

# Variability-selected AGN in dwarf galaxies: the incidence of AGN in dwarf and massive galaxies is similar

S. Kaviraj<sup>1</sup>,<sup>1</sup>★ D. De Cicco,<sup>2</sup> I. Lazar<sup>1</sup>,<sup>1</sup> B. Bichang’a,<sup>1</sup> A. E. Watkins,<sup>1</sup> G. Martin<sup>3</sup> and S. Koudmani<sup>1,4</sup>

<sup>1</sup>Centre for Astrophysics Research, Department of Physics, Astronomy and Mathematics, University of Hertfordshire, Hatfield AL10 9AB, UK

<sup>2</sup>Department of Physics, University of Napoli ‘Federico II’, Via Cinthia 9, I-80126 Napoli, Italy

<sup>3</sup>School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK

<sup>4</sup>St Catharine’s College, University of Cambridge, Trumpington Street, Cambridge CB2 1RL, UK

Accepted 2025 October 30. Received 2025 October 14; in original form 2025 February 11

## ABSTRACT

We use the VST-COSMOS survey to identify, via their optical broad-band variability, 30 active galactic nuclei (AGN) in nearby ( $z < 0.4$ ) dwarf ( $10^8 M_\odot < M_\star < 10^{10} M_\odot$ ) galaxies. VST-COSMOS offers a  $1 \text{ deg}^2$  survey footprint, a single visit depth of 24.6 mag and 68  $r$ -band visits spanning an 11-yr temporal baseline. Compared to a control sample matched in stellar mass and redshift, the dwarf AGN population shows an elevated fraction of early-type galaxies but a similar fraction of interacting objects, suggesting that interactions do not play a significant role in triggering these AGN. Dwarf AGN hosts do not show strong differences in their distances to nodes, filaments, and massive galaxies compared to the controls, which indicates that AGN triggering, at least in this sample, is not strongly correlated with environment. Finally, by combining the true number of galaxies, the detectability of AGN and the measured numbers of variable sources in dwarf and massive ( $M_\star > 10^{10} M_\odot$ ) galaxies, we estimate the relative frequency of AGN in these two mass regimes. Our results suggest that the incidence of AGN in dwarfs and massive galaxies is similar (within less than a factor of 2 of each other), with some evidence that the AGN fraction increases with stellar mass in the dwarf population.

**Key words:** galaxies: active – galaxies: dwarf – galaxies: evolution – galaxies: formation – galaxies: interactions.

## 1 INTRODUCTION

Given their dominance of the galaxy number density (e.g. A. H. Wright et al. 2017; G. Martin et al. 2019), dwarf galaxies are vital for a complete understanding of galaxy evolution. Dwarfs have been studied in detail largely in our immediate vicinity, e.g. in the Local Group or around nearby massive galaxies (e.g. E. Tolstoy, V. Hill & M. Tosi 2009; P.-A. Duc et al. 2015; M. Geha et al. 2017; A. Venhola et al. 2018; I. Trujillo et al. 2021), with the advent of *James Webb Space Telescope* (JWST) (J. P. Gardner et al. 2006) gradually opening up this domain in the high-redshift Universe via its near-infrared capabilities (e.g. R. Endsley et al. 2024; W. M. Baker et al. 2025; C. Witten et al. 2025).

In the context of galaxy evolution, active galactic nuclei (AGN) are thought to significantly influence the formation of massive galaxies (e.g. A. C. Fabian 2012; R. S. Beckmann et al. 2017). Virtually all massive galaxies are thought to host supermassive black holes (BHs) at their centres (e.g. J. Kormendy & D. Richstone 1995; D. Richstone et al. 1998). The negative feedback imparted by these BHs, when they are accreting and active, is commonly employed to regulate star formation activity in theoretical models, in order to bring predicted galaxy properties (e.g. stellar masses, morphologies, and colours) in

line with observational data (e.g. D. J. Croton et al. 2006; S. Kaviraj et al. 2017). However, while BHs and AGN form an integral part of the currently accepted picture of massive-galaxy evolution, many unexplored questions remain about BH–galaxy co-evolution in the dwarf regime. For example, what are the AGN and BH occupation fractions in dwarf galaxies? What is the relative frequency of AGN in dwarfs compared to that in their massive counterparts? Do dwarf AGN hosts show correlations with galaxy morphology and/or local environment, as is observed in the massive-galaxy regime? Is there evidence that AGN feedback regulates star formation activity in dwarfs?

From a theoretical standpoint, BH growth in dwarfs in cosmological simulations appears to be heavily stunted. This is driven both by supernova feedback starving the BH of fuel (e.g. Y. Dubois et al. 2015, 2021; D. Anglés-Alcázar et al. 2017; M. Habouzit, M. Volonteri & Y. Dubois 2017; M. Trebitsch et al. 2018) and because, if not tethered to the barycentre, the shallow dwarf potential wells allow the BHs in some simulations to wander outside the central regions of the host galaxies. This results in the BHs spending most of their lifetimes in regions of low-gas density, often at large distances from the central regions of the host galaxies (e.g. J. M. Bellovary et al. 2021; R. S. Sharma et al. 2022). A consequence of this is that the dwarf AGN fractions in such simulations are very low (if not close to zero). Nevertheless, high-resolution idealized simulations are increasingly demonstrating the likelihood of AGN activity in

\* E-mail: [s.kaviraj@herts.ac.uk](mailto:s.kaviraj@herts.ac.uk)

simulated dwarf galaxies and the potential for AGN feedback to shape their evolution (P. Barai & E. M. de Gouveia Dal Pino 2019; S. Koudmani et al. 2019, 2024; R. S. Sharma et al. 2023; E. Arjona-Galvez, A. Di Cintio & R. J. J. Grand 2024), in a similar fashion to what is seen in massive galaxies.

The behaviour seen in current simulations that employ cosmological volumes appears increasingly in tension with observational work. A growing literature is identifying AGN in nearby dwarf galaxies via an array of methods, e.g. using optical emission-line ratios (e.g. J. E. Greene & L. C. Ho 2007; A. E. Reines, J. E. Greene & M. Geha 2013; E. C. Moran et al. 2014; C. M. Dickey et al. 2019; R. Pucha et al. 2025; M. Mezcua & H. Domínguez Sánchez 2024), broad-band variability (e.g. V. F. Baldassare, M. Geha & J. Greene 2020a; C. J. Burke et al. 2022, 2024), infrared photometry (e.g. T. H. Jarrett et al. 2011; F. R. Marleau et al. 2017; S. Satyapal, N. P. Abel & N. J. Secrest 2018; S. Kaviraj, G. Martin & J. Silk 2019), spectral energy distribution (SED) fitting (e.g. F. Zou et al. 2023; B. Bichang’a et al. 2024), X-ray emission (K. Pardo et al. 2016; C. T. J. Chen et al. 2017; M. Mezcua et al. 2018; K. L. Birchall, M. G. Watson & J. Aird 2020; A. Sacchi et al. 2024), and excess radio emission that cannot be accounted for by star formation alone (e.g. K. Nyland et al. 2017; M. Mezcua, H. Suh & F. Civano 2019; F. Davis et al. 2022).

Several of these studies report appreciable fractions of AGN in the dwarf regime. For example, M. Mezcua & H. Domínguez Sánchez (2024) use integral field spectroscopy, from the MaNGA survey (K. Bundy et al. 2015), of dwarf galaxies drawn from the Sloan Digital Sky Survey (SDSS) to show that at least 20 per cent (and up to  $\sim 54$  per cent) of these systems show signs of AGN activity in their optical emission line ratios. Notably off-nuclear AGN, which are difficult to detect using single fibre spectroscopy, significantly contribute to the statistics in this study. C. M. Dickey et al. (2019) use a similar methodology, implemented via long-slit spectroscopy, to show that  $\sim 80$  per cent of quiescent field dwarfs in the local Universe show AGN signatures in their central regions. SED fitting using deep ultraviolet (UV) to mid-infrared broad-band photometry indicates that around a third of dwarfs show signs of AGN activity (B. Bichang’a et al. 2024), while F. Davis et al. (2022), who combine deep radio and optical data to select radiatively inefficient AGN, conclude that AGN triggering in dwarfs is likely to be stochastic and a common phenomenon.

The existence of AGN in nearby dwarfs appears consistent with recent observations using the *JWST* which reveal the widespread presence of AGN in low-mass galaxies at high redshift (e.g. J. Scholtz et al. 2025), and the conclusions of some studies that AGN activity appears common in such systems (e.g. I. Juodžbalis et al. 2023). While nearby dwarfs are not descendants of galaxies that have similar masses at high redshift, the *JWST* results demonstrate that AGN activity can routinely take place within the gravitational potential wells that host low-mass galaxies (see M. Mezcua et al. 2024 for a discussion of the link between AGN discovered in the high redshift Universe via *JWST* and those that reside in nearby dwarf galaxies). Taken together, these empirical results spanning a range of redshifts raise the possibility that AGN activity (and therefore BH growth) may have existed in the dwarf regime over cosmic time. This aligns with the fact that nearby dwarfs appear to lie on an extrapolation of the  $M-\sigma$  relation defined by massive galaxies (e.g. Z. Schutte, A. E. Reines & J. E. Greene 2019; T. A. Davis et al. 2020; V. F. Baldassare et al. 2020b), a correlation that is expected to arise naturally if AGN activity regulates the growth of its host galaxy (e.g. J. Silk & M. J. Rees 1998; A. King & R. Nealon 2021).

As we explore in detail in Section 4, while many AGN have indeed been discovered in dwarfs, accurate measurements of the

AGN fraction remain challenging because AGN detectability in the dwarf regime is often low. For example, typical dwarfs outside the local neighbourhood are too faint to be detectable in past wide-area surveys like the SDSS because they are shallow (S. Kaviraj et al. 2025). The dwarfs that are detected in shallow surveys outside the very local Universe are biased towards star-forming galaxies, in which young stars boost the galaxy luminosities above the survey detection thresholds of shallow data sets like the SDSS. Unbiased statistical studies of dwarfs require surveys that are both deep and wide. The depth is required for detecting typical dwarfs outside the local neighbourhood and large areas are needed to create statistical samples of dwarfs, particularly in low-density environments. While this will become routinely possible using new and forthcoming surveys like the Legacy Survey of Space and Time (LSST; Ž. Ivezić et al. 2019) and *Euclid* (R. Laureijs et al. 2011), some precursor data sets already exist (such as in the COSMOS field) where such studies are possible, albeit with smaller numbers of galaxies.

The use of shallow surveys also complicates the detection of AGN activity in dwarf galaxies, since it can be swamped by the high levels of star formation that exist in the dwarfs that are present in these data sets (e.g. M. Mezcua & H. Domínguez Sánchez 2024). In a similar vein, the depth of any ancillary data (e.g. radio fluxes) that is being used to identify AGN adds an extra layer of complexity to the detection process and can further reduce the detectability of AGN in dwarf galaxies. A consequence of these two issues, which is particularly important in the dwarf regime where both the host galaxy and the AGN can be relatively faint, is that an accurate measurement of the AGN fraction demands knowledge of both host and AGN detectability. Measuring detectability, however, is challenging, as a result of which this quantity is typically ignored in some studies. We explore this issue further, in the context of the data used in this study, to derive an estimate of the relative frequency of AGN in the dwarf and massive-galaxy regimes, in Section 4 below.

Given the biases noted above, techniques such as variability (e.g. E. Elmer et al. 2020; Y. Kimura et al. 2020; C. J. Burke et al. 2024), that can identify AGN in dwarfs using (deep) broad-band photometry alone, are desirable. Deep photometry makes it possible to detect faint galaxies outside the local neighbourhood, and a survey which is both deep and wide can then be used to construct statistically meaningful samples of dwarfs outside the local Universe. Indeed, the shape of the galaxy mass function means that even modest survey areas can yield large samples of dwarfs. The use of broad-band optical variability holds particular promise in the near future with the advent of the LSST (Ž. Ivezić et al. 2019; A. E. Watkins et al. 2024), which will offer a 3-day cadence, a single-visit depth of  $r \sim 24.7$  mag, an  $18\,000\text{ deg}^2$  footprint and, at completion, a 10-yr temporal baseline.

In this paper, we use optical broad-band variability to identify AGN in nearby ( $z < 0.4$ ) dwarf ( $10^8 M_\odot < M_* < 10^{10} M_\odot$ ) galaxies, using the VST-COSMOS survey. This survey offers a virtually identical single-visit depth and temporal baseline to what will be available from the LSST at completion. Notwithstanding its  $1\text{ deg}^2$  area, VST-COSMOS enables us to extract statistically significant samples of nearby dwarfs and identify AGN within them via variability, offering a preview of the game-changing science that will be possible using the LSST in the near future.

The plan for this paper is as follows. In Section 2, we describe the selection of dwarf galaxies that host AGN using VST-COSMOS, the morphological classification of these dwarfs via visual inspection and the estimation of environmental parameters (distances to the nearest node, filament, and massive galaxy) using the DisPerSE algorithm (T. Sousbie 2011). In Section 3, we study the properties of our AGN and compare them to the general dwarf population, using control

samples matched in stellar mass and redshift. We explore the fraction of the optical flux that might be contributed by the AGN, the role of interactions in AGN triggering and the morphologies, star formation rates (SFRs) and environments of our AGN hosts compared to the general dwarf population. In Section 4, we use the VST-COSMOS data to quantify the relative frequency of AGN in dwarfs compared to that in massive ( $M_* > 10^{10} M_\odot$ ) galaxies. We summarize our findings in Section 5.

## 2 DATA

### 2.1 A catalogue of AGN selected via variability from the VST-COSMOS survey

The sample of AGN in dwarf galaxies that underpins this study is constructed from observations of the COSMOS field using OmegaCAM (K. Kuijken 2011) on the VST (M. Capaccioli & P. Schipani 2011) at the Cerro Paranal observatory. The construction of the AGN catalogue follows the procedure described in D. De Cicco et al. (2015, 2019) and is briefly outlined here. Light curves are calculated using VST observations for an initial sample of 22 927 sources defined in D. De Cicco et al. (2019). The VST-COSMOS data set used here consists of 68  $r$ -band observations spanning 11 yr. A variability threshold is computed by calculating the average magnitude over the 11-yr baseline and the corresponding root mean square (RMS) deviation of each source from its light curve. The sample of variable sources is then (conservatively) defined as those in which the RMS deviations are in excess of the 95th percentile of the distribution of these values. In order to identify spurious sources, the snapshots and light curves of each variable source are visually inspected and each source is flagged with a quality label from 1 to 3. Sources with a flag of 1 are strong candidates without any obvious problems, those with a flag of 2 are very likely to be variable but have a neighbour, while those with a flag of 3 could be spurious, mostly because of the presence of a very close neighbour. Although the single-visit depth is of 24.6 mag, we limit our analysis to sources brighter than 23.5 mag, in order to minimize contamination from noisy objects. Following D. De Cicco et al. (2015, 2019), a ‘robust’ sample is then defined using sources with a flag of 1 and 2 only.

We briefly discuss other potential sources of galaxy variability – e.g. tidal disruption events (TDEs), supernovae, and host variability due to fluctuations in the star formation history – and describe why they do not contaminate the AGN sample used here. TDEs exhibit well-defined peaks and decay smoothly over time (e.g. A. Zabludoff et al. 2021). In addition, they are also initially much more luminous than what is seen in AGN, with the object brightening by several magnitudes. However, objects in our variable sample exhibit stochastic variability over an 11-yr time-scale and, as discussed in Section 3.2, the RMS of the variable magnitude is less than 0.15 mag. In a similar vein, supernovae are also ruled out because their light-curve shapes, which involve a rapid rise followed by a slower decay, are qualitatively different to those of AGN. For example, in D. De Cicco et al. (2015), which employs the same AGN selection procedure as in this work and has a 5-month baseline, the rate of contamination by supernovae is  $\sim 10$  per cent. However, in D. De Cicco et al. (2019), which utilizes a 3.3 yr baseline (significantly shorter than the baseline in this study) the supernova contamination rate drops to zero.

Finally, the variability is very unlikely to be driven by the star formation history (SFH) of the galaxies due to two reasons. First, the dynamical time-scales of dwarf galaxies, over which we would expect to observe variability driven by the star formation rate (SFR),

are hundreds of Myr (orders of magnitude longer than our 11-yr baseline). Second, if the variability was driven by fluctuations in the SFR then we would expect a correlation between variability and colour and, in particular, see frequent (and preferential) variability in blue galaxies. However, the blue fraction is significant (around 50 per cent, see e.g. S. Kaviraj et al. 2025) and much larger than the fraction of variable objects, making it unlikely that the variability is driven by host galaxy fluctuations due to changes in the SFR.

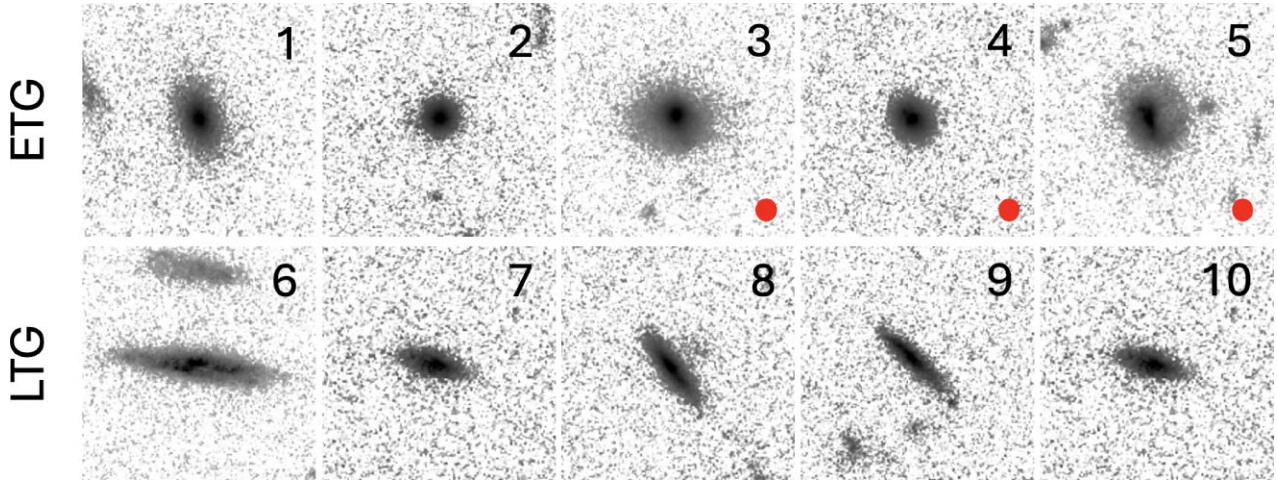
### 2.2 A catalogue of dwarf galaxies that host AGN

To construct our catalogue of dwarf AGN, we combine the variable sample described above with physical parameters (photometric redshifts, stellar masses, rest-frame colours, and SFRs) from the classic version of the COSMOS2020 catalogue (J. R. Weaver et al. 2022). The physical parameters in this data set are calculated by applying the LEPHARE SED-fitting algorithm (S. Arnouts et al. 2002; O. Ilbert et al. 2006) to deep, multiwavelength UV to mid-infrared photometry in around 40 broad and medium band filters, from the following instruments: *GALEX* (M. A. Zamojski et al. 2007), *MegaCam/Canada–France–Hawaii Telescope* (CFHT) (M. Sawicki et al. 2019), *ACS/Hubble Space Telescope* (*HST*; A. Leauthaud et al. 2007), *Hyper Suprime-Cam* (H. Aihara et al. 2019), *Subaru/Suprime-Cam* (Y. Taniguchi et al. 2007, 2015), *VIRCAM/VISTA* (H. J. McCracken et al. 2012), and *IRAC/Spitzer* (M. L. N. Ashby et al. 2013, 2015, 2018; C. L. Steinhardt et al. 2014).

A particular novelty is the inclusion of optical ( $i$ ,  $z$ ) data from the ultradeep layer of the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP) for object detection, which has a point-source depth of  $\sim 28$  mag (H. Aihara et al. 2019). As a comparison, this is  $\sim 10$  mag deeper than the magnitude limit of the SDSS spectroscopic main galaxy sample (MGS; e.g. S. Alam et al. 2015). The accuracy of the photometric redshifts in COSMOS2020 is better than 1 and 4 per cent for bright ( $i < 22.5$  mag) and faint ( $25 < i < 27$  mag) galaxies, respectively.

To construct the sample of dwarf AGN that underpins this study we proceed as follows. We first select objects which are both classified as galaxies by LEPHARE ( $\text{‘type’} = 0$  in the COSMOS2020 catalogue) and as ‘extended’ (i.e. galaxies) in the HSC  $griz$  filters. We then select galaxies which have stellar masses and redshifts in the ranges  $10^8 M_\odot < M_* < 10^{10} M_\odot$  and  $z < 0.4$  respectively, which lie within the HSC-SSP footprint and outside bright-star masks and which have both  $u$ -band and mid-infrared photometry (since a long wavelength baseline improves the accuracy of the parameter estimation, e.g. O. Ilbert et al. 2006). The choice of redshift and stellar mass ranges is driven by the fact that, as described in Section 3, the galaxy population within these ranges is complete. This parent dwarf sample is cross-matched with the sample of variable systems, producing 30 dwarf AGN within the stellar mass and redshift ranges described above. For the purposes of Section 4, we also construct a sample of massive galaxies (which are defined throughout this study as those with  $M_* > 10^{10} M_\odot$ ) in an identical fashion and cross-match it with the variable sample to yield 10 massive galaxies which host AGN at  $z < 0.4$ . The AGN in massive galaxies will be studied in detail in a forthcoming paper (Bichang’a et al., in preparation).

The SED fitting used to derive the physical parameters we use here does not incorporate AGN templates. It is worth noting in this context, however, that current AGN templates are largely built and calibrated based on luminous AGN in massive galaxies and are unlikely to be appropriate for dwarf galaxies. Here, we do not consider X-ray sources, which could host powerful AGN (which might potentially



**Figure 1.** Example images of our dwarf AGN (each image has a linear size of 4 arcsec). The top and bottom rows show early-type galaxies (ETGs) and late-type galaxies (LTGs), respectively. Interacting systems are indicated using a red filled circle in the lower right-hand corner of the image. Galaxies 3 and 5 show internal asymmetries, while galaxy 4 appears to be accreting a smaller companion in its central region. This figure is better viewed online than in print.

affect the SED fitting).<sup>1</sup> Appendix A shows a comparison between the COSMOS2020 photometric redshifts and spectroscopic redshifts of 16 variable dwarfs which have available spectroscopic redshifts, from the recent compilation of A. A. Khostovan et al. (2025). The photometric redshifts show excellent correspondence with their spectroscopic counterparts, with a median difference between the photometric and spectroscopic redshifts of  $\sim 0.0066$ . This suggests that, in spite of the presence of an AGN, the SED fitting produces reliable physical parameters in our sample of galaxies. Note that the upper mass limit of the dwarf regime can vary in the literature, with slightly lower limits adopted by some studies (e.g. R. P. der Marel et al. 2002; F. Davis et al. 2022). Together with the fact that stellar mass errors from SED fitting can be of the order of a few tenths of a dex (e.g. J. R. Weaver et al. 2022) and that the presence of AGN may lead to slight overestimates of stellar mass (e.g. M. Siudek et al. 2024), our choice of upper mass limit for the dwarf regime is reasonable.

### 2.3 Morphological classifications using visual inspection

We use visual inspection of *HST* F814W (*I* band) images in the COSMOS field (A. M. Koekemoer et al. 2007; R. Massey et al. 2010) to morphologically classify our dwarf AGN and control samples (described in Section 3.1 below). We classify our dwarfs by eye into two broad morphological classes: early-type galaxies (ETGs) and late-type galaxies (LTGs). We also identify galaxies that show evidence of a current or recent interaction, such as internal asymmetries or an ongoing merger with another galaxy. Fig. 1 presents example images of our ETGs (top row) and LTGs (bottom row). Interacting systems are indicated using a red filled circle in the lower right-hand corner. Galaxies 3 and 5 show internal asymmetries,

while galaxy 4 appears to be accreting a smaller companion in its central regions.

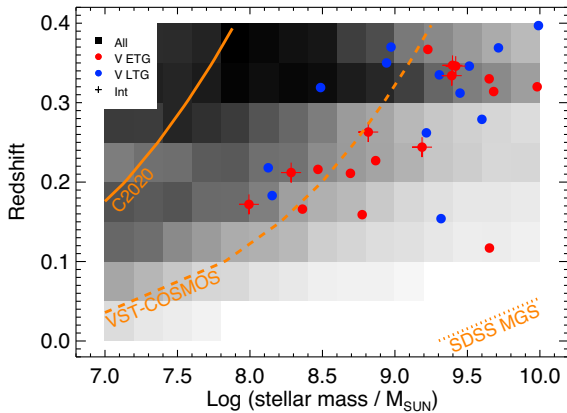
We note that an array of methods have been used in the literature to classify galaxy morphologies, ranging from visual inspection of images to algorithmic techniques which employ morphological parameters. However, visual inspection is typically considered the most accurate technique of morphological classification (e.g. C. Lintott et al. 2011; S. Kaviraj 2014), against which other commonly used methods like morphological parameters (e.g. C. J. Conselice 2003; J. M. Lotz, J. Primack & P. Madau 2004) are calibrated. While morphological parameters are typically employed for large data sets for which visual inspection can be prohibitively time consuming, the size of our data set lends itself well to morphological classification via this method. More importantly, recent work has shown that classical parameters like the CAS system,  $M_{20}$ , and the Gini coefficient are not effective at separating dwarf galaxies into ETGs and LTGs (I. Lazar et al. 2024a). It is also worth noting that, unlike their massive counterparts, dwarf ETGs and LTGs overlap significantly in their rest-frame optical colour distributions (I. Lazar et al. 2024b), as a result of which colour information cannot be reliably used to separate morphological types in this regime.

### 2.4 Environmental parameters calculated using DISPERSE

We use DISPERSE, a structure-finding algorithm, to measure the locations of galaxies within the cosmic web, such as their distances to the nearest nodes, filaments, and massive galaxies. The algorithm measures the density field via Delaunay tessellations, which are calculated using the positions of galaxies (W. E. Schaap & R. van de Weygaert 2000). It then uses this density field to identify the positions of critical points i.e. minima, maxima, and saddles, which correspond to the locations of voids, nodes, and the centres of filaments, respectively. Segments are then used to connect pairs of saddle points and local maxima, forming a set of ridges that describe the network of filaments that define the cosmic web.

A ‘persistence’ parameter sets a threshold value for retaining pairs of critical points within the density map. For example, a persistence of ‘ $N$ ’ results in all pairs which have Poisson probabilities below  $N\sigma$  from the mean being removed. Here, we set the persistence equal to 2, following the methodology of C. Laigle et al. (2018), who have

<sup>1</sup>Note that C. J. Burke et al. (2024) have also recently studied optically variable AGN in the COSMOS field. Their sample is based on the catalogue produced by Y. Kimura et al. (2020), who derive optically variable galaxies from the HSC-SSP using a 3-yr baseline. However, around 90 per cent of the Y. Kimura et al. (2020) sample has X-ray detections, which rules them out of our selection criteria. Furthermore, the C. J. Burke et al. (2024) study focuses on massive galaxies (see for example their fig. 10) and therefore does not overlap with the stellar mass regime studied in this paper.



**Figure 2.** Redshift versus stellar mass of dwarfs in the COSMOS2020 catalogue (shown using the heatmap) and our dwarf AGN (filled circles). ETGs and LTGs are shown colour-coded (see legend). Interacting galaxies are indicated using crosses. The solid, dashed and dotted orange lines show the redshifts below which galaxies of a given stellar mass are complete in COSMOS2020, VST-COSMOS and the SDSS MGS, which have magnitude limits of  $r = 28, 24.6$ , and  $17.77$  respectively.

implemented DISPERSE on redshift slices of similar widths as in our analysis, constructed from the COSMOS2015 (C. Laigle et al. 2016) catalogue. We note that the same methodology has been used to do an identical density analysis using the COSMOS2020 catalogue by I. Lazar et al. (2023) and B. Bichang’a et al. (2024). We refer readers to T. Soubie (2011) for further details about the algorithm.

The accurate COSMOS2020 redshifts enable us to employ relatively narrow redshift slices to build density maps. We only use massive galaxies for this purpose, because they have the smallest redshift errors and dominate the local gravitational potential well. Individual massive galaxies are weighted by the area under their redshift probability density functions that is contained within the redshift slice in question. This takes into account the fact that the photometric redshifts in COSMOS2020, albeit very accurate, do have associated uncertainties.

Finally, we consider the types of environments that are likely to be present in the VST-COSMOS footprint in our redshift range of interest, by considering the  $M_{200}$  values (which are proxies for the virial masses) of groups identified in the literature (A. Finoguenov et al. 2007; M. R. George et al. 2011; G. Gozaliasl et al. 2014, 2019). The virial masses of groups in VST-COSMOS lie in the range  $10^{12.9} M_{\odot} < M_{200} < 10^{13.8} M_{\odot}$ , with a median value of  $10^{13.4} M_{\odot}$ . For comparison, a small cluster like Fornax has a virial mass of  $\sim 10^{13.9} M_{\odot}$  (M. J. Drinkwater, M. D. Gregg & M. Colless 2001), while large clusters like Virgo and Coma have virial masses of  $\sim 10^{15} M_{\odot}$  (e.g. P. Fouqué et al. 2001; R. Gavazzi et al. 2009). The galaxy population considered in this study therefore spans a wide spectrum of environments from large groups to the field but does not reside in rich clusters.

### 3 PROPERTIES OF DWARF GALAXIES THAT HOST AGN

In Fig. 2, we present the stellar masses and redshifts of our dwarf AGN. The heatmap shows the COSMOS2020 galaxy population, while the filled circles indicate our dwarf AGN. Different morphological classes are shown colour-coded (see legend), while crosses indicate interacting galaxies. The solid orange line shows the redshift at which a galaxy population of a given stellar mass is likely to be

complete in COSMOS2020, constructed using the methodology of S. Kaviraj et al. (2025). This is defined as the redshift at which a purely old ‘simple stellar population’ (SSP) of a given stellar mass that forms in an instantaneous burst at  $z = 2$ , is detectable, at the depth of the HSC Ultra-deep imaging in COSMOS (which underpins object detection in the COSMOS2020 catalogue<sup>2</sup>). We construct our SSPs using the G. Bruzual & S. Charlot (2003) stellar models, assuming half solar metallicity.

This purely old SSP represents a faintest ‘limiting’ case, since real galaxies, which are not composed uniquely of old stars, will be more luminous than this limiting value. Thus, if this limiting case is detectable at a given survey depth, then the entire galaxy population at a given stellar mass will also be detectable in the survey in question. The parameter space below the solid orange line therefore represents the region where the COSMOS2020 galaxy population is complete. Fig. 2 indicates that galaxy populations in COSMOS2020 with stellar masses greater down to  $10^8 M_{\odot}$  are complete out to  $z \sim 0.4$  [note that J. R. Weaver et al. (2022) come to a similar conclusion]. As a comparison, the dashed and dotted orange lines show the corresponding completeness thresholds for VST-COSMOS and the SDSS MGS, respectively.

Note that this does not mean that galaxies will not exist in a given survey above the curves that define these completeness regions. Rather, galaxy populations will become progressively more biased towards bluer, star-forming systems as we move above and further away from the curves. Fig. 2 indicates that the vast majority of our dwarf AGN also fall within the completeness region for VST-COSMOS. However, only galaxies at the upper mass end of our study would be complete in the SDSS MGS, and that too only in the very local Universe.

#### 3.1 Construction of control samples

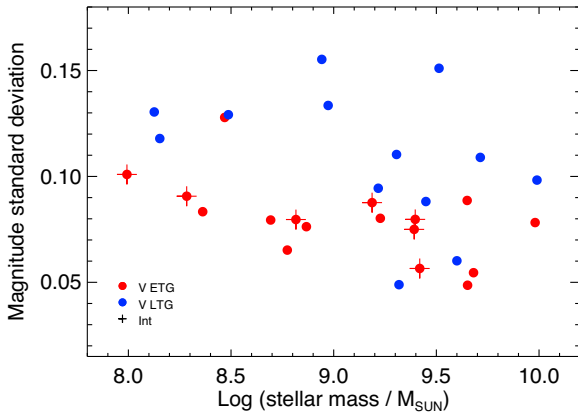
To enable comparisons between our dwarf AGN and the general dwarf population, we construct control samples which consist of populations of non-variable dwarfs with the same mass and redshift distribution and 5 times the number of galaxies as our dwarf AGN. A control population is constructed by identifying, for every AGN, all non-AGN within stellar mass and redshift tolerances of 0.05 dex and 0.03, respectively. We then select, at random, five of these galaxies to be the control counterparts of the AGN in question. No control galaxy is assigned to more than one AGN.

Recall that, as described above, incompleteness will tend to bias the galaxy population towards bluer galaxies. In the massive-galaxy regime, bluer galaxies tend to have more late-type morphology (e.g. R. Buta et al. 1994; I. Strateva et al. 2001), although this effect is much less pronounced in the dwarf regime, where ETGs and LTGs, at least in the nearby Universe ( $z < 0.08$ ), exhibit similar rest-frame colour distributions (I. Lazar et al. 2024a). Thus, to check that incompleteness does not bias our conclusions, we perform our AGN versus control comparisons in two ways. First, we construct a control population that is matched in redshift and stellar mass to our full sample of 30 dwarf AGN. Second, we construct a different control sample that is matched in redshift and stellar mass to the 22 dwarf AGN that reside within the VST-COSMOS completeness region i.e. below the dashed orange line in Fig. 2. We then compare the properties of each control sample and its corresponding dwarf AGN population.

<sup>2</sup>Note that hardly any dwarfs are consistent with a purely old stellar population (e.g. I. Lazar et al. 2024a), making this a conservative criterion.

**Table 1.** Comparisons between our dwarf AGN and control galaxies. The upper section of this table considers all dwarf AGN, while the lower section considers the case where our dwarf AGN are restricted to the completeness region of VST-COSMOS (i.e. the region below the dashed orange line in Fig. 2). In each section, we show the fractions of galaxies that are ETGs (first row), interacting (second row), and quenched (third row) in the dwarf AGN (left-hand column) and the corresponding control population (right-hand column). Uncertainties are calculated following E. Cameron (2011) and shown using superscripts.

	All dwarf AGN	Control
	AGN	
Early-type fraction	$0.57^{+0.09}$	$0.39^{+0.04}$
Interacting fraction	$0.23^{+0.08}$	$0.16^{+0.03}$
Quenched fraction	$0.14^{+0.06}$	$0.23^{+0.03}$
Dwarf AGN in completeness region only		
	AGN	Control
Early-type fraction	$0.64^{+0.10}$	$0.36^{+0.05}$
Interacting fraction	$0.23^{+0.09}$	$0.15^{+0.03}$
Quenched fraction	$0.18^{+0.08}$	$0.32^{+0.04}$



**Figure 3.** The standard deviation of the  $r$ -band magnitude in our dwarf AGN. ‘V ETG’ = variable ETG and ‘V LTG’ = variable LTG. Different morphological classes are shown colour-coded (see legend). Interacting galaxies are indicated using crosses. As described in Section 3.2, if the variability can be attributed to the AGN, then this standard deviation is also an estimate of the fraction of  $r$ -band flux that is likely to be contributed by the AGN.

Table 1 summarizes the comparisons between our dwarf AGN and control galaxies, which are described in detail in Sections 3.3, 3.4, and 3.5 below. The upper section of this table considers all dwarf AGN, while the lower section considers the case where our dwarf AGN are restricted to the completeness region of VST-COSMOS. It is worth noting that, within the uncertainties, the values in both sections of Table 1 are consistent with each other. In other words, restricting our dwarf AGN to the VST-COSMOS completeness region does not alter our conclusions. This is likely driven by the fact that 73 per cent of our dwarf AGN lie within the VST-COSMOS completeness region and, those that do not, still reside close to the completeness threshold (i.e. the dashed orange line in Fig. 2).

### 3.2 AGN flux fractions

It is instructive to first consider the fractions of flux that may be contributed by the AGN in our variable dwarf population. Fig. 3 presents the standard deviation of the  $r$ -band magnitude in our

variable dwarfs over our 11-yr baseline, as a function of their stellar mass. This quantity traces the extent of the variability in our dwarf AGN. Differentiating the magnitude equation,  $m = -2.5 \log(F) + \text{constant}$ , where  $m$  is the magnitude and  $F$  is the flux, yields  $\Delta m = 1.09 \times (\Delta F)/F$ . If the variability is attributed to the AGN, then the standard deviation ( $\Delta m$ ) provides an estimate of the fraction of flux that is likely to be contributed by the AGN. The  $r$ -band flux fractions of our dwarf AGN are modest and range from  $\sim 5$  to  $\sim 15$  per cent. The median flux fraction in AGN that are hosted by LTGs is around a factor of 1.4 higher than that in their ETG counterparts.

Finally, it is interesting to note that the flux fractions in dwarf AGN that are classified as interacting are not larger than in their non-interacting counterparts. If the flux fractions indicate the AGN accretion rates relative to the brightness of the host galaxy, then this suggests that interactions are unlikely to drive higher accretion rates in our dwarf AGN (see also A. Erőstegui et al. 2025).

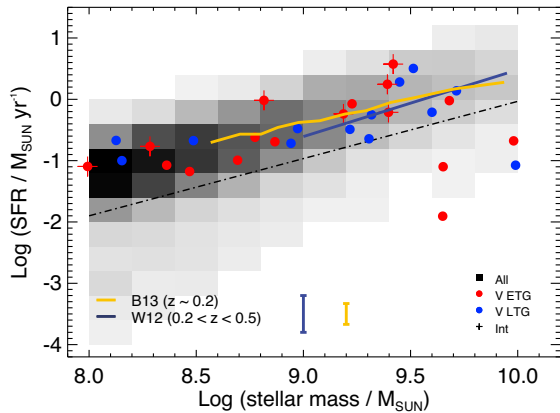
### 3.3 Morphology and the role of interactions

We begin our analysis of how our dwarf AGN compare to the general dwarf population by considering their morphological properties. The first row in both sections of Table 1 indicates that the fraction of ETGs is elevated compared to that in the control sample. However, within the uncertainties, the fraction of interacting galaxies in the AGN and control populations (second row in both sections of the table) are indistinguishable, suggesting that, at least in this sample of dwarf AGN, interactions appear unlikely to play an important role in AGN triggering. This is similar to the findings of other studies in the recent literature (e.g. A. Erőstegui et al. 2025). It is interesting to note that the only interacting systems that host dwarf AGN are those with early-type morphology. This may indicate that the presence of interactions makes it more likely that AGN are triggered in dwarf ETGs, while they do not similarly impact the likelihood of AGN triggering in LTGs.

The comparable interacting fractions in our AGN and control samples are consistent with the findings of B. Bichang’a et al. (2024), who have used an identical visual classification methodology to compare the morphological properties of nearby dwarf AGN (identified via SED fitting) to that in the general dwarf population. However, unlike this study, Bichang’a et al. (2024) do not find a difference in the morphological mix of dwarf AGN and the reasons for this discrepancy are unclear.

### 3.4 Star formation activity

We proceed by considering, in Fig. 4, the SFRs of our dwarf AGN. Recall here that the estimation of SFRs does not include AGN templates but that the excellent correspondence between spectroscopic and photometric redshifts indicates that the physical parameters (e.g. stellar masses, redshifts, and SFRs) derived via SED fitting are robust. The heatmap shows the star formation main sequence of the COSMOS2020 galaxy population in our redshift range of interest ( $z < 0.4$ ), while the filled circles indicate our dwarf AGN. The dashed–dotted line indicates the lower limit of the main locus of the star formation main sequence. We consider galaxies that fall below this threshold to be quenched. Examples of main-sequence loci from P. S. Behroozi, R. H. Wechsler & C. Conroy (2013) and K. E. Whitaker et al. (2012), which agree well with the COSMOS2020 locus, are shown overplotted and their standard deviations are indicated using the error bars. The third row in each section of Table 1 compares the quenched fractions in the dwarf AGN and control samples.



**Figure 4.** The star formation main sequence of galaxies in the COSMOS2020 catalogue (shown using the heatmap) and our dwarf AGN (filled circles). ‘V ETG’ = variable ETG and ‘V LTG’ = variable LTG. The dashed-dotted line demarcates the lower limit of the main locus of the star formation main sequence. Different morphological classes are shown colour-coded (see legend). Interacting galaxies are indicated using crosses. Examples of other main-sequence loci from P. S. Behroozi et al. (2013, B13) and K. E. Whitaker et al. (2012, W12) are shown overplotted and their standard deviations are indicated using the error bars.

Taking the uncertainties into account, the quenched fraction in dwarf AGN does not show differences with that in the control sample. There is, therefore, no indication of significant prompt quenching of star formation in our sample, in a similar vein to the conclusions of recent work (e.g. B. Bichang’a et al. 2024). Note that this result does not preclude the existence of negative AGN feedback, as this process does not necessarily result in lower instantaneous SFRs during individual star formation episodes (e.g. S. R. Ward et al. 2022). Indeed, indicators of quenching tend to correlate more strongly with BH mass, which is a measure of the cumulative output of the AGN over its lifetime (e.g. J. M. Piotrowska et al. 2022).

### 3.5 Locations in the cosmic web

We complete our comparison of the dwarf AGN and control samples by exploring their projected distances to the nearest nodes, filaments, and massive galaxies. Fig. 5 shows the distributions of the distances to the nearest nodes (top row), filaments (middle row), and massive galaxies (bottom row) for the AGN (black) and control (red) populations. The left-hand column presents the case where all dwarf AGN are considered, while the distributions in the right-hand column are restricted to galaxies in the VST-COSMOS completeness region (indicated by ‘CR only’ in the panels). Kolmogorov–Smirnov (KS) tests for the former case produce  $p$ -values of 0.06, 0.92, and 0.89 for the comparison of distances to the nearest nodes (top row), filaments (middle row), and massive galaxies (bottom row), respectively, while the latter case produces  $p$ -values of 0.06, 0.73, and 0.83. Recall that  $p$ -values less than 0.05 reject the null hypothesis that the distributions are drawn from the same parent distribution.

While the median values suggest that dwarf AGN may reside at slightly lower distances to nodes than their control counterparts, the KS tests indicate that the dwarf AGN generally appear to reside in similar environments to the control sample. This appears consistent with the findings of other recent studies (e.g. M. T. Kristensen, K. Pimbblet & S. Penny 2020; M. Siudek, M. Mezcua & J. Krywult 2023; B. Bichang’a et al. 2024), which have employed varying definitions of local environment but also do not find strong

differences between the environments of dwarf AGN and the general dwarf population.

## 4 THE RELATIVE INCIDENCE OF AGN IN THE DWARF AND MASSIVE-GALAXY REGIMES

We complete our study by calculating the relative incidence of AGN in the dwarf and massive-galaxy regimes. As noted above, both host and AGN detectability can impair our ability to calculate the true AGN fraction, particularly in the dwarf regime where both the host galaxy and the AGN can be relatively faint. Therefore, instead of attempting to derive this quantity, we explore the prevalence of AGN in the dwarf regime by estimating the *relative* frequency of AGN in dwarfs and massive galaxies. This allows us to gain insights into the AGN population in dwarfs, given what we already know about the incidence of AGN in the massive-galaxy regime.

If AGN in dwarfs and massive galaxies are selected using an identical technique, then the ratio of the AGN fractions in the two populations can be used to quantify the relative frequency of AGN in these two mass regimes. Extrapolating this further, since it is thought that the BH occupation fraction of massive galaxies is high, and possibly equal to 1, this ratio may then also provide insights into the BH occupation fraction in dwarf galaxies.

Given the completeness issues discussed above, we cannot simply calculate the observed AGN fractions in the dwarf and massive-galaxy regimes and divide one by the other. To calculate this ratio we have to take into account, in both mass regimes, (1) the true number of galaxies, (2) how many of those galaxies are actually classified as AGN via variability, and (3) the detectability of AGN, given the depth of the VST-COSMOS images and the methodology being employed (in our case, variability).

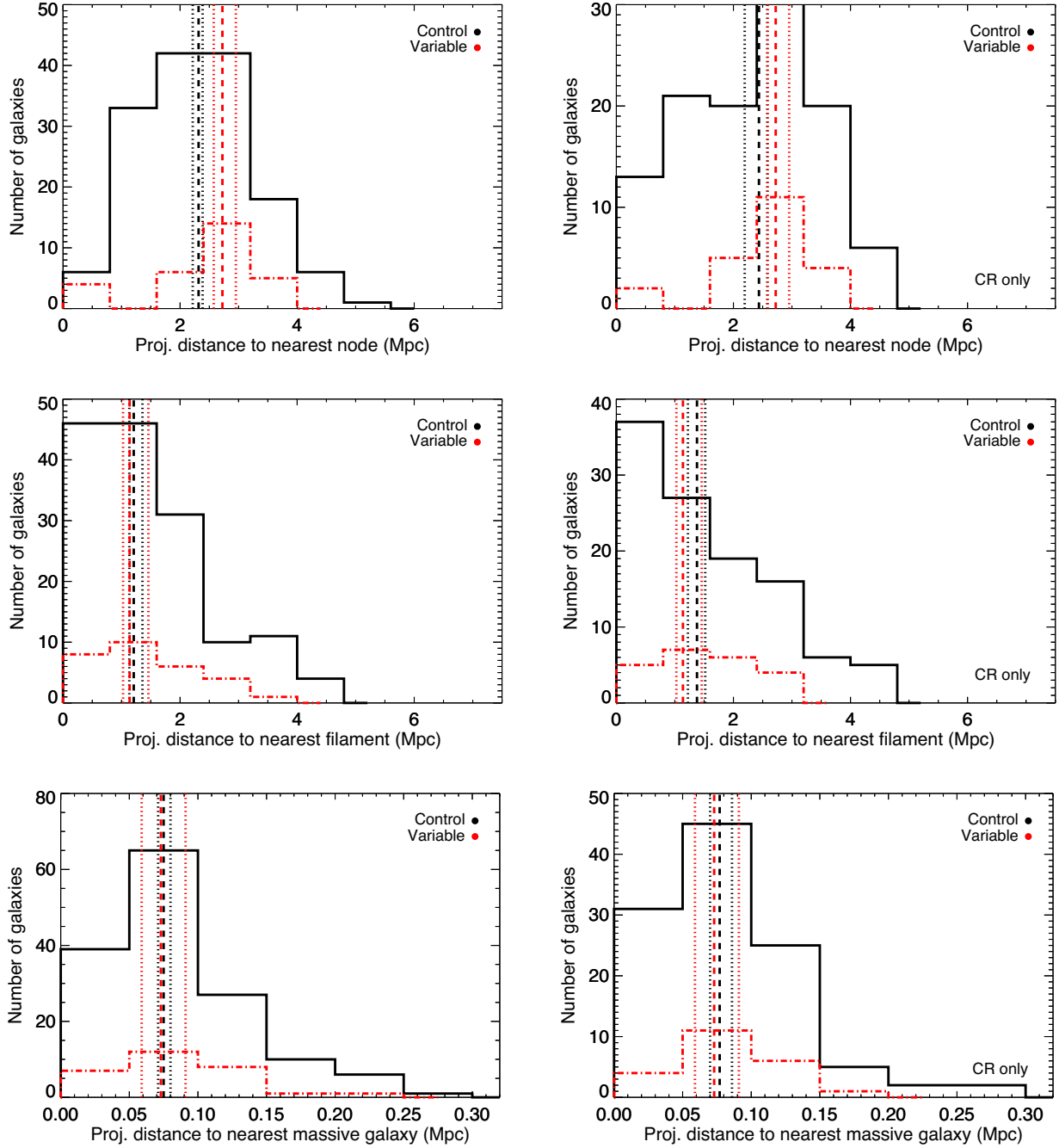
In a given stellar mass and redshift range, if  $N_0$  is the true number of galaxies,  $D$  is the fraction of AGN that are detectable (which we shall use as our definition of detectability) and  $N_{\text{agn}}$  is the actual number of AGN detected, then the true AGN fraction ( $f_{\text{agn}}$ ) is given by

$$f_{\text{agn}} = \frac{N_{\text{agn}}}{D \cdot N_0}. \quad (1)$$

A trivial consequence of this equation is that the measured number of AGN ( $N_{\text{agn}}$ ) leads to an accurate calculation of the true AGN fraction ( $f_{\text{agn}}$ ) only if the detectability ( $D$ ) is equal to 1 and the true number of galaxies ( $N_0$ ) is known.

The detectability ( $D$ ) has two components. The first, which we shall refer to as ‘host detectability’ ( $D_{\text{host}}$ ), is driven by the depth of the (typically optical) survey in which the AGN host galaxies are being identified.  $D_{\text{host}}$  determines which AGN hosts appear in this survey in the first place and, as we show below, can be estimated relatively easily. The second, which we shall refer to as ‘AGN detectability’ ( $D_{\text{agn}}$ ), determines the likelihood of an AGN being identified as such by the method being employed.  $D_{\text{agn}}$  can be more challenging to estimate because it is dependent on the details of the method, how it is being employed, and the types of objects being studied. We illustrate this point further using two examples from the recent literature.

Some AGN-identification methods, such as the commonly used ‘BPT’ technique that employs optical emission-line ratios (e.g. J. A. Baldwin, M. M. Phillips & R. Terlevich 1981; L. J. Kewley et al. 2001), are designed to separate star-forming systems from AGN. However, as noted in the introduction (and apparent from Fig. 2), dwarfs in shallow surveys like the SDSS are biased towards systems which are highly star forming, particularly outside the very local



**Figure 5.** Projected distance to the nearest node (top panels), nearest filament (middle panels), and nearest massive galaxy (bottom panels) of dwarf AGN (red) and their control counterparts (black). The left-hand column presents the case where all dwarf AGN are considered, while the distributions in the right-hand column are restricted to galaxies in the VST-COSMOS completeness region (indicated by ‘CR only’ in the panels). Medians and their uncertainties, calculated using bootstrapping, are shown using the dashed and dotted lines, respectively.

Universe. Applying this technique to such dwarfs using, for instance, SDSS single-fibre spectroscopy which samples a significant fraction of light around the galaxy centre, leads to AGN fractions close to zero (e.g. A. E. Reines et al. 2013). However, if the same technique is applied to SDSS dwarfs in a spatially resolved manner, via integral field spectroscopy, the measured AGN fractions are significantly higher. M. Mezcuca & H. Domínguez Sánchez (2024) have recently explored spatially resolved emission-line diagnostics, using the [N II]-BPT, [S II]-BPT, [O I]-BPT diagrams (L. J. Kewley et al. 2001, 2006; G. Kauffmann et al. 2003) augmented by a WHAN analysis

(R. Cid Fernandes et al. 2010) in the nearby SDSS dwarf population. A conservative classification, which only identifies objects as AGN if they occupy the AGN regions in all three BPT diagrams yields an AGN fraction of  $\sim 20$  per cent. However, if objects that are classified as star forming in the [N II]-BPT diagram but AGN in the [S II]-BPT and [O I]-BPT diagrams are also considered to be AGN, then the corresponding AGN fraction rises to greater than 50 per cent.

In short, a spatially resolved analysis of optical emission-line ratios in the dwarf regime suggests (significantly) higher AGN fractions than those derived using single fibre spectroscopy. This discrepancy

is likely driven by several reasons. First, while star formation will dilute the contribution of the AGN to the emission lines in both single-fibre and spatially resolved measurements, the latter is able to isolate the region around the BH where the gas ionization is dominated by the AGN, increasing the chances of detecting its presence. Second, in many dwarfs the central BHs may wander beyond the 3 arcsec extent of the central SDSS fibre, making it challenging to identify the presence of the AGN. Finally, the accretion disc gets significantly hotter in the dwarf regime, which is likely to cause traditional BPT line ratios to fall outside the parameter space that is typically attributed to AGN (J. M. Cann et al. 2019), resulting in an underestimate of the AGN fraction. The basic point here is that the same technique employed in different ways or in different mass regimes can result in very different values of  $D_{\text{agn}}$ .

In other cases, the ancillary data that are used to identify AGN may restrict the part of the dwarf AGN population that is actually detectable. For example, F. Davis et al. (2022) have combined deep optical and radio data, from the Hyper Suprime-Cam (HSC) and the Low-Frequency Array (LOFAR), respectively, to study radio AGN in nearby dwarfs. While the depth of the optical HSC data results in a relatively unbiased sample of host galaxies at the redshifts in question, the LOFAR detection limit is close to the peak luminosities of the dwarf radio sources that are LOFAR-detectable. As these peak luminosities are expected to decay to values fainter than the LOFAR detection limit over time-scales of a few tens of Myr, the dwarf AGN found by this study are preferentially young. In other words  $D_{\text{agn}}$  in this sample of dwarfs is a strong function of the age of the AGN. Finding the more numerous population of longer-lived dwarf AGN, which have fainter radio fluxes, requires a deeper radio survey. This will be difficult for the foreseeable future, given that the capabilities of the LOFAR surveys will remain unmatched well into the Square Kilometer Array era (e.g. T. W. Shimwell et al. 2017).

In summary, the value of  $D_{\text{agn}}$  is both difficult to measure and, in many cases, very low, because the properties of the dwarf host galaxies and the details of the AGN identification technique (or both) make it challenging to identify these systems. The final detectability ( $D$ ) in equation (1) is a combination of  $D_{\text{host}}$  and  $D_{\text{agn}}$ , i.e.  $D = D_{\text{host}} \times D_{\text{agn}}$ . Given the arguments above, it is reasonable to suggest that the extremely low dwarf AGN fractions reported in some studies in the recent literature are produced by low values of detectability, both in terms of  $D_{\text{host}}$  and  $D_{\text{agn}}$ , and not because AGN do not exist in appreciable numbers in dwarf galaxies.

If we now wish to compare the AGN fractions in dwarfs and massive galaxies, it follows from equation (1) that,

$$\left(\frac{f_{\text{agn,d}}}{f_{\text{agn,m}}}\right) = \left(\frac{N_{\text{agn,d}}}{N_{\text{agn,m}}}\right) \times \left(\frac{N_{0,m}}{N_{0,d}}\right) \times \left(\frac{D_{\text{host,m}}}{D_{\text{host,d}}}\right) \times \left(\frac{D_{\text{agn,m}}}{D_{\text{agn,d}}}\right), \quad (2)$$

where ‘d’ and ‘m’ denote dwarfs and massive galaxies, respectively. Recall that, in our study, dwarfs are defined as galaxies in the stellar mass range  $10^8 M_{\odot} < M_{\star} < 10^{10} M_{\odot}$ , massive galaxies as those in the stellar mass range  $M_{\star} > 10^{10} M_{\odot}$ <sup>3</sup>, and the redshift range of interest is  $z < 0.4$ .

$N_{\text{agn,d}}$  and  $N_{\text{agn,m}}$  are simply the measured number of AGN in the dwarf and massive-galaxy regimes.  $N_{0,d}$  and  $N_{0,m}$  can be calculated directly from the COSMOS2020 catalogue, since this data set is complete down to a stellar mass of  $10^8 M_{\odot}$  and out to  $z = 0.4$  (Fig. 2). We therefore take  $N_{0,d}$  and  $N_{0,m}$  to be the numbers of galaxies

in the dwarf and massive regimes in COSMOS2020, according to the mass range definitions above.  $D_{\text{host,d}}$  and  $D_{\text{host,m}}$  can be derived by calculating the fraction of galaxies, in the dwarf and massive regimes respectively, that are brighter than the magnitude limit of our study ( $r = 23.5$ ). Note that, since we are considering fractions in equation (2), the footprint of the region being considered is not relevant to the calculation. However, cosmic variance may affect the value of  $N_{0,d}/N_{0,m}$ , since the shape of the galaxy mass function at low redshift may show some variation with environment (e.g. C. Papovich et al. 2018). Inserting the relevant values into equation (2) then yields

$$\left(\frac{f_{\text{agn,d}}}{f_{\text{agn,m}}}\right) = 0.65^{+0.52}_{-0.26} \times \left(\frac{D_{\text{agn,m}}}{D_{\text{agn,d}}}\right). \quad (3)$$

The errors bars are derived from the minimum and maximum values that  $f_{\text{agn}}$  can have, given the uncertainties on the individual parameters. Note that the error bars on the central value (0.65) are relatively large, which is driven primarily by the sample size. Finally, we consider the value of  $D_{\text{agn,m}}/D_{\text{agn,d}}$ . In our case, D. De Cicco et al. (2019) have shown that the fraction of known AGN that are detected via variability increases significantly as the magnitudes of the host galaxies become progressively brighter. This indicates that  $D_{\text{agn}}$  is larger for more massive galaxies, since they are brighter and therefore ( $D_{\text{agn,m}}/D_{\text{agn,d}}$ ) is greater than 1. Thus,

$$\left(\frac{f_{\text{agn,d}}}{f_{\text{agn,m}}}\right) > 0.65^{+0.52}_{-0.26}. \quad (4)$$

This, together with the fact that the central value (0.65) indicates that AGN frequencies are within a factor of 2 of each other, suggests that the incidence of AGN in dwarf and massive galaxies (as defined by the stellar mass ranges above) are likely to be similar. It is worth noting that if the lower limit of the dwarf stellar mass range is changed from  $10^8$  to  $10^9 M_{\odot}$ , the lower limit on the right-hand side in equation (4) above changes to  $0.78^{+0.70}_{-0.35}$ . The increase in the value of  $f_{\text{agn,d}}/f_{\text{agn,m}}$  as we consider more massive dwarfs suggests that the AGN fraction in the dwarf regime may increase as a function of stellar mass, similar to what is thought to happen in the massive-galaxy regime (e.g. J. Aird et al. 2012). Given the relatively small number of galaxies in our sample, it is difficult to explore this point in more detail because using narrower stellar mass ranges reduces the number of galaxies, making the calculations noisier. It is worth noting here that a similar lack of evolution in the AGN fraction between dwarf and massive galaxies, when observational biases are properly accounted for, has been found via X-ray studies (e.g. K. L. Birchall et al. 2022).

Finally, as mentioned above, the relative incidence of AGN derived above may offer insights into the BH occupation fractions in the two mass regimes. If the BH occupation fraction in massive galaxies is close to unity and the incidence of AGN in dwarfs and massive galaxies is similar then, given the arguments above, it seems reasonable to suggest that the BH occupation fraction in dwarf galaxies may also be similarly high. This precludes a scenario in which only the dwarfs which show strong variability actually have BHs, which seems unlikely.

## 5 SUMMARY

We have studied 30 nearby ( $z < 0.4$ ) dwarf ( $10^8 M_{\odot} < M_{\star} < 10^{10} M_{\odot}$ ) galaxies, which show strong broad-band variability over an 11-yr baseline, in the VST-COSMOS survey. We have combined physical parameters from the COSMOS2020 catalogue with visual morphological classifications derived from *HST* images, to study

<sup>3</sup>The properties of variability-selected AGN in massive galaxies will be presented in a forthcoming paper (Bichang’a et al., in preparation).

key properties of our dwarf AGN: AGN flux fractions, morphologies and the role of interactions, SFRs and local environment. We have then compared our dwarf AGN to the general dwarf population using control samples that are matched in stellar mass and redshift. Finally, we have quantified the relative frequencies of AGN in the dwarf and massive regimes in the nearby Universe. Our main conclusions are as follows:

- (i) The  $r$ -band flux fractions contributed by the AGN in our sample are modest (between 5 and 15 per cent of the total flux of the system).
- (ii) The flux fractions in dwarf AGN that are interacting are not larger than in their non-interacting counterparts, suggesting that interactions do not drive higher accretion rates in our dwarf AGN.
- (iii) The fraction of ETGs is elevated in dwarf AGN compared to the control sample. However, the fraction of interacting galaxies in the AGN and control populations are indistinguishable within the uncertainties, indicating that interactions do not play an important role in AGN triggering in our sample.
- (iv) Interacting systems only appear within dwarf AGN which are morphologically early-type. This suggests that the presence of interactions makes it more likely that AGN are triggered in dwarf ETGs, while they do not similarly impact the likelihood of AGN triggering in dwarf LTGs.
- (v) We do not find evidence that AGN activity in our dwarf sample gives rise to significant prompt quenching of star formation activity.
- (vi) Dwarf AGN reside at similar distances to nodes, filaments, and massive galaxies as their control counterparts, suggesting that AGN triggering in the dwarf regime is not strongly correlated with local environment.
- (vii) By combining the true number of galaxies, the detectability of AGN and the measured numbers of variability-selected AGN in dwarf and massive galaxies, we have studied the relative frequency of AGN in these two mass regimes. The incidence of AGN in dwarfs and massive galaxies is likely to be similar (within less than a factor of 2 of each other), with some evidence that the AGN fraction increases with stellar mass in the dwarf regime.

Combined with the recent literature, our study suggests that, not only do AGN exist in dwarfs, the incidence of AGN in this regime may be similar to that in massive galaxies. While this empirical result is interesting in its own right, it has strong implications for theoretical models. Given the conclusions of this study (and the recent observational results spanning a range of redshifts), it is unclear whether the lack of BH growth in current cosmological simulations is driven primarily by the characteristics of these simulations, such as their relatively low spatial and mass resolutions, or whether they do in fact faithfully reproduce the behaviour of real dwarf galaxies. It is possible that the implementation of BHs and their interaction with their host galaxies in cosmological simulations will require some revision, informed in part by high-resolution simulations that can probe this process in more detail (e.g. S. Koudmani, D. Sijacki & M. C. Smith 2022). The advent of surveys like the LSST, which will offer optical images of similar depth to that used here, but with a footprint that is four orders of magnitude larger, are keenly anticipated. Studies using such surveys will allow us to study large statistical samples of variability-selected AGN and compare them to more detailed control samples. This is likely to put the results presented here on a firmer statistical footing and offer revolutionary insights into our emerging understanding of the role of AGN in the evolution of dwarf galaxies.

## ACKNOWLEDGEMENTS

We thank the anonymous referee for many constructive comments that helped us improve the original manuscript. We thank Niel Brandt for many interesting discussions. S. Kaviraj, IL, and AEW acknowledge support from the STFC (grant numbers ST/Y001257/1 and ST/X001318/1). S. Kaviraj acknowledges a Senior Research Fellowship from Worcester College Oxford. DD acknowledges PON R&I 2021, CUP E65F21002880003, Fondi di Ricerca di Ateneo (FRA), linea C, progetto TORNADO, and the financial contribution from PRIN-MIUR 2022 and from the Timedomes grant within the ‘INAF 2023 Finanziamento della Ricerca Fondamentale’. BB acknowledges a PhD studentship funded by the Centre for Astrophysics Research at the University of Hertfordshire. S. Koudmani was supported by a Research Fellowship from the Royal Commission for the Exhibition of 1851 and a Junior Research Fellowship from St Catharine’s College Cambridge.

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from the Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper makes use of software developed for Vera C. Rubin Observatory. We thank the Rubin Observatory for making their code available as free software at <http://pipelines.lsst.io/>. This paper is based on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by the Subaru Telescope and Astronomy Data Center (ADC) at NAOJ. Data analysis was in part carried out with the cooperation of Center for Computational Astrophysics (CfCA), NAOJ. We are honoured and grateful for the opportunity of observing the Universe from Maunakea, which has the cultural, historical, and natural significance in Hawaii.

## DATA AVAILABILITY

The galaxy sample used in this study has been constructed using the catalogue from J. R. Weaver et al. (2022). The density maps were created using the DisPerSE algorithm which is described in T. Sousbie (2011).

## REFERENCES

- Aihara H. et al., 2019, *PASJ*, 71, 114
- Aird J. et al., 2012, *ApJ*, 746, 90
- Alam S. et al., 2015, *ApJS*, 219, 12
- Anglés-Alcázar D., Faucher-Giguère C.-A., Quataert E., Hopkins P. F., Feldmann R., Torrey P., Wetzel A., Kereš D., 2017, *MNRAS*, 472, L109
- Arjona-Galvez E., Di Cintio A., Grand R. J. J., 2024, *A&A*, 690, A286
- Arnouts S. et al., 2002, *MNRAS*, 329, 355
- 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
- Baker W. M. et al., 2025, *A&A*, 697, A90
- Baldassare V. F., Geha M., Greene J., 2020a, *ApJ*, 896, 10

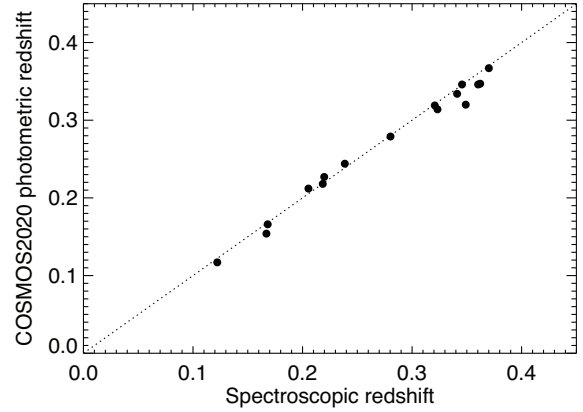
- Baldassare V. F., Dickey C., Geha M., Reines A. E., 2020b, *ApJ*, 898, L3
- Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5
- Barai P., de Gouveia Dal Pino E. M., 2019, *MNRAS*, 487, 5549
- Beckmann R. S. et al., 2017, *MNRAS*, 472, 949
- Behroozi P. S., Wechsler R. H., Conroy C., 2013, *ApJ*, 770, 57
- Bellovary J. M. et al., 2021, *MNRAS*, 505, 5129
- Bichang'a B., Kaviraj S., Lazar I., Jackson R. A., Das S., Smith D. J. B., Watkins A. E., Martin G., 2024, *MNRAS*, 532, 613
- Birchall K. L., Watson M. G., Aird J., 2020, *MNRAS*, 492, 2268
- Birchall K. L., Watson M. G., Aird J., Starling R. L. C., 2022, *MNRAS*, 510, 4556
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Bundy K. et al., 2015, *ApJ*, 798, 7
- Burke C. J. et al., 2022, *MNRAS*, 516, 2736
- Burke C. J., Liu Y., Ward C. A., Liu X., Natarajan P., Greene J. E., 2024, *ApJ*, 971, 140
- Buta R., Mitra S., de Vaucouleurs G., Corwin H. G. J., 1994, *AJ*, 107, 118
- Cameron E., 2011, *PASA*, 28, 128
- Cann J. M., Satyapal S., Abel N. P., Blecha L., Mushotzky R. F., Reynolds C. S., Secrest N. J., 2019, *ApJ*, 870, L2
- Capaccioli M., Schipani P., 2011, *The Messenger*, 146, 2
- Chen C. T. J. et al., 2017, *ApJ*, 837, 48
- Cid Fernandes R., Stasińska G., Schlickmann M. S., Mateus A., Vale Asari N., Schoenell W., Sodré L., 2010, *MNRAS*, 403, 1036
- Conselice C. J., 2003, *ApJS*, 147, 1
- Croton D. J. et al., 2006, *MNRAS*, 365, 11
- Davis F. et al., 2022, *MNRAS*, 511, 4109
- Davis T. A. et al., 2020, *MNRAS*, 496, 4061
- De Cicco D. et al., 2015, *A&A*, 574, A112
- De Cicco D. et al., 2019, *A&A*, 627, A33
- Dickey C. M., Geha M., Wetzel A., El-Badry K., 2019, *ApJ*, 884, 180
- Drinkwater M. J., Gregg M. D., Colless M., 2001, *ApJ*, 548, L139
- Dubois Y., Volonteri M., Silk J., Devriendt J., Slyz A., Teyssier R., 2015, *MNRAS*, 452, 1502
- Dubois Y. et al., 2021, *A&A*, 651, A109
- Duc P.-A. et al., 2015, *MNRAS*, 446, 120
- Elmer E., Almaini O., Merrifield M., Hartley W. G., Maltby D. T., Lawrence A., Botti I., Hirst P., 2020, *MNRAS*, 493, 3026
- Endsley R. et al., 2024, *MNRAS*, 533, 1111
- Erőstegui A., Mezcuca M., Siudek M., Domínguez Sánchez H., Rodríguez Morales V., 2025, *A&A*, 699, A330
- Fabian A. C., 2012, *ARA&A*, 50, 455
- Finoguenov A. et al., 2007, *ApJS*, 172, 182
- Fouqué P., Solanes J. M., Sanchis T., Balkowski C., 2001, *A&A*, 375, 770
- Gardner J. P. et al., 2006, *Space Sci. Rev.*, 123, 485
- Gavazzi R., Adami C., Durret F., Cuillandre J. C., Ilbert O., Mazure A., Pelló R., Ulmer M. P., 2009, *A&A*, 498, L33
- Geha M. et al., 2017, *ApJ*, 847, 4
- George M. R. et al., 2011, *ApJ*, 742, 125
- Gozaliasl G. et al., 2014, *A&A*, 566, A140
- Gozaliasl G. et al., 2019, *MNRAS*, 483, 3545
- Greene J. E., Ho L. C., 2007, *ApJ*, 670, 92
- Habouzit M., Volonteri M., Dubois Y., 2017, *MNRAS*, 468, 3935
- Ilbert O. et al., 2006, *A&A*, 457, 841
- Ivezić Ž. et al., 2019, *ApJ*, 873, 111
- Jarrett T. H. et al., 2011, *ApJ*, 735, 112
- Juodžbalis I. et al., 2023, *MNRAS*, 525, 1353
- Kauffmann G. et al., 2003, *MNRAS*, 341, 54
- Kaviraj S., 2014, *MNRAS*, 440, 2944
- Kaviraj S. et al., 2017, *MNRAS*, 467, 4739
- Kaviraj S., Martin G., Silk J., 2019, *MNRAS*, 489, L12
- Kaviraj S., Lazar I., Watkins A. E., Laigle C., Martin G., Jackson R. A., 2025, *MNRAS*, 538, 153
- Kewley L. J., Dopita M. A., Sutherland R. S., Heisler C. A., Trevena J., 2001, *ApJ*, 556, 121
- Kewley L. J., Groves B., Kauffmann G., Heckman T., 2006, *MNRAS*, 372, 961
- Khostovan A. A. et al., 2025, preprint (arXiv:2503.00120)
- Kimura Y., Yamada T., Kokubo M., Yasuda N., Morokuma T., Nagao T., Matsuoka Y., 2020, *ApJ*, 894, 24
- King A., Nealon R., 2021, *MNRAS*, 502, L1
- Koekemoer A. M. et al., 2007, *ApJS*, 172, 196
- Kormendy J., Richstone D., 1995, *ARA&A*, 33, 581
- Koudmani S., Sijacki D., Bourne M. A., Smith M. C., 2019, *MNRAS*, 484, 2047
- Koudmani S., Sijacki D., Smith M. C., 2022, *MNRAS*, 516, 2112
- Koudmani S., Rennehan D., Somerville R. S., Hayward C. C., Anglés-Alcázar D., Orr M. E., Sands I. S., Wellons S., 2025, *MNRAS*, 540, 1928
- Kristensen M. T., Pimbblet K., Penny S., 2020, *MNRAS*, 496, 2577
- Kuijken K., 2011, *The Messenger*, 146, 8
- Laigle C. et al., 2016, *ApJS*, 224, 24
- Laigle C. et al., 2018, *MNRAS*, 474, 5437
- Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
- Lazar I., Kaviraj S., Martin G., Laigle C., Watkins A., Jackson R. A., 2023, *MNRAS*, 520, 2109
- Lazar I., Kaviraj S., Watkins A. E., Martin G., Bichang'a B., Jackson R. A., 2024a, *MNRAS*, 529, 499
- Lazar I., Kaviraj S., Watkins A. E., Martin G., Bichang'a B., Jackson R. A., 2024b, *MNRAS*, 533, 3771
- Leauthaud A. et al., 2007, *ApJS*, 172, 219
- Lintott C. et al., 2011, *MNRAS*, 410, 166
- Lotz J. M., Primack J., Madau P., 2004, *AJ*, 128, 163
- Marleau F. R., Clancy D., Habas R., Bianconi M., 2017, *A&A*, 602, A28
- Martin G. et al., 2019, *MNRAS*, 485, 796
- Massey R., Stoughton C., Leauthaud A., Rhodes J., Koekemoer A., Ellis R., Shaghoulain E., 2010, *MNRAS*, 401, 371
- McCracken H. J. et al., 2012, *A&A*, 544, A156
- Mezcua M., Domínguez Sánchez H., 2024, *MNRAS*, 528, 5252
- Mezcua M., Civano F., Marchesi S., Suh H., Fabbiano G., Volonteri M., 2018, *MNRAS*, 478, 2576
- Mezcua M., Suh H., Civano F., 2019, *MNRAS*, 488, 685
- Mezcua M., Pacucci F., Suh H., Siudek M., Natarajan P., 2024, *ApJ*, 966, L30
- Moran E. C., Shahinyan K., Sugarman H. R., Vélez D. O., Eracleous M., 2014, *AJ*, 148, 136
- Nyland K. et al., 2017, *ApJ*, 845, 50
- Papovich C. et al., 2018, *ApJ*, 854, 30
- Pardo K. et al., 2016, *ApJ*, 831, 203
- Piotrowska J. M., Bluck A. F. L., Maiolino R., Peng Y., 2022, *MNRAS*, 512, 1052
- Pucha R. et al., 2025, *ApJ*, 982, 10
- Reines A. E., Greene J. E., Geha M., 2013, *ApJ*, 775, 116
- Richstone D. et al., 1998, *Nature*, 385, A14
- Sacchi A., Bogdán Á., Chadayammuri U., Ricarte A., 2024, *ApJ*, 974, 14
- Satyapal S., Abel N. P., Secrest N. J., 2018, *ApJ*, 858, 38
- Sawicki M. et al., 2019, *MNRAS*, 489, 5202
- Schaap W. E., van de Weygaert R., 2000, *A&A*, 363, L29
- Scholtz J. et al., 2025, *A&A*, 697, A175
- Schutte Z., Reines A. E., Greene J. E., 2019, *ApJ*, 887, 245
- Sharma R. S., Brooks A. M., Tremmel M., Bellovary J., Ricarte A., Quinn T. R., 2022, *ApJ*, 936, 82
- Sharma R. S., Brooks A. M., Tremmel M., Bellovary J., Quinn T. R., 2023, *ApJ*, 957, 16
- Shimwell T. W. et al., 2017, *A&A*, 598, A104
- Silk J., Rees M. J., 1998, *A&A*, 331, L1
- Siudek M., Mezcuca M., Krywult J., 2023, *MNRAS*, 518, 724
- Siudek M. et al., 2024, *A&A*, 691, A308
- Sousbie T., 2011, *MNRAS*, 414, 350
- Steinhardt C. L. et al., 2014, *ApJ*, 791, L25
- Strateva I. et al., 2001, *AJ*, 122, 1861
- Taniguchi Y. et al., 2007, *ApJS*, 172, 9
- Taniguchi Y. et al., 2015, *PASJ*, 67, 104
- Tolstoy E., Hill V., Tosi M., 2009, *ARA&A*, 47, 371
- Trebitsch M., Volonteri M., Dubois Y., Madau P., 2018, *MNRAS*, 478, 5607
- Trujillo I. et al., 2021, *A&A*, 654, A40
- van der Marel R. P., Alves D. R., Hardy E., Suntzeff N. B., 2002, *AJ*, 124, 2639

- Venhola A. et al., 2018, *A&A*, 620, A165  
 Ward S. R., Harrison C. M., Costa T., Mainieri V., 2022, *MNRAS*, 514, 2936  
 Watkins A. E., Kaviraj S., Collins C. C., Knapen J. H., Kelvin L. S., Duc P.-A., Román J., Mihos J. C., 2024, *MNRAS*, 528, 4289  
 Weaver J. R. et al., 2022, *ApJS*, 258, 11  
 Whitaker K. E., van Dokkum P. G., Brammer G., Franx M., 2012, *ApJ*, 754, L29  
 Witten C. et al., 2025, *MNRAS*, 537, 112  
 Wright A. H. et al., 2017, *MNRAS*, 470, 283  
 Zabludoff A. et al., 2021, *Space Sci. Rev.*, 217, 54  
 Zamojski M. A. et al., 2007, *ApJS*, 172, 468  
 Zou F. et al., 2023, *ApJ*, 950, 136

## APPENDIX A: COMPARISON OF SPECTROSCOPIC AND PHOTOMETRIC REDSHIFTS

Fig. A1 shows a comparison of spectroscopic redshifts ( $z_{\text{spec}}$ ) and photometric redshifts ( $z_{\text{phot}}$ ) of 16 variable dwarfs which have reliable spectroscopic redshifts (i.e. with confidence levels greater than 90 per cent) from the recent compilation by A. A. Khostovan et al. (2025). This comprehensive catalogue contains 165 312 redshifts of 97 929 unique objects from 108 individual observing programmes. The photometric redshifts, which are derived via SED fitting, show excellent correspondence with their spectroscopic counterparts. The median difference between the photometric and spectroscopic

redshifts is  $\sim 0.0066$ . The median fractional difference, defined as  $(z_{\text{spec}} - z_{\text{phot}})/z_{\text{spec}}$  is 2.8 per cent.



**Figure A1.** Comparison of spectroscopic redshifts and photometric redshifts from COSMOS2020 for 16 variable dwarfs which have reliable spectroscopic redshifts from the compilation by A. A. Khostovan et al. (2025).

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