







# MIGHTEE: the dark matter haloes, duty cycle, and mechanical feedback from radio-AGN up to $z \sim 2.5$

Joel Hamlett <sup>1</sup>★, Catherine L. Hale <sup>1,2</sup>, Matt J. Jarvis <sup>1,3</sup>, David Alonso <sup>1</sup>, Natalia Stylianou <sup>1</sup> and Imogen H. Whittam <sup>1,3</sup>

<sup>1</sup>*Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK*

<sup>2</sup>*School of Physics and Astronomy, Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, EH9 3HJ Edinburgh, UK*

<sup>3</sup>*Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Road, 7535 Bellville, Cape Town, South Africa*

Accepted 2026 March 6. Received 2026 March 5; in original form 2025 October 24

## ABSTRACT

Radio-AGNs (active galactic nuclei) are observed to be more strongly clustered than non-active galaxies, though it is unclear whether this is simply due to their preference for massive host galaxies, or if they reside in distinct environments beyond this mass dependence. Using data from three fields covered by the MIGHTEE survey, we measure the angular two-point cross-correlation functions with a large, stellar mass-limited population of near-infrared selected galaxies, overcoming limitations of previous single-deep-field studies. By fitting halo occupation distribution models, we infer the galaxy bias parameters,  $b$ , for radio-AGN in three redshift ranges with median redshifts of  $z_{\text{med}} = 0.76^{+0.17}_{-0.28}$ ,  $1.25^{+0.14}_{-0.17}$ , and  $1.75^{+0.44}_{-0.18}$ , finding  $b = 1.94^{+0.07}_{-0.07}$ ,  $2.50^{+0.11}_{-0.18}$ , and  $3.38^{+0.27}_{-0.38}$ , respectively. The typical dark matter halo mass decreases with increasing redshift:  $\log_{10}(\langle M_{\text{h}} \rangle / M_{\odot}) = 13.44^{+0.08}_{-0.08}$ ,  $13.17^{+0.07}_{-0.06}$ , and  $13.03^{+0.09}_{-0.10}$ , which we attribute to the increased abundance of cold gas required to fuel AGN activity at earlier times. The AGN duty cycle is determined to be  $\sim 5$ –9 per cent, and we estimate that the total energy radiated by radio-jets over  $0 < z < 2.5$  is  $\sim 10^{53}$  J per halo, which is sufficient to account for the observed excess heating of gas beyond that of gravitational collapse. Comparing the typical dark matter halo masses to the values obtained for the control sample, we find that the halo masses of radio-AGN are  $1.54^{+0.47}_{-0.33}$ ,  $1.11^{+0.25}_{-0.20}$ , and  $1.82^{+1.04}_{-0.57}$  times greater than those of the stellar mass- and redshift-matched galaxies. This difference could arise because AGN feedback suppresses stellar mass growth while leaving halo mass unchanged, or because radio-AGN preferentially reside in earlier forming haloes which are more strongly clustered.

**Key words:** galaxies: active – galaxies: haloes – dark matter – large-scale structure of Universe – cosmology: observations – radio continuum: galaxies.

## 1 INTRODUCTION

It is well-established that the environment in which a galaxy resides can significantly affect its evolution, with active galactic nuclei (AGNs) playing an important role in this process (R. G. Bower et al. 2006; D. J. Croton et al. 2006; P. F. Hopkins et al. 2006; C. M. Harrison et al. 2018). AGNs are powerful astrophysical phenomena, powered by the accretion of matter on to the supermassive black hole (SMBH) at the centre of a galaxy (P. Padovani et al. 2017). Radio-AGN represent the most massive and energetic AGN in the Universe (P. N. Best et al. 2005) with black hole masses  $\gtrsim 10^8 M_{\odot}$  (M. J. Jarvis & R. J. McLure 2002; R. B. Metcalf & M. Magliocchetti 2006). These objects emit radio waves through synchrotron radiation produced by their jets of relativistic charged particles (R. Antonucci 1993). The energetic input of AGN through their jets has a significant impact on both their host galaxies (e.g. S. Rawlings & M. J. Jarvis 2004; V. Springel,

T. Di Matteo & L. Hernquist 2005; C. Cicone et al. 2014; C. M. Harrison et al. 2014) and their large-scale environment (e.g. M. Gitti, F. Brighenti & B. R. McNamara 2012; R. Gilli et al. 2019; D. Eckert et al. 2021). By injecting energy into the surrounding medium, AGN heat the gas in their environment, preventing it from cooling and collapsing to form stars (see B. R. McNamara & P. E. J. Nulsen 2012 for a review). First introduced to resolve discrepancies between observation and simulations (e.g. R. G. Bower et al. 2006; D. J. Croton et al. 2006; I. G. McCarthy et al. 2010), this process is known as AGN feedback, and is now recognized as a key mechanism in regulating galaxy evolution (A. C. Fabian 2012; M. Hardcastle & J. Croston 2020; M. Gaspari, F. Tombesi & M. Cappi 2020). Investigating the environments of AGN helps to determine whether certain environmental conditions trigger or support AGN activity, and provides an insight into how AGN feedback influences their surroundings.

Because radio signals are unattenuated by dust, it is possible to conduct studies on radio-AGN out to high redshifts ( $z \sim 7$ ; e.g. R. Endsley et al. 2022). However, such investigations are often limited by the depths of optical follow-up observations which are

\* E-mail: [joel.hamlett@physics.ox.ac.uk](mailto:joel.hamlett@physics.ox.ac.uk)

necessary to obtain redshifts for these sources (e.g. M. J. Jarvis et al. 2009; K. McAlpine, M. J. Jarvis & D. G. Bonfield 2013; V. Smolčić et al. 2017). Earlier works on the environments of radio sources cross-matched with optical galaxies from wide-area surveys, allowing the clustering of AGN to be measured out to  $z \sim 0.5$  (M. Magliocchetti et al. 2004; K. Brand et al. 2005; D. A. Wake et al. 2008; E. Donoso et al. 2010; S. Fine et al. 2011). More recent efforts have extended clustering analyses to higher redshifts using radio surveys that focus on smaller fields with deeper optical survey coverage (R. C. Hickox et al. 2009; S. N. Lindsay, M. J. Jarvis & K. McAlpine 2014; M. Magliocchetti et al. 2017; C. L. Hale et al. 2018; A. Chakraborty et al. 2020; A. Mazumder, A. Chakraborty & A. Datta 2022).

Radio-AGN are predominantly hosted by the most massive galaxies, both locally (e.g. S. Eales et al. 1997; M. J. Jarvis et al. 2001a; P. N. Best et al. 2005; T. Mauch & E. M. Sadler 2007; J. Sabater et al. 2019; A. Capetti et al. 2022) and at higher redshifts (e.g. V. Smolčić et al. 2009; G. Gürkan, M. J. Hardcastle & M. J. Jarvis 2014; M. Magliocchetti et al. 2016; H. Uchiyama et al. 2022; P. N. Best et al. 2023). The likelihood of a galaxy hosting an AGN is strongly correlated with its stellar mass (P. N. Best et al. 2005; T. Mauch & E. M. Sadler 2007; J. Sabater et al. 2019; M. Magliocchetti et al. 2020; R. Kondapally et al. 2025), with nearly all of the most massive local elliptical galaxies hosting an AGN (M. J. I. Brown et al. 2011; A. Capetti et al. 2022; R. Grossová et al. 2022). However, despite this strong mass dependence, the wide range of radio luminosities observed among AGN appear to be largely uncorrelated with the stellar masses of their host galaxies (T. Mauch & E. M. Sadler 2007).

In contrast to optical-AGN (e.g. G. Kauffmann, T. M. Heckman & P. N. Best 2008; E. Donoso et al. 2010; J. Sabater, P. N. Best & M. Argudo-Fernández 2013; E. Retana-Montenegro & H. J. A. Röttgering 2017), most radio-AGNs are located in overdense and cluster-like structures. This has been quantified in numerous clustering studies, which consistently show that radio-AGN inhabit dark matter haloes with characteristic masses of  $M_h \sim 10^{13} - 10^{14} M_\odot$  across a wide range of radio luminosities, frequencies and redshifts (e.g. R. Allison et al. 2015; M. Magliocchetti et al. 2017; E. Retana-Montenegro & H. J. A. Röttgering 2017; C. L. Hale et al. 2018; G. C. Petter et al. 2024, see the review by M. Magliocchetti 2022). Radio-AGNs have also been found to be more concentrated towards cluster centres, with the brightest cluster galaxy more likely to host a radio-AGN than other galaxies in the same over-density (e.g. P. N. Best et al. 2007; V. Smolčić et al. 2011; N. A. Hatch et al. 2014; W. Mo et al. 2018; J. H. Croston et al. 2019).

Given the strong connection between radio-AGN activity and host galaxy mass, an open question is whether their enhanced clustering is simply due to their preference for massive hosts, or if they occupy distinct environments beyond this mass dependence (i.e. are galaxies that host AGN simply a random subset of massive galaxies?). Differences in the clustering between AGN and non-active galaxies of the same stellar mass may arise from galaxy-level effects, such as feedback reducing star formation or AGN preferentially residing in galaxies at the centres of haloes, as well as from assembly bias (L. Gao, V. Springel & S. D. M. White 2005; R. H. Wechsler et al. 2006), if AGN and stellar mass-matched galaxies tend to reside in haloes with different formation histories at fixed halo mass. Previous works have investigated this by comparing samples of cross-matched AGN to control samples with similar properties (see Section 4.3.2 of the review by M. Magliocchetti 2022). A number of studies find that radio-

AGN inhabit denser environments and more massive dark matter haloes than stellar mass-matched galaxies, even when controlling for quantities such as velocity dispersion (a proxy for black hole mass) or colour (G. Kauffmann et al. 2008; R. Mandelbaum et al. 2009). This environmental enhancement appears to depend on both host-galaxy properties and radio luminosity: passive AGN (hosted by galaxies with little ongoing star formation) are preferentially found in overdense regions, whereas non-passive galaxies hosting AGN do not differ from a control sample (S. Bardelli et al. 2010). At the highest radio luminosities ( $L_{1.4\text{GHz}} > 10^{24.5} \text{ W Hz}^{-1}$ ), the enhanced environmental density was found to disappear (N. Malavasi et al. 2015). More recent work finds that radio-AGN largely follow the environments of mass-matched galaxies, except for a higher prevalence of AGN in galaxy clusters (S. Kolwa et al. 2019).

The connection between the radio luminosity of AGN and their environments remains unclear. Previous works have produced conflicting results on whether the clustering of radio-AGN depends on their luminosity, with some reporting that high-luminosity AGN are found in higher-density regions than their lower luminosity counterparts (e.g. S. Bardelli et al. 2010; S. N. Lindsay et al. 2014; C. L. Hale et al. 2018; J. H. Croston et al. 2019; W. Mo et al. 2020), while others have found no significant dependence of environment on radio luminosity (e.g. M. Magliocchetti et al. 2004; G. Kauffmann et al. 2008; D. Wylezalek et al. 2013; G. Castignani et al. 2014; S. Kolwa et al. 2019). Conversely, a smaller number of studies suggest the opposite trend, with low-luminosity AGN preferentially inhabiting denser regions (e.g. E. Donoso et al. 2010; N. Malavasi et al. 2015; H. Uchiyama et al. 2022).

To investigate the clustering of sources, we measure the two-point correlation function (TPCF), which is a widely used measure of the statistical clustering of galaxies. It can be expressed as the real-space correlation function,  $\xi(r)$ , which measures clustering as a function of three-dimensional separation (e.g. P. J. E. Peebles 1980; M. Davis & P. J. E. Peebles 1983; M. Magliocchetti et al. 2004; A. L. Coil 2013), or the angular correlation function,  $\omega(\theta)$ , which quantifies clustering in two-dimensional projections on the sky when accurate line-of-sight distances are not available (e.g. H. Totsuji & T. Kihara 1969; C. M. Cress et al. 1996; C. Blake & J. Wall 2002; R. A. Overzier et al. 2003; Y. Wang, R. J. Brunner & J. C. Dolence 2013; P. W. Hatfield et al. 2016).

The clustering of galaxies is closely linked to the underlying distribution of dark matter in the Universe, as galaxies form within dark matter haloes. The efficiency of galaxy formation is enhanced in dense environments, where structures of a given mass collapse earlier due to the higher surrounding density. Consequently, galaxies do not trace the underlying dark matter distribution uniformly but are instead biased tracers (H. J. Mo & S. D. M. White 1996; J. A. Peacock 1999; A. Cooray & R. Sheth 2002). The clustering of galaxies,  $\xi_{\text{gal}}$ , and the background dark matter distribution,  $\xi_{\text{DM}}$ , are related by the bias parameter,  $b$ , through

$$b(r, z) = \frac{\delta_{\text{gal}}(r, z)}{\delta_{\text{DM}}(r, z)} = \sqrt{\frac{\xi_{\text{gal}}(r, z)}{\xi_{\text{DM}}(r, z)}}, \quad (1)$$

where  $\delta_{\text{gal}}$  and  $\delta_{\text{DM}}$  are the local galaxy and dark matter overdensities, respectively (N. Kaiser 1984). In the linear regime, the bias is approximately independent of scale,  $b(r, z) \approx b(z)$ . The bias parameter depends on halo mass, with more massive haloes being more strongly clustered, and the epoch of halo collapse,

with haloes that collapse earlier having higher bias (H. J. Mo & S. D. M. White 1996).

The aim of this paper is to investigate whether there is a difference between the environments of radio-selected AGN and a control sample of galaxies matched in both stellar mass and redshift by comparing their projected clustering. The radio data are from the MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE; M. Jarvis et al. 2016) survey, which targets multiple well-studied fields with rich ancillary data. Using a catalogue of radio sources that have been cross-matched with their host galaxies observed at optical and near-infrared (NIR) wavelengths, a sample of  $\sim 2000$  AGN covering  $\sim 7.5 \text{ deg}^2$  and extending to  $z = 2.5$  is constructed by applying a redshift-dependent luminosity threshold. Previous radio-AGN clustering studies have either used wide-area surveys but with higher flux limits (e.g. R. Mandelbaum et al. 2009; R. Allison et al. 2015; E. Retana-Montenegro & H. J. A. Röttgering 2017; G. C. Petter et al. 2024), or considered a single deep field with area  $\sim 2 \text{ deg}^2$  (e.g. M. Magliocchetti et al. 2017; C. L. Hale et al. 2018). MIGHTEE, by spanning three widely separated fields with optical/NIR overlap at high radio sensitivity, both reduces the impact of cosmic variance on the clustering measurements and allows clustering to be measured out to high redshift. To quantify the differences in environment, halo occupation models are fit to the measured correlation functions in order to infer the statistical properties of the dark matter haloes that the samples inhabit.

The paper is organized as follows. In Section 2, the data, the selection of AGN and the construction of the control sample are discussed. In Section 3, we describe the methods that are used to measure clustering and the chosen halo occupation model. The results are presented and discussed in Section 4, and our findings are summarized in Section 5.

For all calculations in this paper, we assume  $\Lambda$  cold dark matter ( $\Lambda$ CDM) cosmology with  $\Omega_\Lambda = 0.7$ ,  $\Omega_m = 0.3$ ,  $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$  with  $h = 0.7$ ,  $\sigma_8 = 0.81$ , and  $n_s = 0.96$ . All magnitudes are given in the AB system (J. B. Oke & J. E. Gunn 1983).

## 2 DATA

### 2.1 Radio data

The radio sources in this work are from the Data Release 1 (DR1; C. L. Hale et al. 2025) catalogues of the  $\sim 1.2 - 1.3 \text{ GHz}$  MIGHTEE survey (M. Jarvis et al. 2016; I. Heywood et al. 2022). Both higher- ( $\sim 5 \text{ arcsec}$ ) and lower-resolution ( $\sim 8 \text{ arcsec}$ ) images are available, and we use the catalogues that were derived from the higher-resolution data. The sources span three fields with a wealth of available ancillary data:  $4.2 \text{ deg}^2$  of the Cosmic Evolution Survey (COSMOS) field,  $14.4 \text{ deg}^2$  of the XMM-Newton Large Scale Structure (XMM-LSS) field, and  $1.5 \text{ deg}^2$  of the Chandra Deep Field-South (CDFS), with central root mean square (r.m.s.) sensitivities of  $\sim 1.2 - 3.6 \mu\text{Jy beam}^{-1}$ . By combining the fields, we increase the cumulative survey area and reduce the impact of cosmic variance on the clustering measurements. The wide separation of the fields further helps to mitigate field-specific systematics that could otherwise affect the results.

The catalogues account for the confusion limited nature of the MIGHTEE images through an iterative source detection procedure, as described in detail by C. L. Hale et al. (2025). The resulting catalogues contain 20 886, 72 187 and 21 152 radio sources in the COSMOS, XMM-LSS and CDFS fields, respectively.

### 2.2 Optical/NIR data

The catalogues of optical/NIR sources that are used in this study have photometry from several surveys. The NIR photometry is provided by the VISTA Deep Extragalactic Observations (VIDEO) survey (M. J. Jarvis et al. 2013) in XMM-LSS and CDFS, and the UltraVISTA survey (H. J. McCracken et al. 2012) in COSMOS which covers the  $YJHK_s$  bands. Further NIR coverage in the *grizy* bands for all three fields is provided by the second data release of the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; H. Aihara et al. 2022). The optical photometry includes data from the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS; J.-C. J. Cuillandre et al. 2012), which provides *ugriz* band data in the COSMOS and XMM-LSS fields, as well as from the VST Optical Imaging of the CDFS and ES1 Fields (VOICE; M. Vaccari et al. 2016) survey, which covers the *ugriz* bands in the CDFS field.

Sources were selected in the  $K_s$ -band using the source finding software SEXTRACTOR<sup>1</sup> (E. Bertin & S. Arnouts 1996). After applying masks to remove areas around bright stars and artifacts, the resulting catalogues contain 419 015, 474 602 and 367 090 sources in the COSMOS, XMM-LSS and CDFS fields, respectively, covering areas of 1.7, 4.3 and  $3.3 \text{ deg}^2$ . Therefore, in the COSMOS and XMM-LSS fields, the area available for clustering analyses is limited by the multiwavelength coverage, while in CDFS it is limited by the area of the radio observations.

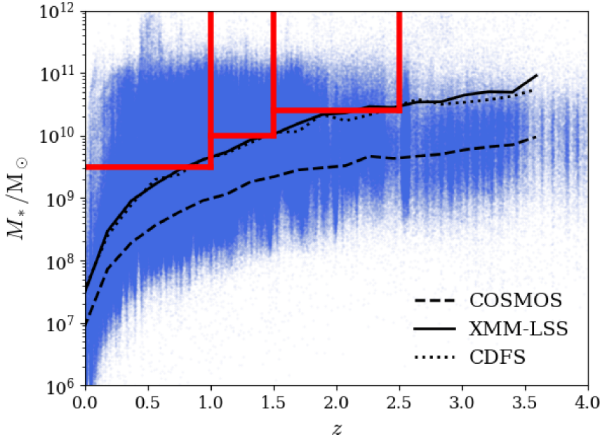
A comprehensive description of the catalogues will be provided by Stylianou et al. (in preparation).

### 2.3 Redshift PDFs

Spectroscopic redshifts are available for a subset of sources, obtained by cross-matching with the merged spectroscopic catalogues of M. Vaccari (2015) using a 1 arcsec matching radius. For the regions overlapping between the multiwavelength and radio data, this results in spectroscopic redshifts for 10 per cent, 13 per cent and 4 per cent of sources in the COSMOS, XMM-LSS, and CDFS fields, respectively.

For the remaining sources, less accurate photometric redshifts are used. To account for the uncertainties in these estimates, the redshifts are represented as probability distribution functions ( $z$ -PDFs). In the CDFS field, the  $z$ -PDFs were derived using LEPHARE (S. Arnouts et al. 1999; O. Ilbert et al. 2006) which uses a  $\chi^2$  minimization approach to determine the most likely redshift by fitting the observed photometry to a suite of galaxy and AGN template spectra. For the COSMOS and XMM-LSS fields,  $z$ -PDFs were found for each galaxy using the methods and catalogues of P. W. Hatfield et al. (2022). They combined two methods for estimating photometric redshifts: template fitting using LEPHARE, and the machine-learning algorithm GPZ (I. A. Almosallam, M. J. Jarvis & S. J. Roberts 2016; Z. Gomes et al. 2018; N. Stylianou et al. 2022), where a machine-learning model was trained using galaxies with photometric data and spectroscopic redshifts. This produced two distinct estimates for the  $z$ -PDFs of each source that were combined using a Hierarchical Bayesian model (T. Dahlen et al. 2013; K. J. Duncan et al. 2018, 2019) to create a single consensus distribution that outperforms each method individually. This is not possible in the CDFS field due to the heterogeneous nature of the optical-wavelength data and the comparative lack of spectroscopic redshift information in this field.

<sup>1</sup><https://www.astromatic.net/software/sextractor/>



**Figure 1.** Stellar mass and redshift distribution of NIR-selected galaxies from all three fields (blue points). The banding effect is caused by the discrete redshift bins of LEPHARE. The black lines are the 90 per cent stellar mass completeness limits in each field, found by following the method of L. Pozzetti et al. (2010). The red boxes are our chosen stellar mass and redshift sub-samples.

## 2.4 Stellar mass completeness

The stellar mass of each optical/NIR galaxy was estimated using the LEPHARE code by fixing the redshift to the peak of the  $z$ -PDF corresponding to the best-fitting galaxy template, and adopting the G. Bruzual & S. Charlot (2003) stellar population synthesis models available in LEPHARE. In the CDFS field, when an AGN template provides a better fit, we scale the stellar mass estimate to the peak AGN  $z$ -PDF redshift using  $M_* \propto D_L^2(z)$  where  $D_L(z)$  is the luminosity distance. This approximation is reasonable because the  $K_s$ -band observations are relatively insensitive to ongoing star formation so consistently trace the bulk of the stellar mass (T. Kodama & R. Bower 2003). Therefore, although the spectral energy distribution of the galaxy may evolve with redshift, the resulting variation in stellar mass estimates for fixed observed  $K_s$ -band magnitude will be relatively small. The stellar masses of sources with spectroscopic redshifts are similarly scaled.

Stellar mass completeness refers to the stellar mass above which all galaxies are detectable given the flux limit of the survey. Below this limit, an increasing fraction of galaxies are too faint to be observed, leading to an incomplete sample that would bias the clustering results. To determine the completeness limit, the method outlined in L. Pozzetti et al. (2010) was adopted. This approach is illustrated in Fig. 1, which presents the stellar masses of the galaxies as a function of peak  $z$ -PDF (or spectroscopic) redshift. The lowest value of stellar mass,  $M_*$ , at which a galaxy could be detected is calculated using its observed 2 arcsec aperture  $K_s$ -band magnitude by

$$\log_{10}(M_{\text{lim}}) = \log_{10}(M_*) + 0.4(K_s - K_s^{\text{lim}}), \quad (2)$$

where the completeness limits are  $K_s^{\text{lim}} = 24.8$  in COSMOS and  $K_s^{\text{lim}} = 23.5$  in XMM-LSS and CDFS.<sup>2</sup> For each field, we calculate the minimum masses at which galaxies could be observed and

<sup>2</sup>The magnitude limit in COSMOS corresponds to the  $5\sigma$ ,  $2''$  completeness limits from the fifth UltraVISTA data release (DR5). Although Stylianou et al. use the deeper DR6 limits, we adopt the DR5 values as this is the release that was used for cross-matching the radio and optical/NIR catalogues (see Hale et al. in preparation). The magnitude limit in XMM-

compute the 90th percentile of these data for bins of redshift, shown by the black lines in Fig. 1. This is used to apply a minimum stellar mass threshold in each redshift bin. The red boxes show the redshift and stellar mass sub-samples that are analysed in subsequent sections.

## 2.5 Cross-matching radio sources with host galaxies

Cross-matching radio sources with their optical/NIR counterparts associates them with their host galaxies for which redshift and stellar mass estimates have been made. Having redshifts enables the conversion of radio-fluxes to luminosities, which are used to identify AGN, and allows the sample to be divided into redshift bins so that the evolution of clustering over cosmic time can be investigated. Without cross-matching, multiple radio components that belong to the same source, for instance the two lobes of a radio jet, could be misidentified as distinct sources with small angular separations when they are in fact part of a single extended source. When calculating the clustering, this would erroneously enhance the signal at small angular scales (e.g. C. M. Cress et al. 1996). Associating radio sources with their true optical/NIR counterparts allows such radio components to be correctly merged into a single catalogue entry, and allows genuinely overlapping sources to be deblended.

To cross-match radio sources with their  $K_s$ -band host galaxies, two methods were used: (i) the automated *Likelihood Ratio* matching method (H. R. de Ruiter, A. G. Willis & H. C. Arp 1977; W. Sutherland & W. Saunders 1992, see e.g. K. McAlpine et al. 2012; R. Kondapally et al. 2021; I. H. Whittam et al. 2024), and (ii) manual cross-matching by MIGHTEE team members via the MIGHTEE zoo, a *Zooniverse* project (C. J. Lintott et al. 2008; L. Fortson et al. 2012) internal to the MIGHTEE consortium. A detailed description of the cross-matching procedure will be provided by Hale et al. (in preparation).<sup>3</sup>

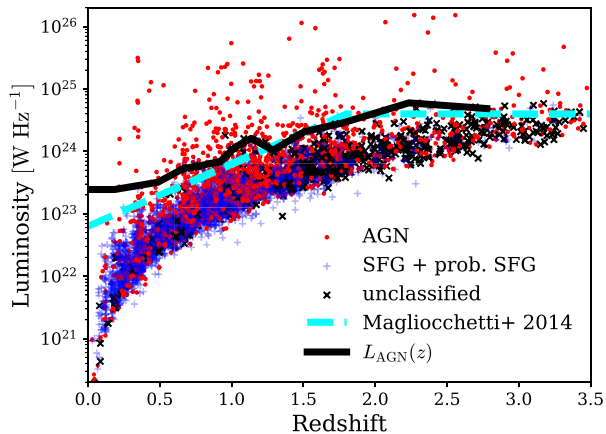
Overall,  $\sim 94$  per cent, 90 per cent, and 88 per cent of radio sources in the area that overlaps with the optical/NIR data were successfully matched to their host galaxy in the COSMOS, XMM-LSS and CDFS fields, respectively, resulting in 12 987, 28 297 and 18 180 cross-matched radio sources.

## 2.6 Selecting AGN

In this work, AGNs are selected by their radio luminosities. Since radio-AGNs are more dominant than star-forming galaxies at high radio luminosity (e.g. M. Magliocchetti et al. 2002; T. Mauch & E. M. Sadler 2007), it is possible to apply a radio luminosity threshold such that the majority of selected sources are AGNs. To convert the radio flux densities of sources to rest-frame luminosities at 1.4 GHz, we assume that the radio sources exhibit a simple power-law radio spectrum  $S_\nu \propto \nu^\alpha$ , where  $S_\nu$  is the flux density of the source with observed effective frequency  $\nu$ , and the standard spectral index of the synchrotron power-law,  $\alpha = -0.7$ , is assumed. The local effective frequency at the location of a source is derived from the effective frequency maps described in Section 2.3 of C. L. Hale et al. (2025). These account for variations

LSS and CDFS corresponds to  $\sim 90$  per cent completeness in VIDEO at that depth (M. J. Jarvis et al. 2013).

<sup>3</sup>This work used a preliminary version of the Hale et al catalogue from August 2025; there will be changes in source association between this and the final catalogue for a small subset of sources.



**Figure 2.** Radio luminosity at 1.4 GHz as a function of redshift for classified sources from I. H. Whittam et al. (2022). The black points are the unclassified sources, blue points are those classified as star-forming galaxies or probable star-forming galaxies, and red points are those classified as AGN. The star-forming galaxy points are partially transparent to improve clarity. The solid black line is the radio luminosity threshold,  $L_{\text{AGN}}(z)$ , above which 95 per cent of sources are classified as AGN. This same threshold is applied to the DR1 radio catalogue to create the AGN sample. For comparison, the AGN threshold of M. Magliocchetti et al. (2014) is denoted by the dashed cyan line.

across the survey field introduced by instrumental factors such as variations in observed frequency due to the primary beam, where sensitivity decreases across the field, and flagging, which removes data affected by interference or instrument issues (B. V. Hugo et al. 2022). Including the  $k$ -correction factor  $(1+z)^{-(1+\alpha)}$  to account for the shift in peak flux frequency due to redshift leads to

$$L_{1.4\text{GHz}} = \frac{4\pi D_L^2(z)}{(1+z)^{1+\alpha}} \left( \frac{\nu}{1.4\text{GHz}} \right)^{-\alpha} S_\nu, \quad (3)$$

where  $z$  is the redshift of the source, and  $D_L(z)$  is the corresponding luminosity distance.

To determine the luminosity threshold that separates AGN from star-forming galaxies, we utilise the classified radio sources from I. H. Whittam et al. (2022). Their study successfully cross-matched 5223 out of the 6102 (86 per cent) radio sources from the MIGHTEE Early Science radio continuum data release (I. H. Whittam et al. 2024) with their host galaxies, and classified 88 per cent of these cross-matched sources as AGN, star-forming galaxies, or probable star-forming galaxies within the central  $0.86\text{ deg}^2$  of the COSMOS field. The classification was done using a combination of several methods: radio-excess, mid-infrared colours, optical morphology, characteristic X-ray emission, and very long baseline interferometry measurements. We bin these sources in redshift bins containing approximately equal numbers of objects, and find the minimum  $L_{1.4\text{GHz}}$  value above which 95 per cent of sources are classified as AGN. To create a continuous luminosity threshold in redshift,  $L_{\text{AGN}}(z)$ , we interpolate in  $\log_{10}(L_{1.4\text{GHz}})$  between the median redshifts of the sources in each bin. The 95 per cent threshold was chosen as a compromise between providing a high-purity AGN sample while maintaining a sufficiently large population for statistical analysis. Since in this method AGN are selected by their intrinsic properties (luminosity and redshift) the same threshold is assumed to hold for the DR1 data.

Fig. 2 presents the redshift and luminosity distribution of the classified Early Science sources along with our measurements of

$L_{\text{AGN}}(z)$ . For comparison, we also include the AGN luminosity threshold of M. Magliocchetti et al. (2014), which roughly corresponds to the break of the local radio luminosity function of star-forming galaxies. This threshold has been used in several previous clustering studies (e.g. M. Magliocchetti et al. 2017; A. Chakraborty et al. 2020; A. Mazumder et al. 2022), and is similar to ours.

Of the DR1 radio sources that were not successfully cross-matched with their optical/NIR host galaxies, only 75 have radio flux densities that are high enough to potentially exceed  $L_{\text{AGN}}(z)$  assuming they have  $z < 2.5$ . The incompleteness in cross-matching is therefore not expected to have a significant impact on the clustering measurements.

## 2.7 Redshift bins and weights

We measure the clustering of galaxies for three bins of redshift:  $0 < z < 1$ ,  $1 < z < 1.5$ , and  $1.5 < z < 2.5$ . These ranges were selected to contain a sufficiently large number of AGN while maintaining a redshift width of  $\Delta z \leq 1$  to avoid excessive cosmic evolution within a single bin. The numbers of sources in each bin are given in Table 1.

Since the photometric redshifts are represented as PDFs, the approach of S. Arnouts et al. (2002) is followed and every source is assigned a weight for each redshift bin corresponding to the probability that the true redshift of the source lies within that redshift range. When calculating the clustering, every galaxy appears in every redshift bin, with its contribution weighted accordingly. For the optical/NIR galaxies, the weights are determined by integrating the normalized  $z$ -PDF of each source,  $p_i(z)$ , over the redshift range of the bin,  $[z_{\text{min}}, z_{\text{max}}]$ :

$$w_i = \int_{z_{\text{min}}}^{z_{\text{max}}} p_i(z) dz. \quad (4)$$

For sources with spectroscopic redshifts,  $w_i = 1$  if  $z_{\text{spec}}$  is between the redshift bin limits and 0 otherwise.

A radio source will only be classified as an AGN if its luminosity exceeds the redshift-dependent luminosity threshold. To incorporate this into the weighting, for each radio source 1000 random redshift samples are drawn from its  $z$ -PDF using an inverse cumulative distribution function sampler. At each sampled redshift, the corresponding radio luminosity is computed from the observed flux of the source (scaled to 1.4 GHz), and the sample is retained if the AGN luminosity threshold is exceeded. The weight assigned to the source,  $w_{\text{AGN}}$ , is then taken to be the fraction of samples that both fall within the redshift bin and meet the AGN luminosity criterion. For sources with a spectroscopic redshift,  $w_{\text{AGN}} = 1$  if its  $L_{1.4\text{GHz}} > L_{\text{AGN}}(z_{\text{spec}})$  and 0 otherwise.

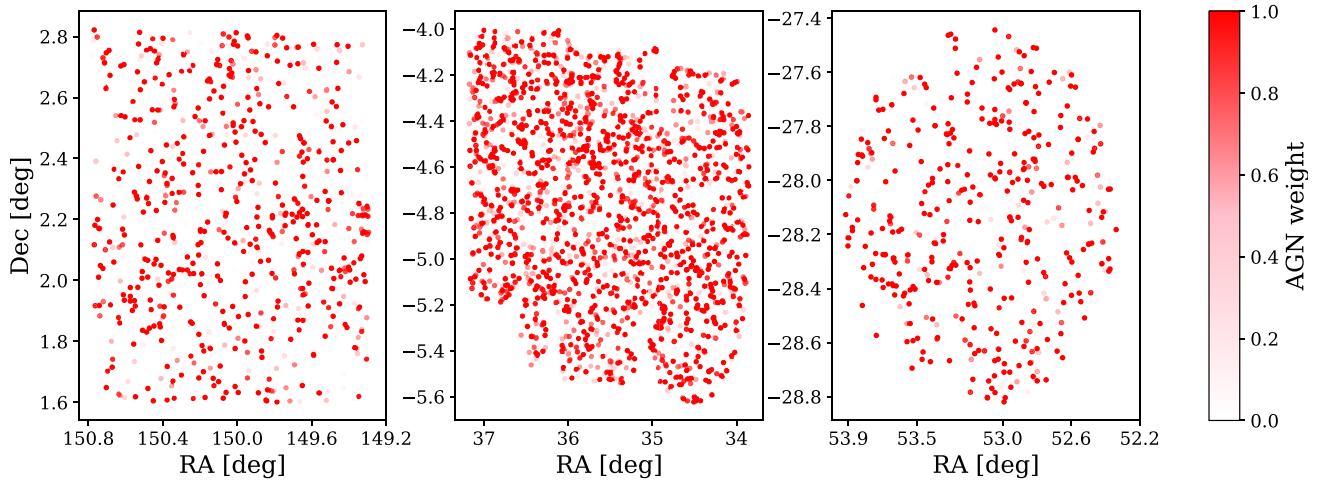
In Fig. 3, we show the distribution of radio sources on the sky in each field. The transparency of each radio source corresponds to the probability that it exceeds the AGN luminosity threshold based on its assigned weight.

## 2.8 The matched galaxy sample

Since our clustering measurements are based on angular separations (see Section 3.1), it is essential that the radio-AGN and comparison galaxy sample have similar redshift distributions; differences in redshift would lead to variations in the physical scales corresponding to a given angular separation. Additionally, it is well-established that galaxies with higher stellar masses reside in more massive haloes which exhibit stronger clustering signals

**Table 1.** The number counts ( $N$ ), median redshift ( $z_{\text{med}}$ ) and best-fitting HOD parameters for the full optical/NIR samples ('All'), the AGN sample and the matched galaxy sample, along with the corresponding reduced  $\chi^2_v = \chi^2/\text{d.o.f.}$ , for each stellar mass threshold and redshift bin. All masses are given as  $\log_{10}$  values in units of  $M_{\odot}$ . Note that the number counts are actually the sums of the weights in each subsample. The quoted uncertainties are the 16<sup>th</sup> and 84<sup>th</sup> percentiles of the posteriors.

$M_*$ threshold	Sample	$N$	$z_{\text{med}}$	$M_{\text{min}}$	$M_0$	$\sigma_{\ln M}$	$M_1$	$\alpha_s$	$\chi^2_v$
$0 < z < 1$									
9.5	All	186 954	$0.71^{+0.21}_{-0.31}$	$11.57^{+0.06}_{-0.02}$	$9.50^{+1.06}_{-1.06}$	$0.46^{+0.42}_{-0.33}$	$12.75^{+0.02}_{-0.03}$	$1.07^{+0.03}_{-0.04}$	2.66
	AGN	952	$0.76^{+0.17}_{-0.28}$	$12.81^{+0.09}_{-0.13}$	—	1	$13.62^{+0.35}_{-0.52}$	$0.93^{+0.40}_{-0.33}$	0.93
	Matched	1 720	$0.78^{+0.15}_{-0.30}$	$12.52^{+0.12}_{-0.21}$	—	1	$13.46^{+0.42}_{-0.79}$	$0.84^{+0.42}_{-0.25}$	0.84
$1 < z < 1.5$									
10	All	69 518	$1.21^{+0.21}_{-0.16}$	$12.03^{+0.08}_{-0.05}$	$9.91^{+1.19}_{-1.28}$	$0.68^{+0.40}_{-0.42}$	$13.20^{+0.02}_{-0.02}$	$1.24^{+0.03}_{-0.04}$	0.34
	AGN	567	$1.25^{+0.14}_{-0.17}$	$12.88^{+0.11}_{-0.35}$	—	1	$14.11^{+1.24}_{-1.16}$	$0.89^{+0.45}_{-0.33}$	1.50
	Matched	1 229	$1.22^{+0.19}_{-0.16}$	$12.64^{+0.24}_{-0.50}$	—	1	$13.42^{+1.06}_{-0.76}$	$1.05^{+0.30}_{-0.37}$	0.99
$1.5 < z < 2.5$									
10.4	All	36 207	$1.85^{+0.38}_{-0.26}$	$12.41^{+0.07}_{-0.06}$	$10.17^{+1.67}_{-1.51}$	$0.74^{+0.31}_{-0.45}$	$13.52^{+0.03}_{-0.03}$	$1.48^{+0.08}_{-0.18}$	1.99
	AGN	498	$1.75^{+0.44}_{-0.18}$	$12.87^{+0.18}_{-0.32}$	—	1	$13.93^{+1.39}_{-0.75}$	$0.99^{+0.42}_{-0.42}$	1.13
	Matched	990	$1.76^{+0.44}_{-0.19}$	$12.51^{+0.25}_{-0.42}$	—	1	$13.97^{+1.31}_{-0.90}$	$0.90^{+0.46}_{-0.37}$	3.43



**Figure 3.** The positions of radio sources on the sky after applying the star mask in each field. The transparency of each point reflects the weight assigned to that source, which represents the probability that the source exceeds the AGN luminosity threshold. The fields from left to right are COSMOS, XMM-LSS, and CDFS. Note that XMM-LSS (middle) is much larger than the other two fields.

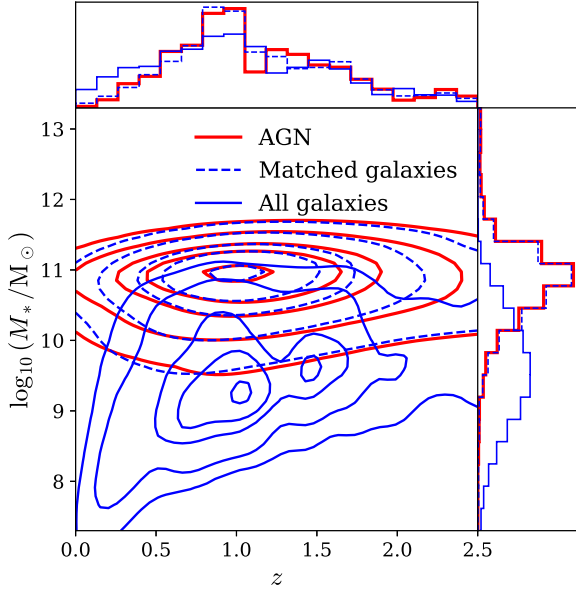
(e.g. H. J. McCracken et al. 2015; P. W. Hatfield et al. 2016). Therefore, to isolate the effects of the presence of an AGN on the clustering signal, we construct a control sample (hereafter the ‘matched galaxy sample’) from the full galaxy population, such that the sum of the  $z$ -PDFs and the stellar mass distribution of the selected sources closely match those of our AGN sample.

The first step in constructing the matched galaxy sample is to measure the AGN host distribution in redshift and stellar mass. The peak  $z$ -PDF values and their corresponding stellar masses for radio sources with luminosities that exceed  $L_{\text{AGN}}(z_{\text{peak}})$  are binned in a regular grid in the  $z$ - $\log_{10} M_*$  plane. The binning is performed over the ranges  $0 < z < 4$  and  $8 < \log_{10}(M_*/M_{\odot}) < 13$ , using bin widths of  $\Delta z = 0.1$  and  $\Delta \log_{10} M_* = 0.05$ . This produces a two-dimensional histogram that describes the joint redshift-stellar mass distribution of AGN host galaxies. The

same binning procedure is applied to the full optical/NIR galaxy sample.

The matched galaxy sample is then constructed by randomly selecting optical/NIR galaxies without replacement on a bin-by-bin basis. In each bin containing at least one AGN, up to twice as many optical/NIR galaxies as AGN are selected, subject to the number of available galaxies in that bin. The resulting numbers of galaxies in the matched galaxy sample in each redshift bin are shown in Table 1.

Although the construction of the matched sample was performed using the peak  $z$ -PDF values, we validate the selection by accounting for the redshift uncertainties using the full  $z$ -PDFs. To do this, we sample once from the  $z$ -PDF of each galaxy in the AGN and matched galaxy sample. The stellar masses are scaled to these sampled redshifts using  $M_* \propto D_L^2(z)$  where  $D_L(z)$



**Figure 4.** The redshift and stellar mass distributions of galaxies before (solid blue line) and after (dashed blue line) matching to the AGN sample (solid red line), constructed by sampling from the  $z$ -PDF of each galaxy. The main panel shows the contour plots of these distributions, while the top and right panels display the corresponding normalized histograms for redshift and stellar mass, respectively. The contours are the 10, 25, 50, 68, and 95 per cent levels.

is the luminosity distance (as in Section 2.4). We then perform a Kolmogorov–Smirnov (KS) test on the resulting redshift and stellar mass distributions between  $0 < z < 2.5$ . Typically, below a  $p$ -value of 0.05, the null hypothesis that both samples come from the same parent distribution is rejected. Sampling many times, we find averages of  $p = 0.10$  and  $0.07$  for the resulting redshift and stellar mass distributions, respectively, so they are consistent with being drawn from the same underlying distributions.

The joint redshift and stellar mass distribution of the full optical/NIR sample before matching, the AGN sample and the matched galaxy sample are shown in Fig. 4.

### 3 METHODS

#### 3.1 Angular two-point correlation function

In this work, the projected angular clustering is measured rather than the full three-dimensional spatial clustering because we lack precise spectroscopic redshifts for most sources. Without accurate redshifts, estimates of  $\xi(r)$  could be biased due to uncertainties in the line-of-sight distances.  $\omega(\theta)$  is defined as the excess probability of finding a pair of galaxies from two populations separated by an angle  $\theta$  on the sky, over the probability that would be obtained if the galaxies were distributed randomly:

$$\delta P = \sigma [1 + \omega(\theta)] \delta \Omega, \quad (5)$$

where  $\delta P$  is the probability of finding two galaxies separated by angle  $\theta$ ,  $\sigma$  is the surface density of sources on the sky, and  $\delta \Omega$  is the solid angle (P. J. E. Peebles 1980). If the separation between pairs of galaxies from the same population is measured, then  $\omega(\theta)$  is known as the *auto*-correlation function, whereas if the two populations are distinct, it is known as the *cross*-correlation function. In the case of the cross-correlation between two populations with

surface densities  $\sigma_1$  and  $\sigma_2$ , the excess probability is given by

$$\delta P_{12} = \sigma_1 \sigma_2 [1 + \omega(\theta)] \delta \Omega_1 \delta \Omega_2. \quad (6)$$

##### 3.1.1 Estimating $\omega(\theta)$ from data

The TPCF can be estimated using the data by comparing the pair counts of separations between data sources for bins of angular separation to pair counts of separations between sources in a catalogue of sources with randomly generated positions. We use the estimator of I. Szapudi & A. S. Szalay (1998):

$$\omega(\theta) = \frac{\overline{D_n D_m}(\theta) - \overline{D_n R_m}(\theta) - \overline{D_m R_n}(\theta) + \overline{R_n R_m}(\theta)}{\overline{R_n R_m}(\theta)}, \quad (7)$$

where  $\overline{D_n D_m}(\theta)$ ,  $\overline{D_n R_m}(\theta)$  and  $\overline{R_n R_m}(\theta)$  are the data–data, data–random and random–random pair counts for separations within  $\theta$  to  $\theta + d\theta$  between the two populations, denoted by  $n, m = 1, 2$ . For  $n = m$ , this expression reduces to the autocorrelation estimator of S. D. Landy & A. S. Szalay (1993), who showed that the inclusion of the cross-term,  $\overline{D_n R_m}(\theta)$ , in the estimator produces errors that are closer to their expected Poissonian values. When calculating the pair counts, each galaxy contributes only its assigned weight. Since  $\mathcal{O}(10)$  times more random positions than data are used in order to minimize statistical noise in pair counts, all pair counts are normalized by the total number of counts over all bins of angular separation so that they sum to unity.

Since the bin counts are discrete, it may naively be assumed that the uncertainties on  $\omega(\theta)$  for the  $i^{\text{th}}$   $\theta$  bin are Poissonian, and given by

$$\delta \omega_i = \frac{1 + \omega_i}{\sqrt{DD_i}}, \quad (8)$$

where  $DD_i$  are the unnormalized data–data pair counts for that bin. However, this underestimates the uncertainties as it fails to take into account both the fact that the  $\theta$  bins are correlated, and the contribution of cosmic variance from large-scale matter fluctuations which dominates the uncertainty on large angular scales. Instead, the uncertainties in  $\omega(\theta)$  are calculated using the jackknife sampling method (M. H. Quenouille 1956). Here, the survey field is split into  $N_{\text{patch}}$  subregions of roughly equal area. Each patch is then excluded in turn from the clustering measurements resulting in  $N_{\text{patch}}$  estimates of  $\omega_i$ . The elements of the covariance matrix are then estimated to be

$$\text{COV}_{ij}(\omega) = \frac{N_{\text{patch}} - 1}{N_{\text{patch}}} \sum_{k=1}^{N_{\text{patch}}} (\omega_i^k - \bar{\omega}_i) (\omega_j^k - \bar{\omega}_j), \quad (9)$$

where  $\omega_i^k$  is the value of  $\omega_i$  in the  $k^{\text{th}}$  jackknife realisation, and the expectation value,  $\bar{\omega}_i$ , is the mean value of  $\omega_i$  over all  $N_{\text{patch}}$  jackknife estimates,  $\bar{\omega}_i = \sum_{k=1}^{N_{\text{patch}}} \omega_i^k / N_{\text{patch}}$ . The uncertainty on  $\omega_i$  is the standard deviation of the individual jackknife realizations from the mean value multiplied by  $\sqrt{N_{\text{patch}} - 1}$  which qualitatively accounts for the fact that the individual jackknife samples are not independent (P. Norberg et al. 2009).

To calculate the clustering results in this work, the `TreeCorr` <sup>4</sup> PYTHON package (M. Jarvis, G. Bernstein & B. Jain 2004) is used, which takes  $N_{\text{patch}}$ , and the RA and Dec of the data and random sources as inputs. We also provide the weight for each data source, as described in Section 2.7. The estimator of equation (7) is then

<sup>4</sup><https://rmjarvis.github.io/TreeCorr>

used to calculate  $\omega(\theta)$ . For the jackknife uncertainties, we use  $N_{\text{patch}} = 100$  patches spread over the three fields.

### 3.1.2 Random catalogues

As discussed in Section 3.1.1, to estimate the TPCF of the data, a catalogue of randomly distributed sources is required. The selection of objects in these catalogues should mirror the selection of the data to ensure that the physical clustering of sources is measured, rather than observational selection effects which will affect the observed density of sources.

To generate sources for the radio random catalogues, the approach of C. L. Hale et al. (2018) is followed. First, random points in RA and cos(Dec) are generated, and any points that lie outside of the r.m.s. map of the fields are removed. This is repeated until there are  $\sim 2 \times 10^7$  points in each field. Next, each point is randomly assigned a flux density and redshift from an entry in the Square Kilometre Array Design Study (SKADS) 100 deg<sup>2</sup> Simulated Sky simulations component catalogue (R. J. Wilman et al. 2008, 2010). This flux is scaled to 1.4 GHz using the maps of local effective frequencies (as discussed in Section 2.6) and assuming the power-law relation  $S_\nu \propto \nu^\alpha$ .

The noise across each field is not uniform due to the variation in beam sensitivity and artifacts around bright sources. This introduces deviations to the observed fluxes, causing some faint sources to appear to fall below the detection threshold, whereas others may be boosted above it. This affects the completeness of the catalogue and the distribution of sources, especially near the edges of the fields where the noise is greatest. To replicate these effects in the random catalogue, a noise term is added to the assigned flux of each point. This is sampled from a Gaussian distribution with a mean of zero and standard deviation equal to the corresponding r.m.s. value at the location of the point. The total flux of each point is then calculated as the sum of the assigned flux and this noise, and a point is included in the random catalogue only if its total flux is greater than a  $5\sigma$  threshold, where  $\sigma$  is the r.m.s. at the location of the point. For the remaining points, their 1.4 GHz luminosities are calculated using their scaled fluxes and assigned redshifts, and the AGN luminosity threshold of Section 2.6 is applied.

More advanced processes exist to account for other factors that affect the selection of sources, such as extended sources having lower signal-to-noise ratios (SNR), and smearing due to time and bandwidth averaging, calibration errors and ionospheric distortion reducing peak flux densities (C. L. Hale et al. 2024). However, 99 per cent of sources with  $w_{\text{AGN}} > 0.1$  have a  $\text{SNR} > 28\sigma$ , so these effects will be minimal.

After repeating this process for each of the three fields, a star mask is applied to exclude random sources located near bright stars where photometric measurements in the optical/NIR data are not possible. Finally, we randomly select ten times as many random sources as there are data sources in each field. Using a significantly larger number of random sources than data reduces the statistical noise in the TPCF estimation.

## 3.2 HOD modelling

To connect the AGN clustering measurements to the statistical properties of the dark matter haloes they inhabit, we make use of a halo-occupation distribution (HOD; A. A. Berlind & D. H. Weinberg 2002; A. Cooray & R. Sheth 2002; I. Zehavi et al. 2005)

model. The HOD framework links the typical number of galaxies in a halo with its mass, providing a statistical description of how galaxies occupy dark matter haloes. This relies on knowledge of both the halo profile and the halo mass function, which must be calibrated using numerical simulations. By assuming a model of how the bias varies with halo mass, the correlation function of galaxies can be predicted and compared to observational data. The parameters of the HOD model, such as the minimum halo mass required to host a central galaxy and how the number of satellite galaxies increases with mass, can then be constrained by fitting the correlation function predicted by the model to the observed clustering measurements (I. Zehavi et al. 2005; Z. Zheng et al. 2005).

In HOD modelling, each dark matter halo is assumed to be populated by either zero or one central galaxy with a number of orbiting satellite galaxies. The occupation of galaxies depends solely on halo mass, with no additional dependence on its environment or formation history. Central galaxies are present in haloes with sufficiently large mass, and satellite galaxies only occupy haloes that already host a central galaxy. The number of satellites then increases monotonically with halo mass (Z. Zheng et al. 2005).

For a choice of concentration–mass relation (how the concentration of dark matter haloes depends on their mass), halo mass function (the number density of haloes as a function of mass), and halo bias model (how dark matter haloes are distributed relative to the overall matter distribution of the Universe), the HOD model can be used to predict the galaxy power-spectrum, and then projected to give the angular correlation function (K.-W. Ng & G.-C. Liu 1999; G. Chon et al. 2004).

Within the HOD framework, the TPCF can be decomposed into two contributions: the *one*-halo term and the *two*-halo term (A. A. Berlind & D. H. Weinberg 2002). The one-halo term describes the clustering signal from pairs of galaxies that reside within the same dark matter halo. This term dominates at small angular scales, typically corresponding to physical scales of less than  $\sim 1$  Mpc (A. Cooray & R. Sheth 2002), and reflects the internal structure of haloes, including the distribution of satellite galaxies around the central galaxy. In contrast, the two-halo term accounts for the clustering of galaxies that reside in separate dark matter haloes. It becomes significant at larger angular scales and is governed by the large-scale distribution of dark matter haloes, tracing the underlying cosmic web.

We note, however, that HOD modelling tends to underpredict the clustering signal in the ‘quasi-linear’ transitional regime between the one- and two-halo terms. This is due to halo-exclusion effects (haloes cannot overlap) and the breakdown of linear perturbation theory on these scales (e.g. C. Fedeli et al. 2014; A. J. Mead et al. 2015). While there is no simple correction, this caveat should be kept in mind when interpreting our results.

### 3.2.1 The HOD model

For the HOD modelling in this work, we use the LSST Dark Energy Science Collaboration’s Core Cosmology Library (CCL)<sup>5</sup> PYTHON package (N. E. Chisari et al. 2019). We assume the A. R. Duffy et al. (2008) concentration–mass relation, the G. Despali et al. (2016) halo mass function, and the J. L. Tinker et al. (2010) halo bias model, which were calibrated using simulations. It is

<sup>5</sup><https://github.com/LSSTDESC/CCL>

also assumed that the radial distribution of dark matter within haloes has a Navarro–Frenk–White profile (NFW; J. F. Navarro, C. S. Frenk & S. D. M. White 1996), and that the distribution of satellite galaxies directly follows this profile. Halo masses are taken to be the virial mass, defined with respect to the critical density following G. L. Bryan & M. L. Norman (1998).

The HOD model used by CCL draws from several papers, including Z. Zheng et al. (2005), S. Ando, A. Benoit-Lévy & E. Komatsu (2017), and A. Nicola et al. (2020), and has the following five parameters of interest:

- (i)  $M_{\min}$ , the characteristic halo mass at which half of haloes host a central galaxy;
- (ii)  $M_0$ , the minimum halo mass required for a halo to host a satellite galaxy;
- (iii)  $M_1$ , the characteristic halo mass at which a halo typically contains a single satellite galaxy;
- (iv)  $\sigma_{\ln M}$  describes the smoothness of the transition in halo mass for central galaxy formation;<sup>6</sup> and
- (v)  $\alpha_s$ , the power-law index governing how the number of satellites scales with the halo mass.

For a given halo mass,  $M_h$ , the average number of central galaxies is given by an expression that smoothly transitions from 0 to 1:

$$\bar{N}_c(M_h) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{\ln M_h - \ln M_{\min}}{\sigma_{\ln M}} \right) \right], \quad (10)$$

where the error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x dt e^{-t^2}. \quad (11)$$

Above the minimum mass required to host a satellite galaxy,  $M_0$ , the average number of satellites in a halo that contains a central galaxy increases as a power law:

$$\bar{N}_s(M_h) = \Theta(M_h - M_0) \left( \frac{M_h - M_0}{M_1} \right)^{\alpha_s}, \quad (12)$$

where  $\Theta(x)$  is the Heaviside step function.

The average total number of galaxies in a halo of given halo mass is then,

$$\bar{N}_{\text{tot}}(M_h) = \bar{N}_c(M_h) \times [1 + \bar{N}_s(M_h)]. \quad (13)$$

Note that  $\bar{N}_{\text{tot}}$  is defined so that a halo must have a central galaxy in order to have a satellite contribution to the total number of galaxies.

Once the best-fitting HOD parameters have been determined, the following values can be derived: the mean halo mass,

$$\langle M_h \rangle(z) = \int dM M n(M, z) \frac{\bar{N}_{\text{tot}}(M)}{\bar{n}_g(z)}, \quad (14)$$

the fraction of galaxies that are satellites,

$$f_{\text{sat}}(z) = 1 - f_{\text{cen}}(z) = 1 - \int dM n(M, z) \frac{\bar{N}_c(M)}{\bar{n}_g(z)}, \quad (15)$$

and the mean galaxy bias,

$$b(z) = \int dM b_h(M, z) n(M, z) \frac{\bar{N}_{\text{tot}}(M)}{\bar{n}_g(z)}, \quad (16)$$

<sup>6</sup>Note that  $\sigma_{\ln M}$  is defined such that all logarithms of mass entering  $\bar{N}_c(M_h)$  are natural logarithms. This differs from the convention in some papers where  $\log_{10}$  is used.

where  $n(M, z)$  is the halo mass function,  $b_h(M, z)$  is the halo bias model, and the integrals are normalized by the mean number density of galaxies,

$$\bar{n}_g(z) = \int dM n(M, z) \bar{N}_{\text{tot}}(M). \quad (17)$$

One additional consideration when modelling the cross-correlation functions is the overlap between the samples. In other clustering studies, it is usually assumed that the two populations being cross-correlated are independent, or that the overlap between the two samples is small enough that it can be ignored. However, our AGNs are a subset of the optical/NIR galaxies, and the matched galaxy sample is drawn from the full optical/NIR population. This introduces additional self-pairing terms when measuring the cross-correlation that correspond to galaxies that are present in both samples. To account for these effects in the model, we modified the CCL code, with the theory and changes described in Appendix A.

### 3.2.2 Model fitting and integral constraints

Since the observed fields have finite sizes, the maximum possible angular separation of two sources is limited. The observed TPCF,  $\omega_{\text{obs}}$ , will therefore be underestimated at large angular scales and negatively offset from the true TPCF,  $\omega_{\text{true}}$ :

$$\omega_{\text{obs}}(\theta) = \omega_{\text{true}}(\theta) - \sigma_{\text{IC}}^2, \quad (18)$$

where the offset,  $\sigma_{\text{IC}}^2$ , is known as the integral constraint. Its value is given analytically by the expression from E. J. Groth & P. J. E. Peebles (1977),

$$\sigma_{\text{IC}}^2 = \frac{1}{\Omega^2} \iint \omega_{\text{true}}(\theta) d\Omega_1 d\Omega_2, \quad (19)$$

where  $d\Omega_1 d\Omega_2$  denotes integrating over the field solid angle twice. This can be estimated numerically using the random-random pair counts (N. Roche & S. A. Eales 1999):

$$\sigma_{\text{IC}}^2 = \frac{\sum_i \bar{R}_n \bar{R}_m(\theta_i) \omega_i^{\text{true}}}{\sum_i \bar{R}_n \bar{R}_m(\theta_i)}. \quad (20)$$

To fit our model to the observed TPCF, we use the Markov chain Monte Carlo (MCMC) sampling method provided by the PYTHON package `emcee`<sup>7</sup> (D. Foreman-Mackey et al. 2013). The starting positions of the walkers are drawn randomly from uniform priors over  $11 < \log_{10}(M_{\min}/M_{\odot}) < 15$ ;  $8 < \log_{10}(M_0/M_{\odot}) < \log_{10}(M_1/M_{\odot})$ ;  $0 < \sigma_{\ln M} < 1.4$ ;  $\log_{10}(M_{\min}/M_{\odot}) < \log_{10}(M_1/M_{\odot}) < 16$ ; and  $0.4 < \alpha_s < 1.6$ . Although the parameters are fit separately in each of the redshift bins, they are assumed to be constant within each bin.

The two parameters,  $M_{\min}$  and  $\sigma_{\ln M}$ , that control the shape of  $\bar{N}_c(M_h)$ , are highly degenerate since both have qualitatively similar effects on  $\omega(\theta)$  (A. N. Salcedo et al. 2020). For the optical/NIR galaxy TPCFs, this degeneracy is broken using  $\bar{n}_g$  by fitting simultaneously to both the measured  $\omega(\theta)$  using the covariance matrix defined in equation (9), and the observed number density of galaxies. A Gaussian likelihood is defined,  $\mathcal{L} = \exp(-\frac{1}{2}\chi^2)$ ,

<sup>7</sup><https://emcee.readthedocs.io/en/stable/>

where the total  $\chi^2$  is the sum of these two contributions:

$$\chi^2 = \sum_i \sum_j (\omega_i^{\text{true}} - \omega_i^{\text{model}}) \text{Cov}_{ij}^{-1} (\omega_j^{\text{true}} - \omega_j^{\text{model}}) + \frac{(N_g^{\text{obs}} - N_g^{\text{mod}})^2}{\sigma_{\text{gal}}^2}. \quad (21)$$

In the first term,  $\omega_i^{\text{model}}$  is the value of the TPCF from the HOD model for the  $i^{\text{th}}$   $\theta$  bin,  $\omega_j^{\text{true}}$  is related to the measured TPCF by equation (18), and  $\text{Cov}_{ij}^{-1}$  are the components of the inverse of the covariance matrix. In the second term, the observed number of galaxies,  $N_g^{\text{obs}}$ , is the sum of the assigned weights for galaxies above the stellar mass threshold (i.e.  $N_g^{\text{obs}} = \sum_i w_i$ ), and the number of galaxies predicted by a given model is found by multiplying the predicted number density by the comoving volume covered by the fields:

$$N_g^{\text{mod}} = \bar{n}_g \times \frac{\Omega}{4\pi} \times \frac{4}{3} \pi [d^3(z_{\text{max}}) - d^3(z_{\text{min}})], \quad (22)$$

where  $d(z)$  is the comoving distance to redshift  $z$ , and  $\Omega$  is the solid angle covered by the fields in steradians. The uncertainty on the number of galaxies,  $\sigma_{\text{gal}}^2 = \sigma_{\text{Pois}}^2 + \sigma_{\text{cv}}^2$ , contains contributions from Poisson noise,  $\sigma_{\text{Pois}} = \sqrt{N_g^{\text{mod}}}$ , and cosmic variance,  $\sigma_{\text{cv}} = b \sigma_{\text{dm}} N_g^{\text{mod}}$ . The galaxy bias is calculated using equation (16) at each step in the MCMC chain, and the fractional dark matter root cosmic variance,  $\sigma_{\text{dm}}$ , in each field is estimated using the cosmic variance cookbook of B. P. Moster et al. (2011) which scales with the width of the redshift bin as  $\sigma_{\text{dm}} \propto \Delta z^{-1/2}$ . The total cosmic variance for the combined fields is then given by a volume weighted sum,

$$\sigma_{\text{cv}}^2 = \frac{\sum_i V_i^2 \sigma_i^2}{(\sum_i V_i)^2}, \quad (23)$$

where  $V_i$  is the comoving volume of each field between the redshift bin limits, calculated as in equation (22).

For the AGN and matched galaxy sample TPCFs, the model is fitted solely to  $\omega(\theta)$  (the first term of equation 21) with the  $N_{\text{gal}}$  terms omitted. This is because the observed AGN population is not complete, since not all radio sources are successfully cross-matched with their host galaxies, which will decrease the inferred number density. Additionally, the number of galaxies in the matched sample is chosen arbitrarily, and so cannot be used in the model fitting. When fitting to the TPCF of the optical/NIR galaxies, we include both terms.

Since the number of AGN is relatively small, the uncertainties on the AGN autocorrelation are large, leading to weak constraints on the HOD parameters when fitting to it alone. For this reason, we fit to the cross-correlation between the AGN (or matched galaxy) sample and the optical/NIR galaxies. This requires knowledge of the best-fitting HOD parameters for the optical/NIR sample. To properly propagate uncertainties, in each redshift bin, a joint fitting is performed to the optical/NIR auto-correlation, the AGN cross-correlation, and the matched galaxy cross-correlation, simultaneously. The total likelihood is the product of the individual likelihoods:

$$\mathcal{L}_{\text{tot}}(\phi_{\text{AGN}}, \phi_{\text{mat}}, \phi_{\text{opt}}) = \mathcal{L}_{\text{opt}}(\phi_{\text{opt}}) \mathcal{L}_{\text{AGN} \times \text{opt}}(\phi_{\text{AGN}}, \phi_{\text{opt}}) \mathcal{L}_{\text{mat} \times \text{opt}}(\phi_{\text{mat}}, \phi_{\text{opt}}), \quad (24)$$

where  $\phi_{\text{opt}}$ ,  $\phi_{\text{AGN}}$ , and  $\phi_{\text{mat}}$  are the HOD parameters for the optical/NIR, AGN, and matched galaxy samples, respectively. This

assumes that the uncertainties on the cross-correlations are independent of those on the optical/NIR autocorrelation. Given that the optical/NIR sample is over an order of magnitude larger than the AGN and matched galaxy samples, the uncertainties on the cross-correlation measurements will be dominated by the shot noise of the latter, rather than by fluctuations in the optical/NIR population. The covariance between the uncertainties will therefore be small, and we find that fitting to the autocorrelation function alone gives the same best-fitting parameters as those obtained when fitting to the three sets of measurements simultaneously.

When fitting to the optical/NIR autocorrelation functions, all five HOD parameters are fit. However, for the AGN and matched galaxy samples, the fit is restricted to only three free parameters:  $M_{\text{min}}$ ,  $M_1$  and  $\alpha_s$ , while fixing  $M_0 = M_{\text{min}}$  and  $\sigma_{\text{ln}M} = 1$ , chosen to be consistent with the values found for high-luminosity radio-galaxies by G. C. Petter et al. (2024). We find that allowing  $M_0$  to vary has a negligible impact on the resulting model correlation function and the best-fitting values of the other parameters. Fixing  $\sigma_{\text{ln}M}$  is necessary because, for these samples, number density constraints cannot be reliably used to break its degeneracy with  $M_{\text{min}}$ .

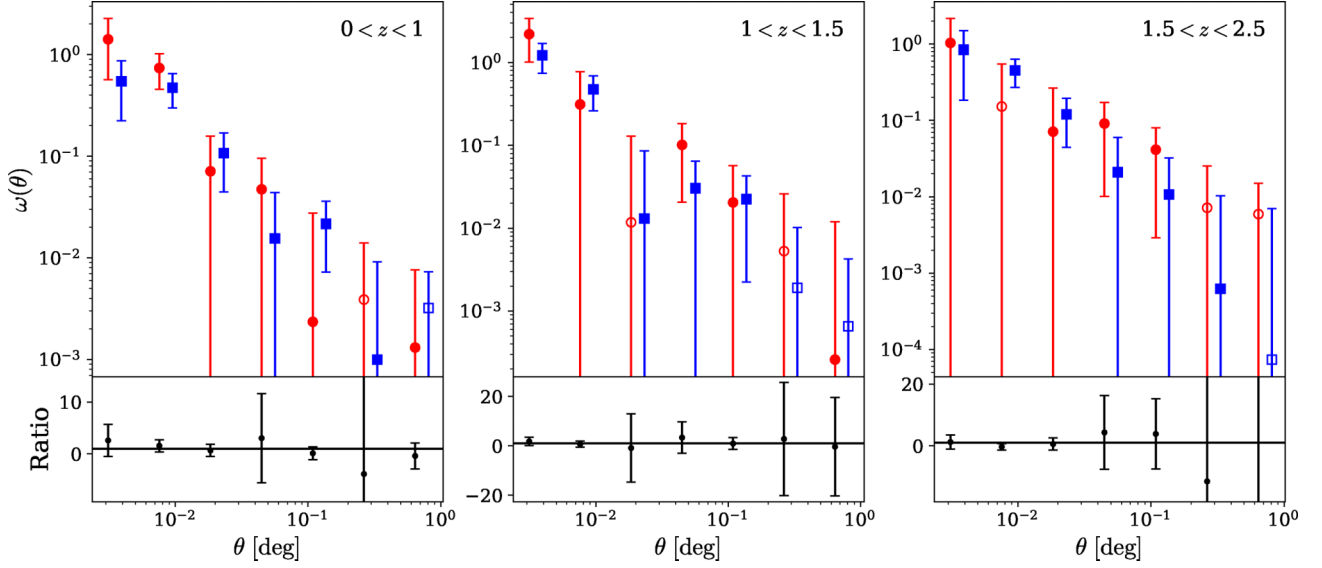
We use 60 walkers (5 per model parameter) with a 500 step burn-in phase and 2500 iterations. The fitting for the  $1.5 < z < 2.5$  bin was also repeated with twice as many iterations which produced almost identical best-fitting HOD parameters and uncertainties, so we are confident that the number of iterations here is sufficient to reach convergence. The quoted upper and lower uncertainties for the HOD parameters are the 16<sup>th</sup> and 84<sup>th</sup> percentiles of the posterior distributions.

## 4 RESULTS AND DISCUSSION

### 4.1 Autocorrelation functions of AGN and matched galaxies

Fig. 5 presents the measured autocorrelation functions of the AGN and matched galaxy samples for each redshift bin. The bottom panels in each plot are the ratio of the two correlation functions. The uncertainties are found using jackknife patches, as discussed in Section 3.1.1. We use  $\theta$  bins ranging from  $\theta = 2 \times 10^{-3} - 1^\circ$ , producing 7 values of  $\omega_i$  that are uniformly spaced in  $\log \theta$  space. The lower  $\theta$  limit was chosen to be approximately twice the 5 arcsec resolution of the radio data, below which radio source blending may become an issue, while the upper limit was chosen to be slightly greater than the angular scale where the measurements drop to zero due to the finite sizes of the fields.

All of the individual points of the measured autocorrelation function for the two samples overlap within the uncertainties. To assess whether the differences between the two curves are statistically significant, we perform a  $\chi^2$  test, accounting for correlations between  $\theta$  bins using the covariance matrices that were calculated with equation (9). We take  $\chi^2 = \mathbf{\Delta}^T [\text{Cov}(\mathbf{\Delta})]^{-1} \mathbf{\Delta}$ , where the total covariance matrix,  $\text{Cov}(\mathbf{\Delta})$ , is the sum of the individual matrices for the two samples, and  $\mathbf{\Delta} = \omega_{\text{AGN}} - \omega_{\text{mat}}$  is the difference between the two curves. The corresponding  $p$ -value is computed using the  $\chi^2$  cumulative distribution function, which quantifies the probability of obtaining the observed difference under the null hypothesis that the two samples have the same underlying clustering.



**Figure 5.** The measured angular two-point autocorrelation functions,  $\omega(\theta)$ , for AGN (red circles) and the stellar mass-matched galaxies sample (blue squares) as functions of angular separation,  $\theta$ , for each redshift bin. Left to right, these are:  $0 < z < 1$ ,  $1 < z < 1.5$  and  $1.5 < z < 2.5$ . Open markers are negative values. For clarity, the matched galaxy correlations are offset by  $+0.1$  dex in  $\theta$ . The bottom panels give the ratio of the two correlation functions, with the black horizontal line at  $\omega_{\text{AGN}}/\omega_{\text{mat}} = 1$ .

From the lowest to highest redshift bin, we obtain  $p$ -values of 0.91, 0.99, and 0.82, which are all much higher than the significance threshold. There is therefore no strong evidence for significant difference between the autocorrelations of the two samples. This motivates our use of the cross-correlation functions with the full galaxy population.

#### 4.2 Cross-correlation functions of AGN and matched galaxies with optical/NIR galaxies

To further investigate the environmental differences between AGN and the matched galaxy sample, we measure their cross-correlations with the full galaxy population. Since the optical/NIR galaxy catalogue is significantly larger than the AGN sample, computing the cross-correlation greatly increases the number of pair counts. This reduces statistical uncertainties and increases the likelihood of detecting significant differences in clustering between the two samples. The smaller error bars will also produce tighter constraints on the best-fitting HOD parameters. As before, the measurements are limited by the resolution of the radio data, so we adopt the same  $\theta$  range as for the auto-correlation measurements while increasing the number of bins to 10.

Fig. 6 shows the measured cross-correlations of AGN with optical/NIR galaxies and the cross-correlations of the matched galaxy sample with optical/NIR galaxies as functions of angular separation for each of our redshift bins and stellar mass thresholds. The uncertainties are found using the jackknife method described in Section 3.1.

For the  $0 < z < 1$  redshift bin at small angular separations, the cross-correlation signal for AGN is stronger than that of the matched galaxy sample, suggesting that galaxies are more densely clustered around AGN within the same halo than non-active galaxies of similar stellar mass. The ratio plots shows that for  $0 < z < 1$ , the AGN clustering signal is  $\sim 1.5$  times greater on scales  $\lesssim 0.1^\circ$ . For higher redshifts, the difference between the

AGN and matched galaxy signals is less pronounced; though, the AGN measurements tend to lie consistently above those of the matched sample.

For larger angular separations, however, the cross-correlations for AGN and the matched galaxies are indistinguishable within the uncertainties in all redshift bins. On large scales where the clustering of galaxies is well described by linear perturbation theory, the autocorrelation of AGN is proportional to the square of the bias parameter,  $b_{\text{AGN}}^2$ , whereas the cross-correlation scales as the product of the biases of the two populations,  $b_{\text{AGN}}b_{\text{opt}}$  (A. Cooray & R. Sheth 2002). As a result, the relative difference in bias between the AGN and matched galaxy sample is less pronounced in the two-halo term of the cross-correlation.

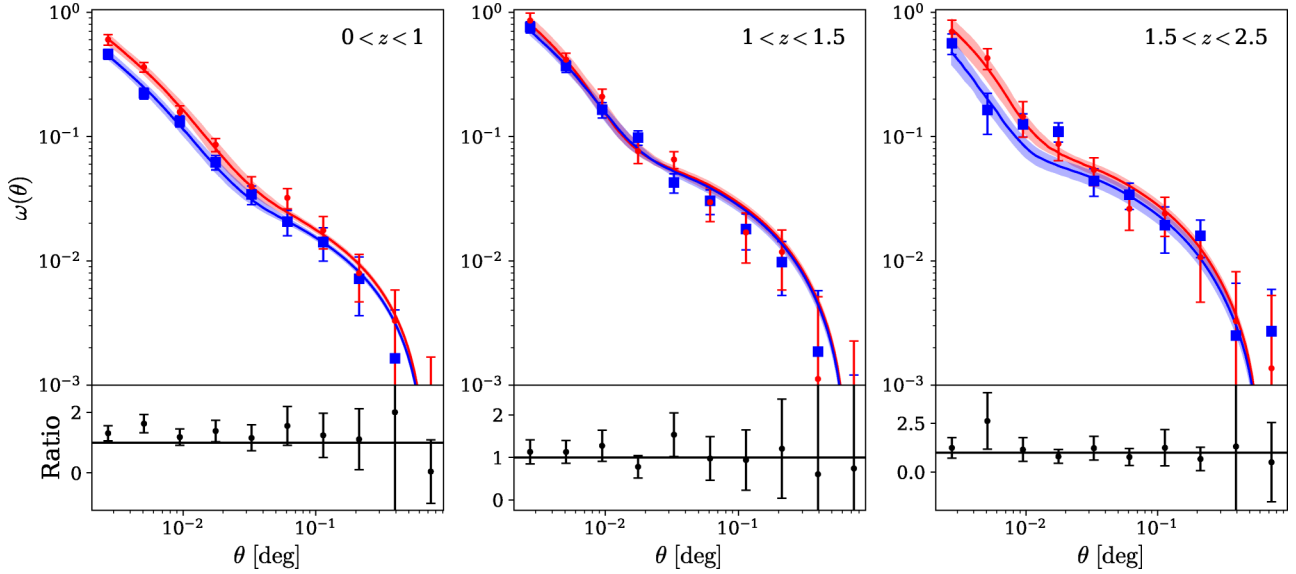
To test whether the differences between the two curves are statistically significant, we perform the same  $\chi^2$  test as described in Section 4.1. Since in each bin the AGN and matched galaxy samples are cross-correlated with the same galaxy sample, their measurements are not independent, and this must be accounted for in the covariance matrix. The total covariance matrix is

$$\begin{aligned} \text{Cov}(\omega_{\text{AGN}} - \omega_{\text{mat}}) &= \text{Cov}(\omega_{\text{AGN}}) + \text{Cov}(\omega_{\text{mat}}) \\ &- [\text{Cov}(\omega_{\text{AGN}}, \omega_{\text{mat}}) + \text{Cov}^T(\omega_{\text{AGN}}, \omega_{\text{mat}})], \end{aligned} \quad (25)$$

where  $\text{Cov}(\omega_{\text{AGN}})$  and  $\text{Cov}(\omega_{\text{mat}})$  are the covariance matrices for the AGN and matched galaxy cross-correlations estimated using equation (9), and the elements of the cross-covariance matrix are given by P. Norberg et al. (2009)

$$\begin{aligned} &\text{Cov}_{ij}(\omega_{\text{AGN}}, \omega_{\text{mat}}) \\ &= \frac{N_{\text{patch}} - 1}{N_{\text{patch}}} \sum_{k=1}^{N_{\text{patch}}} (\omega_{\text{AGN}}^k - \bar{\omega}_{\text{AGN}})_i (\omega_{\text{mat}}^k - \bar{\omega}_{\text{mat}})_j, \end{aligned} \quad (26)$$

with  $\text{Cov}_{ij}^T(\omega_{\text{AGN}}, \omega_{\text{mat}}) = \text{Cov}_{ji}(\omega_{\text{AGN}}, \omega_{\text{mat}})$  its transpose. Performing the test, we find a  $p$ -value of  $7.9 \times 10^{-3}$  for the  $0 < z < 1$  redshift bin which is significantly below the  $p = 0.05$  significance threshold, so the correlation functions differ significantly. However, for the  $1 < z < 1.5$  and  $1.5 < z < 2.5$  bins, we find  $p = 0.23$



**Figure 6.** The measured angular two-point cross-correlation functions,  $\omega(\theta)$ , between AGN and optical/NIR galaxies (red circles), and the cross-correlation of stellar mass-matched galaxies with optical/NIR galaxies (blue squares) as functions of angular separation,  $\theta$ , for each redshift bin. Left to right, these are:  $0 < z < 1$ ,  $1 < z < 1.5$  and  $1.5 < z < 2.5$ . The solid lines are the best-fitting HOD models and the shaded regions are the  $1\sigma$  confidence intervals. The bottom panels give the ratio of the two correlation functions, with the black horizontal line at  $\omega_{\text{AGN}}/\omega_{\text{mat}} = 1$ .

and 0.06, respectively, so there is no significant difference between the two clustering signals.

The enhanced clustering in the lowest redshift bin implies that AGN are more biased tracers of the underlying dark matter distribution and, in turn, reside in more massive dark matter haloes compared to non-active galaxies of the same stellar masses, in agreement with the findings of previous works (e.g. G. Kauffmann et al. 2008; R. Mandelbaum et al. 2009; S. Bardelli et al. 2010; N. Malavasi et al. 2015).

In addition to measuring the clustering of the AGN and matched sample, we also investigated whether the cross-correlations of AGN with the optical/NIR galaxy sample depend on their radio luminosity by dividing the sample into low- and high-luminosity sub-samples. Since the AGN selection threshold is redshift-dependent, a corresponding minimum luminosity was applied in each bin so that the subsamples were drawn from comparable redshift distributions. New AGN weights were recomputed for each radio source following a similar procedure to that in Section 2.7, giving the probability that the source both exceeds the luminosity threshold and falls within the luminosity range of the subsample. Applying the same significance test as above, no significant differences were found between the clustering of low- and high-luminosity AGN. While this null result may indicate that the environments of AGN are genuinely independent of radio luminosity, it is more likely a consequence of the large statistical uncertainties resulting from the small sample sizes. Therefore, a larger dataset is required to establish this conclusively.

### 4.3 HOD modelling results

Having determined that galaxies hosting AGN are significantly more clustered than non-active galaxies matched in stellar mass, we now present the results of HOD modelling to infer the properties of the dark matter haloes that host our galaxy samples. We first present the best-fitting HOD parameters for the optical/NIR autocorrelations function for each redshift and stellar mass bin,

which must be known to measure the HOD parameters of the AGN and matched galaxy samples from their cross-correlations. We then derive the best-fitting AGN HOD parameters and compare to previous results in the literature. Finally, we compare the best-fitting HOD parameters of the AGN and matched galaxy samples.

Table 1 presents the best-fitting HOD parameters for the optical/NIR, matched galaxy and AGN samples for each stellar mass and redshift bin along with the number of objects in each sample and the median redshifts.

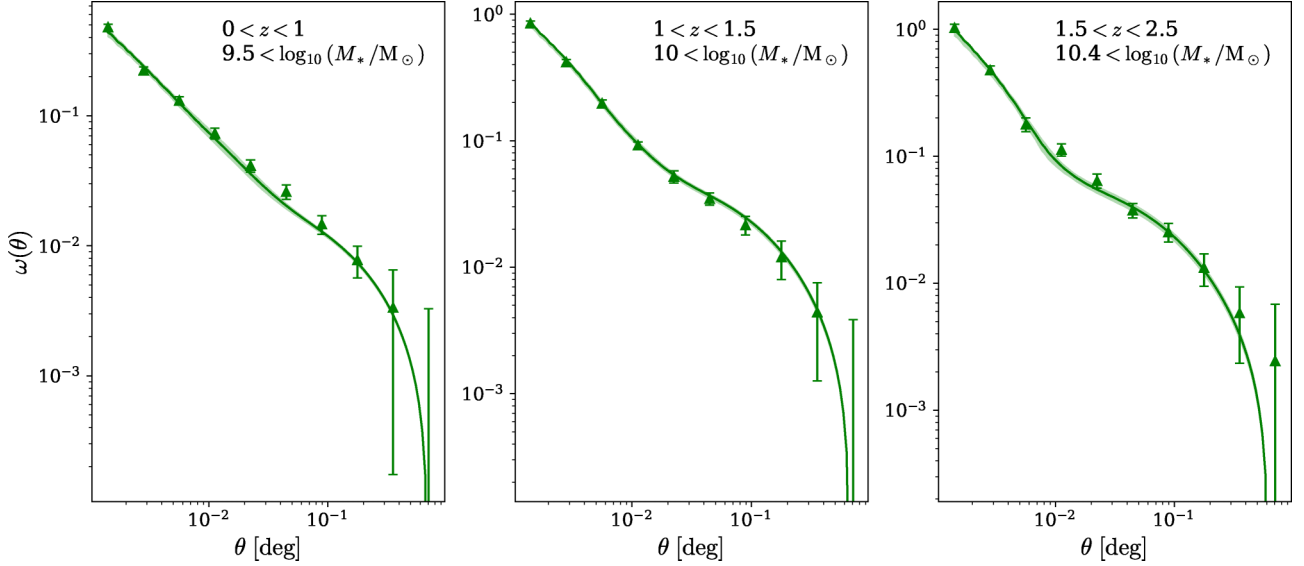
#### 4.3.1 Optical/NIR sample HOD results

In this section, the best-fitting HOD parameters for the autocorrelations of the optical/NIR samples are discussed and compared to the values found by previous works. The measured autocorrelation functions and best-fitting models are shown in Fig. 7.

For the best-fitting HOD parameters, reduced  $\chi^2_{\nu} = 2.66, 0.34$ , and 1.99 were obtained for the  $0 < z < 1$ ,  $1 < z < 1.5$  and  $1.5 < z < 2.5$  redshift bins, respectively. The number of degrees of freedom is calculated as the number of  $\theta$  bins (10) plus one (for fitting to  $\bar{n}_g$ ), minus the number of model parameters (5).

Table 2 presents the characteristic halo masses,  $M_{\text{min}}$  and  $M_1$ , from previous studies that use the same five parameter HOD model. We include the results for their closest matching redshift ranges and stellar mass thresholds to the three used in this work. In particular, P. W. Hatfield et al. (2016) measured the autocorrelation function of galaxies from the first data release of the VIDEO survey out to  $z \sim 1.7$  over an area of  $1.5 \text{ deg}^2$ , while H. J. McCracken et al. (2015) measured the clustering of galaxies from UltraVISTA to  $z \sim 2$  over  $1 \text{ deg}^2$ .

Overall, we find that our measurements of  $M_{\text{min}}$  and  $M_1$  are broadly consistent with those studies across the full redshift range considered. In the lowest redshift bin, we find a slightly lower value of  $\log_{10}(M_{\text{min}}/M_*) = 11.57^{+0.06}_{-0.02}$  compared to the  $\log_{10}(M_{\text{min}}/M_*) = 11.80^{+0.09}_{-0.08}$  reported by P. W. Hatfield et al.



**Figure 7.** The measured angular two-point autocorrelation functions,  $\omega(\theta)$ , for the full optical/NIR galaxy populations (upwards pointing green triangles) as functions of angular separation,  $\theta$ , for each stellar mass and redshift bin. Left to right, these are:  $0 < z < 1$ ,  $1 < z < 1.5$ , and  $1.5 < z < 2.5$  with stellar mass thresholds of  $\log_{10}(M_*/M_\odot) > 9.5$ ,  $10$  and  $10.4$ , respectively. The solid lines are the best-fitting HOD models and the shaded regions are the  $1\sigma$  confidence intervals. Note that the model is fitted to the observed number densities of galaxies in addition to the clustering measurements shown in this figure.

**Table 2.** Characteristic halo masses,  $M_{\min}$  and  $M_1$ , of optical/NIR galaxies from previous works for comparable redshift ranges and stellar mass thresholds to the three used in this study.

Survey	Reference	Redshift	$M_*$ threshold	$\log_{10}(M_{\min}/M_\odot)$	$\log_{10}(M_1/M_\odot)$
VIDEO	P. W. Hatfield et al. (2016)	$0.5 < z < 0.75$	9.6	$11.80^{+0.09}_{-0.08}$	$12.80^{+0.22}_{-0.16}$
		$1 < z < 1.25$	10.1	$12.00^{+0.06}_{-0.08}$	$13.20^{+0.14}_{-0.24}$
UltraVISTA	H. J. McCracken et al. (2015) (see their fig. 9)	$1.5 < z < 2.5$	10.4	$\sim 12.3$	$\sim 13.5$

(2016), although their result corresponds to a stellar mass threshold that is higher by 0.1 dex. The overall agreement in the inferred HOD parameters, together with the acceptable reduced  $\chi^2$  values, improves confidence in the reliability of the AGN HOD parameters derived from the cross-correlations.

#### 4.3.2 AGN sample HOD results

This section discusses the best-fitting HOD parameters for the AGN cross-correlation functions presented in Table 1, and compares the derived typical halo masses and bias values to previous results in the literature.

Good fits to the AGN cross-correlation were obtained for all redshift bins with reduced chi-squared values of  $\chi^2_v = 0.93$ , 1.50 and 1.13, for the  $0 < z < 1$ ,  $1 < z < 1.5$  and  $1.5 < z < 2.5$  redshift bins, respectively, for 7 degrees of freedom.

The best-constrained parameter in terms of relative uncertainty is  $M_{\min}$ , which remains consistent across the redshift bins. It ranges from  $\log_{10}(M_{\min}/M_\odot) = 12.81^{+0.09}_{-0.13}$  in the  $0 < z < 1$  bin to  $\log_{10}(M_{\min}/M_\odot) = 12.87^{+0.18}_{-0.32}$  in the  $1.5 < z < 2.5$  bin. This suggests that the characteristic halo mass required to host a radio-AGN does not evolve strongly over the redshift range considered. In the intermediate-redshift bin,  $1 < z < 1.5$ , with  $z_{\text{med}} = 1.25^{+0.14}_{-0.17}$ , we obtain  $\log_{10}(M_{\min}/M_\odot) = 12.88^{+0.11}_{-0.35}$ . This is lower than the value reported by M. Magliocchetti et al. (2017)

of  $\log_{10}(M_{\min}/M_\odot) = 13.6^{+0.3}_{-0.6}$  for a Very Large Array (VLA) radio sample in COSMOS at  $z \approx 1.25$ . They assumed that all radio-AGN are central galaxies (i.e. one AGN per halo) and fit only to the two-halo term, for which angular correlation scales with the bias as  $\omega_{2h} \propto b^2$ . Satellites preferentially reside in higher mass haloes (typically around  $M_1$ ) that have larger  $b_h(M_h, z)$ . Including satellites therefore raises the mean bias above what would be predicted for centrals alone. To reproduce the observed two-halo amplitude with central galaxies only,  $M_{\min}$  must be driven to higher values so that galaxies occupy more massive, and hence more biased, haloes. This may explain our lower  $M_{\min}$  value, although we note the large uncertainty associated with the measurement of M. Magliocchetti et al. (2017) due to the limited area coverage within just the COSMOS field.

The parameters that determine the number of AGN satellite galaxies,  $M_1$  and  $\alpha_s$ , are poorly constrained with large uncertainties relative to those of the full optical/NIR sample. For increasing redshift bins, we find  $\log_{10}(M_1/M_\odot) = 13.62^{+0.35}_{-0.52}$ ,  $14.11^{+1.24}_{-1.16}$  and  $13.93^{+1.39}_{-0.75}$ . If the fraction of AGN hosts that are satellite galaxies is small, variations in  $M_1$  or  $\alpha_s$  will have minimal effect on the modelled correlation function (see the flat prior distributions in Appendix B). Evaluating equation (15), we derive satellite fractions of  $f_{\text{sat}} = 0.18^{+0.16}_{-0.09}$ ,  $0.06^{+0.21}_{-0.05}$  and  $0.06^{+0.09}_{-0.05}$ . This aligns with observations that the vast majority of radio-loud AGNs are lo-

**Table 3.** The values of the typical halo mass,  $\langle M_h \rangle$ , the satellite fraction,  $f_{\text{sat}}$ , and the bias parameter,  $b$ , for the AGN and matched galaxy samples in each redshift bin. To obtain a single value in each bin, we marginalize  $\langle M_h \rangle(z)$ ,  $f_{\text{sat}}(z)$ , and  $b(z)$  over the respective redshift distributions. Uncertainties are obtained by sampling from the posteriors of the HOD parameters and taking the 16th and 84th percentiles.

Sample	$\log_{10}(\langle M_h \rangle / M_\odot)$	$f_{\text{sat}}$	$b$
$0 < z < 1$			
AGN	$13.44^{+0.08}_{-0.08}$	$0.18^{+0.16}_{-0.09}$	$1.94^{+0.07}_{-0.07}$
Matched	$13.25^{+0.08}_{-0.09}$	$0.18^{+0.18}_{-0.10}$	$1.69^{+0.07}_{-0.06}$
$1 < z < 1.5$			
AGN	$13.17^{+0.07}_{-0.06}$	$0.06^{+0.21}_{-0.05}$	$2.50^{+0.11}_{-0.18}$
Matched	$13.12^{+0.06}_{-0.06}$	$0.15^{+0.16}_{-0.12}$	$2.33^{+0.14}_{-0.19}$
$1.5 < z < 2.5$			
AGN	$13.03^{+0.09}_{-0.10}$	$0.06^{+0.09}_{-0.05}$	$3.38^{+0.27}_{-0.38}$
Matched	$12.77^{+0.14}_{-0.17}$	$0.05^{+0.08}_{-0.05}$	$2.79^{+0.32}_{-0.38}$

cated at the centres of their haloes (e.g. N. A. Hatch et al. 2014; W. Mo et al. 2018; J. H. Croston et al. 2019).

Table 3 presents the values of the derived parameters,  $\langle M_h \rangle$ ,  $f_{\text{sat}}$ , and  $b$ , for the AGN and matched galaxy samples. These were evaluated by sampling from the posterior distributions of the best-fitting HOD parameters, and a single value for each redshift bin was found by marginalizing  $\langle M_h \rangle(z)$ ,  $f_{\text{sat}}(z)$ , and  $b(z)$  over the redshift distribution of the sample. Although all three of the AGN HOD parameters are consistent between the redshift bins, the corresponding values of  $\langle M_h \rangle$  vary due to the evolving halo mass function. For  $1.5 < z < 2.5$ , we find  $\log_{10}(\langle M_h \rangle / M_\odot) = 13.03^{+0.09}_{-0.10}$  which increases to  $\log_{10}(\langle M_h \rangle / M_\odot) = 13.44^{+0.08}_{-0.08}$  by  $0 < z < 1$ . This suggests that radio-AGN in our sample typically reside in galaxy groups, which have halo masses of  $M_h \sim 10^{11} - 10^{14} M_\odot$ , rather than larger galaxy clusters with  $M_h \gtrsim 10^{14} M_\odot$  (S. H. Lim et al. 2017; V. F. Calderon & A. A. Berlind 2019). This is in agreement with studies that investigate galaxy densities around radio-galaxies (e.g. P. N. Best 2004; J. H. Croston et al. 2019).

At low redshift, our estimate of  $\langle M_h \rangle$  for  $0 < z < 1$  is in line with the findings of previous works. For example, R. C. Hickox et al. (2009) investigated the clustering of radio sources with  $L_{1.4 \text{ GHz}} > 10^{23.8} \text{ W Hz}^{-1}$  and  $0.25 < z < 0.8$  using the Westerbork Synthesis Radio Telescope (W. H. de Vries et al. 2002), and found radio-AGNs are strongly clustered with halo masses  $\log_{10}(M_h / M_\odot) \approx 13.3$  (converted to our cosmology using  $h = 0.7$ ). R. Mandelbaum et al. (2009) measured the clustering and lensing of radio-AGN with  $z \lesssim 0.3$  from the Faint Images of the Radio Sky at Twenty-Centimetres radio survey (FIRST; R. H. Becker, R. L. White & D. J. Helfand 1995), and found that they reside in less massive haloes with  $\log_{10}(M_h / M_\odot) \approx 13.0$ .

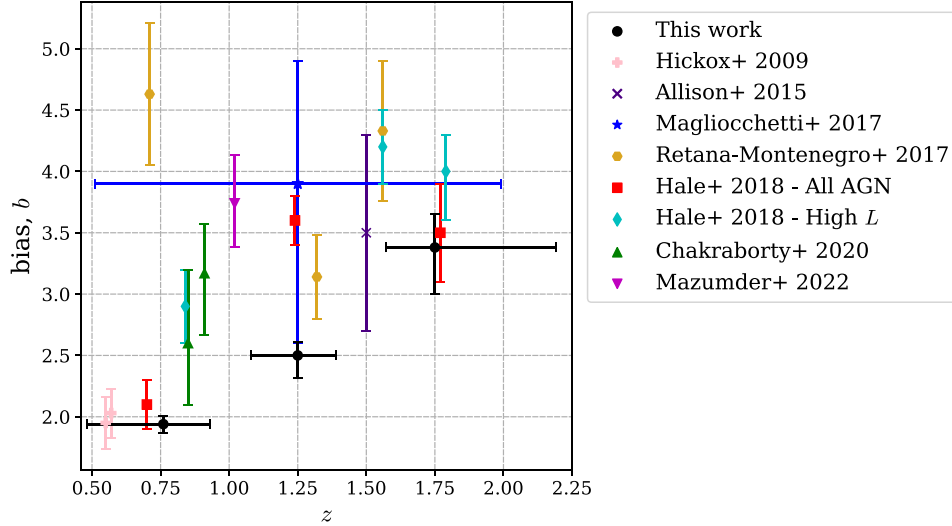
Our estimates of  $\langle M_h \rangle$  at higher redshift are similar to the findings of previous clustering studies, though they lie at the lower end of the reported range. For example, R. Allison et al. (2015) cross-correlated radio-AGN from FIRST with lensing of the cosmic microwave background and found a typical halo mass of  $\log_{10}(\langle M_h \rangle / M_\odot) = 13.6^{+0.3}_{-0.4}$  at a redshift of  $z \approx 1.5$ . This is consistent with our value at  $z_{\text{med}} = 1.25^{+0.14}_{-0.17}$  of  $\log_{10}(\langle M_h \rangle / M_\odot) = 13.17^{+0.07}_{-0.06}$ . E. Retana-Montenegro & H. J. A. Röttgering (2017) measured the projected correlation function of radio-loud quasars from FIRST over the redshift range  $0.3 <$

$z < 2.3$ , and found a higher average dark matter halo masses of  $\log(\langle M_h \rangle / M_\odot) \sim 13.5$  at  $z \sim 1.5$ .

One possible explanation for our slightly lower halo mass estimates is the difference in flux limits across surveys. Both of these studies adopted a flux density threshold of  $S_{1.4 \text{ GHz}} > 1 \text{ mJy}$ . Similarly, M. Magliocchetti et al. (2017) imposed a  $S_{1.4 \text{ GHz}} > 0.15 \text{ mJy}$  flux density limit. The minimum fluxes selected by our luminosity threshold for increasing redshift bins are  $S_{1.4 \text{ GHz}} > 0.18, 0.13$  and  $0.17 \text{ mJy}$ . Higher flux thresholds will preferentially select more luminous radio-AGN, especially at high redshift, than the deeper MIGHTEE sample (compare fig. 3 of E. Retana-Montenegro & H. J. A. Röttgering 2017 to our Fig. 2). Previous works have found that the clustering strength is correlated with luminosity (e.g. S. Bardelli et al. 2010; S. N. Lindsay et al. 2014; C. L. Hale et al. 2018; J. H. Croston et al. 2019; W. Mo et al. 2020), implying that more luminous AGN occupy more massive haloes. In support of this explanation, C. L. Hale et al. (2018), who measured the angular autocorrelation of radio sources from VLA observations of the COSMOS field at a lower flux limit of  $\sim 13 \mu\text{Jy beam}^{-1}$  at 3 GHz ( $0.02 \text{ mJy beam}^{-1}$  at 1.4 GHz assuming  $\alpha = -0.7$ ) also found lower halo mass estimates of  $\log_{10}(M_h / M_\odot) \sim 13.1 - 13.3$  for AGN with  $L_{3 \text{ GHz}} > 10^{23} \text{ W Hz}^{-1}$ . This would also explain the lower mass estimates of R. Mandelbaum et al. (2009) whose sample extended to  $L_{1.4 \text{ GHz}} \approx 10^{23} \text{ W Hz}^{-1}$  compared to our threshold of  $L_{\text{AGN}}(z = 0) \approx 10^{23.4} \text{ W Hz}^{-1}$ .

Our finding that the typical host halo mass of radio-AGN increases at later times is an important result of this work. This trend may reflect the greater abundance of cold gas for accretion at earlier epochs which would allow for radio-AGN activity to occur in lower-mass haloes. Such a scenario would account for the strong evolution in the comoving space number density of powerful radio sources (J. S. Dunlop & J. A. Peacock 1990; M. J. Jarvis et al. 2001b; E. E. Rigby et al. 2011). Previous clustering studies have found little evolution in the environments of radio-AGN out to  $z \sim 2.5$  (see Section 3.2 of M. Magliocchetti et al. 2017). Though C. L. Hale et al. (2018) observed a flattening of the bias parameter with increasing redshift which would suggest a decrease in typical host halo mass.

Fig. 8 presents our estimates of the AGN bias parameter as a function of redshift, showing a clear increase in bias towards earlier times. Included in the plot are results obtained by previous radio-AGN clustering studies at a range of observing frequencies. These include the results from R. C. Hickox et al. (2009), R. Allison et al. (2015), M. Magliocchetti et al. (2017), E. Retana-Montenegro & H. J. A. Röttgering (2017), and C. L. Hale et al. (2018). Also included are the bias estimates of A. Chakraborty et al. (2020), who investigated the clustering of radio-selected AGN at 400 MHz and 612 MHz using archival data from the upgraded Giant Metrewave Radio Telescope (uGMRT; A. Chakraborty et al. 2019); and A. Mazumder et al. (2022), who investigated the clustering of radio sources at 325 MHz using archival GMRT data (G. Swarup et al. 1991). Excluded from the plot are the anomalously high bias values of S. N. Lindsay et al. (2014) who cross-correlated VLA radio sources with optical and NIR galaxies and found  $b(z = 1.35) = 7.62 \pm 1.27$ ,  $b(z = 1.55) = 9.91 \pm 2.48$  and  $b(z = 1.77) = 11.14 \pm 3.01$  for luminosity thresholds of  $L_{1.4 \text{ GHz}} > 10^{23}, 10^{23.5},$  and  $10^{24} \text{ W Hz}^{-1}$ , respectively. The most directly comparable study to this work is that of C. L. Hale et al. (2018), which used a source sample of similar depth and comparable r.m.s. radio sensitivity to that of the MIGHTEE dataset.



**Figure 8.** Radio-AGN bias estimates from this work (black circles) and previous clustering studies as a function of redshift. These include bias estimates from: R. C. Hickox et al. 2009 (pink pluses); R. Allison et al. 2015 (purple cross); M. Magliocchetti et al. 2017 (blue star); E. Retana-Montenegro & H. J. A. Röttgering 2017 (gold hexagons); the full AGN population from C. L. Hale et al. 2018 (red squares), and high-luminosity AGN ( $L_{3\text{GHz}} > 10^{23} \text{ W Hz}^{-1}$ ; cyan diamonds); A. Chakraborty et al. 2020 (green triangles); and A. Mazumder et al. 2022 (downwards-pointing magenta triangles).

For our lowest redshift bin, we find a bias of  $b = 1.94^{+0.07}_{-0.07}$  at  $z_{\text{med}} = 0.76^{+0.17}_{-0.28}$ , in agreement with the value reported by C. L. Hale et al. (2018) of  $b = 2.1 \pm 0.2$  for AGN with  $z < 1$  at a comparable median redshift of  $z_{\text{med}} = 0.70$ . C. L. Hale et al. also report a bias of  $b = 2.9^{+0.3}_{-0.3}$  at  $z_{\text{med}} = 0.84$  for a high-luminosity subsample selected with  $L_{3\text{GHz}} > 10^{23} \text{ W Hz}^{-1}$ , equivalent to  $L_{1.4\text{GHz}} > 10^{23.2} \text{ W Hz}^{-1}$ . That luminosity cut lies below our  $L_{\text{AGN}}(z)$  threshold over the redshift range of the bin, so this is the population that is most similar to the sample used in this work at this redshift, and we find our measured bias to be lower than theirs. In the intermediate redshift range,  $1 < z < 1.5$ , we find  $b = 2.50^{+0.11}_{-0.18}$  at  $z_{\text{med}} = 1.25^{+0.14}_{-0.17}$  which is significantly lower than the  $b = 3.6 \pm 0.2$  at  $z_{\text{med}} = 1.24$  reported by C. L. Hale et al. for their full AGN sample across their whole redshift range ( $\Delta z \approx 4$ ). For our highest redshift bin,  $1.5 < z < 2.5$ , we obtain  $b = 3.38^{+0.27}_{-0.38}$  at  $z_{\text{med}} = 1.75^{+0.44}_{-0.18}$ , in agreement with their  $b = 3.5 \pm 0.4$  at  $z_{\text{med}} = 1.77$  for AGN with  $z \geq 1$ . Our result is also consistent with their high-luminosity sample bias value of  $b = 4.0^{+0.3}_{-0.4}$  at  $z_{\text{med}} = 1.79$ .

The lower bias values found in this work may stem from several factors. First, we used a narrower redshift bins which reduces the blending of potentially evolving clustering signals over cosmic time. Since radio-AGN are selected by their luminosity alone, another possibility is that the sample is contaminated with star-forming galaxies, which are known to be less clustered than AGN (e.g. M. Magliocchetti et al. 2017; C. L. Hale et al. 2018; A. Chakraborty et al. 2020). To test this, we experimented with increasing the luminosity threshold used to select AGN from the minimum luminosity above which 95 per cent of sources in I. H. Whittam et al. (2022) were classified as AGN, to the luminosity above which 100 per cent of sources were classified as AGN (see Section 2.6). No significant increase in the bias parameter was found.

C. L. Hale et al. (2018), A. Chakraborty et al. (2020), and A. Mazumder et al. (2022) all obtain their bias estimates by fitting a power law to the angular autocorrelation function of the form  $\omega(\theta) \propto \theta^{-0.8}$ . The most likely explanation for the lower bias values is that, instead of fitting a simple power law to the entire

clustering signal, in this work the one- and two-halo term contributions are modelled separately, resulting in a more accurate representation of the true clustering signal. Fitting a power-law (i.e. a straight line in  $\log \omega - \log \theta$ ) does not account for the dip in clustering amplitude at intermediate scales where the one- and two-halo regimes meet. Because the majority of the data points lie near the peaks of the one- and two-halo contributions, the power-law fit will be pulled up to match those measurements. This would result in a poorer fit at intermediate scales and an overestimation of the clustering amplitude, leading to a higher inferred bias.

#### 4.3.3 AGN duty cycle

In addition to constraining the halo mass and bias from clustering, the HOD framework can be used to estimate the AGN duty cycle. This is the fraction of time during which a SMBH is observable as a radio-AGN given the luminosity threshold. At any redshift, only a fraction  $f_{\text{DC}}$  of galaxies hosting a sufficiently massive SMBH will be radio-loud and included in the sample. Assuming that the fraction of active SMBHs is random, and not correlated with halo mass, this intermittency can be implemented in the HOD framework by rescaling the central galaxy occupation (e.g. T. Miyaji et al. 2011; M. Krumpel et al. 2015):

$$\bar{N}_c(M_h) \rightarrow f_{\text{DC}} \bar{N}_c(M_h). \quad (27)$$

However, introducing a mass-independent multiplicative factor simply renormalizes both the model pair counts and the mean number density. Pair counts in the one- and two-halo terms scale with the products of the occupation numbers (for example,  $\bar{N}_c \times \bar{N}_c$  or  $\bar{N}_c \times \bar{N}_s$ ; see Appendix A), so they pick up the same overall factor of  $f_{\text{DC}}$  as the mean galaxy number density (equation 17). Because the TPCF is normalized by the mean density (A. Nicola et al. 2020), this global factor cancels out, and the normalized clustering signal is unchanged. Consequently, a mass-independent  $f_{\text{DC}}$  cannot be constrained by fitting to the clustering measurements alone.

Instead,  $f_{\text{DC}}$  can be estimated from the number densities. A straightforward approach is to compare the ratio of the number density of observed AGN to that of potential host galaxies (i.e. galaxies in haloes massive enough to host such AGN). The observed number is found by summing the weights, while the number density of potential host galaxies is found by marginalizing equation (17) for the best-fitting AGN HOD parameters over the redshift distribution, such that

$$f_{\text{DC}} = \frac{\sum_i w_{\text{AGN},i}}{\bar{n}_{\text{AGN}}^{\text{mod}} V_{\text{tot}}}, \quad (28)$$

where  $V_{\text{tot}}$  is the total comoving volume covered by the fields between the limits of the redshift bin. The observed AGN number density can be directly taken from the weights provided the selection is volume-limited for the chosen luminosity threshold and redshift bin. In this case, the sample is complete across the bin and no AGN are missed purely due to sensitivity. If the selection of AGN was flux-limited, the completeness would vary with redshift, and the direct counts would underestimate the true number density. As evident in Fig. 2, the selection threshold,  $L_{\text{AGN}}(z)$ , remains above the survey flux limit at all redshifts in each bin.

One caveat of this method is the fact that approximately 10 per cent of the radio sources were not successfully cross-matched with their optical counterparts and are excluded from the sample, slightly lowering the duty cycle estimates. These unmatched sources are predominantly low-SNR detections, while the luminosity threshold primarily selects high-SNR sources (see Section 3.1.2), so the impact of this incompleteness is minimal.

Sampling from the HOD parameter posteriors, we find  $f_{\text{DC}} = 0.09^{+0.03}_{-0.03}$ ,  $0.08^{+0.04}_{-0.06}$ , and  $0.05^{+0.04}_{-0.03}$  for the  $0 < z < 1$ ,  $1 < z < 1.5$ , and  $1.5 < z < 2.5$  redshift bins, respectively. These estimates are consistent with the results of G. C. Petter et al. (2024), who calculated the duty cycle of low-frequency radio-galaxies with  $L_{150\text{MHz}} \gtrsim 10^{25.25} \text{ W Hz}^{-1}$  (equivalent to  $\gtrsim 10^{24.6} \text{ W Hz}^{-1}$  at 1.4 GHz) which is similar to  $L_{\text{AGN}}(z)$  in the  $1.5 < z < 2.5$  redshift bin. They found that the duty cycle increases from  $\sim 5$  per cent at  $z \sim 1.5$  up to  $\sim 10 - 20$  per cent for  $z \lesssim 1$ .

The fraction of galaxies that host radio-AGN is a strong function of galaxy mass, increasing as  $\propto M_*^{2.5}$  (e.g. P. N. Best et al. 2005) in the local Universe. It is therefore important to quote host galaxy stellar masses alongside duty cycle estimates. Sampling from the  $z$ -PDFs of radio sources and retaining only samples where the corresponding luminosity exceeds  $L_{\text{AGN}}(z)$ , we find median stellar masses for increasing redshift bins of  $\log_{10}(M_{*,\text{med}}/M_{\odot}) = 10.95^{+0.29}_{-0.52}$ ,  $10.82^{+0.34}_{-0.47}$  and  $10.82^{+0.35}_{-0.52}$ .

Given this strong stellar mass dependence, an effective duty cycle can also be inferred by calculating the ratio of the number of radio-AGN to the total number of galaxies within stellar mass ranges. At the median stellar masses of our samples, we find that this alternative approach yields duty cycle estimates that are lower by a factor of  $\sim 2 - 4$ , similar to the AGN fractions reported in previous studies at these masses (e.g. P. N. Best et al. 2005; J. Sabater et al. 2019; R. Kondapally et al. 2025). In Section 4.3.5, we show that stellar mass-matched control galaxies have different HOD parameters than AGN hosts, so AGN halo occupation is not determined by stellar mass alone. The denominator in the stellar mass-based duty cycle therefore includes galaxies in a larger range of halo environments, whereas the HOD estimate is restricted to the subset of haloes consistent with the clustering signal, leading to a higher inferred duty cycle.

The duty cycle can be used to calculate a characteristic time-scale,  $\tau$ , of the observable AGN phase. It can be estimated by multiplying the measured duty cycle by the cosmic time interval spanned by the redshift bin limits,  $\Delta t$ :  $\tau = f_{\text{DC}} \Delta t$ . This represents the total time that an AGN would be radiating above the luminosity threshold between the redshift limits of the bin, not the duration of a single episode of activity. We find  $\tau \approx 670$  Myr for  $0 < z < 1$  (which spans  $\Delta t = 7.7$  Gyr),  $\tau \approx 130$  Myr for  $1 < z < 1.5$  ( $\Delta t = 1.6$  Gyr) and  $\tau \approx 76$  Myr for  $1.5 < z < 2.5$  ( $\Delta t = 1.6$  Gyr). The sum of these characteristic time-scales of  $\sim 880$  Myr over  $0 < z < 2.5$  is similar to the value of  $\sim 1$  Gyr derived by M. Magliocchetti et al. (2017) for  $0 < z < 2.3$  and a similar luminosity threshold. Given that the lifetimes of radio-bright AGN are limited to a few  $\times 10$  Myr (K. M. Blundell & S. Rawlings 1999), this suggests that AGN undergo multiple radio-loud episodes during this period. Evidence for such recurrent activity is observed in ‘restarting’ AGN, where multiple generations of radio jets are seen within the same source (e.g. L. Lara et al. 1999; A. P. Schoenmakers et al. 2000; D. J. Saikia & M. Jamrozy 2009; S. Nandi & D. J. Saikia 2012; V. H. Mahatma 2023).

#### 4.3.4 Energy deposited by jet heating

We now quantify the energy deposited by radio-AGN into the surrounding gas via heating from their jets. The total kinetic energy emitted per halo over the time period  $\Delta t$  is roughly estimated as

$$\Delta E_{\text{kin}} = \frac{\int_{L_{\text{AGN}}(z)}^{\infty} d \log L \Psi(L, z)}{\int_0^{\infty} dM n(M, z) \bar{N}_c(M)} \times \Delta t, \quad (29)$$

where  $\Psi(L, z)$  is the kinetic heating rate density for radio-excess AGN from R. Kondapally et al. (2023), derived by combining the evolving luminosity functions of R. Kondapally et al. (2022) with the 1.4 GHz to kinetic jet power relation of T. M. Heckman & P. N. Best (2014). The latter is based on the results of L. Birzan et al. (2008) and K. W. Cavagnolo et al. (2010), and was calibrated by considering the work required to inflate X-ray cavities. The numerator in the fraction represents the total jet power density of AGN radiating above the luminosity selection threshold  $L_{\text{AGN}}(z)$ , while the denominator gives the comoving number density of haloes capable of hosting such AGN. Since the kinetic heating is dominated by sources with  $L_{1.4\text{GHz}} \gtrsim 10^{25.5} \text{ W Hz}^{-1}$  (R. Kondapally et al. 2023) which is well above  $L_{\text{AGN}}(z)$ , essentially all of the heating power is captured.

For our increasing redshift bins, we find  $\Delta E_{\text{kin}} \approx 4 \times 10^{53}$ ,  $1 \times 10^{53}$  and  $2 \times 10^{53}$  J per halo. Reproducing the observed entropy and thermal properties of the intracluster medium requires additional heating of  $0.5 - 1$  keV per gas particle beyond that of gravitational collapse (S. Borgani et al. 2002; A. V. Kravtsov & S. Borgani 2012). Assuming a halo mass of  $10^{13.3} M_{\odot}$ , a gas mass fraction of 5 per cent (P. Popesso et al. 2024), and  $N_{\text{gas}} = M_{\text{gas}}/(\mu_p m_p)$  with mean molecular mass  $\mu_p = 0.6$  for ionized gas (e.g. X.-P. Wu, Y.-J. Xue & L.-Z. Fang 1999) and proton mass  $m_p$ , this corresponds to  $2 - 3 \times 10^{53}$  J. Therefore, the energy provided by jets over  $0 < z < 2.5$  is sufficient to account for this additional heating.

#### 4.3.5 Comparison of the AGN and matched galaxy sample best-fitting HOD parameters

In this section, we compare the best-fitting HOD parameters and derived quantities between the AGN and matched galaxy sam-

ples. For the matched galaxy sample cross-correlations, our best-fitting models yield reduced chi-squared values of  $\chi^2_\nu = 0.84, 0.99$  and  $3.43$  for increasing redshift bins. We consider the higher  $\chi^2_\nu$  in the highest redshift bin acceptable given that we use a simplified three-parameter model, and fit a single, non-evolving set of HOD parameters over a broad redshift range,  $\Delta z = 1$ .

The only statistically significant difference in the individual HOD parameters between the two samples is in their  $M_{\min}$  values for the  $0 < z < 1$  redshift bin. For the AGN sample,  $\log_{10}(M_{\min}/M_\odot) = 12.81^{+0.09}_{-0.13}$  compared to  $\log_{10}(M_{\min}/M_\odot) = 12.52^{+0.12}_{-0.21}$  for the matched galaxy sample. Due to their large uncertainties, the satellite parameters,  $M_1$  and  $\alpha_s$ , are consistent within the uncertainties between the AGN and matched galaxy sample for all redshift bins.

Despite the overlap in the uncertainties of the individual HOD parameters, the values of the derived parameters given in Table 3 exhibit significant differences. This is because much of the uncertainty in  $M_{\min}$  and  $M_1$  is a result of their degeneracy, as evidenced by the elongated contours in the corner plots (see Appendix B). The largest difference in  $\langle M_h \rangle$  between the AGN and matched galaxy samples is in the highest redshift bin,  $1.5 < z < 2.5$ : for AGN,  $\log_{10}(\langle M_h \rangle / M_\odot) = 13.03^{+0.09}_{-0.10}$  compared to  $\log_{10}(\langle M_h \rangle / M_\odot) = 12.77^{+0.14}_{-0.17}$  for the matched galaxy sample. This means that the typical halo mass hosting galaxies with radio-AGN is approximately  $1.82^{+1.04}_{-0.57}$  times greater than that of non-active galaxies with similar stellar masses. The significant difference in  $\langle M_h \rangle$  persists for the lowest redshift bin,  $0 < z < 1$ . Here, the AGN sample gives  $\log_{10}(\langle M_h \rangle / M_\odot) = 13.44^{+0.08}_{-0.08}$  compared to  $\log_{10}(\langle M_h \rangle / M_\odot) = 13.25^{+0.08}_{-0.09}$  for the matched galaxy sample, differing by a factor of  $1.54^{+0.47}_{-0.33}$ . This is consistent with the results of R. Mandelbaum et al. (2009) who found that the dark matter halo masses of low redshift ( $z < 0.3$ ), radio-loud AGN were about twice as massive as those of a control sample matched in stellar mass. However, for the intermediate redshift bin,  $1 < z < 1.5$ , the two values of  $\langle M_h \rangle$  overlap within the uncertainties:  $\log_{10}(\langle M_h \rangle / M_\odot) = 13.17^{+0.07}_{-0.06}$  for AGN and  $\log_{10}(\langle M_h \rangle / M_\odot) = 13.12^{+0.06}_{-0.06}$  for the matched galaxy sample, giving a ratio of  $1.11^{+0.25}_{-0.20}$ . Though, it is worth noting that all three ratios are statistically consistent with each other within the uncertainties.

It is unclear whether the halo mass excess seen in AGN hosts is a driver of AGN activity, or a consequence of it. On one hand, radio-AGN activity may be enhanced in overdense structures. More massive haloes have deeper gravitational potential wells which will more efficiently channel gas toward the central galaxy, providing fuel for the SMBH. Such environments are also more likely to host mechanisms that promote gas inflows, such as galaxy interactions (e.g. M. S. Alonso et al. 2007; J. Sabater et al. 2013; A. D. Goulding et al. 2018) and mergers (e.g. N. Fanidakis et al. 2010; E. Treister et al. 2012; F. Gao et al. 2020; J. C. S. Pierce et al. 2022) which are known from simulations to be efficient at driving gas towards galaxy centres (V. Springel et al. 2005), triggering AGN activity.

On the other hand, the observed halo mass excess could be due to the cumulative effect of feedback; by suppressing star formation, jet-mode feedback will reduce stellar mass while leaving halo mass unchanged (D. J. Croton et al. 2006; R. G. Bower et al. 2006; T. M. Heckman & P. N. Best 2014; M. Hardcastle & J. Croston 2020; L. Scharré, D. Sorini & R. Davé 2024). As a result, AGN host galaxies would be assigned to a lower stellar mass bin despite residing in more massive halos. At fixed stellar mass, this effect would naturally produce a clustering excess

relative to non-active galaxies. A potential caveat is that the active lifetimes of radio-AGN are on the order of a few  $\times 10$  Myr (K. M. Blundell & S. Rawlings 1999) whereas galaxy quenching occurs on longer time-scales of  $\sim 500$  Myr (J. Lian et al. 2016). However, our estimate of the cumulative active lifetime of radio-AGN of  $\sim 1$  Gyr over  $0 < z < 2.5$  suggests that repeated cycles of activity could be sufficient to sustain this quenching process.

The higher derived halo masses may also reflect assembly bias, whereby clustering depends on properties beyond halo mass. In standard HOD modelling, it is assumed that the clustering of galaxies depends solely on halo mass, but in reality haloes of the same mass can differ in clustering strength depending on their formation history (L. Gao et al. 2005; R. H. Wechsler et al. 2006; D. J. Croton, L. Gao & S. D. M. White 2007; L. Gao & S. D. M. White 2007). At fixed mass, haloes that formed earlier have a higher clustering amplitude than haloes that formed more recently (J. F. Navarro et al. 1996; L. Gao et al. 2005; R. H. Wechsler et al. 2006). If AGN preferentially reside in such early-forming haloes, they would appear both more clustered and have had a longer period over which their central black holes could grow.

Distinguishing between these scenarios is challenging within the standard HOD framework, which assumes that clustering depends only on halo mass, and it is likely that a combination of these effects contributes to the observed difference.

An additional possibility is that differences in black hole mass contribute to the enhanced clustering of radio-AGN. There is some evidence that AGN powered by more massive black holes cluster more strongly (e.g. N. A. Hatch et al. 2014; E. Retana-Montenegro & H. J. A. Röttgering 2017). Since the fraction of galaxies that are radio-loud increases with black hole mass (e.g. R. J. McLure & M. J. Jarvis 2004; P. N. Best et al. 2005; R. B. Metcalf & M. Magliocchetti 2006; I. H. Whittam et al. 2022; C. L. Jackson et al. 2026), and our AGN sample is selected by radio luminosity, the observed clustering excess could therefore reflect systematic differences in black hole mass. However, because black hole mass correlates with host galaxy stellar mass (e.g. J. Magorrian et al. 1998; N. Häring & H.-W. Rix 2004; N. J. McConnell & C.-P. Ma 2013; J. Kormendy & L. C. Ho 2013), our stellar mass matching should largely remove this effect. Consequently, black hole mass would only play an additional role if jet power depends more strongly on black hole mass than black hole mass does on stellar mass.

We also note that it is likely that many of the galaxies in the matched sample also underwent radio-loud AGN activity at earlier epochs. This would also dilute any differences we are able to measure in the clustering of current radio-loud AGN and the radio-dormant AGN in the control sample. The true difference in the halo masses of stellar mass-matched galaxies with and without AGN may therefore be greater.

## 5 CONCLUSION

In this work, we compared the environments of radio-AGN to those of control galaxies matched in both stellar mass and redshift. Although it is well established that radio-AGN reside in the most massive dark matter haloes (M. Magliocchetti et al. 2004; R. Mandelbaum et al. 2009; R. Allison et al. 2015; M. Magliocchetti et al. 2017; E. Retana-Montenegro & H. J. A. Röttgering 2017; C. L. Hale et al. 2018), our aim was to determine whether this is solely a consequence of their preference for massive host galaxies, or

whether radio-AGN activity is linked to some additional environmental preference.

Using radio sources from the MIGHTEE radio survey that had been cross-matched with their optical/NIR counterparts from VISTA and HSC, we constructed a sample of  $\sim 2000$  AGN selected by their radio luminosities. A control sample of galaxies matched in both stellar mass and redshift was drawn from the optical/NIR galaxy catalogue, accounting for the full photometric redshift PDFs. These samples were divided into three redshift ranges:  $0 < z < 1$ ,  $1 < z < 1.5$  and  $1.5 < z < 2.5$ .

Measurements of the angular two-point autocorrelation functions showed no significant difference between the AGN and matched galaxy samples. This motivated our use of the cross-correlations with the full galaxy population, where the increased pair counts reduces statistical uncertainties, and we found that AGN are more clustered than the control sample.

By fitting a HOD model to the measured clustering signals, the statistical properties of the dark matter haloes that our samples inhabit were investigated. AGN were found to occupy dark matter haloes with masses  $\log_{10}(M_h/M_\odot) \gtrsim 12.8$ . Unlike some previous studies that report no significant evolution in the environmental properties of radio-AGN (e.g. M. Magliocchetti et al. 2017), we find a clear decrease in the typical halo mass of radio-AGN with redshift: for increasing redshift bins,  $\log_{10}((M_h)/M_\odot) = 13.44^{+0.08}_{-0.08}$ ,  $13.17^{+0.07}_{-0.06}$ , and  $13.03^{+0.09}_{-0.10}$ . We interpret this trend as a consequence of the higher abundance of cold gas at earlier epochs, which could more readily fuel AGN activity. This would also explain the strong evolution in the comoving space number density of powerful radio sources (J. S. Dunlop & J. A. Peacock 1990; M. J. Jarvis et al. 2001b; E. E. Rigby et al. 2011). The masses we obtain are broadly consistent with previous works (R. C. Hickox et al. 2009; R. Mandelbaum et al. 2009; R. Allison et al. 2015; M. Magliocchetti et al. 2017; E. Retana-Montenegro & H. J. A. Röttgering 2017; C. L. Hale et al. 2018), though lie at the lower end of mass estimates. This may be because MIGHTEE has a lower flux limit than previous surveys which would preferentially select lower-luminosity radio-AGN. Our halo masses suggest that radio-AGN typically reside in galaxy groups rather than higher mass clusters (S. H. Lim et al. 2017; V. F. Calderon & A. A. Berlind 2019).

For increasing redshift bins, we find bias parameters for radio-AGN of  $b = 1.94^{+0.07}_{-0.07}$ ,  $2.50^{+0.11}_{-0.18}$  and  $3.38^{+0.27}_{-0.38}$ . Comparing these bias estimates with those found by C. L. Hale et al. (2018), our measurements in the lowest and highest redshift bins are consistent with their values within the uncertainties. However, for the intermediate redshift bin, our value is significantly lower. We attribute this to the use of full HOD modelling, which explicitly models the one- and two-halo contributions to the clustering signal, rather than using a simple power-law fit.

By comparing the ratio of the number of observed radio-AGN to the number of galaxies in haloes massive enough to host a SMBH predicted by the HOD model, a duty cycle of  $\sim 5 - 9$  per cent is inferred. Given that the active lifetime of radio-AGN is limited to a few  $\times 10$  Myr (K. M. Blundell & S. Rawlings 1999), this implies that each AGN undergoes multiple periods of radio-loud activity over the redshift ranges considered. We estimate that the cumulative energy deposited into the intergalactic medium by radio jets since  $z = 2.5$  is  $\sim 10^{53}$  J. This is sufficient to account for the observed excess energy in the gas of galaxy groups assuming 0.5–1 keV of additional heating per gas particle above that of gravitational collapse.

Comparing the typical halo masses that we derive for the AGN and matched galaxy samples, we find that for  $0 < z < 1$ , the typical dark matter halo mass of radio-AGN is  $1.54^{+0.47}_{-0.33}$  times greater than that of galaxies matched in stellar mass and redshift. A similarly significant excess is seen at  $1.5 < z < 2.5$ , with a ratio of  $1.82^{+1.04}_{-0.57}$ . The halo mass excess of radio-AGN may be a result of AGN feedback which would suppress star formation and result in lower stellar masses for the same halo mass. Alternatively, it could indicate that galaxies hosting radio-AGN preferentially reside in earlier-forming and more clustered haloes, which have had more time to accrete gas and grow their central black hole.

The main limitation of this work is the number of AGN, which leads to large statistical uncertainties in the measured correlation functions. A larger sample would allow for tighter constraints on the HOD parameters and the use of narrower redshift bins to better trace the evolution of radio-AGN environments. It would also allow for a robust comparison between low- and high-luminosity AGN populations. Radio clustering studies have been limited by the lack of overlapping multiwavelength data that are required to obtain the redshifts of host galaxies. Current and upcoming wide-area radio surveys such as with the Low-Frequency Array (LOFAR; M. P. van Haarlem 2005), the Australian Square Kilometre Array Pathfinder radio telescope (ASKAP; R. P. Norris et al. 2011) and ultimately the Square Kilometre Array Observatory<sup>8</sup> (SKAO) will produce vastly increased source counts, with photometric redshift information provided by the next generation of surveys with LSST (LSST Science Collaboration 2009) and Euclid (Euclid Collaboration 2025) together with spectroscopic programmes like DESI (M. Levi et al. 2013) and WEAVE-LOFAR (D. J. B. Smith et al. 2016).

## ACKNOWLEDGEMENTS

JH acknowledges funding from the Science and Technology Facilities Council (STFC) [grant code ST/Y509474/1]. MJJ, CLH, and IHW acknowledge support from the Hintze Family Charitable Foundation through the Oxford Hintze Centre for Astrophysical Surveys. CLH also acknowledges support from STFC [ST/Y000951/1]. MJJ acknowledges the support of the STFC consolidated grant [ST/S000488/1] and [ST/W000903/1]. MJJ and NS acknowledge funding from a UK Research and Innovation (UKRI) Frontiers Research Grant [EP/X026639/1]. DA acknowledges support from the Beecroft Trust. We would like to thank the anonymous referee for their useful comments that have helped to improve this paper.

The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. We acknowledge the use of the ilifu cloud computing facility—[www.ilifu.ac.za](http://www.ilifu.ac.za), a partnership between the University of Cape Town, the University of the Western Cape, Stellenbosch University, Sol Plaatje University and the Cape Peninsula University of Technology. The Ilifu facility is supported by contributions from the Inter-University Institute for Data Intensive Astronomy (IDIA—a partnership between the University of Cape Town, the University of Pretoria and the University of the Western Cape, the Computational Biology division at UCT and the Data Intensive

<sup>8</sup><https://www.skao.int/en>

Research Initiative of South Africa (DIRISA). The authors acknowledge the Centre for High Performance Computing (CHPC), South Africa, for providing computational resources to this research project.

This work is based on data products from observations made with ESO Telescopes at the La Silla Paranal Observatory under ESO programme ID 179.A-2005 (Ultra-VISTA) and ID 179.A-2006(VIDEO) and on data products produced by CALET and the Cambridge Astronomy Survey Unit on behalf of the Ultra-VISTA and VIDEO consortia.

## DATA AVAILABILITY

The MIGHTEE continuum DR1 data used in this work are accessible through <https://doi.org/10.48479/7msw-r692> with information given in C. L. Hale et al. (2025). The cross-matched catalogues will be made available by Hale et al. (in preparation), and the multi-wavelength data catalogues will be made available by Stylianou et al. (in preparation). The results presented in this work can be obtained through a reasonable request to the author.

## REFERENCES

- Afshordi N., Loh Y.-S., Strauss M. A., 2004, *Phys. Rev. D*, 69, 083524
- Aihara H. et al., 2022, *PASJ*, 74, 247
- Allison R. et al., 2015, *MNRAS*, 451, 849
- Almosallam I. A., Jarvis M. J., Roberts S. J., 2016, *MNRAS*, 462, 726
- Alonso M. S., Lambas D. G., Tissera P., Coldwell G., 2007, *MNRAS*, 375, 1017
- Ando S., Benoit-Lévy A., Komatsu E., 2017, *MNRAS*, 473, 4318
- Antonucci R., 1993, *ARA&A*, 31, 473
- Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, *MNRAS*, 310, 540
- Arnouts S. et al., 2002, *MNRAS*, 329, 355
- Bardelli S. et al., 2010, *A&A*, 511, A1
- Becker R. H., White R. L., Helfand D. J., 1995, *ApJ*, 450, 559
- Berlind A. A., Weinberg D. H., 2002, *ApJ*, 575, 587
- Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
- Best P. N., 2004, *MNRAS*, 351, 70
- Best P. N., Kauffmann G., Heckman T. M., Brinchmann J., Charlot S., Ivezić V., White S. D. M., 2005, *MNRAS*, 362, 25
- Best P. N., Von Der Linden A., Kauffmann G., Heckman T. M., Kaiser C. R., 2007, *MNRAS*, 379, 894
- Best P. N. et al., 2023, *MNRAS*, 523, 1729
- Birzan L., McNamara B. R., Nulsen P. E. J., Carilli C. L., Wise M. W., 2008, *ApJ*, 686, 859
- Blake C., Wall J., 2002, *MNRAS*, 337, 993
- Blundell K. M., Rawlings S., 1999, *Nature*, 399, 330
- Borgani S., Governato F., Wadsley J., Menci N., Tozzi P., Quinn T., Stadel J., Lake G., 2002, *MNRAS*, 336, 409
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
- Brand K., Rawlings S., Hill G. J., Tufts J. R., 2005, *MNRAS*, 357, 1231
- Brown M. J. I., Jannuzi B. T., Floyd D. J. E., Mould J. R., 2011, *ApJ*, 731, L41
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Bryan G. L., Norman M. L., 1998, *ApJ*, 495, 80
- Calderon V. F., Berlind A. A., 2019, *MNRAS*, 490, 2367
- Capetti A. et al., 2022, *A&A*, 660, A93
- Castignani G., Chiaberge M., Celotti A., Norman C., De Zotti G., 2014, *ApJ*, 792, 114
- Cavagnolo K. W., McNamara B. R., Nulsen P. E. J., Carilli C. L., Jones C., Birzan L., 2010, *ApJ*, 720, 1066
- Chakraborty A. et al., 2019, *MNRAS*, 487, 4102
- Chakraborty A., Dutta P., Datta A., Roy N., 2020, *MNRAS*, 494, 3392
- Chisari N. E. et al., 2019, *ApJS*, 242, 2
- Chon G., Challinor A., Prunet S., Hivon E., Szapudi I., 2004, *MNRAS*, 350, 914
- Cicone C. et al., 2014, *A&A*, 562, A21
- Coil A. L., 2013, in Oswalt T. D., Keel W. C., eds, *Planets, Stars and Stellar Systems. Volume 6: Extragalactic Astronomy and Cosmology*. Springer, Dordrecht, Netherlands, p. 387
- Cooray A., Sheth R., 2002, *Phys. Rep.*, 372, 1
- Cress C. M., Helfand D. J., Becker R. H., Gregg M. D., White R. L., 1996, in Trimble V., Reisenegger A., eds, *ASP Conf. Ser. Vol. 88, Clusters, Lensing, and the Future of the Universe*. Astron. Soc. Pac., San Francisco, p. 193
- Croston J. H. et al., 2019, *A&A*, 622, A10
- Croton D. J. et al., 2006, *MNRAS*, 365, 11
- Croton D. J., Gao L., White S. D. M., 2007, *MNRAS*, 374, 1303
- Cuillandre J.-C. J. et al., 2012, in Peck A. B., Seaman R. L., Comeron F., eds, *Proc. SPIE Conf. Ser. Vol. 8448, Observatory Operations: Strategies, Processes, and Systems IV*. SPIE, Bellingham, p. 84480M
- Dahlen T. et al., 2013, *ApJ*, 775, 93
- Davis M., Peebles P. J. E., 1983, *ApJ*, 267, 465
- de Ruiter H. R., Willis A. G., Arp H. C., 1977, *A&AS*, 28, 211
- de Vries W. H., Morganti R., Röttgering H. J. A., Vermeulen R., van Breugel W., Rengelink R., Jarvis M. J., 2002, *AJ*, 123, 1784
- Despali G., Giocoli C., Angulo R. E., Tormen G., Sheth R. K., Baso G., Moscardini L., 2016, *MNRAS*, 456, 2486
- Donoso E., Li C., Kauffmann G., Best P. N., Heckman T. M., 2010, *MNRAS*, 407, 1078
- Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., 2008, *MNRAS*, 390, L64
- Duncan K. J. et al., 2018, *MNRAS*, 473, 2655
- Duncan K. J. et al., 2019, *A&A*, 622, A3
- Dunlop J. S., Peacock J. A., 1990, *MNRAS*, 247, 19
- Eales S., Rawlings S., Law-Green D., Cotter G., Lacy M., 1997, *MNRAS*, 291, 593
- Eckert D., Gaspari M., Gastaldello F., Le Brun A. M. C., O’Sullivan E., 2021, *Universe*, 7, 142
- Endsley R. et al., 2022, *MNRAS*, 512, 4248
- Euclid Collaboration, 2025, *A&A*, 697, A1
- Fabian A. C., 2012, *ARA&A*, 50, 455
- Fanidakis N., Baugh C. M., Benson A. J., Bower R. G., Cole S., Done C., Frenk C. S., 2010, *MNRAS*, 410, 53
- Fedeli C., Semboloni E., Velliscig M., Daalen M. V., Schaye J., Hoekstra H., 2014, *J. Cosmol. Astropart. Phys.*, 2014, 028
- Fine S., Shanks T., Nikoloudakis N., Sawangwit U., 2011, *MNRAS*, 418, 2251
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306
- Fortson L. et al., 2012, in Way M. J., Scargle J. D., Ali K. M., Srivastava A. N., eds, *Advances in Machine Learning and Data Mining for Astronomy*. Chapman & Hall, UK, p. 213
- Gao L., White S. D. M., 2007, *MNRAS*, 377, L5
- Gao L., Springel V., White S. D. M., 2005, *MNRAS*, 363, L66
- Gao F. et al., 2020, *A&A*, 637, A94
- Gaspari M., Tombesi F., Cappi M., 2020, *Nat. Astron.*, 4, 10
- Gilli R. et al., 2019, *A&A*, 632, A26
- Gitti M., Brighenti F., McNamara B. R., 2012, *Adv. Astron.*, 2012, 950641
- Gomes Z., Jarvis M. J., Almosallam I. A., Roberts S. J., 2018, *MNRAS*, 475, 331
- Goulding A. D. et al., 2018, *PASJ*, 70, S37
- Grossová R. et al., 2022, *ApJS*, 258, 30
- Groth E. J., Peebles P. J. E., 1977, *ApJ*, 217, 385
- Gürkan G., Hardcastle M. J., Jarvis M. J., 2014, *MNRAS*, 438, 1149
- Hale C. L., Jarvis M. J., Delvecchio I., Hatfield P. W., Novak M., Smolčić V., Zamorani G., 2018, *MNRAS*, 474, 4133
- Hale C. L. et al., 2024, *MNRAS*, 527, 6540
- Hale C. L. et al., 2025, *MNRAS*, 536, 2187
- Hardcastle M., Croston J., 2020, *New Astron. Rev.*, 88, 101539

- Häring N., Rix H.-W., 2004, *ApJ*, 604, L89
- Harrison C. M., Alexander D. M., Mullaney J. R., Swinbank A. M., 2014, *MNRAS*, 441, 3306
- Harrison C. M., Costa T., Tadhunter C. N., Flütsch A., Kakkad D., Perna M., Vietri G., 2018, *Nat. Astron.*, 2, 198
- Hatch N. A. et al., 2014, *MNRAS*, 445, 280
- Hatfield P. W., Lindsay S. N., Jarvis M. J., Häußler B., Vaccari M., Verma A., 2016, *MNRAS*, 459, 2618
- Hatfield P. W., Jarvis M. J., Adams N., Bowler R. A. A., Häußler B., Duncan K. J., 2022, *MNRAS*, 513, 3719
- Heckman T. M., Best P. N., 2014, *ARA&A*, 52, 589
- Heywood I. et al., 2022, *MNRAS*, 509, 2150
- Hickox R. C. et al., 2009, *ApJ*, 696, 891
- Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006, *ApJS*, 163, 1
- Hugo B. V., Perkins S., Merry B., Mauch T., Smirnov O. M., 2022, in Ruiz J. E., Pierfederici F., Teuben P., eds, ASP Conf. Ser. Vol. 532, Astronomical Data Analysis Software and Systems XXX. Astron. Soc. Pac., San Francisco, p. 541
- Ilbert O. et al., 2006, *A&A*, 457, 841
- Jackson C. L., Matthews J. H., Whittam I. H., Jarvis M. J., Temple M. J., Rankine A. L., Hewett P. C., 2026, *MNRAS*, 546, 3
- Jarvis M. J., McLure R. J., 2002, *MNRAS*, 336, L38
- Jarvis M. J., Rawlings S., Eales S., Blundell K. M., Bunker A. J., Croft S., McLure R. J., Willott C. J., 2001a, *MNRAS*, 326, 1585
- Jarvis M. J., Rawlings S., Willott C. J., Blundell K. M., Eales S., Lacy M., 2001b, *MNRAS*, 327, 907
- Jarvis M., Bernstein G., Jain B., 2004, *MNRAS*, 352, 338
- Jarvis M. J., Teimourian H., Simpson C., Smith D. J. B., Rawlings S., Bonfield D., 2009, *MNRAS*, 398, L83
- Jarvis M. J. et al., 2013, *MNRAS*, 428, 1281
- Jarvis M. et al., 2016, *Proc. Sci.*, Vol. 277, *MeerKAT Science: On the Pathway to the SKA*. SISSA, Trieste, PoS(MeerKAT2016)006
- Kaiser N., 1984, *ApJ*, 284, L9
- Kauffmann G., Heckman T. M., Best P. N., 2008, *MNRAS*, 384, 953
- Kodama T., Bower R., 2003, *MNRAS*, 346, 1
- Kolwa S., Jarvis M. J., McAlpine K., Heywood I., 2019, *MNRAS*, 482, 5156
- Kondapally R. et al., 2021, *A&A*, 648, A3
- Kondapally R. et al., 2022, *MNRAS*, 513, 3742
- Kondapally R. et al., 2023, *MNRAS*, 523, 5292
- Kondapally R. et al., 2025, *MNRAS*, 536, 554
- Kormendy J., Ho L. C., 2013, *ARA&A*, 51, 511
- Kravtsov A. V., Borgani S., 2012, *ARA&A*, 50, 353
- Krumpe M., Miyaji T., Husemann B., Fanidakis N., Coil A. L., Aceves H., 2015, *ApJ*, 815, 21
- LSST Science Collaboration, 2009, preprint (arXiv:0912.0201)
- Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
- Lara L., Márquez I., Cotton W. D., Feretti L., Giovannini G., Marcaide J. M., Venturi T., 1999, *A&A*, 348, 699
- Levi M. et al., 2013, preprint (arXiv:1308.0847)
- Lian J., Yan R., Zhang K., Kong X., 2016, *ApJ*, 832, 29
- Lim S. H., Mo H. J., Lu Y., Wang H., Yang X., 2017, *MNRAS*, 470, 2982
- Limber D. N., 1954, *ApJ*, 119, 655
- Lindsay S. N., Jarvis M. J., McAlpine K., 2014, *MNRAS*, 440, 2322
- Lintott C. J. et al., 2008, *MNRAS*, 389, 1179
- Magliocchetti M., 2022, *A&A Rev.*, 30, 6
- Magliocchetti M. et al., 2002, *MNRAS*, 333, 100
- Magliocchetti M. et al., 2004, *MNRAS*, 350, 1485
- Magliocchetti M. et al., 2014, *MNRAS*, 442, 682
- Magliocchetti M., Lutz D., Santini P., Salvato M., Popesso P., Berta S., Pozzi F., 2016, *MNRAS*, 456, 431
- Magliocchetti M., Popesso P., Brusa M., Salvato M., Laigle C., McCracken H. J., Ilbert O., 2017, *MNRAS*, 464, 3271
- Magliocchetti M. et al., 2020, *MNRAS*, 493, 3838
- Magorrian J. et al., 1998, *AJ*, 115, 2285
- Mahatma V. H., 2023, *Galaxies*, 11, 74
- Malavasi N., Bardelli S., Ciliegi P., Ilbert O., Pozzetti L., Zucca E., 2015, *A&A*, 576, A101
- Mandelbaum R., Li C., Kauffmann G., White S. D. M., 2009, *MNRAS*, 393, 377
- Mauch T., Sadler E. M., 2007, *MNRAS*, 375, 931
- Mazumder A., Chakraborty A., Datta A., 2022, *MNRAS*, 517, 3407
- McAlpine K., Smith D. J. B., Jarvis M. J., Bonfield D. G., Fleuren S., 2012, *MNRAS*, 423, 132
- McAlpine K., Jarvis M. J., Bonfield D. G., 2013, *MNRAS*, 436, 1084
- McCarthy I. G. et al., 2010, *MNRAS*, 406, 822
- McConnell N. J., Ma C.-P., 2013, *ApJ*, 764, 184
- McCracken H. J. et al., 2012, *A&A*, 544, A156
- McCracken H. J. et al., 2015, *MNRAS*, 449, 901
- McLure R. J., Jarvis M. J., 2004, *MNRAS*, 353, L45
- McNamara B. R., Nulsen P. E. J., 2012, *New J. Phys.*, 14, 055023
- Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, *MNRAS*, 454, 1958
- Metcalfe R. B., Magliocchetti M., 2006, *MNRAS*, 365, 101
- Miyaji T., Krumpe M., Coil A. L., Aceves H., 2011, *ApJ*, 726, 83
- Mo H. J., White S. D. M., 1996, *MNRAS*, 282, 347
- Mo W. et al., 2018, *ApJ*, 869, 131
- Mo W. et al., 2020, *ApJ*, 901, 131
- Moster B. P., Somerville R. S., Newman J. A., Rix H.-W., 2011, *ApJ*, 731, 113
- Nandi S., Saikia D. J., 2012, *Bull. Astron. Soc. India*, 40, 121
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Ng K.-W., Liu G.-C., 1999, *Int. J. Mod. Phys. D*, 8, 61
- Nicola A. et al., 2020, *J. Cosmol. Astropart. Phys.*, 2020, 044
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *MNRAS*, 396, 19
- Norris R. P. et al., 2011, *PASA*, 28, 215
- Oke J. B., Gunn J. E., 1983, *ApJ*, 266, 713
- Overzier R. A., Röttgering H. J. A., Rengelink R. B., Wilman R. J., 2003, *A&A*, 405, 53
- Padovani P. et al., 2017, *A&A Rev.*, 25, 2
- Peacock J. A., 1999, *Cosmological Physics*. Cambridge Univ. Press, Cambridge, UK
- Peebles P. J. E., 1980, *The Large-scale Structure of the Universe*. Princeton University Press, Princeton, USA
- Petter G. C., Hickox R. C., Morabito L. K., Alexander D. M., 2024, *ApJ*, 972, 184
- Pierce J. C. S. et al., 2022, *MNRAS*, 510, 1163
- Popesso P. et al., 2024, preprint (arXiv:2411.16555)
- Pozzetti L. et al., 2010, *A&A*, 523, A13
- Quenouille M. H., 1956, *Biometrika*, 43, 353
- Rawlings S., Jarvis M. J., 2004, *MNRAS*, 355, L9
- Retana-Montenegro E., Röttgering H. J. A., 2017, *A&A*, 600, A97
- Rigby E. E., Best P. N., Brookes M. H., Peacock J. A., Dunlop J. S., Röttgering H. J. A., Wall J. V., Ker L., 2011, *MNRAS*, 416, 1900
- Roche N., Eales S. A., 1999, *MNRAS*, 307, 703
- Sabater J., Best P. N., Argudo-Fernández M., 2013, *MNRAS*, 430, 638
- Sabater J. et al., 2019, *A&A*, 622, A17
- Saikia D. J., Jamrozny M., 2009, *Bull. Astron. Soc. India*, 37, 63
- Salcedo A. N., Wibking B. D., Weinberg D. H., Wu H.-Y., Ferrer D., Eisenstein D., Pinto P., 2020, *MNRAS*, 491, 3061
- Scharré L., Sorini D., Davé R., 2024, *MNRAS*, 534, 361
- Schoenmakers A. P., de Bruyn A. G., Röttgering H. J. A., van der Laan H., Kaiser C. R., 2000, *MNRAS*, 315, 371
- Smith D. J. B. et al., 2016, in Reylé C., Richard J., Cambrésy L., Deleuil M., Pécontal E., Tresse L., Vauglin I., eds, *SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*. Centre de Recherche Astrophysique de Lyon, p. 271
- Smolčić V. et al., 2009, *ApJ*, 696, 24
- Smolčić V., Finoguenov A., Zamorani G., Schinnerer E., Tanaka M., Gioldini S., Scoville N., 2011, *MNRAS*, 416, L31
- Smolčić V. et al., 2017, *A&A*, 602, A2
- Springel V., Di Matteo T., Hernquist L., 2005, *ApJ*, 620, L79
- Stylianou N., Malz A. I., Hatfield P., Crenshaw J. F., Gschwend J., 2022, *PASP*, 134, 044501
- Sutherland W., Saunders W., 1992, *MNRAS*, 259, 413

- Swarup G., Ananthakrishnan S., Kapahi V. K., Rao A. P., Subrahmanya C. R., Kulkarni V. K., 1991, *Current Science*, 60, 95
- Szapudi I., Szalay A. S., 1998, *ApJ*, 494, L41
- Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, *ApJ*, 724, 878
- Totsuji H., Kihara T., 1969, *PASJ*, 21, 221
- Treister E., Schawinski K., Urry C. M., Simmons B. D., 2012, *ApJ*, 758, L39
- Uchiyama H. et al., 2022, *ApJ*, 926, 76
- Vaccari M., 2015, *Proc. Sci.*, Vol. 267, The Spitzer Data Fusion: Contents, Construction and Applications to Galaxy Evolution Studies. SISSA, Trieste, PoS(EXTRA-RADSUR2015)027
- Vaccari M. et al., 2016, *Proc. Sci.* Vol. 275, The VOICE Survey : VST Optical Imaging of the CDFS and ES1 Fields. SISSA, Trieste, PoS(HEASA 2016)026
- van Haarlem M. P., 2005, in Gurvits L. I., Frey S., Rawlings S., eds, *EAS Publications Series Vol. 15, EAS Publications Series*. Cambridge Univ. Press, Cambridge, UK, p. 431
- Wake D. A., Croom S. M., Sadler E. M., Johnston H. M., 2008, *MNRAS*, 391, 1674
- Wang Y., Brunner R. J., Dolence J. C., 2013, *MNRAS*, 432, 1961
- Wechsler R. H., Zentner A. R., Bullock J. S., Kravtsov A. V., Allgood B., 2006, *ApJ*, 652, 71
- Whittam I. H. et al., 2022, *MNRAS*, 516, 245
- Whittam I. H. et al., 2024, *MNRAS*, 527, 3231
- Wilman R. J. et al., 2008, *MNRAS*, 388, 1335
- Wilman R. J., Jarvis M. J., Mauch T., Rawlings S., Hickey S., 2010, *MNRAS*, 405, 447
- Wu X.-P., Xue Y.-J., Fang L.-Z., 1999, *ApJ*, 524, 22
- Wylezalek D. et al., 2013, *ApJ*, 769, 79
- Zehavi I. et al., 2005, *ApJ*, 621, 22
- Zheng Z. et al., 2005, *ApJ*, 633, 791

## APPENDIX A: CROSS-CORRELATION MODEL

In this section, we discuss the halo model for the cross-correlation between the AGN (or matched galaxy) sample, labelled with superscript A, and the full optical/NIR population, B. It is assumed that sample A is a subset of sample B, so while all central galaxies in A are also in B, only a fraction  $f_c(M)$  of central galaxies in B are also in A. Similarly, only a fraction  $f_s(M)$  of satellite galaxies in B are also in A, and all satellites in A are in B.

The full halo model power spectrum is the sum of the one- and two-halo contributions (A. Cooray & R. Sheth 2002):

$$P_{\text{hm}}(k | z) = P_{\text{1h}}(k | z) + P_{\text{2h}}(k | z). \quad (\text{A1})$$

The two-halo term measures the clustering between galaxies that reside in different haloes, so depends only on the large-scale bias of each sample. It therefore remains unchanged from the standard form:

$$P_{\text{2h}}(k | z) = \frac{1}{\bar{n}_g^A \bar{n}_g^B} I(k, z | U^A) I(k, z | U^B) P_{\text{lin}}(k | z), \quad (\text{A2})$$

where  $\bar{n}_g^A$  and  $\bar{n}_g^B$  are the predicted mean number densities of galaxies in each sample given by equation (17),  $P_{\text{lin}}(k | z)$  is the linear matter power spectrum, and the integral

$$I(k, z | U) = \int dM n(M, z) b_{\text{h}}(M, z) \langle U(k | M) \rangle. \quad (\text{A3})$$

$\langle U(k | M) \rangle$  is the Fourier transform of the mean galaxy density profile for a halo of mass  $M$ :

$$\begin{aligned} \langle n_g(r | M) \rangle &= \bar{N}_c(M) [1 + \bar{n}_s(r | M)] \\ &= \bar{N}_c(M) [1 + \bar{N}_s(M) u_s(r | M)], \end{aligned} \quad (\text{A4})$$

where the normalized distribution of satellites,  $u_s(r | M)$ , is a function of radial distance from the halo centre,  $r$ . The distribu-

tion of satellites in a halo is assumed to follow that of the dark matter, which is modelled as a truncated NFW (J. F. Navarro et al. 1996) profile (see A. Nicola et al. 2020 for the full parametrization). Then

$$\begin{aligned} \langle U(k | M) \rangle &= \bar{N}_c(M) [1 + \bar{n}_s(k | M)] \\ &= \bar{N}_c(M) [1 + \bar{N}_s(M) \bar{u}_s(k | M)], \end{aligned} \quad (\text{A5})$$

where  $\bar{u}_s(k | M)$  is the Fourier transform of  $u_s(r | M)$ .

The one-halo term of the cross-correlation power spectrum accounts for pairs of galaxies that reside in the same halo. It is given by

$$P_{\text{1h}}(k | z) = \frac{1}{\bar{n}_g^A \bar{n}_g^B} \int dM n(M, z) \langle U^A(k | M) U^B(k | M) \rangle. \quad (\text{A6})$$

If the two samples were drawn completely independently of each other, then their joint two-point Fourier moment would factorize:

$$\begin{aligned} \langle U^A(k | M) U^B(k | M) \rangle &= \langle U^A(k | M) \rangle \langle U^B(k | M) \rangle \\ &= \bar{N}_c^A(M) \bar{N}_c^B(M) [(\bar{N}_s^A(M) + \bar{N}_s^B(M)) \bar{u}_s(k | M) \\ &\quad + \bar{N}_s^A(M) \bar{N}_s^B(M) \bar{u}_s^2(k | M)]. \end{aligned} \quad (\text{A7})$$

However, since the samples overlap, this cannot be assumed to be the case.

To derive  $\langle U^A U^B \rangle$  for the case where A is a subset of B, we divide the halo into small volume elements,  $\Delta V$ , with coordinates  $X$ . In the following, we drop the explicit  $M$  notation for brevity. The Fourier transforms then become discrete sums:

$$\langle U^A(k) U^B(k) \rangle = \frac{1}{(\Delta V)^2} \sum_{X, X'} e^{ik(X-X')} (\Delta V)^2 \langle N_X^A N_{X'}^B \rangle. \quad (\text{A8})$$

$\langle N_X^A N_{X'}^B \rangle$  can be decomposed into the contributions from central-central, central-satellite, satellite-central, and satellite-satellite galaxy pairs:

$$\begin{aligned} \langle N_X^A N_{X'}^B \rangle &= \langle N_{\text{cX}}^A N_{\text{cX}'}^B \rangle + \langle N_{\text{cX}}^A N_{\text{sX}'}^B \rangle + \langle N_{\text{sX}}^A N_{\text{cX}'}^B \rangle + \langle N_{\text{sX}}^A N_{\text{sX}'}^B \rangle, \end{aligned} \quad (\text{A9})$$

where, for example,  $N_{\text{cX}}^A \equiv N_c^A(X) \in \{0, 1\}$  denotes the number of central galaxies from sample A in the volume element at position  $X$ , with analogous definitions for satellite galaxies and for sample B.

To evaluate these four terms, we introduce  $\mathcal{P}(N_X)$ , the probability distribution for finding  $N_X$  galaxies in the volume element at position  $X$ . For central galaxies, the expectation value is

$$\sum_{N_{\text{cX}}=0}^1 N_{\text{cX}} \mathcal{P}(N_{\text{cX}}) = \langle N_{\text{cX}} \rangle = \delta_{X0}^K \bar{N}_c, \quad (\text{A10})$$

where  $\bar{N}_c$  is the mean number of central galaxies per halo from equation (10), and the Kronecker-delta function,  $\delta_{X0}^K$ , reflects the assumption that central galaxies reside at the centre of their halo,  $X = 0$ .

For satellite galaxies, the expectation value,

$$\sum_{N_{\text{sX}}=0}^{\infty} N_{\text{sX}} \mathcal{P}(N_{\text{sX}}) = \langle N_{\text{sX}} \rangle = \bar{N}_c \bar{n}_s(X) = \bar{N}_c \bar{N}_s u_s(X), \quad (\text{A11})$$

where  $\bar{N}_s$  is the mean number of satellite galaxies in a halo with a central galaxy from equation (12),  $u_s(X)$  is the normalized spatial profile of satellite galaxies within the halo, and  $\bar{n}_s(X) \equiv \bar{N}_s u_s(X)$ .

The central–central term is given by

$$\begin{aligned}
\langle N_{\text{cX}}^{\text{A}} N_{\text{cX}'}^{\text{B}} \rangle &= \delta_{\text{XX}'}^{\text{K}} \delta_{\text{X0}}^{\text{K}} \sum_{N_{\text{c}}^{\text{A}}, N_{\text{c}}^{\text{B}}} N_{\text{c}}^{\text{A}} N_{\text{c}}^{\text{B}} \mathcal{P}(N_{\text{c}}^{\text{A}}, N_{\text{c}}^{\text{B}}) \\
&= \delta_{\text{XX}'}^{\text{K}} \delta_{\text{X0}}^{\text{K}} \sum_{N_{\text{c}}^{\text{B}}} N_{\text{c}}^{\text{B}} \mathcal{P}(N_{\text{c}}^{\text{B}}) \sum_{N_{\text{c}}^{\text{A}}} N_{\text{c}}^{\text{A}} \mathcal{P}(N_{\text{c}}^{\text{A}} | N_{\text{c}}^{\text{B}}) \\
&= \delta_{\text{XX}'}^{\text{K}} \delta_{\text{X0}}^{\text{K}} \sum_{N_{\text{c}}^{\text{B}}} N_{\text{c}}^{\text{B}} \mathcal{P}(N_{\text{c}}^{\text{B}}) f_{\text{c}} N_{\text{c}}^{\text{B}} \\
&= \delta_{\text{XX}'}^{\text{K}} \delta_{\text{X0}}^{\text{K}} f_{\text{c}} \sum_{N_{\text{c}}^{\text{B}}} N_{\text{c}}^{\text{B}} \mathcal{P}(N_{\text{c}}^{\text{B}}) \\
&= \delta_{\text{XX}'}^{\text{K}} \delta_{\text{X0}}^{\text{K}} f_{\text{c}} \bar{N}_{\text{c}}^{\text{B}},
\end{aligned} \tag{A12}$$

where between the third and fourth lines the identity  $N_{\text{c}}^2 = N_{\text{c}}$  has been used, since  $N_{\text{c}} \in \{0, 1\}$ .

The satellite–satellite term is given by

$$\begin{aligned}
\langle N_{\text{sX}}^{\text{A}} N_{\text{sX}'}^{\text{B}} \rangle &= \sum_{N_{\text{sX}}^{\text{A}}, N_{\text{sX}'}^{\text{B}}} N_{\text{sX}}^{\text{A}} N_{\text{sX}'}^{\text{B}} \mathcal{P}(N_{\text{sX}}^{\text{A}}, N_{\text{sX}'}^{\text{B}}) \\
&= \sum_{N_{\text{sX}}^{\text{A}}, N_{\text{sX}'}^{\text{B}}} N_{\text{sX}'}^{\text{B}} \mathcal{P}(N_{\text{sX}'}^{\text{B}}) N_{\text{sX}}^{\text{A}} \mathcal{P}(N_{\text{sX}}^{\text{A}} | N_{\text{sX}'}^{\text{B}}).
\end{aligned} \tag{A13}$$

If  $X \neq X'$  then the two satellites are distinct, and the probability of finding a satellite at  $X$  from sample A is independent of the probability of finding a satellite at  $X'$  from sample B. However, if  $X = X'$  then they are the same galaxy, and in order for there to be a satellite at  $X$  in sample A, there must also be a satellite from sample B at  $X$ . Therefore,

$$\mathcal{P}(N_{\text{sX}}^{\text{A}} | N_{\text{sX}'}^{\text{B}}) = \begin{cases} \mathcal{P}(N_{\text{sX}}^{\text{A}}), & \text{if } X \neq X' \\ \mathcal{P}(N_{\text{sX}}^{\text{A}} | N_{\text{sX}}^{\text{B}}), & \text{if } X = X' \end{cases} \tag{A14}$$

Separating these two cases, and making use of the fact that satellite galaxies are Poisson distributed so the variance  $\langle N_{\text{s}}^2 \rangle - \langle N_{\text{s}} \rangle^2 = \langle N_{\text{s}} \rangle$ ,

$$\begin{aligned}
\langle N_{\text{sX}}^{\text{A}} N_{\text{sX}'}^{\text{B}} \rangle &= \sum_{N_{\text{sX}}^{\text{A}}} N_{\text{sX}}^{\text{A}} \mathcal{P}(N_{\text{sX}}^{\text{A}}) \sum_{N_{\text{sX}'}^{\text{B}}} N_{\text{sX}'}^{\text{B}} \mathcal{P}(N_{\text{sX}'}^{\text{B}}) (1 - \delta_{\text{XX}'}^{\text{K}}) \\
&\quad + \delta_{\text{XX}'}^{\text{K}} \sum_{N_{\text{sX}}^{\text{B}}} N_{\text{sX}}^{\text{B}} \mathcal{P}(N_{\text{sX}}^{\text{B}}) \sum_{N_{\text{sX}}^{\text{A}}} N_{\text{sX}}^{\text{A}} \mathcal{P}(N_{\text{sX}}^{\text{A}} | N_{\text{sX}}^{\text{B}}) \\
&= \langle N_{\text{sX}}^{\text{A}} \rangle \langle N_{\text{sX}'}^{\text{B}} \rangle (1 - \delta_{\text{XX}'}^{\text{K}}) \\
&\quad + \delta_{\text{XX}'}^{\text{K}} f_{\text{s}} \sum_{N_{\text{sX}}^{\text{B}}} (N_{\text{sX}}^{\text{B}})^2 \mathcal{P}(N_{\text{sX}}^{\text{B}}) \\
&= f_{\text{s}} \langle N_{\text{sX}}^{\text{B}} \rangle \langle N_{\text{sX}'}^{\text{B}} \rangle (1 - \delta_{\text{XX}'}^{\text{K}}) \\
&\quad + \delta_{\text{XX}'}^{\text{K}} f_{\text{s}} \left( \langle N_{\text{sX}}^{\text{B}} \rangle^2 + \langle N_{\text{sX}}^{\text{B}} \rangle \right) \\
&= f_{\text{s}} \langle N_{\text{sX}}^{\text{B}} \rangle \langle N_{\text{sX}'}^{\text{B}} \rangle + \delta_{\text{XX}'}^{\text{K}} f_{\text{s}} \langle N_{\text{sX}}^{\text{B}} \rangle \\
&= f_{\text{s}} \bar{N}_{\text{s}}^{\text{B}2} \bar{n}_{\text{s}}^{\text{B}}(X) \bar{n}_{\text{s}}^{\text{B}}(X') + \delta_{\text{XX}'}^{\text{K}} f_{\text{s}} \bar{N}_{\text{s}}^{\text{B}} \bar{n}_{\text{s}}^{\text{B}}(X).
\end{aligned} \tag{A15}$$

Next we consider the central–satellite contribution with a central galaxy from sample B and a satellite from A:

$$\begin{aligned}
\langle N_{\text{cX}'}^{\text{B}} N_{\text{sX}}^{\text{A}} \rangle &= \sum_{N_{\text{sX}}^{\text{A}}, N_{\text{cX}'}^{\text{B}}, N_{\text{cX}'}^{\text{B}}} N_{\text{cX}'}^{\text{B}} N_{\text{sX}}^{\text{A}} \mathcal{P}(N_{\text{sX}}^{\text{A}}, N_{\text{cX}'}^{\text{B}}, N_{\text{cX}'}^{\text{B}}) \\
&= \sum_{N_{\text{sX}}^{\text{A}}, N_{\text{cX}'}^{\text{B}}, N_{\text{cX}'}^{\text{B}}} N_{\text{cX}'}^{\text{B}} N_{\text{sX}}^{\text{A}} \\
&\quad \times \mathcal{P}(N_{\text{sX}}^{\text{A}} | N_{\text{cX}'}^{\text{B}}, N_{\text{cX}'}^{\text{B}}) \mathcal{P}(N_{\text{cX}'}^{\text{B}} | N_{\text{cX}'}^{\text{B}}) \mathcal{P}(N_{\text{cX}'}^{\text{B}}) \\
&= f_{\text{s}} \sum_{N_{\text{sX}}^{\text{A}}, N_{\text{cX}'}^{\text{B}}} N_{\text{cX}'}^{\text{B}} N_{\text{sX}}^{\text{A}} \mathcal{P}(N_{\text{sX}}^{\text{A}} | N_{\text{cX}'}^{\text{B}}) \mathcal{P}(N_{\text{cX}'}^{\text{B}}) \\
&= \delta_{\text{X'0}}^{\text{K}} f_{\text{s}} \bar{N}_{\text{c}}^{\text{B}} \langle N_{\text{sX}}^{\text{A}} | N_{\text{cX}'}^{\text{B}} = 1 \rangle \\
&= \delta_{\text{X'0}}^{\text{K}} f_{\text{s}} \bar{N}_{\text{c}}^{\text{B}} \bar{n}_{\text{s}}^{\text{A}}(X).
\end{aligned} \tag{A16}$$

And similarly for the satellite–central term:

$$\begin{aligned}
\langle N_{\text{cX}}^{\text{A}} N_{\text{sX}'}^{\text{B}} \rangle &= \sum_{N_{\text{cX}}^{\text{A}}, N_{\text{cX}'}^{\text{B}}, N_{\text{sX}'}^{\text{B}}} N_{\text{cX}}^{\text{A}} N_{\text{sX}'}^{\text{B}} \mathcal{P}(N_{\text{cX}}^{\text{A}}, N_{\text{cX}'}^{\text{B}}, N_{\text{sX}'}^{\text{B}}) \\
&= \sum_{N_{\text{cX}}^{\text{A}}, N_{\text{cX}'}^{\text{B}}, N_{\text{sX}'}^{\text{B}}} N_{\text{cX}}^{\text{A}} N_{\text{sX}'}^{\text{B}} \\
&\quad \times \mathcal{P}(N_{\text{cX}}^{\text{A}} | N_{\text{cX}'}^{\text{B}}, N_{\text{sX}'}^{\text{B}}) \mathcal{P}(N_{\text{sX}'}^{\text{B}} | N_{\text{cX}'}^{\text{B}}) \mathcal{P}(N_{\text{cX}'}^{\text{B}}) \\
&= f_{\text{c}} \sum_{N_{\text{cX}}^{\text{A}}, N_{\text{cX}'}^{\text{B}}} N_{\text{sX}'}^{\text{B}} N_{\text{cX}}^{\text{A}} \mathcal{P}(N_{\text{sX}'}^{\text{B}} | N_{\text{cX}'}^{\text{B}}) \mathcal{P}(N_{\text{cX}'}^{\text{B}}) \\
&= \delta_{\text{X0}}^{\text{K}} f_{\text{c}} \bar{N}_{\text{c}}^{\text{B}} \langle N_{\text{sX}'}^{\text{B}} | N_{\text{cX}'}^{\text{B}} = 1 \rangle \\
&= \delta_{\text{X0}}^{\text{K}} f_{\text{c}} \bar{N}_{\text{c}}^{\text{B}} \bar{n}_{\text{s}}^{\text{B}}(X').
\end{aligned} \tag{A17}$$

Putting these terms into equation (A9):

$$\begin{aligned}
\langle U^{\text{A}}(k | M) U^{\text{B}}(k | M) \rangle &= \sum_{X, X'} e^{ik(X-X')} \\
&\quad \times \left[ \delta_{\text{XX}'}^{\text{K}} \delta_{\text{X0}}^{\text{K}} f_{\text{c}} \bar{N}_{\text{c}}^{\text{B}} + f_{\text{s}} \bar{N}_{\text{c}}^{\text{B}2} \bar{n}_{\text{s}}^{\text{B}}(X) \bar{n}_{\text{s}}^{\text{B}}(X') + \delta_{\text{XX}'}^{\text{K}} f_{\text{s}} \bar{N}_{\text{c}}^{\text{B}} \bar{n}_{\text{s}}^{\text{B}}(X) \right. \\
&\quad \left. + \delta_{\text{X'0}}^{\text{K}} f_{\text{s}} \bar{N}_{\text{c}}^{\text{B}} \bar{n}_{\text{s}}^{\text{B}}(X) + \delta_{\text{X0}'}^{\text{K}} f_{\text{c}} \bar{N}_{\text{c}}^{\text{B}} \bar{n}_{\text{s}}^{\text{B}}(X') \right] \\
&= f_{\text{c}} \bar{N}_{\text{c}}^{\text{B}} + f_{\text{s}} \bar{N}_{\text{c}}^{\text{B}2} \bar{n}_{\text{s}}^{\text{B}*}(k) \bar{n}_{\text{s}}^{\text{B}}(k) + \sum_X f_{\text{s}} \bar{N}_{\text{c}}^{\text{B}} \bar{n}_{\text{s}}^{\text{B}}(X) \\
&\quad + f_{\text{s}} \bar{N}_{\text{c}}^{\text{B}} \bar{n}_{\text{s}}^{\text{B}*}(k) + f_{\text{c}} \bar{N}_{\text{c}}^{\text{B}} \bar{n}_{\text{s}}^{\text{B}}(k) \\
&= f_{\text{c}} \bar{N}_{\text{c}}^{\text{B}} + f_{\text{s}} \bar{N}_{\text{c}}^{\text{B}2} \left( \bar{N}_{\text{s}}^{\text{B}} \bar{u}_{\text{s}}(k) \right)^2 + f_{\text{s}} \bar{N}_{\text{c}}^{\text{B}} \bar{N}_{\text{s}}^{\text{B}} \\
&\quad + (f_{\text{s}} + f_{\text{c}}) \bar{N}_{\text{c}}^{\text{B}} \bar{N}_{\text{s}}^{\text{B}} \bar{u}_{\text{s}}(k),
\end{aligned} \tag{A18}$$

where we have made use of the normalization  $\sum_X u_{\text{s}}(X) = 1$ .

The first and third terms represent central–central and satellite–satellite self-pairings between galaxies common to both samples. These terms are independent of  $k$ , so they correspond to the shot-noise contribution to the power spectrum which affects only the zero-lag correlation function ( $\theta = 0$ ). Since we do not include this in our analysis, these terms can be neglected, so

$$\begin{aligned}
P_{1\text{h}}(k | z) &= \frac{1}{\bar{n}_{\text{g}}^{\text{A}} \bar{n}_{\text{g}}^{\text{B}}} \int dM dz n(M, z) \\
&\quad \times \left[ (f_{\text{s}}(M) + f_{\text{c}}(M)) \bar{N}_{\text{c}}^{\text{B}}(M) \bar{N}_{\text{s}}^{\text{B}}(M) \bar{u}_{\text{s}}(k | M) \right. \\
&\quad \left. + f_{\text{s}}(M) \left( \bar{N}_{\text{c}}^{\text{B}}(M) \bar{N}_{\text{s}}^{\text{B}}(M) \bar{u}_{\text{s}}(k | M) \right)^2 \right].
\end{aligned} \tag{A19}$$

The fractions of central and satellite galaxies in sample B that are also in sample A for a given set of HOD parameters are given by

$$f_{\text{c}}(M) = \frac{\bar{N}_{\text{c}}^{\text{A}}(M)}{\bar{N}_{\text{c}}^{\text{B}}(M)}, \tag{A20}$$

and

$$f_{\text{s}}(M) = \frac{\bar{N}_{\text{c}}^{\text{A}}(M) \bar{N}_{\text{s}}^{\text{A}}(M)}{\bar{N}_{\text{c}}^{\text{B}}(M) \bar{N}_{\text{s}}^{\text{B}}(M)}. \tag{A21}$$

To implement this in CCL, the `fourier_2pt` method of the `Profile2ptHOD` class<sup>9</sup> was modified.

Once the halo model power spectrum has been calculated, the angular clustering can be predicted using CCL (see section 2.4.1 of N. E. Chisari et al. 2019). Using the Limber approximation (D. N. Limber 1954; N. Afshordi, Y.-S. Loh & M. A. Strauss 2004), the angular power spectrum can be written as

$$C_{\ell} = \frac{2}{2\ell + 1} \int dk P_{\text{hm}}(k | z_{\ell}) \Delta_{\ell}^{\text{A}}(k) \Delta_{\ell}^{\text{B}}(k), \tag{A22}$$

<sup>9</sup>[https://github.com/LSSTDESC/CCL/blob/master/pyccl/halos/profiles\\_2pt.py](https://github.com/LSSTDESC/CCL/blob/master/pyccl/halos/profiles_2pt.py)

where a radial distance  $\chi_\ell \equiv (\ell + 1/2)/k$  has been defined and  $z_\ell$  is the corresponding redshift.  $\Delta_\ell^A(k)$  and  $\Delta_\ell^B(k)$  are the transfer functions corresponding to the samples. For galaxy number count tracers, these are given by

$$\Delta_\ell^{\text{NC}}(k) = \frac{H(z_\ell)}{c} p_z(z_\ell), \quad (\text{A23})$$

where  $H(z)$  is the Hubble parameter,  $c$  is the speed of light and  $p_z(z)$  is the normalized redshift distribution of the sample. Finally, the angular correction function,  $\omega(\theta)$ , is obtained by summing over the multipole moments:

$$\omega(\theta) = \frac{1}{4\pi} \sum_\ell (2\ell + 1) C_\ell P_\ell(\cos \theta), \quad (\text{A24})$$

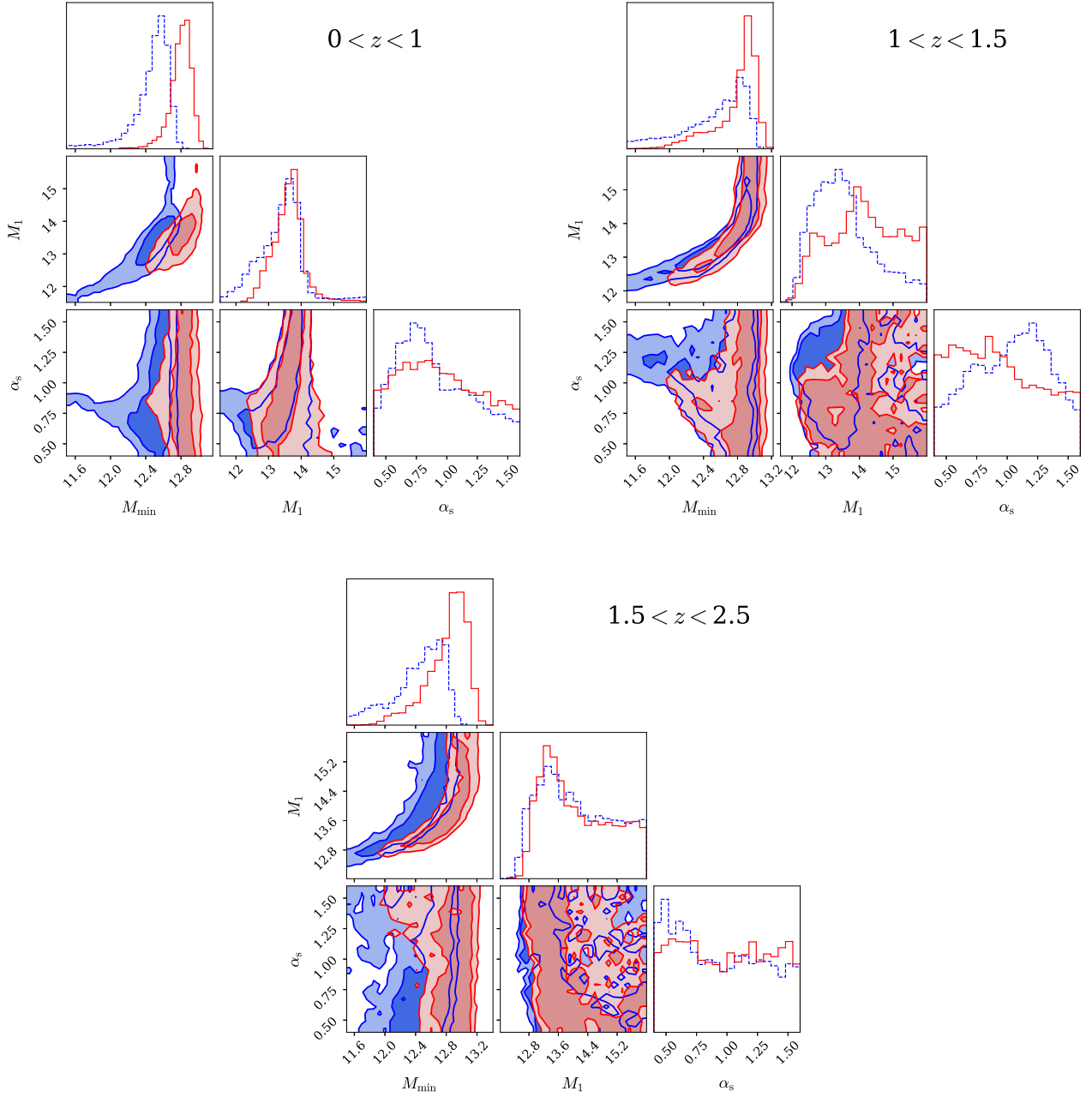
where  $P_\ell(x)$  are the zeroth-order associated Legendre polynomials.

Because stellar mass thresholds are applied to the optical/NIR population but not the AGN and matched galaxy samples, the assumption that A is a subset of B is only approximate. For the two lower redshift bins ( $0 < z < 1$  and  $1 < z < 1.5$ ), fewer than 9 per cent of galaxies in A lie below the stellar mass thresholds, but this rises to  $\sim 18$  per cent for  $1.5 < z < 2.5$ . The only difference in the two-point Fourier moment between the cases where samples A and B are independent (equation A7) and when A is a

subset of B (equation A18) is in the central–satellite and satellite–central contributions, which are proportional to  $\bar{u}_s(k | M)$ . Isolating these terms, we have  $\bar{N}_c^A \bar{N}_c^B \bar{N}_s^B + \bar{N}_c^B \bar{N}_c^A \bar{N}_s^A$  when the samples are independent and  $\bar{N}_c^A \bar{N}_s^B + \bar{N}_c^A \bar{N}_s^A$  when one is a subset of the other. We find that the second term in each of these expressions is much smaller than the first term since  $\bar{N}_s^B \gg \bar{N}_s^A$ , so can be neglected, and the independent case differs only by a factor of  $\bar{N}_c^B$ . In a perfect subset,  $\bar{N}_c^B = 1$  whenever  $\bar{N}_c^A > 0$ , so the two expressions coincide. When some of A’s central galaxies fall below B’s stellar mass limit,  $\bar{N}_c^B < 1$  in those haloes, and exactly that fraction of central–satellite pairs is lost, producing a small, scale-dependent suppression of the one-halo term. However, because  $M_1^B \gg M_{\text{min}}^A$  and  $M_{\text{min}}^B$ , both  $\bar{N}_c^A$  and  $\bar{N}_c^B$  are very close to 1 at the halo masses where B’s satellites first appear, so the loss is negligible in practice.

## APPENDIX B: POSTERIOR PROBABILITIES OF THE HOD PARAMETERS

Fig. B1 shows the two-dimensional posterior distributions of the best-fitting HOD model parameters for the AGN and matched galaxy samples for each redshift bin.



**Figure B1.** The posterior distributions and corresponding histograms of the best-fitting HOD model parameters for the AGN (solid red) and matched galaxy (dashed blue) samples for each redshift bin. The contours are the 68 percent and 95 percent confidence intervals.

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