

The XMM-Newton wide field survey in the COSMOS field: Clustering dependence of X-ray selected AGN on host galaxy properties

A. Viitanen¹, V. Allevato^{2,1}, A. Finoguenov¹, A. Bongiorno³, N. Cappelluti⁴, R. Gilli⁵, T. Miyaji⁶, and M. Salvato⁷

¹ Department of Physics, University of Helsinki, Gustaf Hällströmin katu 2a, 00014 Helsinki, Finland
e-mail: akke.viitanen@helsinki.fi

² Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy

³ INAF – Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monte Porzio Catone, Roma, Italy

⁴ Physics Department, University of Miami, Knight Physics Building, Coral Gables, FL 33124, USA

⁵ INAF – Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Gobetti 93/3, 40129 Bologna, Italy

⁶ Instituto de Astronomía, Universidad Nacional Autónoma de México, 22860 Ensenada, Mexico

⁷ Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse 1, 85748 Garching, Germany

Received 31 January 2019 / Accepted 18 June 2019

ABSTRACT

Aims. We study the spatial clustering of 632 (1130) XMM-COSMOS active galactic nuclei (AGNs) with known spectroscopic or photometric redshifts in the range $z = [0.1-2.5]$ in order to measure the AGN bias and estimate the typical mass of the hosting dark matter (DM) halo as a function of AGN host galaxy properties.

Methods. We created AGN subsamples in terms of stellar mass, M_* , and specific black hole accretion rate, L_X/M_* , to study how AGN environment depends on these quantities. Further, we derived the M_*-M_{halo} relation for our sample of XMM-COSMOS AGNs and compared it to results in literature for normal non-active galaxies. We measured the projected two-point correlation function $w_p(r_p)$ using both the classic and the generalized clustering estimator, based on photometric redshifts, as probability distribution functions in addition to any available spectroscopic redshifts. We measured the large-scale ($r_p \gtrsim 1 h^{-1}$ Mpc) linear bias b by comparing the clustering signal to that expected of the underlying DM distribution. The bias was then related to the typical mass of the hosting halo M_{halo} of our AGN subsamples. Since M_* and L_X/M_* are correlated, we matched the distribution in terms of one quantity and we split the distribution in the other.

Results. For the full spectroscopic AGN sample, we measured a typical DM halo mass of $\log(M_{\text{halo}}/h^{-1} M_\odot) = 12.79^{+0.26}_{-0.43}$, similar to galaxy group environments and in line with previous studies for moderate-luminosity X-ray selected AGN. We find no significant dependence on specific accretion rate L_X/M_* , with $\log(M_{\text{halo}}/h^{-1} M_\odot) = 13.06^{+0.23}_{-0.38}$ and $\log(M_{\text{halo}}/h^{-1} M_\odot) = 12.97^{+0.39}_{-1.26}$ for low and high L_X/M_* subsamples, respectively. We also find no difference in the hosting halos in terms of M_* with $\log(M_{\text{halo}}/h^{-1} M_\odot) = 12.93^{+0.31}_{-0.62}$ (low) and $\log(M_{\text{halo}}/h^{-1} M_\odot) = 12.90^{+0.30}_{-0.62}$ (high). By comparing the M_*-M_{halo} relation derived for XMM-COSMOS AGN subsamples with what is expected for normal non-active galaxies by abundance matching and clustering results, we find that the typical DM halo mass of our high M_* AGN subsample is similar to that of non-active galaxies. However, AGNs in our low M_* subsample are found in more massive halos than non-active galaxies. By excluding AGNs in galaxy groups from the clustering analysis, we find evidence that the result for low M_* may be due to larger fraction of AGNs as satellites in massive halos.

Key words. dark matter – galaxies: active – galaxies: evolution – large-scale structure of Universe – quasars: general – surveys

1. Introduction

Supermassive black holes (SMBHs) with $M \sim 10^{6-9} M_\odot$ reside at the centers of nearly every massive galaxy. Also, SMBHs reach these masses by growing via matter accretion and by simultaneously shining luminously as an active galactic nucleus (AGN). Interestingly, BHs and their host galaxies seem to co-evolve, as suggested by the correlation between the SMBH and the host galaxy properties (velocity dispersion, luminosity, and stellar mass). However, the co-evolution scenario, AGN feedback, and accretion mechanisms are still poorly known (e.g., Alexander & Hickox 2012).

Furthermore, AGNs and their host galaxies reside in collapsed dark matter (DM) structures, such as halos. In the concordance Λ cold dark matter (Λ CDM) cosmology, these halos form hierarchically “bottom up” from the smallest structures, density

fluctuations in the cosmic microwave background, that grow via gravitational instability to the largest galaxy groups and clusters. It is interesting to note that AGNs and DM halos in which they reside are both biased tracers of the underlying DM distribution. By measuring the clustering of AGN, and comparing that to the underlying DM distribution, the AGNs may be linked to their hosting DM halos (e.g., Cappelluti et al. 2012; Krumpe et al. 2014). Recent AGN clustering measurements have not been able to paint a coherent picture of the complex interplay between AGN and their environment. It seems that optically selected luminous quasars prefer to live in halos few $\times 10^{12} h^{-1} M_\odot$ over a wide range in redshift (Croom et al. 2005; da Ângela et al. 2008; Ross et al. 2009) while moderate luminosity X-ray selected AGN prefer larger halos $10^{12.5-13} h^{-1} M_\odot$ at similar redshifts (Coil et al. 2009; Allevato et al. 2011), (Koutoulidis et al. 2013).

Mendez et al. (2016) suggest that the clustering of AGN could be understood as the clustering of galaxies with matched properties in terms of stellar mass, star-formation rate (SFR), redshift, and AGN selection effects. This would indicate that instead of the properties of the AGN itself, the properties of the host galaxy, such as stellar mass M_* or specific black hole accretion rate L_X/M_* , have a more significant role in driving the clustering of AGN.

Many authors have investigated the relation between the stellar mass and the DM halo mass, the so-called M_*-M_{halo} relation for normal non-active galaxies via abundance matching (Moster et al. 2013; Behroozi et al. 2013), in addition to clustering measurements and HOD modeling (Zheng et al. 2007; Wake et al. 2011), or weak lensing (Coupon et al. 2015). For X-ray selected AGNs, the M_*-M_{halo} relation has only recently been studied observationally. Georgakakis et al. (2014) argue that AGN environment is closely related to M_* . However, they do not measure M_* directly, but use the rest frame absolute magnitude in the J band as a proxy for M_* . Very recently, Mountrichas et al. (2019) measured the AGN clustering dependence directly, in terms of M_* , and found that the environments of X-ray AGN at $z = 0.6-1.4$ are similar to normal galaxies with matched SFR and redshift.

In this study, we wish to build upon the previous X-ray selected AGN clustering measurements in XMM-COSMOS (Miyaji et al. 2007; Gilli et al. 2009; Allevato et al. 2011) to investigate the clustering dependence on host galaxy properties (M_* , L_X/M_*). We compare this to the M_*-M_{halo} relation for normal non-active galaxies. In our clustering measurements, we also investigate the new generalized estimator which has been introduced (Georgakakis et al. 2014; Allevato et al. 2016), where photometric redshifts are included in the clustering analysis as probability density functions. Clustering measurements using photometric redshifts will be important in future X-ray AGN surveys, where spectroscopic redshifts are not available either because AGN are optically faint or no extensive spectroscopic follow-up campaigns are available. In eROSITA, for example, spectroscopic redshifts will only be available for a certain portion of the sky and at later stages of the survey (Merloni et al. 2019).

We adopt a flat Λ CDM cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $\sigma_8 = 0.8$, and $h = 0.7$. Distances reported are comoving distances, and the dependence in h is shown explicitly. The symbol “log” signifies base 10 logarithm. Furthermore, DM halo masses are defined as the enclosed mass within the Virial radius, within which the mean density is 200 times more than the background density. Additionally, DM halo masses scale as h^{-1} , while M_* scales as h^{-2} .

2. Data

2.1. XMM-COSMOS multiwavelength data set

To study the dependence of AGN clustering in terms of host galaxy properties, we use the Cosmic Evolution Survey (COSMOS, Scoville et al. 2007), which is a multiwavelength survey over $1.4 \times 1.4 \text{ deg}^2$ field. It is designed to study the evolution of galaxies and AGNs up to redshift $z \sim 6$. To date, the field has been covered by a wide variety of instruments from radio to X-ray bands. XMM-Newton surveyed 2.13 deg^2 of the sky in the COSMOS field in the 0.5–10 keV band for a total of 1.55 Ms (Hasinger et al. 2007; Cappelluti et al. 2007, 2009), providing an unprecedented large sample of point-like X-ray sources (1822).

Brusa et al. (2010) carried out the optical identification and presented the multiwavelength properties ($24 \mu\text{m}$ to UV)

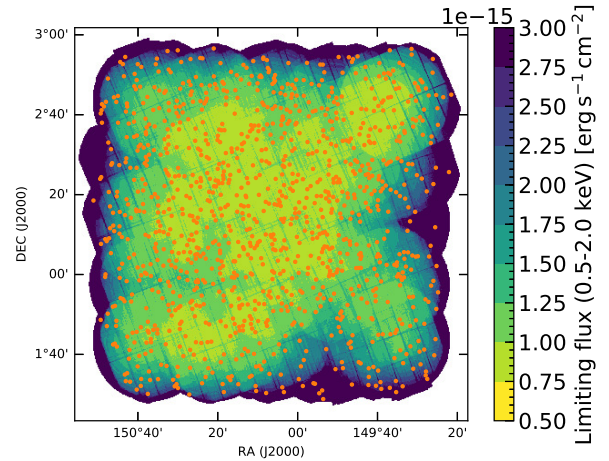


Fig. 1. Distribution of XMM-COSMOS AGN in the COSMOS field used in this study. Orange points mark the positions of 1130 AGNs with $z = [0.1-2.5]$, while the background colors show the sensitivity map in the 0.5–2.0 keV band in $\text{erg cm}^{-2} \text{ s}^{-1}$ (Cappelluti et al. 2009).

of ~ 1800 sources with a spectroscopic completeness of $\sim 50\%$ (e.g., Hasinger et al. 2018). Salvato et al. (2009, 2011) derived accurate photometric redshifts with $\sigma_{\Delta z/(1+z_{\text{spec}})} \sim 0.015$. Bongiorno et al. (2012) used a spectral energy distribution (SED) fitting technique based on AGN and Galaxy template SEDs to estimate the host galaxy properties, that is stellar mass M_* and SFR of ~ 1700 AGN in COSMOS up to $z \lesssim 3$. The quantity L_X/M_* corresponds to the rate of accretion onto the central SMBH scaled relative to the stellar mass of the host galaxy. Assuming a M_*-M_{BH} relation and a constant bolometric correction to convert from L_X to L_{bol} , then the Eddington ratio ($\lambda_{\text{Edd}} \equiv L_{\text{bol}}/L_{\text{Edd}}$) can be expressed as:

$$\lambda_{\text{Edd}} = \frac{A \times k_{\text{bol}}}{1.3 \times 10^{38}} \times \frac{L_X}{M_*}. \quad (1)$$

It is interesting to note that when $A = 500$ and $k_{\text{bol}} = 25$, $L_X/M_* = 10^{34} \text{ erg s}^{-1} M_\odot^{-1}$, this corresponds to accretion at Eddington luminosity, that is $\lambda_{\text{Edd}} = 1$ (Bongiorno et al. 2012).

In this paper we use the catalog presented in Bongiorno et al. (2012), and we focus on 1130 AGN in the redshift range $0.1 < z < 2.5$, with mean $z \sim 1.2$. The redshifts are either spectroscopic (632) or high quality photometric (498) ones. The 2–10 keV luminosity L_X spans $\log(L_X/\text{erg s}^{-1}) = 42.3-45.5$ with a mean $\log(L_X/\text{erg s}^{-1}) = 43.7$. The typical host galaxy of our AGN is a red and massive galaxy with mean $\log(M_*/M_\odot) = 10.7$. However, the host galaxies also span a wide range of stellar masses with $\log(M_*/M_\odot) = 7.6-12.3$. The L_X and M_* distributions for our sample of 1130 XMM-COSMOS AGN are shown in Fig. 2. It would also be of interest to study the clustering as a function of host galaxy SFR or specific SFR, SFR/M_* , as recently done by Mountrichas et al. (2019). However, Bongiorno et al. (2012) conclude for XMM-COSMOS that while stellar masses from SED fitting are relatively robust for both type 1 and type 2 AGNs, SFRs are more sensitive to AGN contamination from type 1 AGN and are unreliable. Thus in order to increase statistics in our clustering analysis, we will not consider the host galaxy SFR, available only for type 2 AGN in XMM-COSMOS.

The recent Chandra COSMOS-Legacy survey (CCLS; Civano et al. 2016; Marchesi et al. 2016) contains the largest sample of X-ray selected AGNs to date. However, for CCLS AGN, host galaxy properties have only been estimated for type 2 AGNs, while Bongiorno et al. (2012) provide the estimates

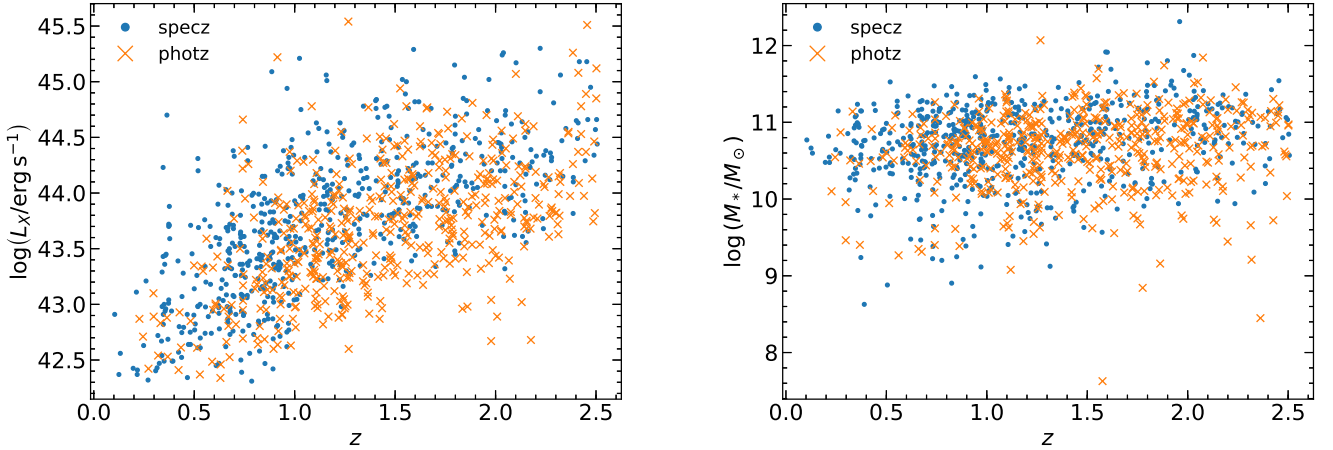


Fig. 2. Distribution of 2–10 keV luminosity (*left*) and host galaxy stellar mass (*right*) as a function of redshift for our sample of 1130 AGNs. Blue (orange) points show 632 (498) AGN with known spectroscopic (photometric) redshifts.

for both type 1 and 2 AGNs. Furthermore, the clustering of XMM-COSMOS AGNs is well studied (Miyaji et al. 2007; Gilli et al. 2009; Allevato et al. 2011, 2012, 2014), but not in terms of host galaxy properties as in this work. For CCLS AGN, Allevato et al. (2016) measured the clustering at $2.9 \leq z \leq 5.5$, and Koutoulidis et al. (2018) used multiple fields including COSMOS to measure the clustering. Thus, there are no clustering measurements for CCLS AGN at the redshift of interest ($z < 2.5$).

2.2. AGN subsamples

The full AGN sample with known spectroscopic redshifts consists of $N = 632$ AGNs with mean $z = 1.19$. For AGNs with only known photometric redshifts, we take into account the full probability distribution function Pdf(z). In this picture, the total weight of an AGN is the integral over z . We limit ourselves to $z < 2.5$ and the combined weighted number of AGNs with photometric redshifts is $N = 488.64$ with weighted mean $z = 1.44$.

To study the dependence on host galaxy properties, we divided our AGN sample effectively in two bins of M_* and L_X/M_* , which we refer to as the low and high subsamples. To give a detailed account of the process, we first binned the distribution of host galaxy stellar mass $\log M_*$ of the sample with bin size 0.1 dex. Then, each bin was split individually and exactly in half based on the logarithm of the specific BH accretion rate $\log L_X/M_*$ to create the low and high L_X/M_* subsamples. The low and high L_X/M_* subsamples consist of 309 objects each. We find the average values for the low (high) L_X/M_* subsamples to be mean $\log L_X/M_* = 32.53$ (33.49), while the difference in mean $\log M_*$ is $\lesssim 0.01$. We then repeated this process by binning the $\log L_X/M_*$ and splitting in terms of $\log M_*$. The number of objects in the low and high M_* subsamples is 309. The average values for the low (high) M_* subsamples are mean $\log M_* = 10.39$ (11.05) and the difference in mean $\log(L_X/M_*)$ is $\lesssim 0.01$.

COSMOS is known to be affected by cosmic variance that influences the clustering measurements (e.g., Gilli et al. 2009; Mendez et al. 2016). This means that it is also important to take into account how our low and high M_* AGN subsamples relate to the large structures in the field. To this end, as an additional test, we associated the AGN sample with known spectroscopic redshifts with the co-added COSMOS galaxy group catalog (see Gozaliasl et al. 2019). An AGN is considered to belong to a

galaxy group if the AGN-group angular separation on the sky is $< R_{200, \text{deg}}$ (the radius of the group in degrees encloses 200 times the critical density) and the radial comoving distance separation is $< \pi_{\text{max}}$ (see Sect. 3). We find 22 (17) AGNs in our low (high) M_* AGN subsamples with spectroscopic redshifts in galaxy groups with a total number of 39 AGNs.

We summarize the properties of the different AGN subsamples in Table 1, and the L_X and M_* distributions are shown in Figs. 2 and 3.

3. Methods

3.1. Two-point statistics

In clustering studies, a widely used measure to quantify clustering is the two-point correlation function $\xi(r)$, which is defined as the excess probability above random of finding a pair of AGNs in a volume element dV at physical separation r , so that

$$dP = n [1 + \xi(r)] dV, \quad (2)$$

where n is the mean number density of AGNs. To estimate $\xi(r)$, we use the Landy & Szalay (1993) estimator

$$\xi(r) = \frac{DD' - 2DR' + RR'}{RR'}, \quad (3)$$

where

$$DD' = \frac{DD}{N_d(N_d - 1)/2} \quad (4)$$

$$DR' = \frac{DR}{N_d N_r} \quad (5)$$

$$RR' = \frac{RR}{N_r(N_r - 1)/2}, \quad (6)$$

and DD, DR, and RR are the number of data-data, data-random, and random-random pairs with physical separation r , respectively. Additionally, N_d and N_r are the total number of sources in the data and random catalogs. This estimator requires the creation of a random catalog to act as an unclustered distribution of AGNs with the same selection effects in terms of right ascension, declination, and redshift, as present in the data catalog (see Sect. 3.4).

As the distances between AGN are inferred from their redshifts, the estimates are affected by distortions due to peculiar

Table 1. XMM-COSMOS AGN subsamples.

	$\sum_i \text{pdf}_i$	$\langle z \rangle$	$\langle \log(M_*/M_\odot) \rangle$	$\langle \log \frac{L_X/M_*}{\text{erg s}^{-1} M_\odot} \rangle$	Type1/type2	χ^2_{min}	d.o.f.	b	$\log \frac{M_{\text{halo}}}{h^{-1} M_\odot}$
<i>Specz</i>									
All	632	1.19	10.72	33.02	351/281	15.14	7	$2.20^{+0.37}_{-0.45}$	$12.79^{+0.26}_{-0.43}$
Low L_X/M_*	309	0.88	10.73	32.53	88/221	6.91	6	$2.14^{+0.35}_{-0.41}$	$13.06^{+0.23}_{-0.38}$
High L_X/M_*	309	1.50	10.73	33.49	253/ 56	1.88	4	$2.95^{+0.93}_{-1.42}$	$12.97^{+0.39}_{-1.26}$
Low M_*	309	0.97	10.39	33.03	134/175	7.96	6	$2.11^{+0.45}_{-0.58}$	$12.93^{+0.31}_{-0.62}$
High M_*	309	1.41	11.05	33.02	211/ 98	10.25	5	$2.69^{+0.61}_{-0.79}$	$12.90^{+0.30}_{-0.62}$
<i>Specz + Photz Pdfs</i>									
All	664	1.20	10.72	33.03	372/292	16.28	7	$2.21^{+0.34}_{-0.40}$	$12.77^{+0.23}_{-0.37}$
Low L_X/M_*	325	0.88	10.73	32.53	95/230	8.45	5	$2.12^{+0.37}_{-0.45}$	$13.03^{+0.25}_{-0.43}$
High L_X/M_*	325	1.52	10.73	33.51	268/57	1.22	4	$2.91^{+0.75}_{-1.03}$	$12.93^{+0.33}_{-0.77}$
Low M_*	323	0.98	10.39	33.03	139/184	9.76	6	$2.25^{+0.40}_{-0.48}$	$13.03^{+0.25}_{-0.43}$
High M_*	323	1.42	11.05	33.03	224/99	13.36	5	$2.18^{+0.57}_{-0.78}$	$12.53^{+0.39}_{-0.98}$
<i>Specz no groups</i>									
Low M_*	287	0.99	10.37	33.06	130/157	6.20	5	$1.69^{+0.49}_{-0.72}$	$12.50^{+0.47}_{-1.67}$
High M_*	292	1.45	11.05	33.05	207/85	9.12	4	$2.48^{+0.55}_{-0.71}$	$12.73^{+0.32}_{-0.64}$

motions of AGNs. To avoid this effect, we express pair separations in terms of distance parallel (π) and perpendicular (r_p) to the line-of-sight of the observer, which is defined with respect to the mean distance to the pair. Then, the projected two-point correlation function (projected 2PCF), which is insensitive to redshift space distortions, is defined as (Davis & Peebles 1983)

$$w_p(r_p) = 2 \int_0^\infty \xi(r_p, \pi) d\pi. \quad (7)$$

In practice, the integration is not carried out to infinity, but to the finite value π_{max} . The estimation of the π_{max} is a balance between including all of the correlated pairs and not including noise to the signal by uncorrelated pairs. For the estimation of the 2PCFs, we used CosmoBolognaLib¹ (Marulli et al. 2016), which is a free (as in freedom) software library for numerical cosmological calculations.

We note that another common way to measure the clustering is to use the cross-correlation function where positions of both an AGN sample and a complete galaxy sample are used to decrease statistical uncertainties (e.g., Coil et al. 2009; Krumpel et al. 2015; Powell et al. 2018; Mountrichas et al. 2019). At our redshift of interest in COSMOS, especially at $1 \lesssim z \lesssim 2.5$, it is difficult to build a complete galaxy sample with known spectroscopic redshifts (see Sect. 3.2 for discussion on the effect of photometric redshift in clustering measurements) with which to measure the clustering. Thus we are limited to the AGN auto-correlation function.

3.2. Generalized estimator

Motivated by recent progress in utilizing photometric redshifts in AGN clustering studies (Georgakakis et al. 2014; Allevato et al. 2016), we used the full probability distribution function Pdf(z) for AGNs with no known spectroscopic redshifts. In this approach, the classic Landy & Szalay (1993) estimator is replaced by a generalized one, where pairs are weighted based on Pdf(z) of the two objects. For more detailed information, we refer the reader to Georgakakis et al. (2014, Sect. 3).

¹ <https://github.com/federicomarulli/CosmoBolognaLib>

For the 498 AGNs with photometric redshifts, we discretized the Pdf(z) by integrating the Pdfs in terms of z with an accuracy of $\delta z = 0.01$, we then normalized the Pdfs to unity. Furthermore, we only considered the part of the Pdf with Pdf(z) > 10^{-5} . Using our redshift limit, we only used the part of the Pdfs with $z < 2.5$. This means that the AGNs with Pdfs spanning over this redshift limit are cut. Also, for these AGNs, the Pdf does not necessarily sum to unity, that is $\sum_i \text{Pdf}(z_i) \leq 1$.

Large uncertainties in photometric redshifts may lead to loss of not only accuracy, but not being able to recover the full clustering signal. This is highlighted by the use of large values of $\pi_{\text{max}} \gtrsim 200 h^{-1} \text{Mpc}$ (Georgakakis et al. 2014; Allevato et al. 2016) versus studies with only spectroscopic redshifts with $\pi_{\text{max}} \lesssim 100 h^{-1} \text{Mpc}$ (e.g., Coil et al. 2009; Allevato et al. 2011; Mountrichas et al. 2016). Therefore, we selected only Pdfs based on the following quality criteria: the comoving distance separation between the z_{min} and z_{max} may not exceed a critical value of $\Delta d = 100 h^{-1} \text{Mpc}$. We defined z_{min} and z_{max} separately for each AGN so that Pdf(z) < 10^{-5} for $z < z_{\text{min}}$ and $z > z_{\text{max}}$.

Furthermore, from the total 498 AGN with photometric redshifts, only 32 AGN passed the quality criterion and are included in the subsample including spectroscopic and photometric redshifts. In terms of our L_X/M_* (M_*) AGN subsamples, a total of 32 (28) AGN with photometric redshifts were kept and divided equally between the low and high subsamples in both cases. The number of AGN in each of our subsamples including photometric redshifts are shown in Table 1.

This quality cut is suggested by the fact that including all phot- z Pdfs leads to large uncertainties in the measured clustering signal for all the AGN subsamples. The investigation of quality criteria for studies including phot- z Pdfs is beyond the scope of this work. However, given the importance of photometric redshifts in future large surveys, such as eROSITA, we will explore clustering phot- z Pdfs in a future study (Viitanen et al., in prep.).

3.3. Halo model

In the halo model (e.g., Cooray & Sheth 2002), the AGN clustering signal is the sum of the 1-halo and 2-halo terms, which arise

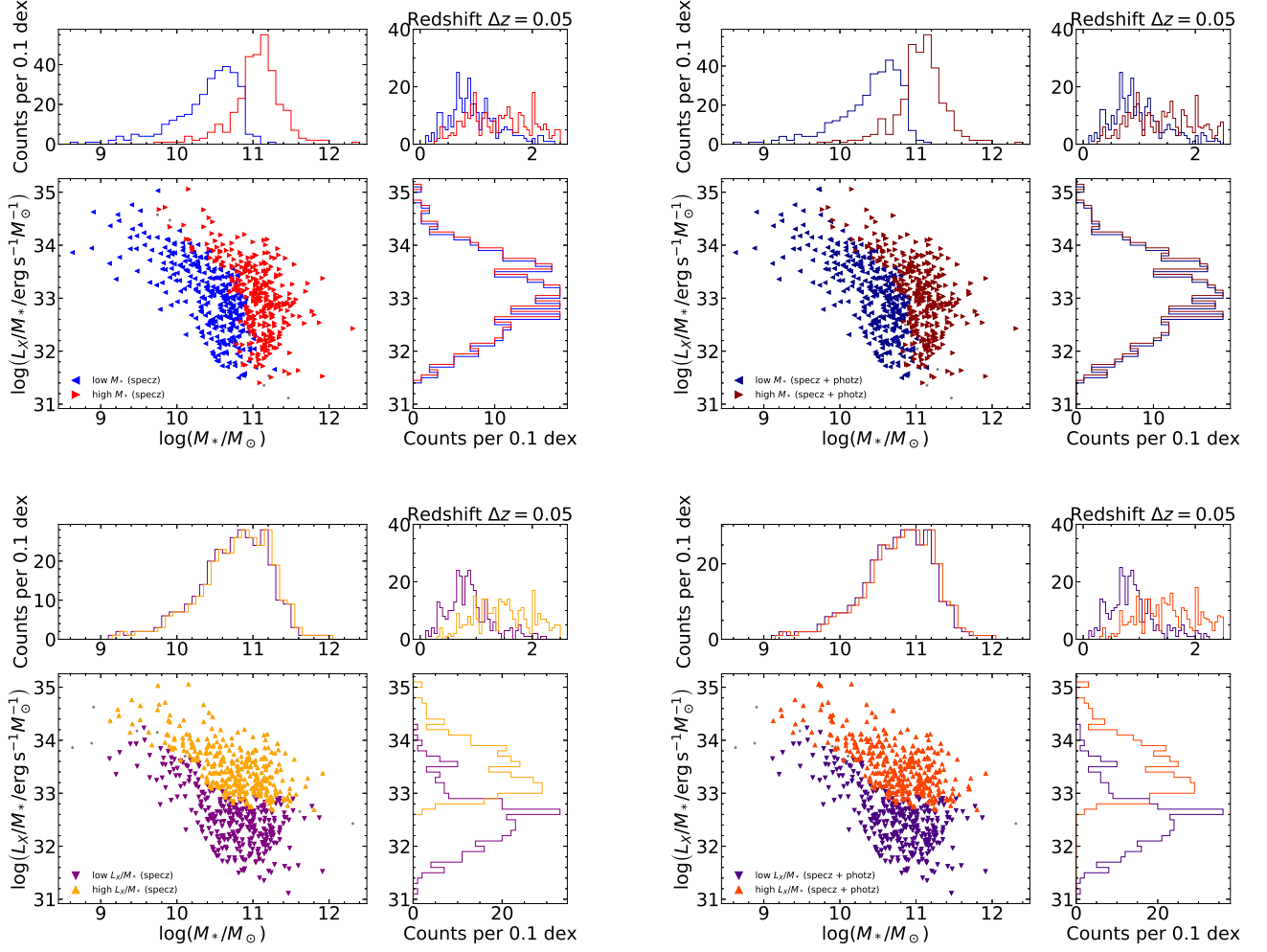


Fig. 3. Distribution in terms of M_* , L_X/M_* , and redshift for XMM-COSMOS AGN with known spec- z (left panels) and spec- z + phot- z Pdfs (right panels). The low and high M_* subsamples are created so that they have exactly the same specific BH accretion rate distribution (upper panels). A similar approach is used in terms of specific BH accretion rate (lower panels). For clarity, when the histograms match exactly, we have slightly offset the bins visually for the high subsample.

from the clustering of AGN that occupy the same halo and two distinct halos, respectively. On large scales ($r_p \gtrsim 1 h^{-1}$ Mpc), the 2-halo term is the dominant term, and the AGN projected 2PCF may be related to the underlying DM projected 2PCF $w_{\text{DM}}^{2\text{-halo}}$ via the linear bias b

$$w_p^{2\text{-halo}}(r_p) = b^2 w_{\text{DM}}^{2\text{-halo}}(r_p), \quad (8)$$

where $w_{\text{DM}}^{2\text{-halo}}$ is estimated at the mean redshift of the corresponding AGN subsample and integrated to the same value of π_{max} . The DM projected 2PCF is related to the DM one-dimensional 2PCF $\xi_{\text{DM}}^{2\text{-halo}}$

$$w_{\text{DM}}^{2\text{-halo}}(r_p) = 2 \int_{r_p}^{\infty} \frac{\xi_{\text{DM}}^{2\text{-halo}}(r) r dr}{\sqrt{r^2 - r_p^2}}, \quad (9)$$

where $\xi_{\text{DM}}^{2\text{-halo}}(r)$ is in turn estimated using the linear power spectrum $P^{2\text{-halo}}(k)$:

$$\xi_{\text{DM}}^{2\text{-halo}}(r) = \frac{1}{2\pi^2} \int P^{2\text{-halo}}(k) k^2 \left[\frac{\sin kr}{kr} \right] dk. \quad (10)$$

We base our estimation of the linear power spectrum on Eisenstein & Hu (1999), which is also implemented in CosmoBolognaLib.

The 1-halo term ($r_p \lesssim 1 h^{-1}$ Mpc) also contains important information on the AGN halo occupation and could contribute toward the clustering signal up to scales $r_p \sim 3 h^{-1}$ Mpc. However, due to low number counts of pairs especially at small scales $r_p \lesssim 3 h^{-1}$ Mpc in our XMM-COSMOS subsamples (see Fig. 5), we are not able to constrain the AGN 1-halo term. Additionally, excluding the 1-halo term from the modeling does not significantly affect our results at large scales.

3.4. Random catalog and error estimation

The random catalog consists of an unclustered set of AGNs with the same selection effects and observational biases. To this end, we follow the work of Miyaji et al. (2007). For our purposes, we drew right ascension and declination at random in the COSMOS field for each random object. Furthermore, right ascension is drawn uniformly, while for declination we drew $\sin(\text{Dec})$ uniformly. Then, we drew a 0.5–2 keV flux from the data catalog, and if the drawn flux was above the limit given by the sensitivity map (Cappelluti et al. 2009, see also Fig. 1), we kept the object. Otherwise, we discarded it. Each random object that was kept was given a redshift drawn from the smoothed redshift distribution of the data catalog with Gaussian smoothing using $\sigma_z = 0.3$. For each of the data catalogs, we created a random catalog with

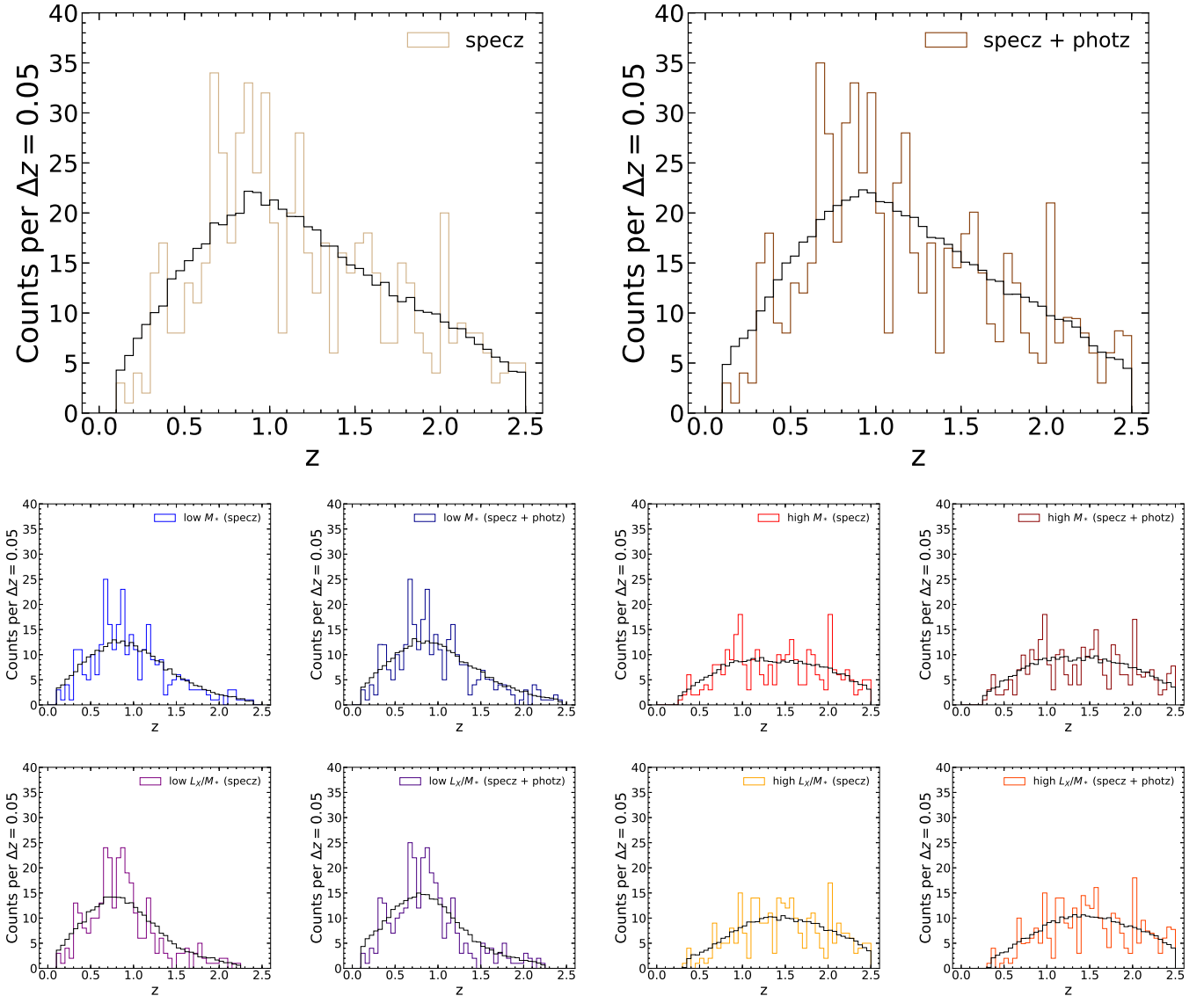


Fig. 4. Redshift distributions of data and random catalogs for our AGN subsamples. The colored histograms show the distribution of the redshifts in the data catalogs. The random redshifts (black histograms) are drawn from the smoothed redshift distribution of the data catalog using a Gaussian smoothing technique with $\sigma_z = 0.3$.

$N_r = 100N_d$. We show the redshift distribution of the data and random catalogs for our AGN subsamples in Fig. 4.

Poissonian errors are readily assigned to the projected 2PCF. However, they are known to underestimate the errors. For this reason, we adopted a Bootstrap resampling technique by dividing the XMM-COSMOS survey into $N_{\text{region}} = 18$ subregions ($3 \times 3 \times 2$ for RA, Dec, and comoving distance, respectively) of roughly equal comoving volumes. We resampled the regions $N_{\text{rs}} = 100$ times. In each of the resamplings, the regions were assigned different weights based on the number of times they were selected (Norberg et al. 2009). The elements of the covariance matrix C are then defined as

$$C_{ij} = \frac{1}{N_{\text{rs}}} \sum_{k=1}^{N_{\text{rs}}} [w_{p,k}(r_{p,i}) - \langle w_p \rangle(r_{p,i})][w_{p,k}(r_{p,j}) - \langle w_p \rangle(r_{p,j})], \quad (11)$$

where i and j refer to the i th and j th r_p bins and the bar denotes the mean over N_{region} resamples. The 1σ error for $w_p(r_{p,i})$ is the square root of the corresponding diagonal element, that is $\sqrt{C_{ii}}$.

4. Results

For each of the AGN subsamples, we estimated the projected 2PCF $w_p(r_p)$ with $r_p = 1.0\text{--}100 h^{-1}$ Mpc using 12 logarithmic bins. We used one bin in the π direction, where the upper limit of this bin is dictated by π_{max} . In order to set π_{max} , we tried out all the values in the range $\pi_{\text{max}} = 20\text{--}75 h^{-1}$ Mpc with an accuracy of $\Delta\pi_{\text{max}} = 5 h^{-1}$ Mpc. For the full spectroscopic AGN sample, we found that the signal converges at $\pi_{\text{max}} = 40 h^{-1}$ Mpc, which was adopted for all the subsamples. This value is similar to previous clustering studies involving XMM-COSMOS AGNs (Gilli et al. 2009; Allevato et al. 2011). The AGN projected 2PCF $w_p(r_p)$ was then estimated using Eq. (7) and the 1σ bootstrap errors were estimated using Eq. (11). We show the estimated projected 2PCF for our subsamples in Fig. 5. Comparison between the spectroscopic subsamples and the specz+photz subsamples are shown in Figs. 5 (full) and 6 (M_* and L_X/M_* subsamples).

We derived the best-fit large-scale bias, defined in Eq. (8), using χ^2 minimization for $r_p = 1\text{--}30 h^{-1}$ Mpc. Furthermore, we utilized the inverse of the full covariance matrix C^{-1} and

minimized $\chi^2 = \Delta^T C^{-1} \Delta$ where Δ was with the same number of elements as the number of r_p bins used in the fit. The symbol Δ is defined explicitly as $\Delta = w_{p, \text{AGN}}^{2\text{-halo}} - b^2 w_{\text{DM}}^{2\text{-halo}}$. With one free parameter, we estimated the 1σ errors on the best-fit bias, given by the lower and upper bounds of the region $(\chi^2 - \chi_{\text{min}}^2)/\nu \leq 1.0$ where $\nu = N - 1$ is the number of degrees of freedom. To exclude noisy bins in the fit, we required that the number of pairs in each bin was >16 . The large-scale bias derived for all the XMM-COSMOS AGN subsamples are summarized in Table 1 and shown in Fig. 7.

For the full spectroscopic AGN subsample (Fig. 5), we found a best-fit bias of $b = 2.20^{+0.37}_{-0.45}$. Following the bias-mass relation described in van den Bosch (2002) and Sheth et al. (2001), this corresponds to a typical mass of the hosting halo of $\log(M_{\text{halo}}/h^{-1} M_{\odot}) = 12.79^{+0.26}_{-0.43}$. It is important to note that in this work we define the typical mass explicitly as the DM halo mass which satisfies $b = b(M_{\text{halo}})$ (e.g., Hickox et al. 2009; Allevato et al. 2016; Mountrichas et al. 2019). Albeit with large uncertainties, we find a small $\lesssim 1\sigma$ difference in the biases of the spectroscopic AGN subsamples split in terms of stellar mass (Fig. 5). The biases are $b = 2.11^{+0.45}_{-0.58}$ for the low stellar mass and $b = 2.69^{+0.61}_{-0.79}$ for the high stellar mass. However, it is worth noting that the two subsamples peak at different redshifts ($z \sim 1.0$ versus $z \sim 1.4$). We find no difference in terms of the typical masses of the hosting halos. For the M_* subsamples, we find that excluding AGNs that are associated with groups has a greater effect on the measured best-fit bias of the low M_* subsample. We measured $b = 1.69^{+0.49}_{-0.72}$ ($b = 2.48^{+0.55}_{-0.71}$) for the low (high) M_* AGN subsample. This lower value for the bias could be an indication that AGNs in galaxies with lower stellar mass are more likely to be satellites in their DM halos compared to AGNs with higher stellar masses.

Moreover, we derived an AGN bias $b = 2.14^{+0.35}_{-0.41}$ (at $z \sim 0.9$) and $b = 2.95^{+0.93}_{-1.42}$ ($z \sim 1.5$) for the low and high L_X/M_* subsamples, respectively (Fig. 5). No significant difference is observed in the typical masses of the hosting halos.

Similar results in terms of bias dependence on M_* and L_X/M_* are found when using phot- z Pdfs in addition to any available spectroscopic redshifts. In particular, in our full AGN subsamples, an increase of $\sim 5\%$ in the weighted number of AGNs introduces no systematic error in the estimation of the bias, but decreases the 1σ error of the bias by $(\delta b_1 - \delta b_2)/\delta b_1 \sim 10\%$ where δb_i is the average error derived from the lower and upper limits of the bias (see Table 1). However, since including photometric redshifts does not change the conclusions drawn from our clustering measurements, we focus on the results from the AGN subsamples with known spectroscopic redshifts in the following sections.

5. Discussion

We performed clustering measurements of 1130 X-ray selected AGN in XMM-COSMOS at $0.1 < z < 2.5$ (mean $z \sim 1.2$) in order to study AGN clustering dependence on the host galaxy stellar mass and the specific BH accretion rate L_X/M_* . For our AGN subsamples, we find a typical DM halo mass $\sim 10^{13} h^{-1} M_{\odot}$ that roughly corresponds to group-sized environments. This is in agreement with similar studies using X-ray selected AGNs at similar redshifts (Coil et al. 2009; Allevato et al. 2011; Fanidakis et al. 2013; Koutoulidis et al. 2013) as well as at lower redshifts $z < 0.1$ (e.g., Krumpke et al. 2018; Powell et al. 2018). We also investigated including photometric redshifts as Pdfs in the analysis in addition to any available spectroscopic redshifts.

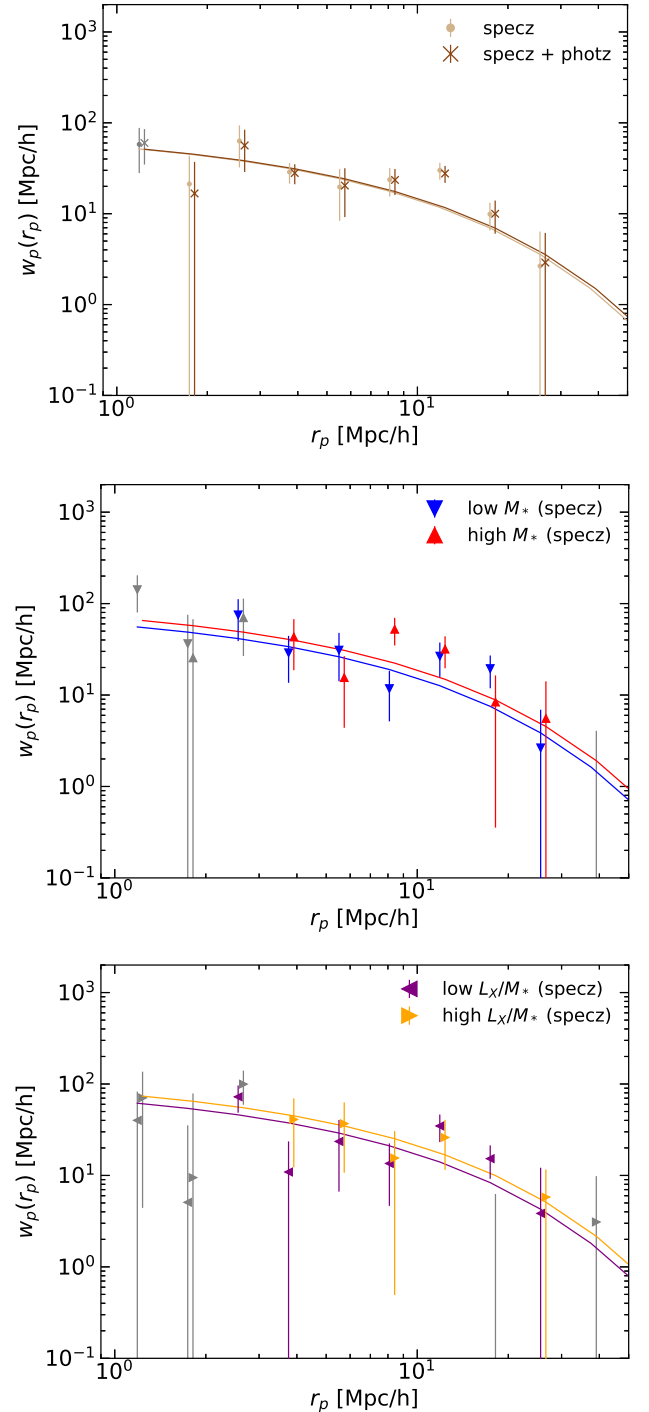


Fig. 5. Measured projected 2PCF for full sample and AGN subsamples. The errorbars correspond to 1σ estimated via the bootstrap method. The solid lines show the squared best-fit bias times the projected DM correlation function estimated at the mean redshift of the particular sample. The gray datapoints are not used in the fit due to a low number of pairs. The excess correlation at $r_p \sim 15 h^{-1} \text{Mpc}$ is likely driven by large structure in the COSMOS field.

In COSMOS, Leauthaud et al. (2015) use weak lensing measurements on X-ray COSMOS AGN at $z < 1$ with $\log L_X/\text{erg s}^{-1} = [41.5-43.5]$ and $\log M_*/M_{\odot} = [10.5-12]$. They infer that 50 percent of AGN reside in halos with $\log M_{\text{halo}}/M_{\odot} < 12.5$. This is not in agreement with the claim that X-ray AGN inhabit group-sized environments with masses $\sim 10^{13} M_{\odot}$. However, they also emphasize that due to the skewed tail in the halo mass distribution, the

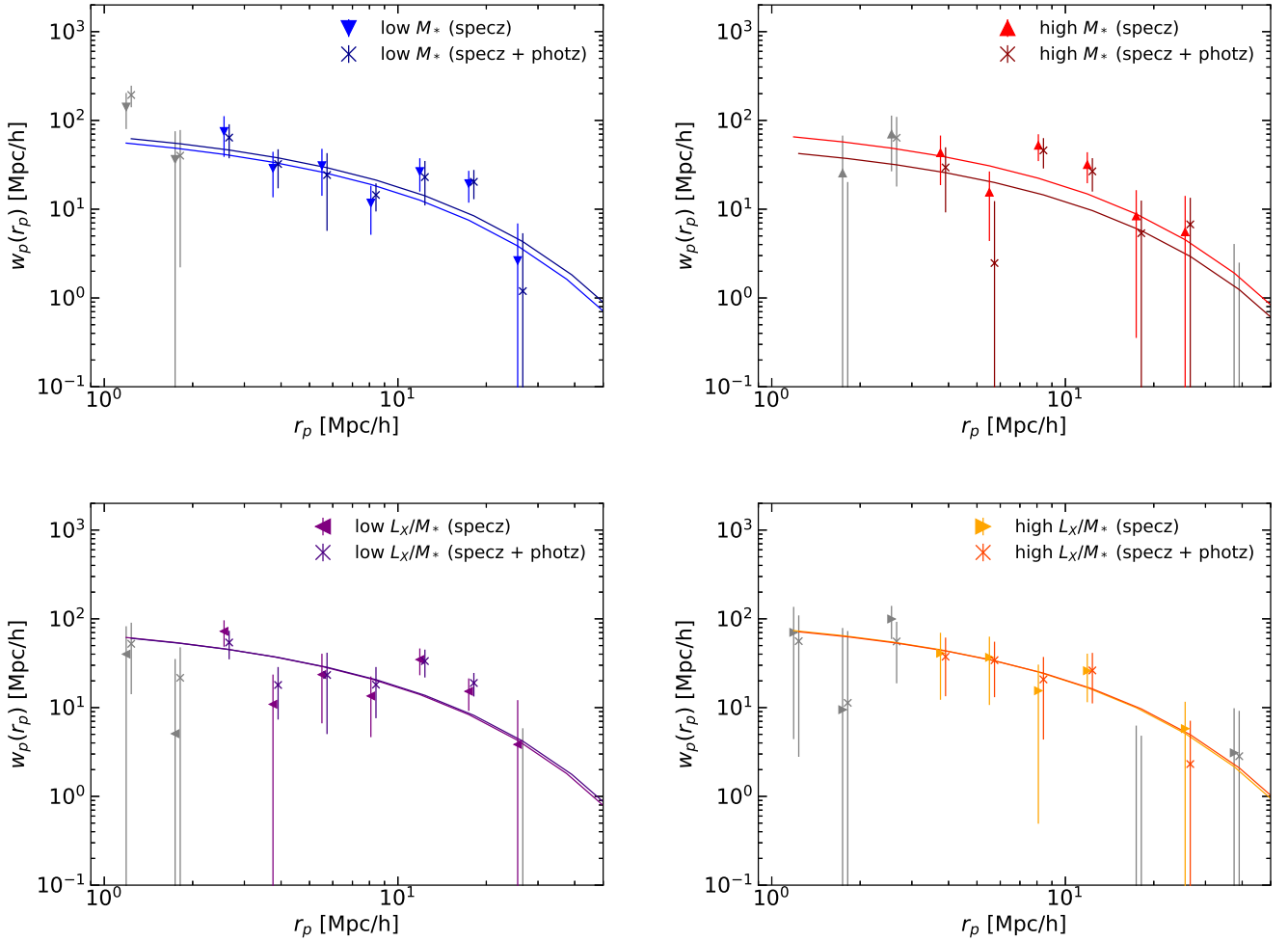


Fig. 6. Effect of including photometric redshifts as Pdfs in estimation of projected 2PCF. Triangles with different orientation mark the different AGN subsamples with spectroscopic redshifts only. Crosses show the projected 2PCF signal when including photometric redshifts as Pdfs. The lines show the squared best-fit bias times the projected DM correlation function. The bins have been slightly offset in the r_p direction for clarity.

typical or the effective halo mass derived from clustering measurements may be markedly different from the median of the distribution.

In fact, they found an effective mass of $M_{\text{eff}} \sim 10^{12.7} M_{\odot}$, which is close to the typical halo masses derived in this work. It is important to specify that they derived the effective halo mass from modeling the AGN halo occupation (see Sect. 5 in Leauthaud et al. 2015), which may differ from the typical halo mass inferred from the 2-halo term as in this work.

Also, they found that the effective DM halo mass of their AGN sample lies between the median and the mean values of the DM halo mass distribution, which are lower and higher than the effective DM halo mass, respectively. Given the statistics in our XMM-COSMOS AGN sample, we are not able to constrain the median or the mean of the DM halo mass distribution. In the future, this could be done with HOD modeling, provided that the 1-halo term is constrained.

Moreover, different cuts in luminosity and host galaxy mass may reflect in different hosting DM halo mass distributions. For instance, our sample of XMM-COSMOS AGN spans a range of host galaxy stellar masses $\log M_*/M_{\odot} = [8-12]$, which also includes low-mass systems with masses $< 10^{10.5} M_{\odot}$ (these are more likely satellite galaxies in galaxy groups) that also probe higher redshifts up to $z = 2.5$.

5.1. Clustering in terms of specific BH accretion rate

We divided the full sample in low and high specific BH accretion rate subsamples with the same M_* distributions and find no significant clustering dependence on L_X/M_* , and thus the Eddington ratio. Krumpe et al. (2015) also found no dependence on λ_{Edd} for their sample of local ($0.16 < z < 0.36$) X-ray and optically selected AGN in the Rosat All-Sky Survey. They concluded that high accretion rates in AGN are not necessarily linked to high-density environments where galaxy interactions would be frequent. Our result provides further evidence that this is also true for non-local AGN at intermediate redshifts $z \sim 1$. Mendez et al. (2016) studied the clustering of AGN in the PRIMUS and DEEP2 surveys (including the COSMOS field) at $z \sim 0.7$ based on multiple selection criteria. In their X-ray selected AGN sample, they did not find a significant dependence on clustering in terms of specific BH accretion rate, which is in line with our results.

5.2. Clustering in terms of host galaxy stellar mass

We also studied the AGN clustering dependence on host galaxy stellar mass by probing the M_*-M_{halo} relation for active galaxies. In Fig. 8, we compare our results for XMM-COSMOS AGN with recent studies in literature using normal, non-active,

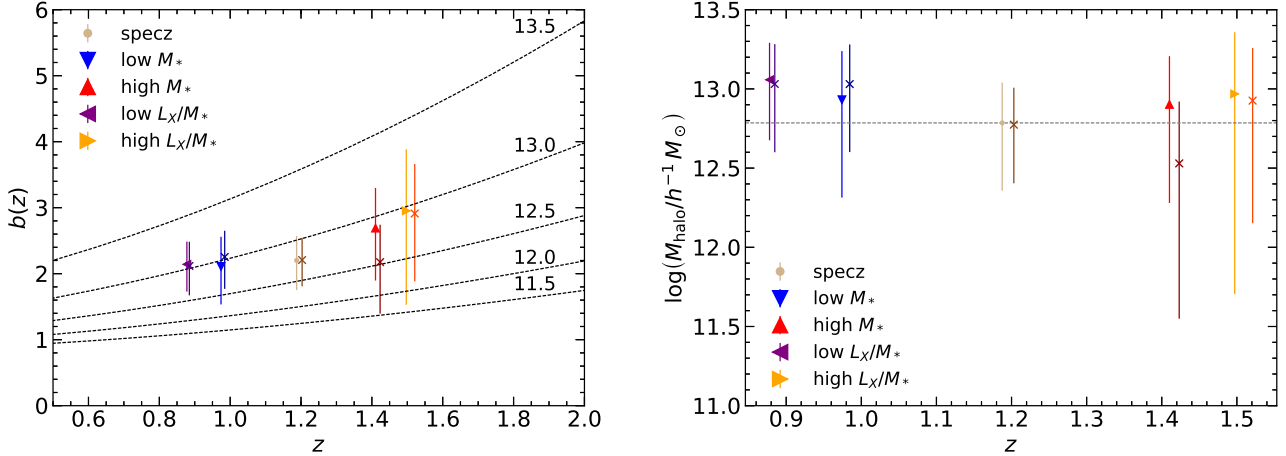


Fig. 7. *Left:* redshift evolution of bias for different XMM-COSMOS AGN subsamples. Crosses mark AGN subsamples where photometric redshifts are included in addition to spectroscopic redshifts. The gray dashed lines correspond to constant halo mass bias evolution $b(z, M_{\text{halo}} = \text{const})$ for $\log M_{\text{halo}} = 11.5, 12.0, 12.5, 13.0, 13.5$ where M_{halo} is given in units of $h^{-1} M_\odot$. *Right:* corresponding typical AGN hosting halo mass evolution with redshift. For visual guidance, the dashed lines show the estimated mass of the halo for the full spectroscopic AGN sample.

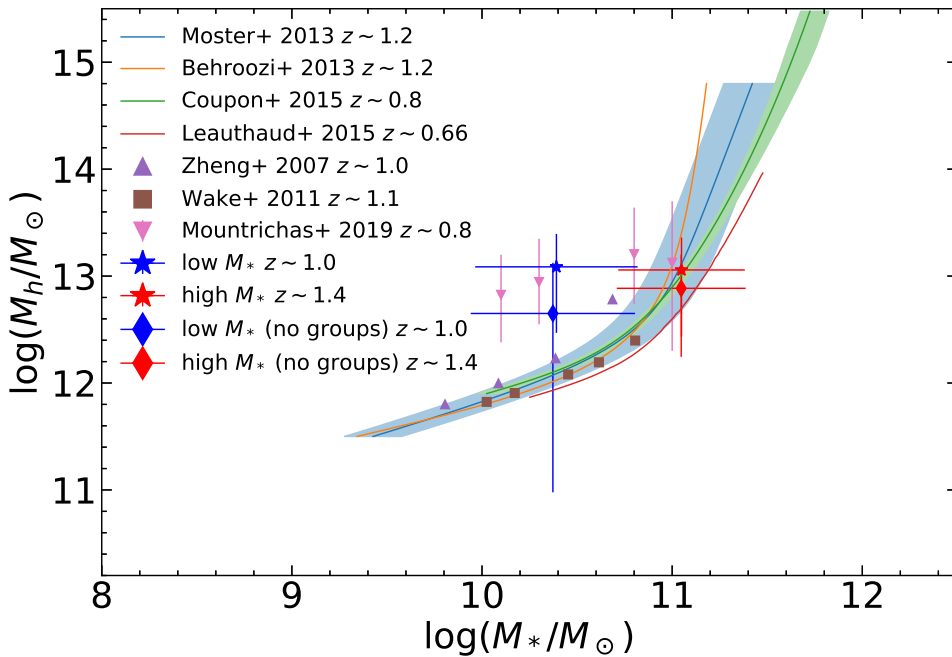


Fig. 8. M_*-M_{halo} relationship for our spectroscopic redshift AGN sample (stars) compared to previous studies in literature according to legend. For each of the M_* subsamples, the horizontal errorbars represent one standard deviation of $\log M_*$ of the sample.

galaxies. For our comparison purposes, we converted the results to our adopted $h = 0.7$ cosmology. Furthermore, we have converted the DM halo masses from different works to be consistent with our definition that is defined with respect to the mean density of the background. The blue curve shows the Moster et al. (2013) M_*-M_{halo} relation for central galaxies estimated using a multi-epoch abundance matching method which we have calculated at the mean redshift $z \sim 1.2$ of our AGN sample. The orange curve shows the galaxy M_*-M_{halo} relation of Behroozi et al. (2013) at $z \sim 1.2$. Coupon et al. (2015) estimated the M_*-M_{halo} relation in the CFHTLenS and VIPERS field at $z \sim 0.8$ using constraints from several different methods including galaxy clustering. Compared to our AGN sample, their sample has a similar range in stellar mass and a slightly lower redshift. Results from HOD modeling of galaxy clustering in DEEP2 (Zheng et al. 2007) and the NMBS (Wake et al. 2011) at comparable redshifts ($z \sim 1.0-1.1$) are shown as well. Using weak lensing methods,

Leauthaud et al. (2015) studied a sample of moderate-luminosity AGN in COSMOS at a lower redshift $z \sim 0.66$ than our sample. At $M_* > 10^{10.5} M_\odot$, they suggest that AGN populate similar DM halos as normal galaxies. Similarly, we find that high M_* ($\gtrsim 10^{10.5} M_\odot$) XMM-COSMOS AGN follow the same M_*-M_{halo} relation as normal non-active galaxies. On the contrary, we estimated that low M_* ($\lesssim 10^{10.5} M_\odot$) AGN are more clustered than normal galaxies. Mountrichas et al. (2019) measured clustering of AGN from the XMM-XXL survey in terms of host galaxy properties (M_* , SFR, and specific SFR) at $z \sim 0.8$ and find a positive dependence on the environment with respect to M_* . Our results at slightly higher redshift are in agreement with their measurements within errors (see Fig. 8).

The M_*-M_{halo} relation obtained from our clustering analysis of XMM-COSMOS AGN is not consistent with results inferred for normal galaxies at similar redshifts, at least for the low M_* bin. In fact, we find that AGN host galaxies with low M_* reside in slightly more massive halos than normal galaxies of similar

stellar mass. On the other hand, at high M_* , our results are in good agreement with the M_*-M_{halo} relation of normal galaxies. As shown in Fig. 8, we do not expect the observed discrepancy at low M_* to be caused by the different mean redshift of the two subsamples ($z \sim 1$ and $z \sim 1.4$). If we exclude AGN that are associated with galaxy groups from our M_* subsamples, we see that this affects our low M_* bin more, while the high M_* bin is left relatively unchanged. This could indicate that XMM-COSMOS AGN with higher M_* are more commonly found in the central galaxies of their respective halos. For lower M_* , the fraction of AGNs as satellites would be higher. Nevertheless, excluding the galaxy groups from the analysis brings our result for the low M_* closer to the M_*-M_{halo} of normal non-active galaxies.

It is important to note that our results for the M_* subsamples include both type1 and type2 AGNs, that is AGNs regardless of obscuration are considered in the same subsample. With the limited sample size of XMM-COSMOS, we are not able to further divide the subsamples and examine the M_*-M_{halo} relation for type1 and type2 AGNs separately, to see whether there are any differences between these two populations. However, this issue can be revisited with *Chandra* COSMOS Legacy Survey AGNs.

6. Conclusions

We measured the clustering of XMM-COSMOS AGN in terms of host galaxy stellar mass M_* and specific BH accretion rate L_X/M_* . Using these two quantities, we created AGN subsamples by splitting the full sample in terms of one quantity, while matching the distribution in the other. In addition, we investigated including AGNs with photometric redshifts as Pdfs in addition to AGNs with known spectroscopic redshifts. From our analysis, we make the following conclusions.

Firstly, XMM-COSMOS AGNs are highly biased with a typical DM halo mass of $M_{\text{halo}} \sim 10^{13} h^{-1} M_{\odot}$, characteristic to group-sized environments and in broad agreement with previous results for moderate-luminosity X-ray selected AGN.

Secondly, we find no significant clustering dependence in terms of specific BH accretion rate, which is consistent with the idea that higher accretion rates in AGNs do not necessarily correspond to more dense environments. Also, we find no significant clustering dependence in terms of host galaxy stellar mass. By comparing our results with various M_*-M_{halo} relations found for normal non-active galaxies, we find that our low M_* AGN subsample is more clustered than what is expected of normal galaxies at similar M_* . We investigated this further by excluding AGNs that are associated with galaxy groups. We find that excluding objects in galaxy groups results in a lower AGN bias for the low M_* AGN subsample, but this does not affect high M_* . This could be due to a higher fraction of satellites for the lower stellar mass systems.

Lastly, our selected quality criterion for including additional photometric redshifts as Pdfs decreases the errors on the measured best-fit bias and does not introduce a bias to the clustering signal. Optimal quality cuts for including photometric redshifts will be studied in a future work.

Acknowledgements. We thank the referee for helpful comments that have improved this paper. AV is grateful to the Vilho, Yrjö and Kalle Väisälä Foundation of the Finnish Academy of Science and Letters. VA acknowledges funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 749348. TM is supported by UNAM-DGAPA PAPIIT IN111379 and CONACyT 252531. RG acknowledges support from the agreement ASI-INAF n.2017-14-H.O.

References

- Alexander, D. M., & Hickox, R. C. 2012, *New Astron. Rev.*, 56, 93
 Allevalo, V., Finoguenov, A., Cappelluti, N., et al. 2011, *ApJ*, 736, 99
 Allevalo, V., Finoguenov, A., Hasinger, G., et al. 2012, *ApJ*, 758, 47
 Allevalo, V., Finoguenov, A., Civano, F., et al. 2014, *ApJ*, 796, 4
 Allevalo, V., Civano, F., Finoguenov, A., et al. 2016, *ApJ*, 832, 70
 Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, *ApJ*, 770, 57
 Bongiorno, A., Merloni, A., Brusa, M., et al. 2012, *MNRAS*, 427, 3103
 Brusa, M., Civano, F., Comastri, A., et al. 2010, *ApJ*, 716, 348
 Cappelluti, N., Hasinger, G., Brusa, M., et al. 2007, *ApJS*, 172, 341
 Cappelluti, N., Brusa, M., Hasinger, G., et al. 2009, *A&A*, 497, 635
 Cappelluti, N., Allevalo, V., & Finoguenov, A. 2012, *Adv. Astron.*, 2012, 85370
 Civano, F., Marchesi, S., Comastri, A., et al. 2016, *ApJ*, 819, 62
 Coil, A. L., Georgakakis, A., Newman, J. A., et al. 2009, *ApJ*, 701, 1484
 Cooray, A., & Sheth, R. 2002, *Phys. Rep.*, 372, 1
 Coupon, J., Arnouts, S., van Waerbeke, L., et al. 2015, *MNRAS*, 449, 1352
 Croom, S. M., Boyle, B. J., Shanks, T., et al. 2005, *MNRAS*, 356, 415
 da Ângela, J., Shanks, T., Croom, S. M., et al. 2008, *MNRAS*, 383, 565
 Davis, M., & Peebles, P. J. E. 1983, *ApJ*, 267, 465
 Eisenstein, D. J., & Hu, W. 1999, *ApJ*, 511, 5
 Fanidakis, N., Georgakakis, A., Mountrichas, G., et al. 2013, *MNRAS*, 435, 679
 Georgakakis, A., Mountrichas, G., Salvato, M., et al. 2014, *MNRAS*, 443, 3327
 Gilli, R., Zamorani, G., Miyaji, T., et al. 2009, *A&A*, 494, 33
 Gozaliasl, G., Finoguenov, A., Tanaka, M., et al. 2019, *MNRAS*, 483, 3545
 Hasinger, G., Cappelluti, N., Brunner, H., et al. 2007, *ApJS*, 172, 29
 Hasinger, G., Capak, P., Salvato, M., et al. 2018, *ApJ*, 858, 77
 Hickox, R. C., Jones, C., Forman, W. R., et al. 2009, *ApJ*, 696, 891
 Koutoulidis, L., Plionis, M., Georgantopoulos, I., & Fanidakis, N. 2013, *MNRAS*, 428, 1382
 Koutoulidis, L., Georgantopoulos, I., Mountrichas, G., et al. 2018, *MNRAS*, 481, 3063
 Krumpe, M., Miyaji, T., & Coil, A. L. 2014, *Multifrequency Behaviour of High Energy Cosmic Sources*, 71
 Krumpe, M., Miyaji, T., Husemann, B., et al. 2015, *ApJ*, 815, 21
 Krumpe, M., Miyaji, T., Coil, A. L., & Aceves, H. 2018, *MNRAS*, 474, 1773
 Landy, S. D., & Szalay, A. S. 1993, *ApJ*, 412, 64
 Leauthaud, A. J., Benson, A., Civano, F., et al. 2015, *MNRAS*, 446, 1874
 Marchesi, S., Civano, F., Elvis, M., et al. 2016, *ApJ*, 817, 34
 Marulli, F., Veropalumbo, A., & Moresco, M. 2016, *Astron. Comput.*, 14, 35
 Mendez, A. J., Coil, A. L., Aird, J., et al. 2016, *ApJ*, 821, 55
 Merloni, A., Alexander, D. A., Banerji, M., et al. 2019, *The Messenger*, 175, 42
 Miyaji, T., Zamorani, G., Cappelluti, N., et al. 2007, *ApJS*, 172, 396
 Moster, B. P., Naab, T., & White, S. D. M. 2013, *MNRAS*, 428, 3121
 Mountrichas, G., Georgakakis, A., Menzel, M.-L., et al. 2016, *MNRAS*, 457, 4195
 Mountrichas, G., Georgakakis, A., & Georgantopoulos, I. 2019, *MNRAS*, 483, 1374
 Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, *MNRAS*, 396, 19
 Powell, M. C., Cappelluti, N., Urry, C. M., et al. 2018, *ApJ*, 858, 110
 Ross, N. P., Shen, Y., Strauss, M. A., et al. 2009, *ApJ*, 697, 1634
 Salvato, M., Hasinger, G., Ilbert, O., et al. 2009, *ApJ*, 690, 1250
 Salvato, M., Ilbert, O., Hasinger, G., et al. 2011, *ApJ*, 742, 61
 Scoville, N., Aussel, H., Brusa, M., et al. 2007, *ApJS*, 172, 1
 Sheth, R. K., Mo, H. J., & Tormen, G. 2001, *MNRAS*, 323, 1
 van den Bosch, F. C. 2002, *MNRAS*, 331, 98
 Wake, D. A., Whitaker, K. E., Labbé, I., et al. 2011, *ApJ*, 728, 46
 Zheng, Z., Coil, A. L., & Zehavi, I. 2007, *ApJ*, 667, 760