

# The clustering of X-ray AGN at $0.5 < z < 4.5$ : host galaxies dictate dark matter halo mass

Charutha Krishnan,<sup>1</sup>★ Omar Almaini<sup>1</sup>,<sup>1</sup> Nina A. Hatch<sup>1</sup>,<sup>1</sup> Aaron Wilkinson,<sup>2</sup> David T. Maltby,<sup>1</sup> Christopher J. Conselice,<sup>1</sup> Dale Kocevski,<sup>3</sup> Hyewon Suh<sup>4</sup> and Vivienne Wild<sup>1</sup><sup>5</sup>

<sup>1</sup>*School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK*

<sup>2</sup>*Department of Physics and Astronomy, Universiteit Gent, Krijgslaan 281, S9, B-9000 Gent, Belgium*

<sup>3</sup>*Department of Physics and Astronomy, Colby College, Waterville, ME 04901, USA*

<sup>4</sup>*Subaru Telescope, National Astronomical Observatory of Japan (NAOJ), National Institutes of Natural Sciences (NINS), 650 North A'ohōku place, Hilo, HI 96720, USA*

<sup>5</sup>*School of Physics and Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK*

Accepted 2020 March 19. Received 2020 March 6; in original form 2019 February 22

## ABSTRACT

We present evidence that active galactic nuclei (AGN) do not reside in ‘special’ environments, but instead show large-scale clustering determined by the properties of their host galaxies. Our study is based on an angular cross-correlation analysis applied to X-ray selected AGN in the COSMOS and UDS fields, spanning redshifts from  $z \sim 4.5$  to  $z \sim 0.5$ . Consistent with previous studies, we find that AGN at all epochs are on average hosted by galaxies in dark matter haloes of  $10^{12}$ – $10^{13} M_{\odot}$ , intermediate between star-forming and passive galaxies. We find, however, that the same clustering signal can be produced by inactive (i.e. non-AGN) galaxies closely matched to the AGN in spectral class, stellar mass, and redshift. We therefore argue that the inferred bias for AGN lies in between the star-forming and passive galaxy populations because AGN host galaxies are comprised of a mixture of the two populations. Although AGN hosted by higher mass galaxies are more clustered than lower mass galaxies, this stellar mass dependence disappears when passive host galaxies are removed. The strength of clustering is also largely independent of AGN X-ray luminosity. We conclude that the most important property that determines the clustering in a given AGN population is the fraction of passive host galaxies. We also infer that AGN luminosity is likely not driven by environmental triggering, and further hypothesize that AGN may be a stochastic phenomenon without a strong dependence on environment.

**Key words:** galaxies: active – galaxies: evolution – galaxies: haloes – large-scale structure of Universe.

## 1 INTRODUCTION

The astronomical community is in consensus that essentially all massive galaxies host supermassive black holes (SMBHs), that are observed as active galactic nuclei (AGN) during their phases of intense mass accretion. Numerous lines of observational evidence point towards a tight correlation between the evolution of galaxies and that of their SMBHs. For instance, the total star formation rate density and the total AGN accretion density appear to follow similar trends from  $z \sim 2$  to  $z \sim 0$  (e.g. Boyle et al. 1998; Franceschini et al. 1999; Silverman et al. 2008; Kormendy & Ho 2013, for a recent

review), implying a link between the mass growth of SMBHs and their host galaxies. The properties of galaxies are also correlated with their environment (e.g. Gisler 1978; Dressler et al. 1999; De Lucia et al. 2006; Conselice 2014), which suggests that SMBHs may also be linked to their large-scale environment. Indeed, several studies have found correlations between the abundance of AGN and their environment. For example, Kauffmann et al. (2004) showed there is an enhancement of AGN in low-density regions compared to high-density regions in the local Universe, whilst at higher redshift the opposite is found, with an enhanced fraction of AGN in high-redshift protoclusters relative to the field (Lehmer et al. 2009; Digby-North et al. 2010; Krishnan et al. 2017).

A useful tool to examine the possible connection between AGN and the large-scale environment in a statistical manner is the angular

\* E-mail: charutha.krishnan@nottingham.ac.uk

and spatial clustering of AGN. As SMBHs populate collapsed dark matter haloes in the  $\Lambda$  cold dark matter paradigm, they can be assumed to reflect the peaks in the spatial distribution of dark matter in the Universe. The 2-point correlation function (2pcf) is the most commonly used tool for large-scale clustering analysis (Peebles 1980). The 2pcf of galaxies expresses the excess probability of finding pairs of galaxies, above a random distribution. Comparison of the observed 2pcf to that of dark matter from the outputs of detailed dark matter simulations allows the determination of the dark matter halo masses of galaxies hosting AGN. In theory, the typical large-scale environments of AGN can then be inferred as a function of cosmic time, providing potential insights into the connection between AGN and their large-scale environments.

The 2pcf of quasars has been studied extensively in the literature and are found to reside in  $10^{12-12.7} M_{\odot}$  haloes out to  $z \sim 4$  (Croom et al. 2005; Shen et al. 2007; Ross et al. 2009; Eftekharzadeh et al. 2015; Ikeda et al. 2015; García-Vergara et al. 2017). Similarly, lower luminosity broad-line AGN out to  $z \sim 0.5$  are hosted by dark matter haloes of  $10^{13} M_{\odot}$  (Miyaji et al. 2011; Krumpe et al. 2012), and X-ray AGN across a wide range of redshifts ( $z \sim 0.05$ ,  $z \sim 0.1$ ,  $z \sim 0.98$ ,  $z \sim 1.25$ , and  $z \sim 3.4$ ) inhabit haloes of  $\sim 10^{13} M_{\odot}$  (Bradshaw et al. 2011; Mountrichas & Georgakakis 2012; Koutoulidis et al. 2013; Allevalo et al. 2016; Powell et al. 2018, respectively). Many of these studies interpret the  $10^{12-13} M_{\odot}$  halo mass as evidence that groups are the ideal environment in which AGN are triggered through galaxy–galaxy interactions (e.g. Miyaji et al. 2011; Ikeda et al. 2015). Inconsistent with this interpretation, however, is the lack of evidence for an enhancement of AGN in groups relative to the field (Shen et al. 2007; Arnold et al. 2009; Oh et al. 2014; Tzanavaris et al. 2014).

Interpretation of the halo mass derived from clustering is complicated because it may not necessarily represent the environments of AGN, unless AGN are a consistent population inhabiting a single environment. If AGN are comprised of different populations of galaxies, each inhabiting a different environment, then the clustering of AGN will indicate the average of these environments. In fact, recent studies at  $z \lesssim 1$  suggest that host galaxies play a major role in clustering measurements of AGN. For example, at  $z \sim 0.7$ , Mendez et al. (2016) find significant differences in the clustering measurements of X-ray, mid-IR, and radio AGN, and that these differences are driven by differences in host galaxy properties (such as stellar mass and star formation rate). Furthermore, at  $z < 0.1$ , Powell et al. (2018) find that, when accounting for host galaxy properties, AGN occupy dark matter haloes consistent with the overall inactive galaxy population.

Drawing a consistent picture of the clustering of AGN is difficult because previous studies select AGN in various methods with different surveys and telescope sensitivities, thus sampling different distributions of host galaxy properties. Furthermore, the clustering is quantified using differing methods (e.g. angular, projected, and real-space correlation functions) and a diverse range of models to estimate halo masses. While several works derive bias values of AGN/quasars using a broad redshift range, it has not been possible as yet for a single study to measure the *evolution* of the AGN bias with redshift from  $z \sim 4.5$  to  $z \sim 0.5$ , due to the large samples required to obtain reliable bias measurements from the AGN autocorrelation function (ACF). In this work, we are able to use the angular cross-correlation function (CCF) using the more numerous underlying galaxy sample from UDS and COSMOS, as these surveys provide the unique combination of depth and area required to detect large numbers of galaxies out to  $z \sim 4$ . This cross-correlation technique allows us to then reliably infer the ACFs of AGN, affording us the

opportunity to split our X-ray AGN sample into several redshift intervals. In this study, we are also able to investigate potential interpretations of the clustering signal of AGN by considering the influence of host galaxy properties such as mass and star formation characteristics.

This paper is structured as follows. We describe the data sets used in Section 2 as well as our sample selection. Section 3 explains the methodology of our clustering analysis, and we present our results in Section 4. We adopt a WMAP9 cosmology (Hinshaw et al. 2013), with  $\Omega_m = 0.3$ ,  $\Omega_{\Lambda} = 0.7$ , and  $h = 0.7$ . All magnitudes are in the AB system. All X-ray luminosities quoted are calculated in rest-frame bands using a power-law model with a photon index  $\Gamma = 2$ . We note that the effect of Galactic absorption on our fluxes is negligible.

## 2 DATA SETS AND SAMPLE SELECTION

In this section, we review the data used in our study and describe our sample selection.

### 2.1 Description of the data

We make use of two deep and wide near-infrared surveys for our galaxy samples (Sections 2.1.1 and 2.1.2) with *Chandra* and *XMM* X-ray coverage.

#### 2.1.1 UDS DR11 catalogue

The first data set used to form our *K* band selected galaxy sample is the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al. 2007) Ultra Deep Survey (UDS). The UDS is a deep photometric survey centred on RA = 02:17:48, Dec. = −05:05:57, covering a survey area of  $0.62 \text{ deg}^2$  after removing masked regions such as bright stars and image artefacts. In this work, we use the 11th data release of the UDS (UDS DR11). The  $5\sigma$  limiting depths in 2 arcsec diameter apertures are 25.6, 25.1, and 25.3 mag in the *J*, *H*, and *K* bands, respectively (Hartley et al. 2013; Almaini et al. in preparation). This catalogue uses deep *Y*-band observations provided by the VISTA VIDEO survey (Jarvis et al. 2013), optical observations (*U*, *B*, *V*, *R*, *i'*, *z'*) from the Subaru *XMM–Newton* Deep Survey (SXDS; Furusawa et al. 2008), as well as IRAC 3.6 and  $4.5 \mu\text{m}$  coverage from the *Spitzer* UDS Legacy Program (SpUDS, PI:Dunlop) and the *Spitzer* Extended Deep Survey (SEDS; Ashby et al. 2013).

As described in Simpson et al. (2013), photometric redshifts were determined by fitting 12-band photometry using a library of templates built from the Bruzual & Charlot (2003) models with range of ages and metallicities, using a Small Magellanic Cloud extinction law. These photometric redshifts have a normalized median absolute deviation of  $\sigma_{\text{NMAD}} = 0.019$  as compared to  $\sim 7000$  secure spectroscopic redshifts, with an outlier fraction of 4.5 per cent (objects with  $|z_p - z_s|/(1 + z_s) > 0.15$ ). Spectroscopic redshifts are used when available. We refer the reader to Almaini et al. (in preparation) for further details on the UDS DR11 catalogue.

#### 2.1.2 COSMOS2015 catalogue

The second data set we use is the Cosmic Evolution Survey (COSMOS; Scoville et al. 2007). COSMOS is a comparable survey to the UDS that reaches shallower depths but covers a larger area of  $1.5 \text{ deg}^2$  in the UltraVISTA-DR2 region centred on

RA = 10:00:28, Dec. = +02:12:21. We draw our galaxy sample from *COSMOS2015* published in Laigle et al. (2016), with PSF-matched photometry from Subaru SuprimeCam in the  $u$ ,  $B$ ,  $V$ ,  $r$ ,  $i$ , and  $z$  bands. UltraVISTA (McCracken et al. 2012) provide NIR photometry in the  $Y$ ,  $J$ ,  $H$ , and  $K_s$  bands. Although UltraVISTA comprises of ‘ultradeep’ stripes over roughly half the area, we limit our sample with  $K$ -band completeness limits corresponding to shallower regions to maximize number statistics while selecting galaxies uniformly across the field. Subaru Hyper Suprime-Cam (HSC)  $Y$ -band imaging and *Spitzer* IRAC observations (3.6 and 4.5  $\mu$ m bands) are also included.

We use photometric redshifts from Laigle et al. (2016) that have made use of Bruzual & Charlot (2003) models, as well as templates of spiral and elliptical galaxies from Polletta et al. (2007). As described in Laigle et al. (2016), the uncertainty of the photometric redshifts varies from  $\sigma_{\text{NMAD}} = 0.007$  with an outlier fraction of 0.5 per cent (for bright, low redshift, star-forming galaxies) to  $\sigma_{\text{NMAD}} = 0.021$  with an outlier fraction of 13.2 per cent (for the  $z > 3$  sample), where outliers are objects with  $|z_p - z_s|/(1 + z_s) > 0.15$ . We use spectroscopic redshifts when available. The reader is referred to Laigle et al. (2016) for further details on *COSMOS2015*.

### 2.1.3 X-ray AGN counterpart catalogues

For both UDS and COSMOS, we use AGN catalogues that identify optical/NIR counterparts for X-ray sources using a maximum-likelihood approach. The UDS field has wide but shallow *XMM* observations, as well as deep *Chandra* coverage of a smaller fraction of the field. Therefore we use the *Chandra* counterparts, but supplement this with *XMM* counterparts outside the *Chandra* covered region.

The X-UDS observations were carried out over 1.25 Ms with *Chandra*’s Advanced CCD Imaging Spectrometer (ACIS; Garmire et al. 2003), 25 ACIS-I pointing positions that cover a total area of roughly 35 arcmin  $\times$  25 arcmin in size. The final X-ray source catalogue, presented in Kocevski et al. (2018), amounts to 868 sources. Counterparts to these X-ray sources were matched to the CANDELS  $H$  band and UDS DR10  $K$ -band catalogues using the likelihood ratio technique of Sutherland & Saunders (1992), following the method outlined in Civano et al. (2012). Spectroscopic redshifts are available for  $\sim 400$  sources. See Hasinger et al. (in preparation) for further details on the counterpart matching procedure.

The UKIDSS UDS survey is also observed by the Subaru-XMM-Newton Deep Survey (SXDS), mapped by seven pointings with *XMM-Newton* covering 0.2–10 keV. The X-ray source catalogue is presented in Ueda et al. (2008), amounting to 1245 sources. *XMM* counterparts have been obtained using the likelihood ratio method to  $R$  band, 3.6  $\mu$ m, near-UV, and 24  $\mu$ m source catalogues. Spectroscopic observations allow the identification of 597 out of 896 total AGN. The remaining AGN have redshifts derived using 15 band photometry, where separate SED templates of QSOs and galaxies are applied to each counterpart. See Akiyama et al. (2015) for further details.

The *Chandra* COSMOS-Legacy is the product of 4.6 Ms of *Chandra* observations over the 2.2 deg<sup>2</sup> COSMOS area. The X-ray source catalogue is described in Civano et al. (2016) and amounts to 4016 sources. Marchesi et al. (2016) identify counterparts to the X-ray sources using the approach of Civano et al. (2012), making use of three different bands:  $i$ ,  $K_s$ , and 3.6  $\mu$ m. Of the 4016 sources,  $\sim 97$  per cent have optical/near-infrared counterparts,

and  $\sim 54$  per cent have spectroscopic redshifts. See Marchesi et al. (2016) for further details on the counterpart catalogue.

These surveys are comparable in the hard band (2–10 keV) and we thus only study AGN detected in this band. Photometric and spectroscopic redshifts of the X-ray sources in X-UDS, SXDS, and *Chandra* COSMOS-Legacy were provided by the most likely counterpart. In all cases, we use the spectroscopic redshift when available.

### 2.1.4 Supercolour catalogues

We determine the spectral class of our AGN and galaxy samples using supercolour classifications that utilize a principal component analysis (PCA) of the optical/near-infrared photometric data in the UDS and COSMOS fields (Wild et al. 2014, 2016). The eigenvectors for the PCA are built from a library of Bruzual & Charlot (2003) spectral synthesis models with a variety of ages, metallicities, star formation histories, and dust contents. The top three eigenvectors are found to represent  $> 99.9$  per cent of the variance in the model SED library and the shape of an observed galaxy SED can be represented by the weights of these eigenvectors, termed ‘supercolours’ (SC1, SC2, and SC3). This method separates a tight red-sequence from star-forming galaxies and also identifies rarer populations such as post-starburst (PSB) and dusty star-forming galaxies, depending on where they lie in supercolour space. We use the updated boundaries presented in Wilkinson et al. (in preparation) to classify our galaxies and AGN host as either passive (red + PSB) or star forming (SF + dusty). The spectral classes and their boundaries, for both the UDS and COSMOS fields, are shown in Table 1.

As the broad-band filters are slightly different between the UDS and COSMOS fields, the PCA analysis is performed separately on each field resulting in different eigenvectors. Wilkinson et al. (in preparation) use models and spectroscopically classified galaxies to ensure consistency in the classification boundaries in SC space between the two fields. Due to deeper data in the UDS field, galaxies can robustly be classified to  $z = 3$ , but galaxies in COSMOS are limited to  $z = 2.5$ . Therefore we adopt a uniform cut in redshift at  $z < 2.5$  when investigating the role of host galaxy properties.

The stellar masses used in this work are determined using this supercolour technique, assuming a Chabrier (2003) initial mass function (IMF) and Bruzual & Charlot (2003) stellar population templates (see Wild et al. 2016 and Wilkinson et al. in preparation for more details). The uncertainty from the supercolour Bayesian stellar mass fitting is typically 0.1 dex, with no indication of additional uncertainty for AGN. This uncertainty allows for the degeneracy between fitted parameters and the photometric redshift uncertainties. Further discussion of the supercolour stellar mass uncertainties can be found in Almaini et al. (2017). Based on this work, we expect our results to be robust to known sources of random and systematic errors.

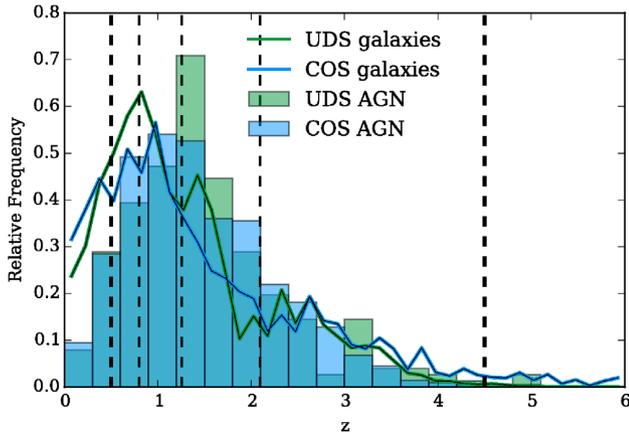
## 2.2 Sample selection

### 2.2.1 Galaxy samples

To study the evolution of clustering of AGN as a function of cosmic time, we split our AGN and galaxy samples in redshift intervals of  $0.5 < z < 0.8$ ,  $0.8 < z < 1.3$ ,  $1.3 < z < 2.1$ , and  $2.1 < z < 4.5$  corresponding to equal cosmic time intervals of 1.8 Gyr. In order to maximize the quality of our galaxy sample, we apply a maximum limit to the minimum  $\chi^2$  values associated with fitting photometric

**Table 1.** Spectral class definitions according to supercolour boundaries.

Spectral class	UDS	COSMOS
SF	$SC2 < 0.4 \times SC1 + 10.86$ $\cap SC2 < 0.783 \times SC1 + 14.83$ $\cap SC2 > -0.2 \times SC1 - 13.75$	$SC2 < 0.38 \times SC1 + 11.67$ $\cap SC2 < 0.82 \times SC1 + 16.39$ $\cap SC2 > -0.18 \times SC1 - 12.12$
Red	$SC2 > 0.783 \times SC1 + 14.83$ $\cap SC2 < -0.34 \times SC1 + 3.19$	$SC2 > 0.82 \times SC1 + 16.39$ $\cap SC2 < -0.26 \times SC1 + 3.94$
PSB	$SC2 > 0.4 \times SC1 + 10.86$ $\cap SC2 > -0.34 \times SC1 + 3.19$	$SC2 > 0.38 \times SC1 + 11.67$ $\cap SC2 > -0.26 \times SC1 + 3.94$
Dusty	$SC2 < -0.2 \times SC1 - 13.75$ $SC1 < -20$	$SC2 < -0.18 \times SC1 - 12.12$ $SC1 < -19.5$



**Figure 1.** Redshift distribution of the AGN in the UDS and COSMOS fields are shown by the green and blue histograms, respectively. The galaxy redshift distributions within these two fields are shown by the solid lines. The black dashed lines represent our redshift intervals of  $0.5 < z < 0.8$ ,  $0.8 < z < 1.3$ ,  $1.3 < z < 2.1$ ,  $2.1 < z < 4.5$ , and the thicker lines denote the redshift limits of our study.

redshifts. We also apply a  $K$ -band magnitude limit of 23.7 to our galaxy sample. After applying these quality cuts, our galaxy sample consists of  $\sim 60\,000$  and  $\sim 170\,000$  galaxies between  $0.5 < z < 4.5$  in the UDS and COSMOS fields, respectively. In addition, we use the methodology from Pozzetti et al. (2010) to apply redshift dependent 90 per cent mass completeness limits to the galaxy sample. We do not apply a mass completeness cut to our AGN sample to maximize the sample size, but we note that 95 per cent of our AGN are above the 90 per cent mass completeness limit of  $10^{10.2} M_{\odot}$  at  $z = 2.5$ . The clustering measurements are robust to more conservative cuts in the  $K$ -band magnitude and mass completeness, as results are consistent within errorbars, albeit with larger uncertainties. The redshift distributions of the galaxy samples are shown by the solid lines in Fig. 1.

### 2.2.2 AGN samples

It is important to select AGN detected to a uniform flux limit to ensure that the clustering measurements are not biased by the varying source density with exposure. We select our hard band AGN to a flux limit of  $3.55 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ . AGN outside the area corresponding to this flux limit are removed, as well as AGN within the region that have fainter fluxes than the flux limit. This limit was chosen to maximize the number of AGN, and applied to AGN in both the UDS and COSMOS fields. In the UDS, the X-UDS coverage is deeper but limited to the central  $\sim 0.33 \text{ deg}^2$ . Outside

this Chandra-covered region, we therefore supplement our X-UDS data with *XMM-Newton* data.

While the UDS achieves greater depths in both  $K$  band and X-rays, the larger number of AGN in the COSMOS field effectively dictates our flux limit. To ensure a consistent depth, we therefore discard the fainter data in the UDS and adopt a shallower X-ray flux limit and  $K$ -band magnitude limit. We also make maximum use of the UDS field by computing the optimum flux limit for the UDS field independently ( $1.25 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ ) and measuring the clustering of AGN in this field. The results are consistent with the measurements using a flux limit of  $3.55 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ , with smaller uncertainties. The redshift distributions of the AGN selected to the latter flux limit are shown in Fig. 1. Fig. 2 shows the spatial distribution of the AGN in UDS (left) and COSMOS (right). The flux limit contours optimized for UDS and COSMOS are overlaid in green and light blue, respectively. AGN selected to these flux limits are highlighted by the respective colours.

In our study of the links between clustering and AGN luminosity, we divide AGN into X-ray luminosity and redshift bins (see boxes in Fig. 3). To the AGN in each bin, we apply a flux limit computed with the lower end of the luminosity bin and median redshift of the AGN within the bin. The diagonal lines traced out by missing points in the bottom right hand corners of the boxes in Fig. 3 correspond to the sources that have been removed as their fluxes were lower than the applied flux limit.

## 3 CLUSTERING METHODS

In this section, we first measure the angular 2pcf of AGN (Section 3.1.1) and that of the underlying dark matter (Section 3.1.2). We then measure the strength of clustering, using the ‘bias’ parameter we describe in Section 3.2, by scaling the dark matter CF to the AGN CF and minimizing  $\chi^2$ . In this section, we also derive the dark matter halo masses of AGN as a function of redshift.

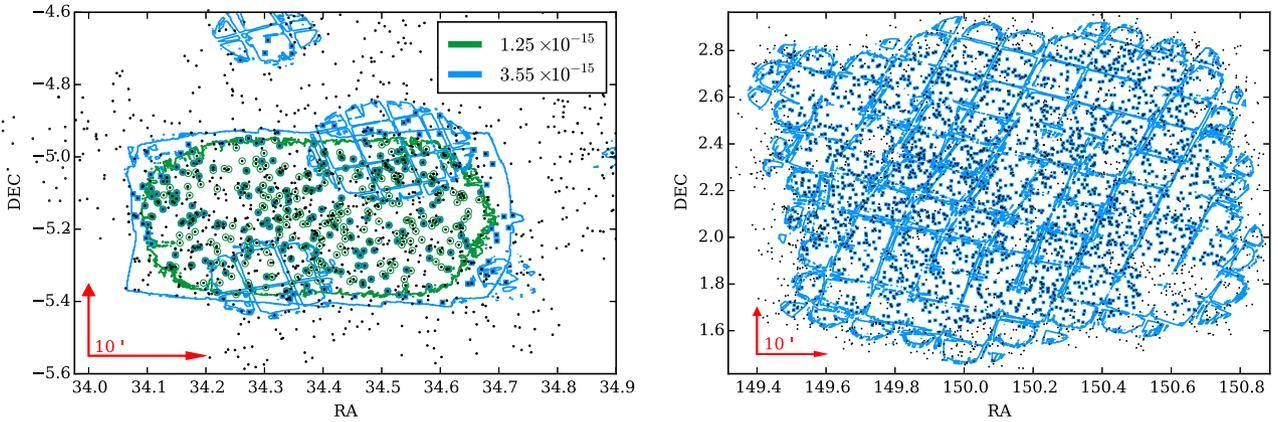
### 3.1 Two-point correlation functions

#### 3.1.1 AGN correlation functions

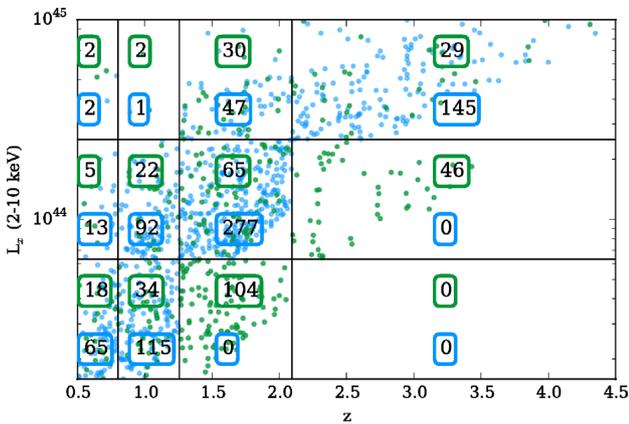
The most commonly used statistical estimator of galaxy clustering uses the two-point ACF. The angular ACF,  $w(\theta)$ , measures the excess probability, above a random distribution, of finding a galaxy at an angular separation  $\theta$  from another galaxy. We use the Landy & Szalay (1993) estimator, described by

$$w(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}, \quad (1)$$

where  $DD(\theta)$ ,  $DR(\theta)$ , and  $RR(\theta)$  are the galaxy–galaxy, galaxy–random, and random–random normalized pair counts, respectively.



**Figure 2.** Hard band X-ray AGN in UDS Chandra and XMM (left) and COSMOS (right). All X-ray sources with optical/near-infrared counterparts are shown as black points, AGN with a flux limit of  $\geq 1.25 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$  and  $\geq 3.55 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$  are highlighted in green circles and blue squares, respectively. The contours corresponding to these flux limits are also shown in the respective colours. Note that the larger size of the COSMOS field may give an appearance of a higher density of sources, but this does not reflect reality. For reference, the red arrows in the bottom left corner of the two plots show a scale of 10 arcmin.



**Figure 3.** The number of AGN in each luminosity and redshift bin in the UDS (green boxes) and COSMOS (blue boxes) fields. We apply a flux limit computed with the lower end of the luminosity bin and the median redshift of the AGN within the bin. The diagonal line traced out by missing points in the bottom right hand corners of the boxes correspond to the sources that have been removed as their fluxes were lower than the applied flux limit.

We choose this estimator because it is more robust to effects that can affect clustering measurements, such as the size of the random catalogue and edge corrections (Coil 2013).

While the clustering of AGN and quasars has been studied using the ACF, this method requires large sample sizes to provide tight constraints on AGN host halo masses. As our AGN sample sizes are limited, we can make use of a close cousin of the ACF; the two-point CCF. We measure the CCF of the AGN with respect to  $K$  band selected galaxy samples provided by the UDS and COSMOS surveys.

We cross-correlate our AGN ( $D_A$ ) with a full volume-limited 90 per cent mass-complete tracer galaxy population ( $D_G$ ), using

$$w(\theta) = \frac{D_A D_G(\theta) - D_A R_G(\theta) - D_G R_A(\theta) + R_A R_G(\theta)}{R_A R_G(\theta)}, \quad (2)$$

where each term is normalized by the total pair counts.  $R_A$  and  $R_G$  denote the random source catalogues that populate the regions from which we select AGN (see Fig. 2) and tracer galaxies, respectively. For the tracer galaxy catalogues, we populate random catalogues

corresponding to the ‘good’ regions of UDS and COSMOS with artefacts such as stars and cross-talk masked out. For our AGN catalogues in UDS and COSMOS, we mask for the bad regions of the relevant optical/near-infrared images, as well as the regions of the X-ray image for which the flux limit is shallower than  $3.55 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$ . For the study of luminosity-dependent clustering, we compute different random catalogues for each of the flux limits derived from the different luminosity and redshift bins. These random catalogues map out the same regions as each of our galaxy catalogues, and the total number of randoms is 50 000 and 100 000 in the UDS and COSMOS fields, respectively. The ratio of randoms to data is roughly consistent for each redshift bin (of order 300:1).

In order to ensure that our clustering measurements are reliable, we impose a lower limit of 30 AGN in a given sample to qualify for our analysis.

As the clustering measured using the CCF is underestimated due to the limited observed field size, we apply a correction factor for the integral constraint  $C$  following Roche & Eales (1999),

$$C_{\text{CCF}} = \frac{\sum R_A R_G(\theta) w(\theta)}{\sum R_A R_G(\theta)}. \quad (3)$$

The integral constraint for the ACF of our tracer sample is similarly defined as

$$C_{\text{ACF}} = \frac{\sum R_G R_G(\theta) w(\theta)}{\sum R_G R_G(\theta)}. \quad (4)$$

We assume that  $w(\theta)$  of AGN traces the angular correlation function of the underlying dark matter distribution, following the method of Hartley et al. (2013) and Wilkinson et al. (2017).

### 3.1.2 Dark matter correlation functions

In order to interpret our angular correlation functions, we compute the angular correlation function of the matter distribution (dominated by dark matter) over the same volume as our tracer sample. We do this following the formalism of Smith et al. (2003) to compute the non-linear dark matter power spectrum at the mean redshift of the sample. This is then Fourier transformed to obtain the 3D non-linear dark matter angular correlation function; i.e. the sum of the correlation functions in the 1-halo regime (pair counts within the

same halo) and the 2-halo regime (pair counts in different haloes). Finally, we project this on to 2D using the redshift distribution of the sample and the relativistic Limber equation (Limber 1954), following (Peebles 1980; Magliocchetti & Maddox 1999). We refer the reader to section 4 of Magliocchetti & Maddox (1999) for a clear account of this implementation. To determine the redshift distribution, we use a top hat redshift probability function for both our AGN and tracer samples, since the AGN redshift probability distributions may not be as reliable as those of the tracer sample. The uncertainty on these redshift distributions is the main limitation in projecting using the Limber equation. In addition, the slope of the real-space correlation function must be known, and a cosmology must be assumed. Although the projection depends on the assumed cosmological parameters, this is a systematic effect such that the comparison of clustering between different populations in the same cosmology is unaffected.

This routine therefore allows us to obtain synthetic  $w_{\text{dm}}(\theta)$  from the halo model that can be fit to the observed AGN correlation functions.

### 3.2 Bias fitting and halo masses

Galaxies are biased tracers of the underlying dark matter density field. The linear galaxy bias parameter  $b$ , defined as the ratio of the overdensity of galaxies to the ( $\delta_g$ ) to the mean overdensity of matter

$$b = \delta_g / \delta, \quad (5)$$

where  $\delta$  is the overdensity, given by  $\delta = \rho / \bar{\rho} - 1$  where  $\bar{\rho}$  is in turn the mean mass density (e.g. Peebles 1980). Therefore galaxies with higher bias have a higher degree of clustering and are more likely to be found near the highest density peaks in the dark matter mass distribution.

We therefore use the bias  $b$  to indicate how strongly clustered our AGN sample is with respect to the underlying dark matter distribution. On linear scales, we can compute this parameter as the square root of the measured observed galaxy correlation function divided by the 2D dark matter correlation function following the definition

$$w_{\text{obs}}(\theta) = b^2 \times w_{\text{dm}}(\theta), \quad (6)$$

where  $w_{\text{obs}}$  denotes the observed AGN CCF and  $w_{\text{dm}}$  denotes the projected correlation function of the dark matter distribution (e.g. Benson et al. 2000).

We therefore fit the dark matter correlation functions multiplied by  $b^2$  to the observed AGN correlation function, and calculate the optimum value of  $b$  using  $\chi^2$  minimization with weights corresponding to the inverse of the uncertainties on the observed correlation function. These uncertainties are calculated using the bootstrap method with 50 repetitions. We therefore resample the tracer/target sample with replacement 50 times, and evaluate  $w(\theta)$  for each of the 50 bootstrap samples. The uncertainties are then given by the standard deviation of resampled values of  $w(\theta)$ . To obtain a combined bias measurement using two separate fields (UDS and COSMOS), we obtain individual  $\chi^2$  measurements of the two fields (using two independent AGN and dark matter correlation functions) and minimize the total  $\chi^2$ , taking into account the correlated uncertainties.

We assume that both AGN and tracer galaxy populations trace the dark matter distribution, and that both populations are linearly biased. This implies that we expect the correlation function to be well described by linear gravity theory ( $w_{\text{linear}}$ ). This assumption

holds true at large scales where both the linear ( $w_{\text{linear}}$ ) and non-linear ( $w_{\text{non-linear}}$ ) correlation functions are mutually consistent, since the non-linear terms are negligible. However, the two CCFs deviate at small scales, (Zehavi et al. 2005), so it is no longer appropriate to assume that the galaxy population traces the dark matter distribution. We therefore adopt a lower limit to  $\theta$  corresponding to  $w_{\text{non-linear}} = 3 \times w_{\text{linear}}$  between the linear and non-linear regimes in order to accurately constrain our bias measurements, following Wilkinson et al. (2017). The angular scale that  $3 \times w_{\text{linear}}$  corresponds to varies with redshift. In Fig. 4, this scale is shown as the minimum scale for which the solid lines are drawn. We choose this limit as a trade-off between minimizing the non-linear effects on small scales as well as maximizing the scales over which we can use the correlation functions. We also employ an upper limit and do not consider pair counts at scales larger than  $\theta = 0.4$  deg as the finite field of view results in unreliable  $w(\theta)$  measurements at large  $\theta$ .

We thus determine the absolute bias of the AGN cross-correlated with tracer galaxies (CCF bias,  $b_{\text{AG}}$ ) and of the tracer galaxy autocorrelation function (ACF bias,  $b_{\text{G}}$ ) with respect to the dark matter. These two absolute bias measurements can then be used to infer the relative bias of the target AGN sample  $b_{\text{A}}$  following  $b_{\text{A}} = b_{\text{AG}}^2 / b_{\text{G}}$ .

The ACF of the target AGN sample can be inferred by multiplying the CCFs by ( $b_{\text{AG}}^2 / b_{\text{G}}^2$ ). These inferred ACFs of UDS and COSMOS X-ray AGN are plotted in Fig. 4 along with their fitted bias.

Finally, we obtain dark matter halo mass estimates using the formalism of Mo & White (2002). The bias of dark matter haloes is dependent on its mass and given epoch. We thus assign dark matter halo masses to our AGN samples by matching our bias measurements at a given redshift to the bias of dark matter haloes of various masses.

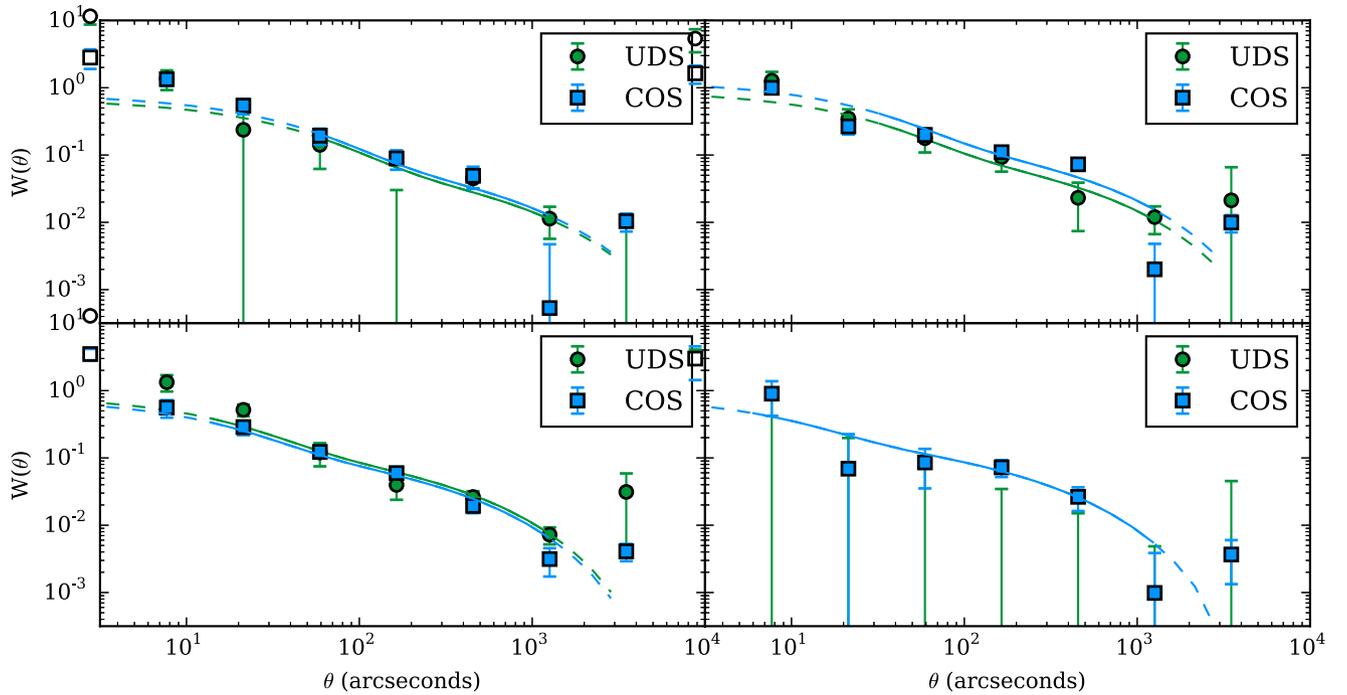
## 4 RESULTS

### 4.1 Interpreting the bias of X-ray AGN

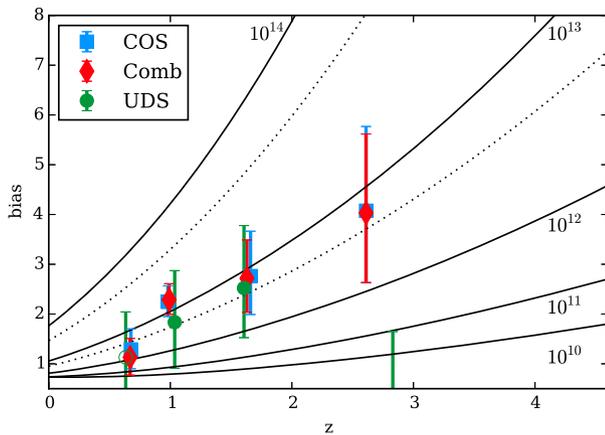
We calculate CCFs in the UDS and COSMOS fields to obtain measurements of the bias and infer the dark matter halo masses of galaxies hosting X-ray AGN as a function of redshift. We plot our measurements in Fig. 5, where the individual COSMOS and UDS measurements are presented, in addition to the combined estimate of the bias. As shown by Fig. 5, the clustering of X-ray AGN suggests that they preferentially reside in ‘group-like’ environments of  $10^{12-13} M_{\odot}$  irrespective of redshift, in agreement with previous studies (Croon et al. 2005; Ross et al. 2009; Shen et al. 2009; Allevato et al. 2011; Krumpke et al. 2012; Allevato et al. 2016; Powell et al. 2018).

In Fig. 6, we compare the bias of AGN to that of star-forming and passive galaxy populations in the UDS and COSMOS fields, selected using the supercolour technique (see Section 2.1.4 and Wilkinson et al. in preparation for more details), matched in mass and redshift distributions to the AGN population.<sup>1</sup> As this technique is only reliable out to  $z \sim 2.5$ , we truncate our final redshift bin to  $z = 2.5$ . We find that AGN are more clustered than star-forming

<sup>1</sup>We note that Wilkinson et al. (in preparation) use the full redshift probability distribution to project the 3D dark matter correlation function, whereas we use a top hat redshift probability distribution for consistency with the method we use for our AGN sample.

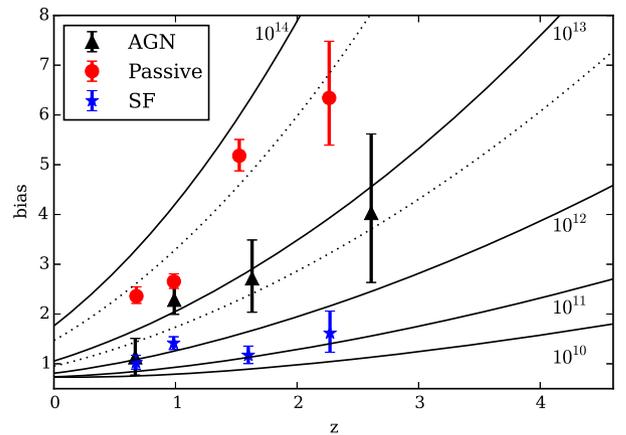


**Figure 4.** Inferred autocorrelation functions of X-ray AGN in the UDS (green) and COSMOS (light blue), as a function of redshift. We present the bias measurement with a flux limit optimized separately for each field. Uncertainties are derived from the standard deviation in resampled  $w(\theta)$ . We note that low number statistics can formally lead to negative  $w(\theta)$  in certain bins, resulting in an underestimation of the bias. The combined measurements in these cases will not be affected by this problem as these points will have larger uncertainties and hence lower weights in the  $\chi^2$  minimization. The dashed lines represent the fits of the dark matter correlation functions to the observed correlation functions (scaled by the square of the bias), and the overplotted solid lines highlight the scales over which the observed correlation function is used in the fitting routine.



**Figure 5.** Redshift evolution of AGN bias in the UDS (green) and COSMOS (light blue). We present the bias measurement with a flux limit chosen to maximize the number of AGN in each field separately, as well as the combined measurement in red optimized for both fields. If fewer than 50 AGN are present in the sample, we plot them in open symbols as these points are potentially less reliable. The redshift evolution of dark matter haloes corresponding to given masses are shown by the solid black lines, and are annotated by their corresponding halo masses in solar mass units. The dotted black lines show the evolution of  $5 \times 10^{12}$  and  $5 \times 10^{13} M_{\odot}$ . On average, therefore, AGN appear to inhabit  $10^{12-13} M_{\odot}$  haloes with no evolution in redshift.

galaxies, but less clustered than passive galaxies of the same mass. As it is well known that AGN are a composite population of star-forming and passive galaxies (e.g. Aird, Coil & Georgakakis 2017),



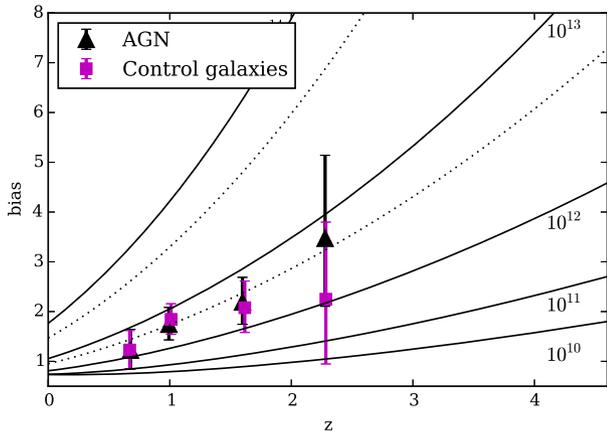
**Figure 6.** Redshift evolution of the bias of AGN (black), along with SF (blue) and passive (red) galaxies, as defined by the supercolour technique, matched in mass and redshift. X-ray AGN inhabit intermediate halo masses relative to star-forming and passive galaxies of the same mass distribution. The solid lines denote the evolution of dark matter haloes, as in Fig. 5.

the clustering signal may indicate this is a mixed population, or possibly a population transitioning from star forming to passive.

To explore this issue, we first classify our AGN host galaxies (for which supercolour classifications are available) into passive and star forming. Properties of these AGN, such as the median X-ray luminosity and passive host fraction, are summarized in Table 2. We compare the clustering of these AGN to the clustering of an analogous mixed population of star-forming and passive inactive galaxies. We ensure that these inactive galaxies are matched in

**Table 2.** Properties of AGN samples (for which supercolour classifications are available) in each redshift bin.

Redshift range	Number of AGN	Median redshift	Fraction with photo-z only	Median stellar mass ( $\log \frac{M_*}{M_\odot}$ )	Passive host fraction	Median X-ray luminosity ( $\log \frac{L_X}{\text{erg s}^{-1}}$ )
$0.5 < z < 0.8$	194	0.677	0.13	10.8	0.19	43.2
$0.8 < z < 1.3$	369	0.995	0.21	10.9	0.11	43.6
$1.3 < z < 2.1$	516	1.599	0.47	10.9	0.09	44.1
$2.1 < z < 2.5$	129	2.278	0.56	11.0	0.06	44.4



**Figure 7.** Redshift evolution of bias of X-ray AGN (for which supercolour classifications are available) in black triangles, compared to a control sample of inactive galaxies (in magenta squares), matched in stellar mass, spectral class (from supercolours), and redshift. The star formation characteristics (i.e. spectral class) of star-forming galaxies in the non-AGN sample matches that of Fig. 6. The solid lines denote the evolution of dark matter haloes, as in Fig. 5. The clustering of AGN is entirely consistent with the clustering of inactive galaxies with similar host galaxy properties.

terms of redshift, spectral class as determined by their supercolours (SC1, SC2, and SC3), and stellar mass. To construct this matched control sample, for each AGN, we identify the closest matching galaxy in a three-dimensional space of stellar mass, spectral class, and redshift (resulting in a sample of the same size as AGN). The precise matching of control galaxies to AGN results in identical distributions of mass, spectral class, and redshift, all consistent with being drawn from the same underlying distribution ( $p = 0.999\,999\,98$ , as probed by a Kolmogorov–Smirnov test). We calculate the bias of this mixed population using the same method applied to the AGN (i.e. by inferring ACF from CCF) and present the results in Fig. 7, showing that the AGN bias measurements in all redshift bins are entirely consistent with inactive galaxies of similar mass and spectral class.

The clustering of the non-AGN control galaxy population is intermediate between passive and star-forming galaxies (see Fig. 6), as expected for a mixed sample, with an average clustering signal consistent with ‘group’ mass haloes. It is therefore possible that the intermediate clustering signal of AGN (corresponding to a halo mass of  $10^{12-13} M_\odot$ ) is produced by an averaging of the clustering signal from the mixed host galaxy population, and that the clustering of AGN predominantly reflects the mix of passive and star-forming host galaxies that occupy a range of different environments.

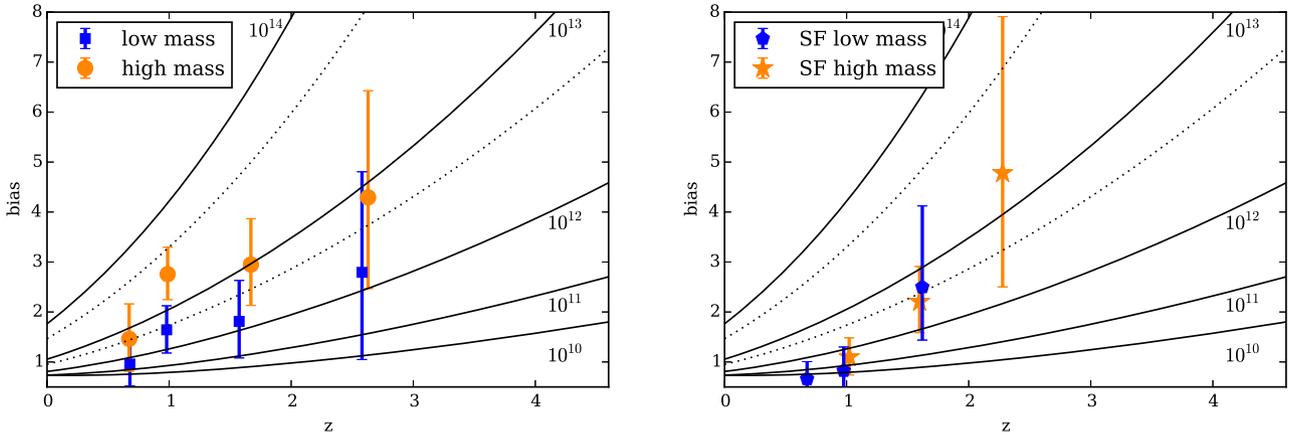
#### 4.2 The role of host galaxy properties on AGN clustering

We now investigate which properties of AGN host galaxies have the dominant influence on the AGN clustering signal. To disentangle the

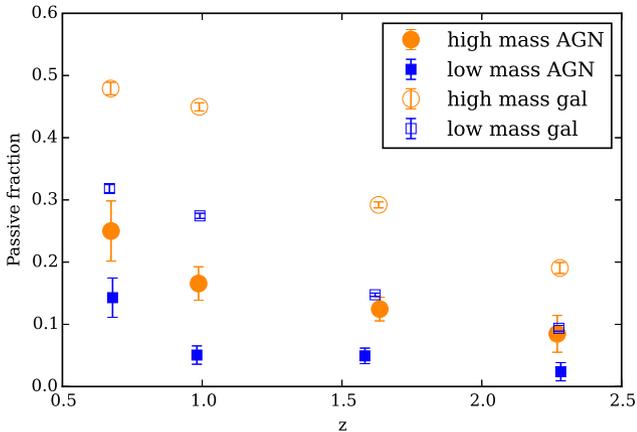
effects of stellar mass and host galaxy spectral class on clustering, we split our AGN into high and low host galaxy stellar mass subsamples around the median mass ( $10^{10.75} M_\odot$ ). In Fig. 8(a), we plot our clustering measurements of the AGN and find that AGN in high-mass galaxies are more strongly clustered and appear to reside in haloes that are almost an order of magnitude (0.9 dex) more massive than AGN in low-mass hosts out to  $z \sim 2$  (at  $1.7\sigma$ ). This is consistent with expectations from the galaxy stellar mass–halo mass relation (e.g. Li et al. 2006; Meneux et al. 2008; Wake et al. 2011; Leauthaud et al. 2012; Marulli et al. 2013; Conselice et al. 2018).

We next consider how host galaxy spectral class affects the clustering strength. To do this, we define the passive fraction as the fraction of a given sample that are classified as either post-starburst or red sequence (based on supercolours). In Fig. 9, we plot the passive fractions of AGN and non-AGN galaxies, within our two mass bins. We only select galaxies above the highest redshift 90 per cent completeness limit of  $10^{10.2} M_\odot$  to construct the low-mass galaxy sample. This plot shows that (a) the AGN detected have significantly lower passive fractions than galaxies of the same mass, and (b) the AGN in high-mass hosts, like high-mass galaxies, have higher passive fractions than their low-mass counterparts. Since passive hosts are more clustered than star-forming hosts (see Fig. 6, Hartley et al. 2013; Coil et al. 2017), the difference in the measured bias between AGN in low- and high-mass hosts may be due to different host spectral class. There appears to be a general correlation between the difference between the passive fractions and clustering signal of AGN in high- and low-mass bins (e.g. the most significant difference between the passive fractions of AGN in low- and high-mass galaxies is at  $z \sim 1$ , as is the most significant difference in bias of the two samples).

To explore this further, we would ideally measure the clustering of star forming and passive AGN separately. Although we lack the sample sizes to explore the clustering of AGN hosted by passive galaxies, we measure the clustering of AGN in star-forming galaxies split around the same median mass, and plot our result in Fig. 8(b). We note that the median stellar masses in each mass/redshift bin between the two panels are consistent within errorbars ( $\Delta \log M \leq 0.04$ ). We find tentative evidence for a shift in the environments of AGN in star-forming host galaxies from higher halo masses at high redshift to field at low redshift. This is consistent with the effect of galaxy downsizing, where star-formation activity shifts from high-mass haloes to low-mass haloes as the Universe ages (Wilkinson et al. 2017). Although the increasing X-ray luminosity limit with redshift implies that we are only sensitive to higher mass black holes at higher redshifts at a given Eddington ratio, we do not expect this to affect our results significantly since we see no correlation between X-ray luminosity and halo mass (see Section 4.3). Fig. 8(b) also shows that we find no difference between the dark matter halo masses of AGN in star-forming host galaxies in low- and high-mass bins, so the excess bias in Fig. 8(a) is likely driven by the higher passive fractions in



**Figure 8.** **Left:** (a) Redshift evolution of X-ray AGN bias in low ( $M_* < 10^{10.75} M_\odot$ ) and high ( $M_* > 10^{10.75} M_\odot$ ) mass host galaxies (blue squares and orange circles, respectively). The solid lines denote the evolution of dark matter haloes, as in Fig. 5. AGN in more massive hosts appear to reside in more massive haloes. **Right:** (b) Bias of X-ray AGN in low ( $M_* < 10^{10.75} M_\odot$ ) and high ( $M_* > 10^{10.75} M_\odot$ ) mass star-forming hosts (blue pentagons and orange stars, respectively). There is now no significant difference between the bias measurements of AGN in low- and high-mass galaxies, indicating that star-forming activity is a more important driver of clustering than stellar mass.



**Figure 9.** Passive fractions of AGN and galaxies split by stellar mass. AGN have significantly lower passive fractions than galaxies of the same mass. AGN in high-mass galaxies ( $M_* > 10^{10.75} M_\odot$ ; filled orange circles) have higher passive fractions than those in low-mass galaxies ( $M_* < 10^{10.75} M_\odot$ ; filled blue squares). In general, the high-mass galaxy population (open orange circles) also has higher passive fractions than the low-mass counterparts (open blue squares). There is a lower passive fraction among AGN hosts than the general galaxy population, so star-forming galaxies are more likely to host AGN than passive galaxies.

the high-mass sample. We therefore find evidence that stellar mass is not the fundamental parameter that drives the excess clustering of AGN in higher mass host galaxies relative to AGN in lower mass host galaxies. Similar results have been obtained for the galaxy population, such as Coil et al. (2017), who find that galaxy clustering does not significantly depend on stellar mass at a given sSFR. This may imply that the stellar mass–halo mass relation is driven by the correlation between passive fraction and stellar mass.

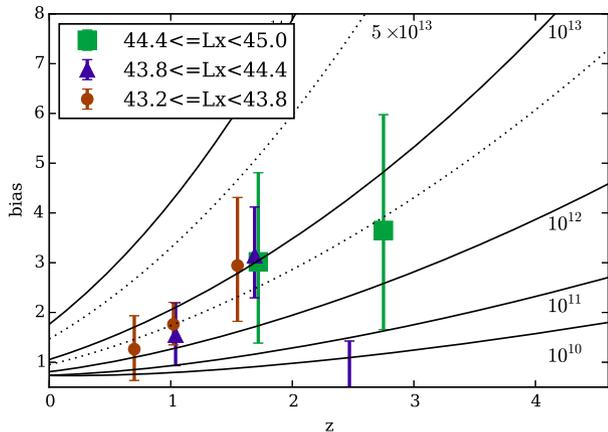
We caution that this analysis could be affected by the possibility that blue light from the AGN contaminates the SEDs and leads to an incorrect star-forming assignment of a passive galaxy. However, this effect is not expected to be significant as supercolour classifications are derived based on filters focused on the rest frame 4000 Å break region. AGN light may also contaminate the mass measurements,

but our AGN have fairly low X-ray luminosities so we do not expect this effect to have a significant impact on our conclusions (Almaini et al. in preparation). In addition, Kocevski et al. (2017) report that colour contamination by lower luminosity AGN in their study is negligible, and Santini et al. (2012) find that only 1.3 per cent of their lower luminosity sources had a difference in their stellar mass larger than a factor of two when an AGN component is added to the stellar template. We have tested the results in Fig. 7 by removing the most luminous AGN ( $L_X > 10^{44.4}$ ) from our sample, creating a new control sample, and repeating the clustering analysis. We find that the results are consistent within errorbars, and thus conclude that any contaminating AGN light does not have a major impact on our results.

### 4.3 Links between clustering and AGN luminosity

In this section, we investigate whether the clustering of AGN is dependent on the power of the AGN determined through the proxy of X-ray luminosity. We study the correlation between the power of the black hole and the inferred dark matter halo mass by splitting our AGN into low, medium, and high X-ray luminosity bins, corresponding to  $10^{43.2} \leq L_x < 10^{43.8}$ ,  $10^{43.8} \leq L_x < 10^{44.4}$ , and  $10^{44.4} \leq L_x < 10^{45.0}$  erg s $^{-1}$ . We cross-check passive fractions and mass distributions between the different luminosity bins and find no significant differences between the different luminosity populations. Passive fractions vary between 5 per cent and 10 per cent at all redshifts and luminosities.

While previous studies across a wide range of AGN luminosities have found similar ‘group-like’ halo masses, we are able to explore this in a robust manner by splitting our AGN sample by X-ray luminosity and redshift to obtain self-consistent estimates of the bias. In Fig. 10, we show that there is no correlation between the power of the black hole on the clustering of the AGN, which is consistent with previous results (e.g. Magliocchetti et al. 2017). Since low, medium, and high X-ray luminosity AGN occupy similar mass haloes, this implies that there may not be an environmental influence on the accretion rate of gas into the central black hole. This also implies that the environments of AGN of all luminosities are driven by the mixed population of their hosts.



**Figure 10.** Bias measurements of AGN with hard band (2–10 keV) X-ray luminosities measured in  $\text{erg s}^{-1}$  in bins of  $10^{43.2} \leq L_x < 10^{43.8}$ ,  $10^{43.8} \leq L_x < 10^{44.4}$ , and  $10^{44.4} \leq L_x < 10^{45.0}$  in red circles, purple triangles, and green squares, respectively. There is no correlation between redshift or luminosity and dark matter halo mass. The outlier point at  $2.1 < z < 4.5$  is likely due to low number statistics, since it contains only 46 AGN (the only bin with  $< 50$  AGN). The solid lines denote the evolution of dark matter haloes, as in Fig. 5.

We also investigated the effects of other AGN properties such as Eddington ratio and hardness ratio, but did not have a large enough AGN sample to obtain robust results.

## 5 DISCUSSION

We have investigated the clustering of AGN with X-ray luminosities between  $43.2 < \log L_x < 45.0$  and redshifts between  $0.5 < z < 4.5$ , and measured a bias corresponding to dark matter haloes of mass  $10^{12-13} M_\odot$ . Similar results have previously been interpreted to indicate a preference for AGN to reside in group-like environments (e.g. Ikeda et al. 2015), we have shown (see Fig. 7) that the same clustering signal can be obtained from a mix of non-AGN star-forming and passive galaxies that populate a range of halo masses. Based on this evidence, we suggest that AGN do not preferentially reside within a particular halo mass, and infer that AGN triggering is not primarily driven by the large-scale environment.

We find corroborating evidence when we divide the AGN by host galaxy property. The clustering of AGN hosted by star-forming galaxies (the dominant host type) have a bias corresponding to only  $10^{11-12} M_\odot$  haloes at  $z < 1.5$ , comparable to the bias obtained for non-AGN star-forming galaxies of similar mass and redshift. Although we do not have a sufficiently large sample to measure the clustering of AGN in passive host galaxies, we note that the clustering of this minor component of the AGN population must be significantly stronger than the AGN in star-forming hosts, because the clustering strength of the combined population average to a halo mass of  $10^{12-13} M_\odot$ . Thus AGN do not preferentially reside within haloes of a certain mass.

Our interpretation of the clustering signal is supported by other observational evidence, such as the lack of enhanced AGN fractions in group-like environments (e.g. Shen et al. 2007; Arnold et al. 2009; Oh et al. 2014; Tzanavaris et al. 2014). Furthermore, if AGN preferentially reside in groups, we would expect the AGN host galaxy properties to vary with redshift since the fraction of star-forming group galaxies increases with redshift (Butcher & Oemler 1978; Giodini et al. 2012; Popesso et al. 2012). Instead we find that the AGN host spectral class does not vary significantly from  $z =$

2.5 to  $z = 0.5$ . We therefore suggest that large-scale environment (e.g. halo mass) is not the dominant factor in triggering an AGN, although it may play a minor role. We also note that the richest and rarest cluster environments (of halo mass  $\sim 10^{15} M_\odot$ ) are not probed by the UDS/COSMOS fields, so we are unable to determine if AGN triggering is enhanced or suppressed in this specific regime.

Models of galaxy evolution, on the other hand, point to an important link between large-scale environment and triggering AGN activity. For instance, less massive groups have been proposed as the ideal environment for AGN activity due to an increased likelihood of mergers (Hopkins et al. 2008a, b). Our finding is in tension with the expectation that AGN should be preferentially triggered in groups if mergers trigger AGN activity. However, we also note that observational evidence that mergers are linked to AGN triggering remains mixed (Ellison et al. 2013; Kocevski et al. 2015; Hewlett et al. 2017; Villforth et al. 2017).

Whilst AGN do not typically reside in a special environment, we find that AGN populate special host galaxies. We have shown that the AGN hosts are not a random subset of the underlying galaxy population within the UDS and COSMOS surveys. Instead, the passive fraction of AGN is lower than the underlying galaxy population at the same stellar mass (Fig. 9), implying that AGN are preferentially hosted by star-forming galaxies, consistent with the findings of Aird et al. (2017). We infer, therefore, that the probability of triggering an AGN is correlated with some properties of the host galaxy. Consistent with this inference, Kocevski et al. (2017) find evidence for a relationship between compactness and AGN triggering. We conclude that the triggering of AGN is likely not a simple single process, but is in fact a complex set (or sets) of conditions.

The set of conditions that trigger AGN are not likely to change drastically across  $z = 2.5$  to  $z = 0.5$  because we find that the stellar mass distribution and class of AGN host galaxies remain approximately constant in this redshift range. On the other hand, the Universe evolves drastically across this period, resulting in a significant change in galaxy properties: star formation declines (Madau & Dickinson 2014), galaxies grow in mass, and the morphological mix of galaxies evolves from predominantly irregulars to spirals and a larger fraction of early types (Mortlock et al. 2013). The passive fraction of AGN and galaxies both increase by a factor of 2–3 from  $z = 2.5$  to  $z = 0.5$  (see Fig. 9). This indicates that the fractions of star-forming and passive galaxies that host AGN do not change significantly with redshift (although we note that Aird et al. 2017 find evidence that quiescent galaxies are more likely to host AGN at higher redshifts).

The lack of dependence of the clustering signal on AGN power, as probed by the X-ray luminosity (Fig. 10), suggests that the accretion rate of AGN does not have a simple dependence or correlation with its large-scale environment. This is in agreement with recent results from Yang et al. (2018). Furthermore, since we have shown that the clustering signal of AGN is primarily driven by the clustering properties of the AGN host population, and the mixture of host spectral classes, we infer indirectly that there is also no strong link between host spectral class and the AGN instantaneous accretion rate.

A correlation between host spectral class and AGN accretion rate is expected because of the evidence for a correlation between the growth of stellar mass and supermassive black holes (Boyle et al. 1998; Franceschini et al. 1999; Silverman et al. 2008). Given that the growth of stellar and black hole mass both rely on the availability of gas, we might expect to find higher accretion rates (through the proxy of X-ray luminosity) for star-forming galaxies

compared to passive galaxies of the same stellar mass. Since we find that the clustering strength of an AGN is primarily driven by its host spectral class, we would therefore expect the clustering signal of more luminous AGN to be lower than less luminous AGN, in discord with our findings.

We can explain the lack of variation in the clustering strengths for high- and low-luminosity AGN with the same phenomenon behind the flat SFR– $L_X$  relationship for AGN found by e.g. Stanley et al. (2015). Hickox et al. (2014) proposed that large X-ray AGN variability on short time-scales (relative to that of star formation) dilutes the intrinsic correlation between SFR and  $L_X$ . To reproduce the underlying relation we must average over the most variable quantity ( $L_X$  in this case). High X-ray variability on short time-scales could also dilute any intrinsic  $L_X$ –host spectral class correlation, which would then dilute any variation in the clustering of AGN of different power. We note that a similar result may also be obtained if star formation and black hole accretion are not coeval (Wild, Heckman & Charlton 2010).

Drawing our interpretations together, the triggering of AGN activity likely depends on a complex set of conditions. We find a lack of correlation between AGN power and clustering signal, and that a mix of non-AGN star-forming and passive galaxies can reproduce the same clustering signal as AGN. Together this suggests that large-scale environment plays at most a minor role in triggering AGN activity. Availability of fuel may be inferred to play a major role since star-forming galaxies are more likely to host AGN at all epochs. This is supported by the well-known increased prevalence of high-luminosity AGN at high redshift compared to low redshift (e.g. Fig. 10). However, environment may be a relevant factor if mergers only trigger the most luminous AGN, as merger rates decline as the Universe ages (Conselice, Rajgor & Myers 2008). Furthermore, discs and bars may trigger nuclear activity (e.g. Knapen, Shlosman & Peletier 2000) as disc instabilities have been proposed to enhance gas flow to the nuclei of galaxies (Dekel, Sari & Ceverino 2009). Therefore, multiple mechanisms could be at play in the triggering of AGN, unlikely to be defined by solely large-scale environment or availability of fuel.

We note that our work is limited to X-ray selection, which tends to select more powerful and rapidly accreting AGN. Selection effects may play an important role in the interpretations and conclusions from AGN clustering results. Previous studies have found that X-ray and radio AGN are more clustered than mid-IR-selected AGN (e.g. Hickox et al. 2009; Mendez et al. 2016). However, when each AGN sample was studied using galaxy samples matched in stellar mass, star formation rate, and redshift, Mendez et al. (2016) find no significant differences between the clustering properties of the AGN samples. They find that AGN selected in different wavelengths appear to have different clustering properties simply because they are sampling different host populations with different stellar mass and SFR distributions. Therefore, the method of AGN selection introduces inherent biases in the host galaxy properties and likely determines the clustering signal. Our study could be repeated using optical, near-infrared, and radio selected AGN with samples matched in mass, (s)SFR, and redshift to obtain a clearer picture of the impact of host galaxies on AGN clustering measurements. The implication for future AGN clustering studies is that samples must be divided by host galaxy properties as the clustering signal from AGN likely represents that of host galaxies.

To summarize our interpretations, the triggering of AGN activity likely depends on a complex set of conditions that do not depend solely on large-scale environment or availability of fuel. The parameters that define the triggering of AGN could have important

implications for our understanding of galaxy evolution, particularly as our most sophisticated galaxy evolution models invoke AGN feedback as a key ingredient in reproducing the stellar mass functions of galaxies.

## 6 CONCLUSIONS

In this work, we study the clustering properties of a flux-limited sample of hard X-ray selected AGN from  $z \sim 4.5$  to  $z \sim 0.5$ , using the COSMOS and UDS multiwavelength surveys. We compare them to a control galaxy sample designed to have similar distributions of stellar mass, spectral class, and redshift. We investigate the role of properties of host galaxies (e.g. stellar mass) and of the central AGN (X-ray luminosity) as a function of redshift. We find that the clustering properties of AGN cannot be naively linked to large-scale environments as host galaxy properties play a significant role in the clustering measurements. To summarize our findings:

(i) We find that hard X-ray selected AGN in the UDS and COSMOS fields have bias parameters that correspond to a typical halo mass of  $10^{12-13} M_\odot$ . No evidence for evolution in halo mass is found between  $0.5 < z < 4.5$ .

(ii) We compare the clustering of AGN to star-forming and passive galaxy populations matched in mass and redshift distributions. We find that the clustering of AGN lies in between the mass-matched star-forming and passive populations.

(iii) We can reproduce the clustering signal of AGN with an inactive galaxy population closely matched in spectral class, mass, and redshift distributions as the AGN host galaxies. We thus find that the mixed population of star forming and passive AGN host galaxies drives the clustering properties of AGN.

(iv) We split AGN by host galaxy stellar mass and find an excess clustering in the high-mass sample. The stellar mass dependence of clustering disappears once passive galaxies are removed from the samples, as we find no difference in their clustering properties. Therefore, we conclude that AGN clustering depends more strongly on the spectral class of the host galaxies than stellar mass.

(v) We find no difference in the clustering properties of low, medium, and high X-ray luminosity AGN. The triggering of AGN activity is likely determined by a complex set of conditions that do not depend solely on large-scale environment.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for their insightful comments, which improved the paper. CK is supported by a Vice-Chancellor’s Scholarship for Research Excellence. We thank Will Hartley for useful discussions. This work is based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme ID 089.A-0126. This work also uses data from ESO telescopes at the Paranal Observatory (programmes 094.A-0410 and 180.A-0776; PI: Almaini). We are grateful to the staff at UKIRT for their tireless efforts in ensuring the success of the UDS project. We also wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has within the indigenous Hawaiian community. We were most fortunate to have the opportunity to conduct observations from this mountain.

## REFERENCES

- Aird J., Coil A. L., Georgakakis A., 2017, *MNRAS*, 465, 3390  
Akiyama M. et al., 2015, *PASJ*, 67, 82

- Allevato V. et al., 2011, *ApJ*, 736, 99
- Allevato V. et al., 2016, *ApJ*, 832, 70
- Almaini O. et al., 2017, *MNRAS*, 472, 1401
- Arnold T. J., Martini P., Mulchaey J. S., Berti A., Jeltema T. E., 2009, *ApJ*, 707, 1691
- Ashby M. L. N. et al., 2013, *ApJ*, 769, 80
- Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, *MNRAS*, 311, 793
- Boyle B. J., Georgantopoulos I., Blair A. J., Stewart G. C., Griffiths R. E., Shanks T., Gunn K. F., Almaini O., 1998, *MNRAS*, 296, 1
- Bradshaw E. J. et al., 2011, *MNRAS*, 415, 2626
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Butcher H., Oemler A. Jr., 1978, *ApJ*, 219, 18
- Chabrier G., 2003, *PASP*, 115, 763
- Civano F. et al., 2012, *ApJS*, 201, 30
- Civano F. et al., 2016, *ApJ*, 819, 62
- Coil A. L., 2013, *The Large-Scale Structure of the Universe*. p. 387, Available at: [https://link.springer.com/referenceworkentry/10.1007%2F978-94-007-5609-0\\_8](https://link.springer.com/referenceworkentry/10.1007%2F978-94-007-5609-0_8)
- Coil A. L., Mendez A. J., Eisenstein D. J., Moustakas J., 2017, *ApJ*, 838, 87
- Conselice C. J., 2014, *ARA&A*, 52, 291
- Conselice C. J., Rajgor S., Myers R., 2008, *MNRAS*, 386, 909
- Conselice C. J., Twite J. W., Palamara D. P., Hartley W., 2018, *ApJ*, 863, 42
- Croom S. M. et al., 2005, *MNRAS*, 356, 415
- De Lucia G., Springel V., White S. D. M., Croton D., Kauffmann G., 2006, *MNRAS*, 366, 499
- Dekel A., Sari R., Ceverino D., 2009, *ApJ*, 703, 785
- Digby-North J. A. et al., 2010, *MNRAS*, 407, 846
- Dressler A., Smail I., Poggianti B. M., Butcher H., Couch W. J., Ellis R. S., Oemler A., Jr., 1999, *ApJS*, 122, 51
- Eftekharzadeh S. et al., 2015, *MNRAS*, 453, 2779
- Ellison S. L., Mendel J. T., Scudder J. M., Patton D. R., Palmer M. J. D., 2013, *MNRAS*, 430, 3128
- Franceschini A., Hasinger G., Miyaji T., Malquori D., 1999, *MNRAS*, 310, L5
- Furusawa H. et al., 2008, *ApJS*, 176, 1
- García-Vergara C., Hennawi J. F., Barrientos L. F., Rix H.-W., 2017, *ApJ*, 848, 7
- Garmire G. P., Bautz M. W., Ford P. G., Nousek J. A., Ricker G. R., Jr., 2003, in Truemper J. E., Tananbaum H. D., eds, *Proc. SPIE Conf. Ser. Vol. 4851, X-Ray and Gamma-Ray Telescopes and Instruments for Astronomy*. SPIE, Bellingham, p. 28
- Giodini S. et al., 2012, *A&A*, 538, A104
- Gisler G. R., 1978, *MNRAS*, 183, 633
- Hartley W. G. et al., 2013, *MNRAS*, 431, 3045
- Hewlett T., Villforth C., Wild V., Mendez-Abreu J., Pawlik M., Rowlands K., 2017, *MNRAS*, 470, 755
- Hickox R. C. et al., 2009, *ApJ*, 696, 891
- Hickox R. C., Mullaney J. R., Alexander D. M., Chen C.-T. J., Civano F. M., Goulding A. D., Hainline K. N., 2014, *ApJ*, 782, 9
- Hinshaw G. et al., 2013, *ApJS*, 208, 19
- Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008a, *ApJS*, 175, 356
- Hopkins P. F., Cox T. J., Kereš D., Hernquist L., 2008b, *ApJS*, 175, 390
- Ikeda H. et al., 2015, *ApJ*, 809, 138
- Jarvis M. J. et al., 2013, *MNRAS*, 428, 1281
- Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, *MNRAS*, 353, 713
- Knapen J. H., Shlosman I., Peletier R. F., 2000, *ApJ*, 529, 93
- Kocevski D. D. et al., 2015, *ApJ*, 814, 104
- Kocevski D. D. et al., 2017, *ApJ*, 846, 112
- Kocevski D. D. et al., 2018, *ApJS*, 236, 48
- Kormendy J., Ho L. C., 2013, *ARA&A*, 51, 511
- Koutoulidis L., Plionis M., Georgantopoulos I., Fanidakis N., 2013, *MNRAS*, 428, 1382
- Krishnan C. et al., 2017, *MNRAS*, 470, 2170
- Krumpe M., Miyaji T., Coil A. L., Aceves H., 2012, *ApJ*, 746, 1
- Laigle C. et al., 2016, *ApJS*, 224, 24
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Lawrence A. et al., 2007, *MNRAS*, 379, 1599
- Leauthaud A. et al., 2012, *ApJ*, 744, 159
- Lehmer B. D. et al., 2009, *ApJ*, 691, 687
- Li C., Kauffmann G., Jing Y. P., White S. D. M., Börner G., Cheng F. Z., 2006, *MNRAS*, 368, 21
- Limber D. N., 1954, *ApJ*, 119, 655
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Magliocchetti M., Maddox S. J., 1999, *MNRAS*, 306, 988
- Magliocchetti M., Popesso P., Brusa M., Salvato M., Laigle C., McCracken H. J., Ilbert O., 2017, *MNRAS*, 464, 3271
- Marchesi S. et al., 2016, *ApJ*, 817, 34
- Marulli F. et al., 2013, *A&A*, 557, A17
- McCracken H. J. et al., 2012, *A&A*, 544, A156
- Mendez A. J. et al., 2016, *ApJ*, 821, 55
- Meneux B. et al., 2008, *A&A*, 478, 299
- Miyaji T., Krumpe M., Coil A. L., Aceves H., 2011, *ApJ*, 726, 83
- Mortlock A. et al., 2013, *MNRAS*, 433, 1185
- Mountrichas G., Georgakakis A., 2012, *MNRAS*, 420, 514
- Mo H. J., White S. D. M., 2002, *MNRAS*, 336, 112
- Oh S. et al., 2014, *ApJ*, 790, 43
- Peebles P. J. E., 1980, *The Large-scale Structure of the Universe*, Princeton University Press, Available at: <https://press.princeton.edu/books/paperback/9780691082400/large-scale-structure-of-the-universe>
- Polletta M. et al., 2007, *ApJ*, 663, 81
- Popesso P. et al., 2012, *A&A*, 537, A58
- Powell M. C. et al., 2018, *ApJ*, 858, 110
- Pozzetti L. et al., 2010, *A&A*, 523, A13
- Roche N., Eales S. A., 1999, *MNRAS*, 307, 703
- Ross N. P. et al., 2009, *ApJ*, 697, 1634
- Santini P. et al., 2012, *A&A*, 540, A109
- Scoville N. et al., 2007, *ApJS*, 172, 1
- Shen Y., Mulchaey J. S., Raychaudhury S., Rasmussen J., Ponman T. J., 2007, *ApJ*, 654, L115
- Shen Y. et al., 2009, *ApJ*, 697, 1656
- Silverman J. D. et al., 2008, *ApJ*, 679, 118
- Simpson C., Westoby P., Arumugam V., Ivison R., Hartley W., Almaini O., 2013, *MNRAS*, 433, 2647
- Smith R. E. et al., 2003, *MNRAS*, 341, 1311
- Stanley F., Harrison C. M., Alexander D. M., Swinbank A. M., Aird J. A., Del Moro A., Hickox R. C., Mullaney J. R., 2015, *MNRAS*, 453, 591
- Sutherland W., Saunders W., 1992, *MNRAS*, 259, 413
- Tzanavaris P. et al., 2014, *ApJS*, 212, 9
- Ueda Y. et al., 2008, *ApJS*, 179, 124
- Villforth C. et al., 2017, *MNRAS*, 466, 812
- Wake D. A. et al., 2011, *ApJ*, 728, 46
- Wild V., Heckman T., Charlot S., 2010, *MNRAS*, 405, 933
- Wild V. et al., 2014, *MNRAS*, 440, 1880
- Wild V., Almaini O., Dunlop J., Simpson C., Rowlands K., Bowler R., Maltby D., McLure R., 2016, *MNRAS*, 463, 832
- Wilkinson A. et al., 2017, *MNRAS*, 464, 1380
- Yang G., Brandt W. N., Darvish B., Chen C.-T. J., Vito F., Alexander D. M., Bauer F. E., Trump J. R., 2018, *MNRAS*, 480, 1022
- Zehavi I. et al., 2005, *ApJ*, 630, 1

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.