High-redshift Galaxies and Black Holes Detectable with the JWST: A Population Synthesis Model from Infrared to X-Rays

The Astrophysical Journal, 849:155 (15pp), 2017 November 10 https://doi.org/10.3847/1538-4357/aa93f1

High-redshift Galaxies and Black Holes Detectable with the *JWST*: A Population Synthesis Model from Infrared to X-Rays

Marta Volonteri[1], Amy E. Reines[2,3], Hakim Atek[1], Daniel P. Stark[4], and Maxime Trebitsch[1]
[1] Institut d'Astrophysique de Paris, Sorbonne Universitès, UPMC Univ Paris 6 et CNRS, UMR 7095, 98 bis Boulevard Arago, F-75014 Paris, France; martav@iap.fr
[2] National Optical Astronomy Observatory, 950 North Cherry Avenue, Tucson, AZ 85719, USA
[3] Department of Physics, Montana State University, Bozeman, MT 59717, USA
[4] Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA
Received 2017 March 31; revised 2017 October 11; accepted 2017 October 12; published 2017 November 10

Abstract

The first billion years of the Universe has been a pivotal time: stars, black holes (BHs), and galaxies formed and assembled, sowing the seeds of galaxies as we know them today. Detecting, identifying, and understanding the first galaxies and BHs is one of the current observational and theoretical challenges in galaxy formation. In this paper we present a population synthesis model aimed at galaxies, BHs, and active galactic nuclei (AGNs) at high redshift. The model builds a population based on empirical relations. The spectral energy distribution of galaxies is determined by age and metallicity, and that of AGNs by BH mass and accretion rate. We validate the model against observations, and predict properties of galaxies and AGN in other wavelength and/or luminosity ranges, estimating the contamination of stellar populations (normal stars and high-mass X-ray binaries) for AGN searches from the infrared to X-rays, and vice versa for galaxy searches. For high-redshift galaxies with stellar ages <1 Gyr, we find that disentangling stellar and AGN emission is challenging at restframe UV/optical wavelengths, while high-mass X-ray binaries become more important sources of confusion in X-rays. We propose a color–color selection in the *James Webb Space Telescope* bands to separate AGN versus star-dominated galaxies in photometric observations. We also estimate the AGN contribution, with respect to massive, hot, and metal-poor stars, at driving high-ionization lines, such as C IV and He II. Finally, we test the influence of the minimum BH mass and occupation fraction of BHs in low-mass galaxies on the restframe UV/near-IR and X-ray AGN luminosity function.

Key words: galaxies: active – galaxies: evolution – galaxies: high-redshift

1. Introduction

The first galaxies and black holes (BHs) within the first billion years of the Universe have set the stage for the ensuing evolution of galaxies. Their radiation has shaped the thermal evolution of the intergalactic medium, ionizing the neutral plasma left over after electrons and protons combined to form neutral hydrogen atoms, and making the Universe transparent to UV radiation. The radiative and kinetic feedback exerted by stellar populations and supernovae, as well as by active galactic nuclei (AGNs) powered by the first BHs, has instead shaped the interstellar medium, influencing how stars and BHs evolve in turn, in a sometimes virtuous and sometime vicious cycle, because feedback can both foster or hinder star formation and BH accretion.

Observational evidence on the first galaxies is growing (for a review, see Stark 2016), and there will be a leap forward when the *James Webb Space Telescope* (*JWST*) is launched in the imminent future. Bright quasars have also been detected at similar cosmic epoch, when the Universe was younger than a billion years (Fan 2001; Fan et al. 2006; Venemans et al. 2013; Bañados et al. 2016; Jiang et al. 2016; Matsuoka et al. 2016), while the population of fainter quasars is still small (Willott et al. 2010; Matsuoka et al. 2016). Currently, high-redshift galaxies and quasars are studied almost separately. We know, however, that in the local Universe, between a completely star-dominated galaxy and a quasar, all sorts of shades are possible. Faint AGNs are now identified even in many dwarf galaxies (for a review, see Reines & Comastri 2016).

At high redshift, $z \gtrsim 6$, the presence or absence of AGNs in the bulk of galaxies is a subject of debate, with few convincing candidates to date (Treister et al. 2013; Giallongo et al. 2015; Weigel et al. 2015; Cappelluti et al. 2016; Vito et al. 2016), with searches focused mostly in the X-rays. The small number of bona fide AGN in Lyman-break galaxies is, however, consistent with theoretical expectations when realistic assumptions are made: Volonteri & Reines (2016) predict that BHs might just be smaller/fainter and below the detection limits when adopting BH-stellar mass relations appropriate for low-mass galaxies (Reines & Volonteri 2015).

The recent discovery of high-ionization lines in UV-selected galaxies is opening a new way of searching for AGNs, or at least of interpreting the relative role of AGNs and hot stars as the powering mechanism. Stark et al. (2015b) revealed the detection of C IV in 1 of 11 known Lyα emitters at $z > 7$, while Sobral et al. (2015) and Bowler et al. (2017) discuss the detection of He II and O III] in a bright Lyα emitter at $z = 6.6$. Mainali et al. (2017) revealed C IV in a gravitationally lensed Lyα emitter at $z \sim 6$. In particular, Mainali et al. (2017) showed that UV line ratios can be used to distinguish between AGNs and massive hot metal-poor stars as a powering mechanism using dedicated emission line models (Feltre et al. 2016). Mainali et al. (2017) argued that based on the presence of C IV and lack of He II, this source was likely to have a break in the ionizing spectrum between 47.9 and 54.4 eV, consistent with a stellar ionizing spectrum and inconsistent with an AGN power-law spectrum. The fact that they detect strong O III] provided further evidence in favor of

metal-poor hot stars, as an AGN spectrum would likely have weaker O III] given that oxygen is triply ionized.

In this paper we follow up on this issue by estimating how frequently a high-z galaxy hosts a BH, an active BH, and we examine the relationship between the two, not only in terms of physical properties, but also in terms of observability. Pacucci et al. (2015, 2016) and Natarajan et al. (2017) recently analyzed the observational properties of "seed" BHs in primeval galaxies; we here take a broader view, moving to later cosmic times and including more mature galaxies and BHs. Hainline et al. (2011) and Hainline et al. (2012) can be seen as observational counterparts to this paper: instead of *assuming* that galaxies are unimportant for quasars and AGNs are unimportant for galaxies, we *assess* their relationship within a statistically relevant population.

In this paper, the first of a series, we present our method and a survey of the main results at $z = 6$. We first calibrate our model, and provide an interpretation tof existing observations in its light. We then advance predictions for galaxies and AGNs that may be observed with the *JWST*, and finally compare the power of optical/near-IR observations versus X-rays in recovering the population properties.

2. Method

We create a population of galaxies, BHs and AGNs, starting from the galaxy stellar mass function. Each galaxy is assigned a metallicity and a star formation rate based on empirical relations with stellar mass (see Section 2.2). Black holes are assigned to galaxies by assuming a relation between black hole and stellar mass, and a luminosity to an active black hole by assuming a duty cycle and a distribution of accretion rates for black holes. The duty cycle is defined as the fraction of black holes with an Eddington ratio above 1%, and the distribution of accretion rates pertains only to these black holes, i.e., those with an Eddington ratio above 1%. The model is constrained by matching the AGN and galaxy luminosity function (see Section 2.1). The model is presented in this paper at $z = 6$ only. It is not conceived as an evolutionary, but as an empirical model; this means that the free parameters need to be fit independently for each redshift where sufficient information for fitting the free parameters is available.

We follow the approach of Volonteri & Reines (2016) here, which is based on Schulze & Wisotzki (2011). We denote BH masses as $\mu = \log M_{\rm BH}$, stellar masses s, with $s = \log M_*$, and the AGN luminosity as $l = \log L_{\rm AGN}$. We adopt a simple functional form for the scaling between BH mass and galaxy stellar mass, $\mu = \gamma + \alpha s$, with lognormal intrinsic scatter σ_μ.

Specifically, we adopt the relationship found by Reines & Volonteri (2015) for moderate-luminosity AGNs, typically in lower-mass host galaxies:

$$\mu = (1.05 \pm 0.11)(s - 11) + (7.45 \pm 0.08), \qquad (1)$$

which was shown to produce BH populations in agreement with constraints by Volonteri & Reines (2016). σ_μ was measured in Reines & Volonteri (2015) to be 0.55 dex, but here we leave σ_μ as a free parameter to be set by fitting the AGN luminosity function (LF). We do not include any redshift evolution, because it is still unclear whether there is an evolution, and whether it is in normalization or slope. This is still discussed both theoretically (Volonteri & Stark 2011; DeGraf et al. 2015; Volonteri et al. 2016; Beckmann et al. 2017) and observationally (Ding et al. 2017, and references therein).

For a given galaxy mass, we assign a BH mass based on Equation (1) and a luminosity through the probability distribution of the logarithmic Eddington ratio λ, recalling that $l = 38.11 + \lambda + \mu$. We consider here a lognormal distribution, motivated by observational (Kauffmann & Heckman 2009; Lusso et al. 2012) and theoretical arguments (Volonteri et al. 2016). We refer to Tucci & Volonteri (2017) for an exhaustive discussion and to Jones et al. (2016) for a different perspective. We set the two parameters $\bar{\lambda}$ and σ_λ by fitting the AGN LF. Finally, we consider a duty cycle, $\mathcal{D}$, giving the fraction of BHs that are active. In this case, we define "active" as the fraction of BHs that are accreting at $\lambda \geqslant -2$. In this paper, where we are mainly interested in AGNs around and below the knee of the LF, we set $\mathcal{D} = 0.25$, based on simulation "D" by Habouzit et al. (2017), which best reproduces observational constraints.

Our starting point is the galaxy mass function (MF), and we create a Monte Carlo simulation of the galaxy+AGN population by assigning a BH mass (μ) and luminosity (ł) through the Eddington ratio (λ). We also assume that only a fraction $\mathcal{D}$ of the BHs are in a luminous phase. With this approach, we can build LFs and samples matching mass/luminosity cuts.

A fraction of AGNs are obscured, and they are missed by observations. When necessary, we include a luminosity-dependent correction[5] for the obscured fraction based on Ueda et al. (2014). The obscured fraction appears to increase with redshift up to at least $z \sim 2$–3 (e.g., Merloni et al. 2014; Vito et al. 2014; Buchner et al. 2015). The evolution at higher redshift is less well constrained, and appears, perhaps, to saturate at $z \sim 4$ (Liu et al. 2017), but it may be affected by incompleteness at faint luminosities (Vito et al. 2017), making a robust assessment hard. For simplicity, we adopt the obscured fraction at $z \sim 2$ also at higher redshift, $z = 6$, but based on the studies above, this may underestimate the obscured fraction. The correction we adopt is based on an X-ray sample, and we use it also in optical/UV. Merloni et al. (2014) show that in about 70% of the sources they study, the optical and X-ray classification of obscured/unobscured AGNs agrees. For the remaining 30%, X-ray unobscured AGNs are obscured in optical/UV at low luminosities, while optical/UV unobscured AGNs may have absorbed X-ray spectra at high luminosities. At low luminosities, the optical/UV obscured characterization is likely induced by line-emission dilution into the dominant host-light, while at high luminosities, the X-ray obscuration may be induced by a higher gas component inducing the obscuration (which does not affect the optical range). When we compared to the X-ray LF, we did not correct for Compton-thin AGNs, with column density $N_{\rm H} = 10^{22}$–10^{24}cm^2, as obscuration is thought to be negligible at $z \gtrsim 6$, although obscuration at the level of $N_{\rm H} = 10^{23}$–10^{24}cm^2 cannot be ruled out. We did correct for Compton-thick sources, however, with column density $N_{\rm H} > 10^{24}$cm^2, which would still be missed. Such a population is expected to account for 30%–50% of the AGN population, based on hard X-ray observations at lower redshift

[5] Recent results, based on a local *Swift*/BAT sample, suggest that the obscuration fraction is related to the Eddington ratio and not to the luminosity (Ricci et al. 2017a). As the authors acknowledge, higher gas fractions and a more turbulent medium in high-z galaxies may induce deviations from this local result.

(Ueda et al. 2014, and references therein) and a synthesis of the X-ray background (Gilli et al. 2007). Specifically, we here follow Ueda et al. (2014) for self-consistency with the absorbed fraction.

In most of this work, we assign a BH in each and every galaxy, although BHs are not necessarily expected to be ubiquitous in low-mass galaxies, and we have adopted a minimum BH mass of $10^2\,M_\odot$. These assumptions should be treated as free parameters, and in Section 3.3 we discuss their importance and variations on the basic models.

Several different measurements and analytical fits to the galaxy stellar MF can be found in the literature. Many of them are summarized in Behroozi et al. (2013), Madau & Dickinson (2014), and Stark (2016) for a focus on high redshift, $z > 4$, where differences and uncertainties are discussed (see Figure 11 in Madau & Dickinson 2014). In the following, we use the MF by Song et al. (2016) as a reference, extrapolated down to $10^5\,M_\odot$ and up to $3 \times 10^{12}\,M_\odot$. The choice of the low-mass end does not impact the results, in the sense that the luminosity functions are not strongly dependent on the minimum value of stellar mass, as long as there is no change in the slope of the MF at the faint end. A slope change like this has not been reported in the literature so far (but the LF is eventually expected to bend). The bright end of the AGN LF is sensitive to the maximum galaxy mass we consider. In practice, we have to include galaxies at least up to the stellar mass at which the stellar mass function, multiplied by the duty cycle, drops to levels consistent with the bright end of the AGN mass function. In this paper we focus on a specific redshift, $z = 6$, as this is the highest redshift for which we have statistical information on the AGN LF. The galaxy LF/MF are currently measured out to $z = 8$, and the model can be extrapolated to the same redshift in order to make predictions for future observations.

2.1. AGN Spectra

In this paper we study the properties of BHs with masses and accretion rates that span a wide range, often covering regions of the parameter space that are far from the typical "quasars" that are used as a benchmark to derive mean spectral energy distributions (SEDs; e.g., Elvis et al. 1994; Vanden Berk et al. 2001; Richards et al. 2006; Elvis et al. 2012) that are then used to derive fits for bolometric corrections (Marconi et al. 2004; Hopkins et al. 2007; Lusso et al. 2012). See Jensen et al. (2016) for a first attempt at using observations to constrain AGN redshift spectral evolution.

We therefore create theoretical AGN spectra that explicitly depend on the physical parameters BH mass (μ) and Eddington ratio (λ), inspired by the Shakura-Sunyaev solution (Shakura & Sunyaev 1973). The mass and accretion rate also determine the total luminosity, $l = 38.11 + \lambda + \mu$. In the classic Shakura-Sunyaev solution, the lower the BH mass and the higher the Eddington ratio, the higher the frequency of the peak: $T_{\rm peak} \propto 10^{(\lambda-\mu)/4}$. We here follow a variant based on the physical models developed by Done et al. (2012). Specifically, we calculate the energy of the SED peak as described in Thomas et al. (2016), but adopt the default functional form of the spectrum used in Cloudy (Ferland et al. 2013):

$$f_\nu = \nu^{\alpha_{\rm UV}} e^{-\frac{h\nu}{kT_{BB}}} e^{-\frac{kT_{\rm IR}}{h\nu}} + a\nu^{\alpha_x}, \quad (2)$$

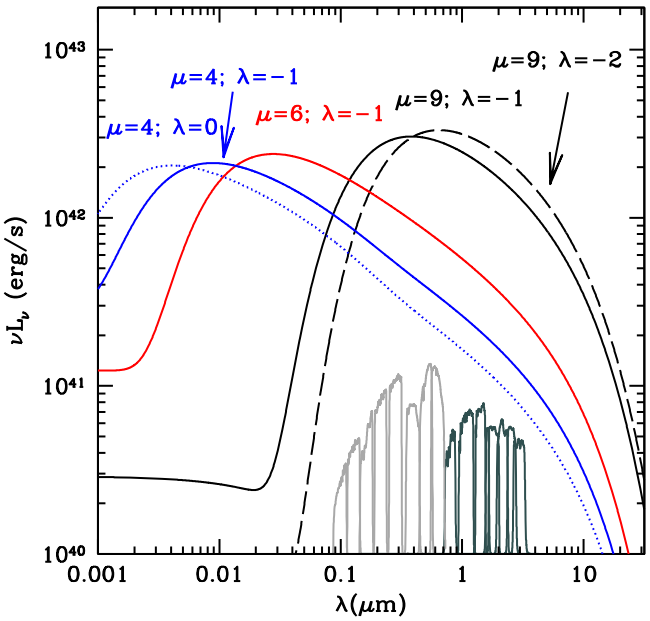

Figure 1. Theoretical AGN spectra, normalized to the same total luminosity. BH masses and Eddington ratios are marked in the figure. Continuum only. We report at the bottom of the figure the NIRCAM (light gray) and MIRI (dark gray) filter curves at $z = 6$ restframe.

with $\alpha_{\rm UV} = 0.5$, $\alpha_X = 1$, and $kT_{\rm IR} = 0.01$ Ryd. Following Ferland et al. (2013), the last term in Equation (2) is set to zero below 1.36 eV (912 nm), and we do not extend the spectrum above 100 keV (12.4 nm) restframe. The normalization of the X-ray component, a, is obtained through $\alpha_{\rm OX}$, the exponent of the power law connecting the continuum between 2 keV and 2500 Å. We assume that $\alpha_{\rm OX}$ depends on mass and Eddington ratio as obtained using the models by Done et al. (2012), see, e.g., Dong et al. (2012), and differently to Ferland et al. (2013), we include the contribution of the Big Bump, i.e., the pseudo-blackbody in the first term of the right-hand side of Equation (2), to the emission at 2 keV, since for low-mass highly accreting BHs, the bump has a very high $T_{\rm peak}$ and it may contribute to the X-rays.

Examples for some masses and Eddington ratios are shown in Figure 1, where we have normalized the SEDs to the same luminosity to facilitate comparison of the spectral shape. The SED is validated in Appendix A against commonly used bolometric corrections and the SED used for their derivation. With this SED, we can calculate monochromatic luminosities as well as broadband ones from infrared to hard X-rays. In principle, with an SED that depends on the BH physical properties, bolometric corrections become distributions, rather than a fixed bolometric correction at a given luminosity. We do not include attenuation for the AGN SED, although we include it statistically as an obscured fraction, an approach that Ricci et al. (2017b) showed to produce a good match between UV and X-ray AGN LFs. We do not include emission lines either, which can boost both the AGN (Vanden Berk et al. 2001; Hainline et al. 2011) and the galaxy magnitude (Reines et al. 2010; Atek et al. 2011; Stark et al. 2013; de Barros et al. 2014). Nebular emission will be studied in a companion paper.

We fit for the parameters describing the BH/AGN population, $\bar{\lambda}$, σ_λ, and σ_μ by minimizing the χ^2 of the distance between the model and both the UV and X-ray AGN LF. For the UV AGN LF, we consider the functional forms proposed by Willott et al. (2010) and Kashikawa et al. (2015), and for the X-ray AGN LF, the upper and lower limits derived from a combination of LFs and upper limits from Vito et al. (2016) and Marchesi et al. (2016).

We include the luminosity range ~8.8–11.8 $L_\odot$ in X-rays and ~11.8–13.25 $L_\odot$ in UV, i.e., only the observed part of the LF, and we exclude the highest luminosities, as a duty cycle $\mathcal{D} = 0.25$ is not appropriate for the most luminous quasars, which should have $\mathcal{D} \sim 0.75$–1. The set of parameters that best allows us to reproduce both the X-ray and UV LF is $\bar{\lambda} = \log(0.40)$, $\sigma_\lambda = 0.40$, $\sigma_\mu = 0.50$. Small variations on the best set are possible, but they do not change the results overall. The range of parameters and their uncertainties as well as variations on the reference model are discussed in Appendix B.

2.2. Galaxy Spectra

Our starting point is the stellar mass and the redshift. To each galaxy we assign an SED from either the Bruzual & Charlot (BC; Bruzual & Charlot 2003, version 2016) or the BPASS (Eldridge & Stanway 2009; Stanway et al. 2016) models, using the minimum set of parameters that allows us to reproduce the observed galaxy UV LF reasonably well. We adopt a Salpeter initial mass function (Salpeter 1955) for consistency with most high-z studies. We assume constant star formation histories and map stellar mass to age through the galaxy main (or mass) sequence, connecting star formation rate (SFR) to galaxy stellar mass, as formulated for galaxies up to $z = 6.5$ by Salmon et al. (2015). Specifically, at $z = 6$, $\log(\mathrm{SFR}) = 0.54s$–3.9, with an intrinsic dispersion of 0.21 dex, and the age is obtained as the ratio of galaxy stellar mass to SFR. As for most other relations and correlations adopted in this paper, the Salmon et al. (2015) relation is determined for a subset of the mass/SFR range we are considering. We then assign a galaxy to a metallicity bin, either $10^{-2.3}\,Z_\odot$, $10^{-0.7}\,Z_\odot$ or solar assuming a mass–metallicity relation. The results are not strongly dependent on the metallicity grid for the range of metallicity expected for galaxies at $z \sim 6$. Since observational constraints at $z = 6$ are unavailable (but see Bian et al. 2017 for local analogs of high-redshift galaxies), we adopt the theoretical results from Ma et al. (2016b). For completeness, we have also performed a test based on the simple stellar populations of BC, but evolved with the code by Maraston (2005), and found results consistent with the standard BC models. Examples of galaxy and AGN SEDs are shown in Figure 2.

When we study the *intrinsic continuum emission*, we adopt unattenuated spectra and apply a magnitude-dependent attenuation (Meurer et al. 1999) only when estimating observable luminosities. We couple the relationship between extinction and UV slope, $A_{\rm UV} = 4.43 + 1.99\beta$ to the correlation between UV slope and magnitude in (Bouwens et al. 2014) to obtain

$$A_{\rm UV} = (0.58 \pm 0.57) - (0.67 \pm 0.28)(M_{\rm UV} + 19.5), \quad (3)$$

where $M_{\rm UV}$ is the intrinsic unattenuated magnitude obtained from the stellar population. The errors are propagated errors and are assumed to be uncorrelated, and no error on the slope of the relation between $A_{\rm UV}$ and β is reported in the literature. This relation produces values similar to the model with no evolution in the relation between infrared excess and stellar mass of Bouwens et al. (2016). We do not decrease $A_{\rm UV}$ below 0.5 in order to obtain a reasonable fit of the faint end of the galaxy LF. When we study the observable emission at wavelengths other than UV (1550 Å), we correct the galaxy luminosity assuming a dust extinction law (Calzetti et al. 2000), renormalized to obtain Equation (3) at 1550 Å. The uncertainty added in the model by extrapolating this and other relationships (e.g., stellar mass-SFR or mass–metallicity) to higher/lower galaxy masses and luminosities is difficult to assess, but we present most results in the following sections as a function of masses, magnitudes, or luminosities, therefore the range of applicability can be inferred from the information given in this and the previous sections. In Figure 3 we compare stellar masses and UV magnitude from the model to the relation proposed by Song et al. (2016). A better agreement at the high-mass end would be obtained by limiting the galaxy age to 0.25 Gyr.

With this setup, we have created galaxy LFs to compare with current observations (Atek et al. 2015; Bouwens et al. 2015) and anchor our model (Figure 4). We obtain a reasonable match around the knee of the galaxy LF for all models, but we underestimate the bright end at $L > 2 \times L_*$ and $L > 4 \times L_*$ for the LF of Atek et al. (2015) and Bouwens et al. (2015), respectively, by ~0.5 dex and the faint end by <0.35 dex (Table 1). The mismatch at the bright end is caused by an underestimate of the galaxy luminosity (after correction for attenuation), while that at the faint end is caused

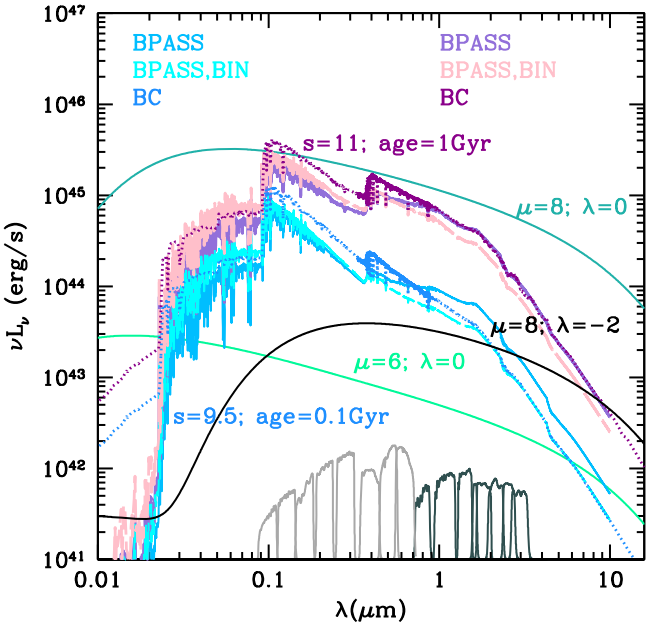

Figure 2. Comparison between galaxy and AGN SEDs. Here the bolometric luminosity of the AGN is the same (black: $\mu = 8$, $\lambda = -2$; turquoise: $\mu = 6$, $\lambda = 0$; and blue-green: $\mu = 8$, $\lambda = 0$), the galaxy mass is $s = 11$ (violet curves) and $s = 9.5$ (blue curves), and the two galaxy spectra are for $Z = 0.2\,Z_\odot$ and different galaxy ages, using the BC or BPASS models with and without binaries, normalized to the same stellar mass. The chosen ages are the expectations from the assumption that the galaxies are on the main sequence, ~0.1 Gyr for $s = 9.5$ and ~1 Gyr for $s = 11.5$. We report at the bottom of the figure the NIRCAM (light gray) and MIRI (dark gray) filter curves at $z = 6$ restframe.

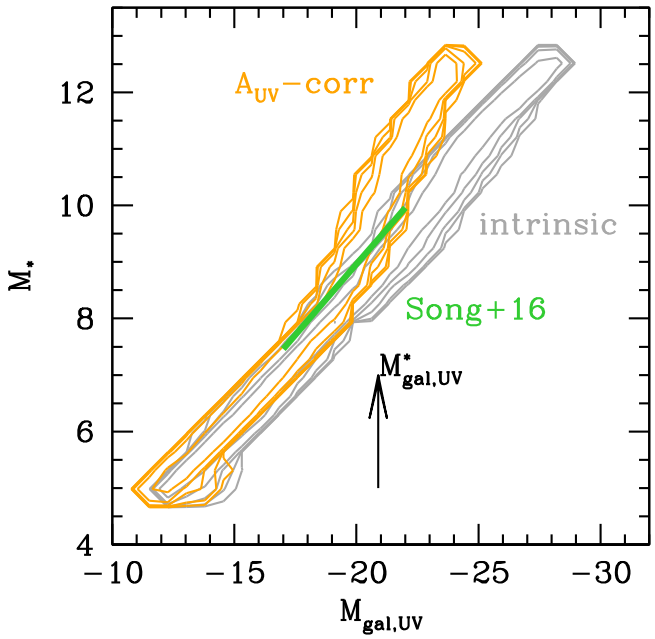

Figure 3. Stellar mass vs. UV magnitude. The gray contours show the intrinsic magnitude, the orange contours the attenuation-corrected magnitude, and the green line is the relation from Song et al. (2016). The arrow shows the location of the break in the galaxy luminosity function.

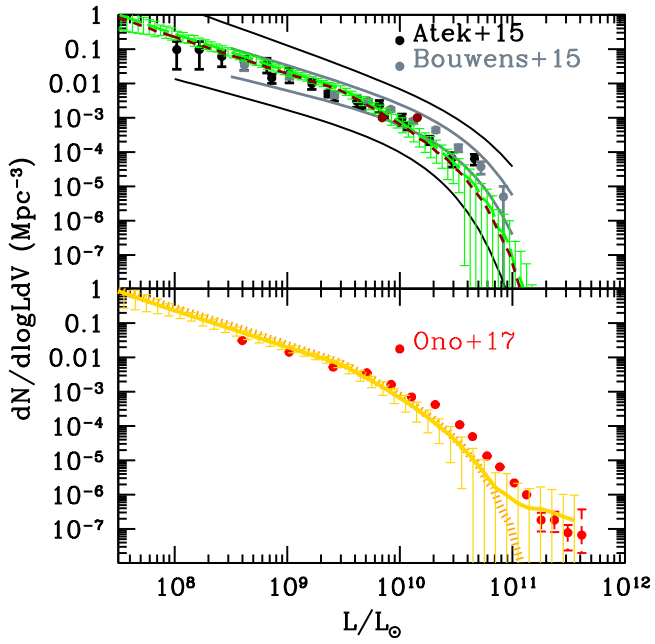

Figure 4. Galaxy LF at 1550 Å. Upper panel: in black and dark gray, we show the error region of the functional forms of the LFs by Bouwens et al. (2015) and Atek et al. (2015) The models are shown as the dotted orange curve for the BC, the short dashed dark red for single stars in the BPASS models, and as long dashed green for the BPASS models with binaries, where we show error bars that are similar for all other cases, but we do not include them for clarity. Lower panel: the yellow curve shows the LF we obtain by summing the galaxy and AGN light, and this should be compared to the red points, which represent the LF in Ono et al. (2017; without correcting for AGN contamination). The orange curve is repeated from the upper panel to guide the eye.

by overestimating the UV luminosity of low-mass galaxies. To improve the match with the faint end of the LF, we would need to either include only older stellar populations (age $\gtrsim 0.1$ Gyr) or set $A_{\rm UV} \geqslant 1.5$, indicating the presence of either non-star-forming galaxies or galaxies with much higher levels of attenuation among low-mass galaxies. The former can be linked to low-mass galaxies being easily affected by stellar (and AGN) feedback as well as to photoionization from the UV background. Our model, based on the mass sequence of star-forming galaxies, does not include this putative population. The extinction we have included in our model is based on that of UV-selected galaxies, which by selection cannot be too dusty. Indications that dusty star-forming galaxies exist at $z \sim 6$ are now seen in ALMA data (Riechers et al. 2013; Ma et al. 2016a; Decarli et al. 2017; Strandet et al. 2017). *Hubble Space Telescope* observations, probing the restframe UV, are biased against quenched or dusty galaxies at high redshift, and the *JWST*, with its optical/near-infrared coverage, will help detect such galaxies. To improve the match at the bright end, we need to assume that galaxies with mass $>10^{10}\,M_{\odot}$ at $z = 6$ are at most 0.25 Gyr old, which is consistent with the stellar ages of slightly less massive galaxies (Curtis-Lake et al. 2013) but is at odds with the age of the stellar populations of some other observed galaxies (e.g., Richard et al. 2011).

Table 1
Reduced χ^2 Separately for the Faint and Bright End of the LF for Models against the Functional Form of Observational LFs

	Bouwens +15 ($L < L_*$)	Bouwens +15 ($L > L_*$)	Atek +15 ($L < L_*$)	Atek +15 ($L > L_*$)
BC	4.11	4.20	4.64×10^{-2}	0.17
BPASS BIN	3.09	3.96	4.21×10^{-2}	0.11
BPASS	2.21	5.19	2.45×10^{-2}	0.25

3. Results

3.1. Galaxy versus AGN: Luminosities and Biases

With the setup described in the previous sections, we have in hand galaxy+AGN populations. We can therefore compare the properties of the two populations. We start with comparing the UV magnitudes[6], and we then move to the *JWST* bands to study which type of AGNs can be more easily detected with future optical/near-IR observations.

In Figure 5 we compare galaxy and AGN magnitudes in the UV (we adopt 1550 Å throughout, although most quasar studies use 1450 Å as a reference wavelength), where we have used Meurer-corrected BC models (the results are qualitatively unchanged for all other stellar population models). We have not corrected the AGN luminosity for attenuation, making this an "optimistic scenario." The currently known population of high-z quasars, detected in shallow surveys, is dominated by high-mass BHs with high accretion rates and a high ratio between BH and galaxy mass. This bias is expected, as a fixed high-luminosity selection picks high-mass BHs that are hosted in low-mass galaxies more frequently than vice versa because of the steep shape of the galaxy mass function (Shields et al. 2006; Lauer et al. 2007; Volonteri & Stark 2011). The population of the first BHs, possibly close to the seed mass, is much harder to disentangle from the host galaxy if it actively forms stars. We refer to Natarajan et al. (2017) for a dedicated study.

[6] We calculate UV magnitudes as monochromatic magnitudes at 1550 Å, while for the NIRCAM and MIRI bands, we convolve the SED with the filter response.

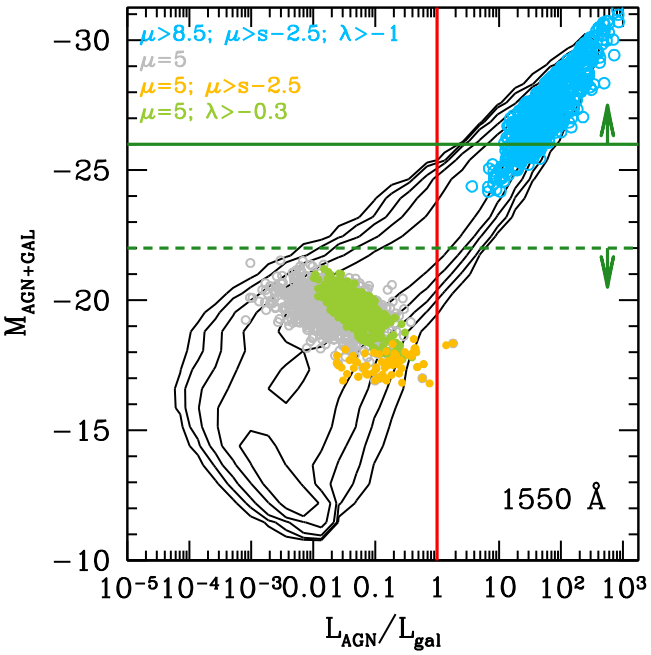

Figure 5. Comparison between galaxy and AGN UV properties. The red line marks equal luminosity in the AGN and in the stellar population. The green solid line is the typical faint limit for quasar searches, while the dashed line marks the typical bright galaxies detected so far at $z \sim 6$. The cyan points show the typical physical properties of the currently detectable quasars, i.e., above the flux limit and above the galaxy luminosity. The gray points mark galaxies hosting BHs with $4.75 < \mu < 5.25$, and the green and orange points are subsets of this population with either high accretion rates or a high ratio between BH and galaxy mass, respectively.

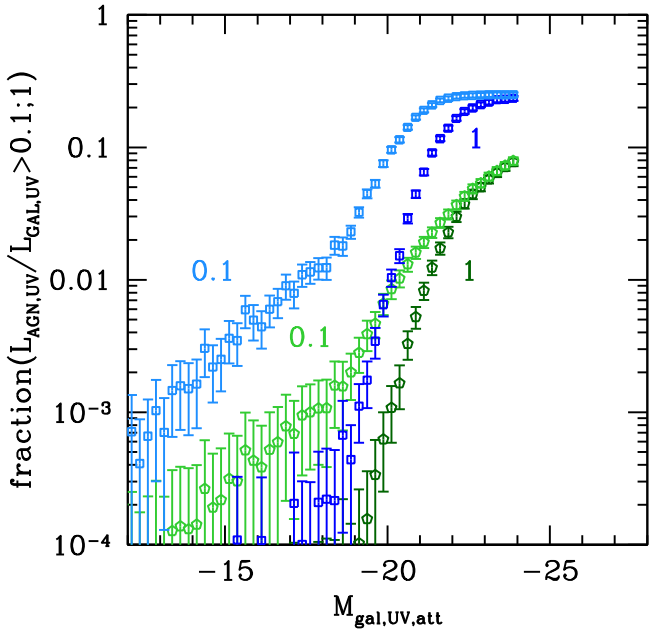

Figure 6. Fraction of galaxies hosting an AGN with luminosity >0.1; 1 times the galaxy luminosity. Here the blue squares represent the intrinsic fraction, while the green pentagons correct for the AGN-obscured fraction. We recall that by construction, only 25% of galaxies host an active black hole in our model, which is taken into account in this figure.

Note also that by selecting sources with a total magnitude brighter than −26 (observations do not have an a priori AGN/galaxy separation), all sources are dominated by the AGN. This explains why no "pure" galaxies have been identified in quasar searches that use photometric information from large-shallow surveys as a starting point. Conversely, since most galaxy searches are based on narrow-deep surveys, galaxies are typically fainter than −22, and in this case, the AGN contribution is small, in agreement with observations (Treister et al. 2013; Giallongo et al. 2015; Cappelluti et al. 2016) and theoretical models (Habouzit et al. 2017; Volonteri & Reines 2016). The recent surveys SHELLQ and GOLDRUSH, interestingly, select some of the most luminous galaxies as well as some of the faintest quasars at $z \sim 6$ (Matsuoka et al. 2016, 2017; Ono et al. 2017). These surveys bridge the region where we predict galaxies and AGN/quasars coexist, and in fact, Ono et al. (2017) find a significant AGN contamination at the bright end of the galaxy UV LF, see Figure 4, in agreement with our model.

The fraction of galaxies hosting an AGN with a luminosity higher than the galaxy UV luminosity, or 10% of it, is shown in Figure 6 for the same model. Intrinsically, a fraction ∼20% of galaxies with UV magnitude ∼−21 hosts an AGN with a UV luminosity >10% of the galaxy, but taking into account the obscured fraction (type 1 versus type 2 AGNs), the detectable fraction decreases by about an order of magnitude. We show here the result obtained with BC models; they are statistically indistinguishable for BPASS models.

For the *JWST* bands, Figure 1 shows that since at high BH masses the spectrum peaks at redder frequencies, high-mass BHs, with the SEDs peaking at ∼1 μm restframe, would be favored over low-mass BHs, where the 1–5 μm observer frame band at $z = 6$ samples a region far from the peak of the SED. Additionally, given the inverse dependence of the peak energy and the Eddington ratio, at fixed BH mass, the SED *shape* favors low Eddington ratios, but the SED *normalization*, via the total luminosity, favors a high Eddington ratio. The latter is the most important of the two at fixed BH mass. However, at fixed total luminosity, proportional to $10^{\mu+\lambda}$, the shape of the SED is dominant in the *JWST* bands. For instance, the luminosity at 1–28 μm is higher for a BH with $\mu = 8$ and $\lambda = -2$ than for a BH with $\mu = 6$ and $\lambda = 0$, although their bolometric luminosities are the same. In summary, at fixed *intrinsic bolometric luminosity, JWST* will be biased toward detecting high-mass BHs, even with low intrinsic accretion rates, over low-mass BHs.

However, we are interested in disentangling AGNs from their host galaxies. A more interesting assessment is on the relationship of AGN and galaxy as a function of physical properties, see for instance Figure 2. We are biased to identify BHs with a low Eddington ratio that are overmassive with respect to their host galaxy, e.g., $s = \mu + 1.5$, with $\mu = 8$, $\lambda = -2$ with respect to BHs with a high Eddington ratio that are "normal" with respect to their host galaxy, e.g., $s = 9.5$, with $\mu = 6$, $\lambda = 0$ in the example of the figure.

Finally, the contrast between AGN and galaxy increases at redder wavelengths, given the young ages of galaxies at $z \gtrsim 6$, therefore AGN searches will be favored at red wavelengths, as shown in Figure 7, where we compare galaxy and AGN magnitudes in a NIRCAM band (F200W, the most sensitive for point sources) and in a red MIRI band (F1800W, more sensitive than F2100W or F2550W). We note also that when we assume AGN dust obscuration, the AGN-to-galaxy contrast will be even higher, enabling the use of shorter-wavelength MIRI filters. The *JWST* can reach an absolute magnitude of about −15, with −12 reachable in the case of strong-lensing

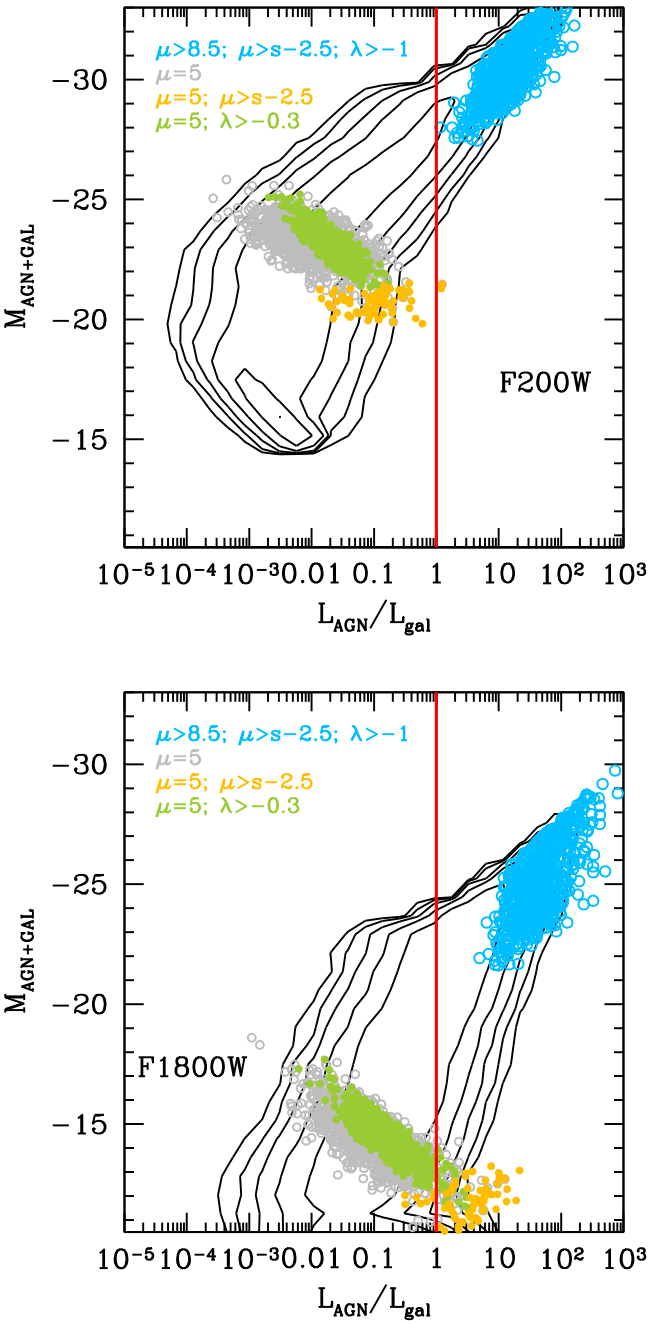

Figure 7. Comparison between galaxy and AGN magnitudes in a NIRcam (F200W) and a MIRI band (F1800W). The red line is the 1:1 relation, and the points highlight the same populations as in Figure 5. AGNs powered by small BHs are more easily identifiable at the reddest wavelengths.

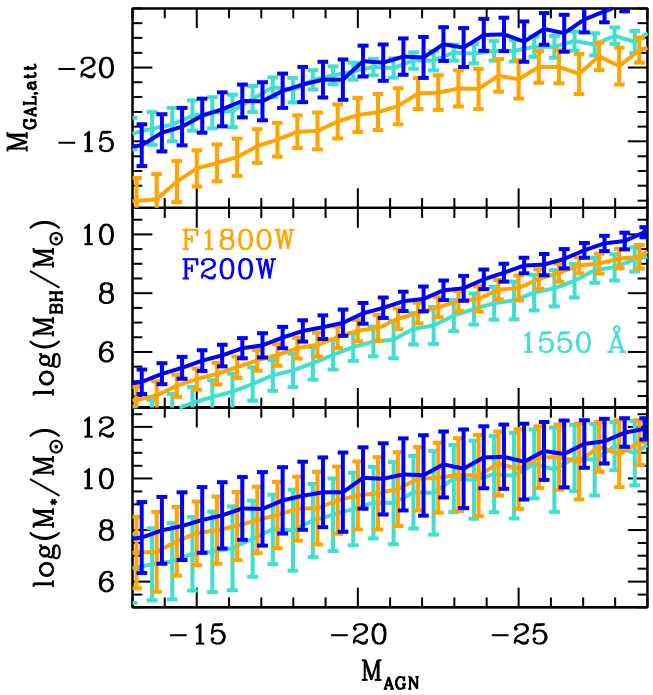

Figure 8. Properties of BHs and galaxies as a function of the AGN luminosity, weighted by number density. The *x*-axis is the non-attenuated AGN magnitude at various wavelengths, 1550 Å(UV), a short-wavelength NIRCAM band (F200W), and a long-wavelength MIRI band (F1800W). The three panels show the logarithmic mean of the galaxy stellar mass (bottom), BH mass (middle), and galaxy magnitude (top) and 1σ dispersion for each quantity.

clusters where we hope to gain a boost of three magnitudes in some areas, provided we observe enough clusters. We have used attenuated galaxy models, but we have not corrected the AGN luminosity for attenuation, making this an optimistic scenario from the point of view of AGNs.

The relationship between BH, galaxy, and their magnitudes is summarized in Figure 8. Here we calculate in bins of non-attenuated AGN magnitude the mean μ, s, and galaxy magnitude, weighted on their number density. The latter takes into account the rarity of the most massive galaxies: a quasar with $M_{UV} = -26$ intrinsically has a higher probability of being hosted in a galaxy with $s = 12$, but such galaxies, if they exist, have a very low number density ($dN/d\log M \sim 10^{-14}$ Mpc^{-3} at $s = 12$ using the mass function of Song et al. 2016) and therefore a low probability of being detected. Weighting by the number density from the MF, which gives the probability of having a galaxy of a given mass in a given volume (i.e., a field at a given redshift), the weight shifts the expectation mass to lower mass, more common galaxies, therefore making the most probable detectable host of a quasar with $M_{UV} = -26$ a galaxy with $s \sim 10.5$. For $M_{\rm AGN} \sim -26$ at 1550 Å, a reference value for the currently known population of $z \sim 6$ quasars, our models predict that $s \sim 10.5$, $\mu \sim 8.5$ and $M_{\rm GAL,att} \sim -21.5$. By including attenuation for the quasar magnitude using the same $A_{\rm UV}$ as we use for galaxies, at an attenuated $M_{\rm AGN} \sim -26$ at 1550 Å, then $s \sim 11$, $\mu \sim 9.3$, and $M_{\rm GAL,att} \sim -22$.

We have explored several color–color combinations that can be used to distinguish bona fide AGNs from galaxies. A possible distinction is the presence of an actively accreting BH with a sizeable mass with respect to the stellar component, e.g., $\lambda > -0.5$ and $\mu > \langle\mu|s\rangle$. At the other end, we envisage a slowly accreting BH with a low mass compared to the stars, e.g., $\lambda < -1.5$ and $\mu < \langle\mu|s\rangle$. In terms of AGN-to-stellar bolometric ratios, this distinction selects the upper cloud of the distribution, but it does not necessarily imply that $L_{\rm AGN} > L_{\rm GAL}$; the proposed cuts roughly correspond (in color–color space) to $L_{\rm AGN} > 0.1\,L_{\rm GAL}$ versus $L_{\rm AGN} < 0.001\,L_{\rm GAL}$. With this practical definition, we searched color–color combinations to find the galaxies that are most promising for a distinction with stronger or weaker AGNs. The best selection we found is a combination of the shortest wavelength NIRCAM filters (F090W to F200W; F070W is also an option, but it has a much lower sensitivity) and

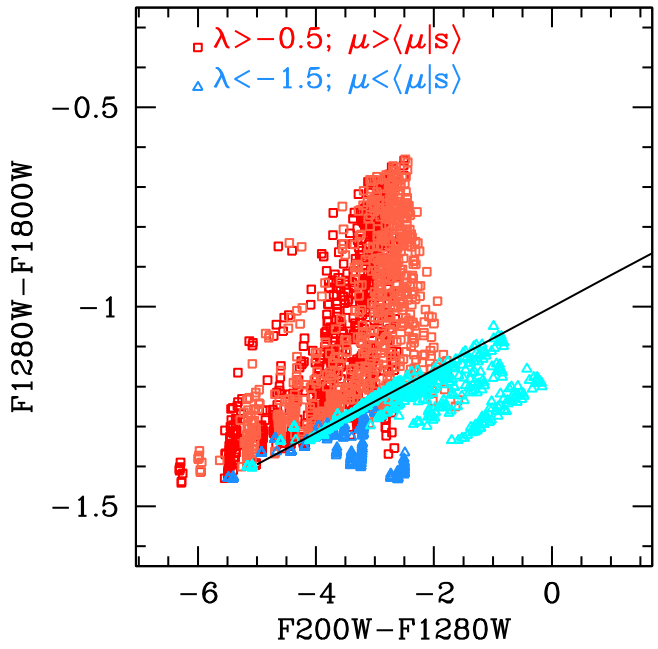

Figure 9. Most promising color–color selection to separate AGNs from galaxies. The blue tracks show colors for pure AGNs with different masses and accretion rates ($4 \leqslant \mu \leqslant 10$; $-2 \leqslant \lambda \leqslant$, as marked in the figure). We recall that the peak temperature of the SED scales as $10^{(\lambda-\mu)/4}$, therefore different combinations of mass and accretion rates can produce the same color. The black curves show galaxy colors (younger to older from left to right) for all the stellar models used in this paper, without including attenuation, while the gray curves include a correction for attenuation as described in Section 2.2. The dark and light red squares show AGN-dominated galaxies with and without attenuation, respectively; the blue and cyan triangles show instead star-dominated galaxies with the same convention for attenuation. A color–color selection F1000W-F2100W vs. F070W-F1000W gives similarly strong diagnostics.

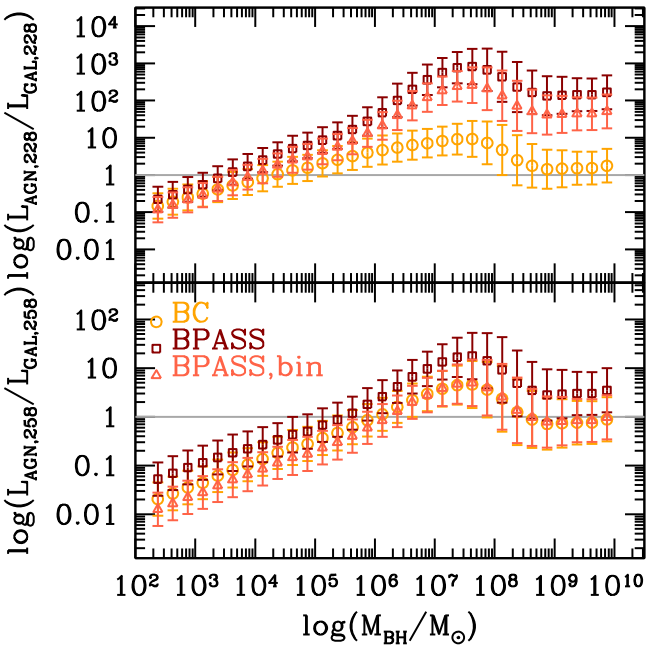

Figure 10. Ratio of AGN and stellar contribution to the emission at 228 Å (54.4 eV) and 258 Å (47.9 eV). We compare BC models to BPASS models with and without binaries.

the longest wavelength MIRI filters (F1800W,; in this case, the choice is also a compromise between clean selection and sensitivity, F2100W and F2550W having a much lower sensitivity but better discriminating power) with an intermediate-wavelength one, namely F1280W, F1000W, or F770W (see Messias et al. 2012, 2014; for the latter choice). These combinations have the least amount of overlap between the two populations, and this is true in both cases, where we correct for attenuation both the galaxy and the AGN or either one of them. For instance, for the example shown in the figure, when we use as a dividing line F1280W − F1800W = 0.08(F200W−F1280W) −0.99, 15% of the red (stronger AGN) points fall below the line, and 19% of the blue (weaker AGN) points fall above the line. The fraction of misclassified sources is similar for the other combinations proposed above. In contrast, with combinations such as F200W-F444W versus F444W-F1800W or F200W-F1500W versus F1500W-F1800W, the fraction of misclassified sources is between 25% and 50%.

As a proof of concept, we corrected AGNs for reddening using the same extinction curve as we used for galaxies, and found that the combinations above remain the cleanest options. We report in Figure 9 tracks of pure galaxies and AGNs as well as a selection from our galaxy+AGN population that matches the above criteria for F1280W-F1800W versus F200W-F1280W. The results might change, however, when gas and dust are included (which will be explored in more detail in another paper).

3.2. Galaxy versus AGN: UV Radiation Powering Emission Lines

Some of the outstanding questions regarding the high-redshift Universe are whether AGNs or galaxies are responsible for reionization, and how many galaxies host an AGN. We will specifically address the first question in a companion paper, but in this study, we start by comparing galaxy and AGN emission in the UV, and we specifically focus on two energies, 54.4 eV and 47.9 eV (228 and 258 Å). Photons with these energies or higher are required to produce the He II and C IV lines that are now observed in high-z galaxy spectra (Stark et al. 2015a, 2015b, 2017; Mainali et al. 2017). In Figure 10 we calculate the median of the ratio of the monochromatic luminosity produced by the AGN and produced by the galaxy stellar population as a function of the BH mass. This is the intrinsic luminosity, i.e., it does not include attenuation.

At 228 Å (He II), the AGN contributes significantly, and sometimes dominantly, in galaxies with mass $8 < s < 11$ because these galaxies normally host BHs with mass $5 < \mu < 8$, where the spectrum has a strong UV component. At lower BH masses, the spectrum becomes too hard, and at higher BH masses, it becomes too soft (cf. Figures 1 and 2). The AGN contribution is lower at 258 Å (C IV), but still significant, above 10%, for galaxies with mass $s \sim 10$ and BHs with mass $\mu \sim 7$. Caution should be taken in interpreting these results quantitatively, as stellar models in the far-ultraviolet are not well calibrated and models are significantly different. The BC models do not tag the Wolf–Rayet phase in the tracks, but for these hot stars, they instead use the default planetary nebulae spectra from Rauch (2002), which are very similar to blackbody spectra. These may overestimate the flux of He ionizing spectra. The BPASS models instead include Wolf–Rayet stars; these can be formed by mass transfer in binaries

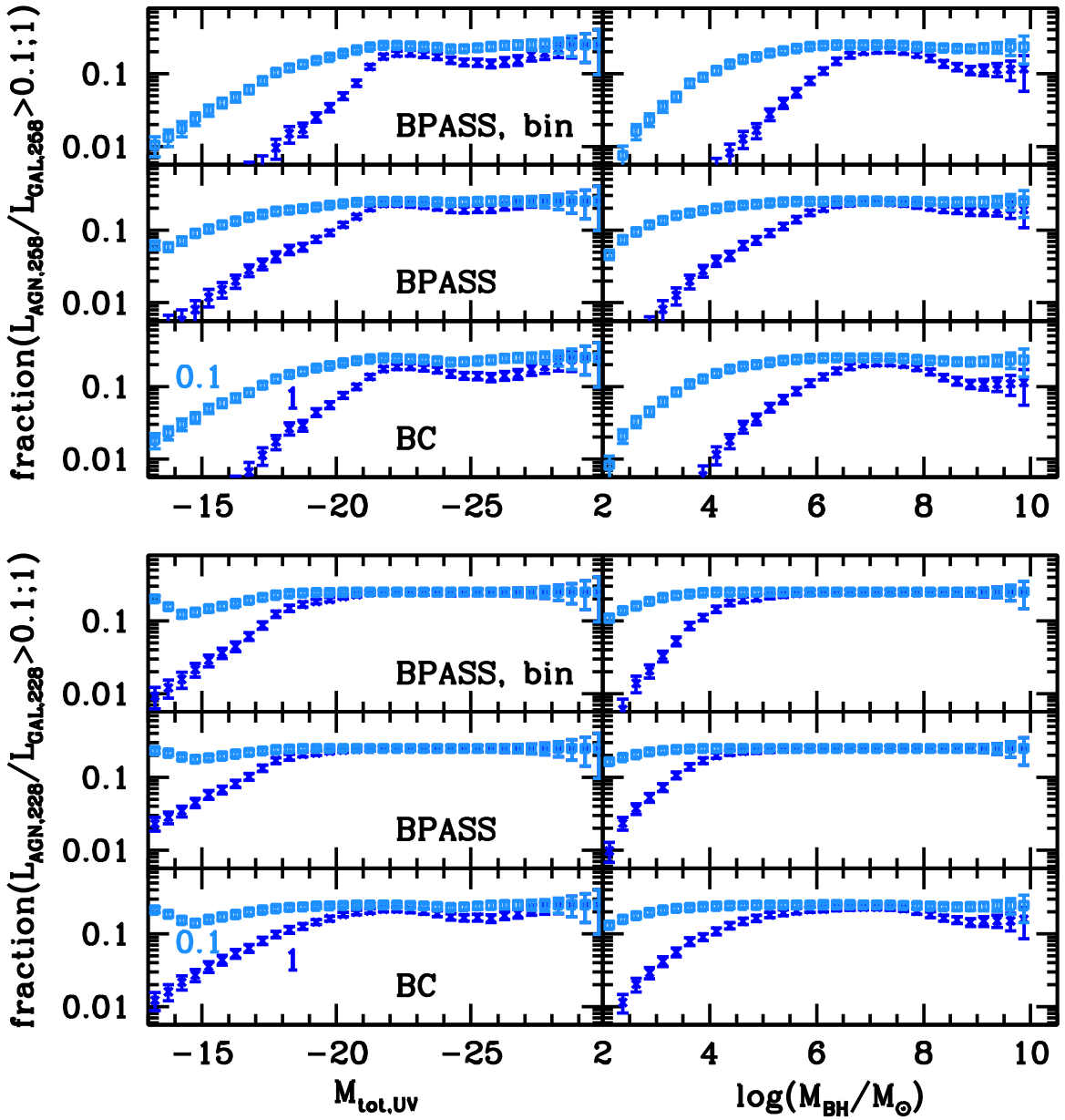

Figure 11. Fraction of galaxies as a function of total UV magnitude (galaxy+AGN, left) or BH mass (right) hosting an AGN with luminosity higher than a fraction 0.1; 1 or of the stellar luminosity at 228 Å (54.4 eV, He II, bottom) or 258 Å (47.9 eV, C IV, bottom). At a magnitude of ~ -20, the fraction of galaxies hosting an AGN contributing more than 10% of the emission is ~25% for He II and 20% for C IV. We recall that we assumed that only 25% of galaxies host an active BH, therefore almost all of the galaxies with an AGN have an AGN contribution >10% at these energies.

and not just the normal Conti mechanism. It is important, however, to draw attention to the possibility that AGN contribution in driving these lines could be substantial, even for young galaxies, and even including stellar binaries.

The previous figures showed the emission contribution when a galaxy contains an active BH, but when estimating the fractional contribution of the AGN emission for a full population, we have to consider that only 25% of the galaxies are assumed to host an active AGN. Figure 11 takes this into account, showing as a function of attenuation-corrected galaxy UV magnitude (without including the AGN) the fraction of galaxies hosting an AGN with a luminosity higher than 0.1; 1 times the galaxy luminosity at the same wavelength. At a magnitude of ~ -20, the fraction of galaxies hosting an AGN contributing more than 10% of the emission is ~20%–25% at 228 Å and >10% at 258 Å.

3.3. Luminosity Functions and Dependence on "Seed" BH Properties

Our models are calibrated by requiring a good match with the region around the knee of the AGN LF in X-ray and UV. The faint end is an unexplored territory at $z \gtrsim 6$, and we investigate here the dependence on BH "seed" properties, namely the minimum BH mass and the occupation fraction (OF) of BHs as a function of galaxy mass, as well as synergies with X-ray observations.

X-rays are usually considered a "clean" way to select or confirm AGNs, as they are less contaminated by the host galaxy. The main source of confusion are high-mass X-ray binaries (HMXBs, e.g., Lemons et al. 2015), which are expected to be abundant in highly star-forming galaxies (Mineo et al. 2012). The contamination, therefore, is likely to

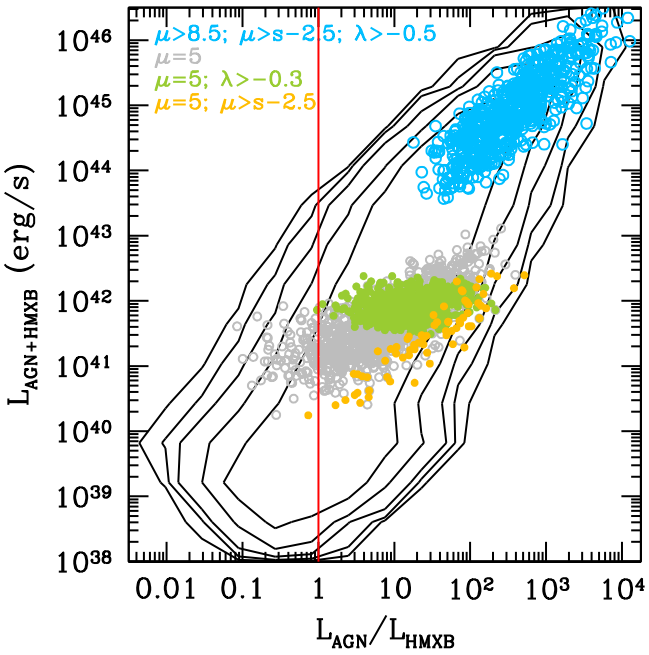

Figure 12. Comparison between AGN luminosity vs. luminosity in HMXBs in the host galaxy at 2–10 keV. Symbols are the same as in Figure 5.

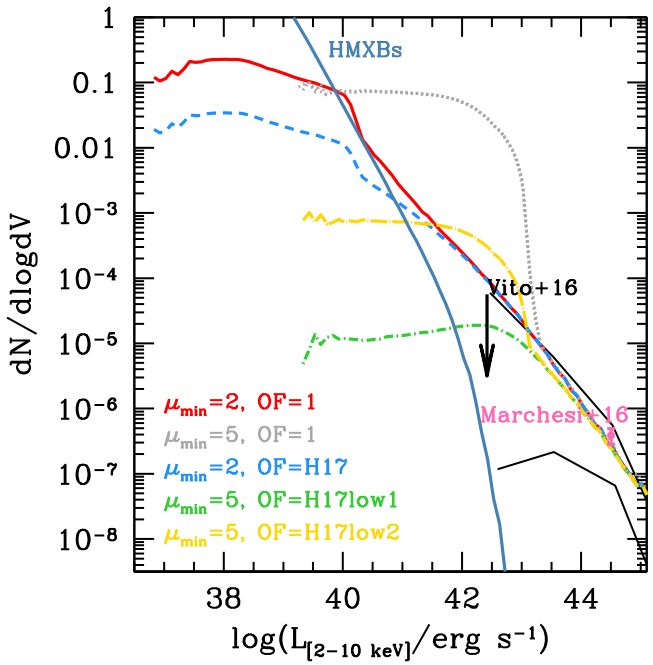

Figure 13. AGN LF at 2–10 keV. The various curves explore different minimum BH mass or occupation fraction, as marked in the figure. The red and gray solid curves assume that every galaxy hosts a BH with minimum mass $\mu_{\rm min} = 2$ and $\mu_{\rm min} = 5$; the blue dashed curve assumes the occupation fraction derived for "light seeds" by Habouzit et al. (2017); the green short dash–dotted curve assumes $\mu_{\rm min} = 5$ and the same functional shape as the occupation fraction, but shifted by a decade in stellar mass and extended to zero (H17low1); the yellow long dash–dotted curve does not decrease the occupation fraction below 0.01 (H17low2). The black curves are upper and lower limits to the AGN LF from Vito et al. (2016), while the pink dot marks the point on the LF derived from the data in Marchesi et al. (2016). The blue-green solid curve shows the LF of HMXBs, derived from model 269 in Fragos et al. (2013) for the galaxy population in our model. On the left of this curve, confusion by HMXBs would hamper an identification of AGN.

be more important in young star-forming galaxies at $z \gtrsim 6$. We assess the level of confusion by estimating the total luminosity of HMXBs as a function of SFR and redshift based on the scaling at 2–10 keV from model 269 in Fragos et al. (2013), shown by Lehmer et al. (2016) to match observations up to $z \sim 2.5$ very well, where good constraints from the X-ray data are available and represent the best characterization of the scaling relations at all redshifts. The ratio of HMXB emission to SFR in this model and in observations increases with redshift, therefore the difficulty of disentangling AGNs from HMXBs is increased at high redshift, $z \sim 6$, as is evident in Figure 12.

A comparison between Figures 12 and 7 shows that a significant fraction of faint AGNs, powered by low-mass BHs, can hardly be distinguished from the collection of HMXBs in the host galaxies, but this fraction is still much smaller than that where confusion from stellar population at optical/near-IR wavelengths is important, cf. the gray points in the two figures.

The same can be seen in the LF (Figures 13 and 14), where selecting for AGNs with a luminosity greater than that of the galaxy (the stellar population in optical/near-IR bands, HMXBs in X-rays), as a proxy for the population of sources more easily identifiable as AGN, significantly reduces the luminosity range that can be probed (compare the dotted curves, all AGNs, with the solid curves of the same color, uncontaminated AGNs).

In Figure 13 we compare different models, where we vary the minimum BH mass and the OF based on theoretical models of BH formation (for reviews, see Volonteri 2010; Haiman 2013; Natarajan 2014). Our basic model places a BH in every galaxy, with $\mu_{\rm min} = 2$, a mass typical of "light seeds," such as the remnants of Population III stars and those formed by dynamical interactions in low-metallicity stellar clusters. In a variant, "H17," we assume the occupation fraction derived by Habouzit et al. (2017) in a dedicated cosmological hydrodynamical simulation (simulation "D"):

$$\mathrm{OF_D} = 1. - \frac{0.85}{1 + (\mu/\varepsilon)^{\beta}} \quad (4)$$

$$\varepsilon = -0.077(1+z) + 7.71 \quad (5)$$

$$\beta = 2.30(1+z)^{1.32}. \quad (6)$$

When the requirement is included that the AGN outshines the HMXBs, these two models are indistinguishable.

To mimic the existence of more massive "heavy seeds," we include models with $\mu_{\rm min} = 5$ and either OF = 1 (using the H17 occupation fraction results are identical to OF = 1) or variations on "H17" that take into account that the production of more massive BH seeds is rare and should occur in halos more massive than those where light seeds form (Habouzit et al. 2016, and references therein). For this, we shift the functional form of the H17 OF by 1 dex in stellar mass, and modify the plateau at low mass with a linear extrapolation, down to OF = 0 ("H17low1") or OF = 0.01 ("H17low2"). We have compared the shifted scaling to the occupation fraction in two studies that model DCBHs, the simulation of Tremmel et al. (2017), and the semi-analytical model by Hartwig et al. (2016). Occupation fractions are unpublished, but the authors provided us with their values, and the results are similar to the rescaled/shifted valued we estimated. Significant differences between the models appear only at luminosities below $10^8\,L_{\odot}$, and there is degeneracy between seed mass and occupation

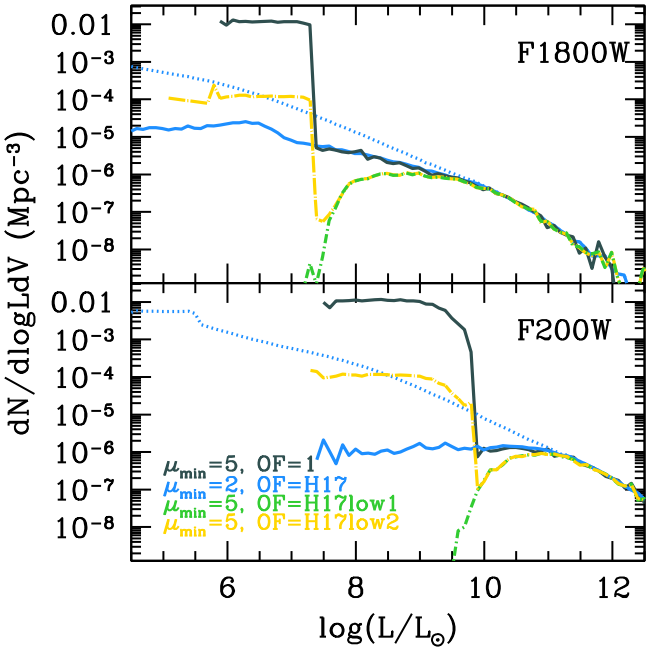

Figure 14. AGN LF at F1800W and at F200W. Line styles are the same as in Figure 13. A model with a moderately high OF for massive seeds (H17low2) shows a dip in the LF when the galaxy outshines the AGN, but at the faint end, the BH is as massive as the galaxy, and therefore the low occupation fraction is compensated for by the AGN outshining the galaxy. When the occupation fraction is lower at low galaxy mass (H17low1), this does not occur, and the faint end of the LF drops monotonically.

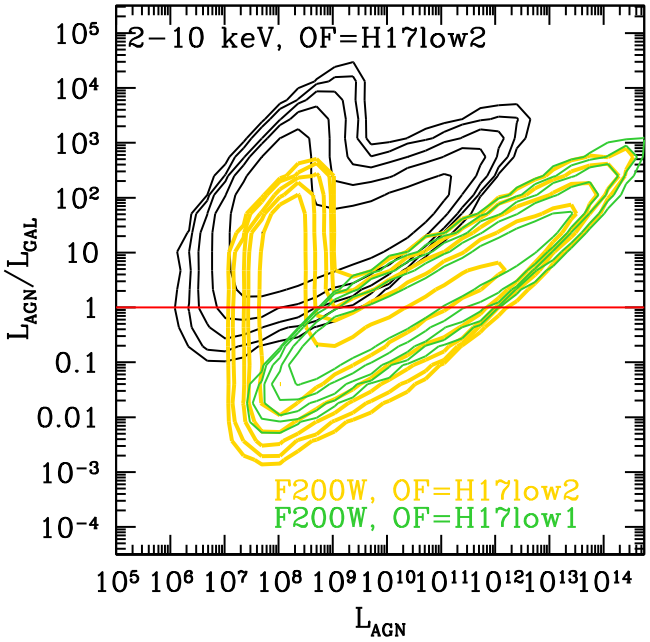

Figure 15. Relation between AGN and galaxy luminosity at 2–10 keV for model H17low2 (black), as well as at F200W, for models H17low2 (yellow) and H17low1 (green). The red line marks equal luminosity.

fraction: more abundant light seeds are indistinguishable from rarer heavy seeds.

After establishing the relevance of different assumptions, in Figure 14 we limit the investigation to the three more physical models. We confirm the impression obtained from Figure 7: a red band enables a better discrimination from the host down to lower luminosities, but at the cost of a lower sensitivity: nominally, the point source detection limit is 8.6 μJy for F1800W and 7.9 nJy for F200W for an exposure time of 10^4 s and a signal-to-noise ratio of 10. The two effects approximately compensate for each other. Figure 14 also shows a non-monotonic behavior of the LF for H17low2: this is caused by a combination of OF and relative luminosity between AGN and galaxy. In this model, at $s \sim 9$, the fraction of galaxies with an AGN with a luminosity higher than the galaxy becomes null, but if the OF does not drop to zero, the average ratio of BH mass to galaxy mass increases soon afterward, as the BH mass cannot dip below $\mu = 5$, and the ratio between BH and stellar mass eventually reaches unity at $s = 5$. This population is a variant of the huge black holes proposed by Agarwal et al. (2013) and Natarajan et al. (2017). If instead the OF continues to decrease, this population does not contribute to the LF. The X-ray LF, instead, decreases monotonically because there is no galaxy mass at which the fraction of galaxies with an AGN with a luminosity higher than the galaxy becomes null. The ratio between AGN and galaxy luminosity at 2–10 keV for H17low2, as well as at F200W for H17low2 and H17low1, is shown in Figure 15 to exemplify the arguments above.

4. Conclusions

We have presented a population synthesis model for galaxies, black holes, and AGNs at high redshift, primarily toward the faint end of the LF. In this first paper we have focused on the description of the method, its validation against observational constraints, and the analysis of the relative properties of galaxies and AGNs in the restframe UV and X-rays where observations are available now, and in the restframe optical/near-IR where observations will soon be available owing to the *JWST*.

The model assumptions are as follows:

1. The model builds a galaxy/AGN population taking as starting point the galaxy MF.
2. Galaxies are characterized by an SED that is determined by age and metallicity. The former is assigned by assuming a relation between stellar mass and SFR (main sequence), and the latter from a mass–metallicity relation.
3. BHs have a mass that scales with the galaxy stellar mass and a luminosity given by a lognormal distribution of the Eddington ratios, with parameters fitted by requiring a good match with upper/lower values to the AGN LF at $z = 6$. The AGN SED depends on the physical properties of BHs, namely their mass and Eddington ratio.

The main results are as follows:

1. For high-redshift galaxies, with stellar ages <1 Gyr, confusion between the galaxy and the AGN is higher at UV and blue optical wavelengths, where uncertainties in dust attenuation are also significant.
2. We propose a color–color selection, e.g., F1280W-F1800W versus F200W-F1280W, to separate galaxies with stronger and weaker AGN in *JWST* photometric observations.
3. We estimate the AGN contribution at the energies driving C IV and He II. At a magnitude of ~ -20, the fraction of galaxies with an AGN contributing more than 10% to driving the He II line is ~20%–25%, and the fraction is ~10%–20% for C IV.

4. We adopt recent determinations of the redshift evolution of the relation between SFR and HMXB luminosity to predict where "stellar contamination" affects X-ray observations, and establish a baseline for multiwavelength studies.
5. For realistic assumptions, the faint end of the X-ray and UV to near-IR LF does not depend appreciably on the minimum BH mass and on the fraction of galaxies hosting a BH, especially considering the degeneracies between these parameters. The difficulty of distinguishing the AGN emission from starlight and HMXBs at low AGN luminosity hinders a clean distinction between these properties.

We are grateful to the reviewer for the suggestions and careful reading of the manuscript. M.V. warmly thanks Matt Lehnert for helping her unravel the mystery of magnitudes, "a quaint unit of historical interest," cit. Cloudy & Associates (http://www.nublado.org), Alice Shapley for thoughtful conversations and her kind help with galaxy spectra, Roberto Gilli, Claudia Maraston, and Roberto Decarli for constructive discussions and comments on the manuscript, and F. Vito and B. Lehmer for help with HMXBs. M.V. and M.T. acknowledge funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013 Grant Agreement no. 614199, project "BLACK"). A.E.R. is grateful for the support of NASA through Hubble Fellowship grant *HST*-HF2-51347.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555.

Appendix A
Validation of the AGN SED

In this appendix we validate the AGN SED that we created against commonly adopted SEDs derived from observations. An important consideration is that our AGNs span a large range in mass and accretion rates, while the observed population of quasars samples a biased region of the μ-λ parameter space.

Davis & Laor (2011) find no evidence of a dependence on the ratio of optical to bolometric luminosity with BH mass (see also Koratkar & Blaes 1999), and it seems clear that a "basic" accretion disk model fails to reproduce all the features of observed spectra. However, we argue in the following that our templates provide a qualitatively good approximation to SEDs in the range of masses and accretion rates probed by observations at a level sufficient for the scope of this study. The assessment below shows that our approach produces reasonable results within a physically motivated framework.

In Figure 16 we compare the shape of our SED to the combination of broken power laws adopted by Marconi et al. (2004) and later by Hopkins et al. (2007). The shape and location of the peak match well when we select only BHs with properties that correspond to the general sample of observed quasars, i.e., with high BH mass and accretion rate. It is particularly encouraging that the geometric mean of the SEDs for a population of BHs with the reference parameters we used, $\bar{\lambda} = \log(0.11)$, $\sigma_\lambda = 0.30$, $\sigma_\mu = 0.75$, and $\mathcal{D} = 0.25$, is in good agreement with the spectrum template.

In more detail, we can appreciate a comparison with the bolometric corrections in the standard reference bands, B band, 0.5–2 keV, and 2–10 keV, using the same conventions as in

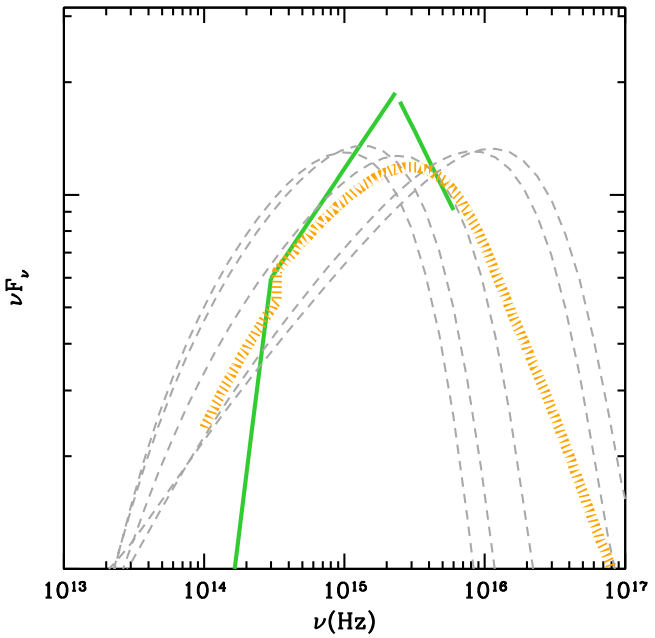

Figure 16. Comparison between the adopted AGN SED with the template spectrum used by Marconi et al. (2004; green broken power laws). The short dashed gray curves show from left to right the SEDs of BHs with $\mu = 9$, $\lambda = -0.52$; $\mu = 8$, $\lambda = -0.52$; $\mu = 8.5$, $\lambda = -0.52$; $\mu = 7.5$, $\lambda = 0$; $\mu = 7$, and $\lambda = 0$. The orange dotted curve is the geometric mean of a suite of SEDs generated for a population of BHs with $\bar{\lambda} = \log(0.40)$, $\sigma_\lambda = 0.40$, $\sigma_\mu = 0.50$, and $\mathcal{D} = 0.25$, selecting only AGNs with $\lambda > -1$ and bolometric luminosity $>10^{44}$ erg s^{-1}. To obtain a sharper cutoff at red wavelengths, we could increase $\tau_{\rm IR}$, set to $k_B T_{\rm IR} = 0.01$ Ryd in the default spectrum.

Marconi et al. (2004), i.e., $L_{\rm bol}/\nu_B L_{\nu_B}$, $L_{\rm bol}/L_{0.5-2\,\rm keV}$, and $L_{\rm bol}/L_{2-10\,\rm keV}$. The comparison is shown in Figure 17, left. Each diagonal sequence is for a different BH mass from $\mu = 2$ to $\mu = 10$, left to right. In B band, for each sequence, the Eddington ratio increases from top to bottom in each sequence, from $\lambda = -2.7$ to $\lambda = 0$. In X-rays, the Eddington ratio instead decreases from top to bottom in the same λ range. Compared also to the recent estimates of bolometric corrections by Lusso et al. (2012), the agreement in B band for the same luminosity range probed by observations, $10^{10} < L_{\rm bol} < 10^{12}\,L_\odot$, is very good, while our X-ray bolometric corrections are higher. In both the comparisons with Marconi et al. (2004) and Lusso et al. (2012), the underestimation of the X-ray luminosity appears to be caused by not having included a reflection component on top of the power law in our SED. When included in a statistical sample, with the population properties with the reference parameters we used, the differences are minimized, as shown in the right panel of Figure 17, since most BHs are accreting at relatively high rates ($\bar{\lambda} = \log(0.11)$). We have checked that this difference does not significantly affect the X-ray LF.

In Figure 18 we compare bolometric corrections at 5100 Å with the results by Jin et al. (2012), who fit the dependence from mass and accretion rate on a large sample of nearby unobscured AGNs. In the left panel, the accretion rate increases from bottom to top, while in the right panel, the behavior is more complicated, but the overall mass increases from top to bottom. Interestingly, while the separate fits for mass and Eddington ratio from Jin et al. (2012) are not in good agreement with our model, the joint μ and λ dependence obtained by combining the two separate fits matches our SED values well.

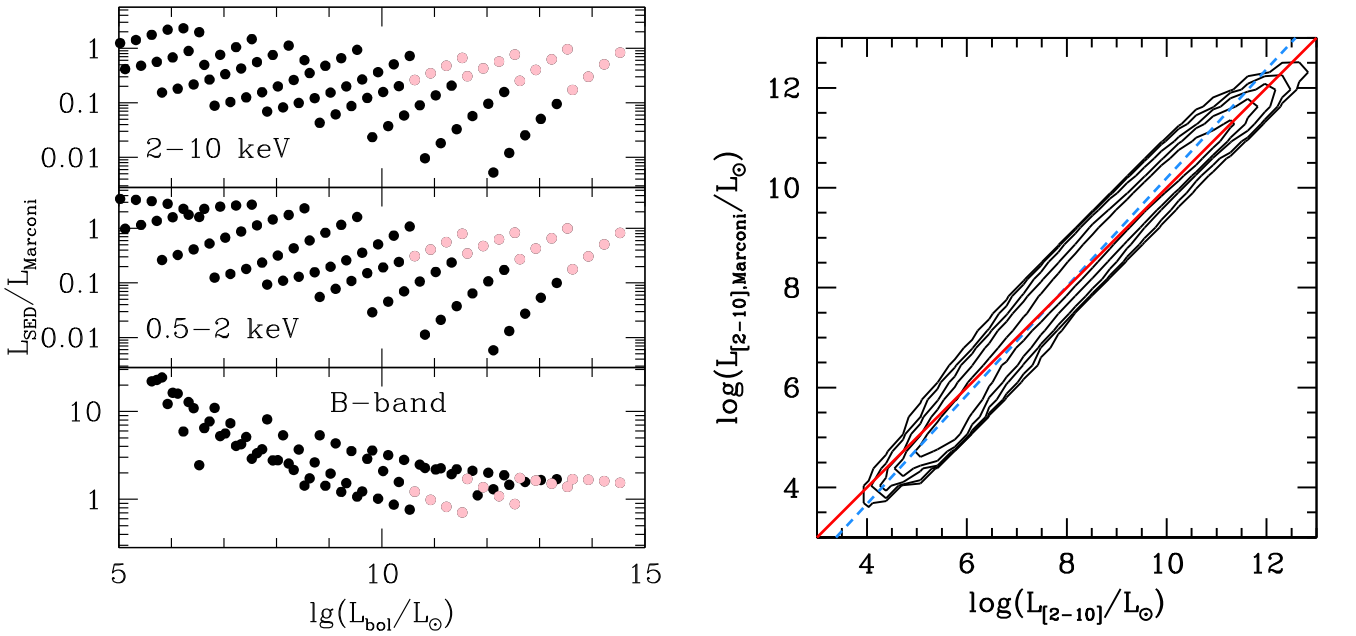

Figure 17. Left: bolometric corrections obtained for our AGN SED compared to those by Marconi et al. (2004). The pink points highlight BHs with $\mu \geqslant 7$ and $\lambda \geqslant -1$, corresponding to the typical masses and accretion rates for quasars for which the standard SEDs have been calculated. Right: comparison of the distribution of X-ray luminosities at 2–10 keV calculated from our SEDs to those obtained from Marconi et al. (2004) for a population of BHs with $\bar{\lambda} = \log(0.40)$, $\sigma_\lambda = 0.40$, and $\sigma_\mu = 0.50$. The red solid line is the 1:1 relation, while the blue dashed line shows a least-squares fit in log space (slope: 1.1, intercept: −0.7).

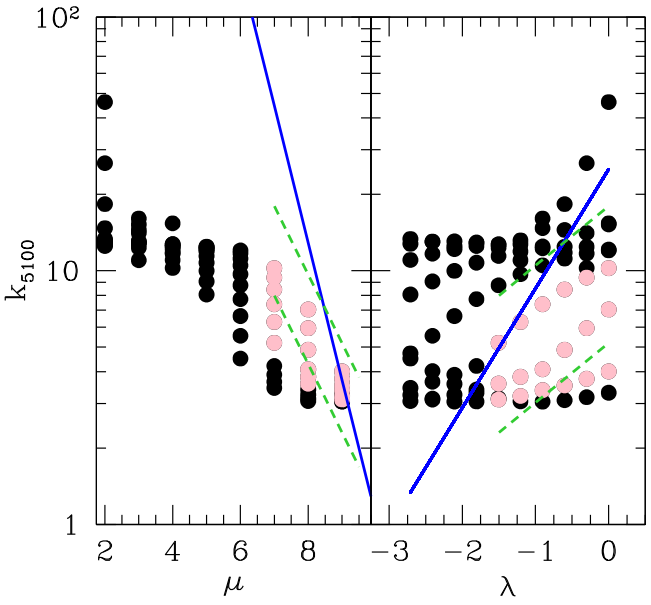

Figure 18. Bolometric correction, k_{5100}, defined as L_{bol}/L_{5100}, where L_{5100} is the monochromatic luminosity at 5100 Å vs. BH mass, μ, or Eddington ratio, λ. The black points are obtained from our SED, with the pink points selecting BHs with $\mu \geqslant 7$ and $\lambda \geqslant -1$. The blue line reports the correlations obtained by Jin et al. (2012) for k_{5100} vs. μ or λ separately. The green dashed lines show the joint μ and λ dependence obtained by combining the two separate fits; these lines are shown only for BHs with $-1.5 < \lambda < 0$, $7 \leqslant \mu \leqslant 9$ to match the mass-accretion range probed by observations.

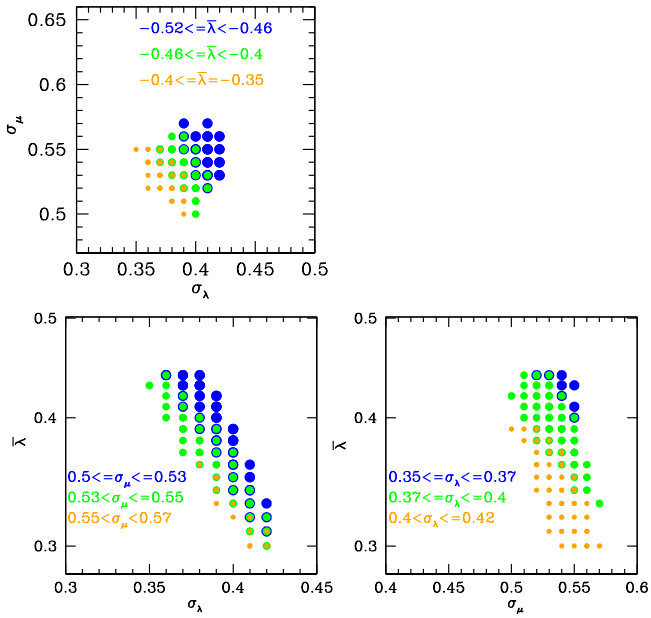

Figure 19. Range of parameters providing a goodness of fit comparable to the set of parameters that best allows us to reproduce both the X-ray and UV LF. In each panel we show how two of the three parameters vary as a function of the third. Parameters are correlated: a smaller $\bar{\lambda}$ requires larger σ_λ and σ_μ.

Appendix B
Best-fit Model Parameters, Uncertainties, and Variations

In Figure 19 we show the parameters giving $\chi^2 < 0.35$ and how they vary in a correlated way. All these models provide similar results for the conclusions of the paper. The three free parameters are not independent. A smaller $\bar{\lambda}$ can be accommodated with larger σ_λ and σ_μ, and vice versa. The reported range is derived from a grid of values, and for combinations of the parameters within the range provided, the chi-squared value is within a similar range.

We discuss here a super-maximal case, where we fit the parameters to the UV LF by Giallongo et al. (2015). This exercise requires a different approach. The AGNs are detected in X-rays, but the LF is provided in the UV for the total UV luminosity, without separating galaxy from AGN. In order to mimic the same approach, we calculate the LF by applying only a correction for Compton-thick AGNs (to reproduce the

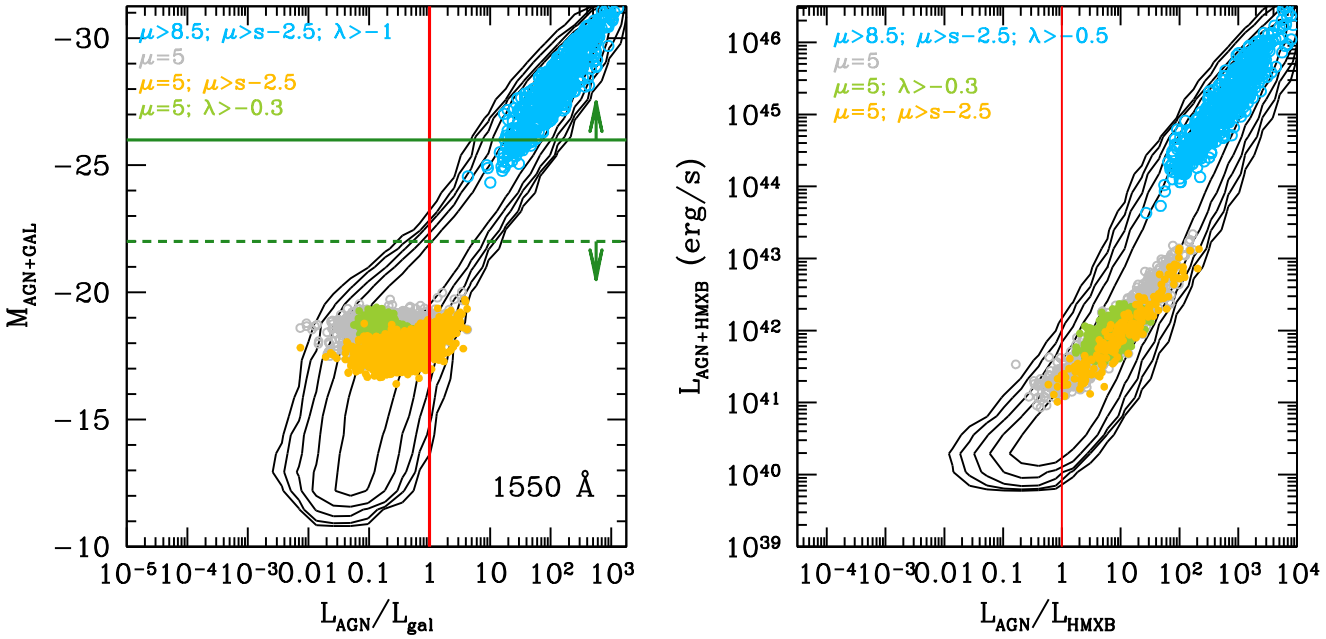

Figure 20. Compare to Figures 5 and 12.

X-ray selection) and by including both AGN and galaxy luminosity when calculating the LF in UV (to reproduce the total UV luminosity used by Giallongo et al. 2015 to estimate the AGN LF). We are unable to find an acceptable fit for this LF for our standard BH-galaxy relationship, although when we limit the comparison only to the magnitude range of the data of Giallongo et al. 2015 (from −19 to −21), i.e., we do not try to fit the full LF, the model parameters of our reference case produce an LF that is compatible with the LF of Giallongo et al. 2015. We confirm the results by Qin et al. (2017) that the UV luminosity is dominated by the galaxy stellar population in this magnitude range, with the AGN-only luminosity function, that is, without including stars in the source luminosity, ×0.04 what we obtain including both stellar and AGN light in the "AGN" luminosity.

To fit for the full range of the LF proposed by Giallongo et al. (2015), i.e., including the bright end, we modify the BH-galaxy relationship using the "vanilla" scaling from Volonteri & Reines (2016), i.e., $\mu = s - 2.7$, and find $\bar{\lambda} = \log(0.75)$, $\sigma_\lambda = 0.4$, and $\sigma_\mu = 0.2$. In this case, most galaxies with a UV magnitude ∼−20 should be AGN dominated if they host an active BH (left panel of Figure 20), and HMXBs would not be a significant contaminant even for BHs with $\mu = 5$ (right panel of Figure 20). Taking into account the assumed duty cycle of $\mathcal{D} = 0.25$, 25% of galaxies brighter than −20 are AGN dominated, and in Figure 21 we show the fraction of galaxies as a function of the galaxy (top) and total (bottom) UV magnitude where an X-ray detection is expected. In the reference model, most galaxies fainter than −22 are star dominated, while in the super-maximal case, a significant fraction, corresponding to all galaxies with an active BH, is AGN dominated down to a magnitude of −20.

ORCID iDs

Marta Volonteri https://orcid.org/0000-0002-3216-1322
Amy E. Reines https://orcid.org/0000-0001-7158-614X
Hakim Atek https://orcid.org/0000-0002-7570-0824

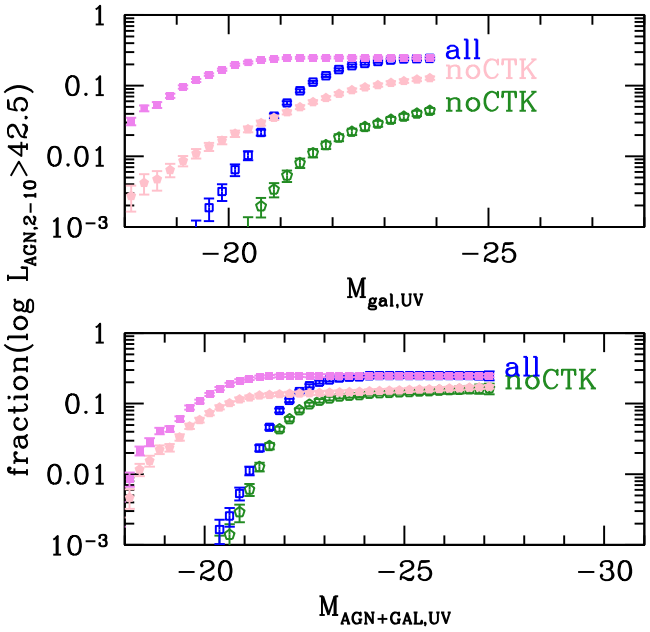

Figure 21. Fraction of galaxies hosting an AGN with luminosity in the 2–10 keV band $>10^{42.5}$ erg s^{-1} vs. galaxy (top) and total (AGN+galaxy, bottom) UV magnitude. The empty symbols (blue: all; green: applying a correction for Compton-thick AGN) refer to the reference model described in the body of the paper, the full symbols (violet: all; pink: applying a correction for Compton-thick AGNs) symbols to the super-maximal case.

References

Agarwal, B., Davis, A. J., Khochfar, S., Natarajan, P., & Dunlop, J. S. 2013, MNRAS, 432, 3438
Atek, H., Richard, J., Jauzac, M., et al. 2015, ApJ, 814, 69
Atek, H., Siana, B., Scarlata, C., et al. 2011, ApJ, 743, 121
Bañados, E., Venemans, B. P., Decarli, R., et al. 2016, ApJS, 227, 11
Beckmann, R. S., Devriendt, J., Slyz, A., et al. 2017, MNRAS, 472, 949
Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57
Bian, F., Kewley, L. J., Dopita, M. A., & Blanc, G. A. 2017, ApJ, 834, 51
Bouwens, R. J., Aravena, M., Decarli, R., et al. 2016, ApJ, 833, 72

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2014, ApJ, 793, 115
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, ApJ, 803, 34
Bowler, R. A. A., McLure, R. J., Dunlop, J. S., et al. 2017, MNRAS, 469, 448
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Buchner, J., Georgakakis, A., Nandra, K., et al. 2015, ApJ, 802, 89
Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682
Cappelluti, N., Comastri, A., Fontana, A., et al. 2016, ApJ, 823, 95
Curtis-Lake, E., McLure, R. J., Dunlop, J. S., et al. 2013, MNRAS, 429, 302
Davis, S. W., & Laor, A. 2011, ApJ, 728, 98
de Barros, S., Schaerer, D., & Stark, D. P. 2014, A&A, 563, A81
Decarli, R., Walter, F., Venemans, B. P., et al. 2017, Natur, 545, 457
DeGraf, C., Di Matteo, T., Treu, T., et al. 2015, MNRAS, 454, 913
Ding, X., Treu, T., Suyu, S. H., et al. 2017, MNRAS, 472, 90
Done, C., Davis, S. W., Jin, C., Blaes, O., & Ward, M. 2012, MNRAS, 420, 1848
Dong, R., Greene, J. E., & Ho, L. C. 2012, ApJ, 761, 73
Eldridge, J. J., & Stanway, E. R. 2009, MNRAS, 400, 1019
Elvis, M., Hao, H., Civano, F., et al. 2012, ApJ, 759, 6
Elvis, M., Wilkes, B. J., McDowell, J. C., et al. 1994, ApJS, 95, 1
Fan, X., Strauss, M. A., Richards, G. T., et al. 2006, AJ, 131, 1203
Fan, X. e. a. 2001, AJ, 121, 54
Feltre, A., Charlot, S., & Gutkin, J. 2016, MNRAS, 456, 3354
Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, RMxAA, 49, 137
Fragos, T., Lehmer, B., Tremmel, M., et al. 2013, ApJ, 764, 41
Giallongo, E., Grazian, A., Fiore, F., et al. 2015, A&A, 578, A83
Gilli, R., Comastri, A., & Hasinger, G. 2007, A&A, 463, 79
Habouzit, M., Volonteri, M., & Dubois, Y. 2017, MNRAS, 468, 3935
Habouzit, M., Volonteri, M., Latif, M., Dubois, Y., & Peirani, S. 2016, MNRAS, 463, 529
Haiman, Z. 2013, ASSL, 396, 293
Hainline, K. N., Shapley, A. E., Greene, J. E., et al. 2012, ApJ, 760, 74
Hainline, K. N., Shapley, A. E., Greene, J. E., & Steidel, C. C. 2011, ApJ, 733, 31
Hartwig, T., Latif, M. A., Magg, M., et al. 2016, MNRAS, 462, 2184
Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, ApJ, 654, 731
Jensen, T. W., Vivek, M., Dawson, K. S., et al. 2016, ApJ, 833, 199
Jiang, L., McGreer, I. D., Fan, X., et al. 2016, ApJ, 833, 222
Jin, C., Ward, M., & Done, C. 2012, MNRAS, 425, 907
Jones, M. L., Hickox, R. C., Black, C. S., et al. 2016, ApJ, 826, 12
Kashikawa, N., Ishizaki, Y., Willott, C. J., et al. 2015, ApJ, 798, 28
Kauffmann, G., & Heckman, T. M. 2009, MNRAS, 397, 135
Koratkar, A., & Blaes, O. 1999, PASP, 111, 1
Lauer, T. R., Faber, S. M., Richstone, D., et al. 2007, ApJ, 662, 808
Lehmer, B. D., Basu-Zych, A. R., Mineo, S., et al. 2016, ApJ, 825, 7
Lemons, S. M., Reines, A. E., Plotkin, R. M., Gallo, E., & Greene, J. E. 2015, ApJ, 805, 12
Liu, T., Tozzi, P., Wang, J.-X., et al. 2017, ApJS, 232, 8
Lusso, E., Comastri, A., Simmons, B. D., et al. 2012, MNRAS, 425, 623
Ma, J., Gonzalez, A. H., Vieira, J. D., et al. 2016a, ApJ, 832, 114
Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016b, MNRAS, 456, 2140
Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415
Mainali, R., Kollmeier, J. A., Stark, D. P., et al. 2017, ApJL, 836, L14
Maraston, C. 2005, MNRAS, 362, 799
Marchesi, S., Civano, F., Salvato, M., et al. 2016, ApJ, 827, 150
Marconi, A., Risaliti, G., Gilli, R., et al. 2004, MNRAS, 351, 169
Matsuoka, Y., Onoue, M., Kashikawa, N., et al. 2016, ApJ, 828, 26
Matsuoka, Y., Onoue, M., Kashikawa, N., et al. 2017, arXiv:1704.05854
Merloni, A., Bongiorno, A., Brusa, M., et al. 2014, MNRAS, 437, 3550
Messias, H., Afonso, J., Salvato, M., Mobasher, B., & Hopkins, A. M. 2012, ApJ, 754, 120
Messias, H., Afonso, J. M., Salvato, M., Mobasher, B., & Hopkins, A. M. 2014, A&A, 562, A144
Meurer, G. R., Heckman, T. M., & Calzetti, D. 1999, ApJ, 521, 64
Mineo, S., Gilfanov, M., & Sunyaev, R. 2012, MNRAS, 419, 2095
Natarajan, P. 2014, GReGr, 46, 1702
Natarajan, P., Pacucci, F., Ferrara, A., et al. 2017, ApJ, 838, 117
Ono, Y., Ouchi, M., Harikane, Y., et al. 2017, arXiv:1704.06004
Pacucci, F., Ferrara, A., Grazian, A., et al. 2016, MNRAS, 459, 1432
Pacucci, F., Ferrara, A., Volonteri, M., & Dubus, G. 2015, MNRAS, 454, 3771
Qin, Y., Mutch, S. J., Poole, G. B., et al. 2017, MNRAS, 472, 2009
Reines, A. E., & Comastri, A. 2016, PASA, 33, e054
Reines, A. E., Nidever, D. L., Whelan, D. G., & Johnson, K. E. 2010, ApJ, 708, 26
Reines, A. E., & Volonteri, M. 2015, ApJ, 813, 82
Ricci, C., Trakhtenbrot, B., Koss, M. J., et al. 2017a, Natur, 549, 488
Ricci, F., Marchesi, S., Shankar, F., La Franca, F., & Civano, F. 2017b, MNRAS, 465, 1915
Richard, J., Kneib, J.-P., Ebeling, H., et al. 2011, MNRAS, 414, L31
Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., et al. 2006, ApJs, 166, 470
Riechers, D. A., Bradford, C. M., Clements, D. L., et al. 2013, Natur, 496, 329
Salmon, B., Papovich, C., Finkelstein, S. L., et al. 2015, ApJ, 799, 183
Salpeter, E. E. 1955, ApJ, 121, 161
Schulze, A., & Wisotzki, L. 2011, A&A, 535, A87
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Shields, G. A., Menezes, K. L., Massart, C. A., & Vanden Bout, P. 2006, ApJ, 641, 683
Sobral, D., Matthee, J., Darvish, B., et al. 2015, ApJ, 808, 139
Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, ApJ, 825, 5
Stanway, E. R., Eldridge, J. J., & Becker, G. D. 2016, MNRAS, 456, 485
Stark, D. P. 2016, ARA&A, 54, 761
Stark, D. P., Ellis, R. S., Charlot, S., et al. 2017, MNRAS, 464, 469
Stark, D. P., Richard, J., Charlot, S., et al. 2015a, MNRAS, 450, 1846
Stark, D. P., Schenker, M. A., Ellis, R., et al. 2013, ApJ, 763, 129
Stark, D. P., Walth, G., Charlot, S., et al. 2015b, MNRAS, 454, 1393
Strandet, M. L., Weiss, A., De Breuck, C., et al. 2017, ApJL, 842, L15
Thomas, A. D., Groves, B. A., Sutherland, R. S., et al. 2016, ApJ, 833, 266
Treister, E., Schawinski, K., Volonteri, M., & Natarajan, P. 2013, ApJ, 778, 130
Tremmel, M., Karcher, M., Governato, F., et al. 2017, MNRAS, 470, 1121
Tucci, M., & Volonteri, M. 2017, A&A, 600, A64
Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, ApJ, 786, 104
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
Vito, F., Gilli, R., Vignali, C., et al. 2016, MNRAS, 463, 348
Vito, F., Maiolino, R., Santini, P., et al. 2014, MNRAS, 441, 1059
Vito, W. N., Brandt, G., Yang, R., et al. 2017, MNRAS, submitted (arXiv:1709.07892)
Volonteri, M. 2010, A&ARv, 18, 279
Volonteri, M., Dubois, Y., Pichon, C., & Devriendt, J. 2016, MNRAS, 460, 2979
Volonteri, M., & Reines, A. E. 2016, ApJL, 820, L6
Volonteri, M., & Stark, D. P. 2011, MNRAS, 417, 2085
Weigel, A. K., Schawinski, K., Treister, E., et al. 2015, MNRAS, 448, 3167
Willott, C. J., Delorme, P., Reylé, C., et al. 2010, AJ, 139, 906