while addressing the outcomes of the MERAXES SAM. In their work, this issue was tackled by employing an inclination-angle dependent obscuration-law for AGN, in order to reconcile the model predictions with observations. We checked that by employing observationally-motivated obscuration laws (i.e. those suggested by Ueda et al., 2014; Merloni et al., 2014, respectively shown as dotted and dashed-dotted black lines in Fig. 3.16), we are unable to reconcile our predictions with current constraints. We ascribe this issue to the interplay between the biases introduced by our grafting procedure (see the discussion in §3.4.1) and the details of the BH-growth model of Izquierdo-Villalba et al. (2020). Indeed, both these aspects of our model need to be modified in order to describe the cosmological evolution of SMBHs over the MR-II dynamical range. We further comment this topic in our discussions (see §5.1) since the improvement of the BH-seeding and growth models we employ need a separate, dedicated analysis.

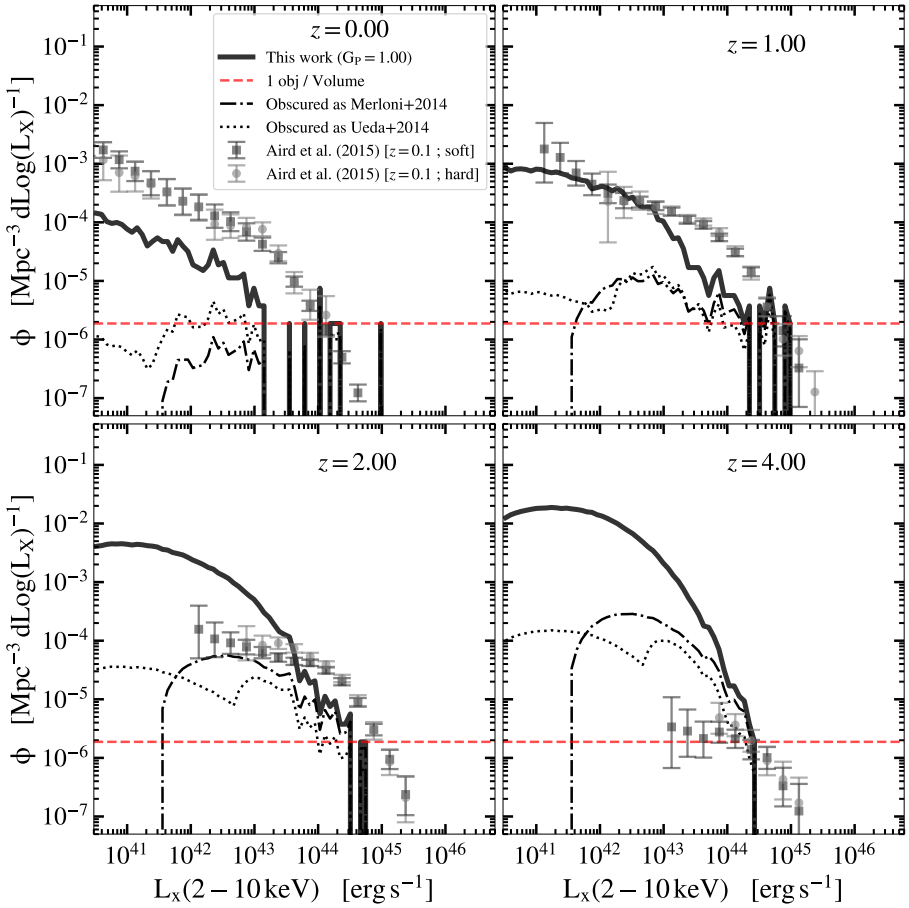

Figure 3.16: X-ray LF (solid black line) predicted by our model at four different redshifts (as in Fig. 3.15). We obtained L_X by exploiting the bolometric correction of Marconi et al. (2004). Current constraints from Aird et al. (2015) are shown as grey dots for either hard and soft X-ray bands (see plot legend in the top-left panel). We note that our model significantly over-predicts the faint end of the XLF, especially at $10^{42} < L_X/[\mathrm{erg\,s^{-1}}] < 10^{44}$ and $z \geq 2$ (where observational constraints are available). The situation reverts at lower redshift, where our predictions lie below current observations. Finally, we note that the obscuration models of Ueda et al. (2014) and Merloni et al. (2014), respectively dotted and dotted-dashed black lines, help to improve our results at high-z but significantly worsen our predictions at $z \geq 1$.

The tension between the mass-growth history predicted by our model and current observational constraints is also apparent in the evolution of both the BH mass density and the BHAR density, namely: $\rho\,(\mathrm{M_{BH}}) = \mathrm{M_{BH}}/\mathrm{V_{box}}$ and $\rho\,(\mathrm{BHAR}) = \mathrm{dM_{BH}/dt} \cdot 1/\mathrm{V_{box}}$. Indeed, as shown in Fig. 3.17, the overall normalization of both quantities is overestimated by our model, particularly by the $\mathrm{G_P} = 1$ run. This is again an effect of our grafting prescription, which extends the GQd BH-occupation to the whole dynamical range of MR-II structures. Indeed, the result of this procedure is to initialize the SMBH population of L-Galaxies with the one modelled by GQd over a specific and relatively-biased environment (i.e. the merger tree of a DM halo reaching $\mathrm{M_{vir}} = 10^{12}\mathrm{M_\odot}$ at $z = 2$). Furthermore, our grafting necessarily happens independently of the local variations of $\mathrm{Z_{IGM\,;\,local}}$ and $\mathrm{J_{LW\,;\,local}}$, which further contributes to over-estimate the formation of SMBHs within the cosmological box of MR-II.

At this point it is worth stressing that our $z = 0$ BHMF at $\mathrm{M_{BH}} \gtrsim 10^7\,\mathrm{M_\odot}$ is in fairly good agreement with current observations (see Fig. 3.8). Therefore, the total mass-growth of SMBHs predicted by our model, when integrated over their whole cosmological evolution, is consistent with current observations. When combined with the results shown in figures 3.15 and 3.16, this argument suggests that our model tends to anticipate in time, with respect to current constraints, the mass-growth of SMBHs which reach $\mathrm{M_{BH}} \gtrsim 10^7\,\mathrm{M_\odot}$ at $z = 0$. This is confirmed by the evolution of $\rho\,(\mathrm{M_{BH}})$ and $\rho\,(\mathrm{BHAR})$ shown in Fig. 3.17. In particular, we note that the evolution of $\rho\,(\mathrm{BHAR})$ for both the $\mathrm{G_P} = 0.01$ and $\mathrm{G_P} = 1$ runs would place the peak of AGN activity at $z \sim 4$, hence anticipating it with respect to the observational constraints of Shen et al. (2020), Aird et al. (2015) and Delvecchio et al. (2020).

Overall, these results show that our mass-growth prescriptions need to be adapted to the MR-II merger

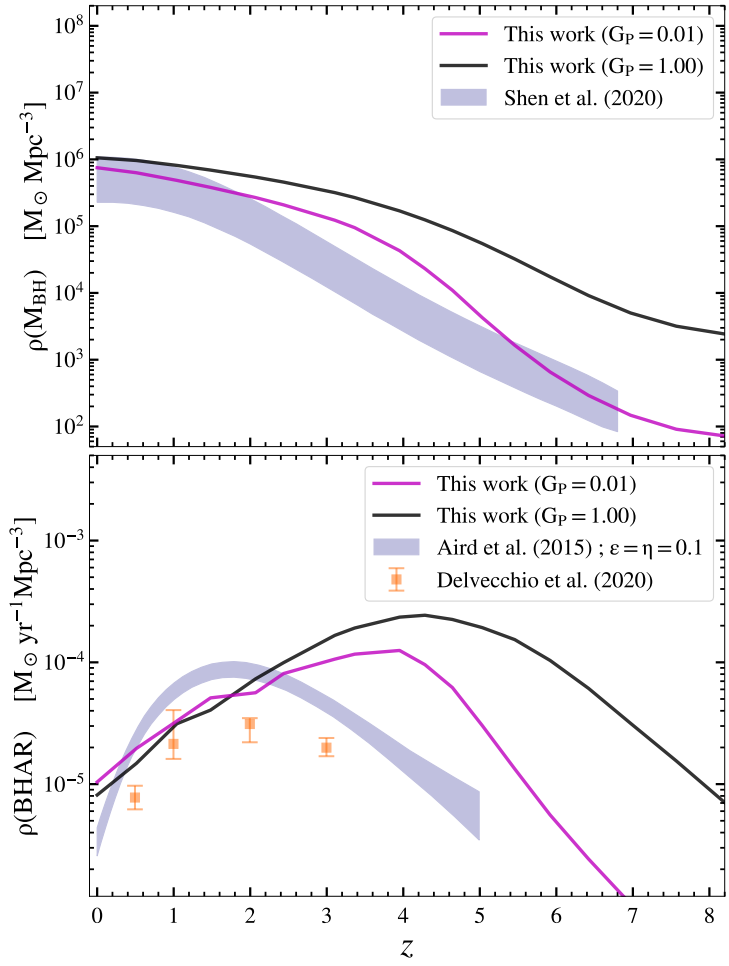

Figure 3.17: Spatial density of $M_\bullet$ and of the $M_\bullet$ accretion-rate (BHAR) onto SMBHs (respectively *upper* and *bottom* panels), computed over the whole MR-II box as a function of redshift. We show the same model runs as in Fig. 3.15, namely: $G_P = 1$ and $G_P = 0.01$ (respectively solid black and solid purple lines). These quantities provide insight about the cosmological mass-growth of SMBHs, showing that BHs in our model grow too rapidly at $z > 2$ with respect to current observational constraints (blue shaded area and dots, respectively in the upper and lower panels).

trees with respect to the version calibrated on the MR, presented in Izquierdo-Villalba et al. (2020). In addition, the comparison between the $G_P = 1$ and $G_P = 0.01$ runs suggest that this parameter might help to break the degeneracy between the efficiency of BH-seeds formation at very high-z and their subsequent growth history (e.g., Volonteri et al., 2008). Indeed, G_P can play a crucial role in improving our predictions for the bLF faint-end at $z \gtrsim 1$ as well as for ρ(BHAR) and $\rho(M_\bullet)$. We further discuss this topic in our conclusions (see§5.1), since the overall improvement of our model and the effect of BH-seeding prescriptions on the growth-history of SMBHs need a separate, detailed analysis.

Ungrown BHs

Interestingly, a fraction of BHs never grow from their initial M_{seed} in our model. According to the Izquierdo-Villalba et al. (2020) prescriptions, the mass growth of SMBHs is primarily driven by either galaxy interactions or secular processes (such as bar-induced inflows; e.g., Anglés-Alcázar et al., 2017a; Li et al., 2017; Jung et al., 2018), which fuel material onto the accretion disk of BHs at the centre of merger remnants. For the MR-II case, the growth of BHs is largely driven by mergers between relatively massive, central galaxies and small satellites (as also shown by Izquierdo-Villalba et al., 2019b, 2020). As a consequence, galaxies associated to *ungrown* BHs are generally isolated structures who never underwent an interaction (not even minor mergers) throughout their evolution. In Fig. 3.18, we show the relative importance of ungrown-BHs within each seed-class population, as a function of the seeds-host M_{vir}, at four different z (as in figures 3.15 and 3.16). The M_{vir} interval marked with a shaded dark-grey area

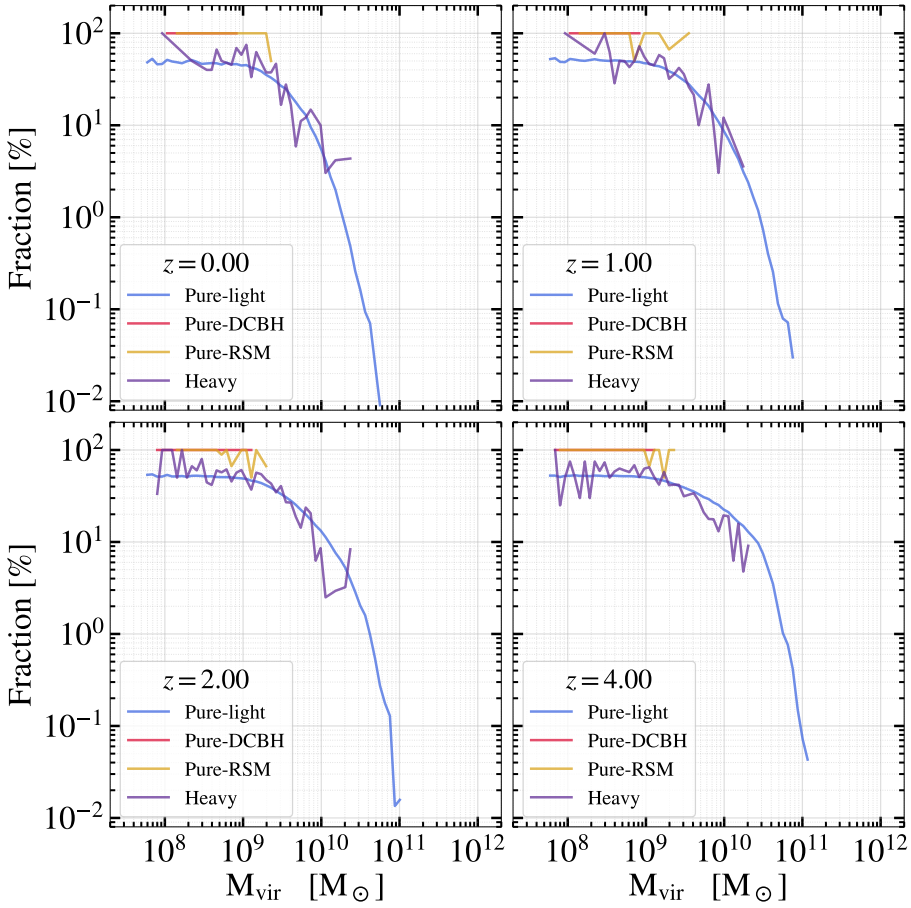

Figure 3.18: Fraction of ungrown BHs (i.e. with $M_\bullet = M_{seed}$ at $z = 0, 1, 2$ and $z = 4$ (as in Fig. 3.15). We only show the run with $G_P = 1$ as a solid black line, splitting the contributions of different seed-types (see plot legend and §3.4.1 for details). The main contribution to the population of ungrown BHs is due to GQd light seeds, bieng these numerically predominant on any other seeding flavour.

refers to virial masses which might be affected by low-resolution effects, namely: at $M_{vir} < 3 \times 10^8 M_\odot$ (corresponding to 25 MR-II particles). We note that *all* the pure-DCBHs which do not mix with other seed-types at each redshift show-up as ungrown BHs, confirming the general picture discussed in §3.4.4 according to which these BHs at $z = 0$ are hosted in galaxies with a very quiet evolution. A similar argument can be made for pure-RSM descendants, for which the ungrown fraction is above 50% at all masses and redshifts. On the contrary, both heavy and pure-light descendants show generally lower fractions of ungrown BHs ($< 50\%$) with a clear decreasing trend towards high M_{vir}. Finally, the overall ungrown fractions for these two latter classes of seed-types diminish with redshift, as an effect of the hierarchical growth of structures.

In order to contextualize the population of ungrown BHs, in Fig. 3.19 we present their contribution (dotted black line in all panels) to the whole BHMF predicted by our model ($G_P = 1$ run). Ungrown BHs become the dominant population at $M_\bullet \lesssim 3 \times 10^4 M_\odot$ for all the redshift we show. In particular, this result at $z \sim 0$ suggests that intermediate-mass BHs ($M_\bullet \lesssim 10^4 M_\odot$) hosted into local dwarf galaxies might still carry significant information about high-z BH-seeding processes (e.g., Volonteri et al., 2008; Greene, 2012; Cann et al., 2021; Mezcua, 2021; Valiante et al., 2021). We underline that our whole BHMF (Fig. 3.19) shows few characteristic features, such as "wiggles" and "humps", especially at $z > 1$ and $M_\bullet \lesssim 10^5 M_\odot$. These are due to the GQd grafting procedures (§3.3.2), from which we inherit an already-evolved population of BHs with an intrinsic mass-distribution (see the two bottom-right panels of Fig. 3.10). In addition, the BHMF of ungrown BHs (dotted black line) shows a clear bi-modality at $M_\bullet \sim 10^5 M_\odot$. This is a consequence of mergers between GQd light and heavy seeds during their evolution within GQd trees (i.e. *before* we graft them into L-Galaxies structures). Indeed, despite we ignore GQd pure-heavy seeds (§3.3.2), these objects might have already merged with light seeds within their GQd

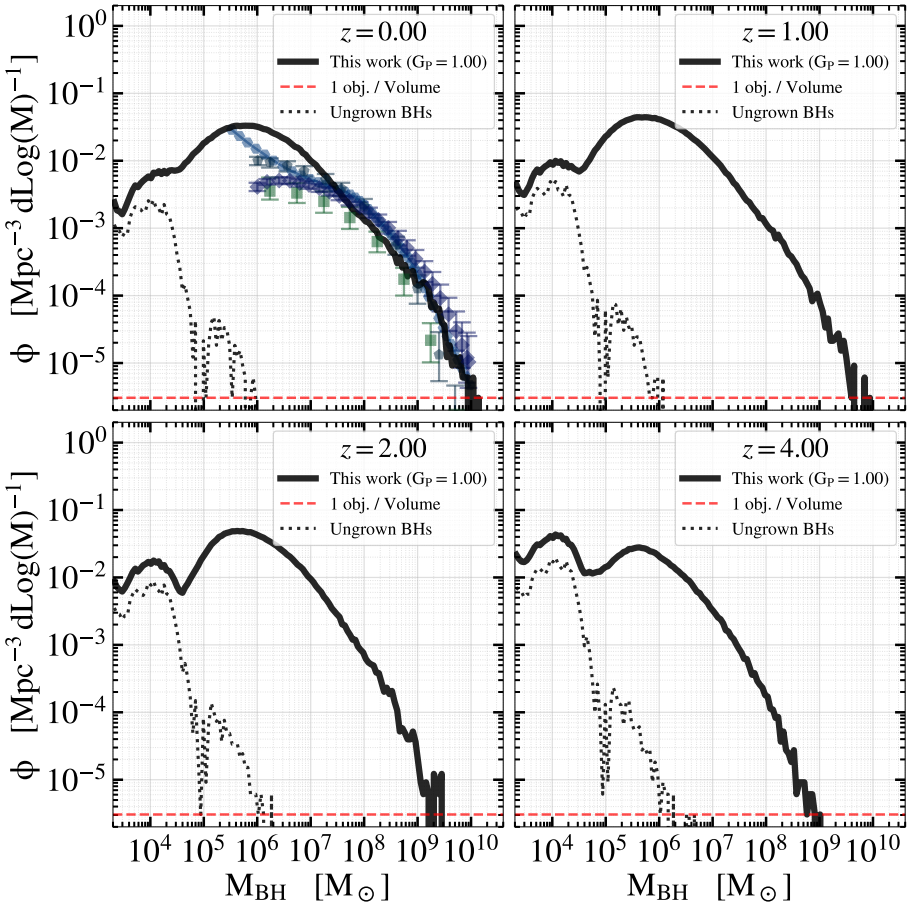

Figure 3.19: Black Hole mass function (BHMF) at four representative redshifts, namely: $z=0,1,2$ and $z=4$ (as in Fig. 3.15 and 3.18). We emphasize the relative contribution of ungrown BHs at any z (i.e. BHs with $M_\bullet = M_{seed}$, dotted black line). We note that these objects provide only a minor fraction to the whole BHMF at $M_\bullet \gtrsim 10^4\,M_\odot$, while quickly becoming predominant at $M_\bullet \lesssim 10^4\,M_\odot$. The interest towards this particular SMBH population reside in the fact that it might carry direct information about BH-formation processes (see §3.4.5).

evolution, hence being labelled as mixed-flavour BHs in GQd (see Valiante et al., 2016), which are not ignored during our grafting procedures.

Finally, we note that our definition of ungrown BHs is only based on the mass-growth we can resolve during the L-Galaxies evolution of BH-seeds descendants. This provides a clear view of the evolutionary path of BH-seeds descendants as modelled by L-Galaxies on the MR-II merger trees. Nevertheless, we graft evolved SMBHs from GQd outputs which have already undergone unresolved mass-growth. Consequently, we over-estimate the fractions of ungrown BHs obtained from our model at the low-mass end of the MR-II dynamic range (Fig. 3.18) or of the BHMF (Fig. 3.19). This supports the idea presented in Mezcua (2019), according to which the average population of low-mass SMBHs hosted in dwarf galaxies might not be a direct tracer of BH-seeding processes due to the inefficient mass-growth experienced within these low-mass systems.

3.5 Discussion and conclusions

This chapter presented the implementation of a physically-motivated set of prescriptions for the formation of BH seeds at high-z. These were embedded in the context of the L-Galaxies SAM, which allow to exploit the N-body, cosmological merger tree of the MR-II simulation (Boylan-Kolchin et al., 2009).

Our model describes the self-consistent formation of BH seeds according to most of the currently-envisioned baryonic channels, namely light, intermediate and heavy scenarios (see §1.2.2 and §3.3). In particular, due to the mas-resolution limitations of MR-II , light-seeds formation is modelled in a sub-grid fashion by exploiting the outputs of the GQd analytic model (§3.3.2). The implementation of

this seeding model is rooted on the modelling we developed for the diffusion of UV photons and metals in the IGM, as well as the build-up of LW and metallicity locally-varying backgrounds. These latter models, in particular, resort to a number of necessary approximations, which on one side are necessary in order to perform our study over the cosmological volume simulated by MR-II , but on the other hand introduce shortcomings in our methods and possibly influence our results.

In particular, the spatial variations of J_{LW} and Z_{IGM} are accounted for by using two lists of metal and LW sources (see §3.2.4). The definition of these lists is subject to thresholds which can lead to including a varying number of MR-II structures in them. Therefore, these thresholds can affect the number of newly-initialized halos subject to favourable local conditions of J_{LW} and Z_{IGM} and therefore influence the abundance of pure-DCBHs and pure-RSMs. Nevertheless, as shown in §3.4, pure-light seeds-descendants are overwhelmingly dominant within our SMBH population at all redshifts. Consequently, although the abundance of pure-DCBHs and pure-RSMs can vary significantly as a function of the threshold we impose, we do not expect a significant impact on our overall results. On the contrary, as commented in the previous sections, the details of our GQd grafting procedures can strongly affect the abundance of pure-light seeds-descendants with tangible consequences on the evolution of SMBH populations at $z<6$.

On top of this, we use the Izquierdo-Villalba et al. (2020) model for the mass-growth of SMBHs, which allows to compute the luminosity of accretion BHs and to build their LF across cosmic times. As shown in previous sections, the modelling of BH seeding processes also impacts the predictions of the LFs at different redshifts, in a non-trivial way. In order to thoroughly account for these complex inter-dependency between BH-seeding and mass-growth prescriptions, we plan to perform an accurate calibration of our model using efficient optimization techniques in the future (se also the discussion in §5.1.4). In this way, also the interplay between BH-seeding and additional complex phenomena related to SMBHs evolution (such as GW emission, merger-induced kicks and BH-spin evolution, see e.g., Izquierdo-Villalba et al., 2020) can be taken into account in the study of the evolution of SMBHs within cosmological contexts.

The inclusion of BH-spin evolution and GW recoils, in particular, might affect the predictions of our model for the BH occupation at low BH-masses. Indeed, light seeds might be expelled easily from low-mass MR-II structures. This is because GW kicks are maximized by BH-BH mergers with similar mass, and light-seeds dominate the relatively narrow mass spectrum inherited by GQd at high-z (see e.g. Fig. 3.10. Therefore, in this regards, the inclusion of the SMBH physics modelled by Izquierdo-Villalba et al. (2020) and its interplay with BH-seeding might provide interesting outcomes for several open questions related to SMBH evolution, such as: the study of wandering BHs around galaxies, the presence of intermediate-mass BHs in local dwarf galaxies and the occupation of SMBHs at the center of massive galaxies.

An interesting result of our model is that heavy-seeds contribution does not appear to be necessary in order to reach the highest BH masses in the MR-II box at $z\sim6$. Although the MR-II box does not allow to sample the space-densities at which extremely bright, high-z QSOs are found (see 1.1), this prediction of our model is in contrast with most current theoretical works. Indeed, the latter show that the most massive SMBHs at high-z should be formed by either super-Eddington accretion onto light seeds (e.g., Wyithe & Loeb, 2012; Inayoshi et al., 2016; Pezzulli et al., 2017; Sugimura et al., 2017) or by the intervening contribution of intermediate-to-heavy seeds (e.g., Valiante et al., 2016; Pacucci et al., 2017; Latif et al., 2018; Sassano et al., 2021). Consequently, we would expect our BH-seeding model to predict a high occupation of BHs descending from heavy or intermediate seed components at the highest BH masses. On the contrary, we observe a clear predominance of pure-light seeds descendants at all masses, at $z\lesssim7$, as shown by the top-left panel of Fig. 3.10. This is a combined effect of our grafting procedures (see §3.3.2) and our self-consistent prescriptions for pure-DCBH and pure-RSM seeds formation. More in detail: given the details of how the GQd and L-Galaxies dynamical ranges overlap, newly resolved structures in L-Galaxies (which have a typical $M_{vir}^{res}\sim2\times10^8M_\odot$) sample different halo populations within the GQd outputs, at different redshifts. Indeed, at $z\gtrsim14$ L-Galaxies are grafted with massive, evolved halos belonging to the main branches of GQd merger trees. On the contrary, around $z\sim10$, L-Galaxies structures sample the low-mass tail of GQd halos, which is increasingly emptier of BH seeds descendants (as shown in Fig. 3.2). On top of this effect, in the outputs of GQd we use, heavy seeds are only formed at $z\lesssim13$, in a relatively small number (see Valiante et al., 2016, 2021). Consequently, before this moment L-Galaxies structures only inherits pure-light light-seeds, which form the vast majority

of the early BH population within the first collapsing MR-II halos. With the progress of grafting, this trend is not interrupted, as pure-light seeds still represent the vast majority of newly-initialized SMBHs down to $z \sim 9$. In these conditions, the small contribution from heavy channels (either inherited from GQd outputs or formed in L-Galaxies self-consistently) is not sufficient to overcome the predominance of pure-light descendants. Therefore, the latter can grow via efficient gas accretion and mergers in L-Galaxies. In particular, the latter process produces interactions between pure-light seeds with a much higher probability than that of mixed-types interactions. This ultimately allows pure-light seeds descendant to reach few $10^8 M_\odot$ by $z \sim 6$ without any heavy contribution.

Possible solutions to this model caveat may exploit new GQd outputs obtained with different resolutions settings (see also §5.1.4 for additional details). Indeed, this can help to explore a wider variety of cosmological environments, rather than only the merger-tree of a specific halo. From this improved sample, then, it would be possible to extract grafting information from a more variate halos sample which can better cover the dynamic range of MR-II structures. A similar approach can be followed by exploiting a higher-resolution N-body, cosmological simulation which can actually resolve the mini-halos mass range (see §1.2). The principle of this "nested" method would be similar to the grafting procedures presented in this chapter, with the exception of using additional information provided by the cosmological nature of the high-resolution simulation (see the discussion in §5.1.4). Indeed, also properties such as e.g. the local density of structures or the amount of neighboring, star-forming galaxies could be used to draw physically-motivated initial conditions for the newly-initialized MR-II structures.

As noted above, the fraction of BH seeds forming in L-Galaxies as an effect of local conditions is always extremely small in comparison to the ones inherited by GQd grafting. We explored the possibility of balancing the different seeding channels by introducing the grafting probability G_P (see §3.3.2). Nevertheless, the overall picture according to which pure-light seeds are always predominant within our global SMBH populations is not significantly affected by this modifications. Indeed, the effect of G_P is to reduce the overall abundance of SMBHs in our box (see Fig. 3.11) rather than allowing for a significantly higher production of SMBHs in L-Galaxies structures. This shows that, in our model, the bottleneck for the formation of pure-RSM or pure-DCBH seeds is the scarce occurrence of favourable local conditions, rather than the occupation of newly-resolved structures by BHs inherited from GQd. In particular, the delay in the build-up of strong LW backgrounds imposed by the mass resolution of MR-II produces a lack of photo-ionized regions which can host DCBH and RSM seeds formation. This shortcoming can be improved by employing less-restrictive thresholds to define our lists of LW sources (see §3.2.4), at the price of significantly longer execution times for our model.

On the contrary, we expect little effects on this aspect from our model of the patchy chemical enrichment of the IGM. Indeed, we neglect energy losses due to e.g. turbulence or to the interaction of metallic-shells fronts during their expansion in the IGM. This overestimates the radius of regions affected by metals around galaxies. Nevertheless, we do not find any halo for which the potential formation of a BH seed is hindered by the direct chemical feedback of neighboring galaxies. Indeed the driving mechanisms of halos pollution in our model are the internal SF and, later in time, the average enrichment of the IGM. Therefore, we do not expect significantly different results by implementing a more accurate model for the propagation of metallic shells. As a matter of fact, energy losses would tend to reduce the radii of expanding fronts, therefore diminishing even further the potential effect of direct chemical feedback onto the regulation of BH seeding processes.

On the contrary, the actual number of metallic sources included in our sources-catalogs (see §3.2.4) might effectively play a role within our BH-seeding recipes. Unfortunately, this approach is computationally prohibitive, mostly due to the very high number of small structures in the MR-II merger trees. A possible solution is offered by the new version of L-Galaxies2020 (see Yates et al., 2021), in which SF histories of galaxies and their associated SNe feedback are efficiently tracked in time. This might allow to follow the ejection of metallic shells (and LW photons emission) without relying on catalogs of metallic (and LW) sources. We plan to explore this possibility in future works.

Conclusions

Our model for the formation and evolution of SMBHs is one of the first semi-analytical approaches in which physically-motivated prescription for light, intermediate and heavy BH-seeds production are embedded in a fully-cosmological context and coupled with detailed mass-growth prescriptions. In particular, the formation of BH seeds is modelled either by resorting to external initial conditions (light-seeds) and self-consistently (intermediate RSM and heavy DCBHs), therefore allowing to study their occurrence over cosmological environments. Our BH-seeding prescription is coupled to the self-consistent modelling of the diffusion of LW photons and metals in the IGM, which produce two spatially- and time-varying background-fields. In this model, SMBHs only form due to physically-motivated conditions which either depend on the resolution mass of newly-virialized structures (see §3.3.2) or local gas properties at their position (see §3.3.3).

By joining this theoretical framework with the mass-growth model detailed in (Izquierdo-Villalba et al., 2020), we can track the build-up in time of the SMBH population resulting solely from high-z BH-seeding. We underline that this is one of the first works in which a comprehensive, physically-motivated model for BH formation at high-z (coupled to spatially-varying LW background and IGM enrichment) is applied to a N-body, cosmological context and follows the evolution of the resulting SMBH population self-consistently down to $z \sim 0$. Consequently, the main scientific outcome of this work, is the prediction of low-z statistical properties of SMBH populations by *only* employing physically-motivated BH formation recipes and mass-growth prescriptions. In other words, we show for the first time over a wide cosmological context that a detailed modelling of high-z BH-seeding, ranging from light to heavy seeds, is able to account for the origin and evolution of the SMBH population observed at low-z.

Our results show a strong predominance of pure-light seeds descendants, whose evolution and growth drives and determines the property of the average SMBHs population at all epochs and BH masses in our model. In particular, light PopIII remnants are able to hierarchically assemble their mass and finally account for the whole BHMF at $z \sim 0$, from relatively small $M_\bullet \sim 10^6 M_\odot$ up to massive $M_\bullet \sim 10^9 M_\odot$ SMBHs. As discussed above, this finding is in apparent contrast the necessity of heavy seeds contribution to reach the highest BH masses at high-z if super-Eddington growth is not taken into account (see e.g., Valiante et al., 2016; Lupi et al., 2021a; Sassano et al., 2021). Nevertheless, in our model, this is a consequence of both the resolution limits of MR-II (which only allow the formation of intermediate or heavy seeds too late in time), and the interplay between the dynamic range of GQd outputs and newly-resolved MR-II structures. Nevertheless, this finding does not exclude that intermediate or heavy seeds might be necessary in order to form the SMBHs powering the brightest $z \gtrsim 6$ QSOs observed, but might play a minor role within global populations of SMBHs within wide cosmological volumes.

According to our model, pure-RSMs and pure-DCBHs preferentially form in overdense regions, as satellites of star-forming galaxies. Indeed, these are responsible for the necessary UV illumination needed for these BH-seeding processes to occur. Furthermore, our results support the close-halo pair scenario (e.g., Dijkstra et al., 2008; Agarwal et al., 2013; Visbal et al., 2014b; Regan et al., 2017), according to which requiring the presence of $J_{LW\,;\,local}$ leads to the formation of DCBHs in the vicinity of actively star-forming galaxies, whose chemical feedback did not yet enrich their surroundings. As expected, this applies also to the formation of pure-RSM, since $J_{LW\,;\,local}$ is also needed for this specific seeding channel (e.g., Omukai et al., 2008; Sassano et al., 2021). This requirement, in particular, has often been relaxed, under the assumption that SF episodes within the structure forming the RSM seed might provide the necessary $J_{LW\,;\,local}$. Therefore, our work allows to test, in a cosmological context, the idea that also RSM might form under the influence of external photo-ionizing fluxes. Nevertheless, we stress that MR-II resolution does not allow to track early SF episodes which might actually provide $J_{LW\,;\,local}$, therefore leaving open the possibility that ionizing fluxes might have either internal or external origins.

Furthermore, our seeding prescriptions leave a clear imprint on both the evolution and low-z properties of the SMBHs population simulated over the MR-II volume. In particular, RSM seeds descendants which underwent at least one merger with pure-light seeds are only found in $M_{vir} \gtrsim 10^9 M_\odot$ halos, while the descendants of light-heavy mergers are widespread across the whole MR-II dynamic range. This difference is related with the fact that heavy components in our model can be either inherited from GQd grafting or formed under the effect of local conditions in L-Galaxies structures. Therefore heavy seeds descendants

are a more heterogeneous class than RSM descendants. Nevertheless, both seed-types classes are more likely found in central galaxies with respect to the average SMBH population. This correlates with the fact that pure-RSMs and pure-DCBHs preferentially form as close satellites of star-forming halos, therefore they can likely merge with their neighboring halo during their evolution. On the contrary, the descendants of pure-RSMs and pure-DCBHs which never experienced a merger with different seed-types are only found in $M_{vir} \lesssim 10^9 M_\odot$ halos and, by $z \sim 0$, only in satellite galaxies. This is in line with the idea that few seeding-remnants are able to actually survive down to $z = 0$ and might be found in dwarf satellite galaxies (e.g., Greene, 2012; Mezcua, 2021).

Thanks to the physical model developed by Izquierdo-Villalba et al. (2020), our SMBHs grow in time by the accretion of gas funneled to the galactic center as a consequence of galactic mergers. Our predictions for the $z \sim 0$ BHMF are in good agreement with current observational determinations, especially at $M_\bullet \gtrsim 10^7 M_\odot$. This shows that the total mass-growth of SMBHs integrated down to $z \sim 0$ is well-predicted by the Izquierdo-Villalba et al. (2020) model. Furthermore, the $z \sim 0$ BHMF shows interesting connections with the details of our BH-seeding model. Indeed, by modulating the initial abundance of pure-light seeds-descendants inherited from GQd we find significant differences on the amplitude of the BHMF at $M_\bullet \lesssim 10^8 M_\odot$. This suggests that the low-mass end of the BHMF carries information about the efficiency of high-z BH-seeding processes. This is not the case at $M_\bullet > 10^8 M_\odot$, where the details of the light-seeds abundance modulation are completely washed out by both the hierarchical and accretion-driven growth of pure-light seeds descendants. We underline that the latter always represent the vast majority of the SMBH population which results from high-z BH-seeding processes. Nevertheless, at least at $z \sim 0$ the contribution of heavy seeds descendants to the BHMF becomes relatively more important at high BH-mass, as an effect of the hierarchical assembly of SMBHs across cosmic times.

Finally, despite the $z \sim 0$ BHMF is substantially well predicted by the combination of our seeding prescriptions and the Izquierdo-Villalba et al. (2020) growth model, the mass-assembly history of our SMBH population shows significant discrepancies with respect to current constraints. In particular, the evolution of the bLF, XLF and the BHAR density collectively show that SMBHs rapidly accrete their mass at early epochs (i.e. $z \gtrsim 3$). Indeed, our model significantly over-estimates the bLF and XLF at these early epochs, especially at the very faint-end (i.e. $L_{bol} < 10^{43} erg\,s^{-1}$ or $L_{[2-10\,KeV]} < 10^{42} erg\,s^{-1}$). On the other hand, at low redshift, SMBHs appear to be much less active than what current observations require, suggesting that the mass-growth of our SMBH population is anticipated in time. This discordance with observations mildly improves by acting on the abundance of pure-light seeds inherited by GQd. Similar mild (but insufficient) improvements can be obtained by acting on the parameters of the SMBH growth model (see Izquierdo-Villalba et al., 2020), without obtaining substantial agreement with current constraints.

This suggests that some physical aspects of the mass-growth model are not sufficient to accurately model the mass-assembly history of SMBHs on MR-II merger trees. Indeed, we stress that the Izquierdo-Villalba et al. (2020) model was calibrated on the MR merger trees, which show a significantly different dynamic range (see e.g., Springel et al., 2005; Henriques et al., 2015). Therefore, our mass-growth model might have to be significantly modified in order to properly treat SMBH evolution within the low-mass halos populations sampled by MR-II. For instance, the inclusion of additional physical processes related to SMBH evolution can significantly influence the mass-assembly history of SMBHs. Indeed, processes such as their spin evolution and its impact on their mass-growth history, or the emission of GWs and the subsequent gravitational recoil after their mergers, are expected to show their imprints on the LF and overall mass-growth history. Since Izquierdo-Villalba et al. (2020) already included these processes, we plan to carry on this thorough analysis in the future, by employing recent techniques for the optimization of the physical-model parameters.

Chapter 4

The Lyman-α luminosity function of $2 < z < 3.5$ quasars

This chapter is devoted to present the photometric determination of the Lyman-α LF of AGN/QSOs at $2<z<3.5$, obtained with data from the J-PLUS survey (see §2.2.4 and Cenarro et al., 2019). This work targeted unexplored intervals of Lyα luminosity ($L_{Ly\alpha}$) and number density (Φ) thanks to the large area sampled by the multi-narrowband photometric observations of J-PLUS. Indeed, this allows to produce flux-limited catalogs of sources without requiring any target pre-selection (see §2.2), therefore potentially including some of the rarest and brightest objects known to date. These features allowed to sample the Lyα luminosity distribution of sources at $L_{Ly\alpha} \gtrsim 10^{44}\,\mathrm{erg\,s^{-1}}$ with sufficient statistics, as compared to previous studies (e.g. Konno et al., 2016; Matthee et al., 2017b; Sobral et al., 2018b). The results discussed here were published in Spinoso et al. (2020) and later-on supported by the spectroscopic determination of Zhang et al. (2021) in the context of the HETDEX survey (Adams et al., 2011).

The discussion within this chapter is organized as follows: §4.1 contextualizes and motivates our study of the Lyα LF, §4.2 presents the technical feature of J-PLUS and §4.3 details the selection rules by which we look for Lyα-emitting sources in J-PLUS. In §4.4 we actually define and analyze our catalogs of Lyα-emitting candidates while §4.5 describes the spectroscopic follow-up programs we undertook at the Gran Telescopio Canarias (GTC), to confirm a small subsample of our candidates. Finally, in §4.6 we describe the methods we use to compute the Lyα LF of our candidates while in §4.7 and §4.8 we respectively discuss and summarize our results.

4.1 Introduction

As detailed in §1.4, the Lyman-α emission of astronomical sources is usually associated to energetic phenomena which produce strong fluxes of ionizing photons (see e.g. Dijkstra, 2017, for a recent review). After being re-emitted at Lyα wavelengths by neutral hydrogen, these can be observed at cosmological distances from the emitting sources, thanks to the globally-ionized state of the $z<6$ Universe (see §1.4). Furthermore, Lyα emission is redshifted into the optical band at $2 \lesssim z \lesssim 5$, hence allowing relatively simple, ground-based observation of high-z sources (as compared to e.g. expensive space-based surveys). As a consequence, Lyα emission is an extremely powerful observable in order to study galaxy formation and evolution, as well as the properties of luminous QSOs, up to very high z (e.g. Charlot & Fall, 1993; Møller & Warren, 1998; Haiman & Rees, 2001; Gawiser et al., 2006; Guaita et al., 2015; Matthee et al., 2016; Erb et al., 2018; Kimm et al., 2018; Khostovan et al., 2019; Santos et al., 2020; Calhau et al., 2020). For the above reasons, an increasing number of recent works has been focusing on the study of high-redshift LAEs (see §1.4.1 for this definition).

On the other hand, AGN/QSOs sources can also emit strongly at Lyα wavelengths due to the reprocessing of the intense ionizing fluxes produced by the accretion of matter onto the SMBHs at their centers. The Lyα emission of AGN/QSOs has been widely used for studying their properties and environments at high-z (e.g. Cantalupo et al., 2012; Borisova et al., 2016a; Kikuta et al., 2017) and

for statistically interpreting their populations at more intermediate redshifts ($z \sim 2$, e.g. Calhau et al., 2017, 2020; Zhang et al., 2021). In addition, the absorption of Lyα photons by the IGM surrounding AGN/QSOs and along their line-of-sight is used to infer the properties of gaseous medium surrounding these sources (ee e.g., Wyithe et al., 2005b; Bolton et al., 2011; North et al., 2012; Croft et al., 2018) and for probing the reionization state of the Universe at different redshift through the so-called Lyα-forest (e.g. Fan et al., 2006a; Becker et al., 2015a; Eilers et al., 2018; Dyer et al., 2019).

Understanding the relative contribution of LAEs and AGN/QSOs to the whole population of Lyα-emitting sources at different luminosity and redshift can shed light on the roles played by star-formation and SMBH mass-growth in the Universe history. This is even more important when considering the peak-epochs of star-formation density and AGN activity in the Universe ($1 \lesssim z \lesssim 3$, see Madau et al., 1996, 1998; Ueda et al., 2014), where the relative fraction of AGN/QSOs and LAEs in photometrically-selected samples of sources is yet poorly constrained (especially at bright $L_{Ly\alpha}$, e.g. Konno et al., 2016; Matthee et al., 2017b; Sobral et al., 2018a; Calhau et al., 2020).

A natural tool for tackling this kind of study is the LF of Lyα-emitting sources, namely their number-density distribution with respect to their measured $L_{Ly\alpha}$ (see §1.5). As previously mentioned, the observational part of this thesis work presents a photometric determination of the Lyα LF by exploiting the wide-area, multi-NB data of the J-PLUS survey. In the context of J-PLUS data, we obtain $L_{Ly\alpha}$ from the sources photometric excess at wavelengths corresponding to those of Lyα emission, redshifted up to $2 \lesssim z \lesssim 3$. Due to the J-PLUS technical features (see §4.2), we can only concentrate on bright regimes, namely $L_{Ly\alpha} \gtrsim 10^{43.5}\,\mathrm{erg\,s^{-1}}$, where a strong contribution from AGN is expected (e.g. Sobral et al., 2018b).

In particular, we exploit the first J-PLUS data-release (DR1 hereafter, Cenarro et al., 2019), which provides measurements of the Northern hemisphere in twelve different photometric bands, as better detailed in §4.2 (see e.g. Fig. 4.1 and Table 2.1). The DR1 covers an area of $> 1000\,\mathrm{deg^2}$, which is unprecedented for NB-surveys of $z > 2$ luminous line-emitters. We exploit these characteristics to obtain large samples of photometrically-selected Ly$\alpha-$emitting sources, and probe the Lyα LF bright-end at four different redshifts (see Table 4.2). Indeed, the combination of large survey area and multi-NB data allows to sample the distribution bright Lyα-emitting sources over an unprecedented luminosity regime and sources number-density interval (Gronwall et al., 2007; Guaita et al., 2010; Blanc et al., 2011; Konno et al., 2016; Sobral et al., 2018b). We complement our study with the results of two follow-up spectroscopic programs aimed at assessing the performance of our method (see §4.5).

Within this chapter, magnitudes are expressed in the AB system (Oke, 1974; Oke & Gunn, 1983). Our results are based on a flat ΛCDM cosmology described by PLANCK15 parameters (Planck Collaboration et al., 2016a,b), namely: $H_0 = 67.3\,\mathrm{Km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{m;0} = 0.315$, $\Omega_{\Lambda;0} = 0.685$.

4.2 The J-PLUS multi-narrowband photometric survey

J-PLUS is an ongoing wide-area photometric survey performed at the Observatorio Astrofísico de Javalambre (OAJ, Cenarro et al., 2014) in Arcos de las Salinas (Teruel, Spain). We introduced its main technical features in §2.2.4, along with those of the main survey of OAJ, i.e.: J-PAS. This section is devoted to contextualize the technical aspects of J-PLUS which play an important role in our analysis (we refer the reader to Cenarro et al., 2019, for additional details), and to describe how we exploit these features to target high-z Lyα-emitting sources.

4.2.1 The J-PLUS catalogs of sources

J-PLUS images are automatically reduced to obtain public catalogs of sources[1]. The results shown in this chapter are based on the recent DR1 results, obtained on 511 pointings with stable pipelines for data reduction and source-extraction, which were specifically calibrated and tested on J-PLUS data (see §2.2.4 and e.g., Cenarro et al., 2019; López-Sanjuan et al., 2019a). J-PLUS catalogs are built by using the SExtractor code (Bertin & Arnouts, 1996), either in single-mode or dual mode. The latter exploits r as the reference band for detecting sources within images and for determining the photometric apertures

[1] J-PLUS catalogs can be found at: http://archive.cefca.es/catalogues

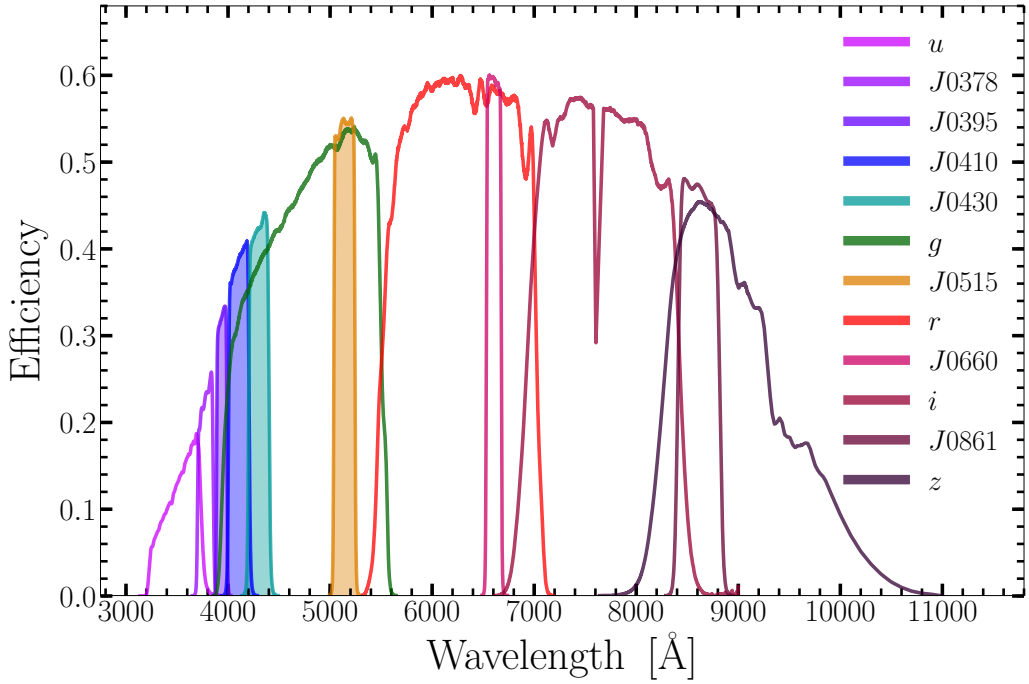

Figure 4.1: The measured transmission curves for the J-PLUS filter set, after accounting for sky absorption, CCD quantum efficiency and the total effect of the JAST/T80 telescope optical system. The four NB we exploit to look for bright Lyα emitters at z>2 (namely, the $J0395$, $J0410$, $J0430$ and $J0515$ NBs) are shown here as filled-area curves. Note how they all share their entire wavelength coverage with the g band (solid empty green line). The three remaining NBs are $J0378$, $J660$ and $J0861$, respectively violet, dark red and brown empty solid lines.

associated to each source position. These apertures are then translated at the same source positions in the images of the remaining 11 filters, in order to perform multi-band photmetric measurements at fixed sky positions. We use measurements obtained with `auto`-apertures[2], which are defined on Kron-like ellipses (Kron, 1980). This photometry definition is motivated by the apparent size of high-z Lyα-emitting sources in J-PLUS data (see §4.2.2), since it ensures to recover the total Lyα line flux of our candidates, as we detail in §4.6.1. We note that using a tabulated definition of photometric aperture (instead of a custom-defined one) allows to exploit the detection-completeness curves provided by the J-PLUS database. Therefore, we can easily account for the bias introduced by the J-PLUS source-detection procedures on our final results, as detailed in §4.6.4. Finally, we note that relying on dual-mode catalogs has non-trivial implications on the total completeness of our final samples of Lyα-emitting candidates, which we also address in §4.6.4. Finally, in our discussions (i.e. §5) we further comment the possible biases introduced by these choices on our final results

Narrow Band	z of contaminant QSO lines			
	SiIV	CIV	CIII]	MgII
$J0395$	$1.82^{+0.03}_{-0.04}$	$1.54^{+0.03}_{-0.04}$	$1.06^{+0.02}_{-0.03}$	$0.41^{+0.01}_{-0.02}$
$J0410$	$1.94^{+0.08}_{-0.07}$	$1.65^{+0.07}_{-0.06}$	$1.15^{+0.06}_{-0.05}$	$0.47^{+0.04}_{-0.03}$
$J0430$	$2.08^{+0.07}_{-0.07}$	$1.78^{+0.07}_{-0.06}$	$1.25^{+0.05}_{-0.05}$	$0.54^{+0.04}_{-0.04}$
$J0515$	$2.68^{+0.07}_{-0.08}$	$2.32^{+0.06}_{-0.07}$	$1.69^{+0.05}_{-0.06}$	$0.84^{+0.03}_{-0.04}$

Table 4.1: Redshift intervals at which we expect to detect strong QSOs lines (e.g., Vanden Berk et al., 2001) which might contaminate our selection, for each of the four NBs we use. We note that, due to the NB filters widths, some of the z intervals overlap among different filters and lines (e.g. CIV in $J0410$ and CIII] in $J0515$). This introduces a further degree of complexity in the contamination we expect from QSO lines.

[2]For details about J-PLUS aperture-photometry definitions see: http://archive.cefca.es/catalogues/jplus-dr1/help_adql.html

4.2.2 Detection of Lyα emission with J-PLUS

The design of J-PLUS filters potentially allows to detect Lyα emission within seven redshift windows, one per NB, respectively centered at z $\sim$ 2.11, 2.24, 2.38, 2.54, 3.23, 4.43 and 6.09. In particular, we employ the *J*0395, *J*0410, *J*0430 and *J*0515 filters (shown as filled-area curves in Fig. 4.1) for targeting z $\sim$ 2.24, 2.38, 2.54 and 3.23 (Table 4.2). Our selection procedure is based on measuring NB excesses with respect to the continuum traced by BB photometry, following the method described in §4.3.1. consequently, due to the limited spectral information retrievable by photometric measurements (see §2.2.3), we cannot determine a priori the nature of the NB excess measured in a given band. Therefore, our selection method is prone to contamination by prominent emission lines at different redshifts with respect to the one of targeted lines.

We expect our samples to be contaminated by both nebular emission due to star-formation (e.g. $\mathtt{H}_\beta$, $\mathtt{[OIII]}_{4959+5007}$ and $\mathtt{[OII]}_{3727}$ lines) and AGN/QSOs ionizing radiation (e.g. $\mathtt{CIV}_{1549}$, $\mathtt{CIII]}_{1909}$, $\mathtt{MgII}_{2799}$ and $\mathtt{SiIV}_{1398}$ lines, see also Stroe et al., 2017a,b). We list the main AGN/QSOs lines from which we expect contamination in Table 4.1, together with their associated redshift intervals in J-PLUS. We note that SiIV and MgII are minor sources of contamination since: i) they are significantly fainter than Lyα (e.g., Telfer et al., 2002; Selsing et al., 2016), ii) J-PLUS probes relatively small cosmological volumes at $0.35 < z < 0.85$ and iii) the number density of bright AGN/QSOs at $z < 1$ is expected to be lower than at $z > 2$ (e.g., Hasinger et al., 2005; Ueda et al., 2014; Miyaji et al., 2015; Palanque-Delabrouille et al., 2016; Pâris et al., 2018).

We exclude the *J*0378 NB from our analysis after checking that our method for detecting and measuring NB excesses does not provide reliable results in this particular NB (see §4.6.1). We also exclude the *J*0660 and *J*0861 NBs since they provide very scarce samples of candidates ($\lesssim$100 sources) whose contamination cannot be reliably estimated, due to the absence of SDSS spectroscopic counterparts (see §4.4). We note that this is in agreement with the work of Sobral et al. (2018b), which shows no significant detection of bright ($L_{Ly\alpha} v \gtrsim 10^{43}\,erg\,s^{-1}$) Ly$\alpha$-emitting sources at $z \gtrsim 3.5$, i.e. at the redshift probed by the *J*0660 and *J*0861 NBs.

$L_{Ly\alpha}$ and $EW_{Ly\alpha}$ detection limits

In order to characterize the limits of J-PLUS towards the detection of high-z Lyα, we are interested to determine $L^{min}_{Ly\alpha}$ for each J-PLUS NB. For this goal we need to measure the minimum Lyα line-flux $F^{min}_{Ly\alpha}$ detectable with J-PLUS NBs. Unfortunately, the relative contributions of source continuum and line emission are degenerate in a single NB measurement (see §2.2.3). As discussed in §2.2.5, the line EW and λ_{peak} are necessary to disentangle these two contributions, but λ_{peak} must be either assumed a priori or measured indirectly. We choose to assume it by relying on reasonable working assumptions, namely: sources emitting faint Lyα fluxes are detected with higher probability at the wavelength of the transmission curve maximum, so that $\lambda_{peak} = \lambda_{T\,max}$. The choice of EW, on the other hand, as a higher degree of arbitrariness. Despite EW as low as 5 Å have been explored in the past (e.g., Sobral et al., 2017), high-z Lyα-emitting sources typically exhibit EW $> 15 - 20$ Å (as in e.g., Gronwall et al., 2007; Guaita et al., 2010; Santos et al., 2020). For the only purpose of computing the theoretical limit $L^{min}_{Ly\alpha}$ to which J-PLUS NBs are sensitive, we select $EW_{min} = 25$Å (see e.g., Ouchi et al., 2008; Santos et al., 2016; Konno et al., 2018). With the above assumptions, we use the detection limits of J-PLUS bands (table 2.1) to obtain the minimum line-flux ($F^{min}_{Ly\alpha}$) measurable with each NB. More in detail, we use the method detailed in §2.2.5 to estimate $F^{min}_{Ly\alpha}$ and then link the latter to $L^{min}_{Ly\alpha}$ by using the assumed values of λ_{peak} and EW_{min}. We note that, as an effect of its detection limits, J-PLUS is sensitive to very bright Lyα emission, namely: $L_{Ly\alpha} > 10^{43.3}\,erg\,s^{-1}$ (see Table 4.2). Few studies have explored this range of $L_{Ly\alpha}$ with sufficient statistics, being generally based on deep but narrow photometric surveys (see e.g., Blanc et al., 2011; Konno et al., 2016; Matthee et al., 2017b; Sobral et al., 2018b). On the contrary, J-PLUS DR1 provides multi-band imaging over $\sim$1000 deg^2, which is unprecedented for studies targeting high-z Lyα-emitting sources.

Narrow Band	Properties of Lyα detection in J-PLUS							
	$\langle z \rangle$	z_p	$[z_{min}\ ;\ z_{max}]$	A_{eff} [deg^2]	Vol [cGpc^3]	$F^{min}_{Ly\alpha}$ [erg cm^{-2} s^{-1}]	$L^{min}_{Ly\alpha}$ [erg s^{-1}]	Log $(L^{min}_{Ly\alpha})$
*J*0395	2.24	2.25	2.20 – 2.28	897.44	0.961	5.23×10^{-16}	2.16×10^{43}	43.33
*J*0410	2.38	2.37	2.29 – 2.46	897.46	1.917	4.46×10^{-16}	2.21×10^{43}	43.34
*J*0430	2.54	2.53	2.46 – 2.62	897.41	1.907	3.99×10^{-16}	2.25×10^{43}	43.35
*J*0515	3.23	3.24	3.14 – 3.31	965.99	2.044	2.65×10^{-16}	2.68×10^{43}	43.43

Table 4.2: Properties of J-PLUS NB filters related to the Lyα line detection. From left to right: median redshift in the filter bandwidth, redshift associated to the filter *pivot* wavelength (see Tokunaga & Vacca, 2005, and §2.2.1), redshift interval covered by the NB FWHM, effective DR1 area ad volume associated to our selection and finally minimum detectable line flux and luminosity (both in linear and logarithmic units).

Expected morphology of Lyα-emitting sources in J-PLUS data

As discussed in §4.2.1, the DR1 dual-mode catalog is based on r-band detections, which probe the rest-frame UV-continuum of $z \gtrsim 2.2$ sources. UV observations of star-forming Lyα emitters (SF LAEs) at these high redshifts show typical rest-frame half-light radii of about $r_{50} \lesssim 2$ kpc (see e.g., Venemans et al., 2005; Taniguchi et al., 2009; Bond et al., 2009, 2012; Kobayashi et al., 2016; Ribeiro et al., 2016; Paulino-Afonso et al., 2017, 2018). This translates into apparent sizes comparable to the spatial resolution of T80Cam (R = 0.5"/pixel) and to the typical J-PLUS seeing (i.e. $s \lesssim 1$", Cenarro et al., 2019). On the other hand, QSOs are point-like sources by definition, therefore we expect both SF LAEs and QSOs at $2.2 \lesssim z \lesssim 3.3$ to show compact r-band morphology. On one side, this justifies our usage of auto-aperture photometry (see §4.2.1); on the other hand, we exploit this morphological assumption to look for low-z interlopers (see §4.3.4).

We note that, despite extended Lyα emission being usually too faint for J-PLUS detection limits (see §1.4.2), extremely rare but sufficiently bright Lyα-emitting extended sources might still be observed within the very large area of J-PLUS DR1. These should be targeted by not relying on dual-mode catalogs but instead on analysing the 511 continuum-subtracted NB images of J-PLUS DR1 and applying specific source extraction criteria (as in e.g., Sobral et al., 2018b). We did not focus on these tasks since they deserve a separate and detailed analysis which lies outside the goals of our Lyα-emitting sources analysis.

4.3 Lyα-emitting candidates selection

Given the its photometric bands and nominal depth, we expect J-PLUS survey to be prone to non-negligible contamination by the most prominent QSOs lines listed in table 4.1, and to detect Lyα line at the redshift and luminosity detailed in table 4.2. Therefore, as a general procedure, we first look for secure *NB emitters*, i.e. objects showing a reliable NB excess for each of the four NBs we use. For this task, we use the method described in §4.3.1 (see also §2.2.5 for further details). We then exploit cross-matches with external databases and the remaining J-PLUS NBs to remove the highest possible number of low-z interlopers. This combination of photometric cuts and excluding conditions defines our final lists of candidates. Our selection rules are fully detailed in §4.3.2 and 4.4, while §4.4 describes the properties of the selected samples of candidates.

As detailed in §1.4.1, recent works have shown clear hints about a strong prevalence of AGN/QSOs in the distribution of Lyα-emitting sources at $L_{Ly\alpha} \gtrsim 3 \times 10^{43}$ erg s^{-1}. We broadly expect these findings to hold valid over the DR1 area, although the latter is bigger by a factor of ~ 500 with respect to commonly-surveyed areas in NB surveys of Lyα-emitting sources. Therefore, we expect to select a mixture of extremely Lyα-bright, rare star-forming galaxies (e.g., Sobral et al., 2016; Hartwig et al., 2016; Cai et al., 2017a; Shibuya et al., 2018a; Cai et al., 2018; Marques-Chaves et al., 2019) and luminous AGN/QSOs, numerically dominated by the latter. Indeed, our work selects objects showing strong and reliable NB excess, without employing any further criterion to disentangle its nature. Figure 4.2 shows typical spectra of high-z SF galaxies and QSOs, pointing out their significant diversity (see e.g., Hainline et al., 2011, for a comparison with narrow-line AGN spectra). Ideally, this difference should be mirrored by bi-modalities in the photometric properties of our selected samples, assuming that i) both the Lyα -emitting source classes are significantly present in our selection and ii) J-PLUS filters can effectively

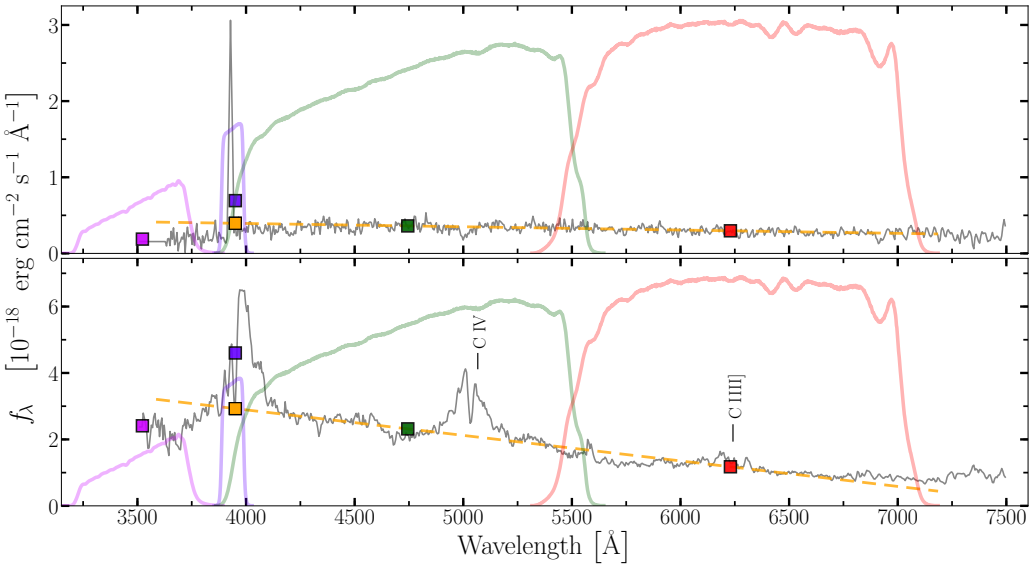

Figure 4.2: Representation of our NB excess detection method (§2.2.5) applied to J-PLUS bands. Grey lines in both panels show the observed spectra of typical z ∼ 2 Lyα-emitting sources (from the publicly available VUDS DR1 spectroscopic dataset, see e.g., Le Fèvre et al., 2015; Tasca et al., 2017). *Upper panel*: a SF LAE spectrum showing a single, prominent Lyα line (here redshifted at the observed wavelength $\lambda_{\rm obs} \sim 3900$ Å) and no other significant features. *Bottom panel*: a QSO spectrum with evident CIV and CIII] lines in addition to Lyα (here at $\lambda_{\rm obs} \sim 4000$ Å). We show the transmission curves and associated synthetic photometry of four J-PLUS bands respectively as colored lines and squares. From left to right, in each panel: u (purple), $J0395$ NB (violet), g (green) and r (red). In brief: our method uses g and r photometry for estimating a linear continuum (yellow dashed line in the plots) which is then evaluated at the NB pivot wavelength (yellow square). Finally, the ratio between the latter and the NB measurement (violet square) is used as a proxy for the Lyα line flux (see Eq. 4.1). By using u and g instead of g and r this method would provide a poorer handle of the non-linear continuum in the region affected by the Lyα line profile.

capture their spectral difference. In the following, we name as Lyα-emitting *candidates*, all the sources showing hints of a prominent Lyα line. We then look for bi-modalities in their photometric properties as hints to identify two distinct classes of Lyα-emitting objects. Where needed, we explicitly refer to either QSOs or SF LAEs to clearly state the distinction.

4.3.1 Detection of NB excess with a set of three filters

Our first goal is to determine whether a source exhibits a photometric excess in a given NB with respect to its own continuum emission. For this, it is first necessary to estimate the continuum emission of the source and then evaluate whether the NB measurement shows a significant discrepancy with respect to it. As detailed in §2.2.5, the method developed by Vilella-Rojo et al. (2015) and tested by Logroño-García et al. (2019) allows to directly: i) estimate the sources continuum, ii) evaluate the significance of the eventual NB excess and iii) separate the continuum and line emission from the NB photometry in order to extract a measurement for the line flux, namely: $F_{\rm Ly\alpha}$ in our case.

We apply the method of Vilella-Rojo et al. (2015) to all DR1 sources by using sets of three filters composed as: [NB; g; r], where NB stands for either $J0395$, $J0410$, $J0430$ or $J0515$. We checked that filter-sets defined as [NB; u; g] provide less accurate Lyα flux measurements than [NB; g; r] by employing spectroscopically identified z > 2 QSOs (Pâris et al., 2018). We therefore obtain estimates of the sources continua and line emission, which we ultimately use to determine their nature as NB-emitters or exclude them from our selection.

As detailed in Vilella-Rojo et al. (2015) and §2.2.5, we assume that: i) the emission line profile can be approximated by a Dirac-delta centered at a given wavelength $\lambda_{\rm EL}$, ii) the source continuum is well traced by a linear function over the wavelength range covered by the three filters. This latter hypothesis implies that NB measurements affected by an emission line should exhibit a photometric excess with respect to the straight line traced by g and r photometry. This is shown in Fig. 4.2, which graphically explains our method with two examples: a SF LAE and a QSO spectrum (respectively upper and lower panel) taken from VUDS public data (see Le Fèvre et al., 2015; Tasca et al., 2017). We note that the

NB measurements in both panels (purple squares) show a shift (i.e. a photometric excess) with respect to the linear-continuum extrapolations at $\lambda_{\rm EL}$ (yellow squares). The goal of our method is to measure these photometric excesses and relate them to the line fluxes which are producing them (see §2.2.5 and Vilella-Rojo et al., 2015, for technical details).

By using the equations and terminology detailed in §2.2.5, we can estimate the eventual NB excess due only to the presence of Lyα in a given NB as:

$$\Delta {\rm m}^{\rm NB} = {\rm m}^{\rm NB}_{\rm cont} - {\rm m}^{\rm NB} = 2.5\,{\rm Log}\left(\frac{{\rm f}^{\rm NB}_{\lambda}}{{\rm f}^{\rm NB}_{\lambda\,;\,{\rm cont}}}\right)\,, \tag{4.1}$$

where the last equality follows by the definition of AB magnitudes. We stress that here $\rm m^{NB}_{cont}$ and $\rm m^{NB}$ are respectively the linear extrapolation of the sources continuum (yellow squares in Fig. 4.2) and their NB measurement (purple squares in Fig. 4.2), both expressed in magnitude units. Therefore, $\Delta \rm m^{NB}$ is the photometric excess (or *color*, see §2.2.3) due to $\rm F_{Ly\alpha}$ (see §2.2.5).

We stress that $\Delta \rm m^{NB}$ is not itself a direct measurement (nor an estimate) of $\rm F_{Ly\alpha}$. Therefore, we only use $\Delta \rm m^{NB}$ and its associated uncertainty ($\sigma_{\Delta \rm m^{NB}}$) to isolate sources showing a reliable NB excess from the whole DR1 catalog (see §4.3.2). As a second step, we estimate $\rm F_{Ly\alpha}$ only for our *secure* candidates (defined in §4.4) by using the equations of the three-filters method (3FM for brevity, see §2.2.5 Vilella-Rojo et al., 2015), i.e. by computing $\rm F^{3FM}_{Ly\alpha}$. We anticipate that, in general, $\rm F^{3FM}_{Ly\alpha}$ measurements are a biased estimate of $\rm F_{Ly\alpha}$, as briefly introduced in §2.2.5. The methods we apply to account for these biases when targeting high-z Lyα emission are detailed in §4.6.1.

4.3.2 Selection function

We extract our $2.2 \lesssim {\rm z} \lesssim 3.3$ LAE candidates from a parent sample of ${\rm N} \sim 1.1 \times 10^7$ sources, obtained from the `J-PLUS` DR1 r-band selected, dual-mode catalog (see §4.2.1 and Cenarro et al., 2019). Our selection is designed to target strong NB excesses with respect to the BB-estimated continuum and to remove secure contaminants (see §4.4). Its overall performance was significantly improved thanks to the spectroscopic follow-up programs described in §4.5. The selection results are presented in §4.4, while the implications of using dual-mode catalogs are addressed in §4.6.1 and 4.6.4.

Magnitude cut in g and r bands

The photometry of too-bright or too-faint objects is likely to be either saturated or severely affected by noise. Hence we apply a very broad cut on g, r magnitudes and their associated errors (σ_g and σ_r), namely:

$$14 \leqslant g \leqslant 24 \;\bigwedge\; \sigma_g < 1 \quad ; \quad 14 \leqslant r \leqslant 24 \;\bigwedge\; \sigma_r < 1.$$

We check that these conditions do not significantly affect the final number of our candidates. Nevertheless, we account for eventual losses of continuum-faint sources with relatively bright Lyα emission (see §4.6.4). Spurious detections eventually included in these g and r intervals are removed by adequate SNR cuts (see below).

Detection confirmation in the three-filters set

We additionally require single-mode detection in each of the NB, g and r bands, since all are necessary for our excess-detection method. For this, we exploit the detection flags provided by the DR1 database[3]. This condition implies that we are only sensitive to low EW at faint Lyα flux; we account for this in our completeness estimates (see §4.6.4).

Effective exposure time cut

The normalized effective exposure time $\rm t^{eff}_{exp}$ (provided in the DR1) can be used as a proxy for the number of exposures contributing to the photometry of each source. The limit $\rm t^{eff}_{exp} > 0.5$ excludes objects whose detection is affected by the dithering pattern of `J-PLUS` pointings, which might compromise the removal of cosmic rays or their extraction process.

[3]For details see the information provided at: `http://archive.cefca.es/catalogues/jplus-dr1/help_adql.html`

`MANGLE` ***mask***

Sources' photometry can be affected by optical artifacts or bright stars. `J-PLUS` makes use of the `MANGLE` software (Swanson et al., 2008) in order to mask-out areas affected by these defects. For each of our selection, we apply the cumulative `MANGLE` mask associated to the three-filters [NB; g; r]. This reduces the total sky-coverage of our data to an effective area of $A_{\rm eff} \sim 900\,{\rm deg}^2$ (see Table 4.2 for details).

Pointing-by-pointing selection

The combined action of the previous cuts produce four different lists (one per NB) of $N \gtrsim 2 \times 10^6$ sources each (see Table 4.3). To proceed, we take into account that `J-PLUS` DR1 is composed by 511 different pointings (or *tiles*) which exhibit e.g. varying depths, source counts and colors. Consequently, we apply the following conditions on each tile separately build a selection function as uniform as possible.

NB excess significance

In order to select line-emitters candidates, we look for outliers in the $\Delta m^{\rm NB}$ vs. $m^{\rm NB}$ distribution of each tile, after considering photometric uncertainties (as in e.g., Bunker et al., 1995; Fujita et al., 2003; Sobral et al., 2009b; Bayliss et al., 2011; Matthee et al., 2017b, and Fig. 4.3). In particular, using Eq. 4.1 we compute the error:

$$\sigma_{\Delta m^{\rm NB}}(m^{\rm NB}) = \sqrt{\sigma^2_{m^{\rm NB}_{\rm cont}} + \sigma^2_{m^{\rm NB}}} \,, \tag{4.2}$$

and identify reliable NB-emitters as the objects satisfying:

$$\Delta m^{\rm NB} > \Sigma \cdot \sigma_{\Delta m^{\rm NB}} + \langle \Delta m^{\rm NB} \rangle \,, \tag{4.3}$$

with $\Sigma = 3$. We account for pointing variations by anchoring our cut to the average color $\langle \Delta m^{\rm NB} \rangle$ of each tile, which acts as a rigid offset. Figure 4.3 shows the results of this procedure on a `J-PLUS` tile with $\langle \Delta m^{\rm NB} \rangle = -0.27$. As expected, only $\lesssim 10 - 15\%$ of our parent-sample pass this cut (see Table 4.3).

NB signal-to-noise

We explicitly exclude objects with low-SNR NB measurements by imposing $m_{\rm NB} > m^{\rm NB}_{\rm cut}$, where $m^{\rm NB}_{\rm cut}$ is the NB magnitude at which the average NB SNR of each pointing is equal to 5. This threshold is relatively impacting on the whole DR1, since only $\sim 35\%$ of sources is able to pass it. We checked that imposing $\langle {\rm SNR} \rangle = 3$ would lead to significantly higher contamination of our selected samples.

BB signal-to-noise

Clean BB photometry is a key requirement to estimate the sources NB excess. We exclude objects with $g > g_{\rm cut}$ and $r > r_{\rm cut}$, where $g_{\rm cut}$ and $r_{\rm cut}$ are defined as the magnitudes at which $\langle {\rm SNR} \rangle = 5$ in each BB and pointing. Despite its effect on the parent samples being small (see Table 4.3), this cut might exclude genuine continuum-faint candidates with bright Lyα. We account for this as described in §4.6.4.

Minimum NB-color cut

In principle, Lyα can be distinguished from e.g. `CIV` and `CIII]` of AGN/QSOs spectra (e.g., Stroe et al., 2017a,b) or nebular H_β, $[OIII]_{4959+5007}$ and $[OII]_{3727}$ by exploiting its generally higher intrinsic strength and EW (e.g., Vanden Berk et al., 2001; Hainline et al., 2011; Selsing et al., 2016; Nakajima et al., 2018). Therefore, we impose a NB-color cut defined by assuming a minimum *rest-frame* EW for our candidates (as in, e.g. Fujita et al., 2003; Gawiser et al., 2006; Gronwall et al., 2007; Hayes et al., 2010; Adams et al., 2011; Clément et al., 2012; Santos et al., 2016). Observed- and rest-frame EWs (respectively $EW_{\rm obs}$ and EW_0) are related via:

$$EW_{\rm obs} = EW_0 \, (1 + z) \,. \tag{4.4}$$

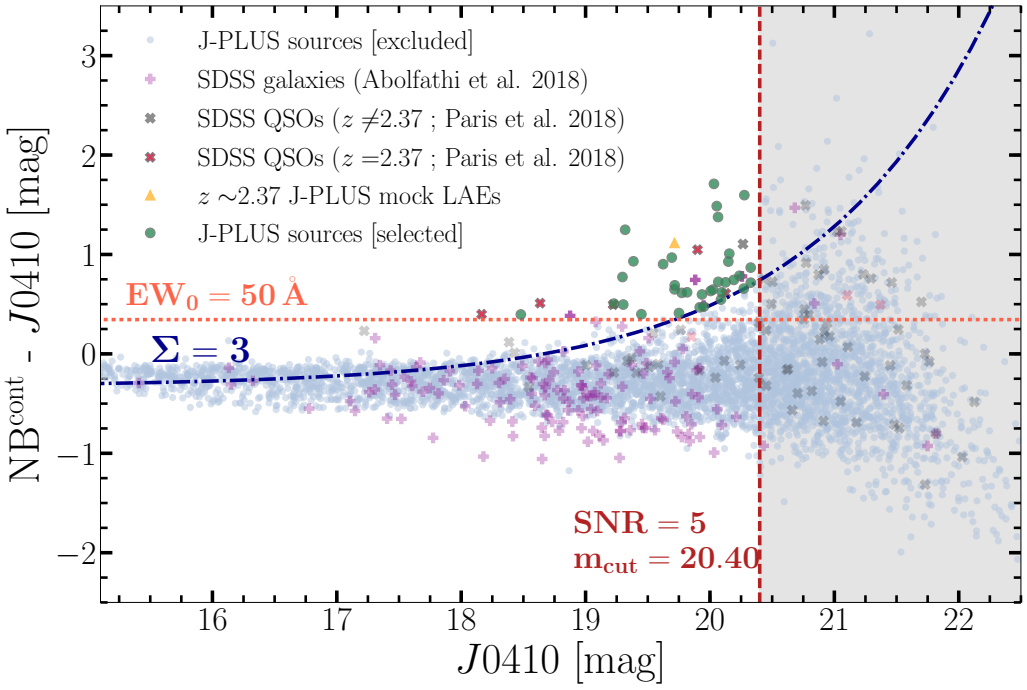

Figure 4.3: Example of a color-magnitude diagram obtained for the NB filter $J0410$ on a DR1 pointing (out of 511). Our photometric cuts are summarized as follows: the blue dashed-dotted line shows the Δm^{NB}-significance threshold, while the vertical red line marks the NB SNR limit. We exclude sources below the blue dashed-dotted line and inside the grey shaded area. The orange horizontal dotted line shows Δm^{NB} associated to EW = 50Å (see Eq. 4.5). Grey-blue dots mark all the `J-PLUS` detections in the pointing, while red and purple crosses show $z \sim 2.4$ QSOs and low-z galaxies from `SDSS DR14`. Yellow triangles show `J-PLUS` mock data of $z \sim 2.4$ SF LAEs (Izquierdo-Villalba et al., 2019b). Finally, our Lyα-emitting candidates are shown as green dots.

We set $EW_0^{min} = 50\,Å$ and obtain the corresponding EW_{obs}^{min} from Eq. (4.4). We then link EW_{obs} and Δm^{NB} (defined in Eq. 4.1) with the analytic expression:

$$\Delta m_{min}^{NB} = 2.5\,\mathrm{Log}\,\left(1 + \beta_{NB} \cdot EW_{obs}^{min}\right) + \langle m^{NB} \rangle\ , \tag{4.5}$$

(see Guaita et al., 2010, and §2.2.5), where β_{NB} is defined in Eq. 2.41 and $\langle m^{NB} \rangle$ is the average color in the tile. By requiring $\Delta m^{NB} > \Delta m_{min}^{NB}$ (orange horizontal dotted line in Fig. 4.3) we exclude $\gtrsim 96\%$ of our DR1 parent sample, since most sources do not show line-emission. We note that the choice of EW_0^{min} has a certain degree of arbitrariness indeed past works have explored a wide range of limiting values (see e.g., Gronwall et al., 2007; Ouchi et al., 2008; Bond et al., 2009; Nilsson et al., 2009b; Guaita et al., 2010; Konno et al., 2016; Matthee et al., 2016; Bădescu et al., 2017; Sobral et al., 2017). We fix $EW_0 = 50$Å after checking our EW estimates on publicly-available spectroscopic catalogs of $z \gtrsim 2$ SF LAEs and QSOs (namely `DR14`, VUDS and VVDS Cassata et al., 2011; Le Fèvre et al., 2015; Pâris et al., 2018) and on the confirmed $z \sim 2$ QSOs in our follow-up data (see Table 4.7 in §A). In particular, 50Å provides a good compromise between the retrieval of $z \gtrsim 2$ sources and the exclusion of $z < 2$ interlopers. We note that this relatively high EW_0^{min} is still close to the lower limits of EW distributions usually measured for high-z Lyα-emitting sources (e.g., Nilsson et al., 2009a; Bond et al., 2012; Amorín et al., 2017; Hashimoto et al., 2017; Santos et al., 2020). Besides, low EWs can be accessed with very-narrow bands (FWHM $\lesssim$ 50Å) and deep observations ($r > 22$, e.g., Sobral et al., 2017), which both act as limiting factors in our case. Finally, we stress that this condition is not directly applied on EW_0, hence it does not pose a strict limit on the measured EW of our candidates (see Ouchi et al., 2008, for a similar discussion).

These cuts select $< 1\%$ of the parent catalog respectively for each NB (see Table 4.3 for the exact number). These samples are still likely to be contaminated by interlopers, such as lower-z QSOs, ELGs and faint blue stars, which are usually targeted with BB-based color cuts (e.g., Ross et al., 2012; Ivezić et al., 2014; Peters et al., 2015; Richards et al., 2015). We checked that, in our case, these methods significantly affect

also the number of selected $z \gtrsim 2$ QSOs from SDSS DR14. We hence decided to drop any color cut because of its non-trivial effect on our selection.

4.3.3 Definition of Lyman-α emitting candidates

Despite efficiently identifying NB-emitters, the conditions in §4.3.2 might also select line-emitting interlopers (see §4.2.2). Previous works based on similar methods have usually explored limited sky regions already surveyed by deep multi-wavelength data, which supported the identification of contaminants (e.g. COSMOS, UDS, SXDS, SA22 and Boötes fields, see Warren et al., 2007; Scoville et al., 2007; Furusawa et al., 2008; Geach et al., 2008; Kim et al., 2011; Bian et al., 2012; Stroe & Sobral, 2015). Unfortunately, few previous surveys uniformly cover the very wide area of J-PLUS DR1, hence limiting our ability to identify contaminants.

Cross-matches with public external databases

Interlopers with a secure identification (either spectroscopic, astrometric or photometric) can be removed via cross-matches with public catalogs. We employ a radius of $r_{max}^{match} = 3.5''$ after checking that this provides a high matching completeness while keeping low the number of multiple matches, for all the matched databases. More in detail, we recover the 80% (95%) of all QSOs from SDSS DR14 (within the DR1 footprint) respectively at $r \sim 21.25$ ($r \sim 20.80$) and $\mathrm{Log}\,(L_{Ly\alpha}) \sim 44.25$ (~ 44.70).

SDSS DR14

We exploit the lists of spectroscopically-identified galaxies (Bundy et al., 2015; Hutchinson et al., 2016), stars (Majewski et al., 2017) and QSOs (Pâris et al., 2018) provided by the recent SDSS-IV DR14 (DR14 hereafter, Blanton et al., 2017; Abolfathi et al., 2018). Given the wide overlap with J-PLUS DR1 and the higher depth of DR14 (Cenarro et al., 2019), this cross-match ensures the removal of secure contaminants from our selection. As discussed in §4.2.2, QSOs can act as both interlopers and genuine candidates depending on their z, hence we need to rely on a list of *securely* identified QSOs. The Pâris et al. (2018) catalog includes $N \gtrsim 5.3 \times 10^5$ sources observed by BOSS and eBOSS surveys (Dawson et al., 2013, 2016) and confirmed as QSOs by careful inspection. We keep genuine Lyα-emitting sources at the z sampled by each NB, while the rest are identified as contaminants and removed. The cross-match with DR14 shows a generally low contamination (table 4.4), with low-z galaxies accounting respectively for 5.1%, 4.3%, 5.3% and 3.1% of our *J*0395, *J*0410, *J*0430 and *J*0515 NB samples. On the other hand, the $z \lesssim 2$ QSOs fraction drops from 11.1% to 0.3%, paralleling the drop of DR14 $z \gtrsim 2.2$ QSOs. Finally, SDSS stars account for $\lesssim 2\%$ of our samples. These fractions are likely to be underestimated, given the different depth of the two surveys and eventual mis-matches between DR14 and DR1 catalogs. Nevertheless, being measured on spectroscopically confirmed sources, these are secure contamination estimates.

Gaia DR2

Our spectroscopic follow-up program 2018A (see §4.5) showed a non-negligible contamination from stars in our samples. To limit this issue, we built a specific criterion for excluding stars, based on the very accurate measurements offered by Gaia DR2 data (Gaia Collaboration et al., 2018). We identify *secure stars* by using the significance of proper-motion assessments. More in detail, we exclude the J-PLUS sources with a counterpart in Gaia DR2, showing significant measurements ($\sigma > 3$) in each proper motion component, i.e.:

$$\sigma_{\mathrm{pm}} = \sqrt{\sigma^2_{\mathrm{pmra}} + \sigma^2_{\mathrm{pmdec}} + \sigma^2_{\mu}} > \sqrt{27} \sim 5.2\ , \tag{4.6}$$

where σ_{pmra}, σ_{pmdec} and σ_{μ} are respectively the errors on proper motion (ra and dec) and parallax. With this cut, we explicitly remove objects showing significant apparent motion from our list of LAE candidates. The good performance of this criterion was confirmed by the results of our second follow-up program, whose targets were selected from the results of our updated pipeline (see §4.5 for details). The contamination from Gaia DR2 is presented in Table 4.4.

Filters	DR1 parent sample	Δm^{NB} significance	NB SNR	BB SNR	EW_{obs}^{min}	First selection
*J*0395	2,036,657	348,613 (17.1%)	1,324,373 (65.0%)	2,017,720 (99.1%)	57,800 (2.8%)	12,251 (0.6%)
*J*0410	2,730,135	232,753 (8.5%)	1,846,144 (67.6%)	2,679,515 (98.2%)	150,321 (5.5%)	19,905 (0.7%)
*J*0430	3,015,684	235,685 (7.8%)	2,024,629 (67.2%)	2,930,026 (97.2%)	173,388 (5.8%)	24,813 (0.8%)
*J*0515	4,520,911	244,550 (5.4%)	2,956,154 (65.4%)	3,797,178 (84.0%)	143,662 (3.2%)	15,213 (0.3%)

Table 4.3: Number counts of sources passing each cut of our selection, for the four J-PLUS NBs we use. Here we report the effects of each cut separately to highlight its effect, hence the fractions reported in the Table do not add to 100%. The most impacting cuts are those on EW_{obs}^{min} cut and on NB excess significance. The number of sources passing all these conditions, for each NB, produce our final samples of LAE candidates and is shown in the last column to the right. These *partial* samples are likely to be highly contaminated by interlopers showing reliable NB excess. Table 4.4 shows a summary of the samples contamination and the final number of selected sources.

GALEX-UV

Lyα-emitting sources at $z > 2$ are generally expected to appear faint at (observed) UV wavelengths due to the dimming action of the Lyα -break and Lyman-break (e.g., Steidel & Hamilton, 1992; Steidel et al., 1996, 1999; Shapley et al., 2003). On the contrary, $z < 2$ AGN/QSOs, blue stars and low-z star-forming galaxies can show significant UV emission. We exploit this property for removing $z < 2$ interlopers by cross-matching our catalogues with GALEX all-sky UV observations (Gil de Paz et al., 2009). In particular, we remove sources with a $SNR > 3$ detection in either of the two FUV and NUV GALEX bands (see e.g., Ciardullo et al., 2012, and table 4.4). In order to check our assumption according to which only $z < 2$ sources are expected to be significantly observed in UV, we additionally matched the J-PLUS sources with counterparts in GALEX to the spectroscopic sample of DR14. This analysis confirmed that $> 99.5\%$ of sources with UV-bright GALEX detection show a spectroscopic $z < 2$, hence act as contaminant in our selection.

LQAC-3

The third release of the Large Quasar Astrometric Catalog (Souchay et al., 2015a,b) is a complete archive of spectroscopically identified QSOs. By combining data from available catalogs, it provides the largest complement to the DR14 list (Pâris et al., 2018). We exclude sources included in LQAC-3 with spectroscopic z lying outside the range probed by each NB. As expected, this step identifies only few additional interlopers (see Table 4.4).

Multiple NB excesses

We target additional interlopers by exploiting the whole set of J-PLUS NBs. Indeed, we expect SF LAEs to not show any additional NB feature (e.g., Shapley et al., 2003; Nakajima et al., 2018), while QSOs at the targeted z can exhibit only particular combinations of NB excesses. Consequently, we remove candidates with multiple NB excesses incompatible with $z > 2$ spectral features (e.g., Matthee et al., 2017b). On the other hand, excesses consistent with $z > 2$ can also be produced by low-z interlopers. As an example, Fig. 4.4 shows the photo-spectra of a $z \sim 0.05$ galaxy (upper panel) and a $z \sim 2.25$ QSO (bottom panel) from the DR14 spectroscopic samples. Both sources show simultaneous excesses in *J*0395 and *J*0515 filters (respectively purple and yellow empty squares) with respect to the linear continuum traced by *g* and *r* BBs (yellow dashed line). On top of this, both photo-spectra exhibit comparable BB colors and might hence be confused by our selection. Since J-PLUS data do not allow to tell apart these objects and measure this source of contamination, we estimate a statistical correction as explained in §4.6.3.

4.3.4 Morphological criterion

We expect our $z \gtrsim 2.2$ candidates to appear compact in J-PLUS data (see §4.2.2), hence sources showing extended morphology are likely to be low-z interlopers. The DR1 catalog provides a parameter $\mathcal{C}$ which allows to discriminate between compact ($\mathcal{C} \sim 1$) and extended objects ($\mathcal{C} \sim 0$, see López-Sanjuan et al., 2019b, for details). By cross-matching the whole DR1 sample to SDSS spectroscopic catalogs of galaxies and QSOs, we checked that more than $\gtrsim 90\%$ of galaxies in SDSS ($z \lesssim 1$, see Hutchinson et al., 2016) and

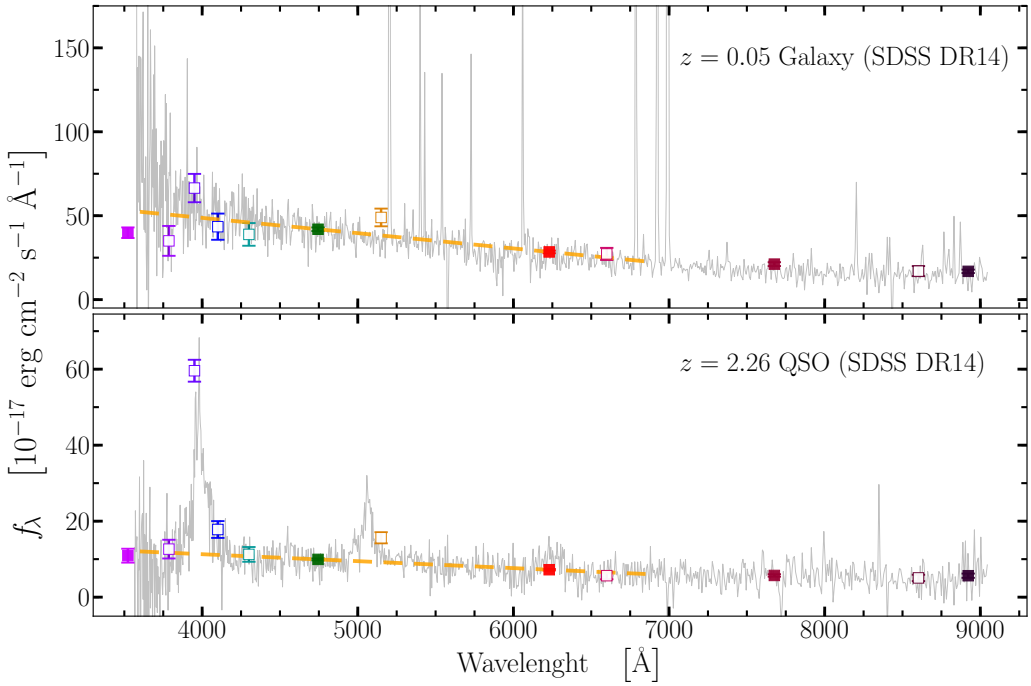

Figure 4.4: Examples of multiple NB excess in J-PLUS photo-spectra. Empty and filled squares respectively mark NB and BB photometry, while the dashed yellow line shows the linear continuum we estimate through g and r BBs (in green and red, respectively). This comparison shows that both a z$\sim$0.05 galaxy (upper panel) and a z$\sim$2.25 QSO (lower panel) exhibit significant excesses in $J0395$ and $J0515$ NBs (respectively second and fifth empty squares from the left). The $J0515$ excess is produced respectively by H$_\beta$ at z$\sim$0.05 and CIV line at z$\sim$2.25, but its nature is hardly distinguishable by J-PLUS photometry.

only $\lesssim 5\%$ of DR14 QSOs (at any z) are found at $\mathcal{C} \leqslant 0.1$. We then remove objects with $\mathcal{C} \leqslant 0.1$ from our selection (see previous-to-last column to the right in table 4.4).

4.4 Analysis of the selected samples of Lyα-emitting candidates

The final results of our selection procedure are four samples of z$>$2 Lyα-emitting candidates which meet all the following requirements: i) reliable excess in the NB used for their selection, ii) secure detection and photometry in the filter triplet [NB; g; r], iii) no spectroscopic counterparts in DR14 with redshift outside the ranges probed by each NB, iv) no apparent motion according to Gaia DR2 data, v) no significant observed-frame UV detection in GALEX, vi) compact morphology and, eventually, vii) multiple NB excesses compatible with being z$\gtrsim$2 sources. These lists account for 2547, 5556, 4994, 1467 sources respectively for $J0395$, $J0410$, $J0430$ and $J0515$ NBs (see Table 4.4), which translates into approximately 2.8, 6.2, 5.6 and 1.5 objects per squared degree, respectively. We underline that these samples are the largest-to-date collections of Lyα-emitting candidates within the narrow redshift bins we can access to (see e.g., Guaita et al., 2010; Cassata et al., 2015; Konno et al., 2016; Matthee et al., 2017b; Sobral et al., 2018b).

The drop of number counts for $J0515$ NB can be ascribed to the combination of J-PLUS data depth and the cosmological decrease of bright SF LAEs and AGN/QSOs number densities at z$\gtrsim$2.5 (e.g., Nilsson et al., 2009b; Ciardullo et al., 2012; Sobral et al., 2018b). Indeed, the right panels of Fig. 4.5 show that $J0515$ NB can only access to ranges of Log (L$_{\mathrm{Ly}\alpha}$) and Lyα Log (EW$_0$) which are significantly higher than the other NBs. In general, filters sampling smaller wavelengths can access to fainter Lyα luminosity and smaller EW$_0$, as a result of the combination between J-PLUS depth and the probed z interval.

EW$_0$ and L$_{\mathrm{Ly}\alpha}$ distributions

The left panels in Fig. 4.5 show the distribution of EW$_0$ measured on our samples, as a function of both r magnitude and $g-r$ color. As commented in §4.3.2, our cut on Δm^{NB} derives from a theoretical expected

Filters	First selection	SDSS spectra	GALEX	Gaia DR2 stars	LQAC QSOs	Multiple NB	Extended	Final [N; deg^{-2}]
*J*0395	12,251	2,192 (17.9%)	2,003 (16.4%)	857 (7.0%)	87 (0.7%)	1,312 (10.7%)	6,307 (51.5%)	2,547 ; 2.8
*J*0410	19,905	1,983 (9.9%)	2,003 (10.1%)	2,738 (13.8%)	56 (0.3%)	16,48 (8.3%)	9,557 (48.0%)	5,556 ; 6.2
*J*0430	24,813	2,083 (8.4%)	2,597 (10.5%)	2,441 (9.8%)	40 (0.2%)	3,313 (13.4%)	15,468 (62.3%)	4,994 ; 5.6
*J*0515	15,213	523 (3.4%)	1,249 (8.2%)	531 (3.5%)	7 (0.05%)	1,282 (8.4%)	12,992 (85.4%)	1,467 ; 1.5

Table 4.4: Number counts (and fractions) of secure interlopers among the sources passing our photometric selection, for each J-PLUS NB we use (see discussion in §4.4). We note that the extended fraction of our samples is particularly high for the *J*0515 NBs, suggesting that this filter is affected by high level of contamination from extended low-z interlopers. Indeed, this is the only NB among the four which is susceptible to contamination from the strong [OIII]$_{4959+5007}$ doublet and H$_\beta$ line, in addition to [NeIII] and [OII]. Sources with at least one identification as secure interloper are excluded; the final number counts of Lyα-emitting candidates are shown in the last column to the right. The average sky density of these objects shows significant variation among the four lists, with an average of $\sim 4\,\mathrm{deg}^{-2}$ sources, per filter. The complete catalogs of genuine candidates (i.e. after excluding securely identified interlopers) can be found at: https://www.j-plus.es/ancillarydata/dr1_lya_emitting_candidates

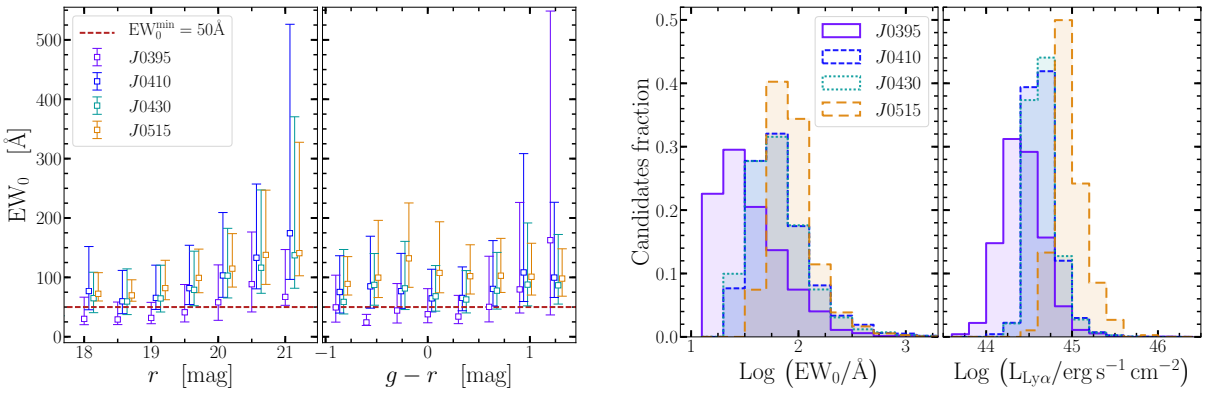

Figure 4.5: *Left figure:* EW$_0$ distribution of our selected candidates as a function of r and $g-r$ color (left and right panels, respectively). Squared points and error bars show respectively the distribution median and 16th-84th percentiles, in each magnitude and color bin. Points have been artificially shifted for a better visualization. The values for *J*0395 filter at $r < 20$ ($g - r < 0.75$) are systematically below the theoretical EW$_0^{\rm min} = 50$Å cut we apply (§4.3.2). This is due to the little overlap beteween this NB and the g BB, which is reflected into a poor extrapolation of the linear-continuum at the NB filter pivot wavelength (see e.g., Ouchi et al., 2008, and the discussion in §4.3.2). *Right figure:* Normalized distributions of our candidates in EW and Lyα luminosity, for each filter. This result clearly shows that filters sampling higher redshifts also sample brighter Lyα luminosity. This is a direct effect of J-PLUS detection limits which only allow to observe brighter and rare objects at higher redshifts. We address this issue by applying the completeness corrections described in §4.6.4.

limit of EW$_0 = 50$ Å. Nevertheless, not all selected sources display EW$_0 > 50$ Å (see also Ouchi et al., 2008, for a similar discussion). This is evident for the *J*0395 NB candidates, whose EW$_0$ distribution is systematically below 50 Å for $r \lesssim 20$ and $g - r \lesssim 0.75$. Indeed, the little overlap between *J*0395 and g transmission curves ultimately provides a relatively poor extrapolation of the linear continuum up to the pivot wavelength of *J*0395, which translates into an under-estimation of EW$_0$. This induces a bias on our Lyα luminosity measurement, which we account for as described in §4.6.1. On the other hand, no significant nor systematic bias affects EW$_0$ with respect to color, as shown by the flat $g-r$ distribution in Fig. 4.5. We confirmed this by using the spectra of z$\sim$2 DR14 QSOs, but we do not show the results for the sake of brevity.

Overall, our distributions are broadly consistent with previous determinations of the rest-frame EW of z$\sim 2-3$ LAEs (either SF LAEs and AGN/QSOs, see e.g., Gronwall et al., 2007; Guaita et al., 2010; Hainline et al., 2011; Ciardullo et al., 2012; Shibuya et al., 2014; Hashimoto et al., 2017; Santos et al., 2020). Interestingly, our samples include a moderate fraction of sources ($\lesssim 7\%$, on average) showing EW$_0 > 240$ Å (e.g., Ouchi et al., 2008; Santos et al., 2020). High-EW LAEs have been studied with particular interest (e.g., Cantalupo et al., 2012; Kashikawa et al., 2012; Shibuya et al., 2014) since nebular emission of Pop-II stellar populations can only account for EW$_0^{\rm Ly\alpha} \lesssim 500$Å (e.g., Charlot & Fall, 1993; Hernán-Caballero et al., 2017). At the same time, high Lyα EWs can be easily produced by AGN/QSOs which are likely to dominate our selected samples. Since analysing high-EW LAEs would require a careful separate analysis, we refrain to comment further on this topic. Nevertheless, we underline that our lists of selected candidates can provide catalogs of high-EW LAE targets for upcoming studies.

Estimate of the relative abundance of QSOs and SF LAEs

The design of J-PLUS filters potentially allows to capture at the same time peculiar combination of high-z lines with different NBs (see §4.3.3). For instance, QSOs emitting Lyα at z$\sim$2.3 could show simultaneous NB excesses in $J0410$ and $J0515$ NBs (the latter being due to CIV emission). This offers the possibility of investigating the relative fraction of AGN/QSOs and SF LAEs in our samples, since the latter should not exhibit such double-NB emission. We hence separate the DR14 QSOs selected with $J0410$ from the rest of $J0410$ candidates and plot the color distribution of these two source classes.

Figure 4.6 shows the color space defined by $J0410$ and $J0515$ NBs with respect to g BB. In this plane, both SF LAEs and QSOs should exhibit $g-J0410>0$ due to Lyα emission, but the CIV line (usually much stronger in QSOs than in SF LAEs, see e.g., Hainline et al., 2011; Stroe et al., 2017a,b; Nakajima et al., 2018) should displace QSOs at $g-J0515>0$. Nevertheless, our color distribution (blue solid histogram) does not show any evident bimodality and no significant overdensity at $g-J0515\sim0$, as expected for SF LAEs. This suggest that either i) our $J0410$ candidates are mostly dominated by z$\sim$2 QSOs or ii) the J-PLUS filter set does not reliably disentangle the different spectral features of high-z SF LAEs and AGN/QSOs.

We further analyse the nature of our candidates by exploiting the cross-match with the all-sky WISE data (Wright et al., 2010) provided by the J-PLUS DR1 database. In particular, we compare the $r-$W1 color of our candidates with WISE counterparts to the synthetic-photometry color-tracks of galaxy and QSOs templates (respectively from Polletta et al., 2007, and Hernán-Caballero et al. 2016). Figure 4.7 shows how the color of our candidates are clearly compatible with the ones of QSOs, while being significantly different from the galaxy ones. We also highlight the SDSS QSOs (black-contoured dots) and the confirmed QSOs of our GTC programs (yellow-contoured dots, only for $J0395$ NB) to underline the comparability of our candidates properties with those of spectroscopically-confirmed QSOs.

Interestingly, by joining these evidences with the results of our spectroscopic follow-up programs, we expect our samples of candidates to be dominated by z > 2 QSOs which yet lack a spectroscopic identification. Indeed, by considering the number of our genuine candidates without SDSS identification (namely 2057, 4959, 4494 and 1377 respectively for $J0395$, $J0410$, $J0430$ and $J0515$ NBs) and conservatively applying a residual contamination of $\gtrsim35\%$ (as suggested by our GTC follow-up) our method identifies for the first time as z>2 QSOs respectively $\gtrsim1300, 3200, 2900$ and ~900 z>2 sources in each NB. This is shown in both Fig. 4.6 and 4.7 by the wide difference between the number counts of DR14 QSOs within our selection and our remaining genuine candidates. We interpret this as an effect of the NB-based selection we perform, which efficiently targets the line-emission features of these objects, eventually missed by previous target-selections based on BB-colors (e.g., Richards et al., 2009; Ross et al., 2012; Ivezić et al., 2014). Nevertheless, a systematic and uniform spectroscopic confirmation of our samples is needed to validate these findings.

4.5 Spectroscopic follow-up at the Gran Telescopio Canarias

This section presents two spectroscopic follow-up programs executed at the Gran Telescopio Canarias (GTC) telescope[4] in the semesters 2018A and 2019A. The spectroscopic confirmation of a sub-sample of our candidates allowed to assess the performance of our selection, to refine our methodology and to estimate its residual contamination. Overall, these programs confirmed 45 sources selected among our $J0395$ NB-emitters.

4.5.1 Programs description

To ensure uniform observations and comparable results, we performed the same target selection and required identical observing conditions for both programs (namely GTC2018A and GTC2019A). In particular, we randomly selected a sample of 24 (21) Lyα-bright candidates ($L_{Ly\alpha}>10^{43.5}\,erg\,s^{-1}$) for program GTC2018A (GTC2019A), spanning the entire luminosity range covered by our candidates. We stress that targets for GTC2019A were selected after refining our selection with the help of GTC2018A

[4]Observatorio del Roque de los Muchachos, La Palma, Canary Islands

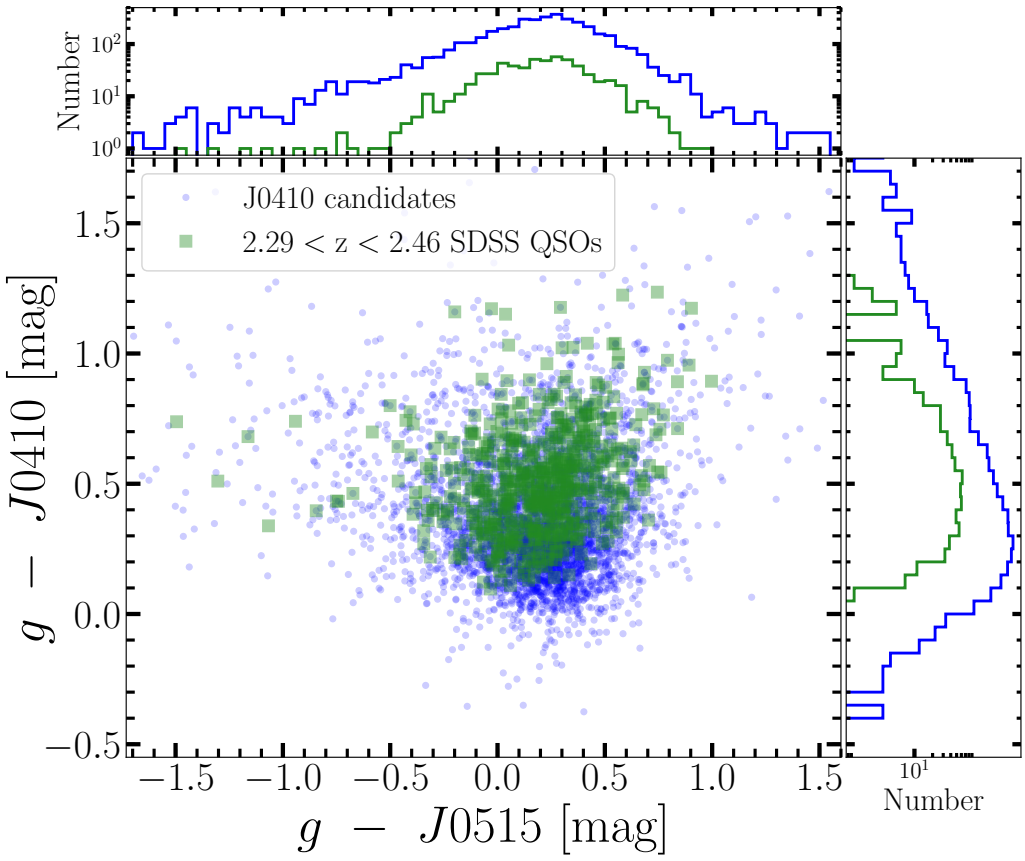

Figure 4.6: Comparison between the color-color distributions of DR14 QSOs (green squares) within our *J*0410 sample and of the remaining *J*0410 genuine candidates (blue dots, after removing known interlopers; §4.4). The two source classes occupy comparable color regions, suggesting that our selection results might be effectively dominated by z∼2.3 AGN/QSOs. This scenario is also supported by the results of our spectroscopic program (§4.5).

results. We requested to use the OSIRIS spectrograph and the R500B grism, in order to exploit its good spectral resolution (R∼500, which translates to $\Delta\lambda\,\mathrm{pixel}^{-1}\sim3.65$Å for the 0.8" slit width we requested). The exposure times for our targets were computed by assuming the observing conditions summarized in the header of Table 4.7 (§A). These were calibrated to achieve SNR $\geqslant 3$ (in each λ bin) over the whole OSIRIS spectral range, in order to identify eventual emission lines and measure their integrated flux.

We limited our programs length to <20 hours, to ensure their completion. Due to the high observing times required by our targets, we followed-up only candidates selected by *J*0395 NB. The target selection balanced the total observing time and the uniform sampling of our candidates $L_{Ly\alpha}$ distribution. Finally, we excluded objects with previous spectroscopic identifications (at any z). Our proposals were respectively awarded with 11.56 and 18.95 hours of observations and were both fully executed.

4.5.2 Spectroscopic results

The results of both programs are shown in Table 4.7. Overall, we identified 29/45 targets (64.4%) as genuine z∼2.2 Lyα-emitting sources, 8/45 (17.7%) as z∼1.5 QSOs emitting CIV at $\lambda_{obs}\sim4000$ Å, 1/45 (2.2%) Lyβ-emitting QSO at z∼2.76, 5/45 (11.3%) blue stars and 2/45 (4.4%) low-z galaxies selected because of their narrow emission lines. As an example, Fig. 4.8 shows a spectra for each different source class together with its associated J-PLUS photometry. Both z∼2.2 and z∼1.5 QSOs show prominent line emission at $\lambda_{obs}\sim3950$ and are consequently selected as genuine *J*0395 NB-emitters. The same applies to the z∼2.8 QSO emitting Lyβ at $\lambda_{obs}\sim3950$. On the contrary, the remaining sources do not show significant spectral features, indeed their selection is due to strong blue colors combined to a barely-significant NB-excess (see e.g. third panel from above). In particular, the star and galaxy interlopers (i.e. third and fourth panels from the top) were picked as targets before we refined our selection rules and the J-PLUS DR1 was re-calibrated (López-Sanjuan et al., 2019a). With the current

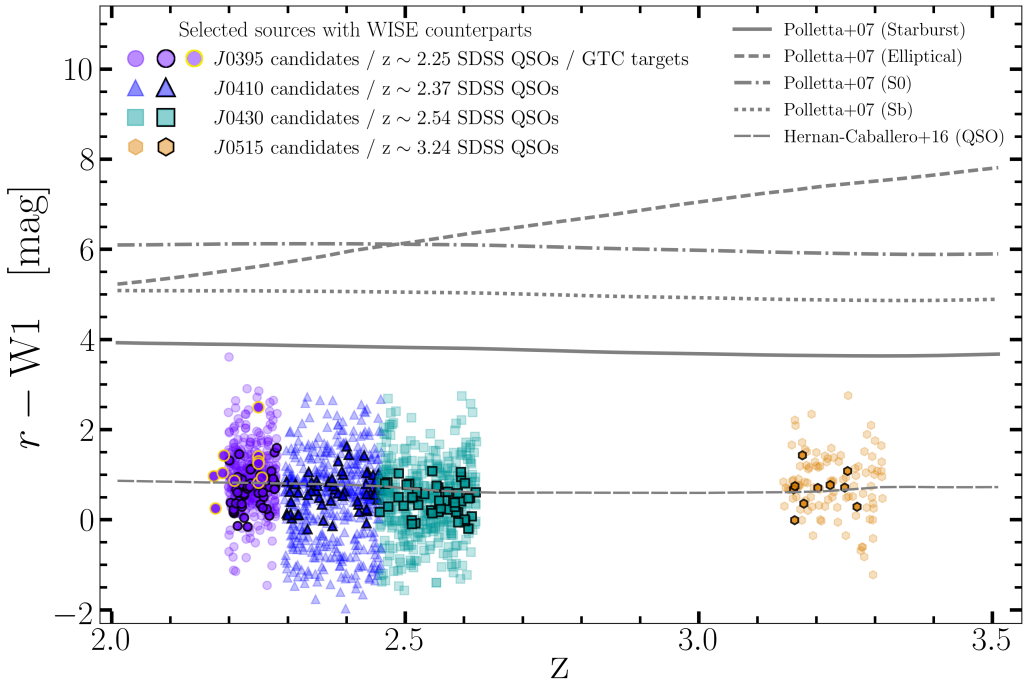

Figure 4.7: *Colored points*: $r - \mathrm{W1}$ color versus redshift of all our candidates with a counterpart in WISE. Our sources are uniformly spread within the z interval sampled by each NB. *Grey lines*: tracks for different galaxy templates (from the SWIRE library, Polletta et al., 2007) and the QSO template of Hernán-Caballero et al. (2016). Black-contoured and yellow-contoured points respectively mark the SDSS QSOs in our selection and the GTC targets (only for *J*0395 NB). Our candidates are all compatible with the high-z QSO template, hence suggesting that the fraction of SF LAEs in our sample is very low. In addition, this suggests that we identify as high-z QSOs a large number of sources without previous spectroscopic identification.

`J-PLUS` photometry and our updated selection these objects are not re-selected (right column of Table 4.5). Given the absence of emission lines at $\lambda_{\rm obs}\sim$3900Å for these objects, their low-significance NB-excess is likely due to imperfections in their photometry. In the case of the z$\sim$0.5 galaxy (fourth panel from the top in Fig. 4.8), we additionally observe a discrepancy between the spectrum and `J-PLUS` data. A number of possible explanation can account for this, such as errors in the spectrum extraction and calibration, too-low spectroscopic SNR at $\lambda_{\rm obs} \lesssim 4500$ Å or artifacts biasing only the *J*0395 photometry. On the contrary, the excess of the z $\sim$ 2.8 QSO (bottom panel in Fig. 4.8) is due to the Lyβ line redshifted at $\lambda_{\rm obs} \sim 3950$ in the observed spectrum, although in tension with `J-PLUS` photometry. In this case, QSO variability might play a role (e.g., Hook et al., 1994; Kozłowski, 2016) as well as photometric imperfections.

Overall, 40/45 targets (88.9%) are genuine line emitters, hence confirming the efficiency of our selection. Moreover, the stars contamination is reduced from $\sim$17% to $\lesssim$5% between the two programs (see Tables 4.5 and 4.5). Indeed, guided by the `GTC2018A` results, we i) excluded sources with significant apparent motion according to `Gaia DR2` and ii) selected $\mathrm{EW}_0 = 50$Å as our limiting value for defining the $\Delta\mathrm{m}^{\rm NB}$ cut (see §4.3.2). Our improved methodology retrieves 15/24 original `GTC2018A` targets, with 11/15 ($\sim$ 74%) being z $\sim$ 2.2 QSOs and 4/15 ($\sim$ 26%) being z $\sim$ 1.5 QSOs (i.e. no star is re-selected). Nevertheless, four out of 15 z$\sim$2.2 QSOs are not re-identified as line emitters. This is partly due to the new calibration of the entire `J-PLUS` data (occurred after `GTC2018A` López-Sanjuan et al., 2019a), indeed two out of the four z$\sim$2.2 QSOs which are not re-selected fail to pass the NB SNR criterion due to their recomputed NB photometry. Finally, the fraction of genuine NB-emitters significantly increased from $\sim$74% for `GTC2018A` to over 95% for `GTC2019A` thanks to our methodology improvements.

4.6 The computation of the Lyman-α luminosity function

The luminosity function Φ(L) of a given class of sources is usually defined as their comoving number density per unit luminosity (see e.g., Schmidt, 1968). Following a common convention in literature, we

Objects spectroscopic class	GTC2018A results			GTC2019A results	
	Fraction #	Percentage (%)	Retrieved after improvement	Fraction #	Percentage (%)
z $\sim$ 2.2 QSOs	15/24	62.5%	11/15	14/21	66.6%
z $\sim$ 1.5 QSOs	4/24	16.7%	4/15	4/21	19.0%
z $\sim$ 2.8 QSOs	-	-	-	1/21	4.8%
Stars	4/24	16.7%	0/15	1/21	4.8%
Low-z Galaxies	1/24	4.1%	0/15	1/21	4.8%

Table 4.5: Number counts resulting from the GTC2018A (left) and GTC2019A (right) programs. Targets are divided in: z $\sim$ 2.2 QSOs and four contaminant species. Among these, z $\sim$ 2.76 QSOs are only present in the GTC2019A results. All targets except those confirmed as stars are genuine $J0395$ NB-emitters due to, respectively: Lyα , CIV, Lyβ and [OIII] emission lines. The contamination from blue stars in GTC2019A significantly dropped to $\lesssim$ 5% (from $\sim$ 17% in GTC2018A results), mainly due to the cross-match with Gaia DR2 data.

express our LFs in logarithmic units of luminosity and hence use the following definition:

$$\Phi[\,\mathrm{Log}\,(\mathrm{L_{Ly\alpha}})] \;=\; \frac{\sum_{\mathrm{i}}\,(\mathrm{P_i}\,/\,\mathrm{C_i})}{\mathrm{V}\;\cdot\;\Delta\,\mathrm{Log}(\mathrm{L_{Ly\alpha}})}\;, \tag{4.7}$$

where the sum at the numerator is extended to all the objects in a given bin of (logarithmic) luminosity $\Delta\,\mathrm{Log}(\mathrm{L_{Ly\alpha}})$, while the coefficients P_i and C_i are statistical weights that account respectively for the sample purity and completeness (as detailed below). We exploit our lists of candidates selected with $J0395$, $J0410$, $J0430$ and $J0515$ NBs to build four determinations of the Lyα LF at the redshifts given by Table 4.2. The next sections detail the steps we perform for assessing the reliability of our Lyα flux measurements (§4.6.1), measuring the Lyα luminosity of our candidates and the cosmological volume probed by J-PLUS NBs (§4.6.2) and estimating the purity (§4.6.3) and completeness (§4.6.4) of our selection.

4.6.1 Retrieval of the total Lyα flux from J-PLUS photometry

As introduced in §2.2.5 and §4.3.1, our $\mathrm{F^{3FM}_{Ly\alpha}}$ measurements are likely affected by systematic uncertainties. These can be due to both the J-PLUS aperture photometry and our measuring method (see Eq. 2.42 and Eq. 2.38. Nevertheless, due to the high number of sources in the DR1 catalog and the substantial ignorance about the details of the intrinsic spectra of each Lyα-emitting candidate, we cannot estimate a flux-correction for each candidate. Therefore, we apply Eq. (2.42) to all our selected candidates and then statistically correct $\mathrm{F^{3FM}_{Ly\alpha}}$ to account for the biases on our line-flux estimate.

More in detail, we first study the differences between $\mathrm{F^{3FM}_{Ly\alpha}}$ and a corresponding spectroscopic measurement (i.e. $\mathrm{F^{spec}_{Ly\alpha}}$), assuming that the latter provides a reliable estimate of the total emitted Lyα flux (i.e. $\mathrm{F^{spec}_{Ly\alpha}} = \mathrm{F_{Ly\alpha}}$). We then compute statistical corrections to account for the bias between $\mathrm{F^{3FM}_{Ly\alpha}}$ and $\mathrm{F^{spec}_{Ly\alpha}}$. For this step, we use spectroscopically identified QSOs (Pâris et al., 2018) at the redshift sampled by each NB and their counterparts in the DR1 catalog. We measure $\mathrm{F^{spec}_{Ly\alpha}}$ from QSOs spectra with the methodology shown in §A **change this with the correct introduction-referencec**.

Photometric aperture correction

To make sure that auto-aperture photometry (see §4.2.1) do not introduce any bias on the photometry of our candidates, we compare the synthetic flux $\langle f_\lambda{}^r \rangle^{\mathrm{synth}}$ of SDSS QSOs to the analogous measurements in J-PLUS DR1. For the sake of brevity, here we summarize our findings and present the details of this check in the §4.A. In general, no significant bias ($\lesssim 0.2\,\sigma_r$) affects the auto-aperture flux of point-like sources for each NB. This is expected, since SDSS fiber-based photometry is normalized to PSF apertures[5] and J-PLUS auto-apertures generally encompass the whole PSF of point-like objects, hence making the two measurements effectively comparable. Consequently, we do not apply aperture corrections to $\mathrm{F^{3FM}_{Ly\alpha}}$. On

[5] see e.g., https://www.sdss.org/dr15/algorithms/spectrophotometry/

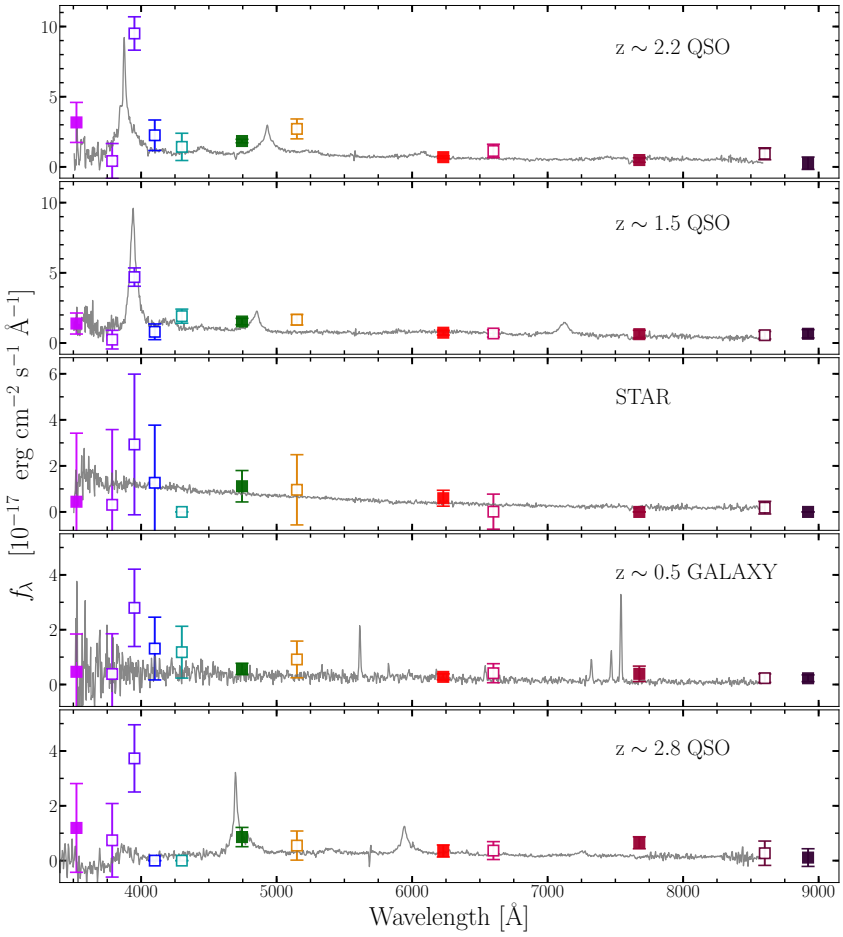

Figure 4.8: Summary of our spectroscopic results, showing one spectrum (grey line in each panel) for each source classes identified in our target lists. From top to bottom: z ∼ 2.2 QSO, z ∼ 1.5 QSO, star, z ∼ 0.5 galaxy and Lyβemitting QSO. The corresponding `J-PLUS` photometry is shown as coloured squares. The star and galaxy targets show low-significance excesses in $J0395$ NB (third square from the left). Indeed, these interlopers were selected as targets by the first version of our methodology, i.e. before applying the improvements due to `CTG2018A` results and the re-calibration of `J-PLUS` data (López-Sanjuan et al., 2019a).

the other hand, the flux comparison points out the need for an additional statistic uncertainty on top of the `J-PLUS` photometric errors (see 4.A for details). We then re-scale the r uncertainties and propagate them on $F_{Ly\alpha}^{3FM}$. Finally, we account for the latter on our LF determinations as discussed in §4.6.4.

Filter width correction

A fraction of the flux of broad lines (i.e. broader than the FWHM of the measuring NB) can be systematically lost by photometric measurements, especially if the line-peak is displaced at the edge of the NB transmission curve. SF LAEs usually show a narrow Lyα as opposed to the usually broad line profile of QSOs (e.g., Vanden Berk et al., 2001; Telfer et al., 2002; Selsing et al., 2016, and Fig. 4.2). For these reasons, we expect this bias to significantly affect the $F_{Ly\alpha}^{3FM}$ measurements of QSOs, while not influencing those of SF LAEs. At the same time, no SF LAEs were observed among our followed-up targets (§4.5.2), in line with previous results suggesting that AGN/QSOs dominate the samples of photometrically-selected LAEs at $L_{Ly\alpha} \gtrsim 2 \times 10^{43}\,\mathrm{erg\,s^{-1}}$ (see e.g., Santos et al., 2004; Konno et al., 2016; Matthee et al., 2017b; Sobral et al., 2018b; Calhau et al., 2020). Furthermore, the (expected) low fraction of SF LAEs in our final selection cannot be reliably disentangled from QSOs by `J-PLUS` photometry (§4.4). This hinders the possibility of applying a flux correction exclusively to a sub-class of our candidates. Consequently, we consider valid our method for measuring $F_{Ly\alpha}^{3FM}$ and then apply a statistical correction to all our

candidates. In particular, we obtain the corrected Lyα flux as follows (see §4.A):

$$F_{Ly\alpha}^{3FM\,;\,corr} = (1 - \Delta F) \cdot F_{Ly\alpha}^{3FM} \ . \tag{4.8}$$

The quantity ΔF is a rigid offset obtained from the normalized distribution of flux difference: $(F_{Ly\alpha}^{3FM} - F_{Ly\alpha}^{spec})/F_{Ly\alpha}^{3FM}$, where $F_{Ly\alpha}^{3FM}$ is our Lyα flux estimate and $F_{Ly\alpha}^{spec}$ is its spectroscopic analog measured on SDSS QSOs (see §4.A for details). We obtain a ΔF for each NB and then use the corrected values $F_{Ly\alpha}^{3FM\,;\,corr}$ for our luminosity function computation. With this a analysis we also obtain a correction for the error on $F_{Ly\alpha}^{3FM\,;\,corr}$, which we propagate on our Lyα LF determination (see §4.6.4).

Finally, the $F_{Ly\alpha}^{3FM}$ obtained with $J0378$ NB are affected by a significant bias ($\Delta F = 1.75 \pm 0.35$). This can be ascribed to wavelength separation between this NB and g, which reflects into a poor extrapolation of the linear continuum approximation. Consequently, we exclude $J0378$ from the list of NBs we use.

4.6.2 Computation of $L_{Ly\alpha}$ and cosmological volume

We compute the Lyα luminosity as:

$$L_{Ly\alpha}^{3FM} = 4\,\pi\,d_L^2(z)\ F_{Ly\alpha}^{3FM\,;\,corr} = 4\,\pi\,[d_c\,(1+z)]^2\,F_{Ly\alpha}^{3FM\,;\,corr}\ , \tag{4.9}$$

where $d_L(z) = d_c(z) \cdot (1+z)$ and $d_c(z)$ are the luminosity and comoving distances of our sources, computed by assuming `PLANCK2015` cosmology (Planck Collaboration et al., 2016a,b).

In order to compute the $d_L(z)$ of our candidates without spectroscopic determination it is necessary to assume a value of z. Being blind towards their nature, we use the z_p obtained by shifting the Lyα rest-frame wavelength to the pivot wavelength (Tokunaga & Vacca, 2005) of the NB used for selection (see Table 4.2). Consequently, the uncertainty σ_z is obtained from the half-width of each NB (see Table 4.2). This does not apply to the candidates with a spectroscopic counterpart, as in these cases we use the `DR14` z and σ_z. Finally, we propagate the redshift errors on the total $L_{Ly\alpha}$ uncertainty and on our LF determinations (see §4.6.4).

The cosmological volume sampled by our data depends on the z windows associated to Lyα detection and the DR1 area not affected by masking, for each NB. In our case, the redshift intervals are given by the FWHM of each NB (see Table 4.2) and converted to cosmological volumes by assuming the `PLANCK2015` cosmology (Planck Collaboration et al., 2016a,b). On the other hand, the effective area observed in a given band can be obtained with the `MANGLE` software (Hamilton & Tegmark, 2004; Swanson et al., 2008). Since we require single-detection in [NB; g and r], we computed the intersection between the three associated `MANGLE` masks (see Table 4.2). We assume negligible errors on volume estimates for our LF computation.

4.6.3 Estimate of the samples contamination

The steps detailed in §4.4 do not ensure to identify *all* the contaminants, as confirmed by our follow-up results (§4.5.2). For this reason, we estimate the residual contamination of our samples by computing a statistical *purity weight* for our candidates as a function of their r-magnitude:

$$P(r) = 1 - \frac{N_{interlopers}(r)}{N_{total}(r)} \ . \tag{4.10}$$

$N_{interlopers}(r)$ and $N_{total}(r)$ are respectively the number of *secure* interlopers (see §4.4) and the total number of candidates at a given magnitude. We then fit $P(r)$ with an error-function and use the latter to obtain the statistical weight of each genuine candidate to the final LF. Figure 4.9 shows the error-function fits for each NB (solid colored lines) and the computed $P(r)$ values for the $J0430$ filter, as an example (dotted grey line). The purple empty square shows the average purity measured on the complete sample of both our spectroscopic follow-up programs. This is in good agreement with the statistical weights of each NB (i.e. $P(r) \gtrsim 60\%$ at $r \gtrsim 18.5$). The high values reached by $J0515$ ($\sim 80\%$) are driven by the drop of interlopers with spectroscopic identification at $z \gtrsim 3$.

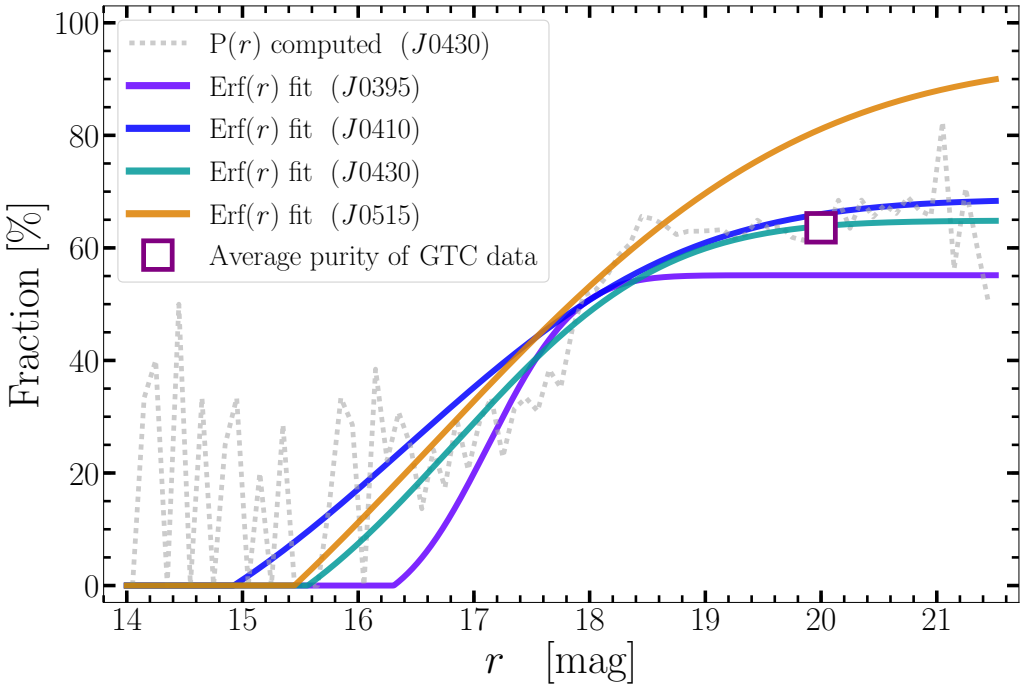

Figure 4.9: Statistical purity weight for each NB (coloured solid lines), as estimated by fitting an error-function to the computed purity. The grey-dotted line shows the computed purity of $J0430$ NB as an example. All filters show similar purity weights, rising to $\gtrsim 60\%$ at $r \gtrsim 18.5$. This is in agreement with the average purity of our 45 spectroscopic targets (§4.5.2), shown as a purple empty square.

4.6.4 Estimate of the samples completeness

Genuine line-emitting candidates might be lost by our selection due to the `J-PLUS` detection limits and source extraction, the effect of photometric errors and the r-band pre-selection of our parent samples (see e.g., Geller et al., 2012; Loveday et al., 2012; Gunawardhana et al., 2013). In order to correct for these known issues, we estimate the completeness[6] of our samples by considering three different components. In detail, we account for: i) the DR1 source-extraction process (i.e. *detection weight* C^{d}), ii) our selection methodology (i.e. *selection weight* C^{s}) and iii) r band pre-selection of dual-mode catalogs (i.e. *dual-mode weight* C^{dm}). We obtain the total completeness weight of each candidate to the final LF as: $C_i = C_i^{d} \cdot C_i^{s} \cdot C_i^{dm}$.

Detection completeness

The detection completeness of each `J-PLUS` pointing (for each filter) is automatically computed by the standard source-extraction pipeline as:

$$C_i^{d}(r) = 1 - \frac{1}{e^{-k_s \cdot (r - r_s)} + 1} , \tag{4.11}$$

where k_s and r_s are computed for each pointing. They are respectively the decay-rate of $C_i^{d}(r)$ and the magnitude at which $C_i^{d}(r)$ reaches 50%. All details of this computation are provided by the `J-PLUS` DR1 datababse. We obtain C_i^{d} from the $[k_s; r_s]$ parameters and r corresponding to each DR1 source.

Selection completeness

Starting from r-detected catalogs, our selection makes use of NB-excess significance and a linear estimate of the sources continuum slopes, related to their $g-r$ color (see §4.3.1, 4.3.2 and figures 4.2 and 4.3). In

[6] We define the completeness C as the ratio between the number of *genuine* targets effectively selected (*true positives*, TP) and the *total* number N_{tot} of genuine targets in the survey footprint, either detected or undetected. N_{tot} is generally unknown and can be thought as the sum of TP, *false negatives* (i.e. genuine targets detected but lost by the selection) and *undetected* candidates. In other words: $C = N_{TP}/(N_{TP} + N_{FN} + N_{UD})$.

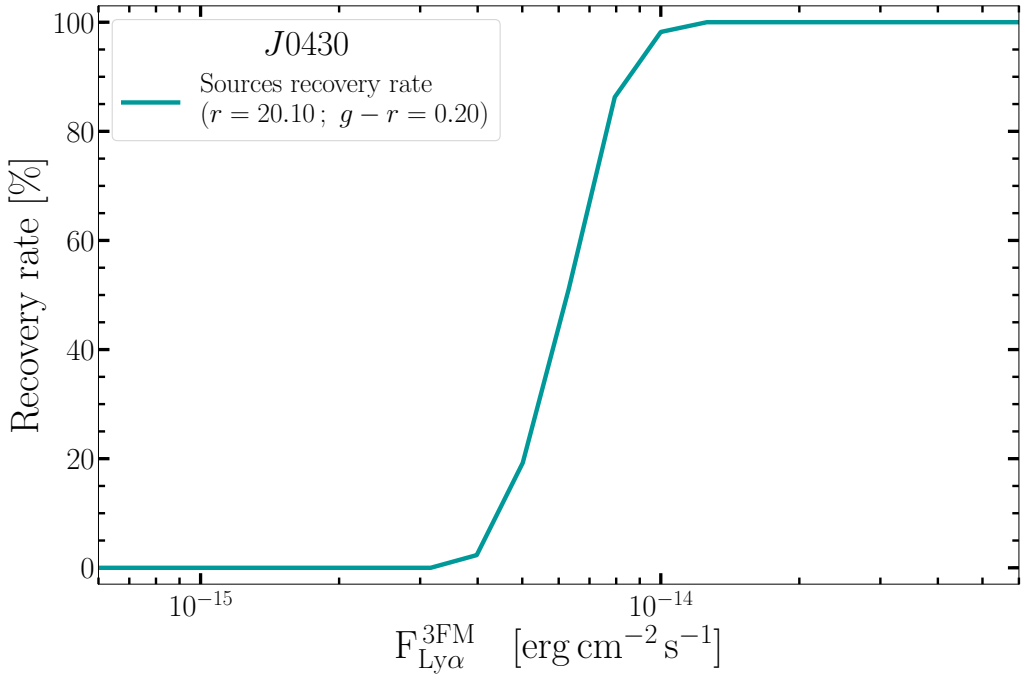

Figure 4.10: Example of the recovery fraction of our selection as a function of Lyα flux, computed for $J0430$ NB in a bin of r magnitude and $g-r$ color (namely at $r = 20.1$ and $g-r = 0.2$). The full 3D grid is shown in §4.C for the same NB. All filters show comparable values of recovery fractions, hence we just report the case of $J0430$ for brevity.

order to capture its multiple dependencies, we test the retrieval efficiency of our selection as a function of r magnitude, Lyα flux and $g-r$ color. In particular, we compute the recovery rate of simulated candidates over wide ranges of these three quantities, by re-applying each of our selection rules. This accounts for source loss at different Lyα flux, continuum and EW. We organize the measured recovery rates in a 3D-grid which we interpolate at the measured position of each genuine candidate to compute its selection weight C_i^s. The details of this computation are given in §4.C, while Fig. 4.10 shows an example of the estimated selection completeness in a bin of r magnitude and $g-r$ color for the $J0430$ NB.

r - $L_{Ly\alpha}$ bivariate completeness

The use of r-band detected catalogs makes our selection prone to the loss of continuum-faint $z \gtrsim 2$ Lyα-emitting sources, with non-trivial effects on the EW distribution of our selected samples. At low Lyα flux, for instance, the r-detection requirement might favour the selection of high-EW Lyα-emitting sources. This issue has been pointed out by previous works whose selection function was built on the convolution of r-band detection and NB-excess significance. In particular, Gunawardhana et al. (2015) showed that accounting for this effect requires a multi-variate approach. In other words, the fraction of undetected continuum-faint line-emitters can be estimated by modelling the full-2D luminosity function of candidates in the r vs. line-luminosity plane.

We closely follow the methods of Gunawardhana et al. (2015) applying their computations to the r vs $L_{Ly\alpha}$ space. The details of this procedure and its main equations are presented in §4.C. In brief, we assume that the 2D LF can be modelled by the product of two functions, describing respectively the r and $\mathrm{Log}(L_{Ly\alpha})$ distributions (see also Corbelli et al., 1991). We combine a Schechter (in logaritmic form) and a Gaussian (in $\mathrm{Log}\, L_{Ly\alpha}$) functions (as in Gunawardhana et al., 2015, see §4.C). By fitting this 2D model to our measured 2D LF, we can model the number density of sources in regions of the r vs $L_{Ly\alpha}$ plane affected by our incompleteness. Finally, the ratio of our data to the the 2D model (in the 2D space r vs. $L_{Ly\alpha}$) allows us to compute the statistical weight $C^{dm}(r, Ly\alpha)$ for each source, which accounts for the loss of r-faint Lyα-emitting sources. Figure 4.11 shows the results of our 2D modelling for the $J0430$ filter. In particular, the top and right panels show the projection of both our 2D LF (green solid lines) and 2D model (red dashed line) respectively along the $\mathrm{Log}(L_{Ly\alpha})$ and r axis. It is clear how the model extrapolates our measurements at $r > 19.5$ and $\mathrm{Log}(Ly\alpha/\mathrm{erg\,s^{-1}}) < 44.2$.

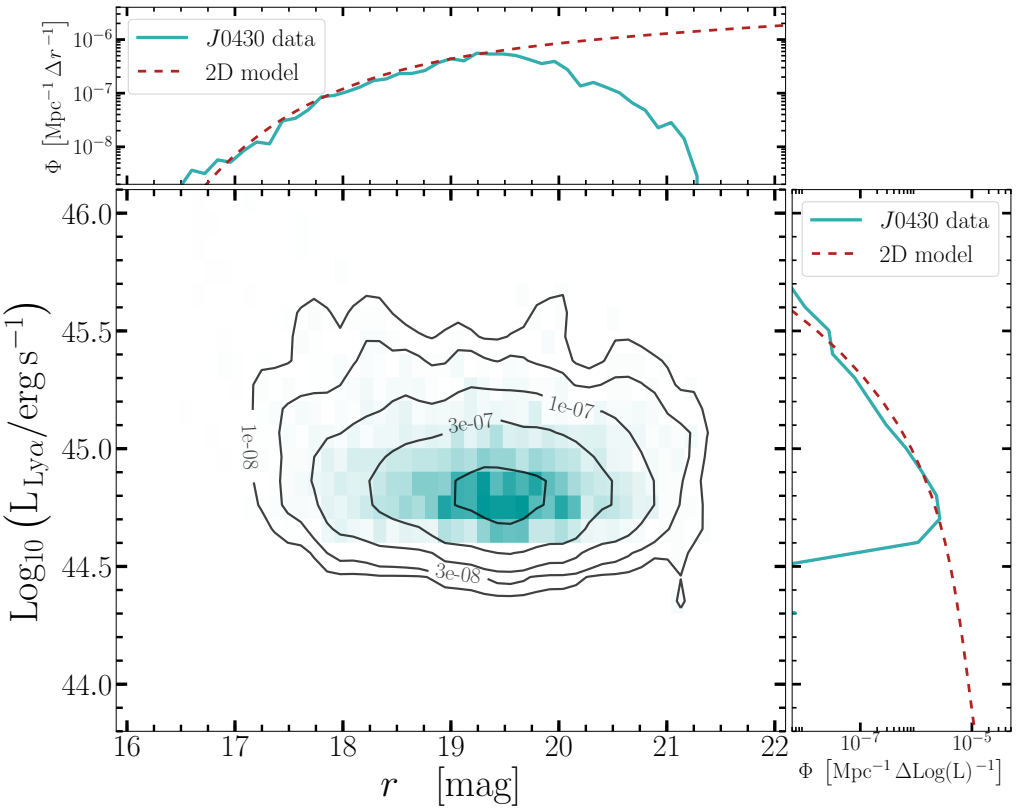

Figure 4.11: Central panel: full-2D luminosity function of our $J0430$ candidates, as a function of r and Log($\mathrm{L_{Ly\alpha}}$). Green solid lines in the top and right panels show the projections of the 2D LF respectively along the Log($\mathrm{L_{Ly\alpha}}$) and r axis. The red dashed lines show the projection of the 2D model along the same axis. This model was fitted on the 2D distribution shown in the central panel (see §4.C for computational details) and it allows to extrapolate our data distribution at faint r and Lyα luminosity. We use the data-to-model ratio (in the r vs.Log($\mathrm{L_{Ly\alpha}}$) 2D plane) to compute the $\mathrm{C_i^{dm}}$ weight of each candidate.

Errors on the Lyα luminosity function

The uncertainties on sources redshift and Lyα flux, the binning in Lyα luminosity and the internal variance of the samples (due to differences among each `J-PLUS` pointing) jointly contribute to the errors on our final LFs (e.g., Sobral et al., 2018a). We measure separately each source of uncertainty and finally sum in quadrature their different contributions. To account for $\mathrm{L_{Ly\alpha}}$ uncertainties, we repeat the determination of our LF 1000 times by perturbing each time the sources flux according to its uncertainty. During this procedure, we keep the sources redshift fixed to $\mathrm{z_p}$ (see §4.6.2) in order to evaluate only the contribution of flux perturbations to the final errors on our LFs. We then compute the asymmetric errors from the percentiles of the LFs distribution as $\sigma^- = 50^{\mathrm{th}} - 16^{\mathrm{th}}$ and $\sigma^+ = 84^{\mathrm{th}} - 50^{\mathrm{th}}$, where 84^{th}, 50^{th} and 16^{th} are the corresponding distribution percentiles. The contribution of redshift errors is accounted in the analogous way by fixing the flux measurements. To account for the internal variance of our LAE candidates sample due to field variations in `J-PLUS` DR1 we perform random realizations of the luminosity function by splitting our samples into ten independent sub-samples and computing a LF for each sub-sample. We repeat this process 1000 times and ultimately extract the errors from the 16^{th} and 84^{th} percentiles of the LFs distribution (see above). Finally we also add the poissonian errors ($\sqrt{N}$) associated to the sources number counts in each bin to the total LF uncertainties.

4.7 Results and discussion

In this section we present the four Lyα LFs we compute. We compare our measurements to previous results in the literature (in §4.7.2), we describe the computation of its Schechter parameters (in §4.7.3) and finally we estimate the fraction of AGN/QSOs as a function of luminosity (§4.7.4).

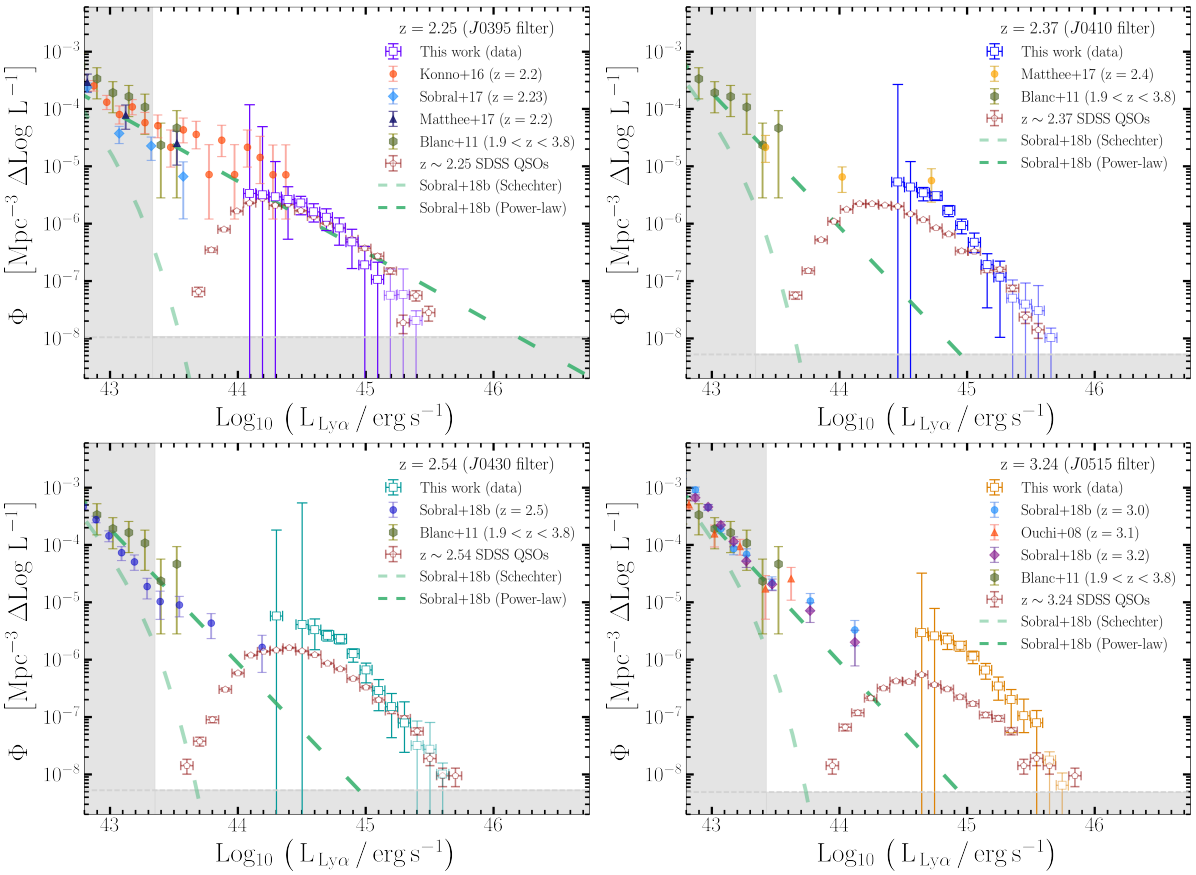

Figure 4.12: Lyα luminosity functions for each of the NB filter we used in our study (colored squares). The grey shaded areas show the Lyα luminosity limit (vertical limit) and the limiting number density measurable by `J-PLUS` (horizontal limit). The wide area explored by the narrow-bands of `J-PLUS` survey allow to remarkably extend the range of luminosity sampled by previous studies (coloured circles, triangles hexagons and diamonds in each plot) and to explore previously-unconstrained $L_{Ly\alpha}$ intervals. Dashed lines marks the best-fit determinations by Sobral et al. (2018b), split respectively into a Schecther (light green) and a power-law contribution (dark green). Our results provide tight constraints at $44.5 \lesssim \mathrm{Log}\,(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) \lesssim 45.5$, a regime currently unexplored by previous Lyα LFs determinations. Our errors are dominated by the completeness correction at low luminosity, while poor statistics due to low number counts (i.e. poissonian errors) dominate the bright tail of our distributions.

4.7.1 The Lyα luminosity functions at $2 < z < 3.3$

Figure 4.12 shows the four determinations of the Lyα LF we compute at $z \sim 2.25$, 2.37, 2.54 and $z \sim 3.24$ (colored empty squares). For each NB, we only consider the candidates with a total completeness weight $C = C^{d} \times C^{s} \times C^{dm} > 0.85$ (see §4.6.4 and §4.C). This excludes sources whose contribution is severely affected by the completeness correction, especially at $\mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) \lesssim 44$. Overall, our results probe a luminosity interval of $\sim$1.5 dex, from $\mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) \sim 44$ to $\mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) \sim 45.5$. These regimes are expected to be significantly populated by Lyα-emitting AGN/QSOs (e.g., Borisova et al., 2016b; Matthee et al., 2017b; Sobral et al., 2018a; Calhau et al., 2020). Interestingly, our results extend by $\sim$1 dex into a previously-unconstrained Lyα luminosity range, allowing to probe it with high precision. In addition, our data extend down to $\sim 10^{-8}\,\mathrm{Mpc^{-3}}$, a limit which is hardly reached by previous studies (see e.g., Sobral et al., 2018b). These remarkable features are ultimately attained because of the very wide area covered by `J-PLUS` NB imaging (unprecedented for Lyα LF determinations), which balances the `J-PLUS` depth ($r < 22$).

The shaded grey areas in each panel of Fig. 4.12 mark the regions which are not accessible by our data, respectively due to the limiting $L_{Ly\alpha}$ (vertical limit, see Table 4.2) and the survey area (horizontal limit). In particular, the latter marks the comoving number density (per $\Delta\,\mathrm{Log}\,L_{Ly\alpha}$) obtained if only a single object were detected in the whole survey footprint. Errors on $\Phi_{L_{Ly\alpha}}$ are computed as described in §4.6.4, and show a clear prevalence of the completeness correction at the lowest luminosity bins. On the other hand, the bright-end of our LFs are dominated by the internal variance of our samples, as the

number density of our candidates approaches the survey limit. To stress the impact of low-statistics on the bright tail, we marked with faded colors the data points at $\leq$1 dex above the density limit.

4.7.2 Comparison with previous determinations

We compare our Lyα LFs to a collection of previous determinations at similar z, after uniforming their underlain cosmology to the PLANCK2015 one. This comparison is complicated by the significant differences between the technical features of J-PLUS and previous high-z Lyα surveys (as noted in e.g., Blanc et al., 2011). Indeed, these can reach up to $\sim$5 magnitudes in depth and a factor of 10^3 on the surveyed area (see e.g., Ouchi et al., 2008; Konno et al., 2016). Nevertheless, the comparisons at z$\sim$2.25 and z$\sim$2.37 (respectively, *J*0395 and *J*0410 NBs) are remarkable, showing an overlap of our faint-end to the works of Konno et al. (2016) and Matthee et al. (2017b).

We report the best fits from Sobral et al. (2018b) at each redshift since these highlight both the Schechter and power-law components of the LFs (respectively, light-green and dark-green dashed lines in Fig. 4.12). These are obtained from a mixed Schechter/power-law model adapted to $\mathrm{Log}(\mathrm{L_{Ly\alpha}}/\,\mathrm{erg\,s^{-1}}) \lesssim 44.5$ data, showing a transition between the two regimes at $\mathrm{Log}(\mathrm{L_{Ly\alpha}}/\,\mathrm{erg\,s^{-1}}) \sim 43.5$. Despite the small overlap of luminosity regimes, our z $\sim$ 2.25 LF shows a remarkably good agreement with the power-law of Sobral et al. (2018b), as shown in the upper-left panel of Fig. 4.12. Interestingly, this component well accounts for the population of X-ray bright objects in their samples, suggesting that these sources might belong to a separate class described by a different luminosity distribution than SF LAEs at $\mathrm{Log}(\mathrm{L_{Ly\alpha}}/\,\mathrm{erg\,s^{-1}}) \lesssim 43.3$. On the contrary, a significant discrepancy between our data and the power-law components is evident at higher z. We ascribe this to the wider separation between the $\mathrm{L_{Ly\alpha}}$ ranges probed by our data and those on which the fits of Sobral et al. (2018b) are obtained at these z.

We note that our z$\sim$2.25 data nicely complement also the bright-end determination of Konno et al. (2016) (orange dots in the upper-left panel of Fig. 4.12). This work clearly showed an excess with respect to the exponential decay of a Schechter function at $\mathrm{Log}(\mathrm{L_{Ly\alpha}}/\,\mathrm{erg\,s^{-1}}) \gtrsim 43$. Their explanation relied on the contribution of a population of Lyα-emitting AGN/QSOs, as in e.g. Matthee et al. (2017b) and Sobral et al. (2018b). By joining these hints to the results of our spectroscopic follow-up and our sample analysis (detailed in §4.5 and §4.4), our work further supports the picture according to which Lyα -emitting AGN/QSOs are responsible for the bright-end excess observed on the $2 \lesssim z \lesssim 3$ Lyα luminosity function at $43.3 \lesssim \mathrm{Log}(\mathrm{L_{Ly\alpha}}/\,\mathrm{erg\,s^{-1}}) \lesssim 44.5$.

Comparison with SDSS DR14 QSOs

Figure 4.12 additionally shows the Lyα LF of all the DR14 QSOs in the J-PLUS footprint (from Pâris et al., 2018), with spectroscopic redshift in the intervals sampled by each NB (red pentagons). We obtain this determination by performing synthetic photometry of SDSS QSOs with J-PLUS filters and applying the same flux corrections as those computed for our data (see §4.6.1). For simplicity, we only associate poissonian errors to the SDSS LF.

Despite the comparison being only qualitative, the agreement between the SDSS QSOs distribution and our data is good, especially at low z. Interestingly, the fraction of our genuine candidates showing SDSS QSOs counterparts at the redshift probed by each NB is $\lesssim$ 30%, in each NB. Assuming that the Pâris et al. (2018) catalog represents a $\sim$100% complete sample of QSOS and considering the low fraction of SDSS QSOs in our data, the agreement between the two LFs could be explained in terms of a significant residual contamination of our samples ($\sim$ 70%). Nevertheless, this is in contrast with both our purity estimates and our spectroscopic follow-up (§4.6.3 and §4.5.2). A more interesting explanation is that our NB-based selection might actually be sensitive to high-z QSOs which lack spectroscopic determination in SDSS (due e.g. to their BB colors, see Ross et al., 2012; Richards et al., 2015), as those confirmed by our follow-up programs. Indeed, their previous classification based on SDSS photometry and morphology would identify most of them just as compact objects (namely stars, see Table 4.7). We suggest that this mis-classification might originate from the SDSS target selection, based on BB-colors, which might miss the presence of emission lines. On the contrary, our selection targets photometric excesses with respect to a continuum estimate, hence it can efficiently select high-z line emitters.

4.7.3 Lyα LF parameters

The bright-end excess: power-law or Schechter function?

As suggested by e.g. Konno et al. (2016); Matthee et al. (2017b); Sobral et al. (2018b,a) and Calhau et al. (2020), the population of bright Lyα-emitting sources at $\mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) > 43$ is likely to be composed by a mixture of SF LAEs and AGN/QSOs. In particular, Matthee et al. (2017b) and Sobral et al. (2018b) suggest that the two source classes might be described by substantially different distributions in terms of typical number density and Lyα luminosity. Interestingly, the power-law component of their studies can be explained as the faint-end of a Schechter function (Schechter, 1976, see also Eq. 4.17) describing the QSOs luminosity distribution. Our data can effectively support this hypothesis by providing the bright-end complement to the AGN/QSOs Schechter distribution. At the same time, our analysis limited by the J-PLUS depth which prevents us to constrain its the faind-end slope at $\mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) \lesssim 44$. This might significantly influence the determination of our Schechter paramters given their mutual correlation. Instead of fixing the faint-end slope to a fiducial value (as in e.g., Gunawardhana et al., 2015; Sobral et al., 2018b), we compute it by jointly exploiting our data and previous Lyα LF determinations, over the whole interval $41.5 \lesssim \mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) \lesssim 44$. More in detail, we make use of the Schechter component from Sobral et al. (2018b) at each redshift to describe the Lyα LF at $\mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) \lesssim 43.3$, and combine it to a second Schechter function to account for $\mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) \gtrsim 44$.. We then vary the faint-end slope of the latter and, for each α, we jointly fit the complete double-Schechter model to both our data and all the literature determinations (see Fig. 4.13). Finally, for each NB we obtain α and its errors from the reduced χ^2 distribution of the double-Schechter fits. The α values are shown in table 4.6.

We further assume no evolution of α with respect to redshift since neither our data nor previous works would allow to constrain it. Under this assumption, we obtain our final α as the weighted average of the above values: $\alpha = -1.35 \pm 0.84$. This high uncertainty is expected, given the limited amount of data populating the transition-regime between the two Schechter functions at $\mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) \sim 43.5$ (see Fig. 4.13). Nevertheless, our procedure consistently accounts for available data over ~ 3 dex in luminosity, providing one of the first estimates of α for the Schechter LF of Lyα-emitting sources at $\mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) \gtrsim 44$. Few works have currently estimated the LF shape at these very bright regimes by usually performing a power-law fit (e.g., Matthee et al., 2017b; Sobral et al., 2018b). Interestingly, these works respectively determined values of $(\alpha+1) = -0.75^{+0.17}_{-0.17}$ and $(\alpha+1) = -0.74^{+0.17}_{-0.17}$ at $z \sim 2.2$, which are both consistent with our faint-end slopes determinations at $z < 2.5$ within 1σ. This suggests that the power-law component observed at the bright end by previous works might be explained as the faint-end of a Schechter function describing the distribution of extremely luminous Lyα-emitting sources (i.e. AGN/QSOs). In other words, the full Lyα luminosity function at $41.5 \lesssim \mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) \lesssim 44$ could be effectively described by a double-Schechter model.

Constraints on Φ^* and L^*

We employ the fixed α computed with the above procedure to fit our data with a single-Schechter model and constrain Φ^* and L^* at $\mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) > 44$. We stress that for this step we explicitly use only our data points. The results of this procedure are compared to literature data in Fig. 4.14, while the left panel of Fig 4.15 directly compares our four redshift bins. We account for correlations between α and the remaining parameters by sampling the error of α (assumed to be Gaussian) with 50,000 monte-carlo realizations oof the single-Schechter fits, from which we extracting our final values and errors for Φ^* and L^*. Our results are listed in Table 4.6 and shown in the right panel of Fig. 4.15.

Under the hypothesis that our samples are greatly dominated by AGN/QSOs, our results show that their LF is described by a clearly distinct distribution with respect to SF LAEs (see also Matthee et al., 2017b). In particular, by comparing our Φ^* and L^* to previous determinations at $\mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) < 43$ (Gronwall et al., 2007; Ouchi et al., 2008; Konno et al., 2016), we measure a typical density and luminosity of AGN/QSOs respectively $\sim$3 dex lower and $\sim$2 dex higher, as already suggested by e.g. Matthee et al. (2017b) and Sobral et al. (2018b). In turn, this would suggest that the transition between the regime dominated respectively by SF LAEs and AGN/QSOs would fall at $\mathrm{Log}(L_{Ly\alpha}/\mathrm{erg\,s^{-1}}) \sim 43.5$, as also highlighted by Sobral et al. (2018a) and Calhau et al. (2020).

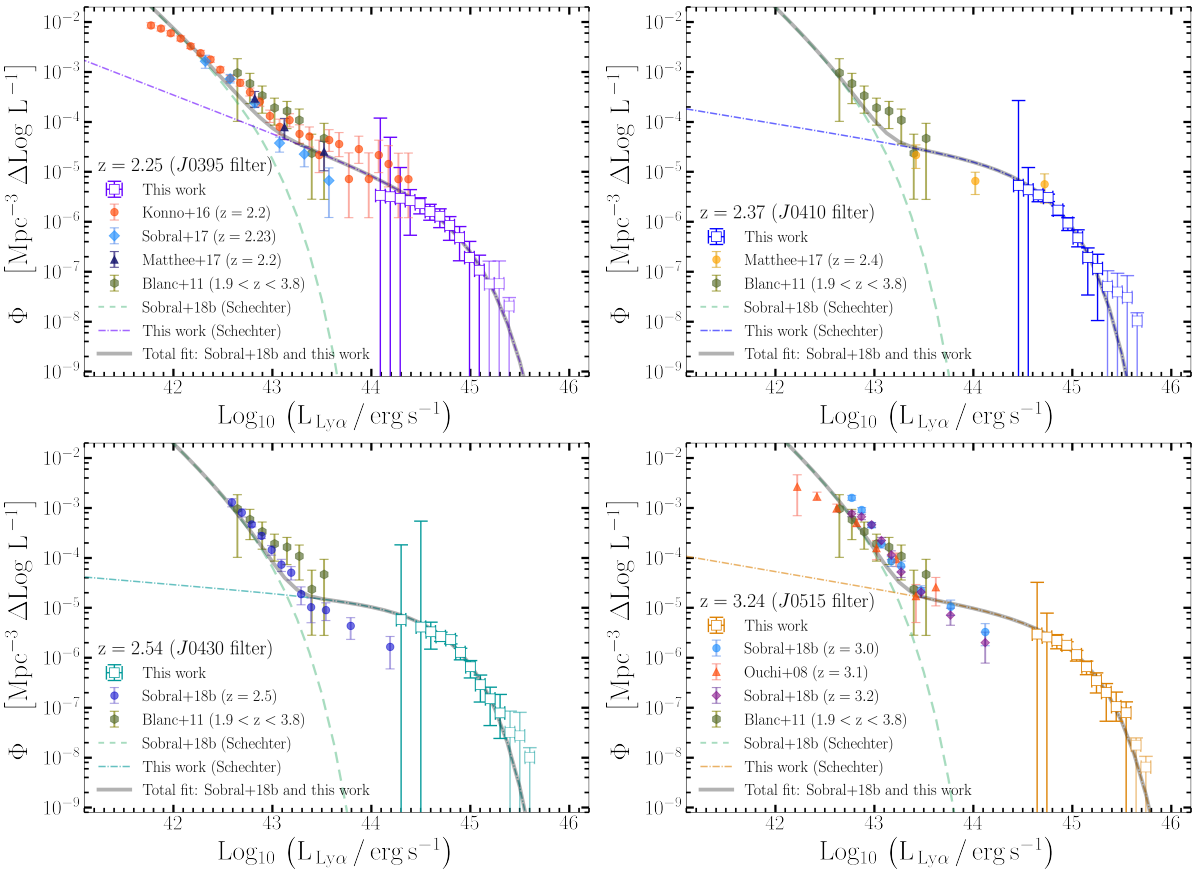

Figure 4.13: Joint fit of our Lyα luminosity functions and literature data with a double-Schechter model (grey solid lines in each panel). This is obtained by joining the best Schechter fit from Sobral et al. (2018b) at each redshift (green dashed lines in each panel) and a second Schechter function (coloured dashed-dotted lines). We jointly fit this double-Schechter model to both our data and the literature ones leaving free the parameters of the second Schechter, in order to constrain its faint-end slope α at each redshift.

Finally, our data do not allow to constrain the evolution of our Lyα LFs determinations. Indeed the Φ^* and L^* we obtain are statistically consistent (at $\sim 2\sigma$) among the four filters, with average values $\Phi^* = (3.33 \pm 0.19) \times 10^{-6}$ Mpc^{-3} and $L^* = 44.65 \pm 0.65$ erg s^{-1}. This is shown in the right panel of Fig. 4.15, where the faint and dark contours for each filter respectively mark the 2-σ and 1-σ levels (i.e. the 86% and 39% iso-contours) of the parameters distributions obtained from monte-carlo realizations. The wide overlap between the four filters shows the low constraining power of our data towards the evolution of the LF parameters with redshift. This was anticipated by the significant variation among the distributions of $L_{Ly\alpha}$ and EW at each z shown in Fig. 4.5, which ultimately hinders the possibility to disentangle the intrinsic variations of our sample properties from systematic effects.

4.7.4 The AGN fraction of $z \gtrsim 2$ LAEs

By assuming that our Lyα LF describes the distribution of only AGN/QSOs, we can build a simple toy model to estimate the relative fraction AGN/QSOs and SF LAEs as a function of Lyα luminosity. We define the latter as:

$$q_{AGN} = \frac{LF^{SF\ LAEs}}{LF^{SF\ LAEs} + LF^{AGN/QSOs}}\ , \tag{4.12}$$

where $LF^{AGN/QSOs}$ is one of the four determinations of the Schechter function computed from our data, while $LF^{SF\ LAEs}$ is the best fit of Sobral et al. (2018b) at the corresponding redshift. We use the latter since it is obtained by excluding LAE candidates with X-ray counterparts from the determination of the Schechter fit. Consequently, we assume it provides a fair estimate for the luminosity distribution of only SF LAEs. We underline that our estimate of q_{AGN} is an illustrative application of our results rather

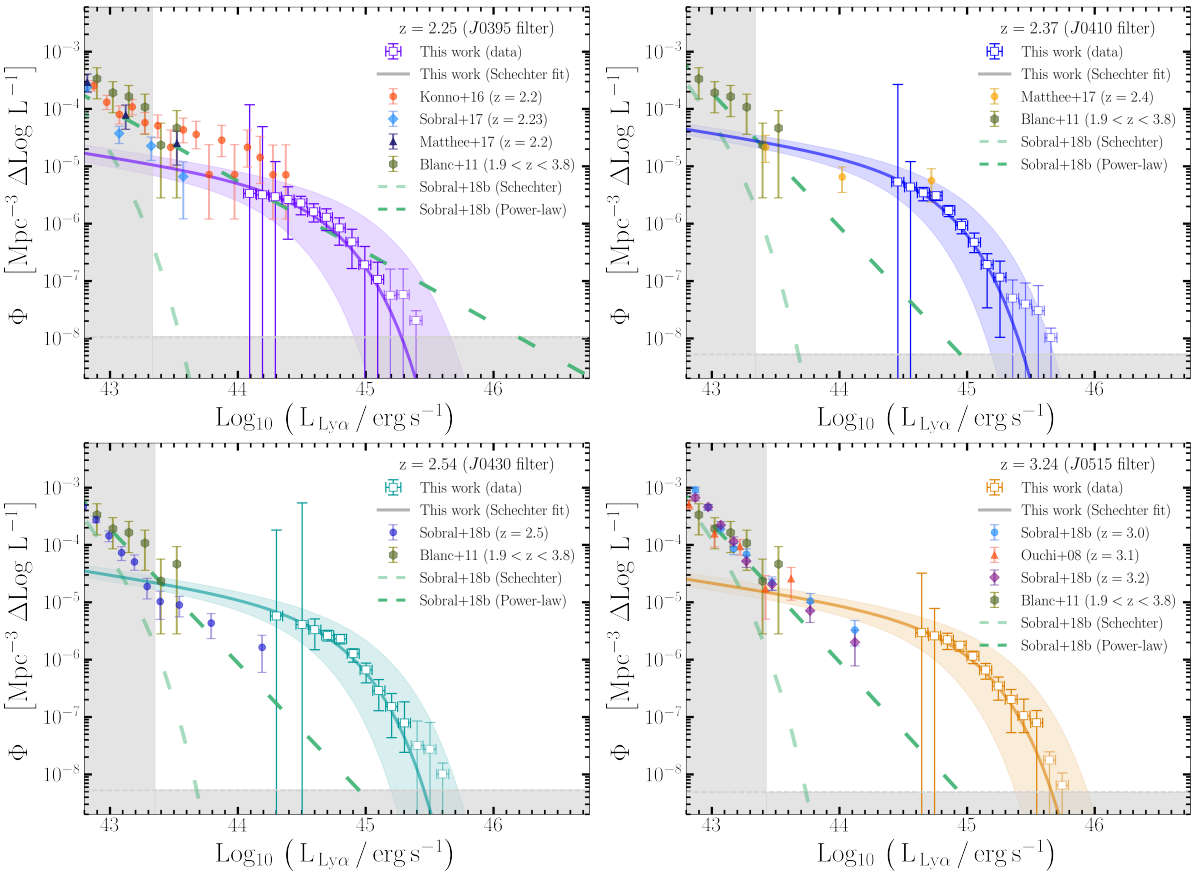

Figure 4.14: Final Schechter fits of our Lyα LFs (colored solid lines in each panel) performed by keeping the faint-end slope fixed to $\alpha = -1.35 \pm 0.84$. The colored shaded regions in each panel mark the 1σ confidence regions for the Φ^* and L^* parameters obtained by sampling their associated errors, obtained via monte-carlo simulations (see §4.7.3). The literature data shown in each panel are the same as in Fig. 4.12.

than a rigorous measurement, given the strong assumptions on which it is based.

The AGN/QSOs fractions for all the redshifts we probe are shown in figure 4.16. Despite our simplifying assumptions, we find a good agreement (within $1\,\sigma$) with the measurements of Sobral et al. (2018a), which are obtained from a spectroscopic follow-up of Lyα-selected targets. On the contrary, the works of Matthee et al. (2017b) and Calhau et al. (2020) (also shown in Fig. 4.16 for comparison) are based on photometric selections which identify AGN/QSOs candidates on the basis of their X-ray and/or radio-loudness. The latter are likely to be significant only for a sub-sample of AGN/QSOs (as suggested by e.g., Sobral et al., 2018b, and Calhau et al. 2020), hence the discrepancy with our estimates might also be explained in terms of this incompleteness effect.

To conclude, the good agreement between our AGN/QSOs fraction estimates and the data of Sobral et al. (2018a) supports the scenario by which our samples are strongly dominated by Lyα-emitting AGN/QSOs. Furthermore, the discrepancy with respect to X-ray/Radio selected AGN candidates suggests that the latter are likely a sub-sample of the whole high-z AGN/QSOs population. Our selection, on the contrary, is only based on Lyα emission, hence it is likely to detect previously-unidentified high-z AGN/QSOs. This is also in line with the results of our spectroscopic follow-up program (§4.5.2).

4.8 Conclusions

This chapter presents the determination of the bright-end of the Lyα luminosity function at four redshifts in the interval $2 \lesssim z \lesssim 3.3$, namely $z = 2.25$, 2.37, 2.54 and $z = 3.24$ (see also Table 4.2). We obtain our LFs by employing four lists of Lyα-emitting candidates selected in the DR1 catalog of the `J-PLUS` survey, according to the significance of their photometric excess in the *J*0395, *J*0410, *J*0430 and *J*0515 narrow-bands. The effective `J-PLUS` survey area sampled by each of these NBs sums up to $\sim 900\,\mathrm{deg}^2$

Filters	z	Schechter parameters		
		α (varying)	Φ^* [10^{-6} Mpc^{-3}] ($\alpha = -1.35 \pm 0.84$)	Log(L*/erg s^{-1}) ($\alpha = -1.35 \pm 0.84$)
*J*0395	$2.25^{+0.03}_{-0.05}$	$-1.77^{+0.09}_{-0.07}$	$1.86^{+4.14}_{-1.60}$	$44.54^{+0.43}_{-0.35}$
*J*0410	$2.37^{+0.09}_{-0.08}$	$-1.33^{+0.50}_{-0.22}$	$4.66^{+6.03}_{-3.25}$	$44.60^{+0.29}_{-0.21}$
*J*0430	$2.53^{+0.09}_{-0.07}$	$-1.17^{+0.19}_{-0.13}$	$3.61^{+4.40}_{-2.57}$	$44.63^{+0.30}_{-0.22}$
*J*0515	$3.24^{+0.07}_{-0.10}$	$-1.34^{+0.12}_{-0.09}$	$2.12^{+3.56}_{-1.55}$	$44.87^{+0.32}_{-0.26}$

Table 4.6: Schechter parameters computed on our data. The average of the α values obtained for each filter (third column, see §4.7.3) and its errors were used as a fixed constraint to obtain Φ^* and L* via monte-carlo sampling. Errors on the latter parameters are obtained from their corresponding 1-D distributions.

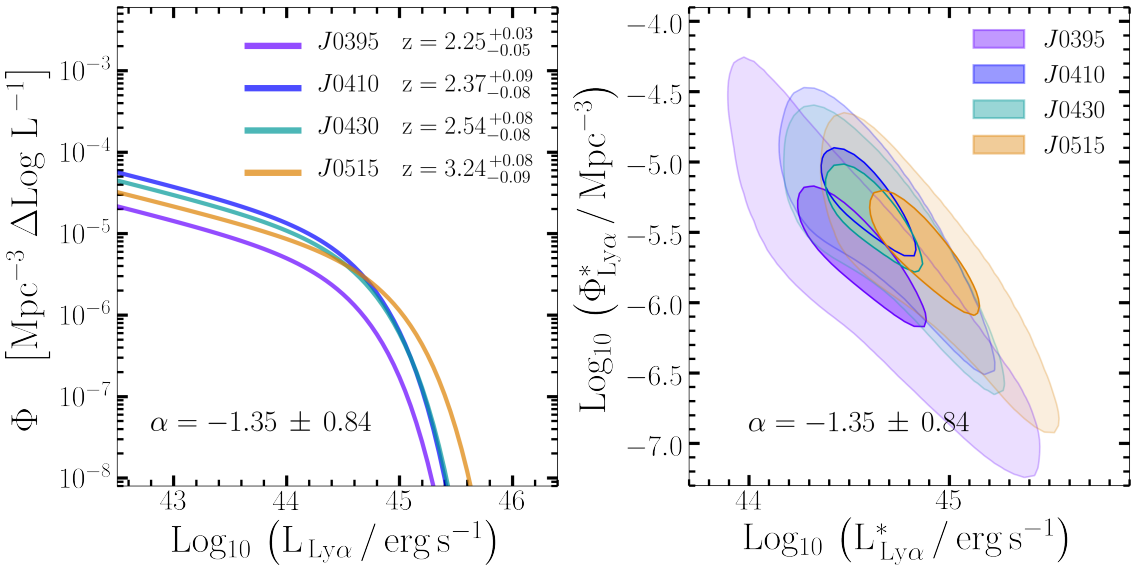

Figure 4.15: Left panel: single-Schechter fits to our data computed with the fixed faint-end slope $\alpha = -1.35 \pm 0.84$ obtained as in §4.7.3. We note that the difference among the four determinations (factor of $\sim$2 both in luminosity and normalization) are absorbed by the errors on the Schechter parameters (right panel). Right panel: distribution of Φ^* and L* obtained from the monte-carlo sampling of α errors. The contours mark the levels including 86% and 39% of the monte-carlo realizations (respectively faint and dark contours). This analysis shows that the parameters of the four determinations are statistically consistent, hence we do not observe hints for an evolution of the $2 \lesssim z \lesssim 3.3$ Lyα LF at Log(L$_{\rm Ly\alpha}$/erg s^{-1}) $\gtrsim$ 43.5.

after masking artifacts and bright stars, which correspond to $\gtrsim 1\,{\rm Gpc}^3$ (comoving) at the z windows corresponding to the NBs we use. These volumes are significantly wider than those probed by previous works at similar redshift, and allow to measure with high precision the Lyα luminosity function at $2.2 \lesssim z \lesssim 3.3$ and $L_{Ly\alpha} \gtrsim 2 \times 10^{43}\,{\rm erg\,s^{-1}}$.

We select 2547, 5556, 4994 and 1467 bright candidates ($L_{Ly\alpha} > 2 \times 10^{43}\,{\rm erg\,s^{-1}}$), which jointly represent the largest sample of photometric Lyα-emitting candidates at $2 \lesssim z \lesssim 3.3$ to date. We expect these lists to include both bright star-forming LAEs (SF LAEs) and Lyα-emitting AGN/QSOs. To identify either of these source classes in our samples, we follow-up spectroscopically a random sub-sample of our candidates (§4.5). The spectroscopic data confirmed 40 out of 45 targets as genuine high-z line-emitters (with 29 out of 45 being $z > 2$ Lyα-emitting QSOs) and found no star-forming LAE. In addition, we look for bi-modalities in the photometric properties of our candidates, such as Lyα luminosity and EW (§4.4) or colors (§4.4). Overall, the properties of our candidates are consistent with those of spectroscopically-confirmed QSOs (Fig. 4.6) and high-z QSO templates (Fig. 4.7), suggesting that the fraction of SF LAEs in our samples is negligible.

We use our candidates samples to compute the Lyα LF at extremely-bright luminosity regimes for the first time, namely at $44 \lesssim {\rm Log}(L_{Ly\alpha}/{\rm erg\,s^{-1}}) \lesssim 45.5$, and extend by $\gtrsim$1.5 dex the intervals covered by previous determinations. The extensive area observed by J-PLUS DR1 allows to access wide cosmological

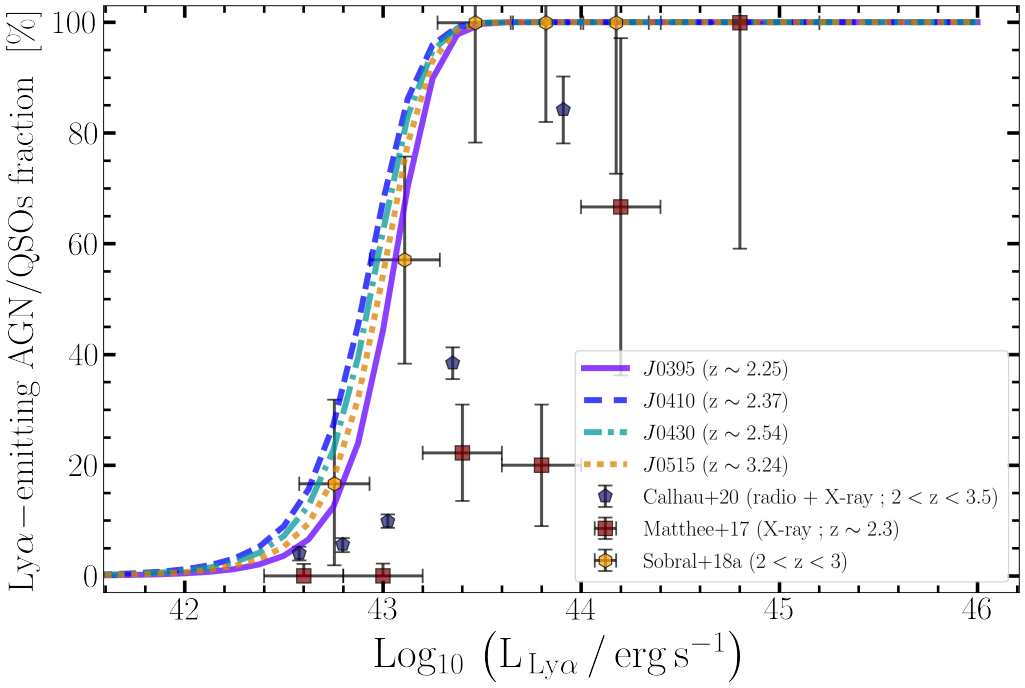

Figure 4.16: The AGN/QSO fraction as a function of luminosity for each NB. We estimated this quantity by assuming that our results are entirely dominated by AGN/QSOs and that the best Schechter fit of Sobral et al. (2018b) describes the distribution of SF LAEs (see Eq. 4.12). Our results are in agreement with the spectroscopic determination of Sobral et al. (2018a), which only employs Lyα emission pre-selection for their targets. On the other hand, the estimates of Matthee et al. (2017b) and Calhau et al. (2020) are based on the detection of either X-ray or radio counterparts for their Lyα-emitting candidates.

volumes ($\gtrsim 1\,\mathrm{Gpc}^3$), hence to probe number densities as low as $\sim 10^{-8}\,\mathrm{Mpc}^{-3}$. This parameters-space region is unprecedented for surveys focused on bright photometrically-selected Lyα-emitting sources. Interestingly, our Lyα LFs are in line with previous results at $\mathrm{Log}(\mathrm{L_{Ly\alpha}}/\mathrm{erg\,s^{-1}}) \gtrsim 43.5$, prolonging their power-law end into a full-developed Schechter function. We derive the redshift-averaged parameters $\Phi^* = (3.33 \pm 0.19) \times 10^{-6}\,\mathrm{Mpc}^{-3}$, $\mathrm{L}^* = 44.65 \pm 0.65\,\mathrm{erg\,s^{-1}}$ and $\alpha = -1.35 \pm 0.84$ for our Schechter best-fits. This shows that the whole Lyα LF, i.e. from $\mathrm{Log}(\mathrm{L_{Ly\alpha}}/\mathrm{erg\,s^{-1}}) < 42$ up to $\mathrm{Log}(\mathrm{L_{Ly\alpha}}/\mathrm{erg\,s^{-1}}) > 45$, can be effectively described by a composite model of two Schechter functions, respectively accounting for the distribution of SF LAEs and bright AGN/QSOs. These distributions appear structurally different, with $\mathrm{L^*_{QSOs}} \sim 100\,\mathrm{L^*_{SF\ LAEs}}$, $\Phi^*_{\mathrm{QSOs}} \sim 10^{-3}\,\Phi^*_{\mathrm{SF\ LAEs}}$ and a transition-regime centered at $\mathrm{Log}(\mathrm{L_{Ly\alpha}}/\mathrm{erg\,s^{-1}}) \sim 43.5$ (in line with e.g., Konno et al., 2016; Matthee et al., 2017b; Sobral et al., 2018a; Calhau et al., 2020). Overall, our results support the scenario suggested by e.g. Konno et al. (2016); Matthee et al. (2017b) and Sobral et al. (2018b), according to which the excess of bright LAEs measured at $\mathrm{Log\,L_{Ly\alpha}} \gtrsim 43$ with respect to a Schechter distribution is due to a population of AGN/QSOs (see also Calhau et al., 2020). Our findings characterize for the first time this population as being ~ 100 times more luminous and ~ 1000 times less dense than that of SF LAEs at comparable redshifts.

In addition, $\sim 70\%$ of our Lyα-emitting candidates lacks any spectroscopic confirmation by current surveys. Based on our spectroscopic follow-up results, we suggest that our samples are dominated by high-z QSOs which are not yet identified as such, but rather mis-classified as stars by current archival data, due to their photometric colors. Indeed, even accounting for a conservative residual contamination of $\sim 35\%$ in our final samples, the number of genuine $z \gtrsim 2$ QSOs identified for the first time by our methodology would be approximately 1300, 3200, 2900 and 900, respectively for *J*0395, *J*0410, *J*0430 and *J*0515 J-PLUS NBs. We ascribe this possibility to the narrow-band excess detection of our methodology, which can be particularly effective in targeting and selecting the strong line-emission features of $z > 2$ AGN/QSOs. Indeed, these might be missed by spectroscopic target selection based only on broad-band colors (e.g., Richards et al., 2009; Ivezić et al., 2014; Richards et al., 2015). We stress that the confirmation of this speculative hypothesis must rely on a systematic and extensive confirmation of our candidates. The latter might be obtained via either spectroscopic analysis or by exploiting the very efficient source

identification yielded by multi-NB imaging. Indeed, the upcoming J-PAS survey can provide a natural setting to extend our work.

Finally, our data do not show significant evolution of the LF over the probed redshifts. Despite X-ray studies suggest little evolution of the $2 < z < 3.3$ AGN/QSOs population (e.g., Hasinger et al., 2007), our findings might also be affected by J-PLUS detection limits. This factor could be mitigated by deeper photometric imaging. Unfortunately, this is hardly attainable by future J-PLUS data releases, since the technical features of the T80 (80cm) telescope hinder the possibility of reaching higher image-depths over very wide sky areas. On the contrary, future multi-NB wide-area photometric surveys can provide a valid tools to test the LF evolution at Log $L_{Ly\alpha} \gtrsim 43.5$.

4.A Appendix: Retrieval of the total line flux

Here we describe how we characterize and correct the bias on our measurements of Lyα flux (namely $F_{Ly\alpha}^{3FM}$). We compute a correction for each NB by exploiting the comparison between J-PLUS sources and their counterparts in the SDSS QSOs catalog (Pâris et al., 2018), within the redshift windows sampled by the NB. For the sake of brevity, we only show the results for either $J0410$ or $J0430$ filters. As a first step, we introduce the Lyα flux quantities obtained from SDSS spectra in order to perform our comparison (see Fig. 4.17):

- $\langle f_\lambda^{x} \rangle^{synth}$: synthetic flux-density, in erg cm^{-2} s^{-1} Å^{-1}, measured by convolving SDSS spectra with the transmission curve of a given J-PLUS filter "x", as in eq. 2.26. Coloured crosses in the bottom panel of Fig. 4.17 mark the synthetic photometry of the SDSS QSO taken as example.
- $F_{Ly\alpha}^{spec}$: spectroscopic measurement of the wavelength-integrated Lyα flux, in erg cm^{-2} s^{-1}. This is obtained by fitting the QSOs continuum and integrating the spectra above it, on the wavelength range affected by the *whole* Lyα line profile (see §4.B for details). This measurement does not involve the convolution of SDSS spectra with the filters transmission curves.
- $F_{Ly\alpha}^{spec\,;\,NB}$: a version of $F_{Ly\alpha}^{spec}$ obtained by integrating the spectra above the continuum-fit exclusively over the wavelength range covered by a J-PLUS NBs. If the Lyα line of a given QSO is wider than the NB, this measurement is just a fraction of $F_{Ly\alpha}^{spec}$, since the Lyα flux lying outside the transmission curve would not be accounted for (see Fig. 4.17).
- $F_{Ly\alpha}^{3FM\,;\,synth}$: this photometric quantity is analogous to $F_{Ly\alpha}^{3FM}$, with the only difference of being computed on the synthetic photometry of SDSS QSOs (coloured crosses in the bottom panel of Fig. 4.17).

The photometric quantity $F_{Ly\alpha}^{3FM}$ is directly comparable to the spectroscopic $F_{Ly\alpha}^{spec;\,NB}$, since the method of Vilella-Rojo et al. (2015) is designed to remove the effect of the filter transmission curve (see also §2.2.5). Moreover, any photometric measurement is only sensitive to the flux received within a given band, hence to $F_{Ly\alpha}^{spec;NB}$ and not to $F_{Ly\alpha}^{spec}$, in our case.

r-band `auto`-aperture flux

Figure 4.18 shows the comparison between r-band flux and $\langle f_\lambda{}^{r} \rangle^{synth}$ for J-PLUS sources and their QSOs counterparts with Lyα in the $J0410$ filter. The left panel displays no systematic shift, hence no strong aperture bias ($\lesssim 0.2\,\sigma_r$) affects the `auto`-aperture flux of point-like sources. Since $z \gtrsim 2$ sources appear point-like in J-PLUS (see §4.2.2), we conclude that `auto`-aperture photometry collects the total light of our Lyα -emitting candidates. On the other hand, the spread of the distribution in the right panel is significantly greater than one, hence we need to account for this additional statistic uncertainty on top of J-PLUS photometric errors when computing $F_{Ly\alpha}^{3FM}$. We re-scale the photometric errors of r band photometry and propagate the resulting σ_r on $F_{Ly\alpha}^{3FM}$. The latter is accounted for in the errors of our LFs as discussed in §4.6.4.

Filter-width effect on the line flux

Since the observed Lyα line profile of QSOs is generally wider than the FWHM of J-PLUS NBs, we need to account for line-flux losses. For this, we obtain two corrections, respectively for $F_{Ly\alpha}^{3FM}$ and its errors[7] $\sigma_{F_{Ly\alpha}^{3FM}}$. These affect the Ly$\alpha$ luminosity of our candidates and its errors, hence we account for them on our final LFs (as discussed in §4.6.4). Figure 4.19 shows the comparison between $F_{Ly\alpha}^{3FM}$ and $F_{Ly\alpha}^{spec}$ for $J0430$ filter. Their flux-difference is presented in the left and middle panel, respectively normalized by $F_{Ly\alpha}^{3FM}$ and by $\sigma_{F_{Ly\alpha}^{3FM}}$. The left panel shows clear evidences of a systematic offset ΔF between the two flux quantities. We measure ΔF with a gaussian fit (see Fig. 4.19) and use it to correct $F_{Ly\alpha}^{3FM}$ as follows:

$$F_{Ly\alpha}^{3FM\,;\,corr} = (1 - \Delta F) \cdot F_{Ly\alpha}^{3FM} \,. \tag{4.13}$$

[7] We stress that $\sigma_{F_{Ly\alpha}^{3FM}}$ is the error on $F_{Ly\alpha}^{3FM}$ computed by propagating the photometric errors in Eq. 2.42, as detailed in §2.2.5.

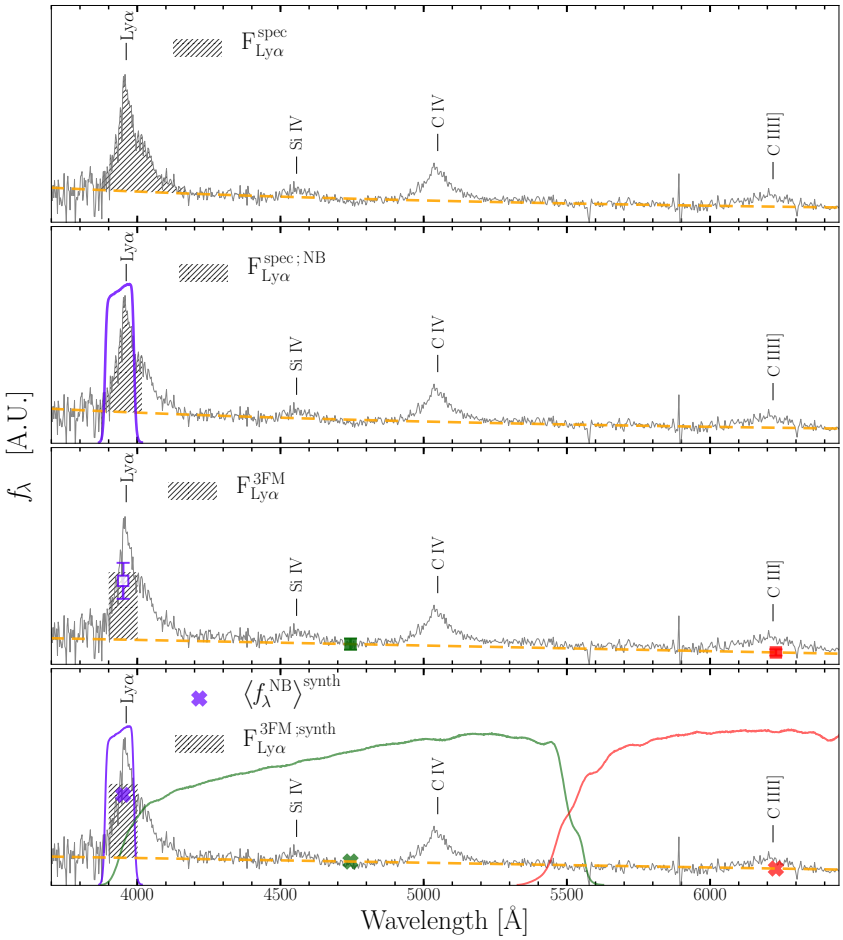

Figure 4.17: Graphic definition of the quantities we use for comparing Lyα flux. The spectrum of a z$\sim$2.2 QSO from SDSS DR14 is used as example in all panels (grey lines). Each Lyα flux definition is outlined by a grey shaded area (see text for details). Yellow lines in each panel show the reconstruction of the source continuum (power-law and linear approximation respectively in the first two and last two panels from above). Finally, colored squares and crosses (respectively third and last panel from the top) show respectively J-PLUS measurements and synthetic photometry performed on the SDSS spectrum with J-PLUS transmission curves (Eq. 2.26).

We stress that by directly comparing $F^{3FM}_{Ly\alpha}$ to $F^{spec}_{Ly\alpha}$, our statistical correction accounts for any systematic bias of our measurements, such as the linear-continuum approximation or the line-peak position of Lyα within the NB. The spread of the distribution in the middle panel of Fig. 4.19 is significantly bigger than unity. This shows that $\sigma_{F^{3FM}_{Ly\alpha}}$ cannot fully account for the difference between $F^{3FM}_{Ly\alpha}$ and $F^{spec}_{Ly\alpha}$. Consequently, we re-scale $\sigma_{F^{3FM}_{Ly\alpha}}$ according to the measured spread of the distribution in the middle panel of Fig. 4.19. Finally, we account for these errors on our final LFs (see §4.6.4). To conclude, the comparison between middle and right panels of Fig. 4.19 clearly shows how $F^{3FM}_{Ly\alpha}$ better compares to the fraction of spectroscopic line-flux measured *only* on the wavelength range covered by the J-PLUS NB (i.e. $F^{spec\,;NB}_{Ly\alpha}$, see Fig. 4.17 and §4.A). This is a direct effect of the filter-width bias, because our methodology is only sensitive to the flux captured by J-PLUS photometry within the NB wavelength range, as discussed in §4.A.

4.B Appendix: Measurement of GTC spectra and follow-up results

We measure the redshift of all the 37 sources identified as QSOs (either at z$\sim$1.5 or at z$\sim$2.2) in both spectroscopic programs. We do not aim at reaching a higher precision than $\sigma_z = 10^{-2}$, since the main goal of our follow-up programs is the spectroscopic confirmation of our targets. We additionally extract the Lyα EW and integrated line flux $F_{Ly\alpha}$ for the 29 z$\sim$2.2 QSOs, from which we compute the sources

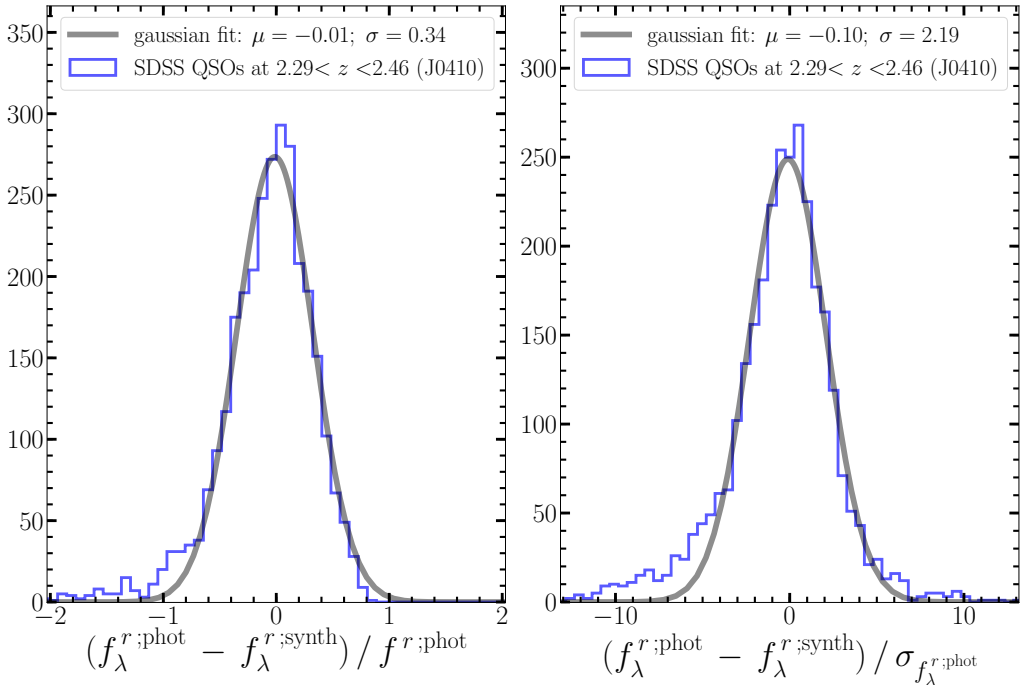

Figure 4.18: Blue solid lines in both panels show the histograms of the difference between J-PLUS r flux (i.e. f_λ^r) and $f_\lambda^{r\,;\mathrm{synth}}$ obtained from SDSS QSOs spectra (see Fig. 4.17). The distributions are normalized by respectively f_λ^r (left panel) and its photometric error $\sigma_{f_\lambda^r}$ (right panel). Both distributions are centered in zero (see plot legends), meaning that f_λ^r and $f_\lambda^{r\,;\mathrm{synth}}$ values are statistically equivalent. On the other hand, the distribution spread in the right panel is significantly bigger than one, hence photometric errors do not fully account for the flux difference. Consequently, we re-scale $\sigma_{f_\lambda^r}$ to the value obtained by the Gaussian fit (right-panel legend).

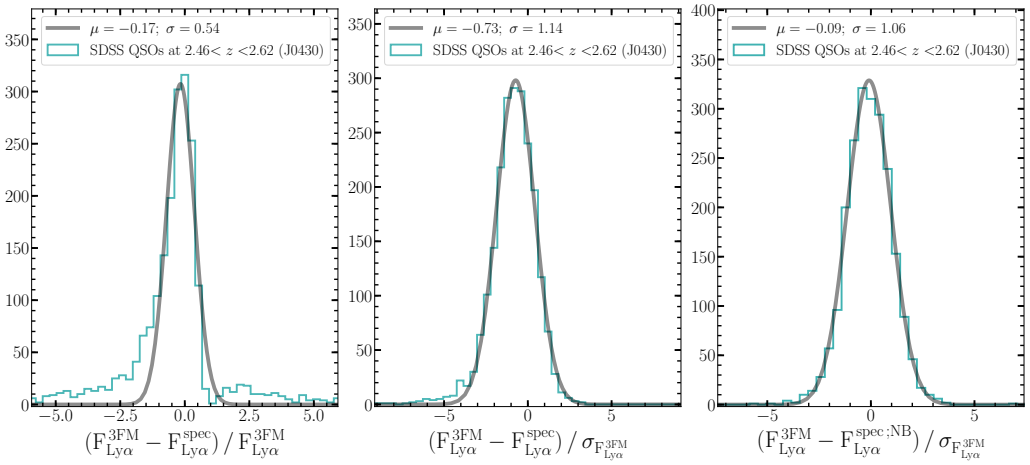

Figure 4.19: Cyan solid lines in all panels show the histograms of differences between line flux measurements performed on J-PLUS photometry (see §2.2.5 and §4.3.1) and on SDSS spectroscopy (Fig. 4.17), for the case of $J0430$. The difference between $\mathrm{F}_{\mathrm{Ly}\alpha}^{\mathrm{3FM}}$ and $\mathrm{F}_{\mathrm{Ly}\alpha}^{\mathrm{spec}}$ is normalized respectively by $\mathrm{F}_{\mathrm{Ly}\alpha}^{\mathrm{3FM}}$ and by $\sigma_{\mathrm{F}_{\mathrm{Ly}\alpha}^{\mathrm{3FM}}}$ in the left and middle panels. We use the systematic shift of the distributions in the left panel to statistically correct $\mathrm{F}_{\mathrm{Ly}\alpha}^{\mathrm{3FM}}$, while the distribution spread in the middle panel allows to account for residual statistical errors not included in $\sigma_{\mathrm{F}_{\mathrm{Ly}\alpha}^{\mathrm{3FM}}}$. Section 4.6.4 details how the systematic offset (left panel) and distribution spread (middle panel) concur to the errors on our final LFs. Finally, right panel shows that $\mathrm{F}_{\mathrm{Ly}\alpha}^{\mathrm{3FM}}$ well compares to the spectroscopic measure $\mathrm{F}_{\mathrm{Ly}\alpha}^{\mathrm{spec\,;NB}}$.

Lyα luminosity.

Following well-established procedures (see e.g., Pâris et al., 2011), we first identify the main spectral lines in our QSOs spectra, such as `CIV` and `CIII]`. We then use their profile-peaks to compute our redshift estimate. We discard the Lyα profile for this analysis, since it provides a systematically biased z measure, due to the complex radiative transfer of Lyα photons in the source rest frame and IGM (see e.g., Gronke

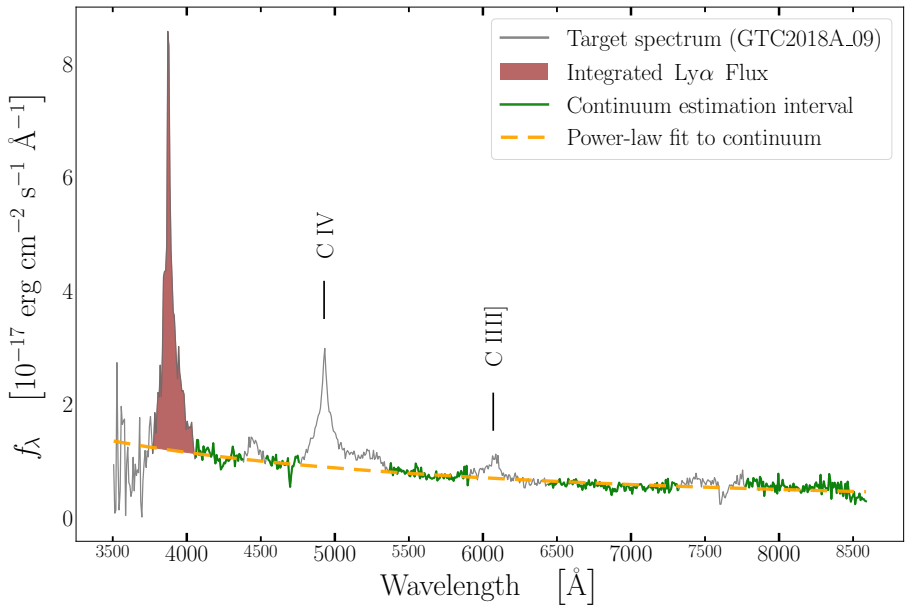

Figure 4.20: Calibrated spectrum (grey solid line) of the GTC2018A_09 target, confirmed as z∼2.2 QSO. The four green regions highlight the intervals used for computing the power-law fit to the continuum (dashed yellow line). Finally, the Lyα integrated flux is highlighted in red.

et al., 2016; Dijkstra, 2017; Gurung-Lopez et al., 2018; Gurung-López et al., 2019, 2020). We fit a double gaussian profile to both `CIV` and `CIII]` profiles, in order to trace at the same time its broad and narrow components. We use the λ position of the narrow-component peaks to obtain two z estimates, whose average provides the final spectroscopic z of our sources. The Lyα line flux can only be obtained after estimating the sources continua. We then fit a power law to the wavelength regions of each spectrum which are not affected by any line feature, as shown by the yellow sections of the spectrum displayed in Fig. 4.20. We use the following simple functional form:

$$\mathrm{f}_\lambda^{\mathrm{c}}(\lambda) = \mathrm{k}\,\lambda_{\mathrm{obs}}^{\alpha}\ , \tag{4.14}$$

where $\mathrm{f}_\lambda^{\mathrm{c}}(\lambda)$ is the spectrum *monochromatic flux density* (in units of erg cm^{-2} s^{-1} Å^{-1})) while k and α are fit parameters. Finally, we measure the total Lyα line flux by integrating the excess above the estimated continuum in the wavelength range affected by the Lyα line, which is shown in Fig. 4.20 as the spectral region highlighted in dark-red. As a last step, we estimate the observed Lyα EW as:

$$\mathrm{EW}_{\mathrm{obs}}^{\mathrm{Ly}\alpha} = \frac{\mathrm{F}_{\mathrm{Ly}\alpha}}{\mathrm{f}_\lambda^{\mathrm{c}}(\lambda_{\mathrm{Ly}\alpha})}\ , \tag{4.15}$$

in which $\mathrm{f}_\lambda^{\mathrm{c}}(\lambda_{\mathrm{Ly}\alpha})$ is the value of the power-law fit to the continuum at the wavelength of the Lyα line-profile peak. Figure 4.20 shows the spectrum of target `GTC2018A_09` as a visual example of our measuring procedure. The results of these measurements are shown in Table 4.7 together with a summary of the spectroscopic follow-up thechnical requrements and additional properties of the observed targets.

4.C Appendix: Multi-variate completeness computation

Here we describe the computation of the corrections accounting for incompleteness due to our selection methodology (§4.C.1) and the use of r-band detected catalogs (§4.C.2).

4.C.1 Selection completeness

To simplify, our selection depends on i) the linear approximation of the sources continuum, ii) their Lyα flux and iii) their Lyα EW. To account for these dependencies, we measure the recovery rate of our

Instrument	OSIRIS spectrograph at Gran Telescopio Canarias (GTC)
Grism	R500B
Seeing	requested: 1.4 (maximum) — effective: 1.03 (averaged on all observations)
Moon	requested: any — effective: >90% Dark
Air mass	requested: 1.5 (maximum) — effective: 1.266 (averaged on all observations)
SNR	$\geqslant 3$ at $\lambda_{obs} \sim 4000$Å

ID	Ra [hh:mm:ss]	Dec [hh:mm:ss]	Time [s]	SDSS class	GTC class	z_{spec}	$L_{Ly\alpha}$	EW_{obs}
GTC2018A_01	22:50:40.27	34:23:43.6	2185	—	QSO/AGN	2.21	5.18e+44	513.98
GTC2018A_02	00:43:38.56	05:41:35.6	2185	GALAXY	QSO/AGN	2.23	1.70e+44	417.21
GTC2018A_03	23:01:08.05	33:44:20.0	2185	STAR	STAR	—	—	—
GTC2018A_04	16:17:15.12	50:25:59.2	2185	STAR	QSO/AGN	1.52	—	—
GTC2018A_05	22:22:06.30	11:07:47.7	2335	STAR	QSO/AGN	2.23	3.87e+43	268.21
GTC2018A_06	18:10:22.34	41:49:25.3	2185	STAR	QSO/AGN	2.22	2.32e+44	741.21
GTC2018A_07	17:35:17.03	31:44:42.8	2185	GALAXY	QSO/AGN	2.23	7.46e+43	665.04
GTC2018A_08	01:31:29.69	33:55:14.9	2185	STAR	QSO/AGN	2.21	1.26e+44	637.59
GTC2018A_09	14:59:37.24	47:15:26.3	2185	STAR	QSO/AGN	2.18	2.13e+44	502.01
GTC2018A_10	16:11:57.72	46:00:45.8	2185	STAR	QSO/AGN	2.21	1.56e+44	263.97
GTC2018A_11	18:32:04.26	39:54:08.8	2185	—	QSO/AGN	1.54	—	—
GTC2018A_12	02:16:13.21	34:28:37.3	2185	—	QSO/AGN	2.21	1.33e+44	272.11
GTC2018A_13	14:32:51.07	52:36:46.7	1954	GALAXY	GALAXY	0.51	—	—
GTC2018A_14	22:01:43.81	28:23:36.5	2335	STAR	QSO/AGN	1.53	—	—
GTC2018A_15	16:09:37.67	45:29:53.6	2245	STAR	STAR	—	—	—
GTC2018A_16	22:43:00.76	34:10:26.4	2335	STAR	QSO/AGN	2.25	—	—
GTC2018A_17	15:59:27.15	57:05:04.6	2245	STAR	QSO/AGN	2.22	1.08e+44	268.69
GTC2018A_18	23:03:24.69	33:20:25.5	2245	STAR	STAR	—	—	—
GTC2018A_19	16:03:33.14	46:11:53.2	2245	—	QSO/AGN	2.25	6.19e+43	351.25
GTC2018A_20	15:19:49.02	53:16:18.4	2335	STAR	QSO/AGN	2.19	2.04e+44	828.73
GTC2018A_21	14:53:19.08	53:02:42.3	2365	STAR	QSO/AGN	2.19	6.15e+43	217.99
GTC2018A_22	17:41:33.43	57:05:11.4	2335	STAR	QSO/AGN	2.27	8.20e+43	481.13
GTC2018A_23	15:38:49.50	48:58:13.1	2014	GALAXY	STAR	—	—	—
GTC2018A_24	14:53:32.94	54:09:44.8	2335	GALAXY	QSO/AGN	1.53	—	—
GTC2019A_01	07:18:49.01	40:50:42.7	1165	STAR	STAR	—	—	—
GTC2019A_02	22:37:58.84	11:41:01.4	1285	STAR	QSO/AGN	2.204	4.56e+44	706.71
GTC2019A_03	02:27:21.78	29:56:23.7	1366	STAR	QSO/AGN	2.202	3.41e+44	859.82
GTC2019A_04	16:14:11.38	53:11:16.6	1426	STAR	BAL QSO	2.174	2.79e+44	290.75
GTC2019A_05	15:31:12.21	48:08:31.4	1576	STAR	QSO/AGN	2.209	2.91e+44	570.59
GTC2019A_06	15:34:37.75	46:42:36.3	1816	STAR	QSO/AGN	2.231	3.02e+44	414.59
GTC2019A_07	12:38:36.94	56:10:39.8	2086	STAR	QSO/AGN	2.177	2.02e+44	183.25
GTC2019A_08	12:29:07.02	56:03:49.2	2086	STAR	QSO/AGN	2.191	1.35e+44	230.15
GTC2019A_09	08:03:53.42	30:46:36.2	2206	STAR	QSO/AGN	2.256	2.14e+44	380.86
GTC2019A_10	07:15:23.13	39:50:57.3	2356	STAR	QSO/AGN	2.189	1.29e+44	625.73
GTC2019A_11	09:00:47.61	32:06:54.9	2806	GALAXY	QSO/AGN	2.253	9.54e+43	364.01
GTC2019A_12	00:51:10.29	03:08:25.4	3406	STAR	QSO/AGN	1.543	8.34e+43	104.21
GTC2019A_13	17:25:01.20	33:46:50.4	3706	STAR	QSO/AGN	1.602	1.27e+44	180.52
GTC2019A_14	10:28:20.14	39:52:42.3	3706	STAR	QSO/AGN	2.271	1.83e+44	121.36
GTC2019A_15	00:32:08.19	39:47:23.6	4476	STAR	QSO/AGN	1.512	1.33e+44	118.74
GTC2019A_16	16:14:30.92	50:12:24.2	4686	GALAXY	QSO/AGN	2.211	1.00e+44	705.82
GTC2019A_17	00:41:06.77	08:02:56.5	4761	STAR	GALAXY	—	—	—
GTC2019A_18	16:10:19.83	45:31:49.4	5136	GALAXY	QSO/AGN	2.271	1.06e+44	644.06
GTC2019A_19	22:56:29.07	09:36:45.5	5136	STAR	QSO/AGN	2.762	8.96e+43	390.88
GTC2019A_20	09:10:21.90	38:39:30.6	6246	GALAXY	QSO/AGN	1.528	1.61e+44	561.05
GTC2019A_21	13:20:29.68	56:31:49.6	6795	GALAXY	QSO/AGN	2.197	8.61e+43	154.68

Table 4.7: Properties retrieved from the follow-up of our 45 spectroscopic targets. These results confirm 29/45 sources (64.4%) as genuine Lyα-emitting QSOs at $z \sim 2$. The most numerous interlopers are CIV-emitting QSOs at $z \sim 1.52$, namely 8/45 targets ($\sim$ 18%, see Stroe et al., 2017a,b), and 5/45 blue stars ($\sim$ 11%). The latter are selected by our pipeline due to their strong color-gradients which mimic a NB photometric excess in the $J0395$ filter. We also note one Lyβ-emitting QSO contaminant at $z \sim 2.76$. We report the measured $L_{Ly\alpha}$ and EW_{obs} only for the confirmed QSOs at $z \sim 2.2$. Among these, the spectrum of GTC 16 could not be calibrated and measured.Finally, all the $z \sim 2.2$ confirmed sources in our sample are QSOs with $L_{Ly\alpha} > 6 \times 10^{43}\,\mathrm{erg\,s^{-1}}$. This supports the results of Konno et al. (2016); Matthee et al. (2017b); Sobral et al. (2018a,b) and Calhau et al. (2020) about the strong contribution of AGN/QSOs to the Lyα LF at $\mathrm{Log(Ly}\alpha) \gtrsim 43.3$.

methodology as a function of i) $g-r$ color, ii) Lyα flux and iii) r-band magnitude. Indeed, $g-r$ color can be thought as a proxy of our linear continuum approximation (see e.g. Fig. 4.2), hence the EW dependence of our selection is accounted for by independently varying the Lyα flux with respect to $g-r$ and r magnitude. More in detail:

1. we first subtract the measured $F_{Ly\alpha}^{3FM}$ from the sources photometry of our candidates, in order to get a list of *non-emitters*, i.e. sources without a significant NB excess according to our measuring method,

2. then we artificially re-add increasing values of line flux to the sources photometry (i.e. to the NB and the g band, since both are affected by the emission line),

3. for each value of the re-added flux, we apply our complete set of selection rules (§4.3.2) and we store the number of re-selected sources as a function of their r magnitude and $g-r$ color,

4. we finally compute the sources-recovery rate of our selection as $C = N_{selected}/N_{total}$ in each bin of r magnitude, $g-r$ color and artificially-injected Lyα flux.

With this method we obtain a 3D grid of recovery rates which can be interpolated in order to compute a selection weight C_i^s for each source. The latter accounts for the loss of candidates due to our selection within the total completeness correction we apply to the final LFs. The computation of C_i^s depends on Lyα-flux since this is the observable tackled by our measuring method (see §2.2.5). Nevertheless, this dependence can be directly converted into a $L_{Ly\alpha}$ dependence by assuming a redshift for every source (see §4.6.2). Figure 4.21 shows the recovery-rates grid for $J0430$ filter as an examples. The values are projected in the planes $F_{Ly\alpha}$ vs. r (left panel) and $F_{Ly\alpha}$ vs. $g-r$ (right panel).

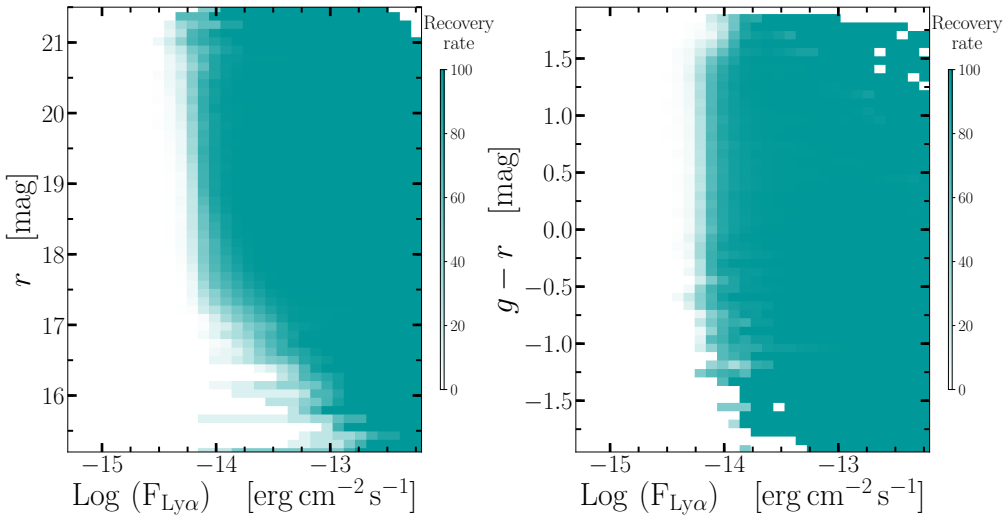

Figure 4.21: 3D grid of recovery rates for the $J0430$ filter, taken as example. Left and right panels respectively show the projections of recovery rates in the Lyα flux vs. r plane and Lyα flux vs. $g-r$ plane. We note that the recovery rates show noisy values at $r<17$ and $g-r<-1.25$ due to the low number of sources in these magnitude and color bins. Nevertheless, these regions of the 3D parameter space are excluded from the LF computation by the purity weight (§4.6.3).

4.C.2 Bivariate completeness model

Since we use r-band detected catalogs, we need to take into account the loss of undetected continuum-faint Lyα-emitting sources. In other words, we need to estimate the distribution of sources in regions of the L_{Lya} vs. r plane which lie outside the thresholds on r and NB-excess significance imposed by J-PLUS detection limits. To carry out this analysis we closely follow the methods of Gunawardhana et al. (2015), who tackled a very similar issue with a multi-variate approach. More in detail, we build the *bivariate* luminosity function of our candidates, defined as the number density of sources in each bin of $\mathrm{Log}\,(L_{Ly\alpha})$

and r, weighted by the purity and completeness corrections:

$$\Phi[\,\mathrm{Log}\,(\mathrm{L_{Ly\alpha}}),\,r\,]_{\,\mathrm{j\,k}} \;=\; \frac{\sum_{\mathrm{i}}\,[\mathrm{P_i}\,/\,(\mathrm{C_i^d}\cdot\mathrm{C_i^s})]}{\mathrm{V}\;\cdot\,\Delta_{\mathrm{j}}\,\mathrm{Log}(\mathrm{L_{Ly\alpha}})\;\cdot\,\Delta_{\mathrm{k}}r}\;, \tag{4.16}$$

where j and k indexes identify the 2D bins of $\mathrm{Log}(\mathrm{L_{Ly\alpha}})$ and r, while the index i runs over the total number of sources in each 2D bin. $\mathrm{P_i}$ is the purity weight of each candidate (see §4.6.3), while $\mathrm{C_i^d}$ and $\mathrm{C_i^s}$ are respectively its detection-completeness (§4.6.4) and selection-completeness (§4.6.4 and §4.C.1) weights. Finally, V is the survey effective volume for a given NB filter (see §4.6.2 and Table 4.1). By following Gunawardhana et al. (2015), we assume that the 2D LF can be modelled by the product of two functions, describing respectively the r and $\mathrm{Log}(\mathrm{L_{Ly\alpha}})$ distributions (see also Corbelli et al., 1991). We choose to employ the combination of a Schechter function (in logaritmic form) for r and a Gaussian in $\mathrm{Log\,L_{Ly\alpha}}$ (as in Gunawardhana et al., 2015):

$$\begin{aligned}
&\Phi(r) = 0.4\,\ln(10)\;\Phi^*_{\mathrm{r}}\;10^{\,0.4(\mathrm{r^*-r})(\alpha_{\mathrm{r}}+1)}\;\exp[-10^{0.4(\mathrm{r^*-r})}]\\
&\Phi(\mathrm{LogL}) = \tfrac{\Phi^*_{\mathrm{L}}}{\sigma_{\mathrm{L}}\sqrt{2\pi}}\exp\left[-\tfrac{1}{2\,\sigma^2_{\mathrm{L}}}\;(\mathrm{LogL}\;-\;\mathrm{LogL^*})^2\right]\;,
\end{aligned} \tag{4.17}$$

where Φ^*_{r}, $\mathrm{r^*}$, α_{r} are the ordinary Schechter parameters (see Schechter, 1976), while Φ^*_{L}, $\mathrm{L^*}$, σ_{L} describe respectively the number-density normalization, the average luminosity and the spread of the $\mathrm{L_{Ly\alpha}}$ distribution in each r bin. In order to obtain the bivariate model, we follow Gunawardhana et al. (2015) and join the two univariate distributions presented in Eq. 4.17 with an equation between their structural parameters:

$$\mathrm{L^*}(r) = 10^{\,\mathrm{A\,(r\,-\,r^*_0)\,+\,B}}\;, \tag{4.18}$$

where A and B are free parameters to be determined by fitting the 2D model to our data, while $\mathrm{r^*_0} = 19.5$ (as in Gunawardhana et al., 2015). By substituting Eq. 4.18 into 4.17 and multiplying the two univariate distributions, we obtain the full bivariate model:

$$\begin{aligned}
&\Phi[\,\mathrm{Log}\,(\mathrm{L_{Ly\alpha}}),\,r\,] = \Phi(r)\;\times\;\Phi(\mathrm{LogL_{Ly\alpha}}\,;\,r) =\\
&0.4\;\ln(10)\;\;\Phi^*_{\mathrm{r}}\;\;10^{\,0.4(\mathrm{r^*-r})(\alpha_{\mathrm{r}}+1)}\;\;\exp\left[-10^{0.4(\mathrm{r^*-r})}\right]\;\;\times\\
&\tfrac{\Phi^*_{\mathrm{L}}}{\sigma\,\sqrt{2\pi}}\exp\left\{-\tfrac{1}{2\,\sigma^2}\;[\mathrm{LogL}\;-\;(\mathrm{A\,(r\;-\;r^*_0)\;+\;B})]^2\right\}\;.
\end{aligned} \tag{4.19}$$

We fit the seven free parameters of this function using our measured 2D luminosity function. In particular, we only use the candidates with a completeness weight $(\mathrm{C_i^d}\cdot\mathrm{C_i^s}) > 0.85$, in order to avoid biasing our fit with regions affected by the incompleteness of our selection. Finally, the ratio between the measured 2D LF of our candidates and the fitted model (over the whole r - $\mathrm{L_{Ly\alpha}}$ plane) provides an estimate of the incompleteness of our selection in each $[\mathrm{Log}(\mathrm{L_{Ly\alpha}}), r]$ bin. We use this ratio to compute a *bivariate weight* $\mathrm{C_i^b}$ for each of our candidates. These weights are then combined to $\mathrm{C_i^d}$ and $\mathrm{C_i^s}$ in order to obtain our total completeness correction (see §4.6.4).

Chapter 5

Discussion and future perspectives

This chapter is devoted to address the main caveats of our methodologies and the possible biases introduced by them on our results. This also allows us to delineate possible future developments which can address and improve these shortcomings, as well as provide new scientific results. We split the discussion of this chapter in three main sections, which respectively discuss our model for SMBH-seeding (§5.1), the construction of the Lyα LF from J-PLUS data (§5.2 and some applications of the work described in previous sections to either ongoing and future projects (§5.3).

5.1 Analysis of our SMBHs formation and evolution model

The physical model described in §3.2 and §3.3 is based on working assumptions and approximations which allow to e.g. account for the un-resolved formation and evolution of SMBHs or to rapidly execute our version of L-Galaxies over the cosmological box of MR-II. **On the other hand, these approximations produce some limitations to our BH-seeding model, as is the case of the interplay between the outputs of GQd and the MR-II merger trees. Indeed, the first ones reproduce the evolution of a biased region of the Universe, therefore they cannot provide a general set of initial conditions for the variety of cosmological environment simulated by the MR-II (see the discussion in §3.3.2 and §3.4.5). Another example is provided by the limitations of our prescriptions for the diffusion of LW photons and metals, which might affect the descriptive power of our BH-seeding model. Indeed, to avoid computationally-expensive approaches, we resorted to simplifications of the physical modelling of both the SNe ejecta thermodynamics and the radiative transfer of UV photons in the IGM. Finally, the interplay between the SMBHs mass-growth model we employ, our grafting procedures and the properties of MR-II merger trees might be at the origin of the discrepancies we observe regarding the mass-growth history of SMBHs. Indeed, the mass-growth prescriptions of Izquierdo-Villalba et al. (2020) were developed and tested on MR structures without yet employing the BH-seeding model we developed. Therefore, both the significantly different merger history of MR structures with respect to MR-II ones (see Izquierdo-Villalba et al., 2019b), and the biases introduced by our grafting procedure (see §3.3.2) might produce the incongruencies discussed in §3.4.5. This might also suggests that the extension of the Izquierdo-Villalba et al. (2020) model to the MR-II dynamic range might need some adaptation to include new physical processes related to SMBH mass-growth which are not yet fully captured.** This section is dedicated to discuss all these critical points and to delineate some possible future development for our BH-seeding model.

5.1.1 Effects of GQd grafting on SMBH evolution in L-Galaxies

Section 3.4.1 shows that the effect of grafting GQd information into newly-initialized L-Galaxies structures is evident at all redshifts in our model. Indeed, pure-light seeds are the dominant class of seed-types in MR-II structures since $\gtrsim$6 down to $z\sim0$. On one side, given the ubiquity of SF processes in the Universe,

it is reasonable to expect that a large fraction of SMBHs observed as AGN at all epochs might have formed as light PopIII remnants (e.g. van Wassenhove et al., 2010). On the other hand, massive SMBHs powering $z \gtrsim QSOs$ are expected to be formed via heavier channels (e.g. RSMs or DCBHs) when super-Eddington accretion is not considered (see e.g. Valiante et al., 2016; Inayoshi et al., 2020; Lupi et al., 2021a; Sassano et al., 2021). Our model, on the contrary, does not predict a strong contribution from heavy seeding channels at any the massive end of the $z \sim 7$ BHMF.

As discussed in §3.4.1, this is entirely explainable as an effect of both our grafting procedures and the interplay between the dynamic ranges of `GQd` and `L-Galaxies`. In particular, the structures within the outputs of `GQd` we use only form heavy seeds at $z \lesssim 13$ (see Valiante et al., 2021). At these redshifts, the `MR-II` and `GQd` M_{vir} ranges overlap completely (see Fig. 3.1). Furthermore, albeit rare, at these redshifts BHs with heavy-seed components (i.e. mixed seeds, see §3.3.2) are distributed over a large fraction of the `GQd` mass range. Consequently, `L-Galaxies` structures have a non-negligible probability of being initialized with a BH whose mass was assembled with an heavy-seed component, during its `GQd` evolution. Since our grafting is based on the random sampling of the `GQd` halos catalog, the possibility of inheriting an heavy-seed component from `GQd` is higher at $M_{vir}^{res} 2 \times 10^8 M_\odot$, i.e. the mass at which most of the `L-Galaxies` halos are newly-resolved. With the progress of cosmic time, this interplay changes due to the evolution of `GQd` mass resolution. Indeed, already by $z \sim 10$ most of `L-Galaxies` structures are initialized as empty or with pure-light remnants, as at now they sample the low-mass part of the `GQd` dynamic range (where contributions from heavy seeds are absent). The net effect of this procedure is to induce a strong predominance of pure-light seed descendants with respect to SMBHs showing a mass contribution from heavy seed-types.

Overall, the bias towards low-mass seeds we observe in our results is produced by the combination of several aspects. The different modelling of the underlying merger trees, for instance, implies an evolving mass resolution for `GQd` and a fixed one for `L-Galaxies`. This determines a short time-window in which the two dynamic ranges overlap completely, so that the whole `GQd` mass distribution and SMBH-population can be sampled by our grafting procedures. Furthermore, the period in which heavy-seeds components evolved in `GQd` can be transmitted to `L-Galaxies` structures is short (only 3 `MR-II` snapshots), so that these cannot effectively inherit mixed seeds descendants. Finally, the bulk of `L-Galaxies` structures are initialized over a mass range which (at $z < 13$) ultimately favours the grafting of pure-light descendants or no BHs.

This *grafting bias* which affects our results can be mitigated by improving the interplay between the dynamic range of halos newly-resolved in `MR-II` and the halos catalogs used as input for the grafting procedures. Indeed, our work is based on 10 repetitions of a specific DM over-density which produces a Milky Way-like halo (i.e. $M_{vir} = 10^{13} M_\odot$) at $z = 2$. Nevertheless, these only reproduce the merger history of a specific, moderately biased region of the Universe. Therefore, a possible improvement of our grafting prescription which would still make use of `GQd` could simulate the evolution of halos with a wider spectrum of final masses, therefore accessing to different environments. These can be then used to feed more accurate grafting information to `L-Galaxies` structures. In §5.1.4 we present possible further solutions which do not necessarily make use of `GQd` outputs.

5.1.2 Physical modelling of LW field and IGM metallicity

Section §3.2 presented the physical models we introduced to account for both uniform and spatially-varying backgrounds of UV photons and metallicity. In particular, we discussed the approximations of Ahn et al. (2009) and Dijkstra et al. (2014a) thanks to which, respectively, the overall transmission of the IGM towards ionizing photons can be accounted avoiding expensive RT computations, and the radii of metallic shells launched by powerful SNe explosions can be computed without recurring to fully-hydrdynamical approaches. On top of the simplifying hypothesis on which these methods are based (see Ahn et al., 2009; Dijkstra et al., 2014a) we further assumed reasonable values for the free parameters and avoid the computationally expensive modelling of complex phenomena. Regarding the latter, for instance, we neglected the thermodynamics of the interaction between the metallic shells fronts and the IGM or the complex modifications induced to the spectrum of the emitted ionizing radiation from galaxies due to ISM radiative-transfer. Improving each of these aspects might have non-straightforward effects

on our results and therefore deserves a separate analysis. Indeed, a more detailed treatment of metal diffusion in the IGM might reduce the radius of regions affected by chemical pollution, hence fostering the production of DCBH seeds. On the contrary, the LW flux received by BH-seeding candidates might be damped due to an improved model for the ISM and IGM radiative transfer of ionizing photons (see e.g. Gurung-López et al., 2019, 2020, for a theoretical approach in the GALFORM SAM).

Furthermore, the outcomes of our model can significantly change by varying the free parameters of the equations governing the diffusion of metals and UV photons. For instance, the rate of H_2-dissociating photons per stellar barion (η in eq. 3.2) can modify $J_{LW\,;\,bg}$ and increase the amount of RSM produced by our model. Analogously, the SNe frequency ν and energy input E_0 (eq. 3.3), as well as the modification n (eq. 3.4) for each SNe feedback episode may play an important role in the final abundance of BH-seeds produced self-consistently in `L-Galaxies`.

Improving the model discussed above by including more accurate details within the context of the analytic recipes employed by `L-Galaxies` is not trivial. On one side, more detailed approach for the diffusion of metals (e.g., Bertone et al., 2005, 2007) or UV photons (e.g., Gurung-López et al., 2019, 2020) have been carried out in semi-analytic contexts. On the other side, it can be argued that these more detailed aspects of the physical model for Z_{IGM} and J_{LW} can be absorbed by the free parameters which are already employed by our model. These, in turn, could be calibrated against numerical simulations which include more detailed treatments for the hydrodynamics and radiative transfer of metals and UV photons diffusion in the ISM and IGM. Nevertheless, numerical simulations rarely reach the joint combination of mass-resolution and volume exhibited by the `MR-II`, which makes this calibration difficult (e.g., Volonteri et al., 2016; Maio et al., 2019; Marshall et al., 2020b). Smaller simulations can provide this comparison (see e.g., Habouzit et al., 2016; Di Matteo et al., 2017), but it is not yet clear how their results at high-z can be generalized to large cosmological volumes. We plan to consider these approaches in our upcoming analysis.

Although the aspects discussed above can effectively modify the prescriptions for $J_{LW\,;\,local}$ and $Z_{IGM\,;\,local}$, the actual driving factor for the approximations we used in our physical model was a technical aspect of the `L-Galaxies` code. Indeed, the latter is structured to follow the evolution of single merger-trees serially. Halos belonging to each particular merger tree are evolved in redshift independently of the eventual presence of neighboring halos belonging to different merger trees. Consequently, at any given time for any given merger tree, `L-Galaxies` cannot access to the evolutionary information of all the halos in the whole box. This clearly hinders the possibility of evaluating self-consistently the contributions to $J_{LW\,;\,local}$ and $Z_{IGM\,;\,local}$, at any redshift, during the code execution. For this reason, we introduced the lists of LW-sources and metal-sources described in §3.2.4, which allow to track the contributions of the whole box at all redshifts. Nevertheless, due to computational time and memory usage requirements, we could include only a limited sample of sources within these lists; namely, the most intensely star-forming galaxies for the LW one, and the galaxies with extended metallic shells for the metallic-sources list (see §3.2.4 for details). Despite the definition of these lists was explicitly determined through the best compromise between execution speed and highest possible number of sources, by construction, our method cannot account exactly for all the possible contributions in the box.

In order to improve the measurement of the contribution to $J_{LW\,;\,local}$ and $Z_{IGM\,;\,local}$ produced by the whole box, different approaches are possible. For instance, these two fields could be evaluated over spatial grids, which could provide their value at each point in the box (and at each redshift). The main limitation of this method would then be the spatial resolution of the spatial grids. Indeed, in order to avoid intensive storage usage, the grids resolution might be kept too low to capture the small-scale variations needed for BH-seeding processes.

Interestingly, a different approach could be provided by the recent version of `L-Galaxies2020` (see Yates et al., 2021). The latter includes an efficient way to store information about the detailed SF histories of galaxies (see also De Lucia et al., 2014, for a similar approach). In principle, this allows to access to the energy and chemical feedback of the whole SF history of galaxies in a very efficient way, along their evolution. Consequently, these SF-history arrays might help to accurately model the contribution provided to $J_{LW\,;\,local}$ and $Z_{IGM\,;\,local}$ by each galaxy in the box, without the need of defining our source-lists. We plan to test this possibility in the near future.

5.1.3 Interplay between BH-seeding and BH-growth in L-Galaxies

In §3.4.5 we showed that, despite the $z\sim0$ BHMF is well-reproduced by the Izquierdo-Villalba et al. (2020) growth model, the combination of the latter with our BH-seeding prescriptions fails to correctly describe the growth history of SMBHs in the MR-II box. In particular, SMBHs appear to grow significantly much earlier than observed, i.e. at $z2$, while being less active than expected at $z<1$. This behavior is clearly evident both in the bolometric and X-ray LFs (respectively Figures 3.15 and 3.16), as well as in the evolution of the BHAR density predicted by our model. In particular, the latter shows a clear peak at $z\sim4$, which is anticipated with respect to current constraints peaking at $z\sim2$ (see the bottom panel of Fig. 3.17). We note that recent theoretical works performed with comparable mass-resolution and simulated-volume setups as ours also found similar mismatches with respect to observed LFs and mass-growth history of SMBHs (see e.g., Marulli et al., 2007; DeGraf & Sijacki, 2019; Marshall et al., 2020a).

The predictions of our model at $z\gtrsim2$ slightly improve by acting on the G_P parameter, as shown by the comparison between the $G_P=1$ and $G_P=1$ runs (see Figures 3.15, 3.16 and 3.17). Nevertheless, this does not allow to reconcile the outcomes of our model with current constraints across the whole cosmological evolution of SMBHs. This suggests that the details of or BH seeding model might play a role on the evolution of SMBHs at high-z but they become less and less important with the progress of cosmic time. Consequently, in order to reconcile our predictions with current constraints, it might be necessary to modify the Izquierdo-Villalba et al. (2020) mass-growth model. Indeed, the latter was tested and calibrated on the MR merger trees, therefore it is possible that new physical processes should be included to better describe the growth of SMBHs at the low-mass dynamical range simulated by MR-II. In particular, we checked that our results on the bLF, Xray LF and BHAR density do not improve by either i) acting on the free parameters of the Izquierdo-Villalba et al. (2020) growth model, ii) applying the obscuration corrections proposed by Ueda et al. (2014) and Merloni et al. (2014), and iii) modifying the physical prescription which determines the gas supply to central BHs after galaxy mergers. None of these cases produced a significant improvement to our results, therefore we do not report the outcomes of these trials. We conclude that further analysis and improvements are needed to reconcile the SMBHs mass-growth history predicted by our model with current constraints. We refrain to comment further this point and present the future perspectives of these ongoing improvement work in §5.1.4.

Although our growth model requires modifications to improve the agreement of the SMBHs mass-growth history, we note that it predicts the total integrated growth of SMBHs. This is shown by both the $z\sim0$ BHMF (Fig. 3.8) and $M_\bullet$ density (upper panel of Fig. 3.17). Therefore, it is interesting to analyze the results of our model at $z\sim0$, where our model is broadly consistent with observations, independently of the G_P value used. Indeed, due to the dynamical range they probe, MR-II merger trees effectively allow to concentrate on the evolution of dwarf galaxies and their central BHs, a currently active field of research (see e.g., Neumayer & Walcher, 2012; Mezcua et al., 2016, 2018, 2019; Mezcua & Domínguez Sánchez, 2020; Pacucci et al., 2021; Volonteri et al., 2020).

We find that all the few DCBHs and RSM descendants which retain memory of their seeding channel (i.e. by not experiencing mergers with different seed-types descendants nor significant mass growth) end up in dwarf galaxies. This is consistent with current pictures which relates local low-mass SMBHs with the remnants of high-z seeding phenomena (see e.g., van Wassenhove et al., 2010; Cann et al., 2021; Mezcua, 2021). On the other hand, Mezcua (2019) argues that the mass-growth of low-mass SMBHs might be hindered by several phenomena such as galaxy interactions and efficient self-regulation, so that the actual connection between the seeding processes and the low-mass end of the SMBHs mass distribution might be blurred by the cosmological evolution of dwarf galaxies. If we focus on the overall population of low-mass SMBHs predicted by our model, this is actually what we find. Indeed, at these low-mass regime our BHMF is dominated by pure-light seeds descendants which experienced a mild mass-growth throughout their history (see §3.4.4). Indeed, given the typical pure-light seed mass of $M_{seed}\sim10^2M_\odot$ modelled in GQd (see Valiante et al., 2016, 2021), the pure-light seeds-descendants with $M_\bullet\sim10^{4-5}M_\odot$ underwent few mass-accretion and mergers episodes during their evolution. This is confirmed by the $z\sim0$ BHMFs shown in the upper-left panel of Fig. 3.19. Indeed, our model predicts that only a low fractions of SMBHs with $M_\bullet\sim10^{4-5}M_\odot$ is ungrown with respect to their seed mass. Furthermore, this result is only referred to the

SMBHs growth we can resolve in L-Galaxies, therefore the fraction of ungrown SMBHs at the low-mass tail of the BHMF would be even lower if we would consider their unresolved growth during their GQd evolution. As a consequence, the significance of growth processes also at low $M_\bullet$ would be higher than what our model predicts. This additionally supports the idea suggested by Mezcua (2019) and, more generally, the picture according to which the details of BH seeding are blurred in the $z \sim 0$ BHMF due to the mass-growth fuelled by either the hierarchical assembly of structures or internal, secular processes within their hosts (e.g., Marulli et al., 2008; Bonoli et al., 2009; Capelo et al., 2015; Qin et al., 2017; Izquierdo-Villalba et al., 2020; Marshall et al., 2020a).

Finally, we underline that our conclusions might be affected by the strong prevalence of pure-light seeds descendants in our model. Indeed, it might be speculated that a higher abundance of pure-RSM or pure-DCBH seeds might reverberate at the low-mass end of the $z \sim 0$ BHMF with a higher significance with respect to what we observe in our results. As already discussed, this possibility cannot be tested in a self-consistent way directly on the MR-II merger trees, but rather requires a higher-resolution simulation which would allow to track self-consistently the formation of PopIII remnants, RSMs and DCBHs at the same time. We further discuss this point in §5.1.4.

5.1.4 Improving the BH-seeding and SMBH evolution model

As pointed out in the above sections, our model is able to reproduce the integrated properties of SMBHs at $z \sim 0$, namely the BHMF and the $\rho(M_\bullet)$, as shown in Fig. 3.8 and in the upper panel of Fig. 3.17. We underline that ours is one of the first models which attains this result on a cosmological volume ($L_{box} = 100\,Mpc\,h^{-1}$ for MR-II) by exploiting a physically-motivated BH-seeding prescription according to which the conditions for the formation of SMBH seeds are only verified at $z \gtrsim 6$. This is substantially different by the commonly followed approach according to which a central BH is associated to every newly-resolved structure, down to $z \sim 0$ (e.g., Bonoli et al., 2009; Croton et al., 2016; Qin et al., 2017; Marshall et al., 2020a; Izquierdo-Villalba et al., 2020). Nevertheless, the current predictions of our model fail to meet current constraints regarding the evolution of the mass-assemply of SMBHs across cosmic time. As discussed in previous sections, these caveats might be produced by the necessary approximations we undertook in order to include the formation of PopIII remnants in our model (see §5.1.1), the diffusion of ionizing photons and metals in the IGM (see §5.1.2) and the extension of the Izquierdo-Villalba et al. (2020) mass-growth model to the dynamical range of the MR-II merger trees (see §5.1.3). For each of these aspects, it is possible to envision a number of future improvements, which could lead to a better description of the cosmological evolution of BH-seeds descendants.

The nested-box approach

Mass-resolution is a key limitation for possible developments of our BH-seeding prescription. For instance, the impossibility to model directly the formation of light PopIII remnants does not allow to study the interplay between light, intermediate and heavy seeding scenario in a fully self-consistent picture. This fostered the introduction of the grafting procedure described in §3.3.2, which introduce the biases discussed in §3.4 and §5.1.1.

These kind of issues could be addressed and mitigated by exploiting a *nested* approach, according to which simulations of increasing size (and decreasing mass-resolution) are initialized with the outcomes of their smaller-scale and higher-resolution counterparts. One possible application of this idea would be to run a suite of N-body simulations of increasing size, from the mass-scale adequate for tracking the small-scale physics of gas-cooling and PopIII star formation up to large cosmological volumes which can access to the low number-density of rare, bright QSOs at $z \gtrsim 6$. These simulations could provide the merger trees over which a specifically-modified versions of L-Galaxies can be run in order to produce physically-motivated initial conditions for the merger trees of larger simulations. In this way, L-Galaxies would be able to directly model self-consistently the formation of BH seeds over the entire mass spectrum envisioned by current models (see §1.2.2). In other words, this procedure would extend the grafting procedure we defined for GQd outputs to a set of nested, simulated boxes, therefore mitigating the bias introduced by the interplay between GQd and MR-II dynamic ranges. In particular, being based on cosmological, N-body simulations, this approach would naturally track the formation of BH seeds on a

variety of environments rather than extending the results obtained on a specific merger tree to the whole box of the MR-II simulation.

In principle, the nested-boxes approach could be used to reach very large volumes, therefore allowing to extend the small-scale processes simulated into small box in a sub-grid fashion over large-scale cosmological environments. In the context of BH-seeding models, this potentially allows to study the role of different seeding channels in the formation of the SMBHs powering $z \gtrsim 7$ QSOs with a much wider statistics than what is currently attainable by single N-body, cosmological simulations. Ideally, this could provide the means to study the impact of small-scale physical models onto very large scales and provide the basis for the theoretical interpretation of future-surveys data focusing on the high-z Universe, such as the JWST or Athena space missions. Furthermore, by making use of a semi-analytic code, the nested-boxes approach potentially allow to efficiently test new physical prescriptions over a very-wide dynamical range. As an example, this method would allow to test the super-competitive accretion scenario for BH seeds formation recently presented by Chon & Omukai (2020) and explored by Sassano et al. (2021). Indeed, this new interesting model has been presented and tested onto a relatively small cosmological box ($20\mathrm{Mpc\,h^{-1}}$, see Chon et al., 2018; Chon & Omukai, 2020) or in the context of the purely-analytic GQd model (see Sassano et al., 2021).

Improving the SMBH mass-growth model

As discussed, both the BHMF (Fig. 3.8) and the $M_\bullet$ density (upper panel of Fig. 3.17) are consistent with $z \sim 0$ observational constraints. This indicates that the net effect of the modifications required by our mass-growth model for SMBHs should be to shift the accretion of mass onto BHs towards $z = 0$. This seemingly simple objective is made complex by the several processes acting at the same time in order to shape the evolution of SMBHs.

It can be argued that by reducing the efficiency of mass accretion, the actual build-up of BH masses can be delayed in time. We tried this approach by relaxing the assumption according to which the first-phase of mass accretion in the Izquierdo-Villalba et al. (2020) model proceeds at the Eddington rate for all BHs in the simulated box (see also Bonoli et al., 2009; Marulli et al., 2008). In particular, we modelled the $f_{\rm Edd}$ distribution of SMBHs by random sampling a log-normal distribution centered at $f_{\rm Edd} = 0.4$, which roughly replicates the observational estimates of Aird et al. (2018). According to this model, after the gas-supply provided by either galaxy mergers or internal disk-instabilities, BHs grow in time with a fixed $f_{\rm Edd}$ extracted by the log-normal distribution This fixed $f_{\rm Edd}$ can change in case a consistent amount of new gas-supply is provided by new mergers or disk instabilities. Although the net effect of these modifications is to slightly improve the bLF we predict at $z \sim 1$, there is no choice of parameters for the log-normal distribution which actually allows to reconcile our predictions to current constraints over a wide redshift interval. We attained similar conclusions by modifying the parameters controlling the quiescent phase of mass accretion (see Marulli et al., 2008; Izquierdo-Villalba et al., 2020, for details), with the aim of prolonging the low-luminosity active-phase of massive BHs at $z \lesssim 1$. Also in this case, the overall improvements to the growth-history of SMBHs are negligible.

Since the significant excess at the faint-end of the $z \gtrsim 2$ bLF in our model is dominated by $M_\bullet < 10^7 M_\odot$ BHs hosted in DM halos with $M_{\rm vir} < 10^{10} M_\odot$, a complementary possibility is to act onto the gas supply of central BHs in dwarf halos. In particular, we implemented a damping of the fuelling efficiency induced by mergers and disk instabilities with an exponential dependence on $M_{\rm vir}$, in order to artificially "starve" this class of BHs. This ad-hoc solution improved our high-z predictions (albeit slightly) but significantly worsened the results of our model at $z \sim 0$. Finally, as already discussed, acting on the abundance of BH-seeds has the strongest impact onto our results, although it does not allow to correctly reconstruct the progress of BH mass accretion within the MR-II box.

As discussed by Izquierdo-Villalba et al. (2020), the results of applying L-Galaxies to the MR merger trees show a slight under-estimation of the predicted bLF at $z < 0.5$ and $10^{44} < L_{\rm bol}/[\mathrm{erg\,s^{-1}}] < 10^{45}$. The authors argue that this might be due to a lack of physical models which can account for this range of luminosity at low-z. This might reflect onto our results even more dramatically, given the different dynamical range probed by MR-II. In particular, they show a clear dependence with galaxy morphology of the main contributions to the bLF. Indeed, at least for the case of MR simulation, the bLF shifts from

being dominated by galaxies hosting pseudo-bulges (at $z!2$) to being dominated by classical bulges and elliptical galaxies (at $z<2$). This also implies that the dominant mechanisms of SMBH fuelling shift from disk-instabilities to mergers. Indeed, as shown by Izquierdo-Villalba et al. (2019b), due to the different dynamical range they model, `MR-II` and `MR` galaxies have profoundly different merger histories. On one side, this affects the fractions of pseudo-bulges, classical bulges and ellipticals but, most importantly for our case, this also deeply affects the fuelling of SMBHs, as testified by the presence of the merger ratio m_r in eq. 2.12 and the possibility of fuelling BHs via disk-instabilities (eq. 2.16). This suggests that the origin of the discrepancies which we observe in our results with respect to current constraints might be ascribed to the interplay between the gas-fuelling model in `L-Galaxies` and the resolution of the underlying N-body simulation. Indeed, this model was tested and calibrated on the `MR` dynamical range, but it might require substantial re-adaptation for the `MR-II` one.

These analysis and considerations suggest that, in order to reconcile the overall growth of SMBHs to current constraints, a multi-variate approach is required. Indeed, it is necessary to focus at the same time on several different aspects of the build-up of the SMBH population across cosmic times. Therefore, in addition to exploring the implementation of additional physically-motivated prescriptions for SMBH growth in low-mass halos, it is fundamental to calibrate the free-parameters of these recipes by keeping into account their high degree of degeneracy. This task can be attained by exploiting efficient methods currently developed and applied in the context of SAMs calibration (Ruiz et al., 2015; Croton et al., 2016; Buchner, 2021). The downside of this approach towards improving the faint-end of the high-z bLF, spacifically, is the lack of observational data at $L_{bol} \lesssim 10^{44} erg\,s^{-1}$. This may hinder the possibility of determining the correct model parameters to improve our model predictions at high-z in favour of a better agreement of our results with current $z\sim0$ constraints.

5.2 Discussion of the photometric determination of the Lyα LF

The observational results described in §4 represent one of the first extension of the Lyα LF to unprecedented regions of the $L_{Ly\alpha}$-Φ parameter space. This task exploited the photoemtric data of the `J-PLUS` survey, wich offers only limited spectral information. Consequently, the processes of source selection, identification and measurement needed to recur to a number of working assumptions. Despite the possible biases introduced by these approximations were addressed in §4, some aspects deserve further discussion. This section is dedicated to present the caveats of our analysis and to introduce some possible future improvements to it.

5.2.1 Selection errors of Lyα-emitting candidates

We base our determination of the Lyα LF on the selection function described in §4.3.2. This is composed of a series of photometric cuts and selection conditions which are adapted to each of the 511 tiles composing the `J-PLUS` DR1 catalog (see §4.2.1 and Cenarro et al., 2019). In particular the NB excess significance cut, the NB and BB signal-to-noise cuts and the minimum NB color cut were measured separately on each of the 511 tiles. On one side, due to the large tile-by-tile variations in the average properties of sources, this ensured to apply homogeneous selection criteria on the whole DR1 footprint. On the other hand, this procedure relied on photometric cuts determined from the sources properties in each tile, therefore the exact values of the thresholds we imposed to the sources photometry are affected by measurement errors. As a consequence, contaminants can be artificially selected and Lyα-emitting sources can be erroneously excluded from the final selected sample.

In order to simplify our approach we ignored *selection errors* on our final results, under the implicit assumption that these were smaller than the final errors on our LF, such as those introduced by e.g. our Lyα flux measurement or by the conversion of the latter into a luminosity estimate (see §4.6). Nevertheless, as discussed in e.g. Sobral et al. (2018b), the effect of selection errors can be taken into account by exploiting the results of multiple realizations of the same photometric selection. In this way, the final sample of candidates is obtained by considering the probability of each source to meet the selection criteria. This method can be performed either by perturbing directly the selection criteria according to their errors or by perturbing the sources photometry according to photometric errors. In

other words, this approach produces a distribution function for each source, describing its probability of being selected as a function of the photometric cuts values, or as a function of the sources photometry.

The effect of selection errors is not straightforward. Indeed, on one side faint sources who are close to the selection cuts might exhibit a low probability of selection and therefore be considered as interlopers. On the other hand, also rare continuum-faint and Lyα-bright sources might be discarded as interlopers. to avoid these undesired effects, the effect of selection errors can be used as a statistical weight for each source when computing the overall properties of the selected sample, such as the LF or the EW distribution (e.g. Fig. 4.5). We underline that the J-PLUS DR2 offers a natural setting to implement and test these techniques in order to improve our selection pipeline for future determinations of the Lyα LF.

5.2.2 Classification errors of Lyα-emitting candidates

As we made explicit in §4.3.2 and §4.3.3, the photometric selection of sources and their definition as genuine candidates are two distinct problems, which we tackled with different methods. In a sense, the two procedures can be seen as two separate methods to respectively build a catalog with the highest possible completeness (photometric selection) and the highest possible purity (candidates definition). Indeed, our definition of genuine candidates is largely based in the attempt to remove the highest possible number of known interlopers by exploiting crossmatches with existing datasets.

The identification of known interlopers can be based on a secure classification, such as in the case of the spectroscopically-selected and visually-confirmed sample of SDSS DR14 QSOs compiled by Pâris et al. (2018) we used. On the other hand, we also employed identification methods tailored for our purposes defined on the properties of sources observed by different surveys, such as Gaia DR2 (Gaia Collaboration et al., 2018) and GALEX (Gil de Paz et al., 2009). Furthermore, we made use of the information provided by J-PLUS photo-spectra and observed morphology to further clean our samples (see §4.3.3 and §4.3.4). As in the case of our photometric selection described above, these definitions are based on selection cuts, which can introduce classification errors. As described in §4.6.3, we attempted to tackle these errors by defining a statistical purity weight for our genuine candidates. Nevertheless, our source classification is still prone to uncertainties, which can be taken into account in future developments of our selection pipeline. For instance, our analysis did not include a method to specifically exclude white dwarfs (WDs), whose blue spectrum can be effectively induce a wrongful classification by our pipeline. More in detail, for the specific case of WDs, the recent work of López-Sanjuan et al. (2021) offers a statistical study which can be exploited in order to identify WDs within J-PLUS data. As for the case of selection errors, the J-PLUS DR2 offers the possibility to also test and improve our methodology by accounting for classification errors.

5.2.3 Effects of luminosity estimates

One of the main sources of uncertainties on our final LF determination is the estimate of the $L_{Ly\alpha}$ of our candidates. Indeed, as discussed in §4.6.4, these are affected by both the uncertainties of our method to measure the Lyα flux of our sources (see e.g. §2.2.5 and §4.6.1) and by our redshift uncertainties (§4.6.2). Although these were accounted by our monte-carlo determination of the final errors on the Lyα LF (see §4.6.4), the estimate of the Lyα luminosity for each source was based on a fixed statistical correction accounting for the loss of Lyα flux due to the filter-width of J-PLUS NBs (see 4.6.1). Since these errors affect the distribution of sources within the luminosity bins used to estimate the Lyα LF, they ultimately produce an impact onto the final shape of the LF and on the estimation of its structural parameters.

Unfortunately, given the very limited spectral information provided by the J-PLUS filter set, the impact of these biases can only be marginally reduced. For instance, by being blind towards the actual nature of our selected candidates, it is extremely difficult to obtain a precise estimate of their redshift. Analogously, the reconstruction of broad line profiles of QSOs (see e.g. Chaves-Montero et al., 2021) could mitigate the effect of flux losses due to the NB widths, but it could only be applied to a limited set of NB filters, i.e. those having redwards adjacent NBs (i.e. $J0395$ and $J0410$), due to the asymmetric shape of the Lyα line profile. We conclude that, for the specific case of the errors introduced by our Lyα luminosity estimate, the monte-carlo determination of our final errors on the LF provides a realistic estimate of the uncertainties affecting the structural parameters of our Lyα LF.

5.3 Future perspectives and applications

This thesis focused on the study of the formation and evolution of SMBHs. We adopted both a theoretical perspective, to study high-z seeding processes and their impact at low-z on a simulated cosmological environment, and observational methodologies, to build the observed LF of $2<z<3.3$ AGN at rest-frame Lyα wavelengths. We reached significant results on both fields: on one side we embedded a comprehensive BH-seeding model in the context of the `L-Galaxies` SAM and applied the latter onto the cosmological, N-body merger trees of the `MR-II` simulation. On the observational side, we extended current constraints on the brightest-end of the Lyα LF, interpreting the excess pointed out by previous works as the exponential-end of a fully-developed Schechter function, extending beyond $L_{Ly\alpha} \gtrsim 10^{45}\,erg\,s^{-1}$. This section is devoted to briefly discuss a few possible developments of our work which either make part of ongoing research projects or might set the basis for upcoming extension of the presented results.

5.3.1 `L-GalaxiesBH`: an extensive model for SMBHs formation and evolution

The properties of the SMBH population at different redshifts can be strongly influenced by the additional physical processes modelled by Izquierdo-Villalba et al. (2020), which we introduced in §2.1.2. In particular, the dependence of the accretion efficiency with the BH spin, the time-delay between the mergers of galaxies and their central BHs, the emission of GWs during the latter process and the associated gravitational kicks can deeply modify both the history of BH-mass assembly and the occupation of SMBHs in low-mass galaxies. This is particularly important for `MR-II` , due to the abundance of low-mass halos with shallow gravitational potential, which might hardly retain kicked SMBHs or re-capture them only after large time-scales (see the discussion in Izquierdo-Villalba et al., 2020). It is therefore straightforward to apply the recipes already implemented in `L-Galaxies` also to the `MR-II` merger trees and combine them with our seeding prescriptions. This step is extremely interesting not only for the possible modifications to our results which can be expected, but also because by joining all the physics introduced by (Izquierdo-Villalba et al., 2020) and our BH-seeding model, `L-Galaxies` would become one of the most comprehensive semi-analytic models for SMBHs formation and evolution currently available in the literature. In addition, by exploiting the nested-boxes idea described above, this SMBH model can be applied to a large cosmological volumes and the diverse environments simulated therein. This might produce a wealth of possible scientific results and applications, ranging from the study of scaling relations between galaxies and SMBHs over a very wide dynamical range up to the production of physically-motivated mocks for future GW-detection experiments such as LISA (see e.g., Valiante et al., 2021, for the application of `GQd` to similar purposes). We underline that our work focused primarily on the development of a physically-motivated BH-seeding model and the study of its impact onto the cosmological evolution of SMBHs. For this reason, we did not apply the prescriptions implemented by Izquierdo-Villalba et al. (2020) to our model and only concentrated on isolating the effect of BH-seeding prescriptions onto the low-z population of SMBHs.

5.3.2 Drawing predictions for the observability of BH-seeding processes with JWST

The content of this section makes part of an ongoing research project directed by Dr. T. Hartwig and Prof. N. Yoshida. This preliminary work was performed during a working-visit at the University of Tokyo and KAVLI IPMU center (Japan)

The combination of our BH-seeding model and the Izquierdo-Villalba et al. (2020) mass-growth model allows to track the radiative emission of accreting SMBHs in time (see §3.4.5). This offers the possibility to simulate the observational counterparts of the first QSOs at $z \gtrsim 7$ and to extract predictions for their observability with future surveys. Furthermore, if the observed properties of these sources are connected to the formation scenario of their active SMBH (i.e. light, intermediate or heavy channels), they might allow to obtain an indirect probe of high-z BH-seeding processes and of the environment in which they happen. To tackle this goal, we built the tools to simulate the observational counterparts of active SMBHs at high-z, connected the properties of the QSOs radiative emission to local properties of their host galaxies and finally applied a machine-learning method (developed by Dr. T. Hartwig, see e.g., Hartwig et al., 2018) in order to extract information and predictions from the simulated observational

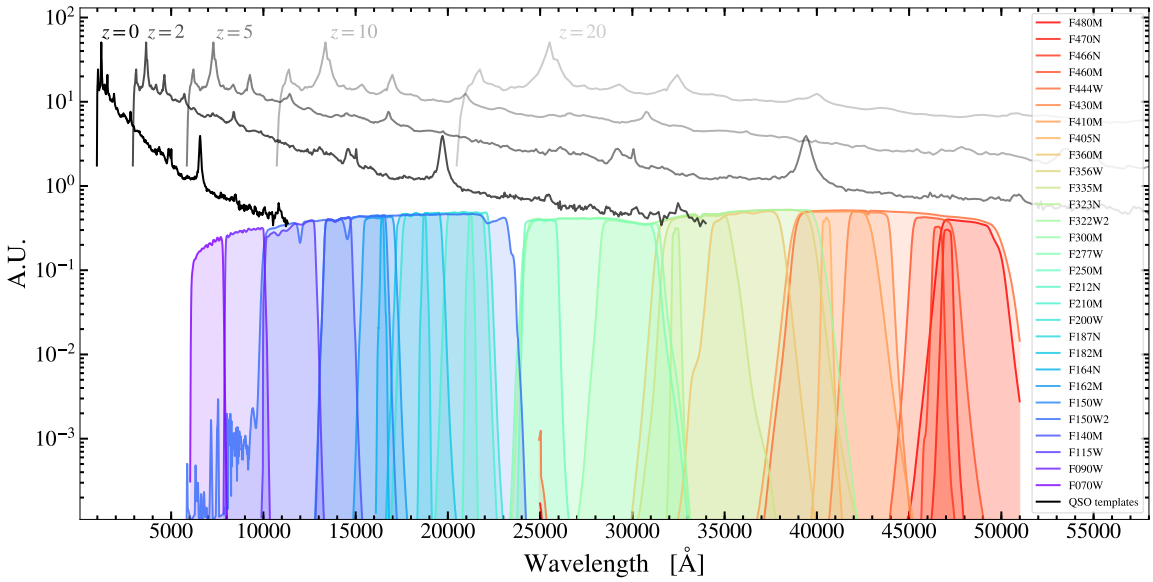

Figure 5.1: Representation of the QSO template we employ in our analysis, shifted at increasing redshifts (solid black and grey lines), compared to the JWST filter-set (shaded colored areas). The spectral template is interrupted at wavelengths shorter than the Lyα break (λ = 912Å rest-frame), therefore at $z > 5$ the bluest filters of JWST lie outside the spectral coverage of the template. This suggests that our template model should be improved in order to include the effect of the IGM onto the radiation emitted by the QSOs at rest-frame UV and X-ray wavelengths.

data. Although this toy model still presents significant room for improvements, the method developed here have the potential to draw predictions for the upcoming James Webb Space Telescope (JWST) survey (see also Natarajan et al., 2017; Volonteri et al., 2017; Vogelsberger et al., 2020; Marshall et al., 2021, for recent works pursuing similar goals).

More in detail, we associated the QSO spectral template of Selsing et al. (2016) to the accreting BHs in L-Galaxies, and complemented it with a Black-Body spectrum of fixed temperature T = 500 K, to simulate the emission of a dusty-torus at IR wavelengths. The IR intensity of the QSO template was then linked to the metallicity of the gas component in the host galaxy of the accreting-BH, assuming fixed dust-to-gas and metals-to-dust ratios, for simplicity. In this way, we obtained a model for the metallicity-dependent emission of high-z QSOs. Since heavy, intermediate and light seeds form in environments with different levels of chemical enrichment, this model has the potential of discerning the seeding channels which led to the formation of an accreting BH. To simulate their observational counterparts, we implemented the computation of synthetic photometry of the QSOs spectra in L-Galaxies through the JWST filter-system. Figure 5.1 shows the latter, together with the QSO template without the Black-Body component, shifted at increasing redshifts (solid dark and grey lines).

Although being yet very simple, this toy-model allows to simulate the observability of high-z QSOs and the ability to infer their formation channel in the upcoming JWST data. Finally, by applying machine-learning techniques, we estimated the discerning power of each JWST filter towards the discrimination of the spectral features of our simulated QSOs. By construction, the latter are linked to the local conditions of the host galaxy, and indirectly to the seeding channel by which their central SMBH was formed in our seeding model. Despite being remarkably predictive, this model is based on simplifying assumptions which limit its descriptive power. For example, we note that for $z > 5$ the Lyman-break of the QSO template starts to affect the bluest filters of JWST. Consequently, the details of the UV and X-ray emission of QSOs, as well as the modelling of IGM absorption, become important in order to predict the photometric properties of QSOs in the JWST upcoming data. To improve this aspect of our model, the UV emission of the QSO template can be linked to the properties of the host-galaxy stellar population, or a more complex treatment of dust could also be implemented (see e.g., Volonteri et al., 2017). Finally, the ISM and IGM absorption (e.g., Wyithe et al., 2005a; Laursen et al., 2009; Gurung-López et al., 2020) and the escape of X-ray/UV photons from the ionized proximity areas of QSOs (e.g., Bajtlik et al., 1988; Haiman & Rees, 2001; Wyithe & Loeb, 2004; Wyithe et al., 2005b; Wyithe & Bolton, 2011) should be taken into account to produce the final observed spectrum. These improvements would ultimately allow

to build a realistic mock for a JWST survey focused on detecting $z \gtrsim 7$ QSOs. In addition, we stress that also the environment in which these accreting SMBHs are observed can help to discern between different seeding channels. Indeed, as suggested by our results discussed in §3.4.2, DCBH and RSM seeds are expected to form in denser environment than light seeds.

Finally, we underline that the L-Galaxies SAM has been already used by Izquierdo-Villalba et al. (2019a) to produce galaxy mocks of photometric multi-band surveys, such as J-PLUS and J-PAS, therefore the same techniques can be naturally extended to this project.

5.3.3 The Lyα LF within the context of the multi-NB J-PAS survey

The content of this section makes part of an ongoing research project which is being carried out by A. Torralba, as part of a Ph.D. thesis project directed by Dr. S. Gurung Lopez and Prof. P. Arnalte (University of Valencia, Spain), with our collaboration.

The procedures and results about the photometric determination of the Lyα LF discussed in §4 find a natural application to the upcoming multi-NB data of the J-PAS survey (see §2.2.4). Indeed, the methods we used to select our Lyα-emitting candidates and obtain their LF have been purposely developed for multi-NB photometric observations. The much finer spectral sampling and the higher depth of J-PAS data, with respect to J-PLUS ones, allow to improve some of the crucial caveats of our pipeline which we addressed in §5.2. More in detail, the depth of J-PAS data allows to push the detection limits towards fainter Lyα luminosity with respect to the results described in §4. This allows to sample the luminosity range where the transition between the Lyα LF of SF LAEs and that of Lyα-emitting AGN is expected to happen (see e.g. Konno et al., 2016; Sobral et al., 2018b; Calhau et al., 2020; Spinoso et al., 2020; Zhang et al., 2021). This can help to study the relative contributions of SF processes and SMBH activity to the Lyα LF. Another technical aspect that can significantly improve our results is the possibility of building a tomographic analysis of the Lyα LF, similarly to what presented in Sobral et al. (2018b). Indeed, we performed our analysis with only 5 (non-contiguous) J-PLUS NBs, while J-PAS offers a striking total of 56 contiguous NBs which can be used to continuously sample the Lyα LF in redshift, potentially between $z \sim 2$ and $z \sim 6$. This can provide extremely useful constraints about the evolution of the Lyα LF, which can shed light onto the progress of star formation and AGN activity from the end of reionization down to $z \sim 2$. The 56 contiguous NBs of J-PAS also allow to detail the continuum emission of sources in greater detail than the two BBs of J-PLUS we used. Consequently, it is possible to abandon the three-filters method described in §2.2.5 in favour of a less-biased estimate of the continuum in the proximity of the targeted emission lines. This allows to significantly improve the selection procedures for Lyα-emitting sources, reducing the selection errors of our method (see §5.2.1) and potentially improving its completeness. Furthermore, albeit with a low spectral resolution ($R \sim 60$, see Bonoli et al., 2020), the J-PAS filter set allows to detail the broad emission lines of high-z QSOs (see e.g., Chaves-Montero et al., 2021), helping to identify them within samples of unknown Lyα-emitting candidates. This represents a significant improvement which can help to reduce the classification bias of our method (see §5.2.2) and improve the selected samples purity. Finally, the possibility of estimating the line profile can also improve the measurement of the Lya flux of broad QSO lines, therefore allowing to abandon the average correction for the filter-width effect we developed (see §4.6.1).

To conclude, the thecnical features of J-PAS (in particular its higher depth and finer spectral coverage with respect to J-PLUS) allow for a wealth of possible improvements to our method. These are currently being developed and tested by A. Torralba (Ph.D. student)[1], on the public data obtained with the pathfinder miniJ-PAS survey (see Bonoli et al., 2020). The main scientific goals of this project are to determine photometrically the Lyα LF exploiting miniJ-PAS observations, analyze the properties of the selected samples of Lyα-emitting candidates and build a stable pipeline for the analysis of the upcoming J-PAS data.

5.3.4 Joining the thesis branches: a theoretical model for the Lyα LF of AGN

A key aspect of physical models is to provide predictions of observable properties and suggest their theoretical interpretation. As discussed in e.g. §3.4.5, the mass-growth model of Izquierdo-Villalba et al.

[1] Under the supervision of Dr. S. Gurung-Lopez, Prof. P. Arnalte and our collanboration

(2020) allows to predict the bolometric luminosity of accreting SMBHs by exploiting the the theoretical picture delineated in §1.3, therefore allowing to interpret their bLF in the context of the cosmological formation and evolution of galaxies. Building on this model, it is possible to use bolometric corrections (see §1.5.1 and e.g. Marconi et al., 2004; Duras et al., 2020) to trasnform the predicted bolometric luminosity of modelled AGN into their luminosity within specific bands (see Fig. 3.16 and e.g., Hopkins et al., 2006a; Aird et al., 2015; Griffin et al., 2019). An alternative possibility to model the AGN luminosity in a given band is to associate spectral templates (Vanden Berk et al., 2001; Selsing et al., 2016; Hernán-Caballero et al., 2016) to the simulated active BHs. By adequately normalizing the templates to the L_{bol} predicted by the AGN-emission model and integrating them within a given wavelength (or frequency) range, it is then possible to estimate the observational counterparts of active SMBHs. This procedure lies at the basis of the research project we described in 5.3.2. The simulation of AGN emission can be made more realistic by adding degree of complexity to the employed spectral template. For instance, this can be attained by modelling additional components due to the host galaxy emission, by adding the presence of broad and narrow components of the emission lines and by considering the intervening action of the IGM absorption effects or the viewing angle under which the AGN is observed.

By following this method, it is possible to model the Lyα LF of AGN if an adequately descriptive model for the emission of Lyα photons from AGN sources is employed. This kind of detailed models have been developed for the Lyα emission and statistical observables of SF galaxies (see e.g., Gronke & Dijkstra, 2014; Gronke et al., 2015a; Gronke & Dijkstra, 2016; Gronke, 2017; Gurung-Lopez et al., 2018; Gurung-López et al., 2019, 2020). On the other hand, the analogous task for AGN is highly complicated by the complexity of the emitting sources of AGN (see §1.3.4 and references therein). Nevertheless, by modelling the processes concurring to the emission and escape of Lyα photons from high-z AGN it is possible, at least in principle, to build a template which can be used to predict the expected Lyα emission from modelled AGN. This procedure could be used to simulate the counterparts of SMBHs formed according to our BH-seeding model (see §3.3) and which accrete mass and shine according to the Izquierdo-Villalba et al. (2020) physical prescriptions. Furthermore, by convolving the predicted spectra of AGN with the filter-set of ongoing or upcoming photometric it is possible to theoretically predict the Lyα LF of AGN and galaxies, as expected by the observational data of these surveys.

This theoretical work has a wealth of applications. Indeed, on one side it allows to extract physical insights about the origin of the Lyα LF and its main contributions at different luminosity, (i.e. AGN or SF LAEs), allowing to determine which physical processes shape its properties evolution across cosmic epochs. For instance, as for the case of our bLF presented in §3.4.5, this allows to study the effects of BH-seeding processes onto the Lyα LF, therefore joining the two main branches of this thesis work. On the other hand, the observational determinations of the Lyα LF could be used in order to constrain the physical model for the Lyα emission of AGN. This might help to understand which small-scale physical processes drive the emission of Lyα photons from the active regions of AGN and govern their radiative transport in the IGM. Given the extremely intriguing outcomes of this research line, we plan to concentrate on it in the next future.

Chapter 6

Conclusions

This thesis presents an analysis centered on the formation and evolution of super-massive black holes (SMBHs) performed through a combination of theoretical and observational methods. On one side, we studied the effect high-redshift BH-seeding processes on the build-up of the global SMBHs population within a simulated cosmological box. In particular, we made use of the `L-Galaxies` semi-analytical model (SAM) applied to the merger-trees of the N-body, cosmological `Millennium-II` simulation (`MR-II`). On the other hand, we performed the photometric determination of the luminosity function (LF) of $2 < z < 3.3$ Active Galactic Nuclei (AGN) at rest-frame Lyman-α (Lyα) wavelengths. For the latter task, we exploited the wide-area, multi narrow-band (NB) photometric data of the Javalambre Physics of the Local Universe Survey (`J-PLUS`). Our results provide a comprehensive view of SMBH formation and evolution which help to shed light on their statistical properties, from their formation at $z \gtrsim 7$ down to $z \sim 0$, as well as on those of their observational counterparts at $z \sim 2 - 3$. In the following, we summarize our methods and results on both sides.

We embedded a set of physically-motivated prescriptions for SMBHs formation (BH-seeding) in a cosmological context, by resorting to a semi-analytical approach. This allowed us to overcome some of the most stringent computational limits for the modelling of BH-seeding over wide simulated volumes. Due to the mass-resolution limits of the `MR-II` simulation, the formation of light seeds (PopIII remnants) is modelled in a sub-grid fashion, by exploiting the outputs of the `GQd` analytical model (see Valiante et al., 2016, 2021). In addition, we modified the `L-Galaxies` SAM by modelling self-consistently the diffusion of metals and UV photons. These prescriptions are necessary to connect the formation of intermediate and heavy BH seeds to the local, physical conditions of the intergalactic medium (IGM). Finally, we included the formation of direct-collapse BHs (DCBHs) induced by gas-rich galaxy mergers (merger-induced DCBHs or miDCBH; see e.g., Mayer et al., 2010; Bonoli et al., 2014; Mayer & Bonoli, 2019). Ours is a comprehensive model which produces high-z BH-seeds over a wide mass range (i.e. from light to heavy seeds), accounting for most of the currently-envisioned scenarios (see e.g., Volonteri, 2010; Latif & Ferrara, 2016; Valiante et al., 2017; Inayoshi et al., 2020, for recent reviews).

Our results underline the strong prevalence of light seeds over the whole simulated volume of `MR-II`. This is a direct effect of the sub-grid implementation of this BH-seeding channel. Indeed, our method for *grafting* the information of `GQd` outputs into newly-initialized `L-Galaxies` structures tends to favour the presence of pure-light seeds remnants with respect to intermediate and heavy seeds (see §3.4.1 for details). On the other hand, these seed-classes form self-consistently in our model, depending on the local conditions of the IGM which we track in time. In particular, the required presence of H_2-dissociating, Lyman-Werner (LW) flux (see §1.2) implies that both intermediate and heavy seeds tend to form in over-dense regions with respect to light seeds. Indeed, the local LW illumination is usually provided by a single luminous neighbor, in accordance with the synchronized-halo-pairs scenario (see e.g., Dijkstra et al., 2008; Agarwal et al., 2013; Visbal et al., 2014b; Yue et al., 2014; Regan et al., 2017). Our results statistically confirm that these conditions for intermediate and heavy seeds formation are found more easily in overdense environments.

The epoch of BH-seeding in our model terminates at $z \sim 6$, when the local IGM conditions modelled self-consistently do not allow anymore the formation of DCBHs nor the occurrence of the runaway-stellar-mergers scenario (RSM, see e.g., Devecchi & Volonteri, 2009; Omukai et al., 2008). In this way, the low-z population of SMBHs in our model is exclusively originated by BH-seeding processes.

This approach is significantly different from the more commonly followed one, according to which each newly-resolved galaxy gets associated with a central SMBH, independently of its resolution redshift. After their formation, our SMBHs accrete mass in time according to the model presented in Izquierdo-Villalba et al. (2020). In particular, the hierarchical growth of structures and the secular processes within the simulated galaxies are the main driver of the SMBHs mass-growth. These processes gradually erase the impact of BH-seeding on the properties of the global SMBHs population. This is clearly visible in our $z\sim0$ BH mass function (BHMF), which shows little-to-none modifications at $M_\bullet \gtrsim 10^8 M_\odot$ when we act on the abundance of the largely predominant light-seeds. This suggests that the statistical properties of the global population of $z\sim0$ SMBHs in our model is the result of significant evolution during their history, which erase the impact of high-z BH-seeding processes. On the other hand, the overall efficiency of BH-seeding reverberates at the low-mass end of the $z\sim0$ BHMF, where our model predictions are significantly sensitive to the abundance of light seeds grafted at $z>8$ from `GQd` outputs. This is consistent with the picture according to which local dwarf galaxies might offer the indirect means to study the low-mass, evolved remnants of BH-seeding processes (e.g., Mezcua, 2021). Indeed, our results show that some direct effects of our BH-seeding model can survive down to low-z. For instance, DCBH and RSM descendants which never experience a merger with different seed-types are generally found in low-mass satellites at $z\sim0$. On the contrary, intermediate and heavy seed-type descendants which experienced mergers with light-seeds remnants can be found even in the most massive halos of the `MR-II` box. We conclude that, on average, the formation scenario of SMBHs is able to influence the environment in which BH seeds are formed. Later on, the hierarchical growth of structures and the internal evolution of galaxies is the main driver of the low-z properties of SMBHs populations and their host galaxies.

Within our model, heavy seeds do not appear necessary to reach the high masses inferred for the SMBHs powering $z\gtrsim6$ QSOs. This is in apparent contradiction with the results of recent works on this topic, according to which the mass contribution of intermediate or heavy seeds is necessary if super-Eddington accretion onto light-seeds is not taken into account (e.g. Valiante et al., 2016; Sassano et al., 2021; Lupi et al., 2021b). Nevertheless, our result is an effect of the specific implementation of light-seeds formation. Indeed, pure-light seeds descendants are grafted in `L-Galaxies` structures with masses that span up to $M_{seed}\sim10^6 M_\odot$, as a consequence of their initial episodes of mass-growth within `gqd`. Therefore, although these SMBHs might descend from pure-light seeds, they effectively act as intermediate or heavy seeds in our model, with initial masses of $M_{seed}\gtrsim10^4 M_\odot$. These objects can then efficiently grow up to $M_\bullet\gtrsim10^7 M_\odot$ by $z\sim6$ according to our mass-growth prescriptions. Finally, the formation of miDCBHs is extremely inefficient in our model. Indeed, although this seeding scenario might offer a viable possibility for the formation of high-z SMBHs, the conditions for its occurrence are extremely rare in the `MR-II` box, due to the dynamical range probed by this simulation.

On the observational side, we exploited the wide-area ($\sim1000\,\mathrm{deg}^2$) covered by the multi-NB photometric observations of the `J-PLUS` DR1 in order to build the Lyman-α luminosity function (Lyα LF). Due to the `J-PLUS` detection limits and the shortcomings of our method to look for emission lines, for our study we only used four out of the 7 available NBs in the `J-PLUS` filter set. This allowed us to target the Lyα emission of astronomical sources in four different redshift bins of width $\Delta z\sim0.1$, within the interval $2\lesssim z\lesssim3.3$. By targeting significant photometric excesses, our pipeline selected a total of ~14500 objects which exhibit photometric and morphological properties compatible with $z>2$ Lyα -emitting sources, as well as non-significant proper-motion measurements and no previous spectroscopic classification.

We confirmed the results of our selection pipeline with two spectroscopic follow-up programs at the GTC telescope which targeted 45 of our genuine candidates, lacking any previous spectroscopic identification. Although the number of spectroscopically followed-up candidates is a factor of ~300 times smaller than the total size of our selected sample, the analysis of our GTC-targets demonstrates that i) our pipeline is able to detect true line-excesses with a very high efficiency ($\sim90\%$) and ii) by selecting sources on the basis of their NB photometric excess, we are able to classify as $2<z<3.3$ QSOs a significant number of previously-unidentified point-like sources. On the other hand, due to the limited spectral information retrivable by `J-PLUS` photometry, our pipeline suffers from significant contamination by genuine line-emitters at different redshift with respect to the targeted one. More in detail, we found that only $\sim65\%$ of our GTC spectroscopic targets are Lyα-emitting QSOs, while the residual contamination

is mostly due to CIV-emitting QSOs at $z \sim 1.52$ ($\sim 20\%$). Blue-stars also contribute significantly to the overall contamination of our spectroscpic results ($\sim 11\%$). This is to the selection bias introduced by our method towards strong gradients between the broad-band (BB) and NB photometric measurements of sources.

We used the spectroscopic results of our follow-up programs as well as cross-matches with publicly-available datasets in order to develop statistical methods to correct for the contamination of our samples and the biases of our selection and Lyα-flux measurements methods. Thanks to these corrections, we are able to build four different determinations of the brightest-end (i.e. at $L_{Ly\alpha} \gtrsim 10^{44} erg\,s^{-1}$) of the Ly$\alpha$ LF. This allowed us to extend previous constraints by ~ 1 dex in Lyα luminosity and by ~ 2 dex in number density. This clear improvement was made possible by both the wide area and the NB nature of the J-PLUS survey.

Although the technical features of the J-PLUS filter-set do not allow to disentangle Lyα-emitting AGN from SF LAEs (see §1.4.1 for details about this definition), our results point towards the absence of the latter class within our selected samples. This allowed us to confirm that the bright-end excess of the Lyα LF can be ascribed to the intervening action of Lyα-emitting AGN populations, as suggested by previous works (see e.g., Konno et al., 2016; Sobral et al., 2018b; Matthee et al., 2017b). In particular, these works identified a deviation with respect to a Schechter-like decay at $L_{Lya} \gtrsim 10^{43.5} erg\,s^{-1}$ and described it through a power-law function. We were able to interpret this power-law as the faint-end of an additional Schechter function, extending well beyond $L_{Ly\alpha} \gtrsim 10^{44}\,erg\,s^{-1}$. Indeed, by estimating the structural parameters of this new, high-luminosity Schechter function, we computed a faint-end slope index of $\alpha = 1.35 \pm 0.84$, which is compatible with the bright-end power-law index inferred by previous works (e.g., Sobral et al., 2018b). We concluded that the overall luminosity distribution of Lyα-emitting sources can be interpreted as a double-Schechter describing two separate classes of sources: SF LAEs and AGN. We point-out that our results on the bright-end of the Lyα LF and on the AGN fractions at $L_{Ly\alpha} \gtrsim 10^{44} erg\,s^{-1}$, presented by Spinoso et al. (2020), were recently confirmed by the spectroscopic study of Zhang et al. (2021).

The results presented here shed light on the formation and evolution of SMBHs over wide, cosmological and statistically significant volumes, either simulated or targeted with astronomical observations. Our analysis have a multitude of possible applications which allow future improvements and scientific results. On one side, this can help to predict the outcomes of upcoming observational programs such as JWST, Athena and LISA. On the other side, this can provide the basis to study the effect of BH-formation and mass-growth models into the observables of AGN evolution, such as their luminosity function at different wavelengths.

Conclusiones

Esta tesis presenta un análisis centrado en la formación y evolución de los agujeros negros supermasivos (SMBHs, de sus siglas en inglés) realizado mediante la combinación de métodos teóricos y observacionales. En la parte teórica hemos explorado cómo la población global de SMBHs modelada en un volumen cosmológico se ve afectada por diferentes mecanismos de formación de agujeros negros en el Unvierso temprano. Para ello, se ha utilizado el modelo de formación de galaxias `L-Galaxies` aplicado en la simulación cosmológica Millennium-II (`MR-II`). En la parte observativa, hemos caracterizado la función de lumionsidad (LF, de sus siglas en inglés) en longitudes de onda Lyman-α (Lyα) de fuentes conocidas como AGN (de la expresión inglés Active Galactic Nuclei). Para esta tarea, utilizamos los datos fotométricos del cartografiado *Javalambre Physics of the Local Universe Survey* (J-PLUS) que utiliza un conjunto de 12 filtros de banda ancha, intermedia y estrecha para observar una amplia área del cielo en el hemisferio norte. Gracias a la sinergia entre teoría y observación, los resultados de esta tesis proporcionan una visión más global y completa a cerca de cómo los SMBHs se forman en el Universo temprano (es decir, a alto desplazamiento al rjo, o alto z) y cómo son capaces de acumular materia a lo largo de toda la historia del Universo (es decir, hasta $z \sim 0$). A continuación, resumimos nuestros métodos y resultados.

Con el objetivo de abordar la primera parte de la tesis, se modificó el código `L-Galaxies` incluyendo la mayoría de los escenarios actualmente aceptados que conducen a la formación de agujeros negros (Volonteri, 2010; Latif & Ferrara, 2016; Valiante et al., 2017; Inayoshi et al., 2020). El hecho de utilizar un código semi-analítico nos permitió superar algunos de los límites computacionales que actualmente obstaculizan la modelización, en simulaciones cosmológicas, de los procesos de formación de BHs. Estos son comunemente conocidos como *BH seeding*, del inglés, y los agujeros negros recién nacidos se les denominan *BH seeds* o, simplemente, *semillas*. Para *BH seeds* de masas intermedia y grande se conectó su formación con las condiciones locales del medio intergaláctico (IGM, de sus siglas en inglés), incluyendo de forma autoconsistente la difusión de metales y fotones ultravioleta (Lyman-Werner, LW), capaces de disociar las moléculas de H_2 y retrasar los procesos de formación estrelar (véase §1.2). Como mecanismo paralelo, se introdujo la formación de *BH seeds* de masa grande mediante fusiones de galaxias ricas en gas (Mayer et al., 2010; Bonoli et al., 2014; Mayer & Bonoli, 2019). A pesar de todas las mejoras en `L-Galaxies`, los límites en la resolución en masas de la simulación `MR-II` (en la cual se cimenta `L-Galaxies`) complican el modelado de *BH seeds* de baja masa (semillas ligeras). Para superar esta limitación se modeló su formación de manera indirecta explotando los resultados del modelo analítico `GQd` (Valiante et al., 2016, 2021). El resultado fue un modelo completo que produce *BH seeds* en edades tempranas del Universo (es decir a $z \gtrsim 6$) en un amplio rango de masas.

Nuestros resultados mostran una fuerte prevalencia de semillas ligeras en todo el volumen simulado. Este hecho se debe a que nuestro método tende a favorecer la presencia de semillas ligeras con respecto a las semillas intermedias y grandes debido a que se utiliza `GQd` como condición inicial para todas las galaxias recien formadas en `L-Galaxies` (véase §3.4.1 para más detalles). Por otro lado, las semillas itermedias y grandes, cuya formación depende de las condiciones locales del IGM, tienden a formarse en regiones sobredensas del Universo como consecuencia de la iluminación LW proporcionada por una única galaxia cercana muy luminosa, de acuerdo con lo que se conoce como "escenario de pares de halo sincronizados" (Dijkstra et al., 2008; Agarwal et al., 2013; Visbal et al., 2014b; Yue et al., 2014; Regan et al., 2017). De este modo, nuestros resultados corroboran de una forma estadísitica las hipótesis anteriores sobre la fácil ocurrencia de semillas intermedias y grandes en entornos sobredensos.

Por otra parte, nuestro modelo apunta a que la época de *BH seeding* termina a $z \sim 6$, dado que las condiciones locales del IGM hacen inviable la formación de semillas intermedias o grandes (Devecchi & Volonteri, 2009; Omukai et al., 2008). De este modo, nuestra pobalción simulada de SMBHs $az < 6$ es el resultado de procesos de *BH seeding* en etapas muy tempranas del Universo. Este resultado contrasta con la metodología comunmente seguida por otros trabajos según la cual, independiemntemente del tiempo cosmológico, cualquier galaxia es capaz de formar y albergar un SMBH. Una vez que las semillas de SMBHs estan formadas, el crecimiento jerárquico de estructuras y procesos seculares dentro de las galaxias causan su crecimiento en masa (Izquierdo-Villalba et al., 2020). De modo interesante, encontramos que

estos procesos de crecimiento son capaces de borrar gradualmente el impacto de los mecanismos de *BH seeding*. Esto es claramente visible en nuestra función de masas de SMBHS (BHMF) a $z \sim 0$ (i.e. en el Universo local), donde se encontran pocas diferencias a masas de SMBHs $\gtrsim 10^8 M_\odot$ cuando, por ejemplo, se varia la eficiencia en producción de semillas ligeras. Sim embargo, el extremo menos masivo de la BHMF es más sensible a la eficiencia con la que se incluian la semillas ligeras desde el modelo analítico `GQd` . Esto es consistente con la hipótesis de que las galaxias de baja masas (galaxias enanas) a $z \sim 0$ serían el perfecto laboratorio para estudiar los procesos de *BH seeding* relacionados con semillas de baja masa (véase por ejemplo, Mezcua, 2021). Además, nuestro modelo apunta a que semillas de masa intermedia y grande con una baja historia de fusiones podrían ser encontradas también en galaxias enanas a $z \sim 0$. Por el contrario, los descendientes de semillas intermedias y grandes que experimentaban un gran número de fusiones con descendientes de semillas ligeras podrían habitar incluso los halos más masivos de nuestro modelo. De esta forma, nuestros resultados nos permitieron sugerir que, en promedio, los procesos de BH-seeding son capaz de influir en el entorno que habitan los SMBHs. Por otro lado también pudimos concluir que el crecimiento jerárquico de las estructuras y la evolución interna de las galaxias son los principales motores a traves de los cuales la poblacion de SMBHs de $z \sim 0$ moldea sus propiedades y las de sus galaxias anfitrionas. Finalmente, nuestro modelo apunta a que las semillas masivas no parecen ser necesarias para alcanzar SMBHS de masa alta que alimentan a los cuásares (QSOs, de su sigla en inglés) más briallantes observados a $z \gtrsim 6$. Esto está en aparente contradicción con los resultados reportados en la literatura, según los cuales la contribución de semillas de masa alta e intermedia es necesaria en caso de que no se modelen eventos de crecimiento muy eficientes (*acreción super-Eddington*, véase Valiante et al., 2016; Sassano et al., 2021; Lupi et al., 2021b). Sin embargo, nuestros resultados parecen ser un efecto de la forma en que se incluyen las semillas ligeras del modelo `GQd`. En concreto todos los descendientes de semillas ligeras incluidos en `L-Galaxies` pueden abarcar masas hasta $10^6\,M_\odot$, como consecuencia de sus episodios iniciales de crecimiento de masa dentro de `GQd`. Por lo tanto, aunque estos SMBHs podrían descender de semillas ligeras, en nuestro modelo actúan como semillas intermedias o masivas, con masas iniciales de $10^4\,M_\odot$. De este modo, estos objetos pueden crecer eficientemente hasta $10^7\,M_\odot$ antes de $z \sim 6$. Finalmente, la formación de semillas grandes a través de fusiones de galaxias ricas en gas es extremadamente ineficiente en nuestro modelo. De hecho, aunque este escenario de formación podría ofrecer una posibilidad viable para obtener SMBHs a $z \gtrsim 6$, las condiciones para su ocurrencia son extremadamente raras en el volumen de `MR-II`.

La parte observacional de la tesis se centró en el cálculo de la función de luminosidad Lyman-α (Lyα LF). Para desarrollar esta tarea, aprovechamos la gran área ($\sim 1000\,\mathrm{deg}^2$) cubierta por las observaciones fotométricas multibanda del cartografiado `J-PLUS` DR1 (véase §2.2.4 y §4.2). Debido a los límites de detección del cartografiado y a las limitaciones de nuestro método, centrado en detectar excesos fotométricos significativos, sólo se utilizaron 4 de las 7 bandas estrechas disponibles en `J-PLUS` . Esto nos permitió centrarnos en la emisión de Lyα de AGN en cuatro franjas diferentes de desplazamiento al rojo entre $2 < z < 3$). Nuestra metodología nos permitió encontrar un total de ~ 14500 objetos cuyas propiedades fotométricas y morfológicas eran compatibles con fuentes emisoras de Lyα a $z > 2$. Además estos candidatos no presentaban evidencias de movimiento propio y no tenían clasificación espectroscópica previa.

Los resultados de nuestra selección fueron confirmados con dos programas espectroscópicos del telescopio GTC, que se centraron en el seguimiento de 45 fuentes sin identificación espectroscópica previa. Aunque el número de candidatos seguidos espectroscópicamente es un factor ~ 300 más pequeño que el tamaño original de nuestra muestra, los resltados del GTC demostran que nuestro metodo es capaz de detectar eficientemente líneas de emisión ($\sim 90\%$ de los casos) y seleccionar fuentes emisoras de Lyα no detectadas anteriormente cuya naturaleza era compatible con QSOs a $2 < z < 3$. A pesar de esto, debido a la limitada información espectral recuperable por la fotometría de `J-PLUS`, nuestro metodo sufre una importante contaminación por emisores a $z < 2$. Tras un detallado análisis encontramos que esta contaminación se debe a un sesgo de selección introducido por nuestro método de búsqueda que tende a seleccionar fuentes con grandes gradientes entre la fotometría en la banda ancha y estrecha. En concreto, encontramos que solo el 65% de nuestros cadidatos observados por GTC son QSOs emisores de Lyα a $z \sim 2.2$. El $\sim 20\%$ de fuentes eran QSOs a $z \sim 1.52$ con emision en `CIV` en vez de Lyα y el otro $\sim 10\%$ corresponde a estrellas azules.

Para solventar el problema de los contaminantes se utilizó los resultados espectroscópicos del GTC y el cotejo con otro datos disponibles en la literatura. Esto nos permitió desarrollar métodos estadísticos capaces de corregir tanto problemas de contaminación como los sesgos introducido por nuestro métodos de selección y análisis. Gracias a estas correcciones, fuimos capaces de construir cuatro Lyα LF capaces de caracterizar con precisión los QSOs más brillantes (es decir, a $L_{Ly\alpha} \gtrsim 10^{44} erg\,s^{-1}$). Esto permitió ampliar hasta $\sim$1 dex las medidas anteriores en la luminosidad Lyα y extender hasta $\sim$2 dex las medidas de densidad numérica, con respeto a risultados presentados anteriormente en la literatura. Enfatizamos que estas claras mejoras fueron posibles tanto por la amplia área como por las observaciones en filtros estrechos de `J-PLUS`.

Pese a los buenos resultados descritos anteriormente, las características de los filtros de `J-PLUS` no nos permitieron separar los QSOs emisores en Lyα de galaxias con intensa formación estelar (LAEs, véase §1.4.1). A pesar de ello, Nuestros resultados apuntan a la ausencia de estas últimas dentro de la muestra seleccionada. Esto corrobora trabajos anteriores que indican que la parte brillante de la funcion de luminosidad Lyα ($L_{Ly\alpha} > 10^{43.5} erg\,s^{-1}$) puede atribuirse inequívocamente a la población de QSOs (véase también Konno et al., 2016; Sobral et al., 2018b; Matthee et al., 2017b). En particular, los trabajos anteriormente publicados en la literatura apuntan que en la LF Lyα existe de una desviación con respecto a un decaimiento exponencial a $L_{Lya} \gtrsim 10^{43,5} erg\,s^{-1}$, más compatible con una ley de potencia. Gracias a nuestros resultados, fuimos capaces de interpretar y parametrizar esta ley de potencia como el extremo débil de una función Schechter adicional que se extende más allá de $L_{Ly\alpha} \gtrsim 10^{44} erg\,s^{-1}$. De hecho, los parametros que encontramos para el estremo débil de nuestra función Schechter son compatibles con trabajos anteriores (véase por ejemplo, Sobral et al., 2018b). Todos estos resultados nos llevaron a la conclusión de que la Lyα LF observada podría interpretarse como una doble función de Schechter que describe dos clases distintas de fuentes: LAEs y QSOs. Cabe destacar que nuestros resultados sobre el extremo brillante de la Lyα LF y sobre las fracciones de QSOs en $L_{Ly\alpha} \gtrsim 10^{44} erg\,s^{-1}$ (publicados en Spinoso et al., 2020), fueron confirmados recientemente por el estudio espectroscópico de Zhang et al. (2021).

Como conclusión general, los resultados presentados con esta tesis han arrojado luz sobre la formación y evolución de SMBHs en volúmenes amplios, cosmológicos y estadísticamente significativos. Sin embargo, el producto científico de esta tesis no se queda solamente en los resultados presentados aquí sino que tiene una multitud de posibles aplicaciones que permitirán obtener resultados científicos adicionales y significativos. Por un lado, la parte teórica permitirá ayudar a interpretar los resultados de los próxoimos observatorios espaciales como JWST, Athena y LISA. Por otro lado, la combinación de teoría y observaciones será la base para estudiar el efecto de los modelos de formación y crecimiento de SMBHs en la evolución de la función de luminosidad de QSO a diferentes longitudes de onda.

Bibliography

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, Phys. Rev. Lett., 116, 061102
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, Phys. Rev. Lett., 119, 161101
Abel, T., Bryan, G. L., & Norman, M. L. 2000, ApJ, 540, 39
Abel, T. & Haiman, Z. 2000, in Molecular Hydrogen in Space, ed. F. Combes & G. Pineau Des Forets, 237
Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, ApJS, 235, 42
Abramowicz, M. A. & Fragile, P. C. 2013, Living Reviews in Relativity, 16, 1
Adams, J. J., Blanc, G. A., Hill, G. J., et al. 2011, ApJS, 192, 5
Agarwal, B., Cullen, F., Khochfar, S., Ceverino, D., & Klessen, R. S. 2019, MNRAS, 488, 3268
Agarwal, B., Dalla Vecchia, C., Johnson, J. L., Khochfar, S., & Paardekooper, J.-P. 2014, MNRAS, 443, 648
Agarwal, B., Davis, A. J., Khochfar, S., Natarajan, P., & Dunlop, J. S. 2013, MNRAS, 432, 3438
Agarwal, B., Khochfar, S., Johnson, J. L., et al. 2012, MNRAS, 425, 2854
Agarwal, B., Regan, J., Klessen, R. S., Downes, T. P., & Zackrisson, E. 2017, MNRAS, 470, 4034
Agarwal, B., Smith, B., Glover, S., Natarajan, P., & Khochfar, S. 2016, MNRAS, 459, 4209
Aguado, D. S., Ahumada, R., Almeida, A., et al. 2019, ApJS, 240, 23
Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, ApJS, 203, 21
Ahn, K., Shapiro, P. R., Iliev, I. T., Mellema, G., & Pen, U.-L. 2009, ApJ, 695, 1430
Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, ApJS, 249, 3
Aird, J., Coil, A. L., & Georgakakis, A. 2018, MNRAS, 474, 1225
Aird, J., Coil, A. L., Georgakakis, A., et al. 2015, MNRAS, 451, 1892
Alexander, D. M. & Hickox, R. C. 2012, New Astron. Rev., 56, 93
Amorín, R., Fontana, A., Pérez-Montero, E., et al. 2017, Nature Astronomy, 1, 0052
Anglés-Alcázar, D., Davé, R., Faucher-Giguère, C.-A., Özel, F., & Hopkins, P. F. 2017a, MNRAS, 464, 2840
Anglés-Alcázar, D., Faucher-Giguère, C.-A., Quataert, E., et al. 2017b, MNRAS, 472, L109
Anglés-Alcázar, D., Quataert, E., Hopkins, P. F., et al. 2021, ApJ, 917, 53
Antonucci, R. 1993, ARA&A, 31, 473
Ao, Y., Matsuda, Y., Beelen, A., et al. 2015, A&A, 581, A132
Ardaneh, K., Luo, Y., Shlosman, I., et al. 2018, MNRAS, 479, 2277
Arrabal Haro, P., Rodríguez Espinosa, J. M., Muñoz-Tuñón, C., et al. 2020, MNRAS, 495, 1807
Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481
Audibert, A., Combes, F., García-Burillo, S., et al. 2019, A&A, 632, A33
Bañados, E., Mazzucchelli, C., Momjian, E., et al. 2021, ApJ, 909, 80
Bañados, E., Venemans, B. P., Decarli, R., et al. 2016, ApJS, 227, 11
Bañados, E., Venemans, B. P., Mazzucchelli, C., et al. 2018, Nature, 553, 473
Babbedge, T. S. R., Rowan-Robinson, M., Vaccari, M., et al. 2006, MNRAS, 370, 1159
Bacon, R., Brinchmann, J., Richard, J., et al. 2015, A&A, 575, A75
Bajtlik, S., Duncan, R. C., & Ostriker, J. P. 1988, ApJ, 327, 570
Barkana, R. & Loeb, A. 2001, Phys. Rep., 349, 125
Baron, D., Netzer, H., Prochaska, J. X., et al. 2018, MNRAS, 480, 3993
Barrow, K. S. S., Wise, J. H., Norman, M. L., O'Shea, B. W., & Xu, H. 2017, MNRAS, 469, 4863
Baugh, C. M. 2013, in The Intriguing Life of Massive Galaxies, ed. D. Thomas, A. Pasquali, & I. Ferreras, Vol. 295, 191–199
Bayliss, K. D., McMahon, R. G., Venemans, B. P., Ryan-Weber, E. V., & Lewis, J. R. 2011, MNRAS, 413, 2883
Bean, R. & Magueijo, J. 2002, Phys. Rev. D, 66, 063505
Becerra, F., Marinacci, F., Bromm, V., & Hernquist, L. E. 2018, MNRAS, 480, 5029
Becker, G. D., Bolton, J. S., & Lidz, A. 2015a, Publ. Astron. Soc. Australia, 32, e045
Becker, G. D., Bolton, J. S., Madau, P., et al. 2015b, MNRAS, 447, 3402
Beckmann, V. & Shrader, C. 2012a, in Proceedings of "An INTEGRAL view of the high-energy sky (the first 10 years)" - 9th INTEGRAL Workshop and celebration of the 10th anniversary of the launch (INTEGRAL 2012). 15-19 October 2012. Bibliotheque Nationale de France, 69
Beckmann, V. & Shrader, C. R. 2012b, Active Galactic Nuclei (Wiley-VCH)
Benitez, N., Dupke, R., Moles, M., et al. 2014, arXiv e-prints, arXiv:1403.5237
Bennett, C. L., Banday, A. J., Gorski, K. M., et al. 1996, ApJ, 464, L1
Bernal, J. L., Raccanelli, A., Verde, L., & Silk, J. 2018, J. Cosmology Astropart. Phys., 2018, 017
Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
Bertone, S., De Lucia, G., & Thomas, P. A. 2007, MNRAS, 379, 1143
Bertone, S., Stoehr, F., & White, S. D. M. 2005, MNRAS, 359, 1201
Best, P., Smail, I., Sobral, D., et al. 2013, in Thirty Years of Astronomical Discovery with UKIRT, Vol. 37, 235
Bhowmick, A. K., Blecha, L., Torrey, P., et al. 2021, MNRAS, 507, 2012
Bian, F., Fan, X., Jiang, L., et al. 2012, ApJ, 757, 139
Blanc, G. A., Adams, J. J., Gebhardt, K., et al. 2011, ApJ, 736, 31
Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, AJ, 154, 28

Bodenheimer, P. H. 2011, Principles of Star Formation (Springer)
Boller, T., Freyberg, M. J., Trümper, J., et al. 2016, A&A, 588, A103
Bolton, C. T. 1972, Nature Physical Science, 240, 124
Bolton, J. S., Haehnelt, M. G., Warren, S. J., et al. 2011, MNRAS, 416, L70
Bonchi, A., La Franca, F., Melini, G., Bongiorno, A., & Fiore, F. 2013, MNRAS, 429, 1970
Bond, N. A., Feldmeier, J. J., Matković, A., et al. 2010, ApJ, 716, L200
Bond, N. A., Gawiser, E., Gronwall, C., et al. 2009, ApJ, 705, 639
Bond, N. A., Gawiser, E., Guaita, L., et al. 2012, ApJ, 753, 95
Bondi, H. 1952, MNRAS, 112, 195
Bongiorno, A., Mignoli, M., Zamorani, G., et al. 2010, A&A, 510, A56
Bonoli, S., Marín-Franch, A., Varela, J., et al. 2020, arXiv e-prints, arXiv:2007.01910
Bonoli, S., Marulli, F., Springel, V., et al. 2009, MNRAS, 396, 423
Bonoli, S., Mayer, L., & Callegari, S. 2014, MNRAS, 437, 1576
Bonoli, S., Mayer, L., Kazantzidis, S., et al. 2016, MNRAS, 459, 2603
Borisova, E., Cantalupo, S., Lilly, S. J., et al. 2016a, ApJ, 831, 39
Borisova, E., Lilly, S. J., Cantalupo, S., et al. 2016b, ApJ, 830, 120
Bosma, A. 1978, PhD thesis, -
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2017a, ApJ, 843, 41
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015a, ApJ, 811, 140
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2017b, arXiv e-prints, arXiv:1711.02090
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015b, ApJ, 803, 34
Bouwens, R. J., Smit, R., Labbé, I., et al. 2016, ApJ, 831, 176
Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, MNRAS, 370, 645
Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, MNRAS, 398, 1150
Boyle, B. J., Shanks, T., & Peterson, B. A. 1988, MNRAS, 235, 935
Boyle, B. J. & Terlevich, R. J. 1998, MNRAS, 293, L49
Bridge, C. R., Blain, A., Borys, C. J. K., et al. 2013, ApJ, 769, 91
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Bromm, V. 2013, Asociacion Argentina de Astronomia La Plata Argentina Book Series, 4, 3
Bromm, V., Ferrara, A., Coppi, P. S., & Larson, R. B. 2001, MNRAS, 328, 969
Bromm, V. & Larson, R. B. 2004, ARA&A, 42, 79
Bromm, V. & Loeb, A. 2003, ApJ, 596, 34
Bryan, G. L. & Norman, M. L. 1998, ApJ, 495, 80
Bădescu, T., Yang, Y., Bertoldi, F., et al. 2017, ApJ, 845, 172
Buchner, J. 2021, arXiv e-prints, arXiv:2101.09675
Buchner, J., Brightman, M., Baloković, M., et al. 2021, A&A, 651, A58
Buchner, J., Georgakakis, A., Nandra, K., et al. 2015, ApJ, 802, 89
Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, ApJ, 798, 7
Bunker, A. J., Warren, S. J., Hewett, P. C., & Clements, D. L. 1995, MNRAS, 273, 513
Butkevich, A. G., Berdyugin, A. V., & Teerikorpi, P. 2005, MNRAS, 362, 321
Cai, Z., Cantalupo, S., Prochaska, J. X., et al. 2019, ApJS, 245, 23
Cai, Z., Fan, X., Bian, F., et al. 2017a, ApJ, 839, 131
Cai, Z., Fan, X., Peirani, S., et al. 2016, ApJ, 833, 135
Cai, Z., Fan, X., Yang, Y., et al. 2017b, ApJ, 837, 71
Cai, Z., Hamden, E., Matuszewski, M., et al. 2018, ApJ, 861, L3
Calderone, G., Nicastro, L., Ghisellini, G., et al. 2017, MNRAS, 472, 4051
Calhau, J., Sobral, D., Santos, S., et al. 2020, MNRAS, 493, 3341
Calhau, J., Sobral, D., Stroe, A., et al. 2017, MNRAS, 464, 303
Cann, J. M., Satyapal, S., Rothberg, B., et al. 2021, ApJ, 912, L2
Cantalupo, S., Lilly, S. J., & Haehnelt, M. G. 2012, MNRAS, 425, 1992
Cantalupo, S., Pezzulli, G., Lilly, S. J., et al. 2019, MNRAS, 483, 5188
Capak, P., Aussel, H., Ajiki, M., et al. 2007, ApJS, 172, 99
Capelo, P. R. & Dotti, M. 2017, MNRAS, 465, 2643
Capelo, P. R., Volonteri, M., Dotti, M., et al. 2015, MNRAS, 447, 2123
Capuzzo-Dolcetta, R. & Tosta e Melo, I. 2017, MNRAS, 472, 4013
Carignan, C. & Freeman, K. C. 1985, ApJ, 294, 494
Cassata, P., Le Fèvre, O., Garilli, B., et al. 2011, A&A, 525, A143
Cassata, P., Tasca, L. A. M., Le Fèvre, O., et al. 2015, A&A, 573, A24
Cenarro, A. J., Moles, M., Cristóbal-Hornillos, D., et al. 2019, A&A, 622, A176
Cenarro, A. J., Moles, M., Marín-Franch, A., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9149, Proc. SPIE, 91491I
Charlot, S. & Fall, S. M. 1993, ApJ, 415, 580
Chaves-Montero, J., Bonoli, S., Trakhtenbrot, B., et al. 2021, arXiv e-prints, arXiv:2111.01180
Chon, S., Hirano, S., Hosokawa, T., & Yoshida, N. 2016, ApJ, 832, 134
Chon, S., Hosokawa, T., & Yoshida, N. 2018, MNRAS, 475, 4104
Chon, S. & Omukai, K. 2020, MNRAS, 494, 2851
Christensen, L., Laursen, P., Richard, J., et al. 2012, MNRAS, 427, 1973
Ciardullo, R., Gronwall, C., Wolf, C., et al. 2012, ApJ, 744, 110
Clark, P. C., Glover, S. C. O., & Klessen, R. S. 2008, ApJ, 672, 757
Clément, B., Cuby, J. G., Courbin, F., et al. 2012, A&A, 538, A66
Clerc, N., Merloni, A., Zhang, Y.-Y., et al. 2016, MNRAS, 463, 4490
Clowe, D., Gonzalez, A., & Markevitch, M. 2004, ApJ, 604, 596
Cole, S., Lacey, C. G., Baugh, C. M., & Frenk, C. S. 2000, MNRAS, 319, 168

Collins, G. W., I. 1978, The virial theorem in stellar astrophysics (Pachart Publishing House)
Cora, S. A., Vega-Martínez, C. A., Hough, T., et al. 2018, MNRAS, 479, 2
Corbelli, E., Salpeter, E. E., & Dickey, J. M. 1991, ApJ, 370, 49
Croft, R. A. C., Miralda-Escudé, J., Zheng, Z., Blomqvist, M., & Pieri, M. 2018, MNRAS, 481, 1320
Croom, S. M., Richards, G. T., Shanks, T., et al. 2009, MNRAS, 399, 1755
Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11
Croton, D. J., Stevens, A. R. H., Tonini, C., et al. 2016, ApJS, 222, 22
Dale, J. E. 2015, New Astron. Rev., 68, 1
Das, A., Schleicher, D. R. G., Leigh, N. W. C., & Boekholt, T. C. N. 2021, MNRAS, 503, 1051
Davies, F. B., Hennawi, J. F., & Eilers, A.-C. 2019, ApJ, 884, L19
Davis, A. J., D'Aloisio, A., & Natarajan, P. 2011, MNRAS, 416, 242
Davis, B. L., Graham, A. W., & Cameron, E. 2018, ApJ, 869, 113
Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44
Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10
De Luca, V., Franciolini, G., Pani, P., & Riotto, A. 2020, J. Cosmology Astropart. Phys., 2020, 044
De Lucia, G., Tornatore, L., Frenk, C. S., et al. 2014, MNRAS, 445, 970
Degraf, C., Di Matteo, T., & Springel, V. 2010, MNRAS, 402, 1927
DeGraf, C. & Sijacki, D. 2019, arXiv e-prints, arXiv:1906.11271
Delvecchio, I., Daddi, E., Aird, J., et al. 2020, ApJ, 892, 17
Devecchi, B. & Volonteri, M. 2009, ApJ, 694, 302
Devecchi, B., Volonteri, M., Rossi, E. M., Colpi, M., & Portegies Zwart, S. 2012, MNRAS, 421, 1465
Di Matteo, T., Croft, R. A. C., Feng, Y., Waters, D., & Wilkins, S. 2017, MNRAS, 467, 4243
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604
Dijkstra, M. 2016, Constraining Reionization with Lyα Emitting Galaxies, Vol. 423 (Mesinger, Andrei), 145
Dijkstra, M. 2017, arXiv e-prints, arXiv:1704.03416
Dijkstra, M., Ferrara, A., & Mesinger, A. 2014a, MNRAS, 442, 2036
Dijkstra, M., Haiman, Z., Mesinger, A., & Wyithe, J. S. B. 2008, MNRAS, 391, 1961
Dijkstra, M., Wyithe, S., Haiman, Z., Mesinger, A., & Pentericci, L. 2014b, MNRAS, 440, 3309
Doi, M., Tanaka, M., Fukugita, M., et al. 2010, AJ, 139, 1628
Dotti, M., Colpi, M., Pallini, S., Perego, A., & Volonteri, M. 2013, ApJ, 762, 68
Drake, A. B., Garel, T., Wisotzki, L., et al. 2017, A&A, 608, A6
Dubois, Y., Pichon, C., Welker, C., et al. 2014a, MNRAS, 444, 1453
Dubois, Y., Volonteri, M., Silk, J., Devriendt, J., & Slyz, A. 2014b, MNRAS, 440, 2333
Dunn, G., Bellovary, J., Holley-Bockelmann, K., Christensen, C., & Quinn, T. 2018, ApJ, 861, 39
Duras, F., Bongiorno, A., Ricci, F., et al. 2020, A&A, 636, A73
Dwelly, T., Salvato, M., Merloni, A., et al. 2017, MNRAS, 469, 1065
Dyer, J. C., Dawson, K. S., du Mas des Bourboux, H., et al. 2019, ApJ, 880, 78
Ebisuzaki, T., Makino, J., Tsuru, T. G., et al. 2001, ApJ, 562, L19
Eddington, A. S. 1913, MNRAS, 73, 359
Eddington, A. S. 1917, MNRAS, 77, 596
Eilers, A.-C., Hennawi, J. F., & Davies, F. B. 2018, ApJ, 867, 30
Einstein, A. 1905, Annalen der Physik, 322, 891
Einstein, A. 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin, 844
Einstein, A. & de Sitter, W. 1932, Proceedings of the National Academy of Science, 18, 213
Ellison, S. L., Patton, D. R., Mendel, J. T., & Scudder, J. M. 2011, MNRAS, 418, 2043
Erb, D. K., Steidel, C. C., & Chen, Y. 2018, ApJ, 862, L10
Erwin, P. & Gadotti, D. A. 2012, Advances in Astronomy, 2012, 946368
Escala, A. 2021, ApJ, 908, 57
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019, ApJ, 875, L1
Faber, S. M. & Jackson, R. E. 1976, ApJ, 204, 668
Fabian, A. C. 2012, ARA&A, 50, 455
Fall, S. M. & Efstathiou, G. 1980, MNRAS, 193, 189
Fan, X., Carilli, C. L., & Keating, B. 2006a, ARA&A, 44, 415
Fan, X., Strauss, M. A., Richards, G. T., et al. 2006b, AJ, 131, 1203
Fan, X., Wang, F., Yang, J., et al. 2019, ApJ, 870, L11
Fanali, R., Dotti, M., Fiacconi, D., & Haardt, F. 2015, MNRAS, 454, 3641
Fanidakis, N., Baugh, C. M., Benson, A. J., et al. 2011, MNRAS, 410, 53
Feldmeier, J. J., Hagen, A., Ciardullo, R., et al. 2013, ApJ, 776, 75
Fernandez, R., Bryan, G. L., Haiman, Z., & Li, M. 2014, MNRAS, 439, 3798
Ferrarese, L. & Merritt, D. 2000, ApJ, 539, L9
Fielding, D., Quataert, E., & Martizzi, D. 2018, MNRAS, 481, 3325
Fielding, D., Quataert, E., Martizzi, D., & Faucher-Giguère, C.-A. 2017, MNRAS, 470, L39
Finkelstein, D. 1958, Physical Review, 110, 965
Finkelstein, S. L., Cohen, S. H., Windhorst, R. A., et al. 2011, ApJ, 735, 5
Fiore, F., Feruglio, C., Shankar, F., et al. 2017, A&A, 601, A143
Fontanot, F., Cristiani, S., Monaco, P., et al. 2007, A&A, 461, 39
Fontanot, F., De Lucia, G., Monaco, P., Somerville, R. S., & Santini, P. 2009, MNRAS, 397, 1776
Frank, J., King, A., & Raine, D. J. 2002, Accretion Power in Astrophysics: Third Edition, 3rd edn. (Cambridge University Press)
Freedman, W. L. 2017, Nature Astronomy, 1, 0121
Fu, H., Yan, L., Scoville, N. Z., et al. 2010, ApJ, 722, 653
Fujita, S. S., Ajiki, M., Shioya, Y., et al. 2003, AJ, 125, 13
Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748

Fumagalli, M., Mackenzie, R., Trayford, J., et al. 2017, MNRAS, 471, 3686
Furusawa, H., Kosugi, G., Akiyama, M., et al. 2008, ApJS, 176, 1
Fynbo, J. P. U., Ledoux, C., Møller, P., Thomsen, B., & Burud, I. 2003, A&A, 407, 147
Fynbo, J. U., Møller, P., & Thomsen, B. 2001, A&A, 374, 443
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gawiser, E., Francke, H., Lai, K., et al. 2007, ApJ, 671, 278
Gawiser, E., van Dokkum, P. G., Gronwall, C., et al. 2006, ApJ, 642, L13
Geach, J. E., Smail, I., Best, P. N., et al. 2008, MNRAS, 388, 1473
Gebhardt, K., Bender, R., Bower, G., et al. 2000, ApJ, 539, L13
Geller, M. J., Diaferio, A., Kurtz, M. J., Dell'Antonio, I. P., & Fabricant, D. G. 2012, AJ, 143, 102
Ghez, A. M., Salim, S., Weinberg, N. N., et al. 2008, ApJ, 689, 1044
Ghisellini, G. 2013, Radiative Processes in High Energy Astrophysics, Vol. 873 (OAB)
Gil de Paz, A., Boissier, S., Madore, B. F., et al. 2009, VizieR Online Data Catalog, J/ApJS/173/185
González Delgado, R. M., Cid Fernandes, R., Pérez, E., et al. 2016, A&A, 590, A44
González Delgado, R. M., Díaz-García, L. A., de Amorim, A., et al. 2021, A&A, 649, A79
Greene, J. E. 2012, Nature Communications, 3, 1304
Greif, T. H. & Bromm, V. 2006, MNRAS, 373, 128
Greif, T. H., Bromm, V., Clark, P. C., et al. 2012, MNRAS, 424, 399
Greif, T. H., Glover, S. C. O., Bromm, V., & Klessen, R. S. 2009a, MNRAS, 392, 1381
Greif, T. H., Johnson, J. L., Klessen, R. S., & Bromm, V. 2008, MNRAS, 387, 1021
Greif, T. H., Johnson, J. L., Klessen, R. S., & Bromm, V. 2009b, MNRAS, 399, 639
Greif, T. H., Springel, V., & Bromm, V. 2013, MNRAS, 434, 3408
Greif, T. H., Springel, V., White, S. D. M., et al. 2011, ApJ, 737, 75
Greig, B. & Mesinger, A. 2017, MNRAS, 465, 4838
Griffin, A. J., Lacey, C. G., Gonzalez-Perez, V., et al. 2019, MNRAS, 487, 198
Gronke, M. 2017, A&A, 608, A139
Gronke, M., Bull, P., & Dijkstra, M. 2015a, ApJ, 812, 123
Gronke, M. & Dijkstra, M. 2014, MNRAS, 444, 1095
Gronke, M. & Dijkstra, M. 2016, ApJ, 826, 14
Gronke, M., Dijkstra, M., McCourt, M., & Oh, S. P. 2016, ApJ, 833, L26
Gronke, M., Dijkstra, M., Trenti, M., & Wyithe, S. 2015b, MNRAS, 449, 1284
Gronwall, C., Ciardullo, R., Hickey, T., et al. 2007, ApJ, 667, 79
Guaita, L., Acquaviva, V., Padilla, N., et al. 2011, ApJ, 733, 114
Guaita, L., Gawiser, E., Padilla, N., et al. 2010, ApJ, 714, 255
Guaita, L., Melinder, J., Hayes, M., et al. 2015, A&A, 576, A51
Gültekin, K., Cackett, E. M., Miller, J. M., et al. 2009a, ApJ, 706, 404
Gültekin, K., Richstone, D. O., Gebhardt, K., et al. 2009b, ApJ, 698, 198
Gunawardhana, M. L. P., Hopkins, A. M., Bland-Hawthorn, J., et al. 2013, MNRAS, 433, 2764
Gunawardhana, M. L. P., Hopkins, A. M., Taylor, E. N., et al. 2015, MNRAS, 447, 875
Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040
Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, MNRAS, 413, 101
Gurung-Lopez, S., Orsi, A. A., & Bonoli, S. 2018, arXiv e-prints, arXiv:1811.09630
Gurung-López, S., Orsi, Á. A., Bonoli, S., Baugh, C. M., & Lacey, C. G. 2019, MNRAS, 486, 1882
Gurung-López, S., Orsi, Á. A., Bonoli, S., et al. 2020, MNRAS, 491, 3266
Haardt, F. & Madau, P. 1996, ApJ, 461, 20
Haardt, F. & Madau, P. 2012, ApJ, 746, 125
Haas, M., Müller, S. A. H., Bertoldi, F., et al. 2004, A&A, 424, 531
Habouzit, M., Volonteri, M., & Dubois, Y. 2017, MNRAS, 468, 3935
Habouzit, M., Volonteri, M., Latif, M., Dubois, Y., & Peirani, S. 2016, MNRAS, 463, 529
Habouzit, M., Volonteri, M., Somerville, R. S., et al. 2019, MNRAS, 489, 1206
Haemmerlé, L., Klessen, R. S., Mayer, L., & Zwick, L. 2021, A&A, 652, L7
Haemmerlé, L., Mayer, L., Klessen, R. S., et al. 2020, Space Sci. Rev., 216, 48
Haemmerlé, L., Meynet, G., Mayer, L., et al. 2019, A&A, 632, L2
Haemmerlé, L., Woods, T. E., Klessen, R. S., Heger, A., & Whalen, D. J. 2018, MNRAS, 474, 2757
Haiman, Z. & Rees, M. J. 2001, ApJ, 556, 87
Hainline, K. N., Shapley, A. E., Greene, J. E., & Steidel, C. C. 2011, ApJ, 733, 31
Hamilton, A. J. S. & Tegmark, M. 2004, Monthly Notices of the Royal Astronomical Society, 349, 115–128
Han, Y., Dai, B., Wang, B., Zhang, F., & Han, Z. 2012, MNRAS, 423, 464
Hao, C.-N., Huang, J.-S., Xia, X., et al. 2018, ApJ, 864, 145
Hao, L., Strauss, M. A., Fan, X., et al. 2005, AJ, 129, 1795
Hartwig, T., Bromm, V., & Loeb, A. 2018, MNRAS, 479, 2202
Hartwig, T., Clark, P. C., Glover, S. C. O., Klessen, R. S., & Sasaki, M. 2015a, ApJ, 799, 114
Hartwig, T., Glover, S. C. O., Klessen, R. S., Latif, M. A., & Volonteri, M. 2015b, MNRAS, 452, 1233
Hartwig, T., Latif, M. A., Magg, M., et al. 2016, MNRAS, 462, 2184
Hashimoto, T., Garel, T., Guiderdoni, B., et al. 2017, A&A, 608, A10
Hasinger, G., Cappelluti, N., Brunner, H., et al. 2007, ApJS, 172, 29
Hasinger, G., Miyaji, T., & Schmidt, M. 2005, A&A, 441, 417
Hayes, M. 2015, Publ. Astron. Soc. Australia, 32, e027
Hayes, M., Schaerer, D., & Östlin, G. 2010, A&A, 509, L5
Hayward, C. C. & Hopkins, P. F. 2017, MNRAS, 465, 1682
Heckman, T. M. & Best, P. N. 2014, ARA&A, 52, 589
Heckman, T. M., Kauffmann, G., Brinchmann, J., et al. 2004, ApJ, 613, 109
Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. 2015, MNRAS, 451, 2663

Herenz, E. C., Wisotzki, L., Saust, R., et al. 2019, A&A, 621, A107
Hernán-Caballero, A., Hatziminaoglou, E., Alonso-Herrero, A., & Mateos, S. 2016, MNRAS, 463, 2064
Hernán-Caballero, A., Pérez-González, P. G., Diego, J. M., et al. 2017, ApJ, 849, 82
Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., & Collins, R. A. 1968, Nature, 217, 709
Hinshaw, G., Weiland, J. L., Hill, R. S., et al. 2009, ApJS, 180, 225
Hirschmann, M., Somerville, R. S., Naab, T., & Burkert, A. 2012, MNRAS, 426, 237
Hook, I. M., McMahon, R. G., Boyle, B. J., & Irwin, M. J. 1994, MNRAS, 268, 305
Hopkins, P. F. & Hernquist, L. 2006, ApJS, 166, 1
Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006a, ApJ, 639, 700
Hopkins, P. F., Narayan, R., & Hernquist, L. 2006b, ApJ, 643, 641
Hopkins, P. F. & Quataert, E. 2010, MNRAS, 407, 1529
Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, ApJ, 654, 731
Hoyle, F. & Lyttleton, R. A. 1941, MNRAS, 101, 227
Hu, E. M., Cowie, L. L., & McMahon, R. G. 1998, ApJ, 502, L99
Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168
Huchra, J., Gorenstein, M., Kent, S., et al. 1985, AJ, 90, 691
Hutchinson, T. A., Bolton, A. S., Dawson, K. S., et al. 2016, AJ, 152, 205
Inayoshi, K., Haiman, Z., & Ostriker, J. P. 2016, MNRAS, 459, 3738
Inayoshi, K., Ichikawa, K., Ostriker, J. P., & Kuiper, R. 2019, MNRAS, 486, 5377
Inayoshi, K., Li, M., & Haiman, Z. 2018, MNRAS, 479, 4017
Inayoshi, K., Visbal, E., & Haiman, Z. 2020, ARA&A, 58, 27
Inayoshi, K., Visbal, E., & Kashiyama, K. 2015, MNRAS, 453, 1692
Inomata, K., McDonough, E., & Hu, W. 2021, arXiv e-prints, arXiv:2104.03972
Irwin, M. J., Webster, R. L., Hewett, P. C., Corrigan, R. T., & Jedrzejewski, R. I. 1989, AJ, 98, 1989
Ivezić, Ž., Brandt, W. N., Fan, X., et al. 2014, in IAU Symposium, Vol. 304, Multiwavelength AGN Surveys and Studies, ed. A. M. Mickaelian & D. B. Sanders, 11–17
Izquierdo-Villalba, D., Angulo, R. E., Orsi, A., et al. 2019a, A&A, 631, A82
Izquierdo-Villalba, D., Bonoli, S., Dotti, M., et al. 2020, MNRAS, 495, 4681
Izquierdo-Villalba, D., Bonoli, S., Spinoso, D., et al. 2019b, MNRAS, 488, 609
Jeans, J. H. 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1
Jeans, J. H. 1928, Astronomy and cosmogony (Cambridge University Press)
Jung, M., Illenseer, T. F., & Duschl, W. J. 2018, A&A, 614, A105
Kashikawa, N., Nagao, T., Toshikawa, J., et al. 2012, ApJ, 761, 85
Kashikawa, N., Shimasaku, K., Malkan, M. A., et al. 2006, ApJ, 648, 7
Katz, H., Sijacki, D., & Haehnelt, M. G. 2015, MNRAS, 451, 2352
Kauffmann, G. & Haehnelt, M. 2000, MNRAS, 311, 576
Kelly, B. C., Vestergaard, M., & Fan, X. 2009, ApJ, 692, 1388
Kennicutt, Robert C., J. 1998, ARA&A, 36, 189
Khostovan, A. A., Sobral, D., Mobasher, B., et al. 2019, MNRAS, 489, 555
Kikuta, S., Imanishi, M., Matsuoka, Y., et al. 2017, ApJ, 841, 128
Kim, C.-G., Ostriker, E. C., & Raileanu, R. 2017a, ApJ, 834, 25
Kim, J.-G., Kim, W.-T., Ostriker, E. C., & Skinner, M. A. 2017b, ApJ, 851, 93
Kim, J. W., Edge, A. C., Wake, D. A., & Stott, J. P. 2011, MNRAS, 410, 241
Kim, W.-T. & Elmegreen, B. G. 2017, ApJ, 841, L4
Kim, Y. & Im, M. 2021, ApJ, 910, L11
Kimm, T., Cen, R., Rosdahl, J., & Yi, S. K. 2016, ApJ, 823, 52
Kimm, T., Haehnelt, M., Blaizot, J., et al. 2018, MNRAS, 475, 4617
Kobayashi, M. A. R., Murata, K. L., Koekemoer, A. M., et al. 2016, ApJ, 819, 25
Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, ApJS, 172, 196
Konno, A., Ouchi, M., Nakajima, K., et al. 2016, ApJ, 823, 20
Konno, A., Ouchi, M., Shibuya, T., et al. 2018, PASJ, 70, S16
Kormendy, J. 2013, Secular Evolution in Disk Galaxies (Falcón-Barroso, Jesús and Knapen, Johan H.), 1
Kormendy, J. 2016, in Astrophysics and Space Science Library, Vol. 418, Galactic Bulges, ed. E. Laurikainen, R. Peletier, & D. Gadotti, 431
Kormendy, J. & Ho, L. C. 2013, ARA&A, 51, 511
Kozłowski, S. 2016, ApJ, 826, 118
Kron, R. G. 1980, ApJS, 43, 305
Kroupa, P., Subr, L., Jerabkova, T., & Wang, L. 2020, MNRAS, 498, 5652
Kudritzki, R. P., Méndez, R. H., Feldmeier, J. J., et al. 2000, ApJ, 536, 19
Kulkarni, G., Worseck, G., & Hennawi, J. F. 2019, MNRAS, 488, 1035
Lacey, C. & Cole, S. 1993, MNRAS, 262, 627
Lacey, C. G., Baugh, C. M., Frenk, C. S., et al. 2016, MNRAS, 462, 3854
Lacy, M., Ridgway, S. E., Sajina, A., et al. 2015, ApJ, 802, 102
Lagos, C. d. P., Tobar, R. J., Robotham, A. S. G., et al. 2018, MNRAS, 481, 3573
Lai, K., Huang, J.-S., Fazio, G., et al. 2008, ApJ, 674, 70
Latif, M. A., Bovino, S., Grassi, T., Schleicher, D. R. G., & Spaans, M. 2015, MNRAS, 446, 3163
Latif, M. A., Bovino, S., Van Borm, C., et al. 2014a, MNRAS, 443, 1979
Latif, M. A. & Ferrara, A. 2016, Publ. Astron. Soc. Australia, 33, e051
Latif, M. A. & Khochfar, S. 2019, MNRAS, 490, 2706
Latif, M. A., Niemeyer, J. C., & Schleicher, D. R. G. 2014b, MNRAS, 440, 2969
Latif, M. A., Omukai, K., Habouzit, M., Schleicher, D. R. G., & Volonteri, M. 2016, ApJ, 823, 40
Latif, M. A. & Schleicher, D. R. G. 2015, MNRAS, 449, 77
Latif, M. A., Schleicher, D. R. G., Bovino, S., Grassi, T., & Spaans, M. 2014c, ApJ, 792, 78

Latif, M. A., Schleicher, D. R. G., Schmidt, W., & Niemeyer, J. C. 2013, MNRAS, 436, 2989
Latif, M. A., Schleicher, D. R. G., Spaans, M., & Zaroubi, S. 2011, A&A, 532, A66
Latif, M. A. & Volonteri, M. 2015, MNRAS, 452, 1026
Latif, M. A., Volonteri, M., & Wise, J. H. 2018, MNRAS, 476, 5016
Laursen, P., Sommer-Larsen, J., & Andersen, A. C. 2009, ApJ, 704, 1640
Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599
Le Fèvre, O., Tasca, L. A. M., Cassata, P., et al. 2015, A&A, 576, A79
Leclercq, F., Bacon, R., Wisotzki, L., et al. 2017, A&A, 608, A8
Leitherer, C. 1994, Reviews in Modern Astronomy, 7, 73
Leitherer, C., Robert, C., & Drissen, L. 1992, ApJ, 401, 596
Li, Z., Sellwood, J. A., & Shen, J. 2017, ApJ, 850, 67
Liddle, A. R. & Lyth, D. H. 2000, Cosmological Inflation and Large-Scale Structure (Cambridge University Press)
Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, ApJS, 172, 70
Liu, C., Mutch, S. J., Angel, P. W., et al. 2016, MNRAS, 462, 235
Liu, J., Bian, L., Cai, R.-G., Guo, Z.-K., & Wang, S.-J. 2021, arXiv e-prints, arXiv:2106.05637
Logroño-García, R., Vilella-Rojo, G., López-Sanjuan, C., et al. 2019, A&A, 622, A180
Longair, M. S. 2008, Galaxy Formation (Cambridge University Press)
López-Sanjuan, C., Tremblay, P. E., Ederoclite, A., et al. 2021, arXiv e-prints, arXiv:2110.14421
López-Sanjuan, C., Varela, J., Cristóbal-Hornillos, D., et al. 2019a, A&A, 631, A119
López-Sanjuan, C., Vázquez Ramió, H., Varela, J., et al. 2019b, A&A, 622, A177
Loveday, J., Norberg, P., Baldry, I. K., et al. 2012, MNRAS, 420, 1239
Lupi, A., Colpi, M., Devecchi, B., Galanti, G., & Volonteri, M. 2014, MNRAS, 442, 3616
Lupi, A., Haiman, Z., & Volonteri, M. 2021a, MNRAS, 503, 5046
Lupi, A., Volonteri, M., Decarli, R., Bovino, S., & Silk, J. 2021b, arXiv e-prints, arXiv:2109.01679
Lupi, A., Volonteri, M., Decarli, R., et al. 2019, MNRAS, 488, 4004
Lusso, E., Fumagalli, M., Fossati, M., et al. 2019, MNRAS, 485, L62
Lyttleton, R. A. & Hoyle, F. 1940, The Observatory, 63, 39
MacLeod, C. L., Green, P. J., Anderson, S. F., et al. 2018, AJ, 155, 6
Madau, P., Ferguson, H. C., Dickinson, M. E., et al. 1996, MNRAS, 283, 1388
Madau, P., Ferrara, A., & Rees, M. J. 2001, ApJ, 555, 92
Madau, P. & Kuhlen, M. 2003, in Texas in Tuscany. XXI Symposium on Relativistic Astrophysics, ed. R. Bandiera, R. Maiolino, & F. Mannucci, 31–44
Madau, P., Pozzetti, L., & Dickinson, M. 1998, ApJ, 498, 106
Madau, P. & Rees, M. J. 2001, ApJ, 551, L27
Maio, U., Borgani, S., Ciardi, B., & Petkova, M. 2019, Publ. Astron. Soc. Australia, 36, e020
Maio, U., Ciardi, B., Dolag, K., Tornatore, L., & Khochfar, S. 2010, MNRAS, 407, 1003
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94
Malbon, R. K., Baugh, C. M., Frenk, C. S., & Lacey, C. G. 2007, MNRAS, 382, 1394
Malhotra, S. & Rhoads, J. E. 2004, ApJ, 617, L5
Malmquist, K. G. 1922, Meddelanden fran Lunds Astronomiska Observatorium Serie I, 100, 1
Malmquist, K. G. 1925, Meddelanden fran Lunds Astronomiska Observatorium Serie I, 106, 1
Manti, S., Gallerani, S., Ferrara, A., Greig, B., & Feruglio, C. 2017, MNRAS, 466, 1160
Mapelli, M. 2016, MNRAS, 459, 3432
Marconi, A., Risaliti, G., Gilli, R., et al. 2004, MNRAS, 351, 169
Marin-Franch, A., Taylor, K., Cenarro, J., Cristobal-Hornillos, D., & Moles, M. 2015, in IAU General Assembly, Vol. 29, 2257381
Markevitch, M., Gonzalez, A. H., Clowe, D., et al. 2004, ApJ, 606, 819
Marques-Chaves, R., Álvarez-Márquez, J., Colina, L., et al. 2020, MNRAS, 499, L105
Marques-Chaves, R., Pérez-Fournon, I., Villar-Martín, M., et al. 2019, A&A, 629, A23
Marshall, M. A., Mutch, S. J., Qin, Y., Poole, G. B., & Wyithe, J. S. B. 2020a, MNRAS, 494, 2747
Marshall, M. A., Ni, Y., Di Matteo, T., et al. 2020b, MNRAS, 499, 3819
Marshall, M. A., Wyithe, J. S. B., Windhorst, R. A., et al. 2021, MNRAS, 506, 1209
Martínez-Solaeche, G., González Delgado, R. M., García-Benito, R., et al. 2021, A&A, 647, A158
Marulli, F., Bonoli, S., Branchini, E., Moscardini, L., & Springel, V. 2008, MNRAS, 385, 1846
Marulli, F., Branchini, E., Moscardini, L., & Volonteri, M. 2007, MNRAS, 375, 649
Marziani, P., Negrete, C. A., Dultzin, D., & Sulentic, J. W. 2011, Baltic Astronomy, 20, 406
Marziani, P., Sulentic, J. W., Negrete, C. A., et al. 2015, Ap&SS, 356, 339
Matsuda, Y., Yamada, T., Hayashino, T., et al. 2004, AJ, 128, 569
Matsuda, Y., Yamada, T., Hayashino, T., et al. 2005, ApJ, 634, L125
Matsuda, Y., Yamada, T., Hayashino, T., et al. 2011, MNRAS, 410, L13
Matthee, J., Pezzulli, G., Mackenzie, R., et al. 2020, MNRAS, 498, 3043
Matthee, J., Sobral, D., Best, P., et al. 2017a, MNRAS, 465, 3637
Matthee, J., Sobral, D., Best, P., et al. 2017b, MNRAS, 471, 629
Matthee, J., Sobral, D., Hayes, M., et al. 2021, MNRAS, 505, 1382
Matthee, J., Sobral, D., Oteo, I., et al. 2016, MNRAS, 458, 449
Matthee, J., Sobral, D., Santos, S., et al. 2015, MNRAS, 451, 400
Matthee, J. J. A., Sobral, D., Swinbank, A. M., et al. 2014, MNRAS, 440, 2375
Matute, I., La Franca, F., Pozzi, F., et al. 2006, A&A, 451, 443
Mayer, L. & Bonoli, S. 2019, Reports on Progress in Physics, 82, 016901
Mayer, L., Fiacconi, D., Bonoli, S., et al. 2015, ApJ, 810, 51
Mayer, L., Kazantzidis, S., Escala, A., & Callegari, S. 2010, Nature, 466, 1082
Mazzucchelli, C., Bañados, E., Venemans, B. P., et al. 2017, ApJ, 849, 91
McGreer, I. D., Fan, X., Jiang, L., & Cai, Z. 2018, AJ, 155, 131

Mei, S., Scarlata, C., Pentericci, L., et al. 2015, ApJ, 804, 117
Merloni, A., Bongiorno, A., Brusa, M., et al. 2014, MNRAS, 437, 3550
Merloni, A. & Heinz, S. 2008, MNRAS, 388, 1011
Mezcua, M. 2019, Nature Astronomy, 3, 6
Mezcua, M. 2021, IAU Symposium, 359, 238
Mezcua, M., Civano, F., Fabbiano, G., Miyaji, T., & Marchesi, S. 2016, ApJ, 817, 20
Mezcua, M., Civano, F., Marchesi, S., et al. 2018, MNRAS, 478, 2576
Mezcua, M. & Domínguez Sánchez, H. 2020, ApJ, 898, L30
Mezcua, M., Prieto, M. A., Fernández-Ontiveros, J. A., et al. 2015, MNRAS, 452, 4128
Mezcua, M., Suh, H., & Civano, F. 2019, MNRAS, 488, 685
Miller, B. P., Gallo, E., Greene, J. E., et al. 2015, ApJ, 799, 98
Miyaji, T., Hasinger, G., Salvato, M., et al. 2015, ApJ, 804, 104
Miyaji, T., Hasinger, G., & Schmidt, M. 2000, A&A, 353, 25
Mo, H. J., Mao, S., & White, S. D. M. 1998, MNRAS, 295, 319
Moles, M., Benítez, N., Aguerri, J. A. L., et al. 2008, AJ, 136, 1325
Møller, P. & Fynbo, J. U. 2001, A&A, 372, L57
Møller, P. & Warren, S. J. 1998, MNRAS, 299, 661
Monaco, P., Møller, P., Fynbo, J. P. U., et al. 2005, A&A, 440, 799
Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, Nature, 474, 616
Myers, A. D., Palanque-Delabrouille, N., Prakash, A., et al. 2015, ApJS, 221, 27
Nakajima, K., Fletcher, T., Ellis, R. S., Robertson, B. E., & Iwata, I. 2018, MNRAS, 477, 2098
Natarajan, P. 2011, arXiv e-prints, arXiv:1105.4902
Natarajan, P., Pacucci, F., Ferrara, A., et al. 2017, ApJ, 838, 117
Negrete, C. A., Dultzin, D., Marziani, P., & Sulentic, J. W. 2014, ApJ, 794, 95
Neistein, E., van den Bosch, F. C., & Dekel, A. 2006, MNRAS, 372, 933
Netzer, H. 2015, ARA&A, 53, 365
Neumayer, N. & Walcher, C. J. 2012, Advances in Astronomy, 2012, 709038
Nilsson, K. K. & Møller, P. 2009, A&A, 508, L21
Nilsson, K. K., Möller-Nilsson, O., Møller, P., Fynbo, J. P. U., & Shapley, A. E. 2009a, MNRAS, 400, 232
Nilsson, K. K., Östlin, G., Møller, P., et al. 2011, A&A, 529, A9
Nilsson, K. K., Tapken, C., Møller, P., et al. 2009b, A&A, 498, 13
Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJ, 660, L43
North, P. L., Courbin, F., Eigenbrod, A., & Chelouche, D. 2012, A&A, 542, A91
Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I., & Stefanon, M. 2018, ApJ, 855, 105
Oh, S. P. & Haiman, Z. 2002, ApJ, 569, 558
Oke, J. B. 1974, ApJS, 27, 21
Oke, J. B. & Gunn, J. E. 1983, ApJ, 266, 713
Omukai, K. 2000, ApJ, 534, 809
Omukai, K. 2001a, Ap&SS, 276, 807
Omukai, K. 2001b, ApJ, 546, 635
Omukai, K. & Palla, F. 2001, ApJ, 561, L55
Omukai, K., Schneider, R., & Haiman, Z. 2008, ApJ, 686, 801
Ono, Y., Ouchi, M., Harikane, Y., et al. 2018, PASJ, 70, S10
Onoue, M., Bañados, E., Mazzucchelli, C., et al. 2020, ApJ, 898, 105
Oppenheimer, J. R. & Volkoff, G. M. 1939, Physical Review, 55, 374
O'Shea, B. W. & Norman, M. L. 2008, ApJ, 673, 14
Ostriker, J. P. & Peebles, P. J. E. 1973, ApJ, 186, 467
Ouchi, M., Harikane, Y., Shibuya, T., et al. 2018, PASJ, 70, S13
Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2005, ApJ, 620, L1
Ouchi, M., Shimasaku, K., Akiyama, M., et al. 2008, ApJS, 176, 301
Ouchi, M., Shimasaku, K., Furusawa, H., et al. 2010, ApJ, 723, 869
Overzier, R. A., Bouwens, R. J., Cross, N. J. G., et al. 2008, ApJ, 673, 143
Pacifici, C., Kassin, S. A., Weiner, B. J., et al. 2016, ApJ, 832, 79
Pacucci, F., Mezcua, M., & Regan, J. A. 2021, arXiv e-prints, arXiv:2107.09069
Pacucci, F., Natarajan, P., Volonteri, M., Cappelluti, N., & Urry, C. M. 2017, ApJ, 850, L42
Pacucci, F., Volonteri, M., & Ferrara, A. 2015, MNRAS, 452, 1922
Palanque-Delabrouille, N., Magneville, C., Yèche, C., et al. 2016, A&A, 587, A41
Palouš, J., Ehlerová, S., Wünsch, R., & Morris, M. R. 2020, A&A, 644, A72
Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, A&A, 613, A51
Pâris, I., Petitjean, P., Rollinde, E., et al. 2011, A&A, 530, A50
Partridge, R. B. & Peebles, P. J. E. 1967, ApJ, 147, 868
Pascual, S., Gallego, J., & Zamorano, J. 2007, PASP, 119, 30
Paulino-Afonso, A., Sobral, D., Buitrago, F., & Afonso, J. 2017, MNRAS, 465, 2717
Paulino-Afonso, A., Sobral, D., Ribeiro, B., et al. 2018, MNRAS, 476, 5479
Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419
Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565
Peters, C. M., Richards, G. T., Myers, A. D., et al. 2015, ApJ, 811, 95
Pezzulli, E., Volonteri, M., Schneider, R., & Valiante, R. 2017, MNRAS, 471, 589
Pilkington, J. D. H., Hewish, A., Bell, S. J., & Cole, T. W. 1968, Nature, 218, 126
Planck Collaboration, Adam, R., Ade, P. A. R., et al. 2016a, A&A, 594, A1
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016b, A&A, 594, A13
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6
Polletta, M., Tajer, M., Maraschi, L., et al. 2007, ApJ, 663, 81

Popa, C., Naoz, S., Marinacci, F., & Vogelsberger, M. 2016, MNRAS, 460, 1625
Portegies Zwart, S. F., Makino, J., McMillan, S. L. W., & Hut, P. 1999, A&A, 348, 117
Portegies Zwart, S. F. & McMillan, S. L. W. 2002, ApJ, 576, 899
Prescott, M. K. M., Dey, A., Brodwin, M., et al. 2012, ApJ, 752, 86
Press, W. H. & Schechter, P. 1974, ApJ, 187, 425
Prialnik, D. 2009, An Introduction to the Theory of Stellar Structure and Evolution (Cambridge University Press)
Qin, Y., Mutch, S. J., Poole, G. B., et al. 2017, MNRAS, 472, 2009
Rasio, F. A., Freitag, M., & Gürkan, M. A. 2004, in Coevolution of Black Holes and Galaxies, ed. L. C. Ho, 138
Rauch, M. 1998, ARA&A, 36, 267
Rauch, M., Haehnelt, M., Bunker, A., et al. 2008, ApJ, 681, 856
Reddy, N. A., Steidel, C. C., Pettini, M., et al. 2008, ApJS, 175, 48
Rees, M. J., Begelman, M. C., Blandford, R. D., & Phinney, E. S. 1982, Nature, 295, 17
Regan, J. A. & Downes, T. P. 2018, MNRAS, 475, 4636
Regan, J. A., Downes, T. P., Volonteri, M., et al. 2019, MNRAS, 486, 3892
Regan, J. A. & Haehnelt, M. G. 2009, MNRAS, 393, 858
Regan, J. A., Johansson, P. H., & Wise, J. H. 2014, ApJ, 795, 137
Regan, J. A., Johansson, P. H., & Wise, J. H. 2015, MNRAS, 449, 3766
Regan, J. A., Johansson, P. H., & Wise, J. H. 2016, MNRAS, 459, 3377
Regan, J. A., Visbal, E., Wise, J. H., et al. 2017, Nature Astronomy, 1, 0075
Regan, J. A., Wise, J. H., Woods, T. E., et al. 2020, The Open Journal of Astrophysics, 3, 15
Reinoso, B., Schleicher, D. R. G., Fellhauer, M., Klessen, R. S., & Boekholt, T. C. N. 2018, A&A, 614, A14
Ressler, S. M., Quataert, E., & Stone, J. M. 2018, MNRAS, 478, 3544
Ribeiro, B., Le Fèvre, O., Tasca, L. A. M., et al. 2016, A&A, 593, A22
Richards, G. T., Fan, X., Newberg, H. J., et al. 2002, AJ, 123, 2945
Richards, G. T., Myers, A. D., Gray, A. G., et al. 2009, ApJS, 180, 67
Richards, G. T., Myers, A. D., Peters, C. M., et al. 2015, ApJS, 219, 39
Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009
Rindler, W. 1956, MNRAS, 116, 662
Ritter, J. S., Safranek-Shrader, C., Gnat, O., Milosavljević, M., & Bromm, V. 2012, ApJ, 761, 56
Rosas-Guevara, Y., Bower, R. G., Schaye, J., et al. 2016, MNRAS, 462, 190
Ross, N. P., McGreer, I. D., White, M., et al. 2013, ApJ, 773, 14
Ross, N. P., Myers, A. D., Sheldon, E. S., et al. 2012, ApJS, 199, 3
Ruiz, A. N., Cora, S. A., Padilla, N. D., et al. 2015, ApJ, 801, 139
Rybicki, G. B. & Lightman, A. P. 1986, Radiative Processes in Astrophysics (Wiley-VCH)
Ryden, B. 2016, Introduction to Cosmology (Cambridge University Press)
Saito, T., Matsuda, Y., Lacey, C. G., et al. 2015, MNRAS, 447, 3069
Saito, T., Shimasaku, K., Okamura, S., et al. 2006, ApJ, 648, 54
Sakurai, Y., Yoshida, N., Fujii, M. S., & Hirano, S. 2017, MNRAS, 472, 1677
Sala, L., Cenci, E., Capelo, P. R., Lupi, A., & Dotti, M. 2021, MNRAS, 500, 4788
Santoro, F. & Shull, J. M. 2006, ApJ, 643, 26
Santos, M. R., Ellis, R. S., Kneib, J.-P., Richard, J., & Kuijken, K. 2004, ApJ, 606, 683
Santos, S., Sobral, D., Butterworth, J., et al. 2021, MNRAS, 505, 1117
Santos, S., Sobral, D., & Matthee, J. 2016, MNRAS, 463, 1678
Santos, S., Sobral, D., Matthee, J., et al. 2020, MNRAS, 493, 141
Sartori, L. F., Schawinski, K., Koss, M., et al. 2016, MNRAS, 457, 3629
Sassano, F., Schneider, R., Valiante, R., et al. 2021, MNRAS, 506, 613
Scannapieco, E., Ferrara, A., & Madau, P. 2002, ApJ, 574, 590
Schaerer, D. 2002a, A&A, 382, 28
Schaerer, D. 2002b, arXiv e-prints, astro
Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521
Schechter, P. 1976, ApJ, 203, 297
Schleicher, D. R. G., Palla, F., Ferrara, A., Galli, D., & Latif, M. 2013, A&A, 558, A59
Schmidt, M. 1959, ApJ, 129, 243
Schmidt, M. 1968, ApJ, 151, 393
Schmidt, M. & Green, R. F. 1983, ApJ, 269, 352
Schneider, R. 2006, New Astron. Rev., 50, 64
Schneider, R., Ferrara, A., Natarajan, P., & Omukai, K. 2002, ApJ, 571, 30
Schneider, R. & Omukai, K. 2010, MNRAS, 402, 429
Schneider, R., Omukai, K., Bianchi, S., & Valiante, R. 2012, MNRAS, 419, 1566
Schneider, R., Omukai, K., Inoue, A. K., & Ferrara, A. 2006, MNRAS, 369, 1437
Schulze, A. & Wisotzki, L. 2010, A&A, 516, A87
Schwarzschild, K. 1916, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin, 189
Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1
Sedgwick, C., Serjeant, S., Pearson, C., et al. 2011, MNRAS, 416, 1862
Selsing, J., Fynbo, J. P. U., Christensen, L., & Krogager, J. K. 2016, A&A, 585, A87
Sesana, A., Barausse, E., Dotti, M., & Rossi, E. M. 2014, ApJ, 794, 104
Sesana, A., Gair, J., Berti, E., & Volonteri, M. 2011, Phys. Rev. D, 83, 044036
Sesana, A., Volonteri, M., & Haardt, F. 2007, MNRAS, 377, 1711
Shakura, N. I. & Sunyaev, R. A. 1976, MNRAS, 175, 613
Shang, C., Bryan, G. L., & Haiman, Z. 2010, MNRAS, 402, 1249
Shankar, F. 2013, Classical and Quantum Gravity, 30, 244001
Shankar, F., Salucci, P., Granato, G. L., De Zotti, G., & Danese, L. 2004, MNRAS, 354, 1020
Shankar, F., Weinberg, D. H., & Miralda-Escudé, J. 2009, ApJ, 690, 20

Shapiro, S. L. & Teukolsky, S. A. 1986, Black Holes, White Dwarfs and Neutron Stars: The Physics of Compact Objects (Wiley-Interscience)
Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65
Sharma, P., Roy, A., Nath, B. B., & Shchekinov, Y. 2014, MNRAS, 443, 3463
Shen, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2020, MNRAS, 495, 3252
Shibuya, T., Ouchi, M., Harikane, Y., & Nakajima, K. 2019, ApJ, 871, 164
Shibuya, T., Ouchi, M., Harikane, Y., et al. 2018a, PASJ, 70, S15
Shibuya, T., Ouchi, M., Konno, A., et al. 2018b, PASJ, 70, S14
Shibuya, T., Ouchi, M., Nakajima, K., et al. 2014, ApJ, 785, 64
Shioya, Y., Taniguchi, Y., Sasaki, S. S., et al. 2008, ApJS, 175, 128
Shlosman, I., Frank, J., & Begelman, M. C. 1989, Nature, 338, 45
Sijacki, D., Vogelsberger, M., Genel, S., et al. 2015, MNRAS, 452, 575
Silk, J. & Rees, M. J. 1998, A&A, 331, L1
Singal, A. K. 2016, ApJ, 827, 66
Singal, J., George, J., & Gerber, A. 2016, ApJ, 831, 60
Singal, J., Petrosian, V., Lawrence, A., & Stawarz, Ł. 2011, ApJ, 743, 104
Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, AJ, 146, 32
Smith, B., Regan, J., Downes, T., et al. 2018, ArXiv e-prints [arXiv:1804.06477]
Smith, B., Sigurdsson, S., & Abel, T. 2008, MNRAS, 385, 1443
Smith, B. D., Turk, M. J., Sigurdsson, S., O'Shea, B. W., & Norman, M. L. 2009a, ApJ, 691, 441
Smith, R. J., Longmore, S., & Bonnell, I. 2009b, MNRAS, 400, 1775
Sobral, D., Best, P. N., Geach, J. E., et al. 2009a, MNRAS, 398, L68
Sobral, D., Best, P. N., Geach, J. E., et al. 2009b, MNRAS, 398, 75
Sobral, D., Best, P. N., Smail, I., et al. 2014, MNRAS, 437, 3516
Sobral, D., Kohn, S. A., Best, P. N., et al. 2016, MNRAS, 457, 1739
Sobral, D. & Matthee, J. 2019, A&A, 623, A157
Sobral, D., Matthee, J., Best, P., et al. 2017, MNRAS, 466, 1242
Sobral, D., Matthee, J., Darvish, B., et al. 2015, ApJ, 808, 139
Sobral, D., Matthee, J., Darvish, B., et al. 2018a, MNRAS, 477, 2817
Sobral, D., Santos, S., Matthee, J., et al. 2018b, MNRAS, 476, 4725
Sobral, D., Smail, I., Best, P. N., et al. 2013, MNRAS, 428, 1128
Soltan, A. 1982, MNRAS, 200, 115
Somerville, R. S., Primack, J. R., & Faber, S. M. 2001, MNRAS, 320, 504
Souchay, J., Andrei, A. H., Barache, C., et al. 2015a, A&A, 583, A75
Souchay, J., Andrei, A. H., Barache, C., et al. 2015b, VizieR Online Data Catalog, J/A+A/583/A75
Spinoso, D., Orsi, A., López-Sanjuan, C., et al. 2020, A&A, 643, A149
Springel, V. 2016, Saas-Fee Advanced Course, 43, 251
Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629
Stahler, S. W. & Palla, F. 2004, The Formation of Stars (Wiley-VCH)
Stefanon, M., Labbé, I., Bouwens, R. J., et al. 2017, ApJ, 851, 43
Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, ApJ, 519, 1
Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2000, ApJ, 532, 170
Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, ApJ, 462, L17
Steidel, C. C. & Hamilton, D. 1992, AJ, 104, 941
Steinborn, L. K., Hirschmann, M., Dolag, K., et al. 2018, MNRAS, 481, 341
Stiavelli, M., Scarlata, C., Panagia, N., et al. 2001, ApJ, 561, L37
Stone, N. C., Küpper, A. H. W., & Ostriker, J. P. 2017, MNRAS, 467, 4180
Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 123, 485
Stroe, A. & Sobral, D. 2015, MNRAS, 453, 242
Stroe, A., Sobral, D., Matthee, J., Calhau, J., & Oteo, I. 2017a, MNRAS, 471, 2558
Stroe, A., Sobral, D., Matthee, J., Calhau, J., & Oteo, I. 2017b, MNRAS, 471, 2575
Sugimura, K., Hosokawa, T., Yajima, H., & Omukai, K. 2017, MNRAS, 469, 62
Sutherland, R. S. & Dopita, M. A. 1993, ApJS, 88, 253
Swanson, M. E. C., Tegmark, M., Hamilton, A. J. S., & Hill, J. C. 2008, MNRAS, 387, 1391
Tanaka, T. L. & Li, M. 2014, MNRAS, 439, 1092
Tanaka, T. L., Li, M., & Haiman, Z. 2013, MNRAS, 435, 3559
Taniguchi, Y., Murayama, T., Scoville, N. Z., et al. 2009, arXiv e-prints, arXiv:0906.1873
Taniguchi, Y., Scoville, N., Murayama, T., et al. 2007, ApJS, 172, 9
Tasca, L. A. M., Le Fèvre, O., Ribeiro, B., et al. 2017, A&A, 600, A110
Taylor, K., Marín-Franch, A., Laporte, R., et al. 2014, Journal of Astronomical Instrumentation, 3, 1350010
Telfer, R. C., Zheng, W., Kriss, G. A., & Davidsen, A. F. 2002, ApJ, 565, 773
Toba, Y., Oyabu, S., Matsuhara, H., et al. 2013, PASJ, 65, 113
Tokunaga, A. T. & Vacca, W. D. 2005, PASP, 117, 421
Tremmel, M., Governato, F., Volonteri, M., Pontzen, A., & Quinn, T. R. 2018, ApJ, 857, L22
Tsuribe, T. & Omukai, K. 2008, ApJ, 676, L45
Tully, R. B. & Fisher, J. R. 1977, A&A, 500, 105
Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, ApJ, 786, 104
Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, ApJ, 598, 886
Urry, C. M. & Padovani, P. 1995, PASP, 107, 803
Valiante, R., Agarwal, B., Habouzit, M., & Pezzulli, E. 2017, Publ. Astron. Soc. Australia, 34, e031
Valiante, R., Colpi, M., Schneider, R., et al. 2021, MNRAS, 500, 4095
Valiante, R., Schneider, R., Salvadori, S., & Bianchi, S. 2011, MNRAS, 416, 1916
Valiante, R., Schneider, R., Salvadori, S., & Gallerani, S. 2014, MNRAS, 444, 2442

Valiante, R., Schneider, R., Volonteri, M., & Omukai, K. 2016, MNRAS, 457, 3356
van Breukelen, C., Jarvis, M. J., & Venemans, B. P. 2005, MNRAS, 359, 895
van Velzen, S., Falcke, H., & Körding, E. 2015, MNRAS, 446, 2985
van Wassenhove, S., Volonteri, M., Walker, M. G., & Gair, J. R. 2010, MNRAS, 408, 1139
Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, AJ, 122, 549
Venemans, B. P., Findlay, J. R., Sutherland, W. J., et al. 2013, ApJ, 779, 24
Venemans, B. P., Röttgering, H. J. A., Miley, G. K., et al. 2005, A&A, 431, 793
Viironen, K., López-Sanjuan, C., Hernández-Monteagudo, C., et al. 2018, A&A, 614, A129
Vilella-Rojo, G., Logroño-García, R., López-Sanjuan, C., et al. 2021, A&A, 650, A68
Vilella-Rojo, G., Viironen, K., López-Sanjuan, C., et al. 2015, A&A, 580, A47
Visbal, E., Bryan, G. L., & Haiman, Z. 2020, ApJ, 897, 95
Visbal, E., Haiman, Z., & Bryan, G. L. 2014a, MNRAS, 442, L100
Visbal, E., Haiman, Z., & Bryan, G. L. 2014b, MNRAS, 445, 1056
Vito, F., Gilli, R., Vignali, C., et al. 2014, MNRAS, 445, 3557
Vogelsberger, M., Nelson, D., Pillepich, A., et al. 2020, MNRAS, 492, 5167
Volonteri, M. 2007, ApJ, 663, L5
Volonteri, M. 2010, A&ARv, 18, 279
Volonteri, M. & Bellovary, J. 2012, Reports on Progress in Physics, 75, 124901
Volonteri, M., Capelo, P. R., Netzer, H., et al. 2015a, MNRAS, 452, L6
Volonteri, M., Capelo, P. R., Netzer, H., et al. 2015b, MNRAS, 449, 1470
Volonteri, M., Dubois, Y., Pichon, C., & Devriendt, J. 2016, MNRAS, 460, 2979
Volonteri, M., Gültekin, K., & Dotti, M. 2010, MNRAS, 404, 2143
Volonteri, M., Haardt, F., & Madau, P. 2003, ApJ, 582, 559
Volonteri, M., Lodato, G., & Natarajan, P. 2008, MNRAS, 383, 1079
Volonteri, M. & Madau, P. 2008, ApJ, 687, L57
Volonteri, M., Pfister, H., Beckmann, R. S., et al. 2020, MNRAS, 498, 2219
Volonteri, M. & Rees, M. J. 2005, ApJ, 633, 624
Volonteri, M., Reines, A. E., Atek, H., Stark, D. P., & Trebitsch, M. 2017, ApJ, 849, 155
Volonteri, M., Sikora, M., Lasota, J. P., & Merloni, A. 2013, ApJ, 775, 94
Volonteri, M., Silk, J., & Dubus, G. 2015c, ApJ, 804, 148
Wang, F., Yang, J., Fan, X., et al. 2021, ApJ, 907, L1
Warren, S. J., Hambly, N. C., Dye, S., et al. 2007, MNRAS, 375, 213
Warren, S. J., Hewett, P. C., & Osmer, P. S. 1994, ApJ, 421, 412
Webster, B. L. & Murdin, P. 1972, Nature, 235, 37
Weinberg, D. H., Davé, R., Katz, N., & Kollmeier, J. A. 2003, in American Institute of Physics Conference Series, Vol. 666, The Emergence of Cosmic Structure, ed. S. H. Holt & C. S. Reynolds, 157–169
Weinberger, L. H., Kulkarni, G., Haehnelt, M. G., Choudhury, T. R., & Puchwein, E. 2018, MNRAS, 479, 2564
White, S. D. M. & Frenk, C. S. 1991, ApJ, 379, 52
White, S. D. M. & Rees, M. J. 1978, MNRAS, 183, 341
Wilkins, S. M., Bunker, A. J., Stanway, E., Lorenzoni, S., & Caruana, J. 2011, MNRAS, 417, 717
Willott, C. J., Albert, L., Arzoumanian, D., et al. 2010, AJ, 140, 546
Wise, J. H. & Abel, T. 2007, ApJ, 671, 1559
Wise, J. H. & Abel, T. 2008, ApJ, 685, 40
Wise, J. H., Demchenko, V. G., Halicek, M. T., et al. 2014, MNRAS, 442, 2560
Wise, J. H., Turk, M. J., Norman, M. L., & Abel, T. 2012, ApJ, 745, 50
Wisotzki, L., Bacon, R., Blaizot, J., et al. 2016, A&A, 587, A98
Wolcott-Green, J. & Haiman, Z. 2011, MNRAS, 412, 2603
Wolcott-Green, J. & Haiman, Z. 2019, MNRAS, 484, 2467
Wolcott-Green, J., Haiman, Z., & Bryan, G. L. 2011, MNRAS, 418, 838
Wolcott-Green, J., Haiman, Z., & Bryan, G. L. 2017, MNRAS, 469, 3329
Wolcott-Green, J., Haiman, Z., & Bryan, G. L. 2021, MNRAS, 500, 138
Woods, T. E., Heger, A., & Haemmerlé, L. 2020, MNRAS, 494, 2236
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Wu, K. K. S., Lahav, O., & Rees, M. J. 1999, Nature, 397, 225
Wu, X.-B., Wang, F., Fan, X., et al. 2015, Nature, 518, 512
Wu, Y., Shi, Y., Helou, G., et al. 2011, ApJ, 734, 40
Wyithe, J. S. B. 2006, MNRAS, 365, 1082
Wyithe, J. S. B., Agol, E., & Fluke, C. J. 2002, MNRAS, 331, 1041
Wyithe, J. S. B. & Bolton, J. S. 2011, MNRAS, 412, 1926
Wyithe, J. S. B. & Loeb, A. 2002a, ApJ, 581, 886
Wyithe, J. S. B. & Loeb, A. 2002b, ApJ, 577, 615
Wyithe, J. S. B. & Loeb, A. 2003, ApJ, 595, 614
Wyithe, J. S. B. & Loeb, A. 2004, Nature, 432, 194
Wyithe, J. S. B. & Loeb, A. 2005, ApJ, 634, 910
Wyithe, J. S. B. & Loeb, A. 2009, MNRAS, 395, 1607
Wyithe, J. S. B. & Loeb, A. 2012, MNRAS, 425, 2892
Wyithe, J. S. B., Loeb, A., & Carilli, C. 2005a, ApJ, 628, 575
Wyithe, J. S. B., Loeb, A., & Carilli, C. 2005b, ApJ, 628, 575
Wyithe, J. S. B. & Padmanabhan, T. 2006, MNRAS, 372, 1681
Wyithe, J. S. B., Webster, R. L., Turner, E. L., & Mortlock, D. J. 2000, MNRAS, 315, 62
Yadav, N., Mukherjee, D., Sharma, P., & Nath, B. B. 2017, MNRAS, 465, 1720
Yan, R., Bundy, K., Law, D. R., et al. 2016, AJ, 152, 197
Yang, Y., Zabludoff, A., Tremonti, C., Eisenstein, D., & Davé, R. 2009, ApJ, 693, 1579

Yao, S., Wu, X.-B., Ai, Y. L., et al. 2019, ApJS, 240, 6
Yates, R. M., Henriques, B. M. B., Fu, J., et al. 2021, MNRAS, 503, 4474
Yonehara, A., Mineshige, S., Manmoto, T., et al. 1998, ApJ, 501, L41
York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579
Yoshida, N., Abel, T., Hernquist, L., & Sugiyama, N. 2003, ApJ, 592, 645
Yoshida, N., Omukai, K., & Hernquist, L. 2007, ApJ, 667, L117
Yue, B., Ferrara, A., Salvaterra, R., Xu, Y., & Chen, X. 2014, MNRAS, 440, 1263
Zahedy, F. S., Rauch, M., Chen, H.-W., et al. 2019, MNRAS, 486, 1392
Zhang, Y., Ouchi, M., Gebhardt, K., et al. 2021, arXiv e-prints, arXiv:2105.11497

List of Figures

List of Tables

Appendix

Appendix A

Measurement of GTC spectra

We measure the redshift of all the 37 sources identified as QSOs (either at z∼1.5 or at z∼2.2) in both spectroscopic programs. We do not aim at reaching a higher precision than $\sigma_z = 10^{-2}$, since the main goal of our follow-up programs is the spectroscopic confirmation of our targets. We additionally extract the Lyα EW and integrated line flux $F_{Ly\alpha}$ for the 29 z∼2.2 QSOs, from which we compute the sources Lyα luminosity.

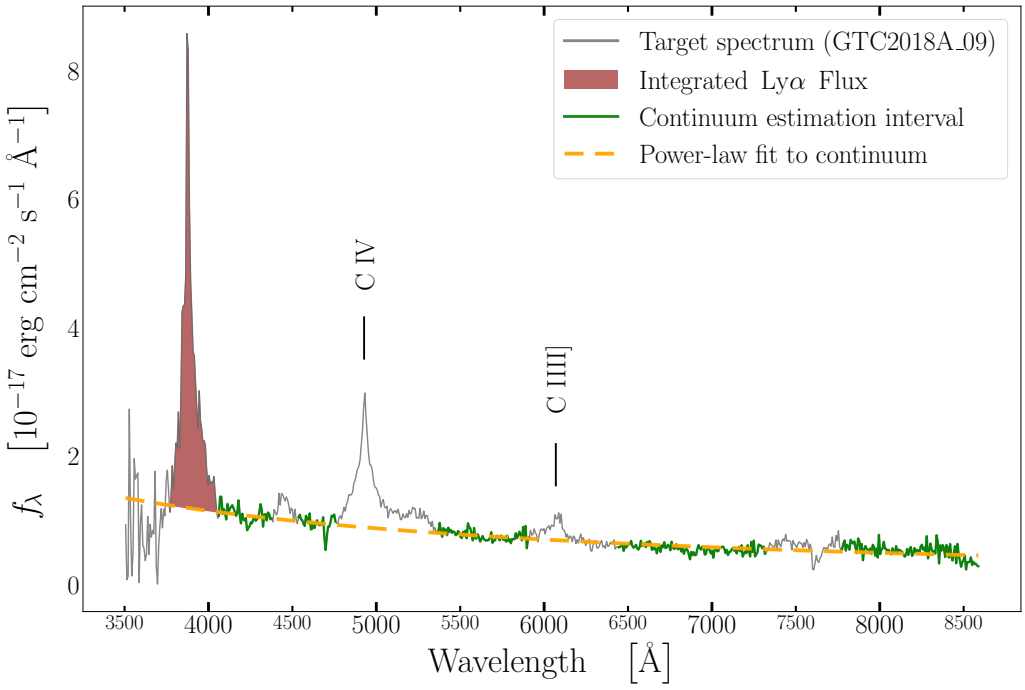

Figure A.1: Calibrated spectrum (grey solid line) of the GTC2018A_09 target, confirmed as z∼2.2 QSO. The four green regions highlight the intervals used for computing the power-law fit to the continuum (dashed yellow line). Finally, the Lyα integrated flux is highlighted in red.

Following well-established procedures (see e.g., Pâris et al., 2011), we first identify the main spectral lines in our QSOs spectra, such as `CIV` and `CIII]`. We then use their profile-peaks to compute our redshift estimate. We discard the Lyα profile for this analysis, since it provides a systematically biased z measure, due to the complex radiative transfer of Lyα photons in the source rest frame and IGM (see e.g., Gronke et al., 2016; Dijkstra, 2017; Gurung-Lopez et al., 2018; Gurung-López et al., 2019). We fit a double gaussian profile to both `CIV` and `CIII]` profiles, in order to trace at the same time its broad and narrow components. We use the λ position of the narrow-component peaks to obtain two z estimates, whose average provides the final spectroscopic z of our sources. The Lyα line flux can only be obtained after estimating the sources continua. We then fit a power law to the wavelength regions of each spectrum which are not affected by any line feature, as shown by the yellow sections of the spectrum displayed in

Fig. A.1. We use the following simple functional form:

$$\mathrm{f}_{\lambda}^{\mathrm{c}}(\lambda) = \mathrm{k}\,\lambda_{\mathrm{obs}}^{\alpha}\ , \tag{A.1}$$

where $\mathrm{f}_{\lambda}^{\mathrm{c}}(\lambda)$ is the spectrum *monochromatic flux density* (in units of $\mathrm{erg\,cm^{-2}\,s^{-1}\,\mathring{A}^{-1}}$) while k and α are fit parameters. Finally, we measure the total Lyα line flux by integrating the excess above the estimated continuum in the wavelength range affected by the Lyα line, which is shown in Fig. A.1 as the spectral region highlighted in dark-red. As a last step, we estimate the observed Lyα EW as:

$$\mathrm{EW}_{\mathrm{obs}}^{\mathrm{Ly}\alpha} = \frac{\mathrm{F}_{\mathrm{Ly}\alpha}}{\mathrm{f}_{\lambda}^{\mathrm{c}}(\lambda_{\mathrm{Ly}\alpha})}\ , \tag{A.2}$$

in which $\mathrm{f}_{\lambda}^{\mathrm{c}}(\lambda_{\mathrm{Ly}\alpha})$ is the value of the power-law fit to the continuum at the wavelength of the Lyα line-profile peak. Figure A.1 shows the spectrum of target `GTC2018A_09` as a visual example of our measuring procedure. The results of these measurements are shown in Table A.1 together with a summary of the spectroscopic follow-up thechnical requrements and additional properties of the observed targets.

Instrument	OSIRIS spectrograph at Gran Telescopio Canarias (GTC)							
Grism	R500B							
Seeing	requested: 1.4 (maximum) — effective: 1.03 (averaged on all observations)							
Moon	requested: any — effective: >90% Dark							
Air mass	requested: 1.5 (maximum) — effective: 1.266 (averaged on all observations)							
SNR	$\geqslant 3$ at $\lambda_{obs} \sim 4000$Å							
ID	Ra [hh:mm:ss]	Dec [hh:mm:ss]	Time [s]	SDSS class	GTC class	z_{spec}	$L_{Ly\alpha}$	EW_{obs}
GTC2018A_01	22:50:40.27	34:23:43.6	2185	—	QSO/AGN	2.21	5.18e+44	513.98
GTC2018A_02	00:43:38.56	05:41:35.6	2185	GALAXY	QSO/AGN	2.23	1.70e+44	417.21
GTC2018A_03	23:01:08.05	33:44:20.0	2185	STAR	STAR	—	—	—
GTC2018A_04	16:17:15.12	50:25:59.2	2185	STAR	QSO/AGN	1.52	—	—
GTC2018A_05	22:22:06.30	11:07:47.7	2335	STAR	QSO/AGN	2.23	3.87e+43	268.21
GTC2018A_06	18:10:22.34	41:49:25.3	2185	STAR	QSO/AGN	2.22	2.32e+44	741.21
GTC2018A_07	17:35:17.03	31:44:42.8	2185	GALAXY	QSO/AGN	2.23	7.46e+43	665.04
GTC2018A_08	01:31:29.69	33:55:14.9	2185	STAR	QSO/AGN	2.21	1.26e+44	637.59
GTC2018A_09	14:59:37.24	47:15:26.3	2185	STAR	QSO/AGN	2.18	2.13e+44	502.01
GTC2018A_10	16:11:57.72	46:00:45.8	2185	STAR	QSO/AGN	2.21	1.56e+44	263.97
GTC2018A_11	18:32:04.26	39:54:08.8	2185	—	QSO/AGN	1.54	—	—
GTC2018A_12	02:16:13.21	34:28:37.3	2185	—	QSO/AGN	2.21	1.33e+44	272.11
GTC2018A_13	14:32:51.07	52:36:46.7	1954	GALAXY	GALAXY	0.51	—	—
GTC2018A_14	22:01:43.81	28:23:36.5	2335	STAR	QSO/AGN	1.53	—	—
GTC2018A_15	16:09:37.67	45:29:53.6	2245	STAR	STAR	—	—	—
GTC2018A_16	22:43:00.76	34:10:26.4	2335	STAR	QSO/AGN	2.25	—	—
GTC2018A_17	15:59:27.15	57:05:04.6	2245	STAR	QSO/AGN	2.22	1.08e+44	268.69
GTC2018A_18	23:03:24.69	33:20:25.5	2245	STAR	STAR	—	—	—
GTC2018A_19	16:03:33.14	46:11:53.2	2245	—	QSO/AGN	2.25	6.19e+43	351.25
GTC2018A_20	15:19:49.02	53:16:18.4	2335	STAR	QSO/AGN	2.19	2.04e+44	828.73
GTC2018A_21	14:53:19.08	53:02:42.3	2365	STAR	QSO/AGN	2.19	6.15e+43	217.99
GTC2018A_22	17:41:33.43	57:05:11.4	2335	STAR	QSO/AGN	2.27	8.20e+43	481.13
GTC2018A_23	15:38:49.50	48:58:13.1	2014	GALAXY	STAR	—	—	—
GTC2018A_24	14:53:32.94	54:09:44.8	2335	GALAXY	QSO/AGN	1.53	—	—
GTC2019A_01	07:18:49.01	40:50:42.7	1165	STAR	STAR	—	—	—
GTC2019A_02	22:37:58.84	11:41:01.4	1285	STAR	QSO/AGN	2.204	4.56e+44	706.71
GTC2019A_03	02:27:21.78	29:56:23.7	1366	STAR	QSO/AGN	2.202	3.41e+44	859.82
GTC2019A_04	16:14:11.38	53:11:16.6	1426	STAR	BAL QSO	2.174	2.79e+44	290.75
GTC2019A_05	15:31:12.21	48:08:31.4	1576	STAR	QSO/AGN	2.209	2.91e+44	570.59
GTC2019A_06	15:34:37.75	46:42:36.3	1816	STAR	QSO/AGN	2.231	3.02e+44	414.59
GTC2019A_07	12:38:36.94	56:10:39.8	2086	STAR	QSO/AGN	2.177	2.02e+44	183.25
GTC2019A_08	12:29:07.02	56:03:49.2	2086	STAR	QSO/AGN	2.191	1.35e+44	230.15
GTC2019A_09	08:03:53.42	30:46:36.2	2206	STAR	QSO/AGN	2.256	2.14e+44	380.86
GTC2019A_10	07:15:23.13	39:50:57.3	2356	STAR	QSO/AGN	2.189	1.29e+44	625.73
GTC2019A_11	09:00:47.61	32:06:54.9	2806	GALAXY	QSO/AGN	2.253	9.54e+43	364.01
GTC2019A_12	00:51:10.29	03:08:25.4	3406	STAR	QSO/AGN	1.543	8.34e+43	104.21
GTC2019A_13	17:25:01.20	33:46:50.4	3706	STAR	QSO/AGN	1.602	1.27e+44	180.52
GTC2019A_14	10:28:20.14	39:52:42.3	3706	STAR	QSO/AGN	2.271	1.83e+44	121.36
GTC2019A_15	00:32:08.19	39:47:23.6	4476	STAR	QSO/AGN	1.512	1.33e+44	118.74
GTC2019A_16	16:14:30.92	50:12:24.2	4686	GALAXY	QSO/AGN	2.211	1.00e+44	705.82
GTC2019A_17	00:41:06.77	08:02:56.5	4761	STAR	GALAXY	—	—	—
GTC2019A_18	16:10:19.83	45:31:49.4	5136	GALAXY	QSO/AGN	2.271	1.06e+44	644.06
GTC2019A_19	22:56:29.07	09:36:45.5	5136	STAR	QSO/AGN	2.762	8.96e+43	390.88
GTC2019A_20	09:10:21.90	38:39:30.6	6246	GALAXY	QSO/AGN	1.528	1.61e+44	561.05
GTC2019A_21	13:20:29.68	56:31:49.6	6795	GALAXY	QSO/AGN	2.197	8.61e+43	154.68

Table A.1: Properties retrieved from the follow-up of our 45 spectroscopic targets. These results confirm 29/45 sources (64.4%) as genuine Lyα-emitting QSOs at $z \sim 2$. The most numerous interlopers are CIV-emitting QSOs at $z \sim 1.52$, namely 8/45 targets ($\sim$ 18%, see Stroe et al., 2017a,b), and 5/45 blue stars ($\sim$ 11%). The latter are selected by our pipeline due to their strong color-gradients which mimic a NB photometric excess in the $J0395$ filter. We also note one Lyβ-emitting QSO contaminant at $z \sim 2.76$. We report the measured $L_{Ly\alpha}$ and EW_{obs} only for the confirmed QSOs at $z \sim 2.2$. Among these, the spectrum of GTC 16 could not be calibrated and measured.Finally, all the $z \sim 2.2$ confirmed sources in our sample are QSOs with $L_{Ly\alpha} > 6 \times 10^{43}$ erg s^{-1}. This supports the results of Konno et al. (2016); Matthee et al. (2017b); Sobral et al. (2018a,b) and Calhau et al. (2020) about the strong contribution of AGN/QSOs to the Lyα LF at $\mathrm{Log}(Ly\alpha) \gtrsim 43.3$.