The AGN-galaxy-halo connection:

The AGN–galaxy–halo connection: The distribution of AGN host halo masses to $z = 2.5$

James Aird[1,2]⋆ and Alison L. Coil[3]

[1]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Edinburgh EH9 3HJ, UK*
[2]*School of Physics & Astronomy, University of Leicester, University Road, Leicester LE1 7RJ, UK*
[3]*Center for Astrophysics and Space Sciences, Department of Physics, University of California,*
9500 Gilman Dr. MC 0424, La Jolla, CA 92093-0424, USA

ABSTRACT
It is widely reported, based on clustering measurements of observed active galactic nuclei (AGN) samples, that AGN reside in similar mass host dark matter halos across the bulk of cosmic time, with log $\mathcal{M}/\mathcal{M}_\odot \sim 12.5 - 13.0$ to $z \sim 2.5$. We show that this is due in part to the AGN fraction in galaxies rising with increasing stellar mass, combined with AGN observational selection effects that exacerbate this trend. Here, we use AGN specific accretion rate distribution functions determined as a function of stellar mass and redshift for star-forming and quiescent galaxies separately, combined with the latest galaxy-halo connection models, to determine the parent and sub-halo mass distribution function of AGN to various observational limits. We find that while the median (sub-)halo mass of AGN, $\approx 10^{12} \mathcal{M}_\odot$, is fairly constant with luminosity, specific accretion rate, and redshift, the full halo mass distribution function is broad, spanning several orders of magnitude. We show that widely used methods to infer a typical dark matter halo mass based on an observed AGN clustering amplitude can result in biased, systematically high host halo masses. While the AGN satellite fraction rises with increasing parent halo mass, we find that the central galaxy is often not an AGN. Our results elucidate the physical causes for the apparent uniformity of AGN host halos across cosmic time and underscore the importance of accounting for AGN selection biases when interpreting observational AGN clustering results. We further show that AGN clustering is most easily interpreted in terms of the relative bias to galaxy samples, not from absolute bias measurements alone.

Key words: galaxies: active – galaxies: evolution – galaxies: haloes

1 INTRODUCTION

Accreting supermassive black holes (SMBHs), observed as active galactic nuclei (AGN), are thought to play a crucial role in the galaxy evolution process, influencing the shape of the galaxy stellar mass function (e.g., Bower et al. 2012; Puchwein & Springel 2013), the morphologies of galaxies (Dubois et al. 2016), and contributing to the quenching of star formation (e.g., Croton et al. 2006; Hopkins et al. 2006; Zubovas & King 2012; Dubois et al. 2013; Beckmann et al. 2017). AGN feedback in particular is an essential element of modern simulations of galaxy formation and evolution (e.g., Di Matteo et al. 2005; Booth & Schaye 2009; Richardson et al. 2016).

While the importance of AGN in the lifecycles of galaxies is well established, it is not understood precisely what triggers AGN activity. The extreme difference in scale between galaxies and supermassive black holes, coupled with the relative rarity of the active accretion phase of SMBHs, has made it difficult to determine the physical mechanisms connecting galaxy and AGN growth. Constraining the triggering and fueling mechanisms of AGNs is crucial to uncovering the relevant physics behind the impact that SMBHs have on their host galaxies.

The spatial distribution or clustering of AGN on large scales is often used to constrain the host dark matter halo properties of AGN and can reveal AGN triggering mechanisms when compared to theoretical models (e.g., Allevato et al. 2011; Gatti et al. 2016). Clustering measurements allow AGN to be placed in a cosmological context and reveal underlying correlations between the large-scale structure of the Universe and AGN fueling.

Most AGN clustering studies, the bulk of which are performed using AGN identified at either optical or X-ray wavelengths, find that AGN – whether high-luminosity quasars or more moderate-luminosity AGN – typically reside in host dark matter halos of mass log $\mathcal{M}_{halo}/\mathcal{M}_\odot \sim 12.5 - 13$ from $z \sim 0$ to $z \sim 4$ (see reviews of X-ray AGN clustering by Cappelluti et al. 2012 and Krumpe et al. 2014 and recent optical quasar clustering results of Timlin et al. 2018 and references therein). In particular, the typical AGN host halo mass is not found to evolve substantially, if at all, with redshift. While small differences in the typical host halo mass have been found when identifying AGN at different wavelengths, luminosities, or obscuration levels (e.g., Hickox et al. 2009; Krumpe et al. 2010; Powell et al. 2018), these differences are relatively small when compared with the striking similarity of the typical halo mass over the bulk of cosmic time. This relative invariance in the halo mass with redshift is difficult to understand, particularly given that the dark matter halo mass of

⋆ james.aird@ed.ac.uk

galaxies is a variable function of redshift as well as galaxy properties such as stellar mass and star formation rate (e.g. Behroozi et al. 2013c; Coil et al. 2017, and references therein). Physically, why would the accretion of mass on the scales of a SMBH at the center of a galaxy be related to a given host halo dark matter mass, and why would the halo mass of AGN not evolve with redshift in a hierarchical galaxy formation model?

A complicating factor in interpreting AGN clustering results is the strong AGN observational selection effect, first shown in Aird et al. (2012), in which SMBHs in more massive galaxies are easier to observe as AGN to a given flux limit of an observational sample. In essence, more massive galaxies host more massive SMBHs (Cisternas et al. 2011; Marleau et al. 2013; Reines & Volonteri 2015; Simmons et al. 2017), therefore to a given flux limit of a survey it is possible to observe AGN at lower Eddington ratios (L_{AGN}/L_{Edd}) in more massive galaxies than in lower mass galaxies. It is simply easier to *detect* AGN in more massive galaxies, as they can be accreting at relatively low rates and still be observable, while in lower mass galaxies AGN must be accreting at a high rate relative to the galaxy mass to be observed. AGN identified in lower mass galaxies will preferentially have high Eddington ratios, as AGN accreting at lower specific accretion rates will be below the flux limit of the observational survey (Aird et al. 2013). When this selection effect is accounted for, it is found that AGN have a wide range of specific accretion rates (L_{AGN}/galaxy stellar mass), which roughly translates to a wide Eddington ratio distribution function, and which depends somewhat on the stellar mass of the host galaxy (Georgakakis et al. 2017; Bongiorno et al. 2016; Aird et al. 2018). This selection effect results in AGN being preferentially identified in high mass galaxies (e.g., Brusa et al. 2009; Xue et al. 2010; Mendez et al. 2013).

This important selection effect is typically not considered when interpreting AGN clustering results, however. Multiple recent AGN clustering studies (Mendez et al. 2016; Powell et al. 2018; Georgakakis et al. 2019, see also Leauthaud et al. 2015) have shown that the clustering of observed AGN samples, whether selected at X-ray, IR, or radio wavelengths, matches the clustering of inactive galaxies of the same stellar mass, star formation rate (SFR) (when calculated), and redshift distributions. Specifically, this shows that the known galaxy stellar mass to halo mass relation and its scatter—which has been constrained from galaxy clustering measurements as a function of stellar mass and redshift (Behroozi et al. 2019)—can be used to infer the dark matter halo mass distribution of AGN host galaxies, once the stellar masses (and ideally SFRs) of AGN host galaxies are determined.

We show here that the measured broad AGN specific accretion rate distribution function, determined for star-forming and quiescent galaxies as a function of stellar mass and redshift, when combined with the latest empirical models for how galaxies populate dark matter halos (UniverseMachine, Behroozi et al. 2019), predicts a roughly constant median dark matter halo mass for observed AGN samples across cosmic time. We predict how the clustering amplitude should vary with AGN luminosity and specific accretion rate, as well as redshift, and compare with current AGN clustering measurements. We further investigate the differences between the parent and sub-halo mass functions of AGN hosts. We also show that the inferred host halo mass derived from observed AGN samples should not be interpreted as reflecting the true AGN host halo mass distribution, which can be lower once the known stellar mass-dependent AGN selection effects are accounted for. We also predict the satellite fraction as a function of halo mass and show that at high halo masses the central galaxy is often not an AGN.

In Section 2 we describe the model and methodology used in this

paper. We present predictions from this model in Section 3, which we compare with observational results in Section 4. We discuss our findings in Section 5 and present conclusions in Section 6. Throughout the paper we adopt a flat ΛCDM cosmology ($\Omega_m = 0.307, \Omega_\Lambda = 0.693, h = 0.678, \sigma_8 = 0.823, n_s = 0.96$: Planck Collaboration et al. 2016). We evaluate h in all quantities unless explicitly stated.

2 MODELLING METHODOLOGY

In this section we describe the methodology we employ to link AGN to dark matter haloes and predict their clustering properties. Our method starts from an advanced *galaxy* model (UniverseMachine: Behroozi et al. 2019), which uses empirical relations to link the assembly of individual galaxies and their dark matter haloes, constrained by a range of galaxy observables (including stellar mass functions, quenched fractions and galaxy clustering measurements). We then use recent, robust measurements of the incidence of AGN as a function of galaxy properties (Aird et al. 2018) to place AGN within these model galaxies and thus—via UniverseMachine—within dark matter haloes. This method allows us to predict AGN clustering properties based purely on the known measured relations between AGN and the properties of their host galaxies, combined with the latest galaxy-halo connection models.

2.1 The UniverseMachine galaxy model

Our starting point is the UniverseMachine empirical galaxy model (Behroozi et al. 2019). Here, we provide a brief summary of the method employed by UniverseMachine to allocate galaxies to dark matter haloes, track their evolution over time, predict observables and constrain the model using measurements of galaxy populations.

For this paper, we adopt UniverseMachine data release 1[1] based on the *Small MultiDark Planck* (SMDPL) dark matter simulation (Klypin et al. 2016; Rodríguez-Puebla et al. 2016). Dark matter haloes are identified within this large N-body simulation (400 h^{-1} Mpc sided co-moving volume with 3840 particles) using the RockStar code (Behroozi et al. 2013a), which identifies both individual gravitationally bound haloes and distinct sub-haloes that lie within these parent haloes. Throughout the remainder of this paper we refer to both individual sub-haloes (containing a satellite galaxy) and the parent haloes (that are associated with a central galaxy) jointly using the specific term "(sub-)haloes" or more generally when using the unqualified term "halo". The merger trees of (sub-)haloes are determined using the Consistent Trees code (Behroozi et al. 2013b).

Galaxies are then associated with these individual (sub-)haloes—and the stellar mass assembly of the galaxy is traced—using empirical relations to assign SFRs. The distribution of SFRs is defined for haloes of a given $v_{\mathcal{M}_{peak}}$ (the maximum circular velocity of the halo, v_{max}, at its peak historical halo mass, $\mathcal{M}_{peak}$) and redshift. SFRs for a given $v_{\mathcal{M}_{peak}}$ and z are modelled as a bimodal distribution corresponding to star-forming galaxies and some fraction of "quenched" galaxies, where the quenched fraction also depends on both $v_{\mathcal{M}_{peak}}$ and z. Individual haloes (at a given $v_{\mathcal{M}_{peak}}$) are ranked based on the relative growth of the halo (traced by the change in v_{max}) over the previous dynamic time, and SFRs are assigned based on this ranking such that haloes that are currently growing faster are assigned higher SFRs

---

[1] http://behroozi.users.hpc.arizona.edu/UniverseMachine/DR1/
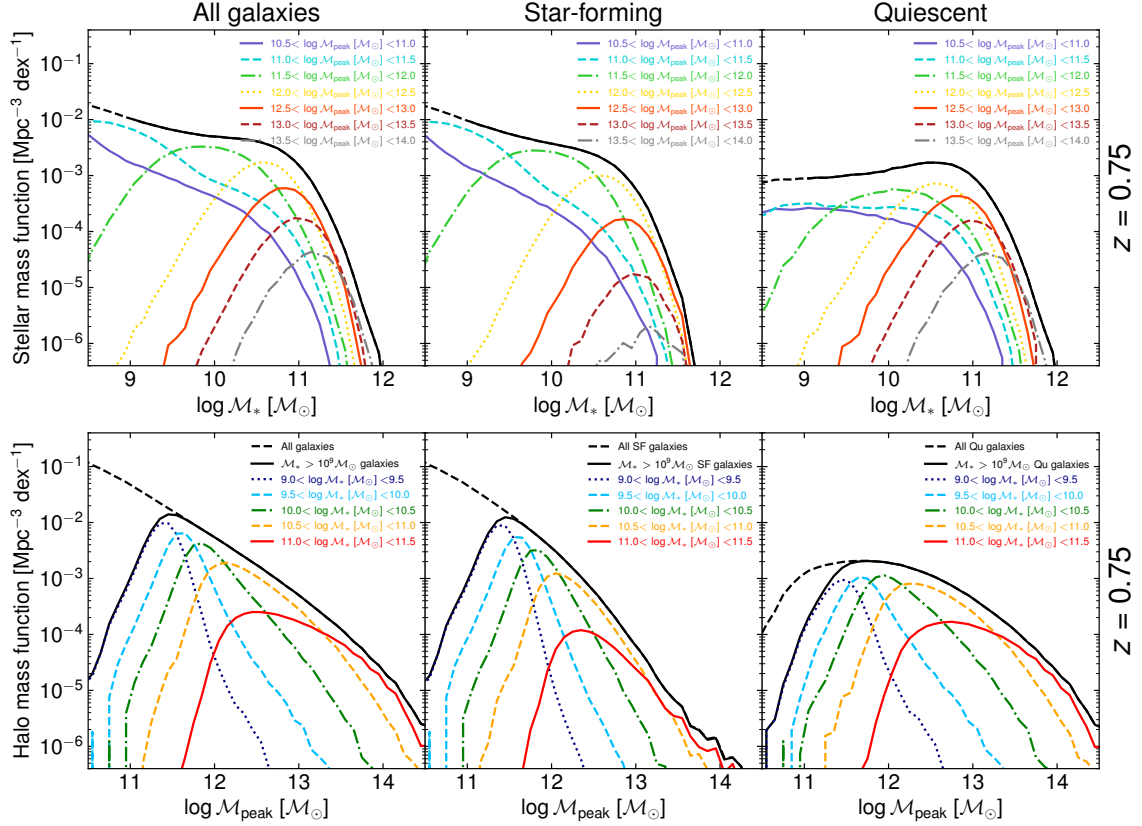
Figure 1. Stellar mass functions of galaxies (top panels) and halo mass functions (bottom panels)—defined here as the differential number density of galaxies as a function of peak historical (sub-)halo mass, $\mathcal{M}_{peak}$—for all galaxies (left), star-forming galaxies (center) and quiescent galaxies (right), as provided by UniverseMachine (Behroozi et al. 2019), shown here at $z = 0.75$. In the top panels, we show the contribution to the total stellar mass function from galaxies in haloes with different $\mathcal{M}_{peak}$ by the colored lines. In the bottom panels, we show the contribution to the total halo mass function from galaxies with different stellar masses by the colored lines. The thick black line in the bottom panel indicates the halo mass function of our "parent galaxy sample" used in this work, which is restricted to galaxies with $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$. This cut introduces a turnover in the halo mass functions of all galaxies and star-forming galaxies at $\mathcal{M}_{halo} \sim 10^{11} \mathcal{M}_\odot$ (cf. the total halo mass function including lower $\mathcal{M}_*$ galaxies shown by the dashed black line). We note that higher $\mathcal{M}_*$ galaxies have a broad range of $\mathcal{M}_{peak}$ and thus make a significant contribution to the total halo mass function across a wide range of $\mathcal{M}_*$. When considering only star-forming galaxies (centre panels) there is a closer mapping between $\mathcal{M}_{peak}$ and $\mathcal{M}_*$, although the intrinsic scatter in the SMHM relation still produces a somewhat broad distribution of $\mathcal{M}_*$ for galaxies of given $\mathcal{M}_{peak}$.

at the current redshift,[2] allowing for scatter in this relation. The total stellar mass growth of individual galaxies over cosmic time is determined by integrating the allocated SFRs over cosmic time and through the mergers of individual (sub-)haloes.

Finally, the model is used to predict a range of galaxy observables, including (but not limited to) stellar mass functions, cosmic star formation rates, specific SFR distributions, quenched fractions, and autocorrelation functions for star-forming and quiescent galaxies across cosmic time. Observational errors and systematic effects are added before comparing the predictions to real data and determining a χ^2 for a given set of input parameters. The input parameters— a total of 44 free parameters controlling the assembly of galaxies

within haloes—are then explored using a Markov Chain Monte Carlo (MCMC) algorithm to determine the best-fit empirical model.

The output of the UniverseMachine model links halo masses to the current stellar mass ($\mathcal{M}_*$) and SFR of individual galaxies, providing robust predictions of the stellar mass–halo mass (SMHM) relation of different galaxy populations and—most crucially—the underlying *scatter* in these relations. In Figure 1 we demonstrate the relationship between stellar mass and (sub-)halo mass, as determined by UniverseMachine. Here, and throughout the remainder of this paper, we adopt the "observed" stellar mass of galaxies that is calculated by UniverseMachine, including the effects of observational errors that can have a systematic impact on the recovered properties. Adopting the observed stellar mass ensures we can directly link the UniverseMachine output to other observables (i.e. AGN fractions as a function of $\mathcal{M}_*$).[3] We use the observationally-determined division of star-forming and quiescent galaxies as adopted by Aird et al. (2018) when measuring AGN fractions (see Section 2.2 below). Specifically,

---

[2] In satellite galaxies whose sub-haloes may be undergoing strong tidal stripping, the change in v_{max} is instead taken over the period since the sub-halo reached its peak historical mass, $\mathcal{M}_{peak}$ (if this exceeds the dynamical time). Changes in v_{max} tend to occur suddenly for such galaxies, followed by longer periods (exceeding the dynamical time) where v_{max} remains relatively stable. Thus, adopting this longer time period prevents such satellite galaxies being artificially assigned high SFRs and ensures that they remain quenched. See section 3.2 of Behroozi et al. (2019) for full details.

[3] Adopting the true stellar mass estimates from UniverseMachine instead of the observed values marginally reduces the breadth of the distributions of stellar masses at fixed halo mass shown in Figure 1 but does not have a significant impact on our results or overall conclusions.

we define quiescent galaxies as having

$$\log \text{SFR} \, [\mathcal{M}_\odot \, \text{yr}^{-1}] < -8.9 + 0.76 \log \frac{\mathcal{M}_*}{\mathcal{M}_\odot} + 2.95 \log(1 + z). \quad (1)$$

This cut cleanly divides the UniverseMachine model galaxy populations for the range of stellar masses ($\mathcal{M}_* > 10^9 \mathcal{M}_\odot$) considered in this paper.

The top row of Figure 1 shows the overall stellar mass functions (SMFs) of all galaxies and separately for the star-forming and quiescent galaxy populations (shown here at $z = 0.75$ only, for demonstration purposes). In each panel we show the contributions to the SMFs from galaxies in haloes with different masses. The bottom row shows the halo mass functions (HMFs) for each galaxy population and the contributions from galaxies with different stellar masses. We adopt the peak historical (sub-)halo mass, $\mathcal{M}_{\text{peak}}$, which traces the overall growth of the halo in which a given galaxy grows and is thus the key quantity to consider for understanding galaxy evolution (Reddick et al. 2013). Figure 1 shows how (sub-)haloes of a given $\mathcal{M}_{\text{peak}}$ host galaxies with a broad range of stellar masses; correspondingly, galaxies of a given $\mathcal{M}_*$ are found in haloes with a broad range of $\mathcal{M}_{\text{peak}}$. When considering star-forming galaxies only there is a closer mapping between $\mathcal{M}_*$ and $\mathcal{M}_{\text{peak}}$, although the distribution of $\mathcal{M}_*$ for a given $\mathcal{M}_{\text{peak}}$ remains fairly broad. For star-forming galaxies with $\mathcal{M}_* = 10^{10-10.5} \mathcal{M}_\odot$, the width of the halo mass function that encloses 90% of galaxies is ~0.6 dex, but increases toward higher stellar masses reaching ~1.1 dex for galaxies with $\mathcal{M}_* = 10^{11-11.5} \mathcal{M}_\odot$. For quiescent galaxies, the distribution of $\mathcal{M}_{\text{peak}}$ (at a given $\mathcal{M}_*$) is broader, spanning ~0.9 dex for galaxies with $\mathcal{M}_* = 10^{10-10.5} \mathcal{M}_\odot$ and increasing to ~1.5 dex for galaxies with $\mathcal{M}_* = 10^{11-11.5} \mathcal{M}_\odot$. Simpler schemes—for example using existing SMHM relations for all galaxies to map stellar masses to halo masses directly—would not capture this diversity of $\mathcal{M}_{\text{peak}}$ across the galaxy population and the differing host haloes of star-forming and quiescent galaxies that is described by the UniverseMachine model.

2.2 Populating galaxies with AGN

The next stage of our method is to populate the model galaxies from UniverseMachine with AGN. To achieve this, we take a purely empirical approach, adopting measurements of the incidence of X-ray selected AGN for samples of galaxies by Aird et al. (2018, hereafter A18), either taking all galaxies (of given $\mathcal{M}_*$ and z) or considering star-forming and quiescent galaxies as distinct populations. We use these observations to populate galaxies in the UniverseMachine snapshot closest to the central redshift of the bins used by A18.[4]

A18 took large, near-infrared–selected samples of galaxies from the CANDELS and UltraVISTA surveys (Grogin et al. 2011; Skelton et al. 2014; McCracken et al. 2012; Muzzin et al. 2013) and extracted X-ray data at the positions of every galaxy from deep *Chandra* imaging (Alexander et al. 2003; Xue et al. 2011; Nandra et al. 2015; Civano et al. 2016). A Bayesian methodology was used to combine the X-ray data from samples of galaxies at particular ranges of $\mathcal{M}_*$ and z and recover robust estimates of the underlying distribution of specific

accretion rates, λ, defined as

$$\lambda = \frac{k_{\text{bol}} L_X}{1.3 \times 10^{38} \, \text{erg s}^{-1} \times 0.002 \frac{\mathcal{M}_*}{\mathcal{M}_\odot}} \quad (2)$$

where L_X is the rest-frame 2–10 keV luminosity (based on 2–7 keV observed X-ray fluxes assuming an unabsorbed spectrum with photon index $\Gamma = 1.9$) and k_{bol} is a bolometric correction (a constant $k_{\text{bol}} = 25.0$ was assumed). The denominator of Equation 2 ensures that $\lambda \approx$ the Eddington ratio, under the assumption that $0.002\mathcal{M}_* \approx \mathcal{M}_{\text{BH}}$, the mass of the central black hole in a galaxy of stellar mass $\mathcal{M}_*$. Constant scaling factors are chosen such that $\lambda \propto \frac{L_X}{\mathcal{M}_*}$, which means λ can be described as the specific accretion rate, the amount of black hole growth relative to the mass of the galaxy, and is thus independent of underlying uncertainties or scatter in the $\mathcal{M}_*$–$\mathcal{M}_{\text{BH}}$ relation. As discussed in Section 1, adopting measurements in terms of specific accretion rate accounts for the $\mathcal{M}_*$-dependent selection bias in AGN samples, whereby an AGN in a more massive galaxy produces a higher observable luminosity and is thus easier to detect than AGN in less massive galaxies. In addition, the Bayesian methodology accurately accounts for the varying sensitivity of the X-ray observations—both within and between different fields—and probes down to ~an order of magnitude below the nominal sensitivity limits of the *Chandra* imaging (see also Aird et al. 2017b).

In this paper, we take the A18 measurements of $p(\lambda \mid \mathcal{M}_*, z)$, which describes the probability that a galaxy with a given stellar mass and redshift hosts an AGN as a function of λ, and integrate them down to fixed limits in λ to provide estimates of the *fraction* of AGN in a given galaxy sample (see also Aird et al. 2019). Such estimates provide a robust definition of the AGN fraction in terms of the number of galaxies that are growing their black holes above a certain threshold in λ, i.e. relative to the mass of the galaxy. In addition, we reverse Equation 2 to recover estimates of the AGN fraction to various limits in L_X. These estimates account for the varying sensitivity of the X-ray observations but do not allow for the broad $\mathcal{M}_*$-dependent selection biases. As such, they provide a robust estimate of the *observed* AGN fraction that allows for known instrumental effects but does not account for intrinsic selection biases in the sample.

Figure 2 shows the AGN fractions (shown here at $z = 0.75$) that we calculate based on the A18 measurements for all galaxies (of given $\mathcal{M}_*$ and z; left panel) and considering star-forming and quiescent galaxies as separate populations with a distinct AGN fraction as a function of $\mathcal{M}_*$ and z within each population (middle and right panels). To determine the AGN fraction within our UniverseMachine model galaxies, we linearly interpolate between the centres of the $\mathcal{M}_*$ bins from A18 to the values of each individual galaxy (extrapolating for $\mathcal{M}_* > 10^{11.5} \mathcal{M}_\odot$). We use the underlying posterior distributions of the $p(\lambda \mid \mathcal{M}_*, z)$ measurements and propagate these uncertainties into our estimates of the AGN fraction for UniverseMachine galaxies. Rather than creating a mock catalogue of AGN, we instead weight each galaxy by the AGN fraction, retaining the full fidelity of the underlying UniverseMachine galaxy sample and allowing us to efficiently propogate uncertainties in the AGN fraction. Shaded regions in Figure 2 (and subsequent figures) encompass the 68% central confidence interval based on the propagated posterior distributions.

As shown in Figure 2, the AGN fraction to fixed λ limits in all galaxies, or when considering star-forming galaxies only, is found to rise with increasing $\mathcal{M}_*$, i.e. there is an intrinsic $\mathcal{M}_*$ dependence such the AGN activity is more common in higher $\mathcal{M}_*$ galaxies, even after accounting for observational selection biases. The trend is slightly weaker when considering a higher $\lambda > 0.1$ limit for the AGN fraction (dark blue region), indicating that high-λ activity does not

---

[4] Specifically, we take the snapshots at scale factor $a = 0.766904, 0.571997, 0.445435, 0.364435$ and 0.308748 that correspond to redshifts $z = 0.3, 0.75, 1.25, 1.75$ and 2.25. We do not explore higher redshifts where the AGN fraction is poorly determined by the A18 observations.
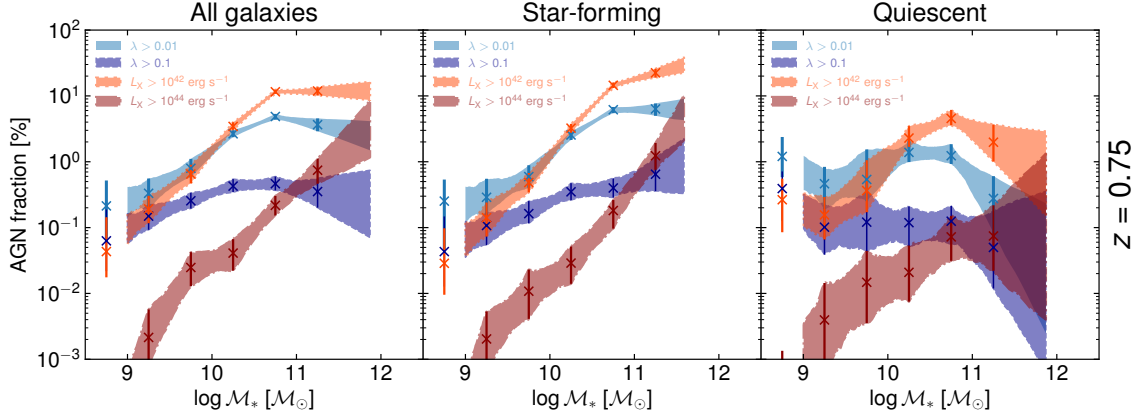
Figure 2. AGN fraction (shown here at $z = 0.75$) as a function of $\mathcal{M}_*$ in all galaxies (left), star-forming galaxies (center) and quiescent galaxies (right) for four "AGN samples" defined to different selection limits in terms of X-ray luminosity (L_X) or specific accretion rate ($\lambda \propto L_X/\mathcal{M}_*$). Crosses mark the robust measurements of the AGN fraction within observed galaxy samples from Aird et al. (2018) along with 1σ-equivalent uncertainties. The colored regions indicate our interpolation (or extrapolation) of the AGN fraction to the $\mathcal{M}_*$ of each galaxy in the UNIVERSEMACHINE sample, with the width of the region indicating the 1σ-equivalent uncertainties that are propoagated from the original measurements. We note the mild $\mathcal{M}_*$ dependence of the AGN fraction to λ limits in all and star-forming galaxies that becomes more exaggerated when adopting L_X limits. In quiescent galaxies the AGN fraction to λ limits is almost constant with $\mathcal{M}_*$; again a $\mathcal{M}_*$ dependence is introduced when adopting L_X limits and neglecting the observational bias against finding growing black holes in lower $\mathcal{M}_*$ galaxies.

depend so strongly on the galaxy mass. The $\mathcal{M}_*$-dependent trends are exaggerated when considering samples to limits in L_X due to the observational selection bias discussed above. For quiescent galaxies, the AGN fraction (to λ limits) is approximately constant with $\mathcal{M}_*$, with tentative evidence for a *decrease* at $\mathcal{M}_* > 10^{11} \mathcal{M}_\odot$. Adopting L_X limited samples introduces an artificial $\mathcal{M}_*$ dependence due to the selection bias. The same pattern is observed in all our redshift slices, although the stellar-mass dependence for λ-limited AGN samples in star-forming galaxies is somewhat milder for our lowest redshift slice ($z = 0.3$, see A18).

Throughout this paper, we do not consider AGN in galaxies with $\mathcal{M}_* < 10^9 \mathcal{M}_\odot$ or their host haloes. Measurements of the AGN fraction in such low-mass galaxies are highly uncertain, especially at $z \gtrsim 0.7$ (although see e.g. Mezcua et al. 2018; Birchall et al. 2020, for recent measurements at lower z). Given the higher space densities of such low-mass galaxies, the large uncertainties in the AGN fraction in this regime begin to dominate recovered quantities such as average halo masses (e.g. Section 5.2). We thus limit our analysis to a parent sample of galaxies with $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$. This mass limit introduces a low-mass turnover in the corresponding HMFs (see solid black lines in the bottom panels of Figure 1) that is propagated to our AGN HMFs (see Section 3). We note that observed AGN samples (e.g. X-ray selected, L_X-limited samples) are dominated by AGN in galaxies with $\mathcal{M}_* \gtrsim 10^{10} \mathcal{M}_\odot$ (Xue et al. 2010; Aird et al. 2013). In addition, only the relatively rare, high-λ AGN in lower mass galaxies ($\mathcal{M}_* \lesssim 10^9 \mathcal{M}_\odot$) would satisfy the luminosity limit of $L_X > 10^{42}$ erg s^{-1} typically adopted for X-ray AGN samples. Thus our model will provide a good description of observed AGN samples and enables us to make robust statements regarding the true properties of AGN within the $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxy population.

We note that being based on 2–7 keV observed energies, the A18 measurements will include both unobscured and moderately obscured AGN but will not include the contribution from heavily obscured, Compton-thick sources. Under the assumption that the intrinsic fraction of Compton-thick fraction of AGN does not strongly depend on λ (e.g. Ricci et al. 2017), our measurements can be used to infer the properties (e.g. halo mass functions) of *all* AGN by applying an appropriate scaling factor ($\sim 1.25 - 2.0$) to $p(\lambda \mid \mathcal{M}_*, z)$ and the

derived AGN fractions, SMFs and HMFs.[5] However, given remaining uncertainties on the overall Compton-thick fraction (ranging from $\sim 20 - 50\%$: Ueda et al. 2014; Ricci et al. 2015; Buchner et al. 2015; Aird et al. 2015; Ananna et al. 2019) and any intrinsic dependence on host galaxy properties (e.g. Koss et al. 2011; Kocevski et al. 2015) we do not apply such a correction here and restrict our analysis and conclusions to the radiatively efficient (i.e. X-ray luminous), Compton-thin AGN population.

Our method shares some similarities with recent work by Georgakakis et al. (2019), who used measurements of AGN specific accretion rate distributions from A18 and complementary work by Georgakakis et al. (2017) to populate mock galaxy catalogues with AGN (see also Comparat et al. 2019). While the conclusions of Georgakakis et al. (2019) are consistent with our model, we focus on a number of different aspects of the resulting halo mass functions and interpretation of the clustering properties of AGN. We use a more sophisticated model to link galaxies and their dark matter haloes, via UNIVERSEMACHINE, that directly models the distinct halo masses of star-forming and quiescent galaxy populations, whereas Georgakakis et al. (2019) adopt a single SMHM relation (with scatter) and separate by galaxy type in post-processing, thus neglecting systematic differences in the host haloes of these populations. In addition, we show explicitly how observational selection biases impact the results and we provide a detailed analysis of how to interpret measurements of AGN clustering and bias.

3 MODEL PREDICTIONS OF AGN HALO MASS FUNCTIONS AND CLUSTERING PROPERTIES

We now use the empirically motivated model detailed in the previous section to predict the HMFs of AGN samples selected to different L_X observational limits and more complete AGN samples selected to a range of λ limits. We also use our model to predict the clustering

---

[5] Assuming no intrinsic dependence of the Compton-thick fraction on host galaxy properties, our predicted clustering measurements and average halo masses should not change.

properties of AGN samples and compare these to the underlying galaxy population.

In this section, we present results primarily at $z = 0.75$, where our measurements of the AGN fraction are best constrained. However, the L_X and λ dependence, explored in Sections 3.1 and 3.2, is similar in the other redshift slices that we investigate. In Section 3.3 we compare the clustering strength of our AGN samples at different redshifts. We defer discussion of the redshift evolution of the HMF itself to Section 5.1 below.

3.1 L_X-limited AGN samples

Figure 3 shows the predicted HMFs of AGN samples selected to various L_X limits, comparing our two different approaches to populate galaxies with AGN. In the left panel, we use measurements of the AGN fraction in all galaxies, adopting the observed $\mathcal{M}_*$ dependence only (see left panel of Figure 2). In the right panel of Figure 3 we adopt the distinct measurements of the AGN fraction (as a function of $\mathcal{M}_*$) in star-forming and quiescent galaxies (see middle and right panels of Figure 2). The recovered HMFs with both modelling approaches show that L_X-limited samples of AGN are predicted to be found in haloes with a broad range of $\mathcal{M}_{\text{peak}}$, regardless of the adopted luminosity limit. For the $\mathcal{M}_*$-dependent only model, the width of the HMF that contains 90% of AGN is ~1.6 dex for $L_X > 10^{42}$ erg s^{-1} AGN, increasing to ~1.9 dex for $L_X > 10^{44}$ erg s^{-1} AGN. Adopting a distinct AGN fraction in star-forming and quiescent galaxies suppresses the number of AGN found in higher mass haloes ($\mathcal{M}_{\text{peak}} \gtrsim 10^{12} \mathcal{M}_\odot$), reducing the the width containing 90% of AGN to ~1.4–1.7 dex. These results differ from the $\mathcal{M}_*$-dependent only model (shown in the left panel) as AGN are preferentially assigned to star-forming galaxies, that tend to lie in lower $\mathcal{M}_{\text{peak}}$ haloes, and AGN are suppressed in quiescent galaxies, that tend to lie in more massive haloes. For both models, the AGN HMFs peak at $\mathcal{M}_{\text{peak}} \approx 10^{12} \mathcal{M}_\odot$, approximately 0.5 dex higher than the underlying $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ parent galaxy population (black solid lines), reflecting the strong preference of L_X-limited AGN samples to be found in higher $\mathcal{M}_*$ galaxies, due to both intrinsic effects and observational biases.

We also predict the *observable* clustering properties of L_X-limited AGN samples for a given simulation snapshot. Following the majority of observational work (see Section 4 below), we determine projected two-dimensional auto-correlation functions,

$$w_p(r_p) = 2 \int_0^{\pi_{\max}} \xi(r_p, \pi) d\pi \qquad (3)$$

where $\xi(r_p, \pi) \equiv \xi(r)$, the two-point correlation function as a function of co-moving distance r, which we separate into the projected on-sky distance, r_p (calculated from the x and y positions in co-moving Mpc in a given UNIVERSEMACHINE snapshot), and line-of-sight distance, π based on the z position in UNIVERSEMACHINE. Integrating $\xi(r_p, \pi)$ in the π direction accounts for redshift-space distortions to the true positions of galaxies and their AGN in real observational datasets. We adopt $\pi_{\max} = 60h^{-1}$ Mpc.

To predict realistic $w_p(r_p)$ measurements using our model, we first add redshift-space distortions to the positions of our UNIVERSEMA-CHINE model galaxies based on the peculiar velocities of their haloes in the z direction (using HALOTOOLS v0.7, Hearin et al. 2017). We then measure $w_p(r_p)$ from the co-moving co-ordinates using the CORRFUNC code (Sinha & Garrison 2020), which accounts for the weighting of each UNIVERSEMACHINE galaxy by the appropriate AGN fraction. The efficiency of the CORRFUNC code enables us to repeat our calculations of $w_p(r_p)$ for each of the posterior draws in our AGN fraction and thus fully propagate these uncertainties.

The top panels of Figure 4 show our model predictions for $w_p(r_p)$ for our various L_X-limited AGN samples. We also show the corresponding $w_p(r_p)$ of the underlying $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxy population (black dashed line), which is also calculated from UNIVERSEMACHINE using CORRFUNC, and for dark matter particles (grey dotted line) which is calculated from the matter power spectrum, including non-linear effects, given by Mead et al. (2016). The bottom panels of Figure 4 show the relative bias of the AGN samples relative to the $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxy population as a function of scale. It is assumed that

$$w_{p,\text{AGN}}(r_p) = b_{\text{rel}}^2(r_p) \, w_{p,\text{gal}}(r_p) \qquad (4)$$

where $w_{p,\text{AGN}}$ and $w_{p,\text{gal}}$ are the projected two-point correlation functions of the AGN sample and parent galaxy sample, respectively, and b_{rel} is the relative bias. With the $\mathcal{M}_*$-dependent only AGN fraction, we find the amplitude of $w_p(r_p)$ for the L_X-limited samples is significantly higher than the parent galaxy sample, reflecting the typically higher $\mathcal{M}_{\text{peak}}$ of the host haloes (as seen in the HMFs shown in Figure 3). We predict a lower amplitude for the relative bias when adopting a distinct AGN fraction in star-forming and quiescent galaxies, reflecting the reduction in the AGN incidence at higher $\mathcal{M}_{\text{peak}}$ seen in Figure 3. Figure 5 shows the average of the relative bias over scales of $1h^{-1}$ Mpc $< r_p < 10h^{-1}$ Mpc as a function of the L_X limit of the AGN samples, comparing our two model approaches. The $\mathcal{M}_*$-dependent only model predicts a mild rise in the relative bias as the L_X limit of the sample increases. Our more realistic model that adopts a distinct AGN fraction in star-forming and quiescent galaxies (as observed) predicts an approximately *constant* relative bias, with tentative evidence of an increase for the highest luminosities ($L_X > 10^{44}$ erg s^{-1}). In both cases we predict a mild— but significant—enhancement of the measured bias of L_X-limited AGN samples by ~3–15% compared to $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxies, reflecting the typically higher halo masses of X-ray selected AGN samples and their enhanced clustering properties.
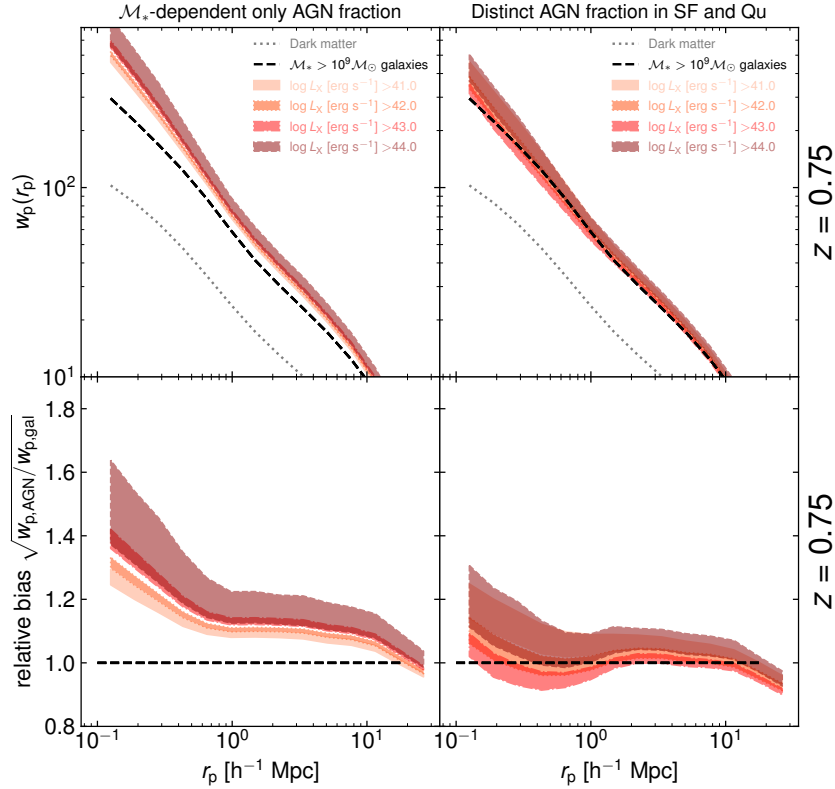
3.2 λ-limited AGN samples

Selecting AGN to a fixed limit in luminosity introduces an observational bias toward higher stellar mass hosts, and thus toward higher halo masses. Here, we instead present HMFs and the clustering properties of AGN samples selected to fixed limits in specific accretion rate, λ. Adopting λ limits allows us to identify galaxies as AGN based on how rapidly they are growing their black holes, countering the observational selection bias toward higher $\mathcal{M}_*$ hosts and providing estimates of the *underlying* distribution of AGN halo masses,[6] in contrast to *observed* samples of AGN explored in Section 3.1 above.

In Figure 6 we present HMFs for λ-limited AGN samples, using our two different methods to populate galaxies with AGN. As for the L_X-limited samples, the HMFs of λ-limited AGN samples are broad, with 90% of the AGN spanning $\gtrsim 1.4$ dex in $\mathcal{M}_{\text{peak}}$. Adopting our more accurate model, where we use a distinct AGN fraction in star-forming and quiescent galaxy populations, reduces the predicted number density of AGN at higher halo masses ($\mathcal{M}_{\text{peak}} \gtrsim 10^{12} \mathcal{M}_\odot$) as fewer AGN are placed in high-$\mathcal{M}_*$ quiescent galaxies, which generally inhabit higher $\mathcal{M}_{\text{peak}}$ haloes. For both models, the peak of the HMF moves toward lower $\mathcal{M}_{\text{peak}}$ as the λ limit is increased. This trend is due to the differences in the measured $\mathcal{M}_*$-dependence of

---

[6] As our analysis is restricted to galaxies with $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$, a more accurate statement is that we present the underlying distribution of AGN halo masses *within this parent galaxy sample*.
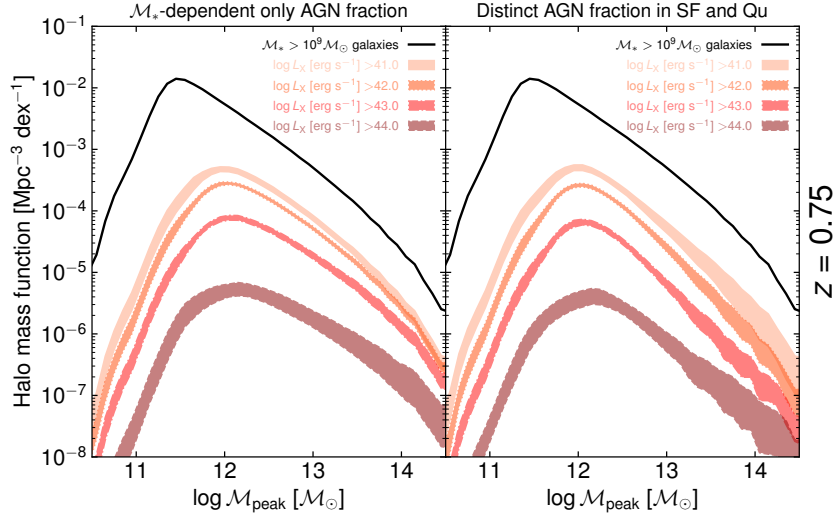
Figure 3. Halo mass functions (differential number density as a function of $\mathcal{M}_{\mathrm{peak}}$) for a range of L_{X}-limited AGN samples, as indicated by different coloured regions, using our two different methods to populate galaxies with AGN. **Left panel:** using measurements of the AGN fraction in all galaxies as a function of $\mathcal{M}_*$ only. **Right panel:** placing AGN in galaxies depending on their classification (star-forming or quiescent) and using the distinct measurements of the AGN fraction (as a function of $\mathcal{M}_*$) in each population. Adopting an $\mathcal{M}_*$-dependent only AGN fraction tends to place a higher number of AGN in high-$\mathcal{M}_*$ quiescent galaxies, which tend to be found in more massive (higher $\mathcal{M}_{\mathrm{peak}}$) haloes. Adopting the distinct measurements of the AGN fraction in star-forming and quiescent galaxies preferentially assigns AGN to massive star-forming galaxies and thus suppresses the high-$\mathcal{M}_{\mathrm{peak}}$ tail of the AGN halo mass function. For both of our methods, the AGN halo mass function reaches a maximum at slightly higher $\mathcal{M}_{\mathrm{peak}}$ than the underlying $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ parent galaxy population (solid black line), reflecting the increased fraction of AGN (to a given L_{X} limit) in higher $\mathcal{M}_*$ galaxies.

Figure 4. Top: Measurements of the projected clustering amplitude as a function of projected distance, $w_P(r_P)$, for a range of L_{X}-limited AGN samples (as indicated by the coloured regions), using our two different methods to populate galaxies with AGN (left versus right panels, as indicated). The black dashed lines show $w_P(r_P)$ of the parent $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxy sample, while the grey dotted lines show $w_P(r_P)$ of dark matter, including non-linear effects (based on Mead et al. 2016). **Bottom:** Relative bias of the AGN samples as a function of r_P, compared to the parent $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxy sample.
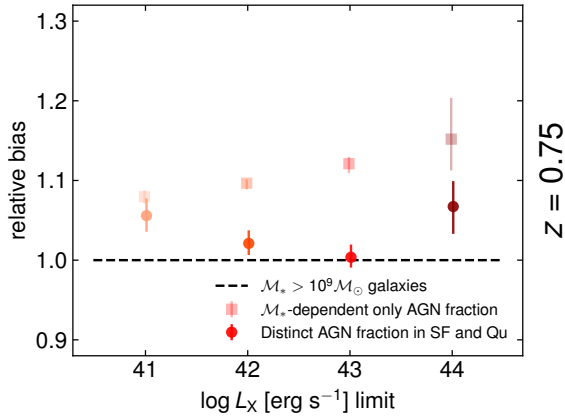
Figure 5. Relative bias of AGN compared to the parent $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxy sample as a function of the L_X limit, averaged over scales of $1h^{-1}$ Mpc $< r_p < 10h^{-1}$ Mpc, comparing our two different methods to populate galaxies with AGN (squares and circles, as indicated). The relative bias shows a mild dependence on L_X for the $\mathcal{M}_*$-dependent only AGN fraction; the relative bias is generally lower and the L_X dependence is weaker when adopting the distinct AGN fractions in star-forming and quiescent galaxies.

the fraction of AGN for different λ limits. The incidence of moderate accretion rate black holes (e.g. $\lambda > 0.01$) is enhanced in higher stellar mass ($\mathcal{M}_* \gtrsim 10^{10.5} \mathcal{M}_\odot$) star-forming galaxies, whereas the incidence of more rapidly accreting black holes does not show this mass dependence and thus the AGN HMFs approach the same shape as the underlying parent galaxy population (albeit with a much lower normalisation). Such behaviour is also seen in the simpler $\mathcal{M}_*$-dependent only model.

We also derive $w_p(r_p)$ for our λ-limited samples as described in Section 3.1 above. In Figure 7 we present the relative bias (compared to the parent galaxy population, averaged over $1h^{-1}$ Mpc $< r_p < 10h^{-1}$ Mpc) as a function of the λ limit for the two model approaches. For our most accurate model (adopting the distinct AGN fractions in star-forming and quiescent galaxies, as observed), we predict a relative bias that is consistent with 1 (at $z = 0.75$), regardless of λ limit, *consistent with there being no observable difference in the clustering properties of AGN and the underlying galaxy population.*

3.3 Comparison between L_X- and λ-limited AGN samples and the underlying galaxy population

In Figure 8 we directly compare the HMFs for L_X-limited and λ-limited AGN samples as predicted by our more sophisticated and accurate model (adopting a distinct AGN fraction in star-forming and quiescent galaxies). Differences in the overall normalisation reflect differences in the fraction of haloes that contain AGN, depending on the definition, with higher L_X and higher λ samples thus corresponding to lower overall normalisations. However, there are also significant differences in *shapes* of the HMFs, seen most clearly in the comparison between the high-luminosity AGN sample ($L_X > 10^{44}$ erg s^{-1}; dark red) and the complete sample of highly accreting black holes ($\lambda > 0.1$; dark blue) which are found in significantly greater numbers in lower mass haloes ($\mathcal{M}_{peak} < 10^{12} \mathcal{M}_\odot$).

In Figure 9 we compare the predicted clustering properties for these four AGN samples at $z = 0.75$. The differences in $w_p(r_p)$ are relatively small but we are able to identify clear systematic differences, which are seen most clearly in the relative bias compared

to the clustering of galaxies (bottom panel). The clustering of λ-limited AGN samples is consistent with the clustering of the parent $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxy population, whereas the introduction of an L_X-limit (corresponding to a observational selection effect) leads to a increase in the bias, of up to $\sim 10\%$ at scales of $\gtrsim 1h^{-1}$ Mpc.

In Figure 10 we compare both the relative bias and the absolute bias (i.e. bias relative to the underlying dark matter distribution) as a function of redshift for the four AGN samples and the underlying parent galaxy sample, where all bias measurements are averaged over scales of $1h^{-1}$ Mpc $< r_p < 10h^{-1}$ Mpc. We see a strong increase in the absolute bias toward higher redshift for all four AGN samples. Such an increase is expected and also seen in the parent galaxy population: at higher redshift, galaxies (of a given $\mathcal{M}_*$) and the dark matter haloes that they lie in correspond to more extreme overdensities relative to the underlying dark matter distribution. However, the *relative* bias of all four AGN samples, compared to galaxies, is also seen to increase with redshift (bottom panel). At $z < 1$, the clustering properties of the λ-limited AGN samples are consistent with the parent galaxy population; at higher redshifts these AGN samples are more strongly clustered, with the relative bias increasing to ~ 1.2 (for $\lambda > 0.01$ AGN) at $z > 2$, although we note the larger uncertainties at higher z (due to the larger uncertainties in measurements of AGN fractions for higher redshift galaxies that are propagated into our model predictions). The L_X-limited samples generally have a higher bias than the λ-limited samples, due to the observational selection effects in such samples, although given the large uncertainties we only find a significant difference for our two best-sampled redshift bins (at $z = 0.75$ and $z = 1.25$).

Overall, at higher redshifts AGN activity shows a slight preference toward higher halo masses and thus denser environments relative to galaxies. Such evolution is due to changes in the $\mathcal{M}_*$-dependence of AGN activity, whereby the fraction of AGN in higher mass starforming galaxies ($\mathcal{M}_* \gtrsim 10^{10.5} \mathcal{M}_\odot$) increases more strongly with redshift than in lower mass galaxies (see e.g. figure 6 of A18), in addition to the increased dominance of such star-forming galaxies compared to quiescent galaxies (at a given $\mathcal{M}_*$) toward higher redshift. The prevalence of AGN activity increases toward higher redshift due to changes in the properties of the *galaxies* they lie in (e.g. higher gas fractions leading to higher SFRs and a greater rate of AGN triggering). We do *not* infer any direct connection between the large-scale environment and an increased incidence of AGN activity.

4 COMPARISON TO OBSERVED AGN CLUSTERING MEASUREMENTS

We next compare the clustering predictions of our model with observed AGN clustering measurements. We focus first on comparing to L_X-limited AGN samples that span a range of luminosity and redshift. In Figure 11 in the left panel we present a variety of published bias measurements for L_X-limited samples from $z \sim 0$ to $z \sim 2$. The observed bias values generally increase with redshift, from ~ 1 at $z \sim 0$ to ~ 3 at $z \sim 2$. The dominant observational errors are cosmic variance and Poisson noise due to small AGN sample sizes. These studies span a range of luminosities, though they typically target AGN with $L_X \sim 10^{43}$ erg s^{-1}. We also show predictions for the absolute bias as a function of redshift for our preferred model, for two values of L_X. In our preferred model the clustering amplitude is a mild function of the L_X limit of the AGN sample, rising only at the highest luminosities. As can be seen in the figure, the predictions from our model match the observational measurements of the absolute bias well, for a wide range of observational studies.
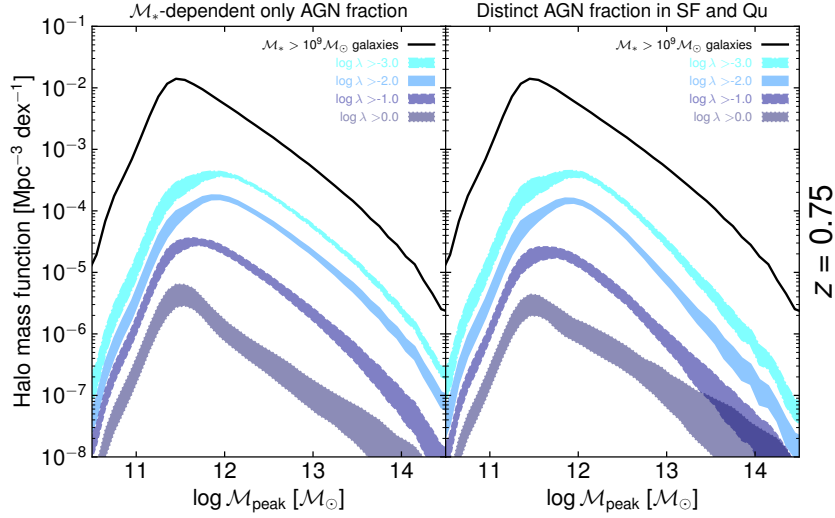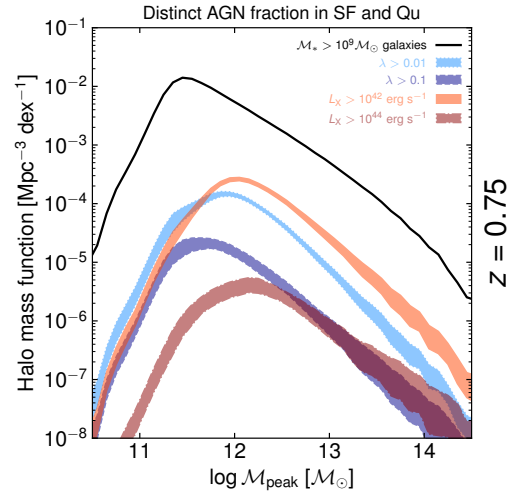
Figure 6. Halo mass functions (differential number density of AGN as a function of $\mathcal{M}_{\text{peak}}$) for a range of λ-limited AGN samples, as indicated by different coloured regions, using our two different methods to populate galaxies with AGN. For higher λ limits the shape of the halo mass function approaches the shape of the underlying $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxy sample (black line)—albeit with a much lower normalisation—reflecting the weaker $\mathcal{M}_*$ dependence of the AGN fraction for such samples.
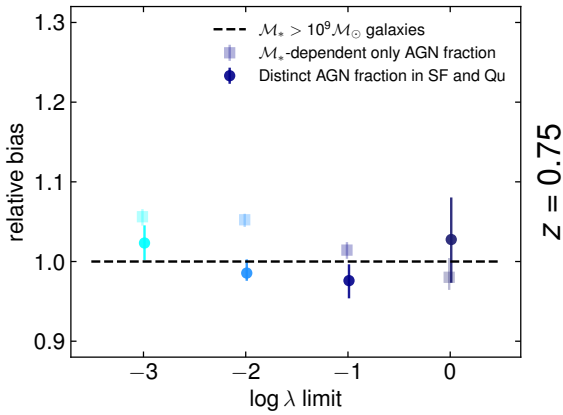
Figure 7. Relative bias of AGN compared to the parent $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxy sample as a function of the λ limit, averaged over scales of $1h^{-1}$ Mpc $< r_p < 10h^{-1}$ Mpc, comparing our two different methods to populate galaxies with AGN. A mild *decrease* in the relative bias with increasing λ is found for the $\mathcal{M}_*$-dependent AGN fraction (squares). When assigning AGN independently in star-forming and quiescent galaxies we find the relative bias is consistent with no dependence on λ and at $z = 0.75$ is consistent with the parent galaxy population.


Figure 8. Halo mass functions for four "AGN samples" to different L_X and λ limits, using our preferred method to populate galaxies with AGN (using the distinct measurements of AGN fraction in star-forming and quiescent galaxies). The halo mass functions for λ-limited samples—that account for $\mathcal{M}_*$-dependent AGN selection biases—peak at lower $\mathcal{M}_{\text{peak}}$ and more closely track the halo mass function of the underlying galaxy population.

In the right panel of Figure 11 we show the same absolute bias values from the literature, here displayed as a function of L_X. To highlight the differences in the redshifts of each study, we colour code the data points from blue at the lowest redshifts to red at the highest redshifts. The coloured regions show our model predictions at different redshifts, as a function of luminosity. As discussed above, our model has little, if any, luminosity dependence, and as can be seen it agrees well with the lack of luminosity-dependence in the observational results. Our model does not extend to $z \sim 0$, where the observational data are shown in dark blue. We note in particular the good agreement with the $z = 1.5 - 2.0$ results (from Krishnan et al. 2020 and Powell et al. 2020, shown in purple in the right panel), which do show a substantial rise in the absolute bias from $L_X \sim 10^{44} - 10^{45}$ erg s^{-1}.

To produce such high luminosities requires massive galaxies (and thus massive black holes) with high accretion rates. Such sources are thus restricted to the most extreme and massive haloes, especially at high-z, leading to our predicted rise in the absolute bias, consistent with the observed data.

We compare in Figure 12 the predictions from our model to clustering measurements of optically-selected quasar samples. Such quasar samples likely suffer from a combination of selection biases, selecting high-luminosity, unobscured quasars whose light dominates that of their host galaxies (i.e. high-λ). In order to compare with these observational samples, we restrict our model predictions in this figure to both a high luminosity *and* high λ AGN sample, with
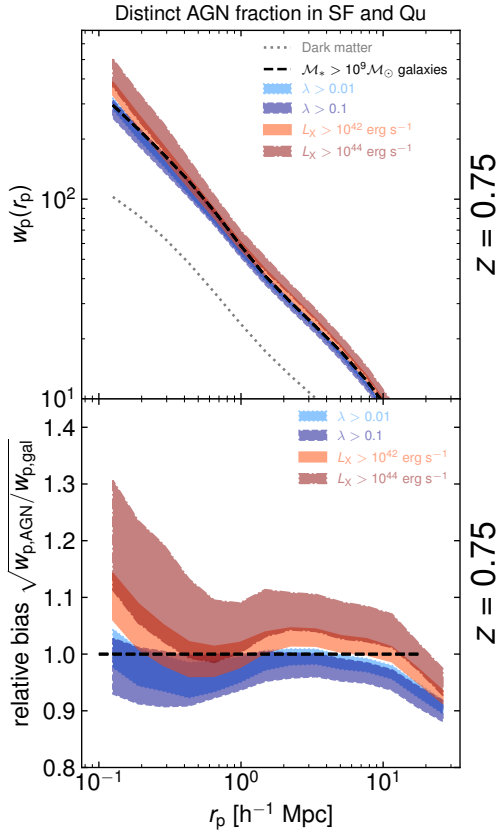
Figure 9. Top: Measurements of $w_p(r_p)$, comparing L_X-limited and λ-limited samples. **Bottom:** Relative bias as a function of r_p, relative to the parent $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxy sample, for the four AGN samples. The clustering amplitude of L_X-limited AGN samples are systematically higher than for λ-limited samples that account for observational biases.

a bolometric luminosity of $L_{\rm bol} > 10^{45}$ erg s^{-1} and $\lambda > 0.1$. This combination of a luminosity cut and a λ limit brings our model into excellent agreement with the observational results and in particular the recent high-precision measurements from the SDSS-IV eBOSS quasar samples at $z > 1.5$ (Laurent et al. 2017). We note that—given the uncertainties in our model and the mild dependence of the bias on either λ of L_X (e.g. Figure 10)—a simple, high luminosity cut can also provide good agreement with the observed data.

It has been noted in the literature that quasar clustering results are surprisingly consistent with quasars residing in halos of roughly constant mass across the bulk of cosmic time (e.g., Porciani et al. 2004; Eftekharzadeh et al. 2015). We show here that this is a natural outcome of the combination of the selection biases inherent to optical quasar samples and the known AGN occupation of galaxies. We also show that the halo mass distribution is wide, and that quasars (and all AGN) reside in a range of halo masses at all redshifts. The width of this distribution is not obvious from a simple statistic like the bias.

As noted above, our model predicts small ($\sim 10\%$) systematic differences between L_X-limited and λ-limited AGN samples. As shown in Figure 10, at intermediate redshift L_X-selected AGN samples have slightly higher bias than λ-selected AGN samples. Observationally, these are similar to typical L_X-limited and optical quasar samples. Our model predictions could therefore explain some of the indications in the literature that X-ray selected AGN samples have slightly elevated clustering compared to optical quasar samples (e.g., Coil et al. 2009). However, to test this observationally one should use AGN
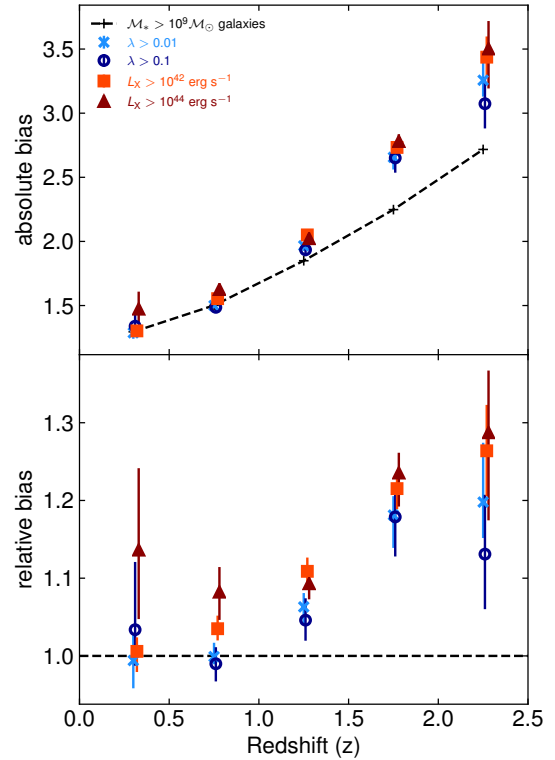

Figure 10. Absolute bias (relative to dark matter, top panel) and relative bias (relative to $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxies, bottom panel) as a function of redshift, averaged over $1h^{-1}$ Mpc $< r_p < 10h^{-1}$ Mpc, for four "AGN samples" defined to different L_X and λ limits, using our second (preferred) method to allocate AGN independently to star-forming and quiescent galaxies (as a function of $\mathcal{M}_*$). The absolute bias of the AGN samples increases to higher redshifts, predominantly driven by the fact that *galaxies* (black dashed line) are more strongly clustered relative to the underlying dark matter. For the L_X-limited samples the bias predicted by our modelling lies systematically above the λ-limited samples. The relative bias of λ-limited samples also increases to higher redshift due to the increased prominence of star-forming galaxies and the stronger $\mathcal{M}_*$ dependence of the AGN fraction in this population.

samples identified from the same parent galaxy survey, with overlap in the optical and X-ray selected AGN samples, and any differences in luminosity between the samples should be taken into account (e.g., Krumpe et al. 2012). Regardless, the differences between the clustering amplitudes of optical quasar and X-ray selected AGN samples should *not* be interpreted as different stages of AGN evolution (e.g., Hickox et al. 2009) but rather as due to differences in the selection biases of AGN identified in different ways.

5 DISCUSSION AND INTERPRETATION

Having established our forward model approach to populate haloes with AGN, based on measurements of AGN fractions as a function of galaxy properties, and shown that the clustering properties predicted by this model are in good agreement with previous studies, we now move on to discuss some of the implications of our model. In Section 5.1 we discuss the properties of the haloes that host AGN, in particular focussing on the important distinction between the (sub-)halo containing an individual galaxy (and its AGN) and the larger parent dark matter halo that defines the broader scale environment. In Section 5.2 we show how commonly-used methods that convert
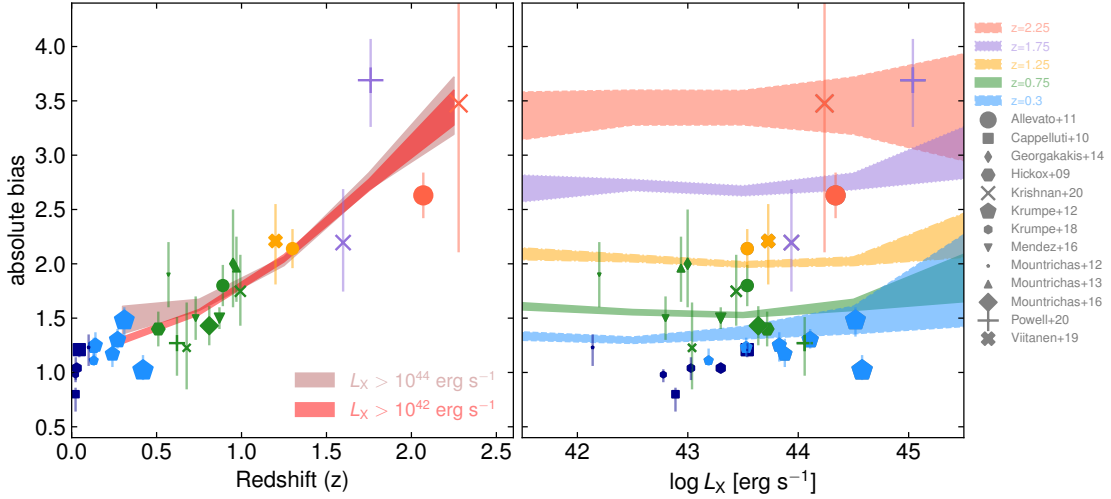
Figure 11. Our model predictions for the absolute bias as a function of redshift (left, coloured regions indicating two L_X-limited AGN samples) and as a function of X-ray luminosity (right, coloured regions indicating different redshifts), compared to observational measurements of the bias from X-ray selected samples. Symbol types indicate the prior study and the sizes are scaled based on the average luminosity of the AGN in these samples. The symbol colours associate the observational measurements to one of our redshift bins (measurements at $z < 0.1$, below the range of our AGN fraction measurements, are indicated in dark blue). Our predictions are consistent with direct measurements of X-ray AGN clustering and the overall increase in the absolute bias to higher redshift. Our model predicts little, if any, dependence of the absolute bias on X-ray luminosity, which is generally consistent with the findings of prior studies.
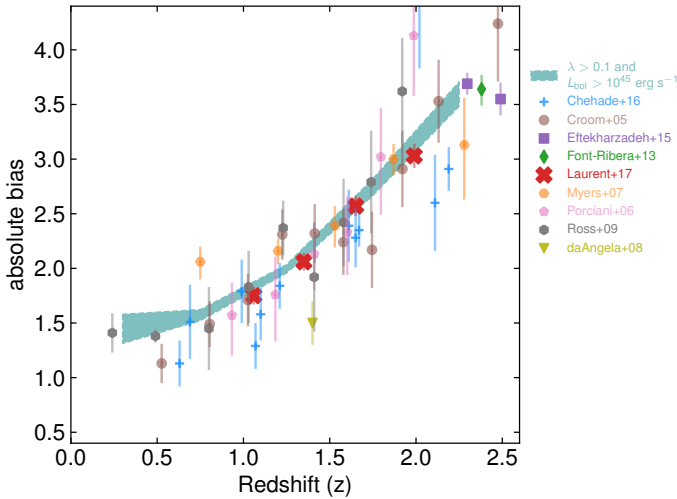
Figure 12. Our model prediction for the absolute bias as a function of redshift for a "quasar sample", which we approximate with cuts of both $\lambda > 0.1$ and bolometric luminosity $L_{bol} > 10^{45}$ erg s^{-1} (teal region), compared to measurements of the bias for optically-selected quasar samples (symbols and colours indicating the different studies, as shown). Our model prediction is generally consistent with observed quasar samples out to $z \sim 2.5$. We do not scale the symbol sizes to indicate differences in the luminosity limits of the observed quasar samples, in contrast to Figure 11 for X-ray selected samples, although we note that none of the studies find a strong luminosity dependence at a fixed redshift, consistent with our model.

measurements of the absolute bias to a typical halo mass provide an inaccurate and incomplete picture of the hosts of AGN, neglecting the underlying distribution of halo masses, the AGN satellite fraction, and AGN selection effects. In Section 5.3 we use our model to predict the halo occupation distribution (HOD) of AGN, present predictions of the AGN satellite fraction, and discuss the advantages and limitations of

HOD modelling in the interpretation of AGN clustering measurements. Finally, we discuss the importance of interpreting AGN clustering measurements compared to the underlying galaxy population in Section 5.4. .

### 5.1 Parent and sub-halo mass distributions

In the results presented in Section 3 above we adopted the peak historical (sub-)halo mass, $\mathcal{M}_{peak}$, as our tracer of the halo mass when presenting the HMFs of AGN. $\mathcal{M}_{peak}$ traces the overall growth of the (sub-)halo in which a galaxy forms and is thus closely related to the overall stellar build-up of a galaxy (i.e. the SFR and total $\mathcal{M}_*$). The triggering of AGN activity requires gas to be driven into the very central region of a galaxy and it is thus the small-scale environment—i.e. the host galaxy properties—that should determine the incidence of AGN; this assumption is inherent in our model. However, it is the mass of the larger parent halo that an AGN host galaxy lies in, $\mathcal{M}_{parent\ halo}$, that traces the large-scale environment and thus determines the *clustering* properties of AGN samples.

In Figure 13 we present HMFs using different definitions of the halo mass. Massive parent haloes contain both a central galaxy and a large number of satellite galaxies. As such, a single parent halo is counted multiple times when constructing the *parent* halo mass function of all galaxies above a chosen $\mathcal{M}_*$ threshold (solid black line), producing a relatively flat distribution extending out to $\mathcal{M}_{parent\ halo} \sim 10^{14} \mathcal{M}_\odot$. When the current virial (sub-)halo mass (i.e. the parent halo mass for central galaxies but the mass of the individual sub-halo of a satellite galaxy) is adopted instead, the HMF has a much steeper high-mass slope and a tail at masses below $\mathcal{M}_{(sub-)halo} \sim 3 \times 10^{11} \mathcal{M}_\odot$ (black dotted line).

AGN samples exhibit a similar behaviour; our prediction for $L_X > 10^{42}$ erg s^{-1} is shown in Figure 13. The parent halo mass function, including AGN in both centrals and satellites, has a broad tail extending to high masses (purple region), whereas the mass function for the individual (sub-)haloes of AGN extends to lower masses (red region), due to differences in the parent halo masses
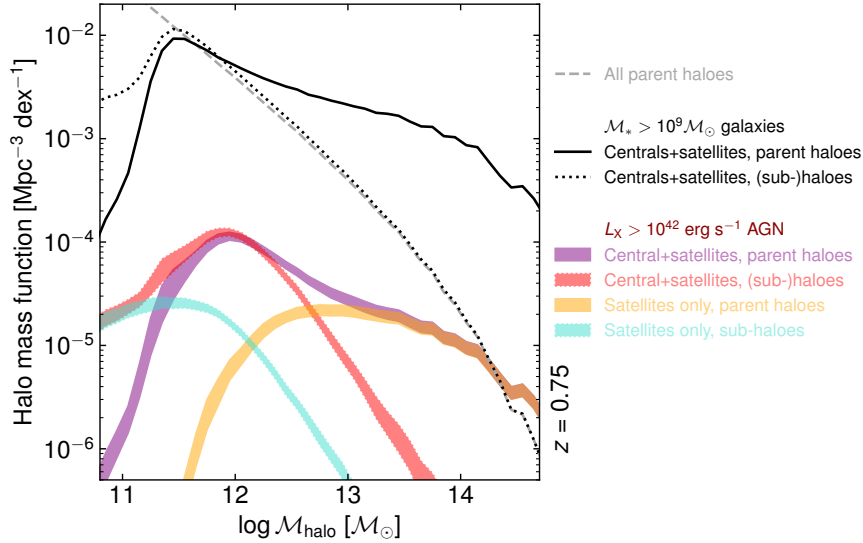
Figure 13. Comparison of predicted HMFs at $z = 0.75$ for different definitions of the halo mass. The grey dashed line shows the mass function of all individual parent haloes in the simulation. The solid black line shows the parent-halo mass function for all $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxies (adopting the current virial mass of the halo); high-mass parent haloes contain multiple galaxies (both centrals and satellites) and are thus counted multiple times for this HMF, resulting in the broad flat shape. The black dotted line shows the halo mass function when adopting the current virial (sub-)halo mass (the parent mass for central galaxies and the sub-halo mass for satellites), which has a much steeper shape at high masses and a tail extending to lower mass. The purple region shows the HMF of parent haloes for $L_X > 10^{42}$ erg s^{-1} AGN (in both centrals and satellites), whereas the red region adopts the current (sub-)halo mass. The yellow and cyan regions show the very different HMFs when considering the parent and sub-halo mass, respectively, for only the AGN that lie in satellite galaxies.
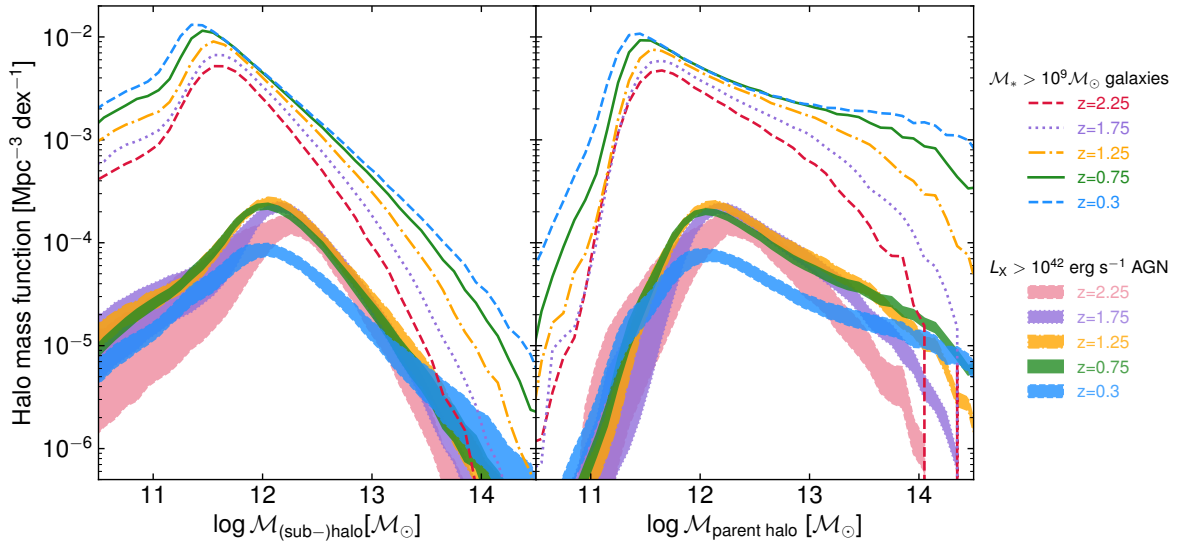
Figure 14. HMFs of galaxies (lines) and $L_X > 10^{42}$ erg s^{-1} AGN (shaded regions) as a function of redshift (as indicated by colours) for the (sub-)halo mass (left) and parent halo mass (right).

(gold region) and sub-halo masses (cyan region) of AGN in satellite galaxies. *The typical parent halo mass of AGN is thus significantly higher than the typical (sub-)halo mass*: averaging over the HMFs to calculate the mean halo mass gives $\langle \mathcal{M}_{\text{parent halo}} \rangle \approx 10^{13} \mathcal{M}_\odot$ versus $\langle \mathcal{M}_{\text{(sub-)halo}} \rangle \approx 2 \times 10^{12} \mathcal{M}_\odot$ for $L_X > 10^{42}$ erg s^{-1} AGN at $z = 0.75$. The difference in these masses reflects the fact that AGN are often found in satellite galaxies that lie within massive parent haloes (see Section 5.3 below, see also Alam et al. 2020a).

In Figure 14 we explore how the parent and (sub-)halo mass functions evolve with redshift. The normalization of the galaxy HMF (coloured lines) increases at all masses as cosmic time progresses

(i.e. as redshift decreases), reflecting the ongoing hierarchical build up of dark matter structure. The rate of increase is especially strong for parent halo mass functions at higher masses ($\mathcal{M}_{\text{parent halo}} \gtrsim 10^{13} \mathcal{M}_\odot$), which is due to the build up of the satellite galaxy population as smaller haloes fall into larger parent haloes. The most massive parent haloes, $\mathcal{M}_{\text{parent halo}} \gtrsim 10^{14} \mathcal{M}_\odot$, are yet to form at $z \gtrsim 2$, leading to an even stronger evolution as this population grows.

The HMFs of AGN follow a different evolutionary pattern. The (sub-)halo mass function (shaded regions in the left panel of Figure 14) remains roughly constant in both normalisation and shape between $z \sim 2$ and $z \sim 0.75$. At this epoch, AGN activity is found across

the galaxy population and over a wide range of halo masses, albeit with an increased triggering rate in higher stellar mass galaxies (an effect that is further exaggerated by observational selection effects). At lower redshift, the overall normalisation of the (sub-)halo mass function drops by a factor ~3. While the large-scale structure of the Universe (and the galaxies that lie within it) builds up progressively with time, AGN are short-lived and the rate at which they are triggered drops significantly at later cosmic times. The peak of the (sub-)halo mass function also shifts toward slightly lower values as the $\mathcal{M}_*$-dependence of the AGN fraction becomes weaker at lower redshifts, as well as becoming broader due to the build-up of the highest mass haloes. The parent halo mass function of AGN (shaded regions in the right panel of Figure 14) follows broadly similar patterns, except at the highest masses ($\mathcal{M}_{\mathrm{parent\ halo}} \gtrsim 10^{13} \mathcal{M}_\odot$) where the build up of the most massive parent haloes—in particular by the accretion of satellite galaxies that often host AGN—leads to an overall increase.

Regardless of which halo mass definition is used, it is clear that AGN are found in haloes with a wide range of masses. While clustering measurements broadly reflect the underlying distribution of parent halo masses, to understand the connections between AGN and galaxy evolution it is vital to understand the sub-haloes that the host galaxies and their AGN reside in.

5.2 Interpretation of AGN bias measurements

Our forward modelling approach allows us to predict the absolute bias that would be measured for different AGN samples using the standard observational approach: measuring the projected two-point correlation function of AGN, $w_p(r_p)$, and comparing to the corresponding correlation function of dark matter at scales of $r_p \gtrsim 1h^{-1}$ Mpc. In Section 4 we showed that our predicted bias measurements are in good agreement—after applying appropriate observational limits—with a wide range of results for both X-ray AGN samples and optically-selected quasar samples out to $z \sim 2.5$. Many prior studies then use their measurements of the absolute bias to infer the "typical" halo mass of AGN using a simplistic comparison to dark matter simulations (e.g. Croom et al. 2005; Cappelluti et al. 2010; Georgakakis et al. 2014). Here, we show explicitly that using a measured bias to directly infer a halo mass, without taking into account the satellite fraction or underlying mass distribution, leads to estimates that are systematically offset from the true averages of *either* the (sub-)halo or parent halo masses (see also DeGraf & Sijacki 2017; Powell et al. 2020).

To convert a measured absolute bias into a typical halo mass, a common approach in AGN clustering studies uses the relationship between halo mass and bias derived from N-body simulations (e.g. Sheth et al. 2001; Tinker et al. 2010; Comparat et al. 2017). Results from simulations estimate the bias for a *narrow* distribution of halo masses, which differs substantially from the broad distributions of halo masses for AGN samples that are predicted by our model. More importantly, such studies implicitly assume one galaxy per halo, i.e. only central galaxies are considered. In Figure 15 we show the Tinker et al. (2010) relationship between $\mathcal{M}_{\mathrm{halo}}$ and absolute bias, evaluated at $z = 0.75$, which we compare to the median (sub-)halo and parent halo masses predicted by our model (black circles and purple triangles, respectively) and the corresponding absolute bias that we measure in a manner analogous to observational studies (i.e. via projected two-point correlation functions). The median halo masses can vary by *up to 0.5 dex* depending on the selection method (L_{X}-limited or λ-limited) and whether parent or (sub-)halo mass is considered. The green crosses show the estimates of $\mathcal{M}_{\mathrm{halo}}$ that would be inferred using the Tinker et al. (2010) relation, which are significantly above
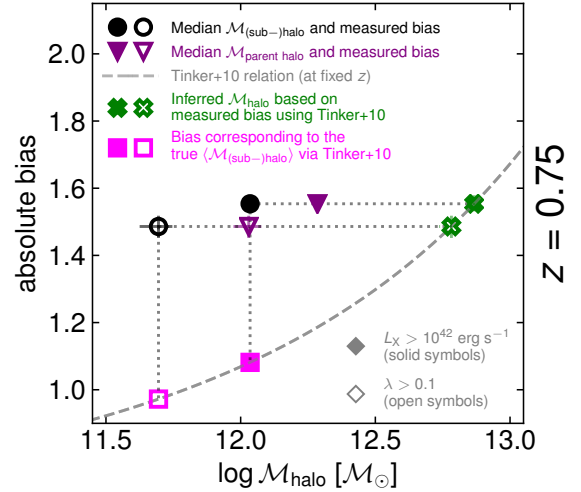

Figure 15. Demonstration of how measurements of the absolute bias may be used to (erroneously) infer the typical halo mass of AGN samples. The solid black circle indicates our model prediction of the measured bias for an $L_{\mathrm{X}} > 10^{42}$ erg s^{-1} AGN sample and the median $\mathcal{M}_{(\mathrm{sub}-)\mathrm{halo}}$ based on our halo mass functions (open symbols show the equivalent measurements for a $\lambda > 0.1$ AGN sample). The purple triangles indicate the median $\mathcal{M}_{\mathrm{parent\ halo}}$ for the same AGN samples. The grey dotted line shows the relationship between bias and halo mass derived by Tinker et al. (2010) using narrow bins in halo mass. This relationship is often used to infer a "typical" halo mass (green cross) for an AGN sample that corresponds to neither the true (sub-)halo mass nor the true parent halo mass. The pink squares shows how we can instead determine a bias corresponding to the true median $\mathcal{M}_{(\mathrm{sub}-)\mathrm{halo}}$ using the Tinker et al. (2010) relation.

the median $\mathcal{M}_{(\mathrm{sub}-)\mathrm{halo}}$ and $\mathcal{M}_{\mathrm{parent\ halo}}$. Using such a relation to infer a halo mass from the absolute bias ignores the broad distribution of halo masses and does not account for the fraction of AGN that lie in satellite galaxies within massive parent haloes. We also note that *mean* halo masses are significantly skewed by the high-mass tail of the HMF, especially in the case of the mean AGN parent halo mass. We choose to report median halo masses as these are more representative of the bulk of AGN and are robust against uncertainties in the shape of the HMF for the most massive host haloes. We note that none of the mean, median or mode halo masses, for parent haloes or (sub-)haloes, are in agreement with the "typical" halo mass inferred from the bias using the Tinker et al. (2010) relation.

In Figure 16 we show how these measures of halo mass—the median $\mathcal{M}_{(\mathrm{sub}-)\mathrm{halo}}$ and $\mathcal{M}_{\mathrm{parent\ halo}}$ as well as the $\mathcal{M}_{\mathrm{halo}}$ that is inferred using the Tinker et al. (2010) relation—vary depending on the chosen L_{X} (left) or λ (right) limit, shown at $z = 0.75$. We find little dependence on the chosen L_{X} limits (at a fixed redshift), reflecting the relatively minor changes in the shapes of the underlying HMFs (see Section 3), although we note that subtle (but significant) changes in the shape are not captured in a single statistic such as the median. In contrast, we find a slight *decrease* in the median (sub-)halo masses with increasing λ limit, which is related to the weaker dependence on stellar mass of the AGN fraction for high-λ samples (see Figure 2 and discussion in Section 2.2) that brings the AGN HMFs closer to the HMF of the underlying galaxy population (see Figure 6).

Figure 17 shows how our different halo mass estimates vary as a function of redshift for the $L_{\mathrm{X}} > 10^{42}$ erg s^{-1} (left) and $\lambda > 0.1$ (right) AGN samples. The median (sub-)halo masses (black circles), are found to be roughly constant with redshift for both samples, reflecting the lack of evolution in the AGN halo mass functions with
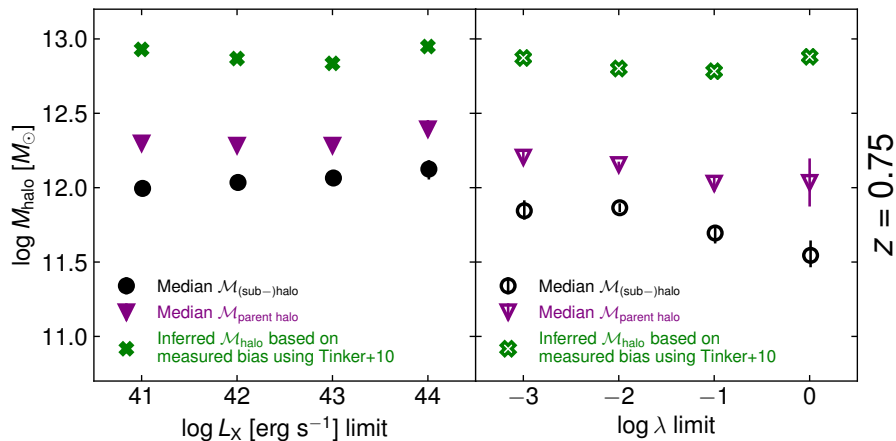
Figure 16. Halo mass as a function of L_X limits (left) and λ limits (right) at $z = 0.75$, showing the true median mass of (sub-)haloes (black circles) and parent haloes (purple triangles) from our model compared to the halo mass that would be inferred based on the measured bias using the Tinker et al. (2010) relation (green crosses).
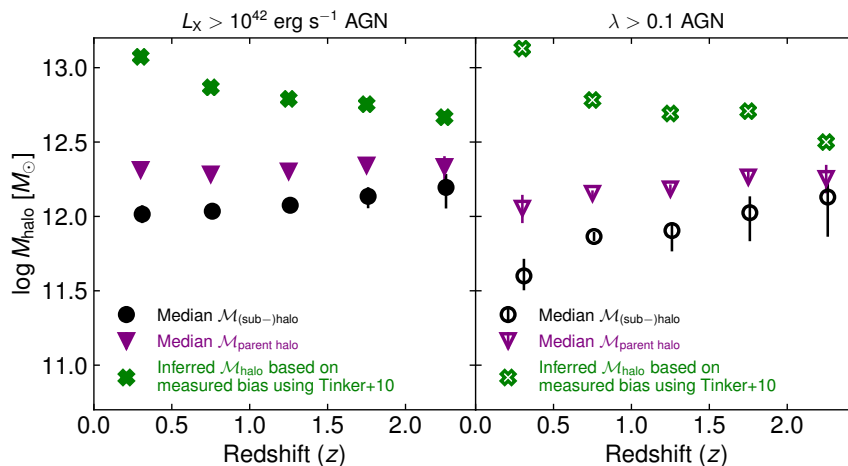
Figure 17. Halo mass as a function of redshift, showing results for the $L_X > 10^{42}$ erg s^{-1} (left panel) and $\lambda > 0.1$ (right panel) samples. The median parent halo mass of AGN samples (purple triangles) is found to decrease with increasing redshift, whereas the mean (sub-)halo mass (black circles) remains relatively constant. Inferring M_{halo} directly from the bias (e.g. using Tinker et al. 2010, ; green crosses) provides an inaccurate estimate of AGN halo masses.

redshift (see Figure 14). However, there is evidence of a decrease at the lowest redshift for the $\lambda > 0.1$ sample that corresponds to the reduced stellar mass dependence of the AGN fraction for such samples at the lowest redshifts. The immediate environment in which AGN activity is triggered, traced by the mass of the host (sub-)haloes, does not appear to change substantially with cosmic time. However, we note that AGN are found over a broad range of (sub-)halo masses and there *are* important changes in the triggering rate depending on the properties of the host *galaxies*, which will be closely related to the immediate environment of the supermassive black holes and the efficacy of different physical processes to bring gas into the central regions to trigger AGN activity.

The median parent halo mass of AGN samples (purple triangles in Figure 14) are ~0.1–0.3 dex higher than the corresponding median (sub-)halo masses and remain approximately constant at all redshifts ($\sim 2 \times 10^{12} M_\odot$ for $L_X > 10^{42}$ erg s^{-1} AGN). Thus, half of the AGN in L_X-limited samples are found in haloes with $M_{parent\ halo} \lesssim 2 \times 10^{12} M_\odot$ at $z < 2.5$. We note that the median $M_{parent\ halo}$ is not affected by the build up of the most extreme, massive parent haloes. In contrast, the *mean* parent halo mass of AGN rises by almost an order of magnitude as cosmic time progresses, increasing from $\sim 10^{12.6} M_\odot$

at $z \sim 2.5$ to $\sim 10^{13.5} M_\odot$ at $z \sim 0.3$ for $L_X > 10^{42}$ erg s^{-1} AGN. As time progresses, the most massive haloes start to assemble and accrete smaller sub-haloes, containing satellite galaxies where AGN activity often occurs (see Figure 14, right). It is the build up of these satellite galaxy populations within massive parent haloes that drives the increase in the mean $M_{parent\ halo}$ with cosmic time. We do not infer any direct correlation between the parent halo—and thus the large-scale environment—and AGN activity.

The green crosses in Figure 17 lie significantly above the median $M_{(sub-)halo}$ and $M_{parent\ halo}$ and show a mild decline with increasing redshift, but do not accurately reflect the evolution of the mean $M_{parent\ halo}$. The satellite fraction and broad parent halo mass distribution are not accounted for when using the absolute bias to infer a halo mass and thus they do not provide an accurate tracer of the typical halo mass.

To further illustrate the dangers of using the absolute bias to simplistically infer a halo mass, in Figure 18 we show the measured bias for an observational, luminosity-limited sample of AGN (red points, as predicted by our model) as a function of redshift, compared to the absolute bias as a function of redshift given by Tinker et al. (2010) for haloes of constant M_{halo} and with a narrow distribution (grey
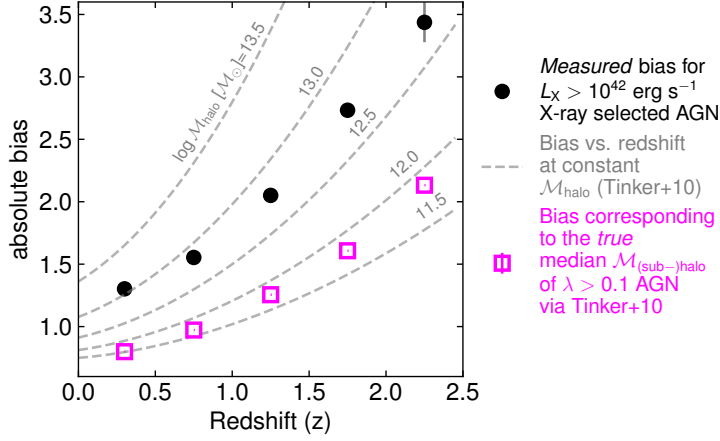
Figure 18. Absolute bias as a function of redshift, showing our predictions of the measured bias for $L_X > 10^{42}$ erg s^{-1} AGN (black circles) compared to the bias of dark matter haloes of constant mass (grey dashed lines: Tinker et al. 2010). Such relations are often used to interpret the measured bias of AGN samples and (erroneously) conclude that AGN all lie in dark matter haloes with log $\mathcal{M}_{halo} \approx 12.5 - 13.0$. The pink open squares show the bias that corresponds to the true median $\mathcal{M}_{(sub-)halo}$ for a λ-limited AGN sample. Using the measured bias of AGN samples to infer the typical halo mass using relations derived from simulations with narrow halo mass bins (e.g. Tinker et al. 2010; Comparat et al. 2017) leads to an overestimate of $\sim 0.5 - 1$ dex. This widely used methodology neglects the distinction between parent haloes and sub-haloes, does not account for known observational selection biases in AGN samples, and does not capture the true breadth of the distribution of AGN host halo masses.

dashed lines). Such plots are often used to summarise observational results and suggest that AGN activity occurs uniformly in haloes of mass $\sim 10^{12.5-13.0} \mathcal{M}_\odot$ throughout cosmic time (e.g. Ross et al. 2009; Chehade et al. 2016; Mendez et al. 2016). The pink squares instead show the absolute bias that corresponds to the median $\mathcal{M}_{(sub-)halo}$ for a complete $\lambda > 0.1$ AGN sample i.e. indicating the absolute bias that *would be measured* for a narrow distribution of haloes at the mass corresponding to the true median $\mathcal{M}_{(sub-)halo}$ (as given by Tinker et al. 2010, see also Figure 15). Using the measured bias to infer a typical halo mass in this simplistic manner gives an inaccurate picture of the haloes that host AGN, neglects the distinction between sub-haloes and parent haloes, does not reflect the underlying broad distribution of host halo masses, and does not account for observational selection effects in AGN samples, all of which are accounted for in our model.

5.3 HOD modelling and the AGN satellite fraction

A more sophisticated interpretation of AGN clustering results can be obtained using halo occupation distribution (HOD) modelling (e.g. Starikova et al. 2011; Shen et al. 2013; Eftekharzadeh et al. 2019). This approach, originally developed for interpreting galaxy clustering, uses analytic functions to describe the average number of galaxies within a given parent halo as a function of mass, $\mathcal{M}_{parent\ halo}$, above a minimum parent halo mass that roughly corresponds to the minimum galaxy stellar mass and thus represents the selection limits of the observed galaxy sample. The HOD model is then used to assign galaxies to haloes in an N-body dark matter simulation. When a single galaxy is assigned to a parent halo, it is designated as the central; additional galaxies assigned to the same parent halo are designated as satellite galaxies and assigned to sub-haloes within the larger parent. By directly linking an observed galaxy sample to a halo model, the clustering properties such as the projected two-point correlation function, $w_p(r_p)$, can be predicted over a wide range of scales. The parameters of the analytic HOD model can be altered until good agreement with the observational results is obtained.

We note that the HOD approach, using analytic functions to directly populate haloes with galaxies, differs from the approach of UNIVERSEMACHINE (and adopted in this paper) which uses abundance matching to assign galaxies (rank-ordered by SFR) to haloes (rank-ordered by halo growth rate in terms of the change of $v_{\mathcal{M}_{peak}}$) and does not assume a functional form for the relation between halo mass and galaxy properties.

The HOD modelling approach can also be applied to AGN samples, with a number of important distinctions: i) the average number of AGN per parent halo is typically much lower than the number of galaxies (i.e. a small fraction of haloes are expected to contain an observable AGN); ii) due to the smaller sample sizes the resulting constraints from clustering measurements on the HOD tend to be weak and thus simpler analytic forms are usually adopted; iii) the link between the minimum halo mass, which must be assumed, and the observational limits of an AGN sample is much less clear; and iv) when a single (or indeed multiple) AGN is assigned to a parent halo it is unclear whether such an AGN should be placed in the central galaxy or satellite galaxies, given that we know the AGN fraction in galaxies is not 100% and AGN activity can be short-lived. Despite these difficulties, a number of studies have derived HODs and placed improved constraints on the halo masses of AGN (e.g. Miyaji et al. 2011; Richardson et al. 2012; Krumpe et al. 2018).

The AGN model developed in this paper is distinct from the HOD modelling approach. We do not link AGN samples to haloes to predict their clustering properties directly. Instead, we use independent measurements of the incidence of AGN as a function of *galaxy* properties to populate UNIVERSEMACHINE model galaxies with AGN. We rely on the UNIVERSEMACHINE to link galaxies and haloes, which is constrained by measurements of *galaxy* clustering (as well as other galaxy observables). From our model, we can *predict* the HOD of AGN. Figure 19 (left) shows our prediction of the mean number of AGN (defined to different limits in L_X and λ) as a function of $\mathcal{M}_{parent\ halo}$. We note the steep rise in the mean number with increasing mass above $\mathcal{M}_{parent\ halo} \sim 10^{13} \mathcal{M}_\odot$, with a slope $\alpha \approx 0.9 - 1.0$ assuming $\langle N(\mathcal{M}_{parent\ halo}) \rangle \propto \mathcal{M}_{parent\ halo}^\alpha$, which is driven by the increasing occupation of parent haloes with satellite galaxies (with $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ and often containing an AGN). At masses $\sim 10^{12-13} \mathcal{M}_\odot$ the slope of the HOD flattens; in this regime there is typically a single AGN in a given halo (with between ~0.1% and 10% of haloes containing an AGN, depending on the chosen limits). The mean
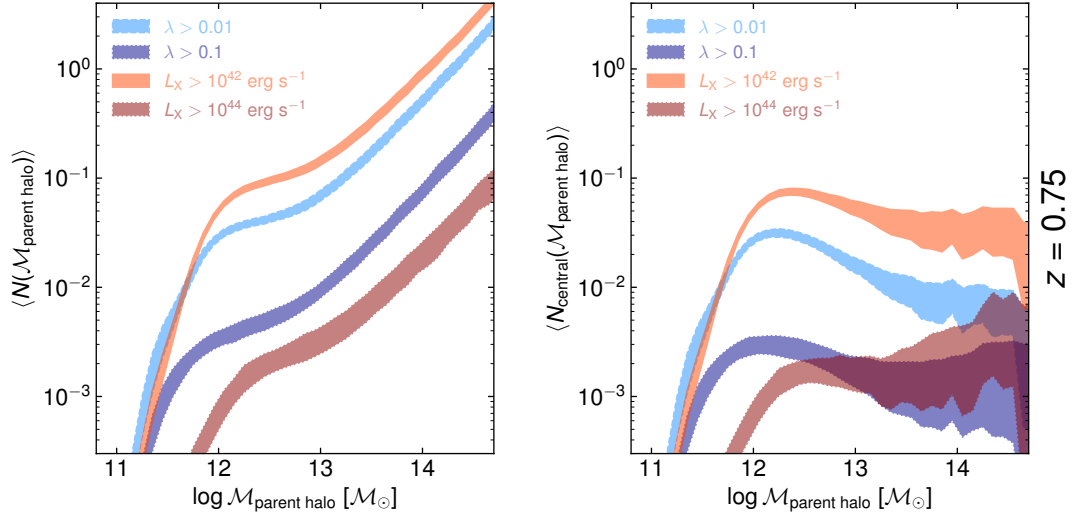
Figure 19. Left: Predicted halo occupation distribution (HOD) functions for AGN samples for different L_X and λ limited samples (as indicated), giving the expected number of AGN as a function of parent halo mass, $\langle N(\mathcal{M}_{\text{parent halo}})\rangle$. **Right:** Predicted AGN HOD functions for central galaxies only, giving the expected number of *central* galaxies that host an AGN as a function of parent halo mass, $\langle N_{\text{central}}(\mathcal{M}_{\text{parent halo}})\rangle$. The expected number peaks at $\mathcal{M}_{\text{parent halo}} \sim 10^{12} \mathcal{M}_\odot$ with a value of ~ 0.1 (for $L_X > 10^{42}$ erg s^{-1} AGN), indicating that only 10% of central galaxies in such haloes are expected to host an AGN, and subsequently *drops* at higher parent halo masses.
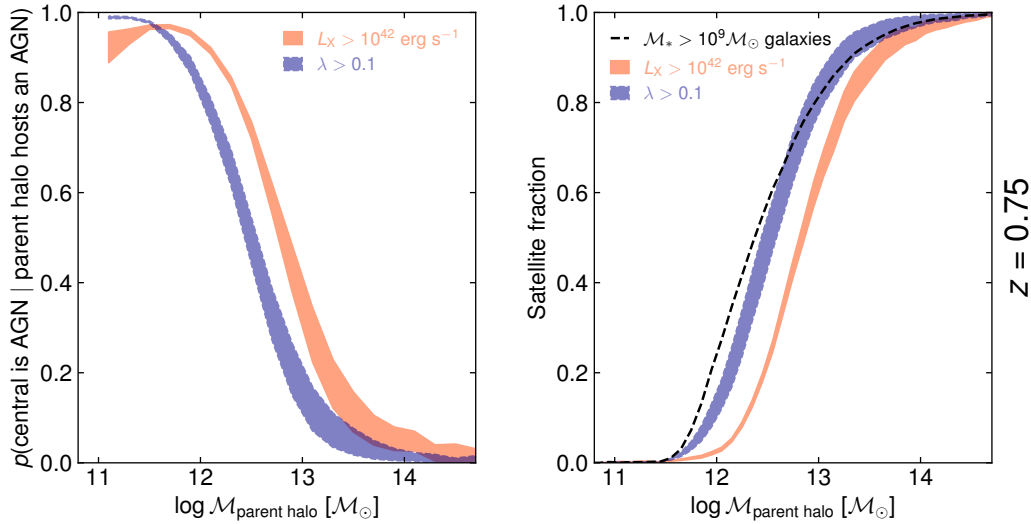
Figure 20. Left: Probability that the central galaxy is an AGN *given* that the parent halo hosts at least one AGN, directly comparing $L_X > 10^{42}$ erg s^{-1} (*red*) and $\lambda > 0.1$ (*blue*) definitions of an AGN. **Right:** The fraction of AGN that are in satellite galaxies as a function of parent halo mass for different L_X and λ limited samples. The black dashed line shows the fraction of $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxies that are satellites for comparison, which also changes with $\mathcal{M}_{\text{parent halo}}$.

occupation number drops rapidly below $\mathcal{M}_{\text{parent halo}} \sim 5 \times 10^{11} \mathcal{M}_\odot$, corresponding to the minimum halo mass found through direct HOD modelling. For the λ-limited samples, we note that this drop is due to the minimum *galaxy* mass assumed in our model and thus reflects the fact that lower mass haloes do not contain $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$ galaxies (e.g. Conroy & Wechsler 2009; Behroozi et al. 2019). The rapid drop thus reflects the observational limits of our input data, not a lack of AGN accreting above these λ limits in such low mass haloes: such AGN may well exist but are difficult to find as they lie in faint, low-$\mathcal{M}_*$ galaxies and produce very low AGN luminosities. For the $L_X > 10^{44}$erg s^{-1}AGN sample the minimum halo mass cut-off is at $\mathcal{M}_{\text{parent halo}} \sim 10^{12} \mathcal{M}_\odot$ and is less distinct, reflecting

the indirect relation between halo mass and an AGN luminosity limit. The complicated shape of the HOD that we recover naturally with our model—reflecting the complex and indirect relation between haloes, galaxies, and the selection biases of different AGN samples—is often not accounted for in AGN HOD models that tend to use simpler analytic parameterisations.

Our model also distinguishes between AGN in central galaxies and satellite galaxies. Figure 19 (right) shows the expected number of AGN in the central galaxy as a function of $\mathcal{M}_{\text{parent halo}}$ (i.e. the HOD of AGN in central galaxies). As a given parent halo will only contain a single central galaxy, the expected number shown in Figure 19 (right) corresponds to the probability of a central galaxy hosting an AGN,

which peaks at $\sim 10\%$ around $10^{12} \mathcal{M}_{\odot}$ (for $L_{\mathrm{X}} > 10^{42}$ erg s^{-1} AGN) and then *decreases* toward higher $\mathcal{M}_{\mathrm{parent\ halo}}$. The central galaxies of higher mass parent haloes tend to be massive quiescent galaxies and thus in our model (based on our input data) the probability of such a galaxy hosting an $L_{\mathrm{X}} > 10^{42}$ erg s^{-1} AGN is $\lesssim 5\%$. This prediction is broadly consistent with direct measurements by Ehlert et al. (2014), who find that $2.9 - 4.6\%$ of Brightest Cluster Galaxies (BCGs), assumed to be the central galaxies of high mass parent haloes, are found to host an X-ray bright AGN. In contrast, Yang et al. (2018) find that $\sim 18\%$ of BCGs host an X-ray luminous AGN ($L_{\mathrm{X}} > 10^{42}$ erg s^{-1}) within a sample of virialised clusters and Noordeh et al. (2020) find 1 out of 7 of the BCGs in their sample has an X-ray luminous AGN, ostensibly higher than our estimate, although these studies are subject to small number statistics and there may be additional selection effects in the construction of high-mass cluster samples that are not incorporated into our model predictions.

The drop in the incidence of AGN in central galaxies is in contrast to the overall HOD (left panel of Figure 19), which continues to increase toward higher $\mathcal{M}_{\mathrm{parent\ halo}}$ and exceeds 1 for $\mathcal{M}_{\mathrm{parent\ halo}} \gtrsim 10^{14} \mathcal{M}_{\odot}$. Such an increase is due to the high incidence of AGN in satellite galaxies, which tend to be moderate-$\mathcal{M}_*$ star-forming galaxies where AGN triggering is common-place. Thus, while a large proportion of massive haloes will contain at least one AGN, the AGN is usually found in a satellite galaxy and not the massive central galaxy.

The left panel of Figure 20 shows the significant drop at high $\mathcal{M}_{\mathrm{parent\ halo}}$ in the probability that the central galaxy is an AGN *given* that a parent halo hosts at least one AGN. Our results indicate that it should not be assumed that the central galaxy hosts an AGN when assigning AGN to high-mass haloes in HOD models.

The right panel of Figure 20 shows our model prediction for the fraction of all galaxies (with $\mathcal{M}_* > 10^9 \mathcal{M}_{\odot}$) or all AGN (to different L_{X} and λ limits) in haloes of a given $\mathcal{M}_{\mathrm{parent\ halo}}$ that are satellites. Despite the apparent similarities in behaviour, this "satellite fraction" is not simply the inverse of p(central is AGN | parent halo is an AGN). The AGN satellite fraction indicates that a high fraction of all AGN (in parent haloes of $\mathcal{M}_{\mathrm{parent\ halo}} \gtrsim 3 \times 10^{12} \mathcal{M}_{\odot}$) are satellites, but the probability that the central galaxy hosts an AGN remains low. This behaviour can be compared to galaxies: they have a high satellite fraction at high halo mass (see right panel of Figure 20), but the probability of having a central galaxy is always 1. It is reasonable to assume that there is always a galaxy at the centre of a massive halo, but there is not always an AGN.

5.4 The clustering of AGN relative to galaxies

In addition to using AGN clustering measurements to determine the absolute bias of AGN samples, which can only be used to infer a typical (sub-)halo or parent halo mass using appropriate modeling as discussed above, it is also useful to determine the bias of AGN samples relative to galaxies. As our understanding of the galaxy-halo connection continues to improve, it is informative to interpret AGN clustering in terms of galaxy clustering.

In our model we find at $z < 1$ that the relative bias compared to $> 10^9 \mathcal{M}_{\odot}$ galaxies is near unity, with only mild dependence on λ and AGN luminosity. There is a small difference in the stellar mass distributions of AGN hosts and the parent galaxy population at $z < 1$, due to AGN residing preferentially in somewhat higher stellar mass galaxies, but generally there are small differences between the clustering of AGN and the parent galaxy population: the relative bias to $> 10^9 \mathcal{M}_{\odot}$ galaxies in our model at $z = 0.75$ is ~ 1.0 for our λ-limited samples. The stellar mass distribution of AGN hosts is further skewed in observational (L_{X}-limited) samples, leading to a slightly

higher relative bias (~ 1.05 at $z = 0.75$ for $L_{\mathrm{X}} > 10^{42}$ erg s^{-1} AGN). We find that at $z > 1$ the relative bias compared to $10^9 \mathcal{M}_{\odot}$ galaxies rises, to ~ 1.2 at $z = 2.3$, due to the increasing preferential occupation of AGN in higher stellar mass galaxies at higher redshift.

These results are qualitatively similar to those of Mendez et al. (2016), who found that while the clustering of X-ray, radio, and mid-infrared AGN samples at $z \sim 0.7$ differ, they all agree with the clustering of galaxies that have the same distributions in both stellar mass and SFR as the AGN host populations. In other words, the relative bias between AGN and matched galaxy samples is unity, if the galaxy population has a matched distribution of stellar mass and SFR to the AGN hosts. Mendez et al. (2016) found that matching on stellar mass alone was insufficient and that the SFR distribution of the galaxy population needed to be matched as well. Krishnan et al. (2020) showed that matching the passive fraction of galaxy samples (i.e. the SFRs) was key to explain the clustering of AGN samples to $z \sim 2$. Powell et al. (2018) also found for a hard X-ray selected AGN sample at $z \sim 0.1$ that AGN clustering matches the clustering of galaxies with the same distribution of stellar mass. These findings are consistent with a relative bias of unity.

These observational papers, as well as the present work, reflect that AGN clustering can be understood simply as galaxy clustering combined with an understanding of which galaxies host AGN. Our model follows this prescription and shows that the predicted clustering matches both X-ray and quasar observations well, such that there is no need for additional components to explain and understand AGN clustering. The agreement between our model and observations (given the observational errors, which can be non-negligible) implies that there is not an additional large-scale environmental effect to AGN triggering.

Our work suggests that using what we know about the galaxy-halo connection, combined with our knowledge of the AGN occupation of galaxies, is a fruitful path forward for constraining halo masses of AGN and understanding the role (if any) of large-scale environment in the triggering of AGN activity (see also Georgakakis et al. 2019; Jones et al. 2019). Measurements of the relative bias of AGN compared to appropriately matched galaxies may provide an additional constraint and test of such models: deviations from our model predictions may indicate that additional large-scale environmental effects play a role in AGN triggering. X-ray selection of AGN is especially powerful for such studies as obscured and low accretion rate objects are identified, allowing host galaxy properties to be measured directly. For quasars, it is difficult to measure the properties of the underlying host galaxy and it is vital to assess the impact of both intrinsic and observational selection effects on the samples. Nevertheless, the cross-correlation of quasar and galaxy samples can help place constraints on where luminous AGN activity occurs within the context of the galaxy population (e.g. Shen et al. 2013; Alam et al. 2020b).

6 SUMMARY AND CONCLUSIONS

In this paper, we have shown how by starting from an empirical galaxy evolution model (UniverseMachine: Behroozi et al. 2019), which is constrained by observations of the galaxy population, and adding AGN based on measurements of their incidence as a function of galaxy properties (Aird et al. 2018), we are able to recover the observed clustering properties of AGN samples out to $z \sim 2.5$ and infer the underlying distribution of AGN host dark matter halo masses. Our approach constitutes a forward model, based on knowledge of the galaxy population, and is not tuned at any stage to reproduce the observed clustering of AGN samples. Our conclusions are as follows:

(i) The triggering of AGN activity—when sufficient quantities of gas are driven into the central regions of galaxies and accreted by the central black hole—is primarily determined by the properties of the galaxies they lie in, specifically their stellar masses and SFRs. An additional, direct dependence on their large-scale environment is *not* required to explain the observed clustering of X-ray AGN and quasar samples.

(ii) Assuming the incidence of AGN depends *only* on host galaxy stellar mass places more AGN in higher mass haloes. Using a distinct AGN fraction in star-forming and quiescent galaxies, as observed, leads to small—but significant—differences in the inferred HMFs.

(iii) AGN have a broad distribution of halo masses, spanning $\gtrsim 3$ orders of magnitude. The peak of the HMF of moderate accretion rate AGN ($\lambda > 0.01$) is ~ 0.3 dex higher than for the underlying galaxy population (with stellar masses $\mathcal{M}_* > 10^9 \mathcal{M}_\odot$) due to the increased incidence of AGN in high-mass star-forming galaxies, an effect which is exaggerated for luminosity-limited samples due to the observational selection bias toward higher stellar mass hosts. The HMFs of higher specific accretion rate sources ($\lambda > 0.1$) tend to follow the shape of HMF of the parent galaxy population more closely, indicating little dependence of vigorous black hole growth on either host galaxy or host halo properties.

(iv) The median (sub-)halo mass of AGN samples is $\sim 10^{12} \mathcal{M}_\odot$ and does not depend strongly on AGN luminosity or redshift, although we do find that higher specific accretion rate AGN have lower median (sub-)halo masses. The median *parent* halo mass is ~ 0.3 dex higher, $\approx 10^{12.3} \mathcal{M}_\odot$, and does not evolve with redshift (from $z \sim 2.5$ to 0.3). At all epochs, half of $L_X > 10^{42}$ erg s^{-1} AGN are in parent haloes with masses $< 2 \times 10^{12} \mathcal{M}_\odot$.

(v) Using a relation that provides the bias for a narrow distribution of halo masses to infer a typical halo mass based on the absolute bias of AGN samples (measured from the observed clustering amplitude) does not provide an accurate indicator of the median (sub-)halo or parent halo mass as such methods neglect the broad distribution of halo masses and do not account for the high fraction of AGN that are found in satellite galaxies.

(vi) The overall fraction of AGN in satellite galaxies is $\sim 25\%$ but depends strongly on parent halo mass. The probability of finding at least one AGN in the most massive parent haloes approaches 100% but the probability that the central galaxy is an AGN remains low ($\lesssim 5\%$).

(vii) The clustering of observed AGN samples is most easily interpreted in terms of the relative bias to galaxy samples, not from absolute bias measurements alone. In particular, matched galaxy samples allow comparisons that reveal whether additional large-scale environmental triggering is at play in AGN physics, beyond what is know about the galaxy-halo connection and how AGN occupy galaxies.

A natural feature of our model is the uniformity of halo masses of AGN over cosmic time and as a function of luminosity or specific accretion rate. This uniformity is due to a combination of the efficiency of galaxy formation as a function of halo mass, the triggering of AGN as a function of stellar mass and SFR, and the flickering of AGN activity over the course of a galaxy's lifecycle, as well as AGN selection biases.

Our work shows how detailed measurements of the AGN fraction within galaxy samples, combined with a sophisticated model of the galaxy–halo connection, provide a key avenue to understanding the clustering properties and host halo masses of AGN. A number of future directions can expand and improve on this approach. As galaxy studies continue to develop our understanding of the connection between large-scale environment and galactic SFRs (e.g. Coil et al. 2017; Berti et al. 2020), more refined measurements of the AGN fraction as a function of SFR (e.g. Aird et al. 2019) may be used to enhance our model. The connection between AGN obscuration properties and both small-scale (i.e. host galaxy) and large-scale (i.e. host halo) environment also remains unclear, especially for the most heavily obscured sources that are missed in the X-ray based measurements of AGN fraction that underpin this work. Furthermore, understanding the complex selection biases of optically identified spectroscopic samples of quasars, both in terms of the underlying host galaxies and how they relate to more complete X-ray selected AGN samples, is vital for the interpretation of future measurements of quasar clustering. Upcoming large-scale X-ray and optical spectroscopic surveys will provide the samples of both AGN and galaxies needed to fully elucidate the AGN–galaxy–halo connection.

DATA AVAILABILITY

The UNIVERSEMACHINE data underlying this article are publicly available at http://behroozi.users.hpc.arizona.edu/UniverseMachine/DR1/. AGN fractions were derived from measurements available at https://doi.org/10.5281/zenodo.1009604 (Aird et al. 2017a). The derived data generated for this paper will be shared on reasonable request to the corresponding author.

REFERENCES

Aird J., et al., 2012, ApJ, 746, 90
Aird J., et al., 2013, ApJ, 775, 41
Aird J., Coil A. L., Georgakakis A., Nandra K., Barro G., Pérez-González P. G., 2015, MNRAS, 451, 1892
Aird J., Coil A. L., Georgakakis A., 2017a, [Dataset] X-rays across the galaxy population: The distribution of AGN accretion rates as a function of stellar mass and redshift, doi:10.5281/zenodo.1009605, https://doi.org/10.5281/zenodo.1009605
Aird J., Coil A. L., Georgakakis A., 2017b, MNRAS, 465, 3390
Aird J., Coil A. L., Georgakakis A., 2018, MNRAS, 474, 1225
Aird J., Coil A. L., Georgakakis A., 2019, MNRAS, 484, 4360
Alam S., Ross N. P., Eftekharzadeh S., Peacock J. A., Comparat J., Myers A. D., Ross A. J., 2020a, arXiv e-prints, p. arXiv:2007.02612
Alam S., Peacock J. A., Kraljic K., Ross A. J., Comparat J., 2020b, MNRAS, 497, 581
Alexander D. M., et al., 2003, AJ, 126, 539
Allevato V., et al., 2011, ApJ, 736, 99
Ananna T. T., et al., 2019, ApJ, 871, 240
Beckmann R. S., et al., 2017, MNRAS, 472, 949
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013a, ApJ, 762, 109
Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013b, ApJ, 763, 18
Behroozi P. S., Wechsler R. H., Conroy C., 2013c, ApJ, 770, 57

Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143

Berti A. M., Coil A. L., Hearin A. P., Behroozi P. S., 2020, arXiv e-prints, p. arXiv:2009.02405

Birchall K. L., Watson M. G., Aird J., 2020, MNRAS, 492, 2268

Bongiorno A., et al., 2016, A&A, 588, A78

Booth C. M., Schaye J., 2009, MNRAS, 398, 53

Bower R. G., Benson A. J., Crain R. A., 2012, MNRAS, 422, 2816

Brusa M., et al., 2009, A&A, 507, 1277

Buchner J., et al., 2015, ApJ, 802, 89

Cappelluti N., Ajello M., Burlon D., Krumpe M., Miyaji T., Bonoli S., Greiner J., 2010, ApJ, 716, L209

Cappelluti N., Allevato V., Finoguenov A., 2012, Advances in Astronomy, 2012, 853701

Chehade B., et al., 2016, MNRAS, 459, 1179

Cisternas M., et al., 2011, ApJ, 741, L11

Civano F., et al., 2016, ApJ, 819, 62

Coil A. L., et al., 2009, ApJ, 701, 1484

Coil A. L., Mendez A. J., Eisenstein D. J., Moustakas J., 2017, ApJ, 838, 87

Comparat J., Prada F., Yepes G., Klypin A., 2017, MNRAS, 469, 4157

Comparat J., et al., 2019, MNRAS, 487, 2005

Conroy C., Wechsler R. H., 2009, ApJ, 696, 620

Croom S. M., et al., 2005, MNRAS, 356, 415

Croton D. J., et al., 2006, MNRAS, 365, 11

DeGraf C., Sijacki D., 2017, MNRAS, 466, 3331

Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604

Dubois Y., Gavazzi R., Peirani S., Silk J., 2013, MNRAS, 433, 3297

Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, MNRAS, 463, 3948

Eftekharzadeh S., et al., 2015, MNRAS, 453, 2779

Eftekharzadeh S., Myers A. D., Kourkchi E., 2019, MNRAS, 486, 274

Ehlert S., et al., 2014, MNRAS, 437, 1942

Font-Ribera A., et al., 2013, J. Cosmology Astropart. Phys., 2013, 018

Gatti M., Shankar F., Bouillot V., Menci N., Lamastra A., Hirschmann M., Fiore F., 2016, MNRAS, 456, 1073

Georgakakis A., et al., 2014, MNRAS, 443, 3327

Georgakakis A., Aird J., Schulze A., Dwelly T., Salvato M., Nandra K., Merloni A., Schneider D. P., 2017, MNRAS, 471, 1976

Georgakakis A., Comparat J., Merloni A., Ciesla L., Aird J., Finoguenov A., 2019, MNRAS, 487, 275

Grogin N. A., et al., 2011, ApJS, 197, 35

Hearin A. P., et al., 2017, AJ, 154, 190

Hickox R. C., et al., 2009, ApJ, 696, 891

Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Springel V., 2006, ApJS, 163, 50

Jones M. L., Hickox R. C., Mutch S. J., Croton D. J., Ptak A. F., DiPompeo M. A., 2019, ApJ, 881, 110

Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340

Kocevski D. D., et al., 2015, ApJ, 814, 104

Koss M., Veilleux S., Baumgartner W., Tueller J., Gehrels N., Valencic a., 2011, The Astrophysical Journal, 739, 57

Krishnan C., et al., 2020, MNRAS, 494, 1693

Krumpe M., Miyaji T., Coil A. L., 2010, ApJ, 713, 558

Krumpe M., Miyaji T., Coil A. L., Aceves H., 2012, ApJ, 746, 1

Krumpe M., Miyaji T., Coil A. L., 2014, in Multifrequency Behaviour of High Energy Cosmic Sources. p. 71 (arXiv:1308.5976), doi:10.14311/APP.2014.01.0071

Krumpe M., Miyaji T., Coil A. L., Aceves H., 2018, MNRAS, 474, 1773

Laurent P., et al., 2017, J. Cosmology Astropart. Phys., 2017, 017

Leauthaud A., et al., 2015, MNRAS, 446, 1874

Marleau F. R., Clancy D., Bianconi M., 2013, MNRAS, 435, 3085

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

Mead A. J., Heymans C., Lombriser L., Peacock J. A., Steele O. I., Winther H. A., 2016, MNRAS, 459, 1468

Mendez A. J., et al., 2013, ApJ, 770, 40

Mendez A. J., et al., 2016, ApJ, 821, 55

Mezcua M., Civano F., Marchesi S., Suh H., Fabbiano G., Volonteri M., 2018, MNRAS, 478, 2576

Miyaji T., Krumpe M., Coil A. L., Aceves H., 2011, ApJ, 726, 83

Mountrichas G., Georgakakis A., 2012, MNRAS, 420, 514

Mountrichas G., et al., 2013, MNRAS, 430, 661

Mountrichas G., Georgakakis A., Menzel M. L., Fanidakis N., Merloni A., Liu Z., Salvato M., Nandra K., 2016, MNRAS, 457, 4195

Muzzin A., et al., 2013, ApJ, 777, 18

Myers A. D., Brunner R. J., Nichol R. C., Richards G. T., Schneider D. P., Bahcall N. A., 2007, ApJ, 658, 85

Nandra K., et al., 2015, ApJS, 220, 10

Noordeh E., et al., 2020, MNRAS, 498, 4095

Planck Collaboration et al., 2016, A&A, 594, A20

Porciani C., Norberg P., 2006, MNRAS, 371, 1824

Porciani C., Magliocchetti M., Norberg P., 2004, MNRAS, 355, 1010

Powell M. C., et al., 2018, ApJ, 858, 110

Powell M. C., Urry C. M., Cappelluti N., Johnson J. T., LaMassa S. M., Ananna T. T., Kollmann K. E., 2020, ApJ, 891, 41

Puchwein E., Springel V., 2013, MNRAS, 428, 2966

Reddick R. M., Wechsler R. H., Tinker J. L., Behroozi P. S., 2013, ApJ, 771, 30

Reines A. E., Volonteri M., 2015, ApJ, 813, 82

Ricci C., Ueda Y., Koss M. J., Trakhtenbrot B., Bauer F. E., Gandhi P., 2015, ApJ, 815, L13

Ricci C., et al., 2017, Nature, 549, 488

Richardson J., Zheng Z., Chatterjee S., Nagai D., Shen Y., 2012, ApJ, 755, 30

Richardson M. L. A., Scannapieco E., Devriendt J., Slyz A., Thacker R. J., Dubois Y., Wurster J., Silk J., 2016, ApJ, 825, 83

Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, MNRAS, 462, 893

Ross N. P., et al., 2009, ApJ, 697, 1634

Shen Y., et al., 2013, ApJ, 778, 98

Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1

Simmons B. D., Smethurst R. J., Lintott C., 2017, MNRAS, 470, 1559

Sinha M., Garrison L. H., 2020, MNRAS, 491, 3022

Skelton R. E., et al., 2014, ApJS, 214, 24

Starikova S., et al., 2011, ApJ, 741, 15

Timlin J. D., et al., 2018, ApJ, 859, 20

Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, ApJ, 724, 878

Ueda Y., Akiyama M., Hasinger G., Miyaji T., Watson M. G., 2014, ApJ, 786, 104

Viitanen A., Allevato V., Finoguenov A., Bongiorno A., Cappelluti N., Gilli R., Miyaji T., Salvato M., 2019, A&A, 629, A14

Xue Y. Q., et al., 2010, ApJ, 720, 368

Xue Y. Q., et al., 2011, ApJS, 195, 10

Yang L., Tozzi P., Yu H., Lusso E., Gaspari M., Gilli R., Nardini E., Risaliti G., 2018, ApJ, 859, 65

Zubovas K., King A. R., 2012, MNRAS, 426, 2751

da Ângela J., et al., 2008, MNRAS, 383, 565

This paper has been typeset from a TeX/LaTeX file prepared by the author.