Obscured AGN enhancement in galaxy pairs at cosmic noon: evidence from a probabilistic treatment of photometric redshifts

Sean L. Dougherty,[1]★ C. M. Harrison [ORCID],[1]★ Dale D. Kocevski[2] and D. J. Rosario[1]

[1]*School of Mathematics, Statistics and Physics, Newcastle University, Tyne NE1 7RU, UK*
[2]*Department of Physics and Astronomy, Colby College, Waterville, ME 04961, USA*

ABSTRACT

Observations of the nearby universe reveal an increasing fraction of active galactic nuclei (AGNs) with decreasing projected separation for close galaxy pairs, relative to control galaxies. This implies galaxy interactions play a role in enhancing AGN activity. However, the picture at higher redshift is less established, partly due to limited spectroscopic redshifts. We combine spectroscopic surveys with photometric redshift probability distribution functions for galaxies in the CANDELS and COSMOS surveys, to produce the largest ever sample of galaxy pairs used in an AGN fraction calculation for cosmic noon ($0.5 < z < 3$). We present a new technique for assessing galaxy pair probability (based on line-of-sight velocities $\pm 1000\,\mathrm{km\,s^{-1}}$) from photometric redshift posterior convolutions and use these to produce weighted AGN fractions. Over projected separations 5–100 kpc, we find no evidence for enhancement, relative to isolated control galaxies, of X-ray ($L_X > 10^{42}\,\mathrm{erg\,s^{-1}}$) or infrared-selected AGN in major (mass ratios up to 4:1) or minor (4:1 to 10:1) galaxy pairs. However, defining the most obscured AGN as those detected in the infrared but not in X-rays, we observe a trend of increasing obscured AGN enhancement at decreasing separations. The peak enhancement, relative to isolated controls, is a factor of 2.08 ± 0.61 for separations < 25 kpc. Our simulations with mock data, indicates this could be a lower limit of the true enhancement. If confirmed with improved infrared imaging (e.g. with *JWST*) and redshifts (e.g. with forthcoming multi-object spectrograph surveys), this would suggest that galaxy interactions play a role in enhancing the most obscured black hole growth at cosmic noon.

Key words: galaxies: active – galaxies: interactions – infrared: galaxies – X-rays: galaxies.

1 INTRODUCTION

There are now several observational results that indicate a co-evolution of supermassive black holes (SMBHs) and their host galaxies. This includes tight correlations between SMBH mass and spheroidal properties (e.g. Magorrian et al. 1998; Gebhardt et al. 2000; Gültekin et al. 2009; Kormendy & Ho 2013) as well as the similar cosmic evolution of SMBH accretion and star formation rate (SFR) density, both of which peak at 'cosmic noon' (i.e. $z \sim 2$; e.g. Madau & Dickinson 2014; Aird et al. 2015).

There is ongoing research on the role of galaxy mergers in driving the co-evolution of SMBHs and galaxies. Theoretically, strong gravitational interactions within a merger could reduce the angular momentum of gas and channel inflows into the nuclear region, both compressing gas into regions of intense star formation and accreting it on to the SMBH itself, consequently triggering an active galactic nuclei (AGNs; e.g. Barnes & Hernquist 1991; Mihos & Hernquist 1996; Blumenthal & Barnes 2018). Major galaxy mergers (i.e. those with mass ratios up to 4:1) undoubtedly impact the final morphology of the combined system (e.g. Darg et al. 2010; Ellison et al. 2010; Casteels et al. 2013), but their implications for star formation (e.g.

Patton et al. 2013; Barrera-Ballesteros et al. 2015; Moreno et al. 2021; Ellison et al. 2022) and AGN activity (e.g. Treister et al. 2012; Ellison et al. 2013b; Villforth et al. 2017; Goulding et al. 2018) are a matter of significant debate in the literature.

In the nearby universe ($z \sim 0$), there is extensive observational evidence for mergers enhancing the number of AGN selected by optical emission lines (Ellison et al. 2011, 2013b; Pierce et al. 2023), mid-infrared colours (Ellison et al. 2013a; Satyapal et al. 2014; Gao et al. 2020) and those with high-excitation radio emission (Ramos Almeida et al. 2011; Pierce et al. 2022). The largest excess of AGN associated with galaxy interactions or mergers is usually found for the most luminous sources; however, it is a matter of ongoing debate if galaxy interactions significantly *increase* accretion rates (Ellison et al. 2019; Pierce et al. 2022; Bickley et al. 2023; Steffen et al. 2023). These findings are well complemented in the cosmological hydrodynamic simulation EAGLE, where McAlpine et al. (2020) report both the highest AGN excess in mergers at low redshifts (0.05 $< z < 0.1$) and the highest merging excess in AGN at high bolometric luminosities ($L_{\mathrm{bol}} \sim 10^{45}\,\mathrm{erg\,s^{-1}}$).

This is all broadly in agreement with a scenario where, in the local Universe, the most luminous AGN phases can be boosted by major mergers whereas less luminous AGNs are predominantly associated with secular processes (e.g. Sanders et al. 1988a, b; Bennert et al. 2008; Urrutia, Lacy & Becker 2008). None the less, the overall importance of galaxy mergers or interactions *for total black hole*

★ E-mail: sldough21@gmail.com (SLD); christopher.harrison@newcastle.ac.uk (CMH)

growth remains unclear (e.g. see discussion in McAlpine et al. 2020; Byrne-Mamahit et al. 2023). Secular processes, such as stochastic bar instabilities, stellar winds or interactions with dark matter haloes, are alternate AGN fueling mechanisms for driving the majority of black hole growth (e.g. Kormendy & Kennicutt 2004; Hopkins & Hernquist 2006).

Meanwhile, the AGN-merger connection is far from established at cosmic noon. Empirical works find conflicting results when starting from similar samples, such as luminous quasars (Glikman et al. 2015; Mechtley et al. 2016; Marian et al. 2019), X-ray selected AGN (Cisternas et al. 2011; Silverman et al. 2011; Kocevski et al. 2012; Lackner et al. 2014; Rosario et al. 2015; Hewlett et al. 2017; Villforth et al. 2017; Shah et al. 2020) and infrared (IR) selected AGN (Schawinski et al. 2012; Donley et al. 2018; Goulding et al. 2018; Silva et al. 2021). While a significant delay between the onset of inflows and AGN activity (∼50–500 Myrs; e.g. Davies et al. 2007; Schawinski et al. 2009; Wild, Heckman & Charlot 2010) may make it difficult to reconcile contradicting findings, there are a couple of important cosmological factors to consider. First, the frequency of merging galaxies and luminous AGN is substantially higher at cosmic noon, relative to the local Universe (∼10×; e.g. Conselice et al. 2003; Kartaltepe et al. 2007; Huško, Lacey & Baugh 2022). Second, the gas fraction is similarly much greater (∼5×; e.g. Daddi et al. 2010; Tacconi et al. 2010; Scoville et al. 2014), which could result in different inflow physics as well as more prominent phases of obscured black hole growth. These factors make a well considered control sample (of non-merging galaxies) crucial for any analyses of the AGN frequency.

Inconsistencies among works may, at least in part, be attributed to the various methods used to classify mergers. Most high-redshift studies rely on the identification of morphological disturbances indicative of mergers, such as tidal bridges or tails, in imaging data with model-fitting (e.g. Lackner et al. 2014), machine learning (e.g. Goulding et al. 2018) or by visual inspection (e.g. Cisternas et al. 2011). However, with a decrease in surface brightness $\propto (1 + z)^{-4}$, this approach becomes exceptionally more challenging and temporally expensive when searching beyond nearby sources. This may lead to incorrect non-merging identifications and restrictive samples, particularly for the most luminous or obscured AGN.

An alternative approach, which is less susceptible to surface brightness dimming effects, is to look at close galaxy pairs and explore AGN fraction as a function of projected separation. This method has been used extensively at low redshift for emission-line- and IR-selected AGN (e.g. Ellison et al. 2011, 2013b; Satyapal et al. 2014; Steffen et al. 2023). Recently, it has been applied at cosmic noon (0.5 < z < 3.0) by Shah et al. (2020), who used spectroscopic redshifts in the CANDELS and COSMOS fields to distinguish between random projections and truly associated galaxies. They find no excess in X-ray- or IR-selected AGN in close pairs compared to a matched control sample of isolated galaxies. This is in contrast to Silverman et al. (2011), who report mild excess of X-ray AGN (∼2×) in 0.25 < z < 1.05 spectroscopic galaxy pairs in COSMOS, albeit, at lower redshift range and when compared to a more general control sample (i.e. to the field). However, spectroscopic completeness at high redshift is poor due to source faintness and many strong emission lines being redshifted out of optical spectroscopy beyond $z \gtrsim 1.5$. Spectroscopic redshift incompleteness is particularly problematic for the most obscured sources (e.g. those not detected in X-ray surveys), which are much less frequently the target of spectroscopic observations. Without redshift information for *all galaxies*, true physical pairs remain inevitably entangled with chance projections.

In this work, we address the issue of spectroscopic completeness at cosmic noon by folding in photometric redshift posterior probability distribution functions (PDFs). We build on previous studies of galaxy pairs and mergers that have utilized photometric redshifts (e.g. Kartaltepe et al. 2007; Bundy et al. 2009; López-Sanjuan et al. 2015; Mundy et al. 2017; Mantha et al. 2018; Duncan et al. 2019), but we develop a new approach, which is made novel by its fair treatment of the full PDF uncertainty.

The structure of this paper is as follows. In Section 2, we describe the relevant archival data and define our parent sample. We outline our pair probability methodology in Section 3. We present our results in Section 4 and discuss their implications for merger-driven SMBH growth at cosmic noon in Section 5. Finally, we summarize our findings in Section 6. Throughout this work, we assume a Lambda cold dark matter (ΛCDM) cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.7$, and $\Omega_M = 0.3$.

2 DATA AND SAMPLE SELECTION

Our study is designed to measure the prevalence of AGN as a function of galaxy separation, using a statistically robust sample of physically associated galaxy pairs and isolated control galaxies at cosmic noon (0.5 < z < 3). To achieve this, we have utilized the extragalactic deep fields from the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011) and the Cosmic Evolution Survey (COSMOS; Scoville et al. 2007), following Shah et al. (2020). These fields have extensive multiwavelength photometry, detailed archival analyses of photometric redshift PDFs and stellar masses (Section 2.1), and multiple spectroscopic redshift studies (Section 2.2). We use these as the basis for our parent galaxy sample selection (Section 2.3) and for identifying both X-ray AGNs and IR AGNs (Section 2.4).

2.1 Photometry and derived quantities

We make use of the five CANDELS extragalactic deep fields, which were all observed with high-resolution near-IR and optical filters from the *Hubble Space Telescope (HST)*/Wide-Field Camera 3 (WFC3) and *HST*/Advanced Camera for Surveys (ACS). The CANDELS team has performed consistent deblending and photometry across fields, which combine for a total sky coverage of ∼960 arcmin². To increase our source statistics, we also make use of the wider ∼2 deg² COSMOS field, the largest continuous area observed by *HST*/ACS to date. Below, we describe the archival ultraviolet (UV) to mid-infrared (MIR) photometry, the photometric redshift information, the derived stellar masses and the X-ray observations that we use in this study.

2.1.1 UV–MIR photometry

The CANDELS photometry of the Great Observatories Origins Deep Survey North (GOODS-N) and South (GOODS-S; Giavalisco et al. 2004), Extended Groth Strip (EGS; Davis et al. 2007), UKIDSS Ultra Deep Survey (UDS; Lawrence et al. 2007) and a central region from the COSMOS field are described in Barro et al. (2019), Guo et al. (2013), Stefanon et al. (2017), Galametz et al. (2013), and Nayyeri et al. (2017), respectively. In short, the photometry is based on broad-band data from the UV to MIR, including deep near-infrared (NIR) *HST*/WFC3 and optical *HST*/ACS CANDELS observations. Ancillary MIR observations were taken with *Spitzer*/IRAC 3.4, 4.5,

5.8, and 8.0 μm imaging of GOODS-N (Dickinson et al. 2003; Ashby et al. 2015), GOODS-S (Dickinson et al. 2003; Ashby et al. 2013), EGS (Barmby et al. 2008; Ashby et al. 2013, 2015), UDS (Ashby et al. 2013, 2015), and COSMOS (Sanders et al. 2007; Ashby et al. 2013). Source detection was performed with SOURCE EXTRACTOR (Bertin & Arnouts 1996) on the WFC3 *F*160*W* (*H*-band) mosaic, which reaches 5σ limiting AB depths of 27.3–27.8 across the fields. Low-resolution photometry was measured with TFIT (Laidler et al. 2007). Briefly, this procedure involves convolving the brightness profile of high-resolution imaging (i.e. WFC3 *F*160*W*; *H*-band) with a kernel to match the point-spread function of the low-resolution imaging (i.e. IRAC channels), from which sources can be deblended and fluxes extracted via best fit.

For the wider COSMOS field, we use the recent COSMOS2020 CLASSIC photometric catalog (Weaver et al. 2022). Once again, the photometry incorporates a wealth of multiwavelength imaging from the UV-MIR, notably with recent optical Subaru/Suprime-Cam and NIR VISTA/ULTRAVISTA observations. In the MIR, IRAC observations were taken from the Cosmic Dawn Survey (Euclid Collaboration 2022), a collation of new and existing IRAC data in the Euclid deep fields, of which COSMOS is included (Sanders et al. 2007; Ashby et al. 2013; Steinhardt et al. 2014; Ashby et al. 2015, 2018). Source detection was performed on a 'chi-squared' *izYJHK$_s$* image (Szalay, Connolly & Szokoly 1999) with SOURCE EXTRACTOR. Low-resolution photometry was performed with IRACLEAN (Hsieh et al. 2012). Briefly, this method treats the high-resolution brightness profile as a prior, from which a fraction of its flux is iteratively subtracted until some minimal pixel brightness criteria is reached. Following Shah et al. (2020), we prioritize using the photometry and quantities derived by the CANDELS team in the region of COSMOS where the surveys overlap.

2.1.2 Photometric redshift PDFs

In this study, we utilize the full photometric redshift PDFs (in addition to spectroscopic redshift information) to calculate the probability of galaxies at close projected separations being in a physically associated pair (see Section 3). That is, we do not simply assume the peak of the PDF as a fixed photometric red-shift or its width as an uncertainty. Across all fields, these PDFs have been carefully constructed in previous works to both be accurate (i.e. the peak PDF estimate is similar to a confirmed spectroscopic redshift) and give realistic widths (i.e. 68 per cent of spectroscopic redshifts should fall within 1σ of the mean PDF estimate).

For the wider COSMOS field, spectral energy distribution (SED) fitting was performed separately with EAZY (Brammer, van Dokkum & Coppi 2008) and LePhare (Arnouts & Ilbert 2011) based on the COSMOS2020 photometry (Weaver et al. 2022). We elect to use the photometric redshift PDFs (and subsequent derived properties) from LePhare due to its greater precision across all magnitudes (see fig. 15 in Weaver et al. 2022). For the CANDELS fields, these PDFs have been independently generated from separate SED-fitting codes with various template configurations by 6 groups in the CANDELS team. These codes are EAZY (Brammer et al. 2008), zphot (Giallongo et al. 1998; Fontana et al. 2000), HyperZ (Bolzonella, Miralles & Pelló 2000), LePhare (Arnouts & Ilbert 2011), and WikZ (Wiklind et al. 2008). In an assessment of these methods, Dahlen et al. (2013) find that the individual redshift posteriors underestimate realistic uncertainties

and that they improve in accuracy when combined. Therefore, for each source, we use the PDF combination assembled by Kodra et al. (2023) using a minimum Frećhet distance method, which was shown to be an improvement in both accuracy and PDF width over any individual group. We note that, despite the specific differences in PDF generation, we verified that our main results are consistent across CANDELS and COSMOS, if they are treated independently.

We perform an assessment of the quality of the photometric redshifts in the context of the spectroscopic redshifts for the global galaxy population and the known AGNs in Appendix A. In summary, we find that the normalized median absolute deviation between the spectroscopic redshifts and photometric redshifts is $\sigma_{NMAD} \sim$ 0.02 and $\sigma_{NMAD} \sim$ 0.03, for the full galaxy parent sample and only the AGNs, respectively. Although these are modestly higher for the AGNs, this is countered somewhat by the higher fraction of spectroscopic redshifts of the AGNs compared to the general galaxy population (see Appendix A). We reiterate that the full PDF information is incorporated into our method (see Section 3), and we further perform tests of including or removing spectroscopic redshifts in our analyses (see Section 4.1).

2.1.3 Stellar masses

We follow Shah et al. (2020) in our choice of archival masses for the sample across fields used in this study. Specifically, in COSMOS, stellar masses were computed using LePhare (i.e. the same as that used for the photometric redshift PDFs) with the configuration described in Laigle et al. (2016). Medians of the photometric redshift PDFs generated by LePhare were taken as the fitting redshift. Stellar masses in the CANDELS fields were computed following the method outlined in Mobasher et al. (2015) and Santini et al. (2015). In short, 10 groups within the CANDELS team estimated stellar mass PDFs from the same photometry and redshifts but with different SED stellar templates and/or codes. The median of the average stellar mass PDF is taken as the best mass estimate.

We note that our results are insensitive to mass variations or un-certainties (e.g. caused by subtle differences in the chosen redshifts) because these are only used to create very broad categories of major and minor galaxy samples, for which we observe no differences in our results (see Section 2.3).

2.1.4 X-ray data

We make use of public *Chandra* point-source catalogues to identify X-ray AGNs in our sample. In GOODS-S, we use the point-source catalogue of Xue et al. (2011). This benefits from the *H*-band counterpart matching (directly to our preferred CANDELS photometric catalogue) performed by Hsu et al. (2014) and reaches observed X-ray luminosities down to $L_X = 10^{42}$ erg s^{-1} at $z \sim$ 3, which is sufficient for our purposes. We use the 2 Ms and 800 ks point-source catalogues of Xue et al. (2016) and Nandra et al. (2015) for GOODS-N and EGS, respectively, and the optical counterparts therein. In COSMOS, we use the 4.6 Ms catalogue from Civano et al. (2016) with optical counterparts identified in Marchesi et al. (2016). Finally, we use the 600 ks source catalogue from Kocevski et al. (2018) in UDS, matching the X-ray point-sources to the *H*-band-selected objects in CANDELS using a maximum-likelihood technique (Sutherland & Saunders 1992; Civano et al. 2012).
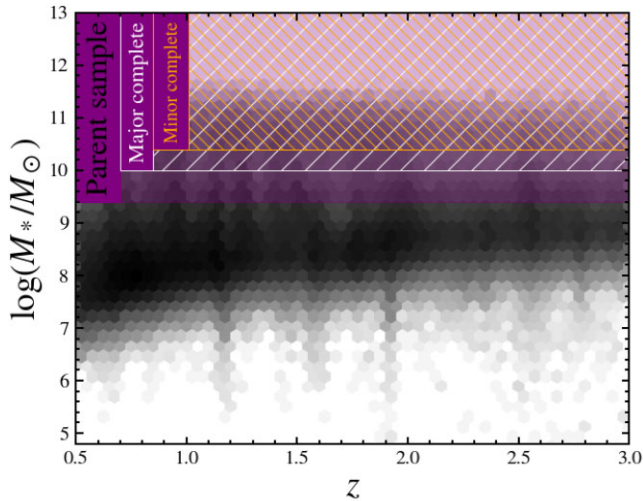
Figure 1. Stellar mass distribution of galaxies in the CANDELS and COSMOS fields. The shaded purple region, $M_* > 10^{9.4}\,M_\odot$, indicates all galaxies included in our parent sample. Fixing that galaxy 1 is always more massive than galaxy 2 (i.e. $M_{*,1} > M_*$,), the white and orange striped regions define the stellar mass ranges of galaxy 1 to ensure a complete major pair sample ($M_* > 10^{10}\,M_\odot$ and $M_{*,1}$: $M_{*,2} < 4:1$) and minor pair sample ($M_* > 10^{10.4}\,M_\odot$ and $4:1 < M_{*,1}$: $M_{*,2} < 10:1$), respectively.

2.2 Spectroscopic redshifts

We compile all publicly available secure spectroscopic redshifts in the CANDELS and COSMOS fields. Those spectroscopic redshifts were obtained from the measurements provided in several published works, which are detailed in full in Appendix A.

There is no consistent quality flagging across these works; however, we select only spectroscopic redshifts measured from multiple emission lines. In the case where an object has multiple secure spectroscopic redshifts and the relative line-of-sight velocity calculated from the maximum difference in redshifts is less than $\pm 1000\,\mathrm{km\,s^{-1}}$ (i.e. the velocity offset used in our pair criteria; see Section 3), we use the mean redshift. If this relative line-of-sight velocity is greater than $\pm 1000\,\mathrm{km\,s^{-1}}$, then we default to the spectroscopic redshift measured from the higher resolution spectrograph. These final cases make up <1 per cent of all spectroscopic redshifts used in this work. There are 32 per cent ($\sim$3500) more spectroscopic redshifts in the parent sample (defined below) when including low-quality redshifts (i.e. those with single line identification or low signal-to-noise spectra). However, we prefer to focus on the high-quality redshifts only, using photometric redshifts otherwise (for which we take into account the full uncertainty). We note that our results are consistent if we only use photometric redshifts.

2.3 Galaxy sample selection

Our parent sample selection, primarily based on stellar mass and redshift, is presented in Fig. 1. The selection is largely motivated by Shah et al. (2020), who also study galaxy pairs in the COSMOS and CANDELS fields. This allows for a more direct comparison of the X-ray and IR AGNs between pairs identified with just spectroscopic redshifts (following Shah et al. 2020) to a more complete sample, folding in the full photometric redshift PDFs for all galaxies, which is the procedure of this work (see Section 4.1).

Starting with the photometric catalogues described in Section 2.1, we first clean the sample for poor photometric measurements. We exclude sources with a SOURCE EXTRACTOR stellarity parameter

greater than 0.9 to account for high-probability stellar contaminants (see Guo et al. 2013), and we remove sources where the photometry is compromised by image artifacts, such as edge effects, nearby bright objects or diffraction spikes.

We then select the parent sample as all galaxies with $M_* > 10^{9.4}\,M_\odot$ and $0.5 < z < 3$ (purple shaded region in Fig. 1). Since we are not limited to sources with spectroscopic redshifts, our sample is large enough to separately investigate AGN activity in both major and minor mergers, which we define as pairs with mass ratio $M_{*,1}$: $M_{*,2} < 4:1$ and $4:1 < M_{*,1}$: $M_{*,2} < 10:1$, respectively. Designating $M_{*,1} > M_{*,2}$, our parent sample is complete for major mergers when $M_{*,1} > 10^{10}\,M_\odot$ and for minor mergers when $M_{*,1} > 10^{10.4}\,M_\odot$ (see white and orange striped regions in Fig. 1). Consequently, we define two pair samples, major complete ($M_{*,1} > 10^{10}\,M_\odot$ and $M_{*,1}$: $M_{*,2} < 4:1$; i.e. most similar to the parent sample of Shah et al. 2020) and minor complete ($M_{*,1} > 10^{10.4}\,M_\odot$ and $4:1 < M_{*,1}$: $M_{*,2} < 10:1$).

2.4 AGN selection

X-ray fluxes in the 0.5–10 keV band were extrapolated from the observed fluxes in the full (0.5–8 keV), hard (2–8 keV), and soft (0.5–2 keV) bands (in that order of priority) assuming an X-ray spectral index of slope $\Gamma = 1.4$ (e.g. Marchesi et al. 2016). The total X-ray luminosity, L_X, is then

$$L_X = F_X \times 4\pi d_L^2 \times (1+z)^{\Gamma-2}, \qquad (1)$$

where F_X is the extrapolated X-ray flux and d_L the luminosity distance (calculated using the peak of the PDF photometric redshift or spectroscopic redshift, if available). We then impose a conservative AGN threshold cut of $L_X > 10^{42}\,\mathrm{erg\,s^{-1}}$ (e.g. Moran, Lehnert & Helfand 1999) to avoid any contamination from high-mass X-ray binary emission in star-forming galaxies (Alexander et al. 2005). These AGNs are represented in an X-ray luminosity-redshift plane in Fig. 2. We note that there is some uncertainty in the X-ray luminosities, particularly due to the photometric redshifts; however, these do not affect our results. The luminosities are only used to define broad X-ray luminosity bins, and, when modifying the X-ray luminosity threshold, our results remain quantitatively and qualitatively consistent (see Section 4.2).

We define two X-ray AGN samples, with different X-ray luminosity ranges, to reflect the varying depths of the *Chandra* observations across the CANDELS and COSMOS fields. We define 'moderate L_X AGN' as those with $L_X = 10^{43.2-43.7}\,\mathrm{erg\,s^{-1}}$, complete across all fields to $0.5 < z < 2$ (see purple shaded region in Fig. 2), and 'high L_X AGN' with $L_X > 10^{43.7}\,\mathrm{erg\,s^{-1}}$, complete over the full redshift range (see blue shaded region in Fig. 2).

Following Shah et al. (2020), we classify IR AGN from IRAC data based on the criteria outlined in Donley et al. (2012). This method combines a colour–colour cut with a power-law selection, excluding star-forming galaxies at this epoch and assembling a highly complete and reliable sample of IR AGNs. Other IRAC selection methods just use a simple colour–colour cut to identify IR AGNs (Lacy et al. 2004; Stern et al. 2005; Lacy et al. 2007), but they become contaminated by star-forming galaxies at $z \gtrsim 0.5$ (e.g. Assef et al. 2010; Park et al. 2010) and are therefore less conservative.

We note that 14.6 per cent of X-ray-selected AGNs are also IR-selected, but 47.7 per cent of IR-selected AGNs are not X-ray selected as AGNs. We investigate this latter population further in Section 4.3. Additionally, given the propensity of X-ray-selected AGNs for spectroscopic follow-up, 30 per cent of X-ray-selected AGNs have a secure spectroscopic redshift compared to (8.3 per cent) 21 per cent of AGN selected (only) in the IR. The redshift distribution (peak
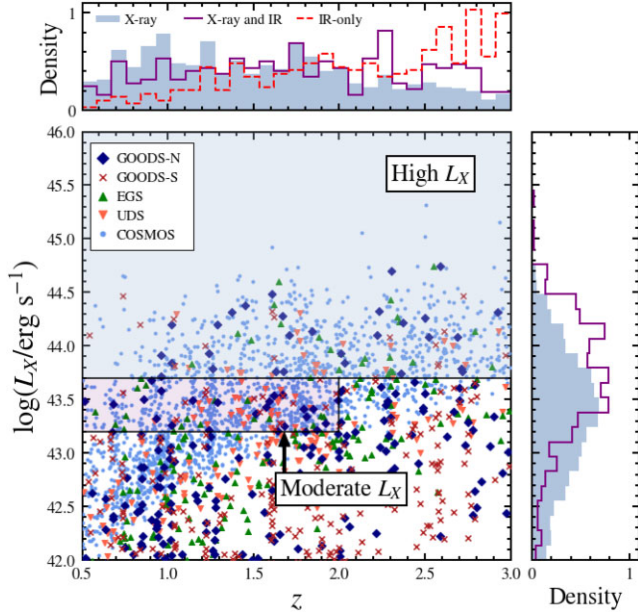
Figure 2. X-ray luminosity (full band, 0.5–10 keV) versus redshift for all X-ray sources ($L_X > 10^{42}$ erg s^{-1}) in our parent sample. The navy diamonds, red crosses, green upward triangles, orange downward triangles and blue dots correspond to X-ray sources in GOODS-N, GOODS-S, EGS, UDS, and COSMOS, respectively. We define two large L_X regimes that are complete across all fields. The purple shaded region corresponds to 'moderate L_X AGN' ($L_X = 10^{43.2-43.7}$ erg s^{-1}), with $0.5 < z < 2$. The blue shaded region defines the 'high L_X AGN' sample ($L_X > 10^{43.7}$ erg s^{-1}), over the full redshift range. Redshift and luminosity distributions are shown in the top and right-hand panels, respectively, for AGN sub-samples: AGN selected in the X-ray (solid blue histogram); those also selected in X-ray and IR (purple line) and those selected only in the IR (i.e. $L_X < 10^{42}$ erg s^{-1} or undetected; red dashed line).

photometric estimate or secure spectroscopic redshift, if available) of X-ray AGNs, those also selected in the IR and IR-only AGNs are characterized in Fig. 2. Photometric PDF validation is performed for these sub-samples using spectroscopic redshifts in Appendix A.

3 PROBABILISTIC PAIR METHODOLOGY

In this section, we outline the method used to assess the probability that a projected pair of galaxies is in a physically associated 'true' pair and how we construct a control sample of isolated galaxies. We use this information to calculate the AGN fraction and AGN enhancement (with respect to the control galaxies) as a function of projected separation. We define a true galaxy pair as two galaxies within projected separation $r_p < 100$ kpc and a relative line-of-sight velocity $|\Delta V| < 1000$ km s^{-1}. Our choice of $|\Delta V| < 1000$ km s^{-1}, to define physically associated pairs, follows that chosen for the primary analyses of Shah et al. (2020). We also note that enhanced AGN fractions are typically found to start in pairs of $r_p \lesssim 50$ kpc (e.g. Ellison et al. 2013b; Satyapal et al. 2014).

Broadly, our approach is as follows. We combine photometric redshift PDFs to compute, for each candidate galaxy pair, the probability that $|\Delta V|$ is within 1000 km s^{-1} ($\mathcal{P}_{\Delta V}$; see Section 3.1) and that their projected separation falls within each separation bin out to $r_p = 100$ kpc ($\mathcal{P}_r^{\text{bin}}$; Section 3.2). Combining $\mathcal{P}_{\Delta V}$ and $\mathcal{P}_r^{\text{bin}}$, we compute the true pair (TP) probability for each pair in each projected separation bin ($\mathcal{P}_{\text{pair}}^{\text{bin}}$; Section 3.3). We then carefully assemble a control sample of physically unassociated projected pairs, matched in redshift, stellar mass, environmental density, and redshift quality

(Section 3.4). Finally, we calculate and compare weighted AGN fractions in the TP and control samples (Section 3.5).

3.1 Relative line-of-sight probabilities

We begin by assembling a pool of candidate projected galaxy pairs, defined as those separated by less than 16.4 arcsec. This distance was chosen because it roughly corresponds to a projected separation of 100 kpc at the lowest redshift of the parent sample ($z = 0.5$). This step ensures that we obtain all pairs with $r_p < 100$ kpc across the full redshift range. Due to our large parameter space and the density of sources in these deep fields, a galaxy has anywhere from 0 to 36 candidate projected pairs in our parent sample within 16 arcsec, all of which we assign a 'TP' probability (described below).

Once we have candidate pairs, we assess the probability that their relative line-of-sight velocity is within 1000 km s^{-1}, $\mathcal{P}_{\Delta V}$. This approach relies on careful incorporation of the photometric redshift PDFs, resulting in the galaxy pairs with wider PDFs receiving lower pair probabilities.

We perform our analysis with and without the available spectroscopic redshifts (see Section 4.1), although we include the spectroscopic information as our primary approach. There are three possible pair types: (1) where both galaxies have a spectroscopic redshift (z_{spec}); (2) both have photometric redshift PDFs; and (3) one of each redshift type. Calculating ΔV and r_p is trivial for type (1), and we assign $\mathcal{P}_{\Delta V} = 0$ or $\mathcal{P}_{\Delta V} = 1$ for these cases, which make up ~1.5 per cent of all initial pair candidates. For evaluating cases (2) and (3), we introduce a probability convolution technique to combine the redshift probability distributions and create a relative line-of-sight velocity PDF, $P(\Delta V)$. We interpret a spectroscopic redshift as a dirac delta distribution, $\delta(z_{\text{spec}})$. Photometric redshift PDFs and $\delta(z_{\text{spec}})$ are henceforth both included in '$P(z)$', as they are treated equally in the methodology.

As we define a TP based on a ΔV threshold, we begin with a simple change of variables from redshift to line-of-sight velocity space, $P(z) \to P(V)$, for galaxies 1 and 2 in a pair. We convolve $P_1(V)$ with $P_2(V)$ to obtain the PDF of their relative line-of-sight velocity, $P(\Delta V)$, following:

$$P(\Delta V) = \int P_1(V_1 = V)\, P_2(V_2 = V + \Delta V)\, dV. \quad (2)$$

We then integrate this distribution within our chosen ΔV threshold ([−1000, 1000] km s^{-1}) to obtain $\mathcal{P}_{\Delta V}$:

$$\mathcal{P}_{\Delta V} = \int_{-1000 \text{ km s}^{-1}}^{1000 \text{ km s}^{-1}} P(\Delta V)\, d\Delta V. \quad (3)$$

In Fig. 3, we provide a visual representation of the probability convolution for the following examples: a high probability case involving a photometric redshift PDF and a z_{spec} (top row); a moderate probability case of two PDFs (middle row); and a low probability case of two PDFs (bottom row). While $\mathcal{P}_{\Delta V}$ increases when the means of the two PDFs are similar, it is appropriately decreased by the individual PDF widths. As a result, we do not need to exclude sources with large redshift uncertainties or correct for chance pairs as is sometimes applied (e.g. Bundy et al. 2009; Mantha et al. 2018). Furthermore, our method penalizes against very broad PDFs regardless of how similar their means are, which is not the case for combined redshift probability methods based purely on PDF overlap (e.g. López-Sanjuan et al. 2015; Duncan et al. 2019, see Section 5.1).

We validate our convolutional pair probabilities with a 'brute-force' Monte Carlo method, where we draw redshifts from their
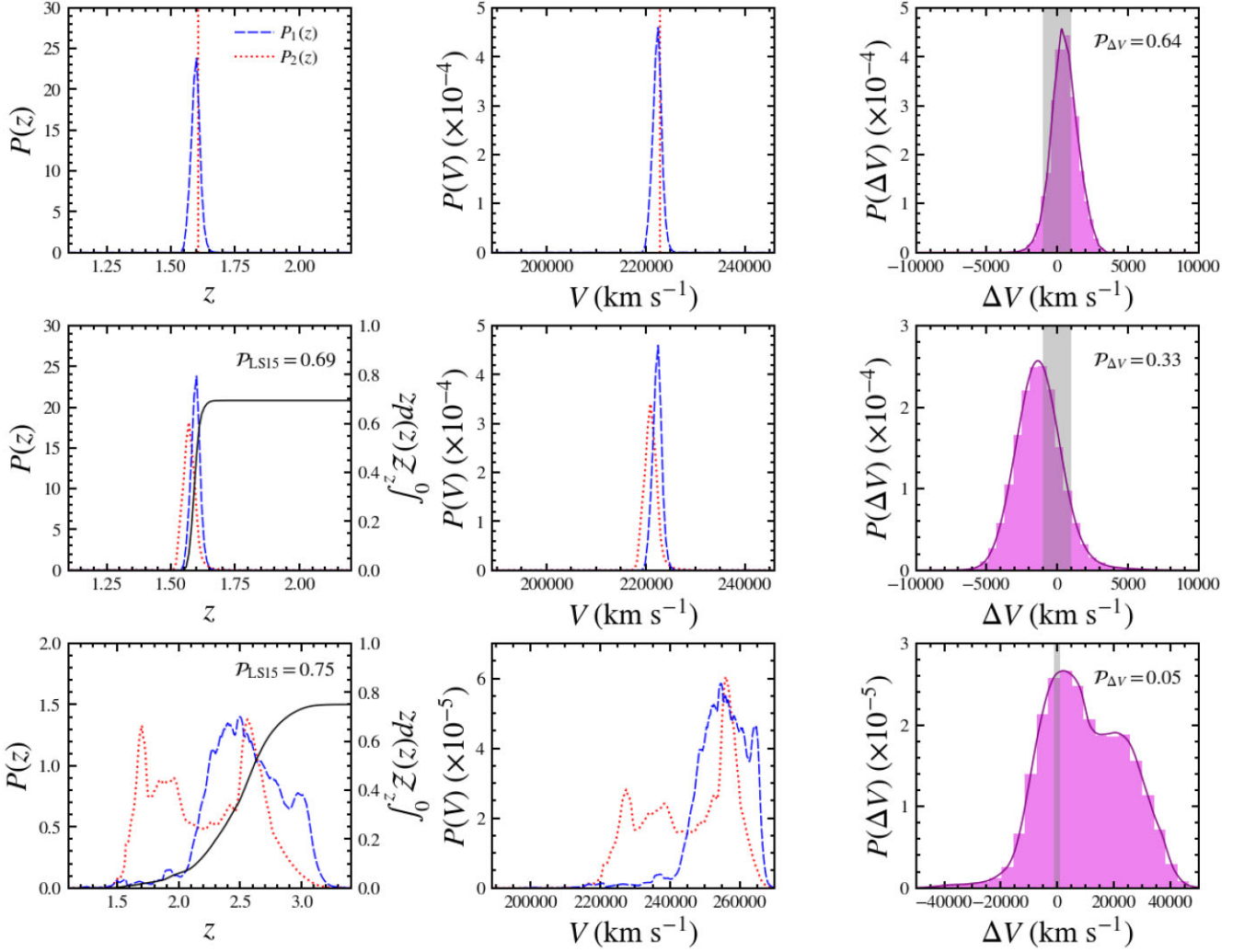
Figure 3. Determination of the relative line-of-sight probability, $\mathcal{P}_{\Delta V}$, for three example projected pairs (one in each row). From top to bottom: *Row 1:* the galaxy pair includes one spectroscopic redshift and one good quality photometric redshift PDF; *Row 2:* two good quality photometric PDFs; *Row 3:* two poor quality PDFs. *Left column*: Overlaid redshift PDFs, $P_1(z)$ (dashed curve) and $P_2(z)$ (dotted curve), where a vertical line corresponds to a spectroscopic redshift. The black curve in the middle and bottom panels is the cumulative integral of the combined PDFs, $\mathcal{P}_{LS15}$, which has been used in previous work to estimate TP probability (see Section 5.1). *Middle column*: line-of-sight velocity distributions. *Right column*: relative line-of-sight velocity distributions computed from equation (2) (solid curve). The purple histogram shows the ΔV distribution from a Monte Carlo simulation, verifying the accuracy of the convolution method. The vertical shaded region indicates the integration bounds of $\mathcal{P}_{\Delta V}$ for our TP definition ($|\Delta V| < 1000\,\mathrm{km\,s^{-1}}$).

underlying PDFs, 10 000 times,[1] with which we calculate ΔV each iteration. These distributions of ΔV (shown as filled histograms) are plotted together with the probability convolutions (shown as curves) in the rightmost panels of Fig. 3, demonstrating their excellent agreement.

3.2 Projected separation probabilities

The projected separation in physical units of a galaxy pair, r_p, is also dependent on redshift (i.e. the conversion from arcseconds to kiloparsecs). Therefore, we require the uncertainties in $P(z)$ to be reflected in our determination of r_p as well, specifically when we wish to assign a galaxy pair into a particular bin of r_p to investigate trends with projected separation as is commonly done (e.g. Ellison et al. 2013b; Shah et al. 2020). Following the logic of the previous

section, we assess the probability, $\mathcal{P}_r$, a pair is within some projected separation, $r_p^{\min}$ and $r_p^{\max}$, from a projected separation PDF, $P(r_p)$. First, we combine $P_1(z)$ and $P_2(z)$ to obtain a distribution of how likely they are to both be at a certain redshift:

$$P(z_1 = z_2) = P_1(z) \times P_2(z). \tag{4}$$

In other words, if the two galaxies were truly associated, this distribution tells us what that shared redshift likely would be. We normalize and perform a change of variables on $P(z_1 = z_2)$ to arrive at $P(r_p)$, which we evaluate between $r_p^{\min}$ and $r_p^{\max}$ to obtain the projected separation probability:

$$\mathcal{P}_r = \int_{r_p^{\min}}^{r_p^{\max}} P(r)\,dr. \tag{5}$$

This process is illustrated in Fig. 4 for the low probability case (bottom row) of Fig. 3.

---

[1] A z_{spec} is drawn each iteration when available for the supplemented sample.
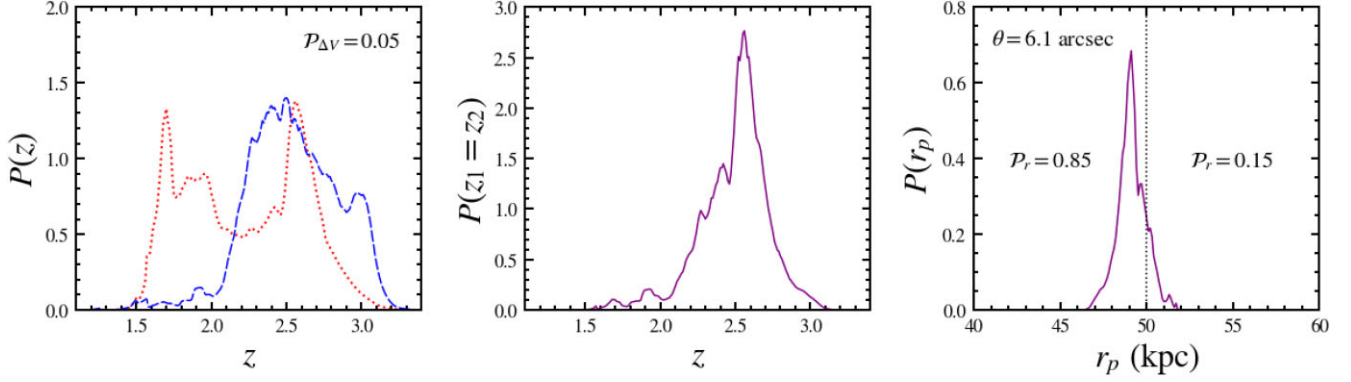
Figure 4. Example to show the method of calculating the projected separation probability, $\mathcal{P}_r$, using the pair shown in the bottom row of Fig. 3, which has an angular projected separation of 6.1 arcsec. *Left-hand panel*: overlaid photometric redshift PDFs, $P_1(z)$ and $P_2(z)$. *Middle panel*: combined redshift distribution (equation 4). *Right-hand panel*: projected separation distribution, $P(r_p)$, obtained from a change of variables of the combined redshift distribution. In this case, there is a probability of $\mathcal{P}_r = 0.85$ that this pair falls in the projected separation bin of 25–50 kpc and a probability of $\mathcal{P}_r = 0.15$ that it falls in the 50–75 kpc bin.

To compare samples versus r_p in a statistically significant way, we sort our pairs into r_p bins of width 25 kpc. To limit the impact of source confusion, we exclude pairs with separations of less than 5 kpc, following previous work on high-redshift pair finding studies (e.g. Mantha et al. 2018; Duncan et al. 2019). The bin boundaries become the integration limits $r_p^{\min}$ and $r_p^{\max}$ in equation (5), from which we determine $\mathcal{P}_r$ for all 4 bin regions (beginning with $\mathcal{P}_r^{5-25\,\mathrm{kpc}}$). This approach ensures that arbitrary bin selection does not misappropriate projected separations estimated from uncertain redshifts.

3.3 True pair probability

To obtain a final TP probability for a candidate galaxy pair in each projected separation bin, $\mathcal{P}_{\mathrm{pair}}^{\mathrm{bin}}$, we combine the relative line-of-sight probability ($\mathcal{P}_{\Delta V}$) and projected separation probability ($\mathcal{P}_r$), following:

$$\mathcal{P}_{\mathrm{pair}}^{\mathrm{bin}} = \mathcal{P}_{\Delta V} \times \mathcal{P}_r^{\mathrm{bin}}. \qquad (6)$$

Continuing the pair example shown in both Figs 3 (bottom row) and 4, with an angular separation of 6.1 arcsec, we calculate $\mathcal{P}_{\mathrm{pair}}^{25\text{-}50\,\mathrm{kpc}} = 0.05 \times 0.85 = 0.0425$ and $\mathcal{P}_{\mathrm{pair}}^{50\text{-}75\,\mathrm{kpc}} = 0.05 \times 0.15 = 0.0075$. We use these pair probabilities as weights to calculate a weighted AGN fraction (see Section 3.5), where more confident pairs (i.e. those with higher $\mathcal{P}_{\mathrm{pair}}$) contribute more, following a similar weighting method previously used by Duncan et al. (2019).

3.4 Control selection

To achieve our goal of establishing the effect of close galaxy pairs on the prevalence of AGNs, we need to compare the AGN frequency in our TP sample to that of a physically unassociated isolated sample controlled for mass, redshift, and environmental density (e.g. Ellison et al. 2013b; Shah et al. 2020). These three properties all may play a role in the rate of AGNs, though they may not be directly related to close galaxy pair separation (e.g. Perez, Tissera & Blaizot 2009). Following Ellison et al. (2013b), we define environmental density as the total number of sources within $|\Delta V| < 1000\,\mathrm{km\,s^{-1}}$ and $r_p < 1\,\mathrm{Mpc}$ over the corresponding projected unit area (calculated using our probabilistic method; see Appendix B). However, we note that this has $10\times$ lower weighting in the matching criteria compared to mass, redshift, and PDF quality. As we interpret

TP probabilities as weights in an ultimate weighted AGN fraction calculation (Section 3.5), we must ensure that any biases associated with PDF quality (i.e. $P(z)$ width) are controlled for as well.

Generally, it is trivial to find several matches for each paired galaxy in the often large pool of isolated galaxies, but $\sim$47 per cent of all galaxies in our parameter space have at least one projected companion with a $\mathcal{P}_{\Delta V} > 0.01$. That is, around half of the galaxies have at least one other galaxy, within 100 kpc, with a projected pair probability greater than 0.01. Since our definition of a TP is not binary but rather continuous (i.e. $\mathcal{P}_{\mathrm{pair}}^{\mathrm{bin}}$ is between 0 and 1), we must treat isolated galaxies as such.

Our full probabilistic approach to selecting control galaxies is supplied in detail in Appendix B. Briefly, for each TP, we select three control pairs (CPs) of physically unassociated galaxies (six unique galaxies; i.e. $\mathcal{P}_r^{0-100} = 0$) matched in mass, redshift, environment, and $P(z)$ width. The three CPs are matched to the relative line-of-sight probability of the TP to ensure statistically comparable weighting (Section 3.5). Normalized distributions of the matched parameters with and without weights are shown in Fig. 5. We report that more than 90 per cent of CPs are matched to within stellar mass of 0.2 dex, redshift of 0.2, environmental density of 4 and $P(z)$ width of 0.2 dex of their TP counterparts (this includes to both galaxy 1 and 2) when $\mathcal{P}_{\Delta V}^{\mathrm{TP}} > 0.01$. This confirms that our control matching process has been successful and that the effects of galaxy proximity on AGN frequency will be isolated in our analysis.

3.5 Weighted AGN fraction and AGN enhancement

With a sample of TPs and a matched sample of unassociated controls (Section 3.4), we can analyse the effects of galaxy separation on AGN fraction. In Sections 3.2, we split our TP sample into four projected separation bins of width 25 kpc by integrating equation (5) between the bin edges. In Section 3.3, we combine these bin probabilities, $\mathcal{P}_r^{\mathrm{bin}}$, with the relative line-of-sight probability of each TP, $\mathcal{P}_{\Delta V}$, to obtain the TP probability for each bin, $\mathcal{P}_{\mathrm{pair}}^{\mathrm{bin}}$ (equation 6). Here, we quantify the frequency of AGN of each bin with a weighted AGN fraction, defined as

$$\text{Weighted AGN Fraction} = \frac{\sum_i (\mathcal{P}_{\mathrm{pair},i}^{\mathrm{bin}} \times N_{\mathrm{AGN},i})}{\sum_i \mathcal{P}_{\mathrm{pair},i}^{\mathrm{bin}}}, \qquad (7)$$

where $N_{\mathrm{AGN},i}$ is the number of AGN (0, 1, or 2) in pair i. Rather than counting AGNs in all paired galaxies, we count AGNs in galaxy
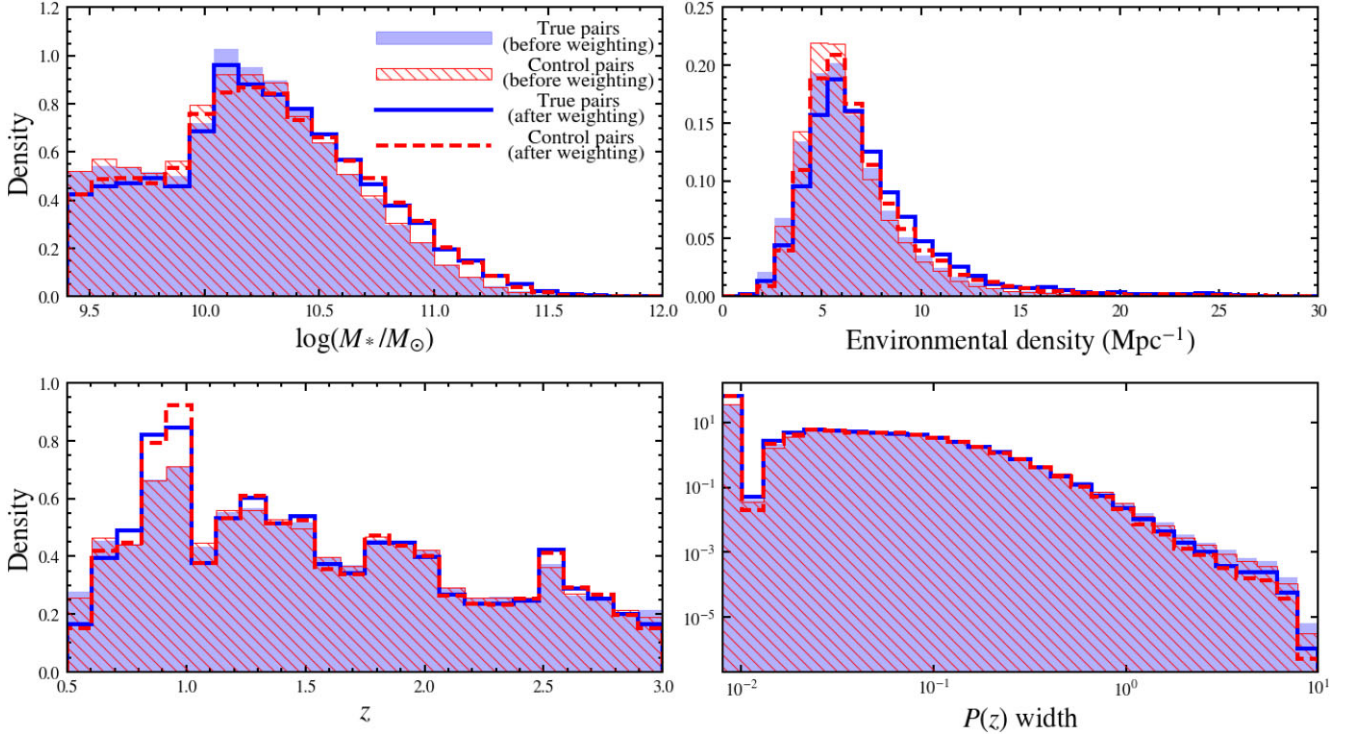
Figure 5. Distributions of stellar mass (*top left*), redshift (*bottom left*), environmental density (*top right*), and $P(z)$ widths (*bottom right*) for all TPs and isolated controls. The shaded blue and striped red histograms correspond to the true and control distributions without weighting by pair probability, $\mathcal{P}_{\text{pair}}$. The solid blue and dashed red step histograms show the TP and CP distributions after applying the weighting used in the analyses, which demonstrates that the controls are well matched in these properties.

pairs weighted by how likely they are to be truly associated. In other words, the weighted AGN fraction in each pair is the weighted mean of AGN counts per pair and ranges from 0 to 2. In order to isolate and quantify the effects of galaxy separation on AGN activity, we need a weighted AGN fraction of the CPs, which we calculate from equation (7) by replacing $\mathcal{P}_{\text{pair}}^{\text{bin}}$ with the control relative line-of-sight probability multiplied by the corresponding TP projected separation probability. Multiplying $\mathcal{P}_{\Delta V}^{\text{CP}}$ by $\mathcal{P}_r^{\text{bin}}$ ensures, for instance, that a TP of $\mathcal{P}_{\Delta V}^{\text{TP}}$ split between the 25–50 and 50–75 kpc bins will have its matched CPs of $\mathcal{P}_{\Delta V}^{\text{CP}} \approx \mathcal{P}_{\Delta V}^{\text{TP}}$ shared between bins in the same way (Fig. 4).

Finally, we define AGN enhancement (also known as AGN excess) as the ratio of the weighted AGN fraction for the TPs and CPs of each bin:

$$\text{AGN Enhancement} = \frac{\text{Weighted AGN Fraction}_{\text{TP}}}{\text{Weighted AGN Fraction}_{\text{CP}}}. \quad (8)$$

We use a common bootstrap technique (Efron 1979, 1981) to estimate the standard error for the weighted AGN fractions, which we propagate to estimate errors on AGN enhancement.

In Appendix C, we apply this method, using only photometric redshifts, to a mock sample of galaxies with a true AGN enhancement. Although it is very difficult to capture the full complexities of the data sets, and caution should be taken in over-interpreting the quantitative simulated results, the simulation aims to roughly mimic the quality of the PDFs used in this work. We find that even with only photometric redshifts, we are able to recover an increase in the weighted AGN fraction of close galaxy pairs. The measured AGN fractions may be moderately underestimated on our main result presented for the obscured AGN in Section 4.3; however, the large error bars on this final result capture some of this uncertainty on the exact level of

enhancement. Only with a large number of spectroscopic redshifts would we be able to provide a more precise final measurement.

4 RESULTS

In this section, we report weighted AGN fractions (i.e. the weighted average AGN count per single galaxy pair) and AGN enhancements (relative to matched control galaxies), as defined in Section 3.5. We show this separately for X-ray and IR AGN samples, and we investigate any trends in these quantities as a function of projected separation (Section 4.1). We go on to define sub-samples to explore any additional trends in these results with X-ray luminosity or redshift (Section 4.2). Finally, we investigate a sample of AGN which are only identified in the IR but not selected in the X-rays (Section 4.3). Our main results incorporate spectroscopic redshifts; however, we present some results using solely photometric redshift PDFs for validation (shown as grey data points in Fig. 6). All data plotted is tabulated in Appendix D.

4.1 Full sample

The weighted AGN fractions in bins of projected separation for all X-ray AGN ($L_X > 10^{42}$ erg s^{-1}, $0.5 < z < 3$) are presented for our major-complete and minor-complete samples in the top-left panel of Fig. 6. For both the major and minor samples (filled coloured points), we see no trend of weighted AGN fraction as a function of projected separation, with all values scattering around $\sim$0.07. The control galaxies (hollow points) are always in good agreement with the TPs. Alongside the primary data points, which include both spectroscopic and photometric information, we show the results from
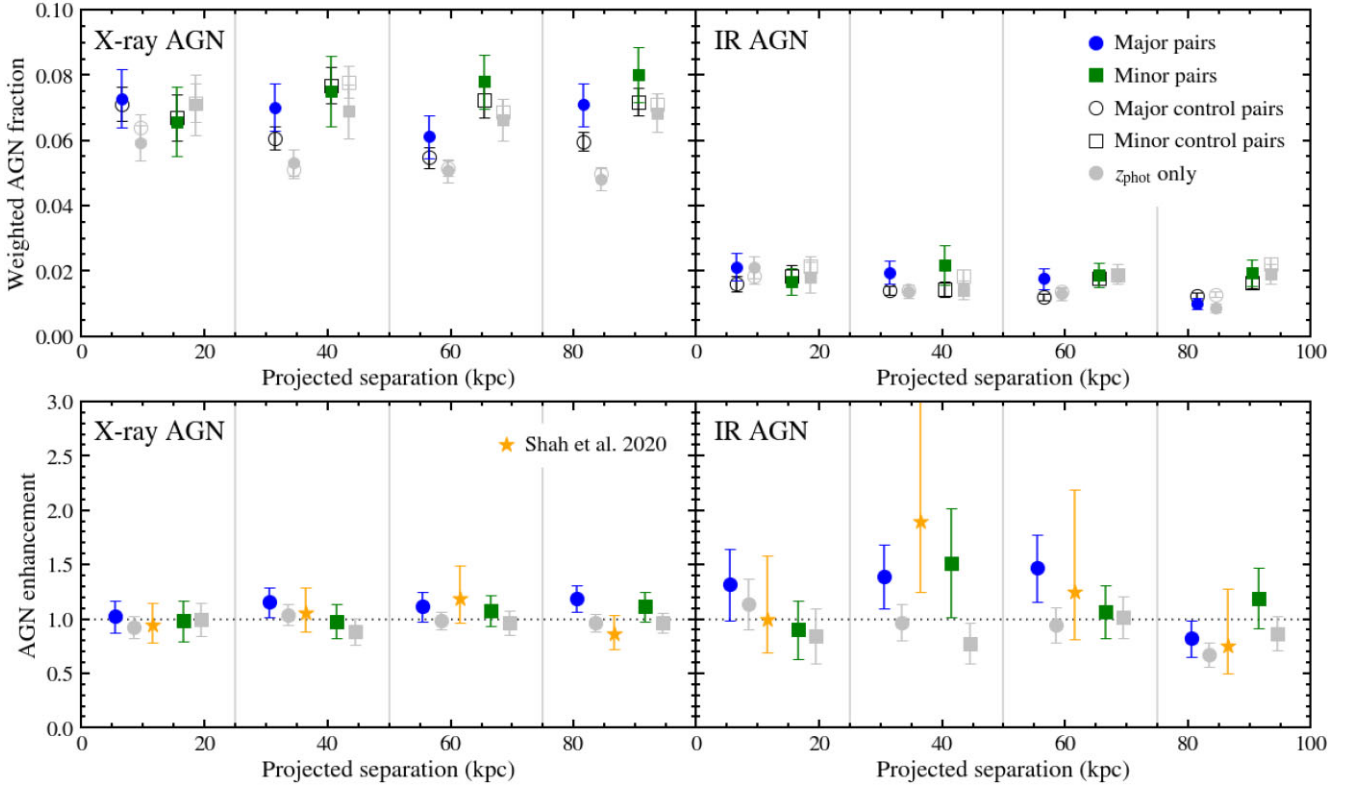
Figure 6. Results of weighted AGN fractions (*top row*) and AGN enhancement (relative to the controls; *bottom row*) as a function of projected separation. AGN fractions and enhancements for all $0.5 < z < 3$ X-ray AGNs and IR-selected AGNs are shown on the left and right, respectively. In all panels, the filled blue circles and filled green squares indicate major and minor galaxy TPs, respectively, as defined in Fig. 1. Open circles and squares indicate the weighted AGN fractions for the corresponding control samples. Grey points show the same weighted AGN fractions found without the inclusion of spectroscopic redshifts, denoted as $z_{\rm phot}$ only. Also plotted are AGN enhancements of major galaxy pairs from Shah et al. (2020), who only used spectroscopic redshifts in the same fields. Vertical grey lines indicate the projected separation bin edges. Data points are horizontally offset for visual aid. The dotted horizontal lines correspond to an AGN enhancement of 1; all data points scatter about this line, indicating no enhancement of AGN in true galaxy pairs relative to the control.

the purely photometric redshift approach, denoted '$z_{\rm phot}$ only' (not using $\delta(z_{\rm spec})$). We remind the reader that spectroscopic surveys also specifically target X-ray sources, which can boost the spectrosopic redshift fraction (see Section 2.4) and thus the chances of X-ray sources being detected in a pair. Although the weighted AGN fraction values are slightly lower ($\sim$0.055) for the major galaxy pair sample when spectroscopic redshifts are not used, this decrease is matched by the corresponding control galaxy pairs.

X-ray AGN enhancement, that is the ratio of the weighted AGN fractions of the TPs and CPs, is plotted for these samples in the bottom-left panel of Fig. 6. The dotted line at an AGN enhancement value of 1 illustrates the case where the weighted AGN fractions are the same for both the TPs and controls (i.e. no enhancement). Both major and minor pairs exhibit no evidence for AGN enhancement at small projected separations ($r_p < 25$ kpc), with values of 1.02 ± 0.15 and 0.98 ± 0.19, respectively. This finding holds regardless of the inclusion of spectroscopic redshifts, with the $z_{\rm phot}$ only method resulting in enhancements of 0.93 ± 0.10 and 0.99 ± 0.15 for major and minor galaxy pairs, respectively. Additionally, this result is consistent with Shah et al. (2020), who find no X-ray AGN enhancement for major pairs while only using spectroscopic redshifts for the same fields (shown with yellow stars in Fig. 6).

We plot the corresponding results for all IR AGNs (following Donley et al. 2012, $0.5 < z < 3$) in the right-hand panels of Fig. 6. Given the conservative selection criteria, there are $\sim$4 $\times$ fewer IR AGNs than X-ray AGNs. As a result, IR AGN enhancements have

larger errors. We find no statistically significant AGN enhancement with decreasing projected separation. For the smallest separations ($r_p < 25$ kpc), the enhancement values are 1.32 ± 0.33 in major and 0.90 ± 0.27 in minor pairs. Once again, these results are consistent to those of the $z_{\rm phot}$ approach, which results in enhancements of 1.14 ± 0.23 and 0.84 ± 0.25 for major and minor galaxy pairs, respectively, at the smallest separations. These findings are consistent with the IR AGN enhancements reported by Shah et al. (2020) in $0.5 < z < 3$ major galaxy pairs using only spectroscopic redshifts, albeit their error bars are significantly larger due to lower number statistics.

The results presented in the following two subsections focus on major galaxy pairs with the incorporation of spectroscopic redshifts where available.

4.2 X-ray luminosity and redshift sub-samples

Given the varying depths of the *Chandra* X-ray observations over the different fields, we investigate AGN enhancement as a function of projected separation for the more complete X-ray AGN samples defined in Section 2.4. These X-ray AGN samples are moderate ($L_X = 10^{43.2–43.7}$ erg s^{-1} at $0.5 < z < 2$) and high L_X samples ($L_X > 10^{43.7}$ erg s^{-1} at $0.5 < z < 3$; see Fig. 2).

Moderate and high L_X AGN enhancements for major galaxy pairs are presented in the left-hand panel of Fig. 7. Just as with the full sample of X-ray AGNs, both the moderate and high L_X AGN samples are consistent with no enhancement across all projected
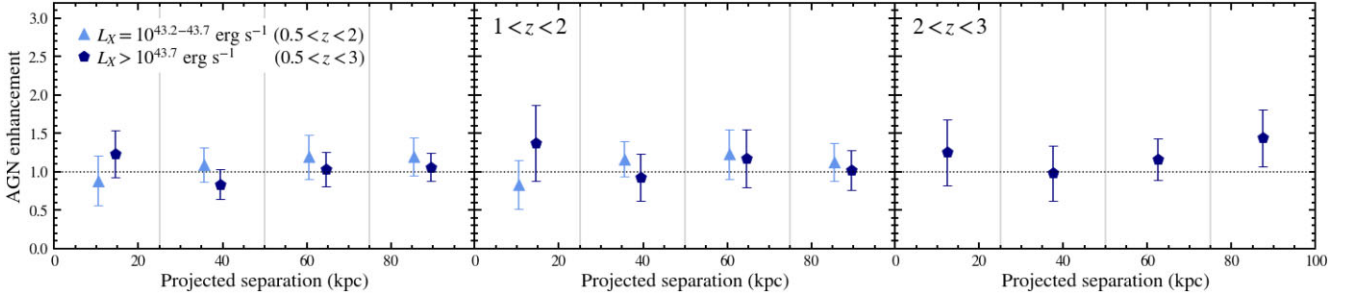
Figure 7. AGN enhancements for our varying X-ray luminosity samples (defined in Fig. 2) in major galaxy pairs in bins of projected separation. *Left-hand panel*: AGN enhancement for moderate L_X (sky-blue triangles) and high L_X AGN (navy pentagons) over their respective complete redshifts, $0.5 < z < 2$ and $0.5 < z < 3$. *Middle and right-hand panels*: AGN enhancements in different redshift range, $1 < z < 2$ and $2 < z < 3$. The moderate L_X AGN sample is not plotted in the right-hand panel as not all fields are complete for these X-ray luminosities at $z > 2$. Vertical grey lines indicate the projected separation bin edges, and each data point is computed from all pairs in its respective bin (horizontally offset for visual aid). The data points scatter around an AGN enhancement of 1 (shown with the dotted lines), indicating that the AGN fraction in the TP and control samples is the same across all separations.

separation bins. For major galaxy pairs within $r_p < 25$ kpc, we report enhancement values of 0.88 ± 0.32 and 1.23 ± 0.30 for moderate and high L_X AGN samples, respectively. We further investigate whether these results are redshift-dependent by measuring the enhancement of moderate and high L_X AGN at $1 < z < 2$ and $2 < z < 3$. As shown in the middle and right-hand panels of Fig. 7, we find no enhancement of X-ray AGN in major galaxy pairs at any redshift. Shah et al. (2020) also report no enhancement across different X-ray luminosity and redshift bins. We note that with these data we are unable to rigorously investigate AGN enhancement at $z < 1$ in a meaningful way due to low number statistics. While the literature provides some evidence on the redshift-dependency of the AGN–merger connection down to lower redshifts (e.g. Silverman et al. 2011; Goulding et al. 2018), we refrain from directly addressing this question here.

4.3 Obscured AGN

Theoretical models predict, and observational works suggest, an increase in nuclear obscuration during advanced stages of a major merger independent of viewing angle (e.g. Hopkins et al. 2008; Kocevski et al. 2015; Lansbury et al. 2017; Ricci et al. 2017; Blecha et al. 2018; Hickox & Alexander 2018; Ricci et al. 2021). As AGN emission is absorbed and re-emitted in the IR by circumnuclear dust, IR AGN selection is less susceptible to nuclear obscuration than selection via shorter wavelengths (e.g. Alonso-Herrero et al. 2006; Donley et al. 2007, 2010, 2012; Hickox & Alexander 2018). We do not have high-quality constraints on X-ray obscuration (N_H) across all fields; furthermore, the most obscured sources could be completely undetected in the X-ray surveys. Therefore, following Satyapal et al. (2014) and Weston et al. (2017), we define an 'obscured' AGN sample as those identified exclusively in the IR. Specifically, we define this sample as those AGNs which are identified in the IR (following Donley et al. 2012) but not detected in the X-rays above the AGN threshold ($L_X > 10^{42}$ erg s^{-1}; 'IR-only AGNs'). While all X-ray-detected AGNs are not necessarily unobscured, for the purposes of this analysis, we assume that the IR-only detected sample is more obscured (Andonie et al. 2022). Of the 709 IR AGN in our full sample, ~48 per cent are either not detected in X-rays or fall below this threshold, which corresponds to 338 'obscured AGN'.

We calculate the fraction of 'obscured' AGN compared to 'unobscured' AGN in bins of projected separation, as

$$\frac{\text{Obscured}}{\text{Unobscured}} = \frac{\sum_i \mathcal{P}_{\text{pair}} \times N_{\text{IR-only AGN},i}}{\sum_i \mathcal{P}_{\text{pair}} \times N_{\text{X-ray AGN},i}}, \quad (9)$$
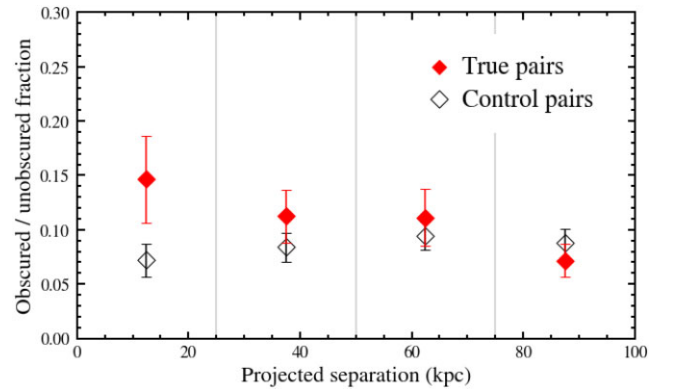

Figure 8. Obscuration fraction in bins of projected separation for major galaxy pairs. Obscuration fraction is defined as the ratio of the number of AGNs selected only in the IR to the number selected in the X-ray, weighted by the appropriate pair probabilities (equation 9). The red and empty diamonds correspond to the obscuration fraction of TPs and CPs, respectively.

where, $N_{\text{IR-only AGNs}}$ and $N_{\text{X-ray AGNs}}$ denote the number of AGNs in a pair (0, 1, or 2) selected exclusively in the IR and those selected in the X-ray (irrespective of IR selection), respectively. Just as in calculating the weighted AGN fraction, we weight our IR-only and X-ray AGN counts by their corresponding pair probabilities, and we estimate the standard error on the obscuration fraction with a common bootstrap technique (Efron 1979, 1981).

In Fig. 8, we plot this ratio for the TP and CP samples as a function of projected separation. For TPs of separation $r_p > 25$ kpc, the obscuration fraction is largely consistent with that of the control, albeit trending upward towards lower projected separations. For TPs of separation $r_p < 25$ kpc; however, the obscuration fraction substantially differs from that of the CPs (14.6 ± 3.8 per cent versus 7.2 ± 1.4 per cent), which supports the presence of increased AGN obscuration at the smallest projected separations (Satyapal et al. 2014; Weston et al. 2017, see discussion in Section 5.2.2).

We determine the weighted AGN fraction and AGN enhancement for our 'obscured' AGN sample and present these results in Fig. 9. While the obscured AGN fraction of the *control group* remains stable at all separations (with 0.51 ± 0.09 at $r_p < 25$ kpc), the obscured AGN fraction in TPs increases continuously from 0.68 ± 0.15 per cent at $50 < r_p < 75$ kpc to 1.06 ± 0.25 per cent at $r_p < 25$ kpc. This corresponds to an increase in obscured AGN *enhancement* at decreasing projected separations peaking at 2.08 ± 0.61 for the
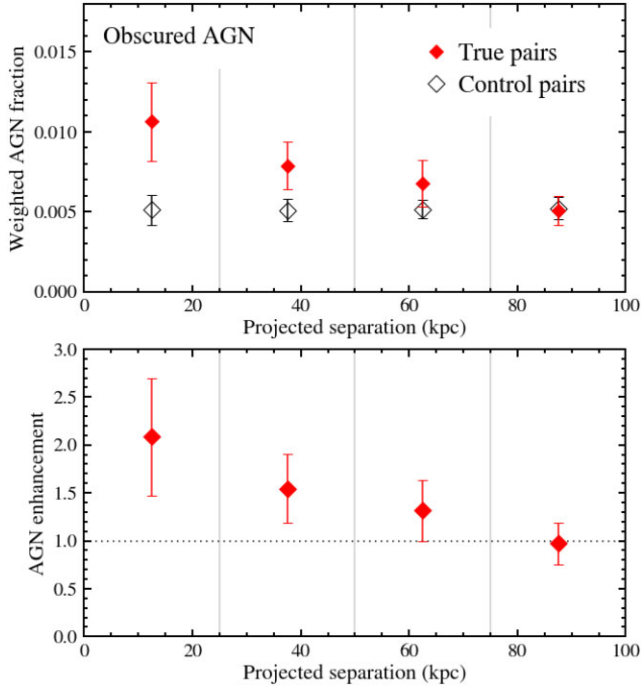
Figure 9. Weighted AGN fraction (*top*) and AGN enhancement (*bottom*) of obscured AGN in major galaxy pairs binned by projected separation. The red and empty diamonds correspond to the weighted AGN fractions for the TPs and CPs, respectively. The AGN enhancement systematically increases from a value of 1 (dotted horizontal line; indicating no enhancement) with decreasing projected separation.

closest projected pairs. However, as detailed in Section 3.5, we cannot rule out that we are slightly underestimating the true AGN enhancement due to the challenges of dis-entangling the TPs and CPs with only photometric redshift PDFs.

5 DISCUSSION

Due to our new pair probability approach, we have been able to incorporate the largest (known) sample of high-redshift galaxy pairs into a merger-induced AGN enhancement calculation.[2] We begin this section by discussing the advantages of our pair probability method compared to other pair-finding techniques that make use of photometric redshift PDFs (Section 5.1). Then, we discuss our findings on the connection between close galaxy pairs and AGN enhancement in the context of previous work and a merger-driven evolutionary scenario for heavily dust-obscured black hole growth (Section 5.2).

5.1 Photometric PDF approaches for pair probabilities

Our method for calculating the probability that two galaxies are in a physically associated pair involves calculating the probability that two galaxies are within a relative line-of-sight velocity of $|\Delta V| < 1000 \, \mathrm{km \, s^{-1}}$ (Section 3.1) and a projected separation of $r_p < 100 \, \mathrm{kpc}$ (Section 3.2) from a convolution of redshift photometric redshift PDFs or spectroscopic redshifts (Section 3.3). A major strength of

our probabilistic approach is that we include the full underlying uncertainties of the photometric redshifts, but we neither need to exclude photometric redshifts based on some arbitrary uncertainty threshold nor apply a correction for chance (not physically associated) projected pairs, which breaks down for dense environments (e.g. Le Fèvre et al. 2000; Kartaltepe et al. 2007; Bundy et al. 2009).

Another method has previously been developed to circumvent the challenges of using photometric redshift PDFs with a broad range of quality/widths (López-Sanjuan et al. 2015). This method was expanded by Mundy et al. (2017) and Duncan et al. (2019) to calculate weighted merger fractions in the CANDELS fields while correcting for stellar mass completeness. For galaxy 1 and 2 in a projected pair, the method compares photometric redshift PDFs, $P(z)$, using the combined redshift probability function, $\mathcal{Z}(z)$, defined as:

$$\mathcal{Z}(z) = \frac{2 \times P_1(z) \times P_2(z)}{P_1(z) + P_2(z)}. \tag{10}$$

A numerical pair probability that these galaxies are physically associated, $\mathcal{P}_{\mathrm{LS15}}$, is then obtained by taking the cumulative integral of $\mathcal{Z}(z)$ over the full redshift range:

$$\mathcal{P}_{\mathrm{LS15}} = \int_0^z \mathcal{Z}(z) dz. \tag{11}$$

We show the resulting integrals and $\mathcal{P}_{\mathrm{LS15}}$ values, for the example PDFs in Fig. 3 (see the black curves and inset text in the middle-left and bottom-left panels). This approach effectively determines pair probability based on the amount of *overlap* of the two photometric redshift PDFs. For a set of narrow PDFs, the method is successful in establishing which of the galaxies are most likely to be physically associated. For example, in the middle-left panel of Fig. 3, the PDFs are well constrained and their peak values are close. The combined redshift method gives a pair probability of $\mathcal{P}_{\mathrm{LS15}} = 0.69$. However, the PDFs of the projected pair in the bottom left panel are very poorly constrained, i.e. they are very broad and feature multiple peaks. In this case, equation (11) gives a probability of $\mathcal{P}_{\mathrm{LS15}} = 0.75$, which is larger than the previous example for the well-defined PDFs. The values of our relative line-of-sight probabilities for the same example pairs are $\mathcal{P}_{\Delta V} = 0.33$ and $\mathcal{P}_{\Delta V} = 0.05$, respectively (for the middle and bottom row examples in Fig. 3). These $\mathcal{P}_{\Delta V}$ values appear to more clearly reflect the relative quality of the underling PDFs, and hence the greater uncertainty in the two galaxies being in physically associated pairs.

In contrast to assessing the overlap between two PDFs, our convolution method results in pair probabilities that are larger when the peaks of the two PDFs are similar but smaller when the individual PDF widths are larger. This guarantees that sources with less certain redshifts will always be assigned a smaller probability of being in a physically associated pair.

Whilst a 'brute-force' Monte Carlo approach can achieve the same results as the convolution method (i.e. by randomly drawing redshifts from PDFs and calculating ΔV for all projected pairs each iteration; see Fig. 3), the convolution approach is significantly more efficient due to the computational ease of calculating a convolution as the product of two fast Fourier transforms. This makes it applicable to the large samples required to rigorously investigate AGN activity, or any other galaxy properties, in galaxy pairs as a function of projected separation.

5.2 The association of galaxy pairs with different AGN types

Studies of low-redshift galaxies find an excess of disturbed morphological features indicative of a galaxy merger or interaction in AGNs,

---

[2]To give an indicative number of the sample size, we report that 22 295 major galaxy pairs of projected separation $r_p < 100 \, \mathrm{kpc}$ have a TP probability $\mathcal{P}_{\mathrm{pair}} > 0.01$.

compared to samples of inactive galaxies (e.g. Guyon, Sanders & Stockton 2006; Ramos Almeida et al. 2011; Goulding et al. 2018; Koss et al. 2018; Ellison et al. 2019; Gao et al. 2020; Pierce et al. 2022, 2023). An enhancement of AGN is also found in physically associated projected pairs, relative to matched control galaxies, but with the quantitative results being dependent on the AGN selection method (e.g. Ellison et al. 2011, 2013b; Satyapal et al. 2014). In this work, we have investigated the connection between physically associated galaxy pairs and the number of X-ray and IR-selected AGNs at cosmic noon ($0.5 < z < 3$). We have built on the work of Shah et al. (2020), who took a similar approach but was limited to using only spectroscopic redshifts to identify which galaxies are in physically associated pairs. In the following, we discuss our results for X-ray AGNs and the most obscured AGNs in the context of an AGN-merger connection at cosmic noon.

5.2.1 X-ray AGNs are not enhanced in close pairs at cosmic noon

Our results are consistent with no enhancement for X-ray-selected AGNs (i.e. $L_X > 10^{42}$ erg s^{-1}) in major (or minor) galaxy pairs at close projected separations (see Fig. 6). This is strong evidence that X-ray AGNs are not preferentially associated with galaxy mergers/interactions in the CANDELS and COSMOS fields. This is in agreement with the spectroscopic redshift sample of Shah et al. (2020). Many other observational studies at high redshift also report no merger-induced AGN enhancement in X-ray AGN populations. Working in the COSMOS field, Cisternas et al. (2011) find no difference in the frequency of morphological disturbances indicative of an ongoing merger between X-ray-selected AGNs and inactive galaxies at $0.3 < z < 1$. Likewise, Kocevski et al. (2012) report that X-ray AGNs and inactive control galaxies at $1.5 < z < 2.5$ are highly disturbed at consistent rates in GOODS-S ($16.7^{+5.3}_{-3.5}$% versus $15.5^{+2.8}_{-2.2}$%). Both these works note that the majority of X-ray AGNs reside in disc galaxies, suggesting secular processes to be the more efficient fueling mechanism for these systems (see also, Rosario et al. 2015). Silva et al. (2021) also find no difference in the fraction of multiwavelength selected AGNs in merger and non-merger galaxies at cosmic noon in the CANDELS fields.

We also find no difference in enhancement between our moderate ($L_X = 10^{43.2-43.7}$ erg s^{-1}) and high L_X ($>10^{43.7}$ erg s^{-1}) AGN samples in major galaxy pairs (see Fig. 7). Specifically, for both samples the AGN enhancement in galaxy pairs separated by less than 25 kpc, is consistent with 1 (i.e. no enhancement; 0.88 ± 0.32 and 1.23 ± 0.30, respectively). No trend of AGN enhancement with X-ray luminosity is similarly reported by other works at cosmic noon (e.g. Villforth et al. 2014, 2017; Hewlett et al. 2017; Shah et al. 2020).

Our results are not unexpected when compared to recent cosmological simulations. AGN frequency in merging galaxies has been investigated in EAGLE (McAlpine et al. 2020), which identifies AGNs based on a bolometric luminosity threshold ($L_{\text{bol}} > 10^{43}$ erg s^{-1}; also see Byrne-Mamahit et al. 2023 for similar results based on IllustrisTNG). If we assume that X-ray emission is a reasonable tracer of instantaneous accretion rates, we can make an apt comparison to this work. For the nearby Universe ($0.05 < z < 0.1$), McAlpine et al. (2020) report a modest AGN enhancement ($\sim$1.3) at physical separations less than 30 kpc. However, at $z > 1$, they report no enhancement even for the most rapidly accreting SMBHs ($\lambda_{\text{edd}} > 0.01$), which is consistent with our analysis of X-ray selected AGNs in galaxy pairs at $0.5 < z < 3$.

All together, the evidence supports that X-ray AGNs are not preferentially associated with galaxy interactions or mergers at

cosmic noon. However, it is critical to consider AGN selection. For example, Secrest et al. (2020) find no statistically significant enhancement (a value of $2.22^{+4.44}_{-2.22}$) of X-ray AGNs in post-mergers at $0.03 < z < 0.15$, which is in stark contrast to the $\sim$17 factor excess of MIR AGN they find in the same sample. Furthermore, Satyapal et al. (2014) report an $\sim$4–6 × excess of IR-selected AGN in spectroscopic galaxy pairs at $0.01 < z < 0.2$ separated by less than 10 kpc, while Ellison et al. (2013b) find only an $\sim$2.5 × excess in AGN selected from optical emission lines at the same redshift. Using identical MIR and optical AGN selection, Ellison et al. (2019) report that almost 60 per cent of MIR AGNs are visually disturbed (indicative of a post-merger), double that of optically identified AGNs (see also, Bickley et al. 2023). Therefore, in the following we discuss the results for the most 'obscured AGN' at cosmic noon.

5.2.2 Obscured AGN enhancement in close pairs at cosmic noon

We find evidence for enhanced AGN obscuration at late merging stages ($<$25 kpc), defined by an increased fraction of AGN identified only through their IR emission and not through X-rays (Fig. 8). Additionally, we observe evidence for an increasing AGN enhancement at decreasing projected separations in this obscured AGN sample (Fig. 9). Our results qualitatively match results at low redshift, which together might favour merger-induced obscuration and enhanced black hole growth. For example, Weston et al. (2017) report a similar excess in obscuration fraction (factor of $\sim$2–6) when comparing the MIR- to optical emission-line-selected AGNs in merging and isolated nearby galaxies. Satyapal et al. (2014) find that the ratio of AGN selected exclusively in the MIR (from *WISE* colour cuts) to that selected in the optical (i.e. emission-line *Seyferts*; may also be selected in the MIR), increases significantly for pairs separated by less than 30 kpc compared to a matched control sample. They report ratio excesses of $>$2 and $\sim$4 for pairs separated by less than 10 kpc and for visually identified post-mergers, respectively.

IR AGN selection is more effective during the most obscured phases of black hole growth, compared to optical emission-line or X-ray selection, which can be faint or completely invisible (e.g. Sanders et al. 1988a, b; Goulding & Alexander 2009; Veilleux et al. 2009; Snyder et al. 2013; Kocevski et al. 2015; Blecha et al. 2018; Hickox & Alexander 2018; Andonie et al. 2022). It follows that if obscured galaxy growth is linked to merging and interacting galaxies, then AGN selection most sensitive to this population would exhibit the strongest AGN enhancement. Indeed, simulations predict that intense nuclear starbursts induced during the late stage of a gas-rich major merger obscure the SMBH with hot dust and gas (Cattaneo et al. 2005; Hopkins et al. 2008) and that the SMBH may accrete significant mass during a relatively short obscured phase (e.g. Fabian 1999; Treister et al. 2009; Draper & Ballantyne 2010). Whilst mergers may not be relevant for the majority of AGNs or black hole growth (e.g. McAlpine et al. 2020; Byrne-Mamahit et al. 2023), these discrete events may play an important evolutionary phase for the host galaxy and host halo, due to boosted AGN luminosities, disruptions of the galaxy disc and consequently effective AGN feedback (Davies, Pontzen & Crain 2023; Quai et al. 2023).

Robust and uniformly measured X-ray column densities are not available for our X-ray AGN sample, which could provide another method to investigate trends with obscuration. However, we note that at low redshift, Koss et al. (2018) report a significant enhancement of late-stage nuclear mergers in AGN selected in *hard* X-rays (14–195 keV from *Swift*/BAT), suggesting soft X-rays (i.e. those observed by *Chandra*) are unable to detect AGNs in the heavily obscured

phase of a merger. Furthermore, for cosmic noon AGNs, using an X-ray spectral analysis of 154 $z \sim 1$ AGNs in the CANDELS fields, Kocevski et al. (2015) find that the frequency of disturbed morphologies, indicative of a merger/interaction, in Compton-thick AGN is nearly three times that of an unobscured AGN sample ($21.5^{+4.2}_{-3.3}\%$ versus $7.8^{+1.9}_{-1.3}\%$). However, many obscured AGN – and the ones most associated with mergers based on these results – may be hidden from the *Chandra* X-ray surveys covering the CANDELS and COSMOS fields (Andonie et al. 2022). Indeed, this is strongly supported by Donley et al. (2018), who find in the CANDELS portion of the COSMOS field that AGN detected only in the IR (i.e. receptive to obscured AGN) are more than twice as likely than those selected exclusively in the X-ray ($75^{+8}_{-13}\%$ versus $31^{+6}_{-6}\%$) to have disturbed host morphologies indicative of a merger/interaction. Therefore, merger-induced obscuration could explain why an enhancement of AGN is not seen for the typical X-ray selected samples at cosmic noon.

We are in agreement with several other works that there seems to be an increased association between obscured AGNs and interactions/mergers, compared to unobscured AGNs. However, it is important to address the differences and limitations among different AGNs selection methods. For example, large samples of IR AGNs are often selected at low redshift via a colour–colour cut. Such a selection requires AGN emission to dominate over that of the host (Mateos et al. 2012; Gürkan, Hardcastle & Jarvis 2014). Beyond $z \gtrsim 1$, strong stellar emission compounds this luminosity bias, making selecting clean samples of IR AGN difficult (Blecha et al. 2018). Stellar contamination is addressed by Donley et al. (2012), who implement a power-law selection, in addition to a colour–colour cut, to identify clean AGN samples from *Spitzer*/IRAC data out to $z \sim 3$. Such a strict criteria again requires dominant AGN emission in the IR, making this selection conservative towards the most luminous AGNs. Therefore, more complete and confident samples of IR AGNs are required to better constrain their intrinsic properties, their connection to interactions/mergers and a robust comparison to luminosity-matched samples of unobscured AGNs. This will be possible with very large samples that are covered by comprehensive SED analyses, with improved constraints on X-ray column densities and with improved quality infrared photometry from *JWST* (e.g. Satyapal et al. 2021; Andonie et al. 2022).

Finally, we note that our results on the most obscured sources would not be possible without the inclusion of photometric redshifts because the obscured AGN sample has significantly less spectroscopic coverage (8.3 per cent) than X-ray AGN (30 per cent). Future spectroscopic surveys, targeting the most obscured AGN, will provide further insight into this important population (e.g. Merloni et al. 2019).

6 CONCLUSIONS

We have developed a novel technique for identifying high probability projected galaxy pairs (i.e. those highly likely to be physically interacting with line-of-sight velocities within $\pm 1000 \, \mathrm{km \, s^{-1}}$) using convolutions of photometric redshift PDFs. This enables us to compile the largest sample of high-redshift galaxy pairs ever used for a merger-induced AGN enhancement calculation. Our work is an important addition to studies limited to only spectroscopic redshifts, as they lack complete information on the line-of-sight velocity for *all* galaxy pairs within a given projected separation.

We use our technique to translate pair probabilities into weights when calculating the fraction of X-ray and IR AGNs in $0.5 < z$

< 3 projected major ($M_{*,1} > 10^{10} \, \mathrm{M_\odot}$ and $M_{*,1}$: $M_{*,2} < 4$: 1) and minor galaxy pairs ($M_{*,1} > 10^{10.4} \, \mathrm{M_\odot}$ and $4{:}1 < M_{*,1}$: $M_{*,2} < 10$: 1) in the CANDELS and COSMOS fields. We take a complementary probabilistic approach to identify highly likely isolated galaxies from which we assemble a control sample matched to the TPs in redshift, stellar mass, environment, and redshift quality. We compute the AGN enhancement as the ratio of the weighted AGN fractions of the truly associated pairs to the corresponding control group. Our key findings are:

(i) We find no evidence of an enhancement of X-ray AGN ($L_X > 10^{42} \, \mathrm{erg \, s^{-1}}$) in major or minor galaxy pairs for projected separations of 5–100 kpc. Specifically, in the closest separation bin of 5–25 kpc we find an AGN enhancement values of 0.98 ± 0.19 and 1.02 ± 0.15 for the minor and major pairs, respectively (Fig. 6).

(ii) We further explore two X-ray AGN samples more complete to the sensitivity of observations across all fields investigated: (1) moderate luminosity ($L_X = 10^{43.2-43.7} \, \mathrm{erg \, s^{-1}}$; $z = 0.5 - 2$) and; (2) high luminosity ($L_X > 10^{43.7} \, \mathrm{erg \, s^{-1}}$; $z = 0.5 - 3$). We still find no evidence for an AGN enhancement in the close galaxy pairs, nor do we observe differences as a function of redshift or X-ray luminosity (Fig. 7).

(iii) Defining the most 'obscured AGN' as those that are identified in the IR (Donley et al. 2012) but are not detected in the X-rays, we find the fraction of obscured AGN increases with decreasing projected separation. Specifically, we find that major galaxy pairs separated by less than 25 kpc are twice as likely to be obscured compared to isolated controls (Fig. 8).

(iv) We find a systematic trend of increasing AGN enhancement as a function of decreasing projected separation for these 'obscured AGN'. This begins at a projected separation of ~ 75 kpc and peaks with an AGN enhancement value of 2.08 ± 0.61 for separations of 5–25 kpc (Fig. 9).

Our results for X-ray selected AGN are consistent with previous work of galaxy pairs at cosmic noon that were limited to only using spectroscopic redshifts (Shah et al. 2020). This indicates no evidence for an association between galaxy interactions/mergers and X-ray selected AGN. In contrast, our results hint towards a potentially strong association between galaxy interactions and the most obscured AGN, which population cannot currently be adequately investigated when limited only to samples with spectroscopic redshifts. These results are broadly consistent with an evolutionary scenario of merger-driven obscured black hole growth.

This work is limited by the quality of infrared photometry and limited spectroscopic redshift confirmation for the most obscured AGN. For example, we cannot rule out that we are underestimating the true AGN enhancement due to the challenges of disentangling truly associated pairs and non-associated control pairs with only photometric redshift PDFs. *JWST* will provide higher quality imaging in the near- and mid-IR, improving source photometry, in particular for close galaxy pairs, and placing higher quality constraints on the SEDs of AGNs (e.g. Satyapal et al. 2021; Rigby et al. 2023). Furthermore, future near-IR spectroscopic surveys (e.g. MOONS; Cirasuolo et al. 2020 and PFS; Tamura et al. 2016) will provide increasingly higher spectroscopic completeness of deep fields, permitting pair identifications with much higher certainty than possible with photometric redshifts. Therefore, the coming years will yield ground-breaking insight into the drivers of black hole growth at cosmic noon, in particular for the most obscured systems.

DATA AVAILABILITY

All photometric data, redshift information, and derived quantities (e.g. stellar masses) used in this work are available in public catalogues as described in Section 3. The derived values of this work, presented in Figs 6–9, are tabulated in Appendix D. Code used in this work is available on github (https://github.com/sldough21/pair-convolutions).

REFERENCES

Aird J., Coil A. L., Georgakakis A., Nandra K., Barro G., Pérez-González P. G., 2015, MNRAS, 451, 1892
Alexander D. M., Smail I., Bauer F. E., Chapman S. C., Blain A. W., Brandt W. N., Ivison R. J., 2005, Nature, 434, 738
Alonso-Herrero A. et al., 2006, ApJ, 640, 167
Andonie C. et al., 2022, MNRAS, 517, 2577
Arnouts S., Ilbert O., 2011, Astrophysics Source Code Library, record ascl:1108.009
Ashby M. L. N. et al., 2013, ApJ, 769, 80
Ashby M. L. N. et al., 2015, ApJS, 218, 33
Ashby M. L. N. et al., 2018, ApJS, 237, 39
Assef R. J. et al., 2010, ApJ, 713, 970
Astropy Collaboration, 2013, A&A, 558, A33
Astropy Collaboration, 2018, AJ, 156, 123
Balestra I. et al., 2010, A&A, 512, A12
Barger A. J., Cowie L. L., Wang W. H., 2008, ApJ, 689, 687
Barmby P., Huang J. S., Ashby M. L. N., Eisenhardt P. R. M., Fazio G. G., Willner S. P., Wright E. L., 2008, ApJS, 177, 431
Barnes J. E., Hernquist L. E., 1991, ApJ, 370, L65
Barrera-Ballesteros J. K. et al., 2015, A&A, 579, A45
Barro G. et al., 2019, ApJS, 243, 22
Bennert N., Canalizo G., Jungwiert B., Stockton A., Schweizer F., Peng C. Y., Lacy M., 2008, ApJ, 677, 846
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Bickley R. W., Ellison S. L. Patton D. R. Wilkinson S. 2023, MNRAS, 519, 6149
Blecha L., Snyder G. F., Satyapal S., Ellison S. L., 2018, MNRAS, 478, 3056
Blumenthal K. A., Barnes J. E., 2018, MNRAS, 479, 3952
Bolzonella M., Miralles J. M., Pelló R., 2000, A&A, 363, 476
Bradshaw E. J. et al., 2013, MNRAS, 433, 194
Brammer G. B., van Dokkum P. G., Coppi P., 2008, ApJ, 686, 1503
Bundy K., Fukugita M., Ellis R. S., Targett T. A., Belli S., Kodama T., 2009, ApJ, 697, 1369
Byrne-Mamahit S., Hani M. H., Ellison S. L., Quai S., Patton D. R., 2023, MNRAS, 519, 4966
Capak P. et al., 2004, AJ, 127, 180
Casey C. M. et al., 2017, ApJ, 840, 101
Casteels K. R. V. et al., 2013, MNRAS, 429, 1051
Cattaneo A., Blaizot J., Devriendt J., Guiderdoni B., 2005, MNRAS, 364, 407
Cirasuolo M. et al., 2020, The Messenger, 180, 10
Cisternas M. et al., 2011, ApJ, 726, 57
Civano F. et al., 2012, ApJS, 201, 30
Civano F. et al., 2016, ApJ, 819, 62

Cohen J. G., 2001, AJ, 121, 2895
Cohen J. G., Hogg D. W., Blandford R., Cowie L. L., Hu E., Songaila A., Shopbell P., Richberg K., 2000, ApJ, 538, 29
Comparat J. et al., 2015, A&A, 575, A40
Conselice C. J., Bershady M. A., Dickinson M., Papovich C., 2003, AJ, 126, 1183
Cooper M. C. et al., 2011, ApJS, 193, 14
Cooper M. C. et al., 2012a, MNRAS, 419, 3018
Cooper M. C. et al., 2012b, MNRAS, 425, 2116
Cowie L. L., Barger A. J., Hu E. M., Capak P., Songaila A., 2004, AJ, 127, 3137
Daddi E. et al., 2010, ApJ, 714, L118
Dahlen T. et al., 2013, ApJ, 775, 93
Darg D. W. et al., 2010, MNRAS, 401, 1043
Davies R. I., Müller Sánchez F., Genzel R., Tacconi L. J., Hicks E. K. S., Friedrich S., Sternberg A., 2007, ApJ, 671, 1388
Davies J. J., Pontzen A., Crain R. A., 2023, preprint(arXiv:2301.04145)
Davis M. et al., 2007, ApJ, 660, L1
Dawson S., Stern D., Bunker A. J., Spinrad H., Dey A., 2001, AJ, 122, 598
Dickinson M. et al., 2003, Great Observatories Origins Deep Survey (GOODS) Validation Observations. Spitzer Proposal ID 196
Donley J. L., Rieke G. H., Pérez-González P. G., Rigby J. R., Alonso-Herrero A., 2007, ApJ, 660, 167
Donley J. L., Rieke G. H., Alexander D. M., Egami E., Pérez-González P. G., 2010, ApJ, 719, 1393
Donley J. L. et al., 2012, ApJ, 748, 142
Donley J. L. et al., 2018, ApJ, 853, 63
Draper A. R., Ballantyne D. R., 2010, ApJ, 715, L99
Duncan K. et al., 2019, ApJ, 876, 110
Efron B., 1979, Ann. Stat., 7, 1
Efron B., 1981, Biometrika, 68, 589
Ellison S. L., Patton D. R., Simard L., McConnachie A. W., Baldry I. K., Mendel J. T., 2010, MNRAS, 407, 1514
Ellison S. L., Patton D. R., Mendel J. T., Scudder J. M., 2011, MNRAS, 418, 2043
Ellison S. L., Mendel J. T., Scudder J. M., Patton D. R., Palmer M. J. D., 2013a, MNRAS, 430, 3128
Ellison S. L., Mendel J. T., Patton D. R., Scudder J. M., 2013b, MNRAS, 435, 3627
Ellison S. L., Viswanathan A., Patton D. R., Bottrell C., McConnachie A. W., Gwyn S., Cuillandre J.-C., 2019, MNRAS, 487, 2491
Ellison S. L. et al., 2022, MNRAS, 517, L92
Euclid Collaboration, 2020, A&A, 642, A192
Euclid Collaboration, 2022, A&A, 658, A126
Fabian A. C., 1999, MNRAS, 308, L39
Fontana A., D'Odorico S., Poli F., Giallongo E., Arnouts S., Cristiani S., Moorwood A., Saracco P., 2000, AJ, 120, 2206
Galametz A. et al., 2013, ApJS, 206, 10
Gao F. et al., 2020, A&A, 637, A94
Garilli B. et al., 2021, A&A, 647, A150
Gebhardt K. et al., 2000, ApJ, 539, L13
Giallongo E., D'Odorico S., Fontana A., Cristiani S., Egami E., Hu E., McMahon R. G., 1998, AJ, 115, 2169
Giavalisco M. et al., 2004, ApJ, 600, L93
Glikman E., Simmons B., Mailly M., Schawinski K., Urry C. M., Lacy M., 2015, ApJ, 806, 218
Goulding A. D., Alexander D. M., 2009, MNRAS, 398, 1165
Goulding A. D. et al., 2018, PASJ, 70, S37
Grogin N. A. et al., 2011, ApJS, 197, 35
Gültekin K. et al., 2009, ApJ, 698, 198
Guo Y. et al., 2013, ApJS, 207, 24
Gürkan G., Hardcastle M. J., Jarvis M. J., 2014, MNRAS, 438, 1149
Guyon O., Sanders D. B., Stockton A., 2006, ApJS, 166, 89
Harris C. R. et al., 2020, Nature, 585, 357
Hasinger G. et al., 2018, ApJ, 858, 77
Hewlett T., Villforth C., Wild V., Mendez-Abreu J., Pawlik M., Rowlands K., 2017, MNRAS, 470, 755
Hickox R. C., Alexander D. M., 2018, ARA&A, 56, 625

Hoaglin D. C., Mosteller F., Tukey J. W., 1983, Understanding Robust And Exploratory Data Analysis. Wiley
Hopkins P. F., Hernquist L., 2006, ApJS, 166, 1
Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008, ApJS, 175, 356
Hsieh B.-C., Wang W.-H., Hsieh C.-C., Lin L., Yan H., Lim J., Ho P. T. P., 2012, ApJS, 203, 23
Hsu L.-T. et al., 2014, ApJ, 796, 60
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Huško F., Lacey C. G., Baugh C. M., 2022, MNRAS, 509, 5918
Inami H. et al., 2017, A&A, 608, A2
Kartaltepe J. S. et al., 2007, ApJS, 172, 320
Kartaltepe J. S. et al., 2010, ApJ, 721, 98
Kartaltepe J. S. et al., 2015, ApJ, 806, L35
Kocevski D. D. et al., 2012, ApJ, 744, 148
Kocevski D. D. et al., 2015, ApJ, 814, 104
Kocevski D. D. et al., 2018, ApJS, 236, 48
Kodra D. et al., 2023, ApJ, 942, 36
Koekemoer A. M. et al., 2011, ApJS, 197, 36
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Kormendy J., Kennicutt R. C. Jr, 2004, ARA&A, 42, 603
Koss M. J. et al., 2018, Nature, 563, 214
Kriek M. et al., 2015, ApJS, 218, 15
Kurk J. et al., 2013, A&A, 549, A63
Lackner C. N. et al., 2014, AJ, 148, 137
Lacy M. et al., 2004, ApJS, 154, 166
Lacy M., Petric A. O., Sajina A., Canalizo G., Storrie-Lombardi L. J., Armus L., Fadda D., Marleau F. R., 2007, AJ, 133, 186
Laidler V. G. et al., 2007, PASP, 119, 1325
Laigle C. et al., 2016, ApJS, 224, 24
Lansbury G. B. et al., 2017, ApJ, 846, 20
Lawrence A. et al., 2007, MNRAS, 379, 1599
Le Fèvre O. et al., 2000, MNRAS, 311, 565
Le Fèvre O. et al., 2013, A&A, 559, A14
Lilly S. J. et al., 2007, ApJS, 172, 70
López-Sanjuan C. et al., 2015, A&A, 576, A53
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Magorrian J. et al., 1998, AJ, 115, 2285
Mantha K. B. et al., 2018, MNRAS, 475, 1549
Marchesi S. et al., 2016, ApJ, 817, 34
Marian V. et al., 2019, ApJ, 882, 141
Masters D. C. et al., 2019, ApJ, 877, 81
Mateos S. et al., 2012, MNRAS, 426, 3271
Matthews D. J., Newman J. A., Coil A. L., Cooper M. C., Gwyn S. D. J., 2013, ApJS, 204, 21
McAlpine S., Harrison C. M., Rosario D. J., Alexander D. M., Ellison S. L., Johansson P. H., Patton D. R., 2020, MNRAS, 494, 5713
McKinney W., 2010, in van der Walt S., Millman J., eds, Proceedings of the 9th Python in Science Conference SciPy Organizers, p. 56, https://conference.scipy.org/
Mechtley M. et al., 2016, ApJ, 830, 156
Merloni A. et al., 2019, The Messenger, 175, 42
Mihos J. C., Hernquist L., 1996, ApJ, 464, 641
Mobasher B. et al., 2015, ApJ, 808, 101
Moran E. C., Lehnert M. D., Helfand D. J., 1999, ApJ, 526, 649
Moreno J. et al., 2021, MNRAS, 503, 3113
Mundy C. J., Conselice C. J., Duncan K. J., Almaini O., Häußler B., Hartley W. G., 2017, MNRAS, 470, 3507
Nanayakkara T. et al., 2016, ApJ, 828, 21
Nandra K. et al., 2015, ApJS, 220, 10
Nayyeri H. et al., 2017, ApJS, 228, 7
Onodera M. et al., 2012, ApJ, 755, 26
Pandas Development Team, 2020, in pandas-dev/pandas: Pandas. Zenodo
Park S. Q. et al., 2010, ApJ, 717, 1181
Patton D. R., Torrey P., Ellison S. L., Mendel J. T., Scudder J. M., 2013, MNRAS, 433, L59
Perez J., Tissera P., Blaizot J., 2009, MNRAS, 397, 748
Pierce J. C. S. et al., 2022, MNRAS, 510, 1163

Pierce J. C. S. et al., 2023, MNRAS, 522, 1736
Quai S., Byrne-Mamahit S., Ellison S. L., Patton D. R., Hani M. H., 2023, MNRAS, 519, 2119
Ramos Almeida C., Tadhunter C. N., Inskip K. J., Morganti R., Holt J., Dicken D., 2011, MNRAS, 410, 1550
Ravikumar C. D. et al., 2007, A&A, 465, 1099
Reddy N. A., Steidel C. C., Erb D. K., Shapley A. E., Pettini M., 2006, ApJ, 653, 1004
Ricci C. et al., 2017, MNRAS, 468, 1273
Ricci C. et al., 2021, MNRAS, 506, 5935
Rigby J. et al., 2023, PASP, 135, 048001
Rosario D. J. et al., 2015, A&A, 573, A85
Sanders D. B., Soifer B. T., Elias J. H., Madore B. F., Matthews K., Neugebauer G., Scoville N. Z., 1988a, ApJ, 325, 74
Sanders D. B., Soifer B. T., Elias J. H., Neugebauer G., Matthews K., 1988b, ApJ, 328, L35
Sanders D. B. et al., 2007, ApJS, 172, 86
Santini P. et al., 2015, ApJ, 801, 97
Satyapal S., Ellison S. L., McAlpine W., Hickox R. C., Patton D. R., Mendel J. T., 2014, MNRAS, 441, 1297
Satyapal S., Kamal L., Cann J. M., Secrest N. J., Abel N. P., 2021, ApJ, 906, 35
Schawinski K., Virani S., Simmons B., Urry C. M., Treister E., Kaviraj S., Kushkuley B., 2009, ApJ, 692, L19
Schawinski K., Simmons B. D., Urry C. M., Treister E., Glikman E., 2012, MNRAS, 425, L61
Scodeggio M. et al., 2018, A&A, 609, A84
Scoville N. et al., 2007, ApJS, 172, 1
Scoville N. et al., 2014, ApJ, 783, 84
Secrest N. J., Ellison S. L., Satyapal S., Blecha L., 2020, MNRAS, 499, 2380
Shah E. A. et al., 2020, ApJ, 904, 107
Silva A., Marchesini D., Silverman J. D., Martis N., Iono D., Espada D., Skelton R., 2021, ApJ, 909, 124
Silverman J. D. et al., 2010, ApJS, 191, 124
Silverman J. D. et al., 2011, ApJ, 743, 2
Silverman J. D. et al., 2015, ApJS, 220, 12
Snyder G. F., Hayward C. C., Sajina A., Jonsson P., Cox T. J., Hernquist L., Hopkins P. F., Yan L., 2013, ApJ, 768, 168
Stanford S. A. et al., 2021, ApJS, 256, 9
Stefanon M. et al., 2017, ApJS, 229, 32
Steffen J. L. et al., 2023, ApJ, 942, 107
Steinhardt C. L. et al., 2014, ApJ, 791, L25
Stern D. et al., 2005, ApJ, 631, 163
Straatman C. M. S. et al., 2018, ApJS, 239, 27
Sutherland W., Saunders W., 1992, MNRAS, 259, 413
Szalay A. S., Connolly A. J., Szokoly G. P., 1999, AJ, 117, 68
Tacconi L. J. et al., 2010, Nature, 463, 781
Tamura N. et al., 2016, in Evans C. J., Simard L., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI. SPIE, Bellingham, p. 99081M
Treister E. et al., 2009, ApJ, 706, 535
Treister E., Schawinski K., Urry C. M., Simmons B. D., 2012, ApJ, 758, L39
Trump J. R. et al., 2007, ApJS, 172, 383
Trump J. R. et al., 2009, ApJ, 696, 1195
Urrutia T., Lacy M., Becker R. H., 2008, ApJ, 674, 80
Urrutia T. et al., 2019, A&A, 624, A141
Vanzella E. et al., 2008, A&A, 478, 83
van der Wel A. et al., 2021, ApJS, 256, 44
Veilleux S. et al., 2009, ApJS, 182, 628
Villforth C. et al., 2014, MNRAS, 439, 3342
Villforth C. et al., 2017, MNRAS, 466, 812
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Weaver J. R. et al., 2022, ApJS, 258, 11
Weston M. E., McIntosh D. H., Brodwin M., Mann J., Cooper A., McConnell A., Nielsen J. L., 2017, MNRAS, 464, 3882
Wiklind T., Dickinson M., Ferguson H. C., Giavalisco M., Mobasher B., Grogin N. A., Panagia N., 2008, ApJ, 676, 781
Wild V., Heckman T., Charlot S., 2010, MNRAS, 405, 933

Wirth G. D. et al., 2004, AJ, 127, 3121
Wirth G. D. et al., 2015, AJ, 150, 153
Xue Y. Q. et al., 2011, ApJS, 195, 10
Xue Y. Q., Luo B., Brandt W. N., Alexander D. M., Bauer F. E., Lehmer B. D., Yang G., 2016, ApJS, 224, 15
Yamada T. et al., 2005, ApJ, 634, 861

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Appendix D. Tabulated Results.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: SPECTROSCOPIC REDSHIFTS AND COMPARISON TO PHOTOMETRIC REDSHIFTS

Here, we provide the references that were used to compile the spectroscopic redshifts described in Section 2.2, followed by a comparison between the spectroscopic redshift and photometric redshifts described in Section 2.1.2. Table A1 summarizes the total number of galaxies (parent sample and AGN), the fraction of these with spectroscopic redshifts and the accuracy of the corresponding photometric redshift PDFs (σ_{NMAD}).

For the CANDELS GOODS-N field, we utilized the spectroscopic redshifts that were obtain on: (1) Keck I, with the Low Resolution Imaging Spectrometer (LRIS; Cohen et al. 2000; Cohen 2001; Dawson et al. 2001; Cowie et al. 2004; Reddy et al. 2006) and the Multi-Object Spectrometer For Infra-Red Exploration (MOSFIRE; Kriek et al. 2015; Wirth et al. 2015); (2) Keck II, with the DEep Imaging Multi-Object Spectrograph (DEIMOS; Cowie et al. 2004; Wirth et al. 2004; Barger, Cowie & Wang 2008; Cooper et al. 2011); and (3) on Subaru with the Multi-Object Infrared Camera and Spectrograph (MOIRCS; Barger et al. 2008).

For the CANDELS GOODS-S field, we used spectroscopic redshifts obtained on: (1) Keck I using MOSFIRE (Kriek et al. 2015); (2) Keck II with DEIMOS (Silverman et al. 2010; Cooper et al. 2012b); and (3) the Very Large telescope (VLT) with the VIsible MultiObject Spectrograph (VLT/VIMOS; Ravikumar et al. 2007; Balestra et al. 2010; Le Fèvre et al. 2013; Garilli et al. 2021), the FOcal Reducer and low dispersion Spectrograph 1 and 2 (FORS1/2; Vanzella et al. 2008; Kurk et al. 2013), and the Multi Unit Spectroscopic Explorer (MUSE; Inami et al. 2017; Urrutia et al. 2019).

The spectroscopic redshifts used in this study for the CANDELS EGS field were obtained on Keck I with LRIS and MOSFIRE (Kriek

Table A1. The number of galaxies (N) in the parent sample and the number of AGN split into CANDELS and COSMOS. f_{spec} is the percentage of sources in each category with a spectroscopic redshift and σ_{NMAD} is the normalized median absolute deviation between z_{spec} and z_{phot} (see equation A1).

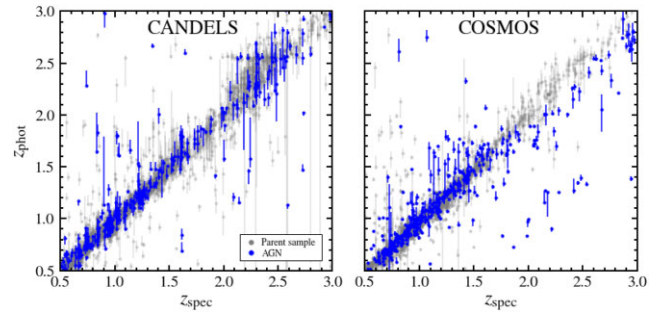|	N	f_{spec} (per cent)	σ_{NMAD}
Parent Sample (CANDELS)	22 093	19.9	0.022
AGN (CANDELS)	1036	34.1	0.033
Parent Sample (COSMOS)	112 436	5.3	0.015
AGN (COSMOS)	2216	24.3	0.033

Figure A1. Spectroscopic redshifts (z_{spec}) versus corresponding photometric redshift PDF peak estimate (z_{phot}) for those done by the CANDELS (*left*; Kodra et al. 2023) and COSMOS teams (*right*; Weaver et al. 2022). The grey and blue points correspond to the parent sample and all AGNs, respectively. Error bars correspond to the upper and lower 68 per cent confidence intervals of the photometric redshift PDFs.

et al. 2015; Masters et al. 2019) as well as on Keck II with DEIMOS (Cooper et al. 2012a; Matthews et al. 2013; Masters et al. 2019).

For the CANDELS UDS field, the data were taken on: (1) Keck I with LRIS and MOSFIRE (Kriek et al. 2015; Masters et al. 2019); (2) Keck II with DEIMOS (Masters et al. 2019); (3) VLT with FORS2 (Bradshaw et al. 2013) and VIMOS (Bradshaw et al. 2013; Scodeggio et al. 2018; Garilli et al. 2021); and (4) Subaru with the Faint Object Camera And Spectrograph (FOCUS; Yamada et al. 2005).

Finally, for all of COSMOS, spectroscopic observations were taken on: (1) Keck I with LRIS and MOSFIRE (Kriek et al. 2015; Nanayakkara et al. 2016; Casey et al. 2017; Masters et al. 2019; Stanford et al. 2021); (2) Keck II with DEIMOS (Capak et al. 2004; Kartaltepe et al. 2010; Hasinger et al. 2018; Masters et al. 2019; Stanford et al. 2021); (3) VLT with FORS2 (Comparat et al. 2015), VIMOS (Lilly et al. 2007; Straatman et al. 2018; van der Wel et al. 2021), and the *K*-band Multi-Object Spectrograph (KMOS; Euclid Collaboration 2020); (4) Subaru with the Fiber Multi-Object Spectrograph (FMOS; Kartaltepe et al. 2015; Silverman et al. 2015) and MOIRCS (Onodera et al. 2012); and (5) the Magellan Telescope with the Inamori-Magellan Areal Camera and Spectrograph (IMACS; Trump et al. 2007, 2009).

We are primarily interested in the fraction of AGN in paired galaxies identified via their photometric redshift PDFs. Therefore, it is important to have an assessment of the quality of the photometric redshifts for AGN, compared to the wider galaxy population. In Fig. A1, we compare spectroscopic redshifts (z_{spec}) to the peak of the PDF estimate of the photometric redshift (z_{phot}). We reiterate that these z_{phot} values are not used in the analyses to identify true galaxy pairs, but instead we use the full PDFs to encapsulate the full uncertainties. Therefore, the comparison of z_{spec} and z_{phot} values can only be considered indicative of the quality of the PDFs and thus hides the full treatment of the uncertainties.

Fig. A1 shows that for the majority of the parent sample (grey data points) and the AGN (blue data points) the two redshifts are in good agreement. One way to quantify this is with the normalized median absolute deviation (NMAD; Hoaglin, Mosteller & Tukey 1983) comparing the spectroscopic and photometric redshifts, which is a common approach for photometric redshift validation (e.g. Guo et al. 2013; Nayyeri et al. 2017; Stefanon et al. 2017; Barro et al. 2019; Weaver et al. 2022). This is defined as

$$\sigma_{NMAD} = 1.48 \times \text{median} \left(\frac{|\Delta z|}{1 + z_{spec}} \right), \tag{A1}$$

where Δz is the difference between the peak photometric redshift PDF estimate and the corresponding spectroscopic redshift.

The CANDELS and COSMOS PDFs are exceptionally accurate for both the parent sample ($\sigma_{\mathrm{NMAD}} \sim 0.02$) and the AGN ($\sigma_{\mathrm{NMAD}} \sim 0.03$), albeit slightly poorer for the AGNs. The AGNs also have higher incidence of spectroscopic redshifts compared to the parent sample (34 per cent compared to 20 per cent for CANDELS and 24 per cent compared to 5 per cent for COSMOS). This will further improve the quality of TP probabilities presented in Section 3.3. Overall, we feel confident the photometric PDFs are of sufficiently high quality for our large statistical study, and future observations with improved infrared photometry and larger spectroscopic samples will provide even tighter constraints (see discussion in Section 6).

APPENDIX B: PROBABILISTIC APPROACH TO SELECTING CONTROL GALAXIES

Here, we provide a comprehensive overview of the matching algorithm used to select control galaxy pairs in Section 3.4.

As we calculate the weighted AGN fraction for galaxy pairs (Section 3.5), we assemble isolated galaxy 'pairs' as our control sample. In other words, we match physically associated pairs, with weights $\mathcal{P}_{\mathrm{pair}}$, to physically unassociated pairs. By construction, we require unassociated pairs to have $\mathcal{P}_{\mathrm{pair}} = 0$, which requires either $\mathcal{P}_{\Delta V} = 0$ or $\mathcal{P}_r^{0\text{-}100\,\mathrm{kpc}} = 0$ (equation 6). In Section 3.1, we described how $\mathcal{P}_{\Delta V}$ is derived from the $P(z)$ of paired galaxies independent of projected separation. Therefore, we select unassociated pairs on the basis of $\mathcal{P}_r^{0\text{-}100\,\mathrm{kpc}} = 0$, and we interpret their $\mathcal{P}_{\Delta V}$ as weights both statistically comparable to those of the TP sample and reflective of the individual unassociated galaxy redshift uncertainties.

For each TP, we select three unassociated CPs matched in redshift, mass, environmental density, $P(z)$ width[3] and relative line-of-sight probability from a pool of galaxies highly likely to be physically isolated. We begin by assuming that every galaxy in our parameter space has some probability of being isolated (i.e. no companion within $r_p < 100$ kpc or $|\Delta V| < 1000$ km s^{-1}), $\mathcal{P}_{\mathrm{iso}}$, defined as

$$\mathcal{P}_{\mathrm{iso}} = \prod_{i=1}^{N} (1 - \mathcal{P}_{\mathrm{pair},i}^{5\text{-}100\,\mathrm{kpc}}), \tag{B1}$$

where N is the total number of its projected pairs and $\mathcal{P}_{\mathrm{pair},i}^{5\text{-}100\,\mathrm{kpc}}$ is the TP probability (equation 6) of pair i evaluated over the full separation range. We create a pool of all galaxies with $\mathcal{P}_{\mathrm{iso}} > 90$ per cent, from which we will select controls.

Following Ellison et al. (2013b), we compute the local environmental density for each galaxy, Σ, defined as

$$\Sigma = \frac{n}{\pi r_n^2}, \tag{B2}$$

where r_n is the projected separation to the nth nearest companion within $|\Delta V| < 1000$ km s^{-1}. We set $r_n = 1$ Mpc and apply our probabilistic methodology to measuring n as the sum of all TP probabilities of all projected companions within 1 Mpc:

$$n = \sum_{i=1}^{N} \mathcal{P}_{\mathrm{pair},i}^{0\text{-}1\,\mathrm{Mpc}}. \tag{B3}$$

In finding the best matches, we would ideally calculate the difference in mass, redshift, environment, $P(z)$ width (for both

---

[3]For matching purposes, sources with secure spectroscopic redshifts are assigned a $P(z)$ width equal to 0.01.
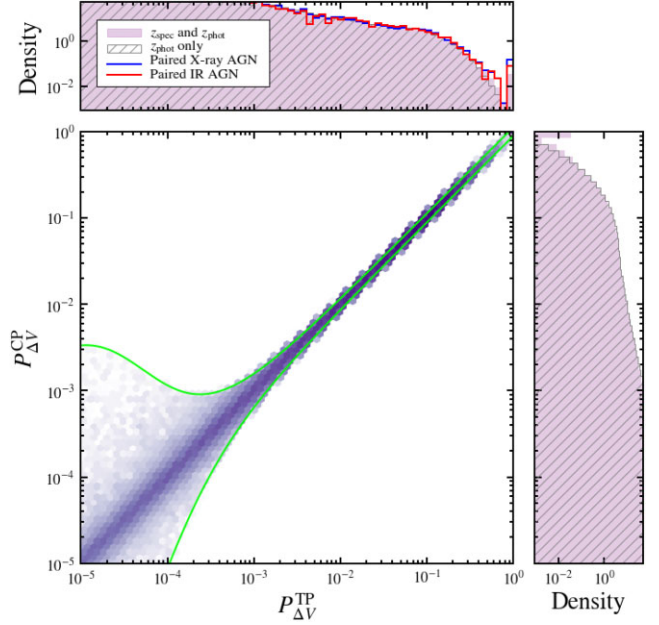
Figure B1. Distribution of the relative line-of-sight probabilities of the CPs, $\mathcal{P}_{\Delta V}^{\mathrm{CP}}$, matched to that of the TPs, $\mathcal{P}_{\Delta V}^{\mathrm{TP}}$. The green lines correspond to the matching function extrema (equation B4) projected onto the $\mathcal{P}_{\Delta V}^{\mathrm{TP}} - \mathcal{P}_{\Delta V}^{\mathrm{CP}}$ plane. TPs with high $\mathcal{P}_{\Delta V}^{\mathrm{TP}}$ are tightly matched by $\mathcal{P}_{\Delta V}^{\mathrm{CP}}$ to ensure equal weighting when calculating, then comparing, weighted AGN fractions (equation 7). Distributions of $\mathcal{P}_{\Delta V}^{\mathrm{TP}}$ and $\mathcal{P}_{\Delta V}^{\mathrm{CP}}$ are shown with (solid purple histogram) and without (grey striped histogram) the inclusion of spectroscopic redshifts, adjacent to the corresponding axes. Normalized distributions of $\mathcal{P}_{\Delta V}^{\mathrm{TP}}$ for pairs featuring X-ray and IR AGNs are given by blue and red lines, respectively.

galaxies) as well as $\mathcal{P}_{\Delta V}$ between the TP and all candidate CPs. This approach would involve first calculating $\mathcal{P}_{\Delta V}$ then sorting for over 6 billion unique galaxy pairs. Rather, we construct an iterative search algorithm to find sufficient matches in a timely manner. For both galaxies in a TP, we assemble initial samples of control candidates with $|\Delta z| < 0.05$, $|\Delta \log(M_*)| < 0.05$, $|\Delta \Sigma| < 2$, and $|\Delta \log(P(z)\,\mathrm{width})| < 0.05$. We generate every combination of galaxies from the two samples separated by more than 20 arcsec to ensure $\mathcal{P}_r^{0\text{-}100\,\mathrm{kpc}} = 0$, and we compute $\mathcal{P}_{\Delta V}$ for each pairing. In order for a pair of relative line-of-sight probability, $\mathcal{P}_{\Delta V}^{\mathrm{CP}}$, to be selected as a match to a TP of relative line-of-sight probability, $\mathcal{P}_{\Delta V}^{\mathrm{TP}}$, it must meet the criteria of our matching function:

$$|\log\left(\mathcal{P}_{\Delta V}^{\mathrm{TP}}\right) - \log\left(\mathcal{P}_{\Delta V}^{\mathrm{CP}}\right)| < \frac{1.54 \times 10^{-4}}{\mathcal{P}_{\Delta V}^{\mathrm{TP}} + 5.24 \times 10^{-5}} + 0.05. \tag{B4}$$

We devised the matching function such that high probability TPs (i.e. the ones that receive the highest weights in the weighted AGN fraction calculation), are strictly matched in $\mathcal{P}_{\Delta V}$ while the matching to low probability TPs (i.e. those whose weights will contribute negligibly to the weighted AGN fraction), is less stringent. Matched $\mathcal{P}_{\Delta V}$ are shown in the bounds of the matching function in Fig. B1.

If more than 3 candidate CPs satisfy the matching function, then we select the 3 best-matched CPs by minimizing the difference:

$$
\begin{aligned}
\mathrm{dif.} = {} & (\Delta z_1)^2 + (\Delta \log(M_{*,1}))^2 + \frac{(\Delta \Sigma_1)^2}{10} + (\Delta \log(P_1(z)\,\mathrm{width}))^2 \\
& + (\Delta z_2)^2 + (\Delta \log(M_{*,2}))^2 + \frac{(\Delta \Sigma_2)^2}{10} + (\Delta \log(P_2(z)\,\mathrm{width}))^2 \\
& + 10(\Delta \log(\mathcal{P}_{\Delta V})^2, \tag{B5}
\end{aligned}
$$

where the subscripts 1 and 2 indicate the difference in matched parameters to galaxy 1 and 2 in a TP. We amplify the difference in $\mathcal{P}_{\Delta V}$ so that the CPs receive the closest weight possible to their matched TP. Additionally, we down-weight the environmental components to place their difference in roughly the same order of magnitude as that of the other matched parameters (i.e. stellar masses range from $\log(M_*/M_\odot) \sim 9.4$–12 while environmental density ranges from $\Sigma \sim 0$–25 Mpc^{-1}; see Fig. 5). If less than 3 CPs satisfy the matching function, then we iteratively broaden the initial z, $\log(M_*)$, Σ, and $\log(P(z)$ width) search thresholds by 0.03, 0.03, 1, and 0.05, respectively, until 3 do. CPs are selected with replacement for different TPs, and we require 6 unique galaxies to comprise the 3 CPs per TP to mitigate duplicate effects. Normalized distributions of the matched parameters with and without weights are shown in Fig. 5, which reveals the matching process was successful.

APPENDIX C: SIMULATED RESULTS

We test the robustness of our probabilistic pair methodology by applying it to a mock data-set with the goal of recreating the observed weighted AGN fractions (Fig. 9) of our obscured AGN sample. We generate 250 000 galaxies in an $\sim$2.4 deg^2 area, from which we draw the RA and Dec. of each galaxy randomly. True redshifts, $z_{\rm true}$, are drawn from a random uniform distribution between 0.5 and 3. To simulate increasing weighted AGN fractions at decreasing separation comparable to that of the observed obscured AGN enhancement, we then randomly assign AGNs to galaxy pairs ($\Delta V < 1000$ km s^{-1}) at 0.02, 0.01, and 0.005 probabilities if their projected separation is within 25 kpc, 50 kpc, or not (i.e. isolated), respectively.

We generate photometric redshift PDFs for each galaxy from a Gaussian distribution of mean $\mu_{\rm sim}$ and standard deviation $\sigma_{\rm sim}$. We fit Gaussian distributions of σ equal to $\sim$0.017 and $\sim$0.068 (both $\mu \sim$ 0) to the observed Δz (i.e. $z_{\rm spec} - z_{\rm phot}$) and $\sigma_{P(z)}$ (i.e. half the upper minus lower PDF 68 per cent confidence intervals), respectively, from which we draw Δz_i shifts and $\sigma_{P(z),i}$ such that $\mu_{{\rm sim},i} = z_{{\rm true},i} + \Delta z_i$ and $\sigma_{{\rm sim},i} = \sigma_{P(z),i}$ for galaxy i.

Finally, we calculate the weighted AGN fractions from the simulated PDFs, which we compare to the true statistics of the mock sample (i.e. from $z_{\rm true}$) as a function of projected separation in Fig. C1. Determining the AGN enhancement via PDF convolutions indeed recovers an increase in the weighted AGN fraction from $\sim$0.5 per cent at projected separations less than 50 kpc to $\sim$10 per cent at projected separations less than 25 kpc, just as in the obscured AGN sample (Fig. 9).

However, based on this simulated experiment, the convolution approach underestimates the true AGN fraction of close galaxy pairs (within 25 kpc) in the mock data set by $\sim$40 per cent. Given that
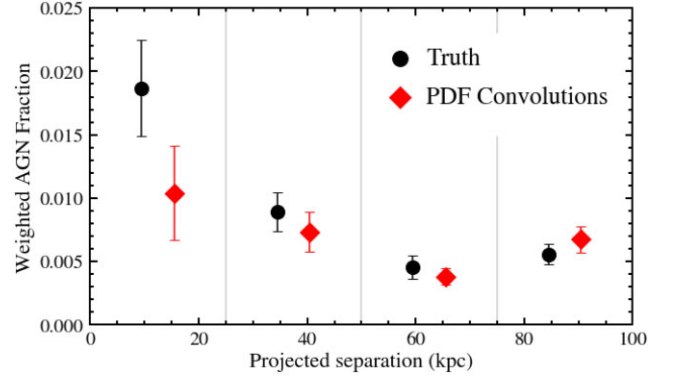

Figure C1. Weighted AGN fraction of simulated galaxy pairs binned by projected separation. The black circles correspond to the weighted AGN fraction ($\mathcal{P}_{\rm pair}$ equal to 0 or 1) calculated from the true redshifts (i.e. $z_{\rm true}$). The red diamonds correspond to the weighted AGN fraction determined using simulated PDFs and the pair convolution method detailed in this work. Standard errors for both groups are calculated from a common bootstrap technique (Efron 1979, 1981).

the X-ray sample has a much higher spectroscopic redshift coverage (i.e. roughly one third have spectrosopic redshifts), and therefore is less affected by uncertain PDFs, and we observe no enhancement at all for this sample (Fig. 6), the observed lack of a trend with projected separation is expected to be robust. For the obscured AGN sample (Fig. 9), our simulations indicate that our finding may be moderately *underestimating* the true AGN enhancements (albeit this is captured somewhat with that large uncertainties) as TP and CP samples inevitably become less clean with a higher fraction of photometric redshifts.

APPENDIX D: TABULATED RESULTS

This tabulated data plotted in Section 4 is provided in the supplementary material.

This paper has been typeset from a T$_{\rm E}$X/L^{A}T$_{\rm E}$X file prepared by the author.