MNRAS 000, 1–23 (2015) Preprint 24 July 2026 Compiled using MNRAS LATEX style file v3.0

MIGHTEE: The Host-Galaxy Associated Catalogue of the Radio Sources in MIGHTEE Continuum Data Release 1

C. L. Hale,[1,2★] J. Hamlett,[2] I. H. Whittam,[2,3] M. J. Jarvis,[2,3] M. J. Hardcastle,[4] S. L. Jung,[2] N. Stylianou,[2] R. G. Varadaraj,[2] M. Vaccari,[5,3,6] L. Barchiesi,[7,5,6] P. N. Best,[1] K. K. L. Charlton,[6,8,7] E. Malefahlo,[9] E. Moravec,[10] J. P. Moss,[11] M. Rakototafika,[7,12] T. F. Rarivoarinoro,[7,5,12] F. Sinigaglia,[13,14,15,16] M. N. Tudorache,[17] S. I. Loubser,[18,19] L. Heino,[5,7,20] A. Saintonge,[21,22] A. Mazumder,[23] N. Netshiavha,[7] D. J. B. Smith,[4] L. Verdes-Montenegro,[24] J. Delhaize,[7] H. Pan,[25,26] F. Pozzi,[8,27] Z. Randriamanakoto,[12,28] L. Stockenstroom,[7] A. A. Vărăşteanu,[2] E. Vardoulaki,[29,30] W. L. Williams,[3] R. A. A. Bowler,[23] L. Marchetti,[7,6,5] A. Matthews[32]

Affiliations are listed at the end of the paper

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Radio continuum surveys provide samples of active galactic nuclei (AGN) and star forming galaxies (SFGs) to high redshifts, free of biases due to dust obscuration. However, radio detected sources require multi-wavelength counterparts to understand their intrinsic properties (e.g. redshift, stellar mass) and to study the evolution of star formation and AGN activity. In this work we present host galaxy counterparts for the MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) survey continuum Data Release 1 in regions with the best ancillary data (totalling 7.5 sq. deg). We combine statistical cross-matching and visual inspection to identify K_s-band selected host galaxies, and additionally combine multiple radio components into single physical objects, where needed. This results in a combined radio catalogue of ~66 000 sources, with host counterparts and redshifts identified for ~95 per cent of sources in the COSMOS field, ~91 per cent in XMM-LSS and ~90 per cent in CDFS-DEEP. This includes a significant fraction of sources with spectroscopic redshifts within the COSMOS field (~50 per cent), with ~30 and ~20 per cent in the XMM-LSS and CDFS-DEEP fields respectively. Using the cross-matched catalogue, we make an initial identification of radio-excess and star forming galaxies based on comparisons of the radio luminosities to host star formation rates. Using this split as a proxy for radio loud AGN or SFGs, we present expectations for the redshift distributions of these sources, finding broad agreement with those from deep radio luminosity functions and simulated catalogues.

Key words: radio continuum: galaxies, general - astronomical data bases: catalogues - galaxies: active

1 INTRODUCTION

We are in an era of radio astronomy where a combination of deep and wide area surveys from telescopes such as the Low Frequency Array (LOFAR; van Haarlem et al. 2013), MeerKAT (Jonas 2009; Jonas & MeerKAT Team 2016), the Australian Square Kilometre Array Pathfinder (ASKAP Hotan et al. 2021), the upgraded Giant Metrewave Radio Telescope (uGMRT Gupta et al. 2017) and the Karl G. Jansky Very Large Array (VLA) are unveiling a wealth of previously undetected radio sources across the Universe. Numerous surveys with these telescopes such as those presented by Jarvis et al. (2016); Smolčić et al. (2017a); Tasse et al. (2021); Sabater et al. (2021); van der Vlugt et al. (2021); de Jong et al. (2024) and Lal et al. (2025), have focussed on deep observations over areas where there is an abundance of multi-wavelength data. This allows such surveys to trace faint extragalactic radio sources over large periods of cosmic time and thus probe the evolution of radio source populations and the properties of their host galaxies. These deeper surveys are complemented by larger area surveys from LOFAR, the VLA and ASKAP (Lacy et al. 2020; Hopkins et al. 2025; Shimwell et al. 2026) which can help to trace both rare and variable radio sources as well as more diverse environments and more local sources. Importantly, extragalactic radio sources can present a unique perspective of the extragalactic sky. Unlike optical/IR emission (where stars, gas and dust dominate the emission), radio emission (at ≲GHz frequencies) predominantly traces non-thermal synchrotron emission (see e.g. Condon 1992). This results in observations that detect two key extragalactic populations: active galactic nuclei (AGN) and star

★ E-mail: Catherine.Hale@ed.ac.uk

forming galaxies (SFGs; see e.g. Mauch & Sadler 2007a; Smolčić et al. 2017b; Whittam et al. 2022; Best et al. 2023).

In AGN, relativistic electrons spiral in powerful jets and the extended structures that they generate, producing radio synchrotron emission. For radio loud AGN, there are two apparent accretion modes which play a role in their fuelling (High/Low excitation radio galaxies; see reviews in e.g. Heckman & Best 2014; Hardcastle & Croston 2020). Whilst initially such modes were believed to accrete at different Eddington rates (see e.g. Hardcastle et al. 2007; Best & Heckman 2012; Mingo et al. 2014), evidence from deep radio surveys has suggested a much broader overlap for intrinsically less luminous populations (see e.g. Whittam et al. 2018, 2022). However, recent work using spectroscopic data has again indicated a more pronounced split between the populations (see Arnaudova et al. 2025) than previous spectroscopic studies (Whittam et al. 2018). Therefore, questions still remain and it is necessary to observe radio sources where deep ancillary data are available to provide information about the properties of host galaxies. At the same time, for radio quiet AGN, the depth of recent surveys has allowed a significant increase in the numbers which have been observed, helping to advance studies into whether jets, star formation, winds or other AGN processes are dominating their emission (see e.g. Laor & Behar 2008; White et al. 2015; Chen et al. 2023; White et al. 2025; Njeri et al. 2026; Jackson et al. 2026).

For SFGs, the synchrotron emission originates from high energy cosmic electrons spiralling in the magnetic fields of supernova remnants. As such, radio emission from SFGs is known to be an excellent (delayed) tracer of star formation (see e.g. Garn et al. 2009; Davies et al. 2017; Smith et al. 2021; Matthews et al. 2021; Cochrane et al. 2023; Cook et al. 2024; Thykkathu et al. 2026), crucially without the effects of dust obscuration which can plague other wavelengths (see discussions in Madau & Dickinson 2014). This can allow star formation to be traced across large periods of the Universe's history, provided the surveys are deep enough. Using the radio emission as a star formation rate (SFR) tracer across cosmic time is one of the science goals for future deep radio surveys with telescopes such as LOFAR2.0 and the Square Kilometre Array Observatory (SKAO). However, considerations may have to be made to account for the effects of inverse Compton scattering from the cosmic microwave background (CMB), which could affect the inference of star formation rates at high redshifts (see e.g. Murphy 2009; Whittam et al. 2025).

Observing both AGN and SFGs is essential to understand how and why galaxies evolve over cosmic time. This is because both star formation and black holes are known to play crucial roles in galaxy properties and their evolution (see e.g. Benson 2010; Kormendy & Ho 2013; Heckman & Best 2014; Hardcastle & Croston 2020, for reviews), and affecting their growth through the effects of feedback and fuelling. Therefore, radio surveys are crucial to build a comprehensive picture of galaxy evolution. However, continuum radio surveys alone are unable to provide the critical source information such as the redshift (z) of a source and the stellar mass of its host galaxy (M_*) which are crucial to study how radio sources evolve (see e.g. Novak et al. 2017; Smolčić et al. 2017c; Smith et al. 2021; Kondapally et al. 2022; Cook et al. 2024; Kondapally et al. 2025). Determining the physical properties first requires the identification of host galaxies in order to user photometric and spectroscopic multi-wavelength data to measure properties of the radio sources through methods such as SED modelling (see e.g. Smolčić et al. 2017b; Whittam et al. 2022; Best et al. 2023; Das et al. 2024; Drake et al. 2024; Arnaudova et al. 2025). Whilst identifying host galaxies may sound like a simple task, there is typically a significant mis-match in the angular resolution of radio surveys to optical/near-IR surveys, with radio generally having poorer resolution (though see the high-resolution, but small area COSMOS-XS survey; van der Vlugt et al. 2021). This can lead to complexity when identifying host galaxies - especially when multiple optical galaxies may exist within the extent of the restoring beam of the radio data. Moreover, radio sources can be extended and composed of physically separated areas of radio emission. This can lead to traditional source finders (e.g. Whiting & Humphreys 2012; Mohan & Rafferty 2015) struggling to associate all the flux into a single physical source (thus inaccurately measure intrinsic properties) and may struggle to identify a radio position which is close to the host galaxy. Therefore, the analysis of numerous deep radio surveys has focused on galaxy identification and classification to maximize the science capabilities of radio data (e.g. Smolčić et al. 2017b; Algera et al. 2020; Kondapally et al. 2021; Whittam et al. 2022; Best et al. 2023; Whittam et al. 2024).

One deep radio survey which is transforming views of the radio sky is the MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) survey (Jarvis et al. 2016). As one of the large survey projects with MeerKAT, its aims span continuum, H I and polarisation science with the over-arching goal of studying galaxy evolution over large periods of cosmic time. It specifically observes four extragalactic fields, each of which has an abundance of multi wavelength data across the electromagnetic spectrum. Initially 5 sq. deg spanning the Cosmic Evolution Survey field (COSMOS, 1.5 sq. deg Scoville et al. 2007) and the XMM-Newton Large-Scale Structure field (XMM-LSS, 3.5 sq. deg) were released in an early science data release (MIGHTEE-ES; Heywood et al. 2021), with a subset of the COSMOS field (∼0.8 sq. deg) cross-matched to host galaxies (Whittam et al. 2024) and classified into AGN and SFGs through a combination of diagnostics (Whittam et al. 2022). The area of MIGHTEE was significantly increased to ∼20 sq. deg in Data Release 1 (hereafter MIGHTEE-DR1; Hale et al. 2025). This release included areas spread across three of the four deep fields which MIGHTEE will observe. Two of the fields, COSMOS (4.2 sq. deg) and XMM-LSS (14.4 sq. deg) represent the final L-band images and the third field was a singular deep pointing over the Chandra Deep Field South field (denoted here as CDFS-DEEP[1], 1.5 sq. deg). The L-band radio coverage will be extended in the future with larger coverage over CDFS and imaging of the fourth field, the European Large Area ISO Survey S1 (ELAIS-S1) field.

In this work we identify the host galaxies and redshifts for radio continuum detected sources in MIGHTEE-DR1. The paper is arranged as follows. First, in Section 2 we introduce the data used in this work, both for the initial radio source catalogues and the multi-wavelength catalogues that we cross-match to. In Section 3 we introduce the methods used to cross-match our radio sources to such host galaxies through both statistical and visual-cross matching. We present our cross-matched catalogue and associated analysis in Section 4 before discussing the results of initial investigations into the host properties of the radio sources in Section 5. Finally, we summarise and draw conclusions in Section 6. Unless otherwise stated, we assume a concordance cosmology of $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_{\rm M} = 0.3$ and $\Omega_\Lambda = 0.7$ and we assume a radio spectral index, $\alpha = 0.7$ where the flux density, S_ν, is related to frequency, ν as $S_\nu \propto \nu^{-\alpha}$. Magnitudes quoted are in the AB system (Oke & Gunn

[1] Throughout the paper we use CDFS-DEEP to refer to the MIGHTEE-DR1 coverage in the field and CDFS to refer to the larger field area with multi-wavelength coverage.

1983) and a Chabrier initial mass function (IMF; Chabrier 2003) is adopted.

2 DATA

2.1 Radio Data

The base radio catalogue for this work comes from the L-band (∼1.3 GHz) MIGHTEE-DR1 radio catalogue from Hale et al. (2025). Each observation was reduced using the oxkat pipeline (Heywood 2020), using a combination of steps to calibrate the data, accounting for both direction independent and direction dependent calibration. Imaging was performed using two Briggs' weighting values (Briggs 1995). These Briggs' weightings produced both a higher (5 arcsec) resolution image, which is typically less sensitive, as well as a lower (7-9 arcsec) resolution image, which is typically more sensitive. Whilst the CDFS-DEEP field observations were all centred on a single pointing, multiple pointings were mosaiced together for the XMM-LSS and COSMOS fields to produce images of larger contiguous regions. Each image has rms levels of $\sim 1-5$ μJy/beam, where a combination of both thermal noise and confusion noise (see e.g. Condon 1974) contributed to the background rms level, with source confusion becoming more significant for the lower resolution images. For CDFS-DEEP the median rms is ∼ 2μJy/beam in both images, this increases in the COSMOS and XMM-LSS fields to ∼ 3μJy/beam in the lower resolution images and ∼ 5μJy/beam in the higher resolution images. The full details of this, alongside the improved rms in the central regions are detailed in Table 1 of Hale et al. (2025). Alongside each continuum image, an associated effective frequency map was also released to reflect the variation in frequency observed across the field of view due to a combination of effects such as differences in the primary beam with frequency and differences in flagging per individual observation.

Source catalogues were extracted using the Python Blob Detection Source Finder (PyBDSF; Mohan & Rafferty 2015). A multi-stage approach to run PyBDSF was used to identify an initial list of sources and a residual map, with sources removed. The residual map was used to generate an rms map which was then supplied back to PyBDSF to detect sources above a peak flux detection limit of 5σ, and modelling the emission as Gaussian components. Further details of the source finding process are given in Hale et al. (2025) and the total number of sources (and Gaussian components) are presented in Table 1 of Hale et al. (2025). Hale et al. (2025) performed no source verification, therefore whilst the majority of sources are genuine, single sources, it was unclear what was the contribution of (i) physical sources split into multiple PyBDSF sources, or (ii) blended sources in the catalogue. The contribution of such sources are expected to be relatively small - the most comparable data which has similar depth, resolution and uses the same source finder is the LoTSS Deep Fields. In their work, Kondapally et al. (2021) find ∼1-2 per cent of PyBDSF sources were merged into a single sources. However, the motivation of the current paper is to produce a catalogue with radio components associated to a single object, and obtain host galaxy positions and redshifts for the combined radio catalogue, where available.

For the radio catalogue we use the high resolution (∼5 arcsec) PyBDSF source catalogues of Hale et al. (2025). This is to reduce the issues arising from source confusion, the blending of radio sources, and the chances of having multiple multi-wavelength sources within the radio extent.

2.2 Multi-wavelength Catalogues

As discussed in Section 1, the MIGHTEE fields were selected to cover four extragalactic fields which have an abundance of multi-wavelength data. The coverage and depth of this multi-wavelength data varies per field, but includes data from X-ray (e.g. Cappelluti et al. 2009; Chen et al. 2018; Iwasawa et al. 2020), UV (e.g. Martin et al. 2005; Morrissey et al. 2007), optical (e.g. Aihara et al. 2022; Vaccari et al. 2016) and IR (e.g. Sanders et al. 2007; Mauduit et al. 2012; Jarvis et al. 2013; McCracken et al. 2012; Lacy et al. 2021) wavelengths, and see compilations of this multi-wavelength data in e.g. Shirley et al. (2019). This is alongside additional radio observations available from surveys such as those of Tasse et al. (2008); Schinnerer et al. (2010); Hales et al. (2014); Smolčić et al. (2017a); Hale et al. (2019); Heywood et al. (2020) and Lal et al. (2025). For the association of host galaxies in this work we focus on K_s band selected catalogues, as this band is often used for measuring stellar mass in galaxies (see e.g. Kauffmann & Charlot 1998) and deep imaging covered large areas of the radio imaging. For this work, we adopt the K_s data from UltraVISTA (McCracken et al. 2012)[2] in the COSMOS field and from the VISTA Deep Extragalactic Observations (VIDEO; Jarvis et al. 2013) survey in the XMM-LSS and CDFS fields. Details of the near-IR catalogues can be found in Varadaraj et al. (2023). To summarise, SExtractor (Bertin & Arnouts 1996) was used to extract catalogues from the K_s band, recording both aperture and auto magnitudes. The K_s band catalogue was then used as a reference catalogue and used SExtractor in dual-image mode to obtain forced photometry across other optical and IR wavelengths.

In order to use these catalogues for statistical matching, we first needed to ensure that these catalogues are robust and have uniform detection across the field of view. Therefore, 5σ magnitude limits were applied to the catalogues, using a limiting 5σ K_s-band magnitude in 2 arcsec apertures of 24.8, 23.8 and 23.7 in the COSMOS, XMM-LSS and CDFS fields respectively. Additionally, as we require a clean sample of galaxies, we removed sources which are likely to be stars from the K_s band catalogues before matching to the radio catalogues. We adopted the stellar locus method of e.g. Jarvis et al. (2013); Hatfield et al. (2022) to identify and remove potential stars using colour cuts on the g-i and J-K_s colours, using UltraVISTA, VIDEO and Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; Aihara et al. 2022) in 2 arcsec apertures.

2.3 Additional Masking across the Field of View

Finally, we restrict cross-matching our radio sources to the regions with the best multi-wavelength data and remove regions where image quality is affected by bright stars (e.g. stellar ghosting) and other artefacts in the fields, as have been implemented in the K_s band catalogues. We applied spatial masks whose effect can be seen in Figure 1. The majority of the MIGHTEE CDFS-DEEP data is unmasked. For the COSMOS and XMM-LSS fields, instead, the MIGHTEE data covers an extended region significantly beyond

[2] We note that at the time of starting the cross-matching, this used catalogues which were derived from Data Release 5 (DR5) of UltraVISTA. The COSMOS UltraVISTA based catalogues have now been updated to include the final data release (DR6). The catalogues which we use to obtain photometric redshifts (see Section 4.3) use DR6 for COSMOS and for CDFS/XMM-LSS updated catalogues were made to allow for better mosaicing. This does not change the majority of our cross-matching, but did lead to a need to match to the updated catalogues when we associated photometric redshifts. This will be discussed in Section 4.3.

where the best multi-wavelength data exists. Therefore, the area over which we can cross-match radio data to K_s band hosts is reduced. For COSMOS, this is driven by the coverage of the UltraVISTA data and for XMM-LSS and CDFS it is driven by the combination of the VIDEO and HSC-SSP data. Combined, this results in a reduction of sources as shown in Table 1 and results in an area with both the multi-wavelength and radio coverage totalling ∼1.5 sq. deg in CDFS-DEEP, ∼1.7 sq. deg in COSMOS and ∼4.3 sq. deg in XMM-LSS.

After applying the masking and reducing contamination from stars, we proceed to cross-match to the radio sources to identify their host galaxies.

3 CROSS-MATCHING TECHNIQUES IN RADIO SURVEYS

Traditionally several methods have been adopted to identify the host galaxies of radio sources, with pros and cons to each approach. The quickest and simplest is to adopt a simple positional match between the radio and potential host galaxies, using a positional match radius which is typically comparable to (or slightly smaller) than the beam size of the radio data. This works well with compact and isolated sources, and when the radio resolution and that of the multi-wavelength imaging being matched to are comparable. However, when the resolution of radio data is poorer or where the multi-wavelength data to be cross-matched to has a much higher source density, statistical methods are more appropriate for cross-matching compact objects. One such statistical method that is commonly used is the likelihood ratio (LR; Sutherland & Saunders 1992, see Section 3.1). This uses the magnitude distribution of galaxies (and those hosting radio sources) combined with positional information to identify probabilistic host matches, and this technique has been adopted for numerous radio catalogues (see e.g. McAlpine et al. 2012; Vernstrom et al. 2016; Kondapally et al. 2021).

However, probabilistic methods typically only work successfully for compact objects, but can encounter challenges for resolved sources, such as large, jetted AGN which may have been measured as separate objects by the radio source finder. Therefore, to identify a host galaxy, all the components of the source need to be associated with a single physical object. For these sources visual inspection is often used to identify the host galaxy (see e.g. Banfield et al. 2015; Williams et al. 2019; Kondapally et al. 2021; Wong et al. 2025). However, visual inspection is a time consuming process, and so filtering the sources which should be visually inspected is useful (see e.g. Kondapally et al. 2021; Whittam et al. 2024).

For the MIGHTEE-ES data, each Gaussian component in the lower resolution image was visually inspected for a subset of the COSMOS field (covering ∼0.8 sq. deg Whittam et al. 2024). This resulted in a catalogue of cross-matched sources for ∼80 per cent of sources in that region. Additionally, the host galaxies of the sources were also associated using the likelihood ratio technique, where ∼94 per cent of radio sources were identified as having a host galaxy using the likelihood ratio technique. They also used this to assess the agreement between the visually inspected and likelihood ratio associated host galaxies. For those sources determined to have a 'good' match (i.e. LR above the relevant threshold and reaching their agreement criteria from visual inspection) in both the visually inspected and likelihood ratio catalogue, there was 95.5 per cent agreement between the host galaxies. This suggests that the likelihood ratio can be used robustly in a number of cases, but there will be sources for which visual inspection is required (see Section 3.2).

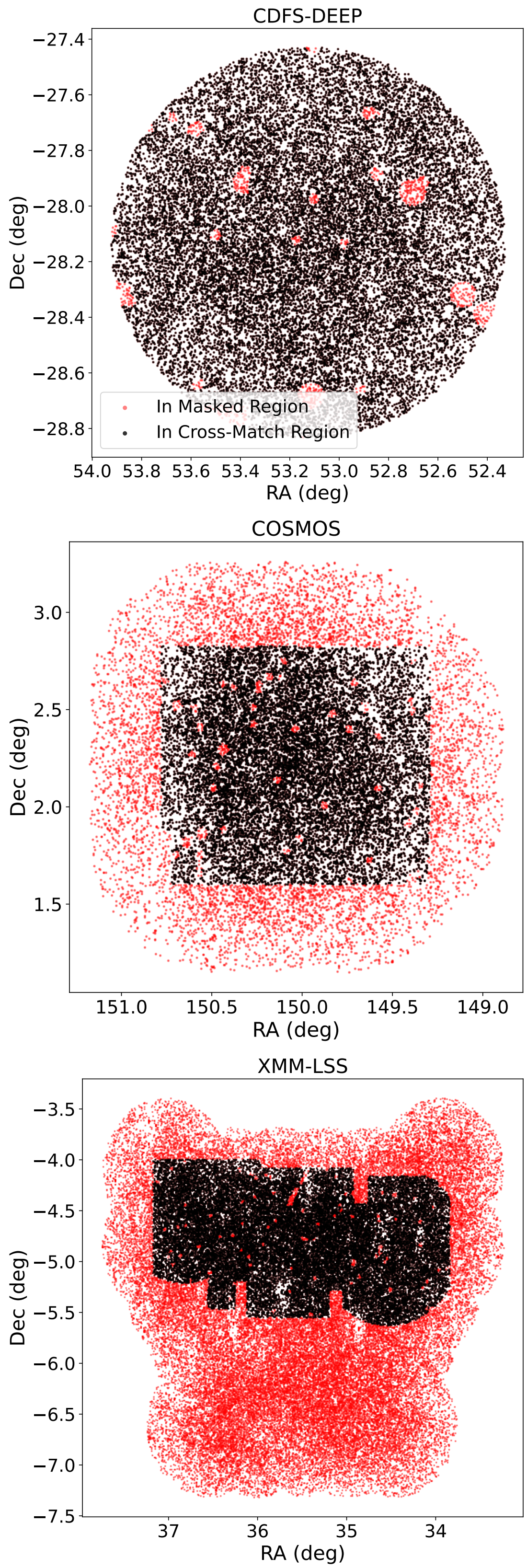

Figure 1. Full MIGHTEE-DR1 source catalogue from Hale et al. (2025) with those radio sources in masked regions (and thus not cross-matched) shown in red, and those within the multi-wavelength regions used for cross-matching shown in black.

3.1 Likelihood Ratio

The likelihood ratio (LR) describes the probability that two sources are related (in this case using the radio sources and the K_s selected catalogue as potential hosts) compared to them being unrelated. This probability takes into account both the proximity of potential host sources to the radio source as well as host galaxy properties. One common method is to consider the magnitude distribution of the potential host galaxies and the background population, which is what we adopt in this work. In this case, the likelihood ratio is given by:

$$\mathrm{LR} = \frac{q(m)f(r)}{n(m)}. \tag{1}$$

Here, $f(r)$ is the radial distribution of K_s sources around radio sources, thus taking into account host proximity. The remaining factor $(q(m)/n(m))$ takes the ratio of the magnitude (m) distribution for K_s sources which are true counterparts of the radio sources $(q(m))$ and normalises this by the magnitude distribution of the complete K_s source catalogue used $(n(m))$. To find $q(m)$ a maximum radius around the radio sources is adopted which accounts for the beam size as well as positional errors and calibration uncertainties. The details of this are available in the source code[3]. By assigning a numerical value for the LR, it is possible to select the best match for the radio source and identify a potential host K_s band galaxy.

We obtain the likelihood ratio of potential host galaxies following the method of works such as McAlpine et al. (2012) and Whittam et al. (2024), using an existing likelihood ratio matching code. The input K_s and radio catalogues have the magnitude and stellar cuts and spatial masking applied, as discussed in Section 2.3. The best K_s band match is considered suitable if the LR is above a threshold. As adopted for the MIGHTEE-ES data (Whittam et al. 2024), this threshold is calculated using the method of Williams et al. (2019) by balancing the completeness and reliability of matches. The completeness, $C(L_{\mathrm{thr}})$, and reliability, $R(L_{\mathrm{thr}})$, at a given likelihood ratio are defined as:

$$C(L_{\mathrm{thr}}) = 1 - \frac{1}{Q_0 N_{\mathrm{radio}}} \sum_{\mathrm{LR}_i < L_{\mathrm{thr}}} \frac{Q_0\,\mathrm{LR}_i}{Q_0\,\mathrm{LR}_i + (1 - Q_0)}, \tag{2}$$

and

$$R(L_{\mathrm{thr}}) = 1 - \frac{1}{Q_0 N_{\mathrm{radio}}} \sum_{\mathrm{LR}_i < L_{\mathrm{thr}}} \frac{1 - Q_0}{Q_0\,\mathrm{LR}_i + (1 - Q_0)}. \tag{3}$$

The LR threshold for determining if a host is a suitable match is taken as the value of LR where the completeness and reliability curves intersect[4]. In Equations 2 and 3 Q_0 is a normalisation fraction used to indicate the number of radio sources expected to have a match compared to the total number of radio sources (N_{radio}). It is calculated in the code adopting the estimation method of Fleuren et al. (2012) and we impose a 5 arcsec maximum radius to fit this. For the three fields, we find Q_0 values for the source catalogues of ~0.91 in CDFS-DEEP, ~0.96 in COSMOS and ~0.92 for XMM-LSS - indicating that we expect to have between 91-96 per cent of sources in the three fields to have associated host galaxies (and is similar to the results found, see Table 2). In general, for sources with a LR above the given threshold we adopt the host galaxy (but see Section 3.2), thus reducing the need for a significant number of sources to be visually inspected.

[3] from https://github.com/lmorabit/likelihood_ratio

[4] The threshold LR for the source catalogue are 0.25, 0.14 and 0.32 for the CDFS-DEEP, COSMOS and XMM-LSS fields respectively, increasing to 0.35, 0.35, and 0.47 for the Gaussian catalogues.

3.2 Likelihood Ratio or Visual inspection?

As we scale to the ~65 000 sources in the high-resolution MIGHTEE-DR1 data within the multi-wavelength regions (as shown in Figure 1), visual inspection of each source (as done in MIGHTEE-ES) becomes impractical. As the early science data suggested that the LR can be trusted in a large number of cases, we use a process of adopting the likelihood ratio match for a source where possible, and using further visual inspection when we believe that the LR may not be as robust. We construct a decision tree (similar to that of Kondapally et al. 2021) to take a series of steps to determine whether to adopt the LR match or send a source for visual inspection. This is presented in Figure 2, and to summarise the steps are:

(i) First, sources which are defined as large (having a non-deconvolved major axis, Maj≥ 2.5× the radio image's FWHM) are automatically sent for visual identification. These sources might constitute lobed AGN or nearby star forming galaxies and so are likely to need combining with other radio objects.

(ii) Secondly, we select sources which have a LR match host ID which is the same as that of another radio source; these are automatically sent for visual inspection.

(iii) Next, we take two approaches based on whether a source is identified as a single Gaussian component source by `PyBDSF` (S_code='S') or not. For sources identified as a single component source and where the LR is above the threshold then this host galaxy is adopted. If not, the radio source is instead sent for visual inspection.

(iv) For those sources which have multiple Gaussian components we consider the LR of all Gaussian components within the source (using a second LR catalogue which used the Gaussian catalogue instead of the radio source catalogue as input). If the LR of the source is greater than the LR of any of the individual Gaussian components and the source LR is greater than the associated LR threshold, then the LR match for the overall source is adopted and the source is not sent for visual inspection. However, if the LR of the source is below the threshold, the source is sent for visual inspection.

(v) Finally, for those sources which have multiple Gaussian components where the LR of all Gaussian components are greater than the LR of the source then that might indicate a blended source. Therefore, if the LR of all the Gaussian components are greater than the respective threshold then the sources are de-blended into their Gaussian components and the LR host is adopted for each of these. Else, the source is sent for visual inspection.

Combined, these steps allow us to identify which sources should be sent for visual inspection through the 'MIGHTEE Zoo' (see Section 3.3). The number of sources which are sent to MIGHTEE Zoo for

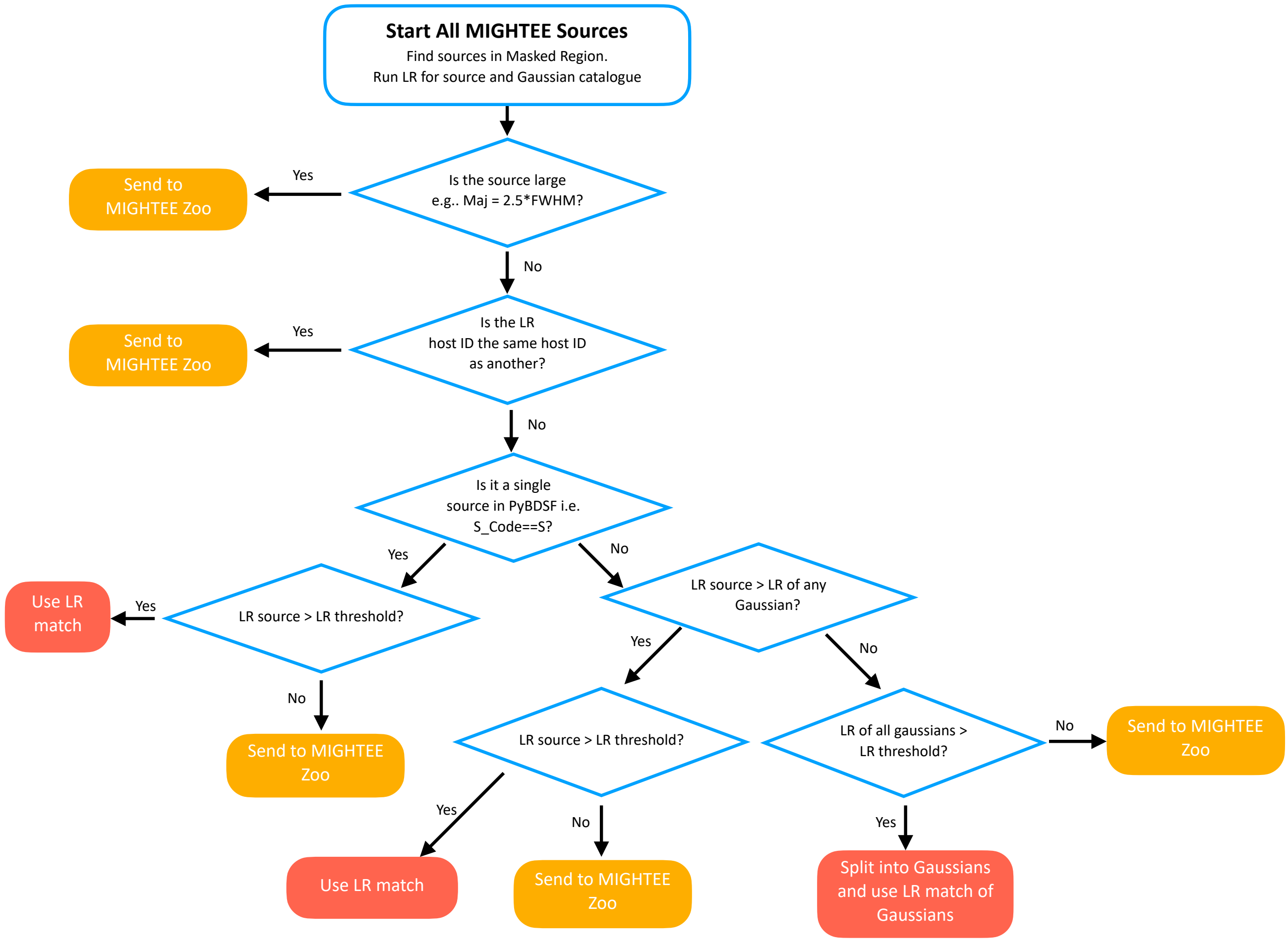

Figure 2. Decision tree flow chart used to determine whether the likelihood ratio matched source is suitable to associate a host galaxy from the near-IR or if the source should be sent to MIGHTEE zoo for visual inspection, see Section 3.2.

each field[5] in each step outlined above (and the total numbers across all steps) are given in Table 1.

3.3 MIGHTEE Zoo

In order to visually cross-match the sources which could not be robustly assigned a host using the likelihood ratio method, we follow the method used by the LOFAR surveys collaboration (see Williams et al. 2019; Kondapally et al. 2021; Hardcastle et al. 2023) and set up a Zooniverse[6] project (see e.g. Lintott et al. 2012) internal to the MIGHTEE consortium (see the process described in Hardcastle et al. 2023). Each source sent to the Zooniverse was visually inspected by at least 5 members of the MIGHTEE collaboration, where they were asked a series of questions. These questions were used to (1) identify additional `PyBDSF` sources which are associated with the same galaxy; (2) if possible, identify a host galaxy; (3) flag the source if necessary to identify cases where e.g. the object was an artefact or the image was too zoomed in to see the full source extent and (4) provide a text box to (optionally) describe the source. Example images which were shown in the zoo are presented for three sources in Figure 3. Four images are created per source: two use i-band images from HSC-SSP as the background and the others use 3.6 μm images from either the COSMOS Spitzer survey (S-COSMOS; Sanders et al. 2007) or the Spitzer Extragalactic Representative Volume Survey (SERVS; Mauduit et al. 2012).

As outlined in Table 1, a total of ∼10 000 sources required visual inspection, resulting in a need for ∼50 000 inspections in order for each source to have at least 5 responses from different people before it could be retired from the zoo. Using the criteria adopted in Williams et al. (2019), a match would be considered good if > 3/5 of the classifications from individuals were in agreement. A total of ∼70 volunteers from the MIGHTEE consortium across ∼10 countries registered to assist with the cross-matching process. Of this, ∼45 people eyeballed over 100 sources, and ∼20 people inspected over 1 000 sources each. The visual identification process was first run only on the COSMOS field in 2023. This allowed core members of the MIGHTEE cross-matching team to investigate results and to identify potential challenges which were encountered by the team or improvements to the guidance which could be given. The larger

[5] We note an update to the LR code which was performed after the COSMOS MIGHTEE zoo was started and prior to the other fields. In the table we include the numbers of source from the flowchart of the updated code, which was predominately adopted to classify sources, but include in brackets the number of sources which were sent to the zoo in reality. This was more sources than necessary from the flow chart, but as these were visually inspected, it should not affect the results.
[6] https://www.zooniverse.org

Table 1. Details of the number of radio sources used throughout the analysis after applying different steps of the cross-matching process. Included are the number of sources (N_s) or Gaussians (N_g) across the full MIGHTEE-DR1 images and over the restricted masked regions indicated in Figure 1 (in red). Also included are details of the number of sources which should be sent to MIGHTEE zoo or for which the likelihood ratio cross-matching could be used, using the different criteria from the flowchart in Figure 2. For the COSMOS field, see the note at the end of Section 3.2.

	CDFS-DEEP	COSMOS	XMM-LSS	Total
N_s Full Radio Image	21 152	20 886	72 187	114 225
N_g Full Radio Image	22 660	22 420	76 567	121 647
N_s Multi-wavelength region	20 335	13 781	31 491	65 607
N_g Multi-wavelength region	21 757	14 581	33 247	69 585
Large N to zoo	1 870	1 380	2 891	6 141
Sources with same Multi-wavelength ID	135	15	198	353
Single Sources to zoo	1 057	238	1 530	3 120
Single Sources use LR	13 360	10 395	23 490	46 942
Multi-component Sources use LR	3 301	1 508	2 899	6 623
Multi-component Sources use Gaussians LR	278	148	164	1 585
Multi-component Sources to zoo	334	102	319	843
Total sources to use LR	16 661	11 902	26 389	53 565
Total sources to use Gaussians	278	148	164	1 585
Total sources to MIGHTEE Zoo	3 396	1731 (2 123)	4 938	10 457

cross-matching effort over the CDFS-DEEP and XMM-LSS fields was then conducted in 2024.

4 COMBINED CROSS-MATCHED CATALOGUE

4.1 Initial combined catalogue from LR and Zoo

To obtain our cross matched catalogue across each of the three fields we combined the results from the LR with those from visual classification in the zoo. Sources which were selected by the decision tree to use the LR criterion (either as sources or Gaussians) were included in the final catalogue. Next, sources which underwent visual inspection needed to be combined into the catalogue. Following the work of Williams et al. (2019) and Kondapally et al. (2021), those sources in which 60 per cent of inspectors agreed on the host galaxy (i.e. 3 out of the 5), as well as the association of any `PyBDSF` sources into the same object, were combined together. Sources in which there was no consensus of a host galaxy from the zoo (including those where there was agreement that multiple `PyBDSF` sources were combined into a single object) were still included in the final catalogue with an updated radio position but without a corresponding host RA/Dec. For those sources in which multiple `PyBDSF` sources were combined to create a new radio source, new source properties were assigned, namely a flux weighted host position, a new total flux (summed from the individual `PyBDSF` sources) and size (see Williams et al. 2019).

We note that for those sources where a consensus host galaxy was assigned in the zoo, ~99 per cent of these had >80 per cent agreement from visual inspectors and ~77 per cent, ~78 per cent and ~ 65 per cent had 100 per cent agreement on the host in CDFS-DEEP, COSMOS and XMM-LSS respectively. Additionally, for those sources where `PyBDSF` source association was performed in the zoo $\gtrsim$ 70 per cent of these had at least 4/5 visual inspectors agree, and ~ 15 – 20 per cent had all five inspectors agree (though see additional classifications in Section 4.2). These suggest the associations are broadly robust and any mis-classifications from the zoo would be a small fraction of the full catalogue.

4.2 Additional Checks on the Catalogues

Finally, additional checks were performed on this merged catalogue to ensure the quality of the catalogue. The checks performed identified a number of corrections which needed to be made to finalise the catalogue. This included identifying:

(i) Sources which were in the original LR catalogue, but had also been incorporated within the zoo catalogue and thus had been left as duplicate sources within the catalogues

(ii) Sources flagged as too zoomed in by ≥60 per cent of inspectors within the Zoo framework (i.e. 'Zoom_prob'≥0.6, as described in Appendix A).

(iii) Sources where the host galaxy assigned was the same as another within the catalogue

(iv) Sources where either the radio positions or host galaxy positions were close to one another (≤1 arcsec)

(v) Sources which consisted of a component which was included in multiple output zoo sources

Of these criteria, the last two typically represented sources which were a combination of (i) sources which were split into Gaussian components, but where multiple of these Gaussians were assigned the same host galaxy or (ii) sources which had been output from the zoo but where there appeared to be some disagreement between classifiers on how many initial radio `PyBDSF` sources to include within the full extent of the source.

Sources which met any of these five criteria were additionally eyeballed by the lead author in order to make a decision on the source (i.e. which `PyBDSF` sources to combine, if sources should be merged and if a host galaxy should be assigned). This led to ~ 50−300 sources being investigated per field. Where any changes to the sources in the catalogues were required, these were applied by either removing or merging sources in the catalogue. These sources were flagged within the catalogue through the 'match_method' column, see Appendix A.

The flux distribution of all sources in this final catalogue is shown in Figures 4 and 5 for those sources which are within the multi-wavelength regions. In Figure 4, the flux density distribution of those sources which were combined together from the initial `PyBDSF` catalogue either through the Zoo or through the associations outlined in this section are presented, and the number of components which

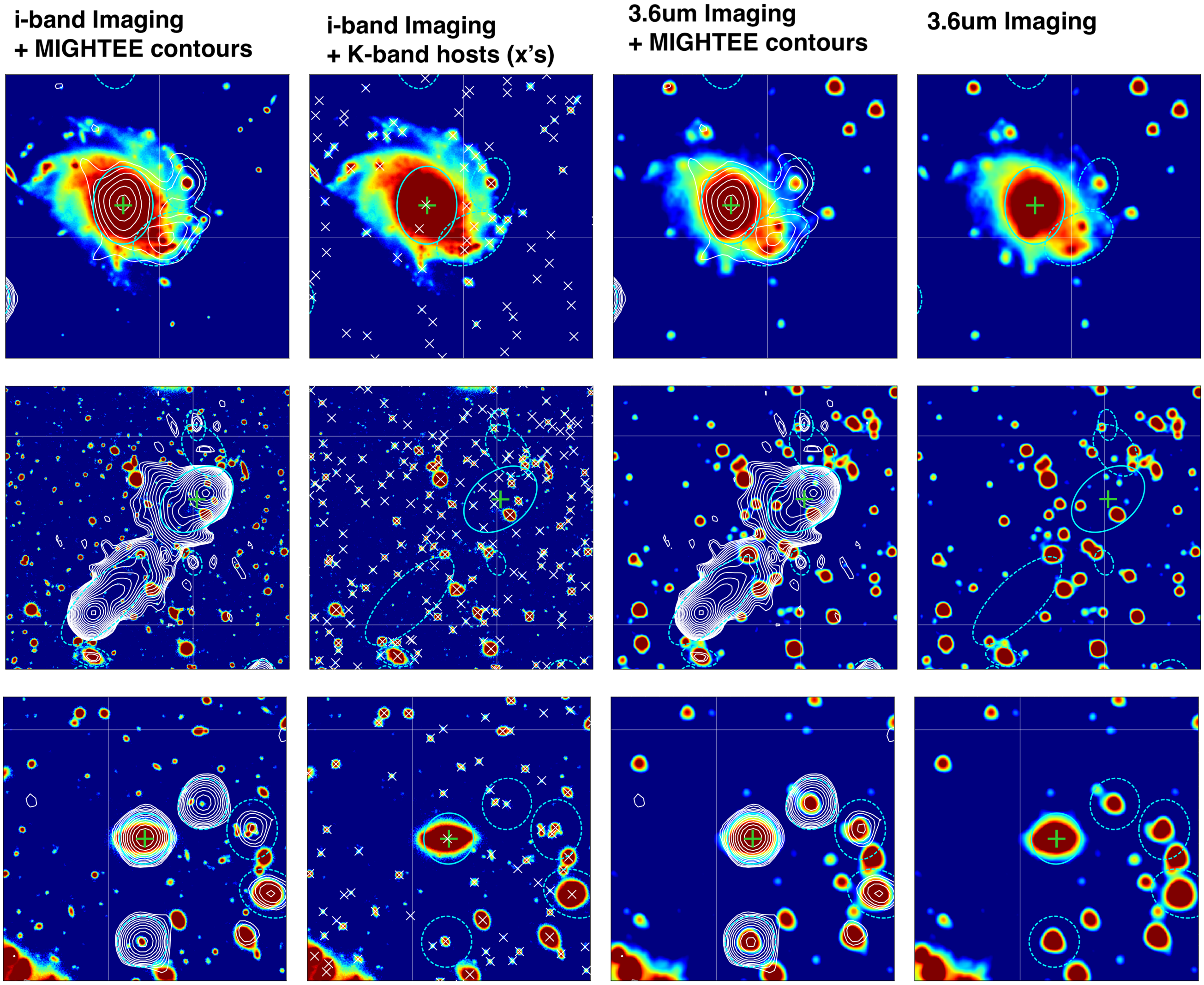

Figure 3. Images for three sources (upper: a nearby spiral galaxy; middle: a jetted AGN; lower: a compact radio source) which were used in the galaxy zoo platform to help identify host galaxies and associate radio components. Shown are the images available to MIGHTEE zoo participants, namely: i-band imaging background with white radio contour overlays (left); i-band imaging background with white crosses indicating the location of potential K-band selected host galaxies (centre left); 3.6 µm imaging from Spitzer as the background with white radio contour overlaid (centre right); and 3.6 µm imaging from *Spitzer* as the background only. On each panel a solid cyan ellipse is used to indicate the radio source to be cross-matched and additional radio sources identified by `PyBDSF` are indicated by cyan dashed ellipses.

were combined together to make these sources is presented. In Figure 5, the flux density distribution is instead shown for all sources and shown separately as those sources with (a) a host assigned and (b) sub-categorised based on the method used to assign a host.

We present the distribution of the positional offsets between the radio positions and the host galaxy positions in Figure 6 per field. For each field, a non-zero mean offset in both RA and Dec is found. These mean offsets are typically $\lesssim$0.2 arcsec, except for the XMM-LSS field, where the absolute mean declination offset is closer of 0.4 arcsec. However, all mean offsets in both RA and Dec are smaller than the pixel size of the MIGHTEE images (1.1 arcsec), and thus are within expected values.

We additionally present the comparison of the fraction of sources which have a MIGHTEE source as a function of K_s band magnitude in Figure 7, in a similar vein to Fig. 7 of Kondapally et al. (2021). This presents both (i) the fraction of the MIGHTEE radio sources within the multi-wavelength regions and (ii) the fraction of the full K_s catalogue of sources which overlap with the radio data. We present this both for each field individually, alongside the total fraction across the three fields. Figure 7 demonstrates that the CDFS-DEEP field detects fainter radio sources than the other two fields, with the fraction of MIGHTEE sources peaking at fainter K_s band magnitudes than the other two fields. The K_s band sources which have the largest fraction of MIGHTEE counterparts have a magnitude of ∼19-20, with between 25-40 per cent of K_s sources in this magnitude bin containing a radio source. However, for the radio population themselves, these peak closer to fainter K_s band magnitudes at ∼21.

4.3 Redshifts

As outlined in Section 1, redshifts are imperative to studies of galaxies in order to accurately place galaxies within the context of cosmic history. In this section we present the combination of photometric and spectroscopic redshifts for the radio sources.

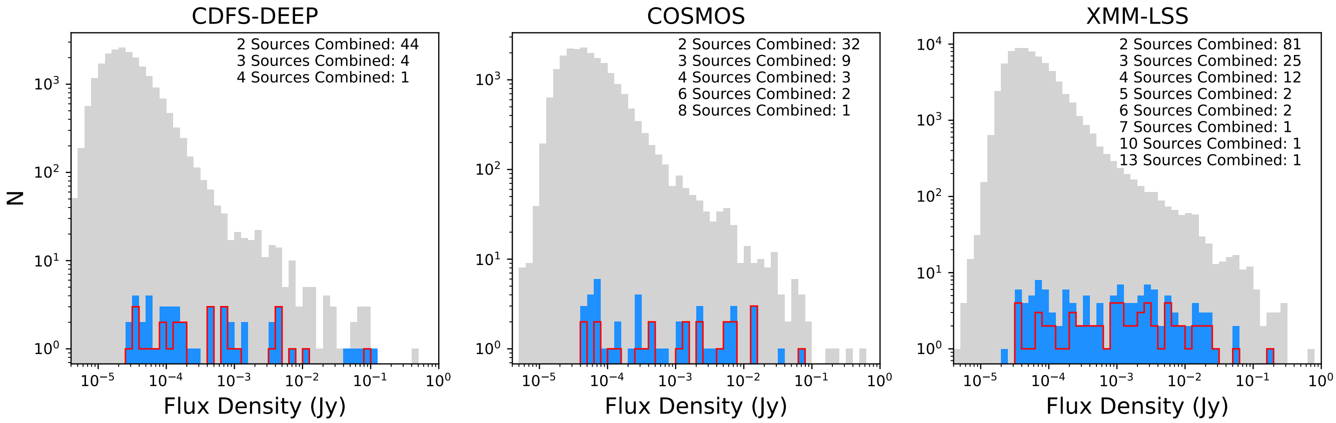

Figure 4. Flux density distribution of sources which were merged together (blue) compared to the final catalogue of sources within the multi-wavelength regions (light grey) for the CDFS-DEEP, COSMOS and XMM-LSS fields (left to right respectively). Also shown are sources which were checked in the additional steps of Section 4.2 (red) and the number of input `PyBDSF` sources which were combined together into individual output sources from the zoo and given in the top right of each figure.

4.3.1 *Photometric Redshifts*

To assign photometric redshifts to the combined radio source catalogue we use the catalogues adopted in Stylianou et al. (in prep). As discussed, at the time of commencing the cross-matching process, a previous version of K_s catalogues were used to assign host galaxies. In COSMOS, the UltraVISTA data were updated to DR6[7] in the catalogues of Stylianou et al. (in prep) and in the XMM-LSS and CDFS fields, the catalogues were also updated between the source cross-matching and redshift association process. For all fields, there were also small differences in the masking which were applied to these catalogues, this is recorded in the catalogue, see Appendix A. Therefore, we match the host galaxy positions to the positions in the updated catalogues of Stylianou et al. using a 1 arcsec match radius. These updated source positions are provided in the catalogues released with this work, as outlined in Appendix A.

For the catalogues of Stylianou et al. (in prep), there are two origins for the photometric redshifts, depending on the field. In the COSMOS and XMM-LSS fields, the redshifts are predominantly from the compilation of photometric redshifts from Hatfield et al. (2022) based on 10 photometric bands: four VISTA filters (from McCracken et al. 2012; Jarvis et al. 2013), five HSC filters and the u-band filter from the Canada-France-Hawaii Telescope Large Area U-band Deep Survey, CLAUDS (Sawicki et al. 2019). Hatfield et al. (2022) uses a Hierarchical Bayesian model (see Duncan et al. 2018, for methodology) to combine photometric redshift estimates from template fitting using the code `LePHARE` (see Ilbert et al. 2006; Arnouts & Ilbert 2011) with machine learning techniques from the code GPz (Almosallam et al. 2016a,b). In the CDFS-DEEP field, however, the catalogues of Hatfield et al. (2022) were not available and so Stylianou et al. (in prep) carried out template fitting of the source fluxes using the code `LePHARE`. For each source, the photometric redshifts were obtained by matching the observed photometry with a set of redshifted galaxy SED templates (with and without the incusion of AGN) and selecting the best-fitting redshift through χ^2 minimisation. The template set adopted here was the COSMOS library included in `LePHARE`, which contains 32 SEDs from the Ilbert et al. (2009) COSMOS photo-z work. The templates were allowed to vary with dust attenuation using the Calzetti et al. (2000) extinction law, and the fitting used only the available photometric bands for each galaxy, with a minimum flux-error floor of 5 per cent. `LePHARE` was additionally run for those sources in the COSMOS and XMM-LSS fields for sources which were not in the catalogues of Hatfield et al. (2022). For sources with `LePHARE` only templates we select either the galaxy only best-fitting redshift or the redshifts where AGN templates were also included, depending on which redshift minimised the χ^2 value.

For each of the redshifts available, an associated probability distribution function (pdf) was available and therefore we record several properties of the pdf for the associated radio source to quantify the uncertainties in the redshift distribution. These properties are: the peak of the photometric redshift, the 1σ confidence interval around the peak redshift, and the values of the 50th, 16th and 84th percentiles of the pdf. Additionally, for the COSMOS field, we match to sources within the Physics of the Accelerating Universe (PAU) survey catalogue (Alarcon et al. 2021a). If available, the PAU redshifts were adopted to supersede those from Hatfield et al. (2022). In total, 18 380 sources in the CDFS-DEEP field had photometric redshifts available (corresponding to ~90 per cent of sources in the masked region), increasing to 91 per cent (28 493 sources) in the XMM-LSS field and 95 per cent (13 117 sources) in the COSMOS field.

As discussed in Hatfield et al. (2022), the combined method of template fitting and machine learning can help improve the accuracy of the redshifts, and so this may lead to more biases within the CDFS redshifts. In general, with photometric redshifts, a lack of availability of spectroscopic redshifts and samples which cover the broad distribution of galaxies (especially at high z) will affect the accuracy of photometric redshifts. Such photometric redshifts will be improved in these fields in the future thanks to additional deep ancillary data provided by the Large Synoptic Survey Telescope (LSST; Ivezić et al. 2019) and Euclid (Euclid Collaboration et al. 2025).

4.3.2 *Spectroscopic Redshifts*

Due to the fact that these fields are well studied across the electromagnetic spectrum, there also exist numerous spectroscopic surveys which provide more accurate redshifts across the fields. We use a

[7] https://doi.org/10.18727/archive/52

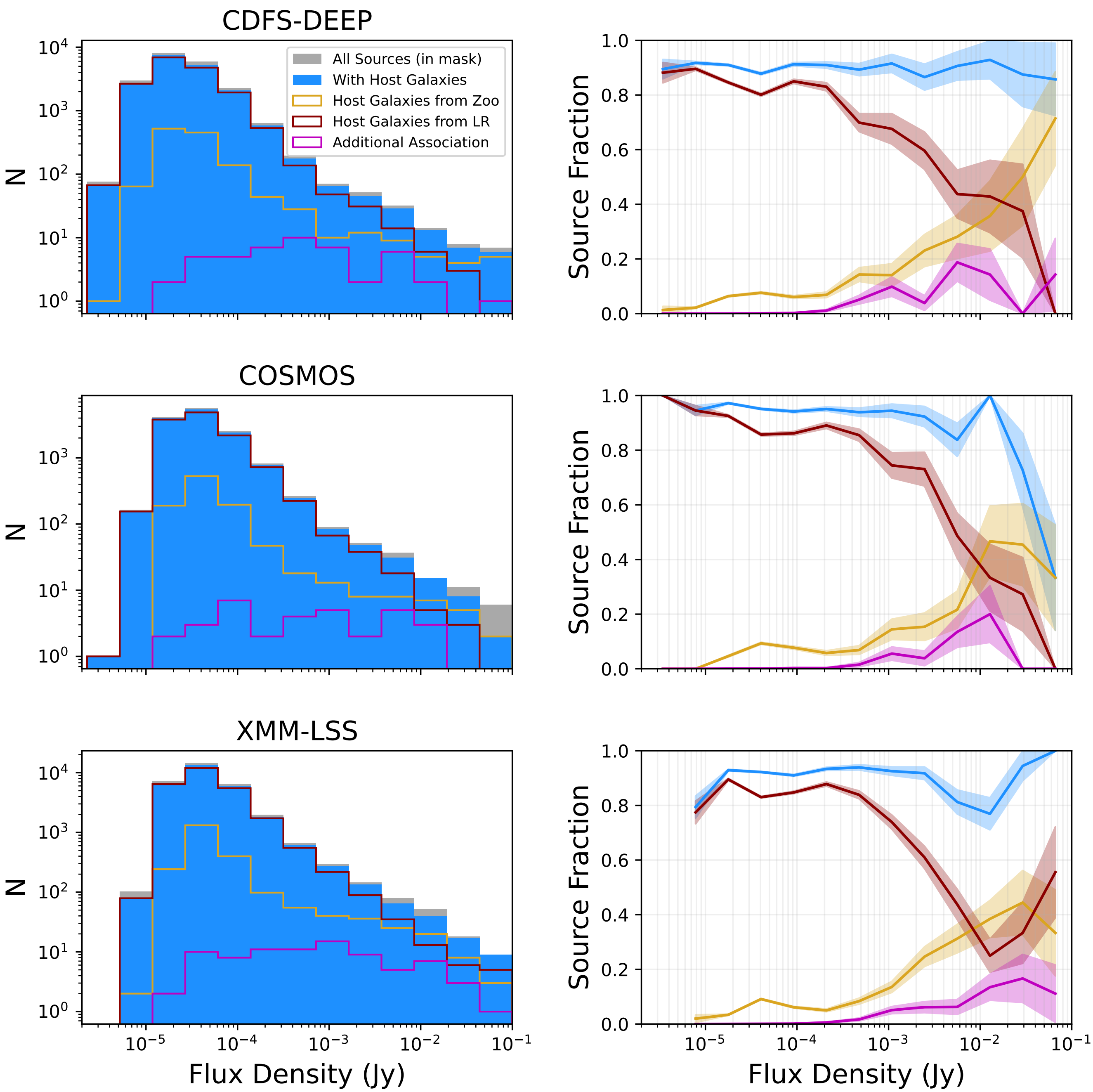

Figure 5. Upper: the flux density distribution of sources in the CDFS-DEEP, COSMOS and XMM-LSS fields (left to right respectively) for all sources which are within the unmasked, multi-wavelength regions (grey) and split into those which have a host galaxy associated with the source (blue) and then subset into those sources which have host galaxies associated from: the zoo (gold), the LR method (red), and the additional steps of Section 4.2 (pick). Lower: the total fraction of sources in the masked region which have host galaxies associated, and the fraction which come from the zoo and LR method, using the same colours as for the histograms. The associated errors provided are from the binomial errors.

compilation of redshifts within the three deep fields[8] (see e.g. Vaccari 2022, 2026) to identify spectroscopic counterparts for our radio sources, using a 1 arcsec match radius. These spectroscopic redshifts are a compilation from a variety of spectroscopic catalogues including (but not limited to) those from Dark Energy Spectroscopic Instrument survey (DESI DESI Collaboration et al. 2025), Galaxy And Mass Assembly survey (GAMA; Driver et al. 2022) and Sloan Digital Sky Survey (SDSS; Ahumada et al. 2020) which are outlined in Appendix A. If a spectroscopic redshift was available, this was adopted as the best redshift for the source.

Across the three fields the spectroscopic coverage varies, leading to a total of 4 393 sources in with spectroscopic redshifts in CDFS-DEEP, 7 256 sources in COSMOS and 9 871 in XMM-LSS. This

[8] Using the version from 20th March 2025

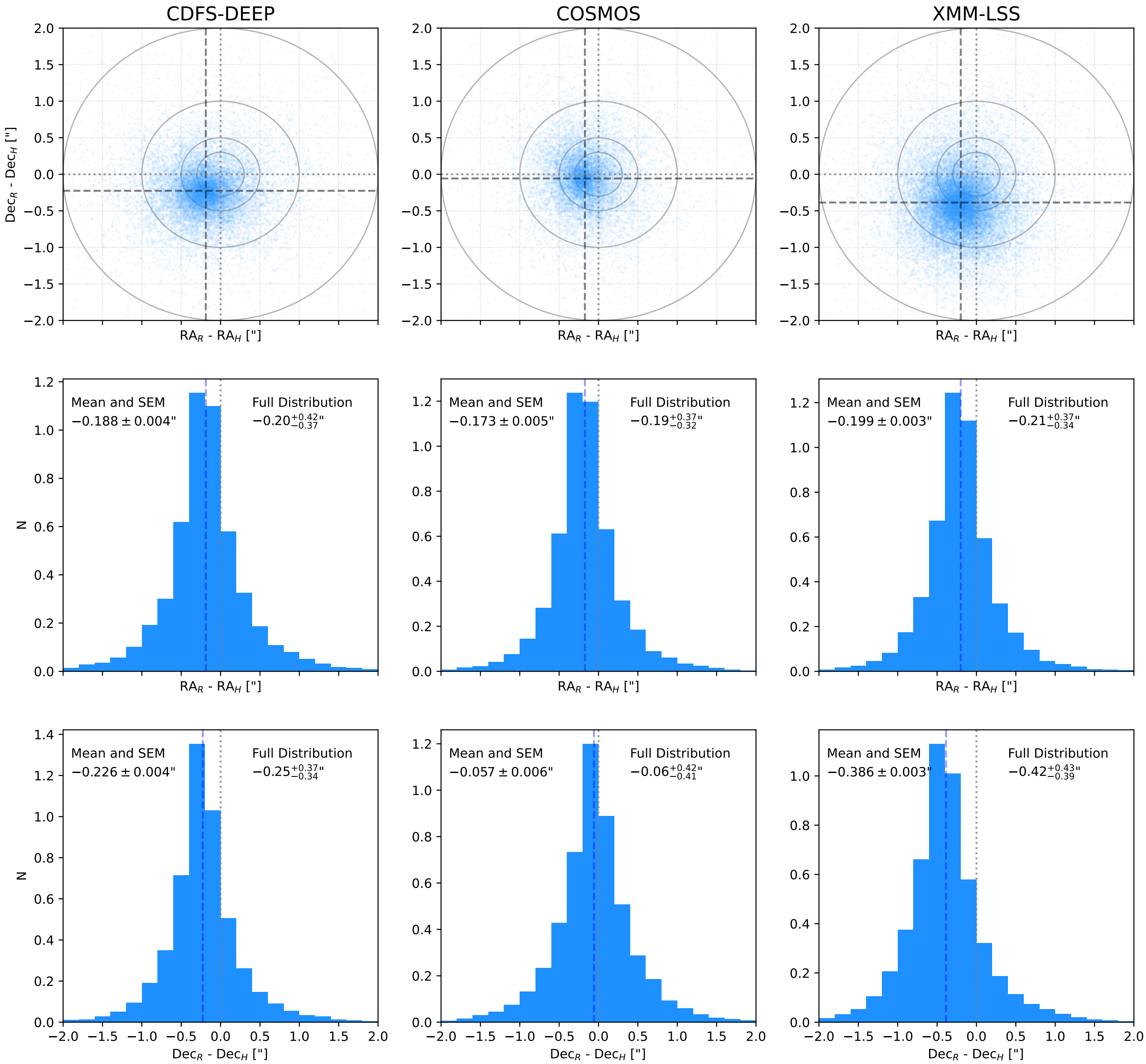

Figure 6. Comparisons of the positions of the radio sources to those of the host multi-wavelength galaxies (where available) from the 'RA_host' and 'Dec_host' columns (see Appendix A). Shown are the positional offsets for the CDFS-DEEP, COSMOS and XMM-LSS fields (left to right) shown as: (i) scatter of the offsets (top row), (ii) histogram of the RA offset (middle row) and (iii) histogram of the Dec offset (bottom row). Each offset is defined as the difference in the RA or Dec values (radio position - host position), given in arcsec. The mean offset and standard error on the mean (SEM) is recorded in the the lower two rows (top left) as is the median and errors from the 16th and 84th percentiles (top right). Angular radii of 0.3 arcsec, 0.5 arcsec, 1 arcsec and 2 arcsec are shown in the top panels. The mean offset is shown as a black dashed line, as well as a grey dotted line to indicate no positional offset (for reference only). We note that a small fraction of sources ($\lesssim$1 per cent per field) have positional offsets larger than the 2 arcsec limits shown.

corresponds to 21.5 per cent, 52.6 per cent and 31.4 per cent of the radio catalogues within the unmasked regions of the CDFS-DEEP, COSMOS and XMM-LSS fields. A comparison of the spectroscopic redshifts to the corresponding photometric redshift (taken as the peak of the pdf) is presented in Figure 8. We also provide values for two quantities which are commonly used to assess the quality of the photometric redshifts: the normalised median absolute deviation ($\sigma_{\rm NMAD}$) and outlier fraction (OLF). $\sigma_{\rm NMAD}$ is given by:

$$\sigma_{\rm NMAD} = 1.48 \times \text{median}\left(\frac{|\Delta z|}{1+z_s}\right) \tag{4}$$

see Hoaglin et al. (1983). Here Δz is the difference between the spectroscopic (z_s) and photometric (z_p) redshifts and $\sigma_{\rm NMAD}$ quantifies the spread in the data between the spectroscopic and photometric redshifts. The OLF is instead the fraction of sources which satisfy:

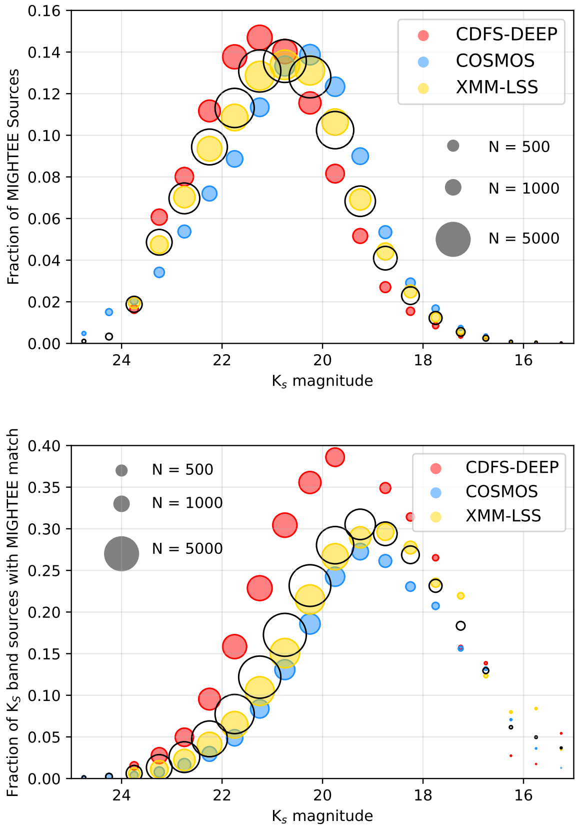

Figure 7. Comparison of the K_s band magnitude distribution of sources which are hosting one of the MIGHTEE sources. *Upper:* Magnitude distribution of our radio sources which have a K_s band magnitude as a fraction of all MIGHTEE sources which have a host galaxy assigned. *Lower:* The fraction of all K_s band sources with an associated MIGHTEE source. Shown in each panel are the CDFS-DEEP (red), COSMOS (blue) and XMM-LSS (yellow) fields and their size indicates the number of radio sources contributing (see the grey circles for an indication of numbers). The black open circles represent the combined values across the three fields.

$$|\Delta z| > 0.15(1 + z_s), \quad (5)$$

see e.g. Dahlen et al. (2013); Duncan et al. (2021) and Hatfield et al. (2022) and the OLF quantifies the fraction of sources with significant deviations (of 15 per cent) between the photometric and spectroscopic redshifts.

Of the fields, CDFS-DEEP has the highest OLF at 0.20 and $\sigma_{\rm NMAD}$ of 0.10, with the COSMOS and XMM-LSS fields having lower OLFs (0.07 and 0.09 respectively) and $\sigma_{\rm NMAD}$ (0.01 and 0.04 respectively). Given that the redshifts for the CDFS-DEEP field are from template fitting alone, whereas the other fields use machine learning techniques to improve the redshifts, this larger OLF and $\sigma_{\rm NMAD}$ are expected and this can also explain the small offset in redshifts from the 1-to-1 line (which is more apparent due to the log scaling on Figure 8). Furthermore, whilst the values of the OLFs in the COSMOS and XMM-LSS fields are larger than for the full population of galaxies found in Hatfield et al. (2022) (closer to $\lesssim$4 per cent), this is expected given the challenges of accurately measuring photometric redshifts from SED fitting when there is a significant population of AGN in the sample (see e.g. discussion in Duncan et al. 2021). We note that in the CDFS-DEEP field, the spectroscopic redshifts appear systematically higher than the photometric redshifts, with an apparent excess of sources in the CDFS-DEEP field with $z_{\rm spec}$~0.6-0.7 which have lower photometric redshifts (see Figure 8). Users of the catalogues should be aware of this.

In total, ~90 per cent of sources in the masked CDFS-DEEP field, ~91 per cent in XMM-LSS field and ~95 per cent in COSMOS have a corresponding redshift associated with them (either spectroscopic or photometric). We present the distribution of the best redshift associated with each source (where available) in Figure 9, separated into those with spectroscopic or photometric redshifts. This demonstrates that the majority of sources for $z \lesssim 1$ in the COSMOS field have spectroscopic redshifts. The COSMOS field has a cumulative fraction of > 80 per cent of sources with spectroscopic redshifts below z ~1, whereas the cumulative fraction of sources with spectroscopic redshifts below z ~1 is closer to ~55 per cent and ~35 per cent in the XMM-LSS and CDFS-DEEP fields respectively. The fractions of sources with spectroscopic redshifts will increase with the commencement of spectroscopic surveys with telescopes such as the 4-m Multi-Object Spectroscopic Telescope (4MOST; de Jong et al. 2019) and the Multi-Object Optical and Near-infrared Spectrograph (MOONS; Cirasuolo et al. 2020) including, most notably, the Optical Radio Continuum and HI Deep Spectroscopic Survey (ORCHIDSS; Duncan et al. 2023) which will specifically follow up MIGHTEE continuum sources with spectroscopy as part of the Wide-Area VISTA Extragalactic Survey (WAVES; Driver et al. 2019). We note that in both Figures 8 and 9 we plot only redshifts up to z = 5. We note that a small number of sources may have photometric redshifts recorded at z > 5. Although some have been confirmed with spectroscopy (e.g. Varadaraj et al. 2026), we urge caution and note that these sources often have a second lower redshift peak in the pdf.

A summary of the number of sources with redshifts can be found in Table 2.

4.4 Final catalogue of radio sources

In the final catalogue we combine together those sources from within the multi-wavelength regions (regardless of whether a host galaxy was identified) along with the sources which were in the masked regions so that no host was ever sought either through the LR technique or the MIGHTEE zoo[9]. A description of the added flags can be found in Appendix A, where we outline the columns within the final catalogue which is released.

5 HOST GALAXY PROPERTIES

Having constructed our radio catalogues with matched host galaxy information we investigate the properties of the radio sources, including the luminosity distribution (Section 5.1) and the additional host galaxy properties from Stylianou et al. (in prep.). Specifically we explore the host galaxy stellar mass (M_*) and SFR distributions for our sample and make initial investigations of the redshift distribution for AGN and SFG analogues selected through whether they appear to have a radio excess in 1.4 GHz luminosity compared to expectations from the star formation rate (Section 5.2).

[9] We note that for a few sources split into Gaussians, their individual flux density could now be below the nominal 5σ individual detection limit, so we urge users to be aware of this if using these sources.

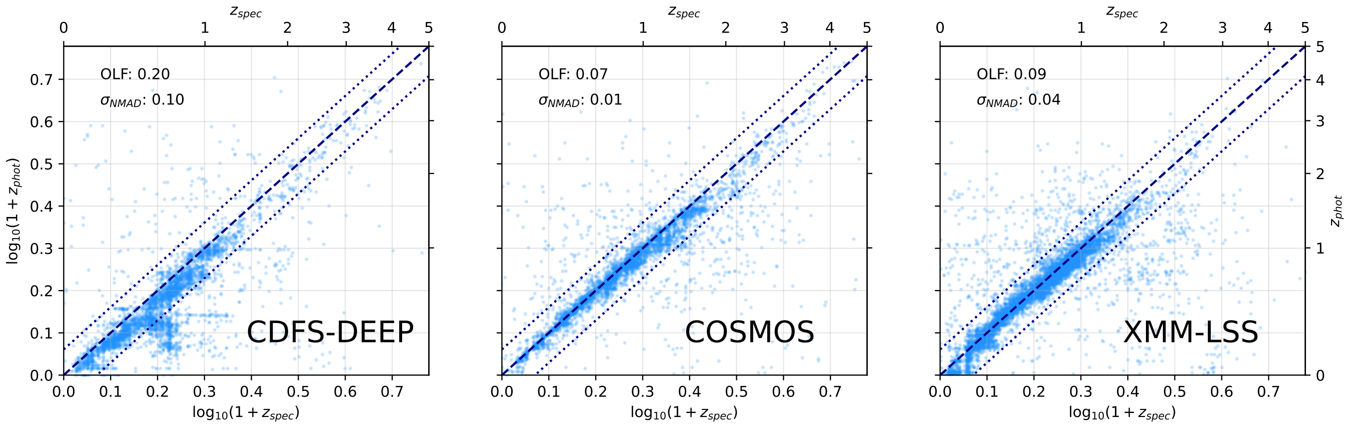

Figure 8. Comparison of the spectroscopic redshifts (x-axis) to the corresponding photometric redshift (y-axis, given as the peak in the redshift pdf) for sources in the CDFS-DEEP (left), COSMOS (centre) and XMM-LSS (right) fields. Also shown to guide the eye are the 1-to-1 line (dashed) and dotted lines which indicate the regime for the outlier fraction (OLF), using Equation 5 (dotted).

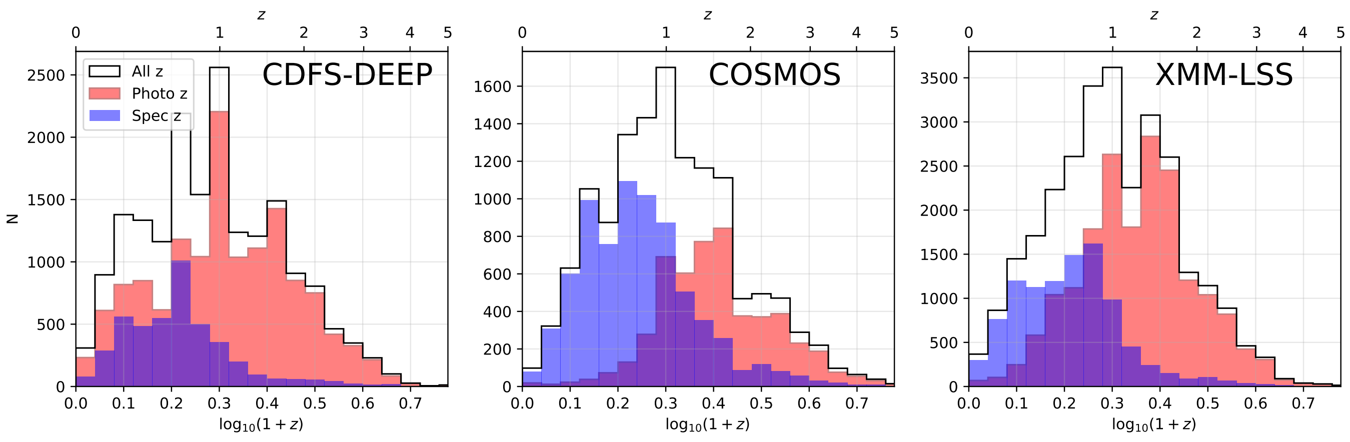

Figure 9. Histogram of the best redshifts for all available sources split into the contributions for photometric redshifts (red) and spectroscopic redshifts (blue) for the CDFS-DEEP (left), COSMOS (centre) and XMM-LSS (right) fields. The combined distribution from the photometric and spectroscopic redshifts are shown in the black outline.

5.1 Radio Luminosity Distributions

First, we present a comparison of the luminosity distribution of the cross-matched sources in Figure 10 both as a function of redshift, as well as the full distribution of luminosities from the three fields - both with a comparison to the MIGHTEE-ES data. Combined, we have a factor of ~11.6 more MIGHTEE sources with host galaxy and redshift associations than in the work of Whittam et al. (2024) (a factor of ~2.5 more sources in the COSMOS field). This is predominantly due to the areal increase (~7.5 sq deg c.f. ~0.8 sq. deg in Whittam et al. 2024). However, there are also some improvements in the sensitivity both within the COSMOS field as well as from the inclusion of the CDFS-DEEP data. This is illustrated through the comparison of the sensitivity limits from the three different fields of MIGHTEE-DR1 compared to the data available from MIGHTEE-ES. For COSMOS, this improved sensitivity is as a result of the increased number of pointings in the MIGHTEE-DR1 COSMOS data (compared to the single pointing in MIGHTEE-ES) which reduces sensitivity loss in the outer edges of the field and improves image sensitivity throughout. For CDFS-DEEP, this is due to the longer integration time within a single pointing centre.

As a result, the median luminosity being probed within each of the fields is reduced compared to the MIGHTEE-ES data. In the COSMOS and XMM-LSS fields, the median luminosities are ~1.5-1.6×10^{23} WHz^{-1}, which is a factor of ~2 lower than the median luminosity of Whittam et al. (2024) (~3×10^{23} WHz^{-1}). In the deeper CDFS-DEEP field, the median luminosity is a factor of four smaller compared to the median luminosity MIGHTEE-ES (with a median $L_{1.4}$ ~7.4×10^{22} WHz^{-1} across the sample). This reduction in median luminosity provides a significant sample of galaxies to study the evolution of AGN and SFGs to lower luminosity sources and to higher redshifts. For example, considering the sensitivity curves in Figure 10, a source with $L_{1.4} \sim 10^{23}$ WHz^{-1} could be observed to $z \approx 2$ in the CDFS-DEEP fields, as opposed to $z \approx 1.5$ in the COSMOS field and $z \approx 1$ in the MIGHTEE-ES data. This adds an extra $\sim 1-2$ Gyr of lookback time to trace the evolution of sources at this luminosity.

Table 2. Details of the number of radio sources over the restricted multi-wavelength regions (see Figure 1) before and after host galaxy association. Also listed are the number of sources post host association with redshifts assigned, and the corresponding subsets of these where the final redshifts (`z_best_final`) are from the photometric ($z_{\rm phot}$), spectroscopic ($z_{\rm spec}$) and, for the COSMOS field, with redshifts from the Pau survey ($z_{\rm Pau}$). For the subset of sources with redshifts we indicate the fraction of the host associated catalogue this represents.

	CDFS	COSMOS	XMMLSS	Total
N_s (pre host association)	20 335	13 781	31 491	65 607
N_s (post host association)	20 458	13 782	31 412	65 652
N_s With z	18 380 (89.8 per cent)	13 117 (95.2 per cent)	28 493 (90.7 per cent)	59 990 (91.4 per cent)
N_s With $z_{\rm phot}$	13 987 (68.4 per cent)	5 224 (37.9 per cent)	18 622 (59.3 per cent)	37 833 (57.6 per cent)
N_s With $z_{\rm Pau}$	0 (0.0 per cent)	637 (4.6 per cent)	0 (0.0 per cent)	637 (1.0 per cent)
N_s With $z_{\rm spec}$	4 393 (21.5 per cent)	7 256 (52.6 per cent)	9 871 (31.4 per cent)	21 520 (32.8 per cent)

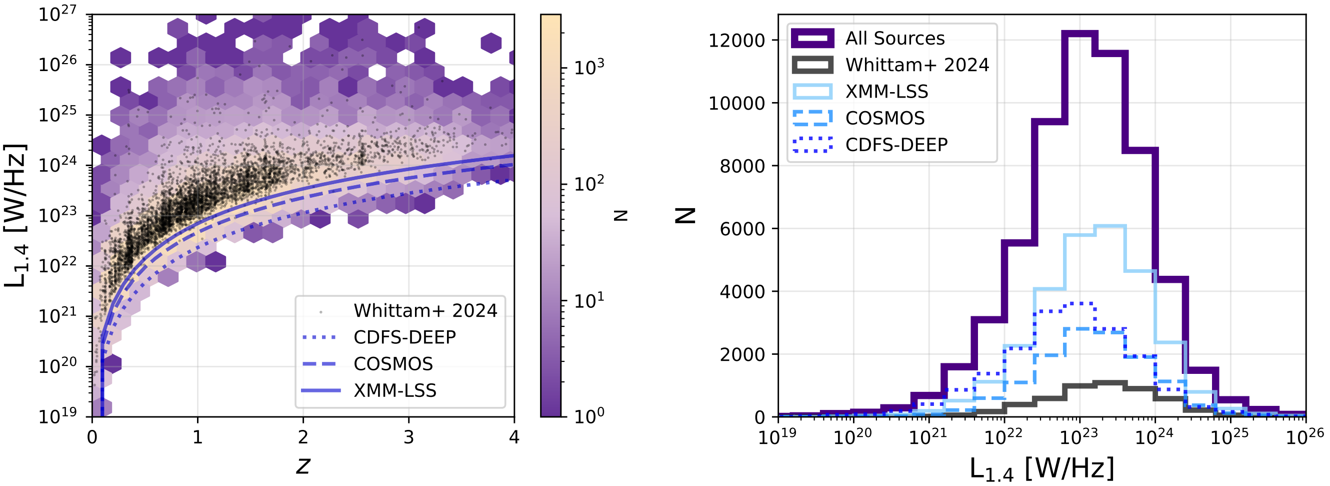

Figure 10. The 1.4 GHz luminosity distribution of sources across the combined DR1 cross-matched catalogues. Left: the distribution of luminosity with redshift up to $z = 5$, compared to the 5σ limiting luminosity (adopting the central rms values of Hale et al. 2025, and an average effective frequency of 1.2 GHz) for the CDFS-DEEP (black dashed), COSMOS (navy dot-dashed) and XMM-LSS (blue dotted) fields scaled to 1.4 GHz. Also shown (black dots) are the corresponding distributions from the MIGHTEE early science data, using the catalogue of visually cross-matched sources that were further used to classify sources (from Whittam et al. 2022, 2024). Right: the combined distribution of luminosity for all sources with a redshift (thick purple line) and separated into those sources from the CDFS-DEEP (dark blue dotted), COSMOS (blue dashed) and XMM-LSS (light blue solid) field, as well as the MIGHTEE-ES visually inspected cross-matched catalogue (grey).

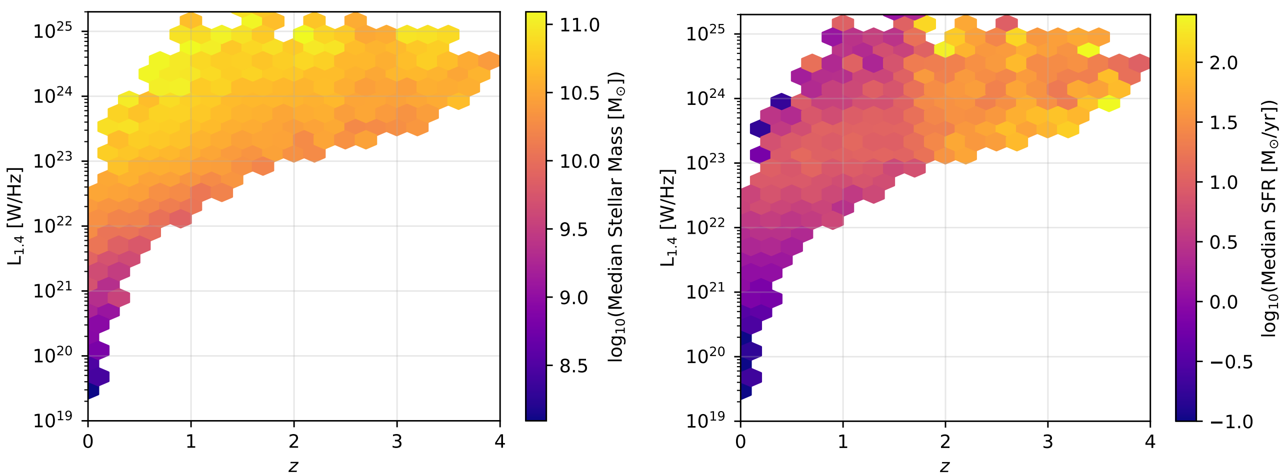

Figure 11. Median stellar mass (left) and SFR (right) for the MIGHTEE-DR1 sources with an associated host galaxy as a function of the corresponding redshift and 1.4 GHz radio luminosity. The stellar mass and SFR are those derived from SED fitting in the work of Stylianou et al. in prep.

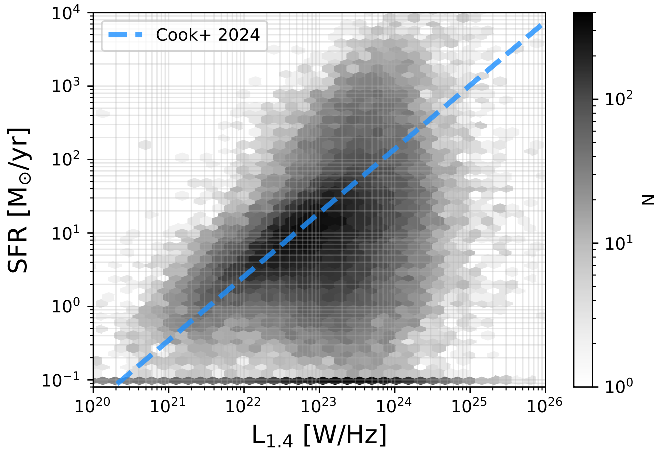

Figure 12. Comparison of the SED measured SFR in the work of Stylianou et al. (in prep) as a function of the 1.4 GHz radio luminosity alongside the relation of Cook et al. (2024, blue dashed line) adopted in this work to define radio excess sources (see Section 5.2). We note that the excess of sources at SFR = 0.1 $M_\odot yr^{-1}$ (representing $\lesssim$10 per cent of the full population) represents the lower limit for the SFR in the SED fitting of Stylianou et al. (in prep).

5.2 Redshift distributions of AGN and SFGs

Accurate classification of radio sources into SFGs, AGN and their subsets relies upon a number of diagnostics factors from across the electromagnetic spectrum (e.g. X-ray counterparts, very long baseline interferometry, VLBI, imaging, emission lines within spectra). This includes requiring detailed modelling of the spectral energy distribution (SED) fitting to include the contribution of any AGN to the full spectrum (see e.g. Calistro Rivera et al. 2016; Das et al. 2024; Buchner et al. 2024). Such classifications are essential to accurately study how SFGs and AGN evolve differently and thus the contributions of star formation and AGN activity across cosmic time. Moreover, given different physical sub-populations of AGN such as high/low excitation radio galaxies (see e.g. Best & Heckman 2012; Heckman & Best 2014; Whittam et al. 2022) and radio quiet AGN all evolve differently, it is crucial to be able to separate AGN into sub-populations and study their evolution (see e.g. Best & Heckman 2012; Pracy et al. 2016; Whittam et al. 2022; Kondapally et al. 2022, 2025). This requires detailed modelling utilising all the information we have on these sources which is beyond the scope of this work and will be studied in future MIGHTEE publications.

However, the catalogues of Stylianou et al. (in prep) do provide some host galaxy properties, including stellar mass and star-formation rate, were derived by Stylianou et al. (in prep) using the SED-fitting code `LePHARE`, with the redshifts fixed for each source[10]. The fits were performed using total-auto fluxes, rather than aperture fluxes, which better capture the total light of each galaxy and are therefore more suitable for estimating integrated galaxy properties. The Bruzual & Charlot (2003) SED templates were adopted for this step and included both solar and 20 per cent solar metallicity populations. A range of star-formation histories was considered, including a constant SFR model and exponentially declining models with timescales from 0.1 to 30 Gyr, with stellar population ages spanning 0.01-13.5 Gyr. We do note, though, that the contamination from AGN in the SED and the lack of far-IR data in their SED fitting may bias the SFR values and affect the ability to accurately constrain it, this will be addressed later in the section. We present the distribution of the median of both M_* and SFR as a function of redshift and radio luminosity for the MIGHTEE DR1 cross-matched sources in Figure 11. As expected, the most massive galaxies appear to trace more luminous radio galaxies, which are more likely to host AGN (see e.g. Smolčić et al. 2017a; Whittam et al. 2022; Best et al. 2023). However, the SFR does not exhibit smooth trends between SFR, z and luminosity. This is because the radio sources represent a mix of populations and whilst the SFR and radio luminosity for SFGs is correlated (see e.g. relations in Bell 2003; Davies et al. 2017; Delhaize et al. 2017; Delvecchio et al. 2021; Cook et al. 2024), this is not the case for AGN.

5.2.1 *Identifying radio excess AGN*

In this work, we use this relationship between radio luminosity and SFR for SFGs to take a first step at studying the redshift distributions of AGN and SFGs separately in the MIGHTEE-DR1 data. For this, we utilize the SFR-$L_{1.4}$ relationship of Cook et al. (2024) which used a combination of MIGHTEE, Deep Investigation of Neutral Gas Origins survey (DINGO; Rhee et al. 2023) and the Deep Extragalactic VIsible Legacy Survey (DEVILS; Davies et al. 2018, 2025) data to measure the SFR-$L_{1.4}$ relationship to $z \sim 1$ (see Figure 12). We compare this to the SFRs provided by the catalogues of Stylianou et al. in prep. However, as both the star formation history (SFH) and the choice of data used to model the SED or spectra can affect the measured SFR, we allow for variations in the relationship to the normalisation of Cook et al. (2024). To do this, we follow the method of Mazzolari et al. (2026) and define a radio excess ratio (REX) which is calculated as the ratio of the SFR predicted by the 1.4 GHz luminosity using the relationship of Cook et al. (2024) compared to the SFR from the SED fitting. Using $\log_{10}$(REX), we model the peak in this distribution ($\mu_{\log_{10}(\mathrm{REX})}$, measured to be ∼0.06) and use those sources with a $\log_{10}$(REX) below this peak (which cannot be radio-excess sources) with a Gaussian to find a standard deviation in the relationship ($\sigma_{\log_{10}(\mathrm{REX})}$, measured to be ∼0.5) for SFGs around the radio excess. Those sources with:

$$\mathrm{REX} > 10^{\mu_{\log_{10}(\mathrm{REX})}+2\sigma_{\log_{10}(\mathrm{REX})}} \quad (6)$$

are denoted to be radio excess sources (and are characterised here as likely radio AGN). Whilst previous studies have instead used 3σ deviations in SFR-radio relations to define their radio excess criteria (e.g. Best et al. 2023; Mazzolari et al. 2026), these find 3σ deviations in the range $\sim 0.7-1$ dex, which is closer to the $\sim 1.5-2\sigma$ relations adopted in this work. However, given that the SFRs calculated by Stylianou et al. (in prep) do not include far-IR data, which can be important in constraining the SFR of galaxies, as has been used in

[10] We note that for their analysis to obtain galaxy properties, Stylianou et al. (in prep) fix the redshifts were fixed to those best redshift solutions using galaxy templates only (i.e. without AGN present). This is that case even if the template with AGN present provided the best fitting redshift, this is as the models with AGN can struggle to constrain the stellar mass as accurately. Therefore for some sources in CDFS (∼20 per cent) where the AGN template was most appropriate and for the sources in COSMOS where the PAU redshifts were adopted, this is different to the z recorded in our catalogue. However, using the same OLF criteria as in Section 4.3.2, of the total sample of sources with redshifts, sources with large OLF due to this represent ∼9 per cent of sources in CDFS-DEEP and <0.5 per cent of sources in COSMOS and therefore the impact of this should not be significant on our work.

previous studies, it is likely we might expect larger scatter in our REX. Therefore, given that 2σ should obtain the majority (~95 per cent) of star forming sources close to the relation and that this compares well to previous values used in the literature, the criteria that we use should be sufficient for identifying star forming and radio excess sources within our sample. We note though that the SFG population will likely also include a small contribution of radio quiet AGN, whose origin of radio emission is still uncertain (see e.g. White et al. 2015, 2025; Njeri et al. 2026), jetted but non radio excess AGN (see e.g. Vardoulaki et al. 2021) and will also struggle to identify dusty star forming galaxies, whose emission is obscured and thus may affect their identification.

Furthermore, as the SFR-$L_{1.4}$ relation of Cook et al. (2024) is restricted to fitting sources at $z \lesssim 1$, this may not fully account for evolution in this relationship, as discussed in Thykkathu et al. (2026). Therefore, we additionally separate into radio-excess and star forming sources using an evolving REX model and fit for both $\mu_{\log_{10}(\mathrm{REX})}$ and $\sigma_{\log_{10}(\mathrm{REX})}$ within redshift bins (with $\delta z = 0.5$). This does indeed exhibit evolution in $\mu_{\log_{10}(\mathrm{REX})}$ to $\lesssim 0.9$ and, as a result of the fewer sources within increasing redshift bins there is also an increase in the scatter around the peak values. Using this evolving split will provide a check on the effect of changes to the SFR-$L_{1.4}$ models on our work. Numerous other relationships between SFR and radio luminosity have been suggested in the literature (see e.g. Delhaize et al. 2017; Delvecchio et al. 2021) and we note we also investigated using these in the same way as we used the relation of Cook et al. (2024) and found that they show broadly similar redshift distributions to obtained here.

5.2.2 Comparison of the redshift distribution for SFG and AGN analogues

Adopting this split to determine radio-excess sources, we consider the redshift distributions of radio AGN and SFGs within our MIGHTEE sample and compare it to predicted distributions from luminosity functions from the VLA 3 GHz COSMOS data (Novak et al. 2018) and from MIGHTEE-ES data (Thykkathu et al. 2026). We also compare to the simulated radio continuum catalogues from the SKA Design Studies (SKADS; Wilman et al. 2008) and from the Tiered Radio Extragalactic Continuum Simulation (T-RECS; Bonaldi et al. 2019, 2023)[11]. We present a comparison of the expected number counts and the normalised redshift probability distributions, $p(z)$, in Figure 13 at a number of limiting flux densities. This allows us to compare how the redshift distributions appear over a range of flux density cuts, where we expect AGN to increase in fraction at higher flux densities (see e.g. Whittam et al. 2022; Best et al. 2023). We present such comparisons at 30 µJy, 50 µJy, 80 µJy, 100 µJy and 200 µJy. These are chosen as similar to the lower flux density limits which are detectable across the three fields (30 μJy), increasing to large flux density limits which are: (i) still deep but where we believe we are ~complete (50 - 80 µJy) and SFGs dominate at the flux density limit; (ii) the flux limit where we believe the AGN and SFG are similarly contributing to the full population (100 µJy, as in Whittam et al. 2022) and (iii) a higher flux density limit where AGN dominate the sample (200 µJy).

As shown in Figure 13, whilst there are differences between adopting the non-evolving and evolving REX in classifying the MIGHTEE-DR1 sources into SFGs and radio-excess AGN, these differences are typically limited and do not substantially change the overall shape of the redshift distribution. At the larger flux densities ($S_{1.4} \geq 100$µJy) the redshift distributions for the SFGs are in excellent agreement with both the predictions from the luminosity functions of both Novak et al. (2018) and Thykkathu et al. (2026). At lower flux densities, the redshift distribution of the MIGHTEE-DR1 SFG sources instead appears to lie between the relationships of Novak et al. (2018) and Thykkathu et al. (2026). This reflects differences in the modelling the luminosity function, where Novak et al. (2018) adopt the faint end slope from the local luminosity functions of Mauch & Sadler (2007b), whereas Thykkathu et al. (2026) adopt this as a free parameter. Thykkathu et al. (2026) exhibit a much larger prediction in the number of SFGs (and total sources) at the faintest flux densities. However, we note that the faint end slope in Thykkathu et al. (2026) is driven by fitting the faintest sources in their lowest redshift bin ($z = 0.2 - 0.4$) which they fit only to a limiting luminosity of 10^{22} WHz^{-1}. This represents flux densities closer to ~ 90µJy, at the lower redshift considered ($z = 0.2$), and ~ 20µJy at their higher redshift range ($z = 0.4$). Moreover, given the MIGHTEE-ES data is noisier and at lower resolution than in this work, this could result in challenges disentangling sources at fainter flux densities or a misestimation of source completeness. Combined this could result in an over-prediction of the faint end slope, leading to an over-prediction of faint sources compared to that found with our MIGHTEE-DR1 catalogue. For AGN, on the other hand, both the predictions from the luminosity functions of Novak et al. (2018) and Thykkathu et al. (2026) reflect in the peak of the distribution from the MIGHTEE-DR1 data and show broad agreement in the shape of the redshift distribution, though again the work of Thykkathu et al. (2026) over-predicts the number of AGN compared to Novak et al. (2018), with the MIGHTEE-DR1 catalogue closer to that of Novak et al. (2018). However, given the caveats discussed above in determining the SFR of galaxies and therefore their REX, we could be mistaking radio excess AGN for SFGs, given the larger spread in our data than in previous work. Moreover, as the work of Thykkathu et al. (2026) accounts for non cross-matched sources using an assumption that these are all at $z > 1$ (in the model compared to in this work), it will likely have larger numbers of sources at $z \gtrsim 1$ than with our catalogue, where do not attempt to account for the up to ~10 per cent of sources without an associated host galaxy and redshift.

Instead, focusing on the $p(z)$ distribution, we find the peak in the $p(z)$ for both the AGN and SFG analogues are in good agreement with those from the luminosity functions. This is regardless of the differences in normalisation in total source numbers, see Figure 13. For the SFGs (star forming) populations, there is excellent agreement between the star forming galaxies in this work and those simulated SFGs in T-RECS, but they show more differences in comparison to SKADS, which appear to have a shallower peak and extend to higher redshifts. This reinforces known concerns from studies of the deep source counts from radio surveys which have highlighted the under prediction of galaxies in SKADS at the faintest flux densities, where SFGs dominate (see e.g. Smolčić et al. 2017a; Hale et al. 2023, 2025). For the AGN, these again broadly follow the models from the luminosity functions, though around $z \sim 1$ and $z \sim 1.5$ there are noticeable differences in the $p(z)$ when the evolving and non-evolving REX models are adopted, and also there is an excess of AGN compared to the models from the luminosity functions. This may be as a result of aliasing in the photometric redshifts, as we adopt the peak in the redshift pdf for this work, as this is the redshift adopted in Stylianou et al. (in prep) to obtain the SFRs. These effects might therefore be smoothed out should the full pdf be considered. The predictions of AGN distributions for both the T-RECS and SKADS simulations are both in broadly good agreement with the radio-excess

[11] Using the continuum clustered output simulation of Bonaldi et al. (2023).

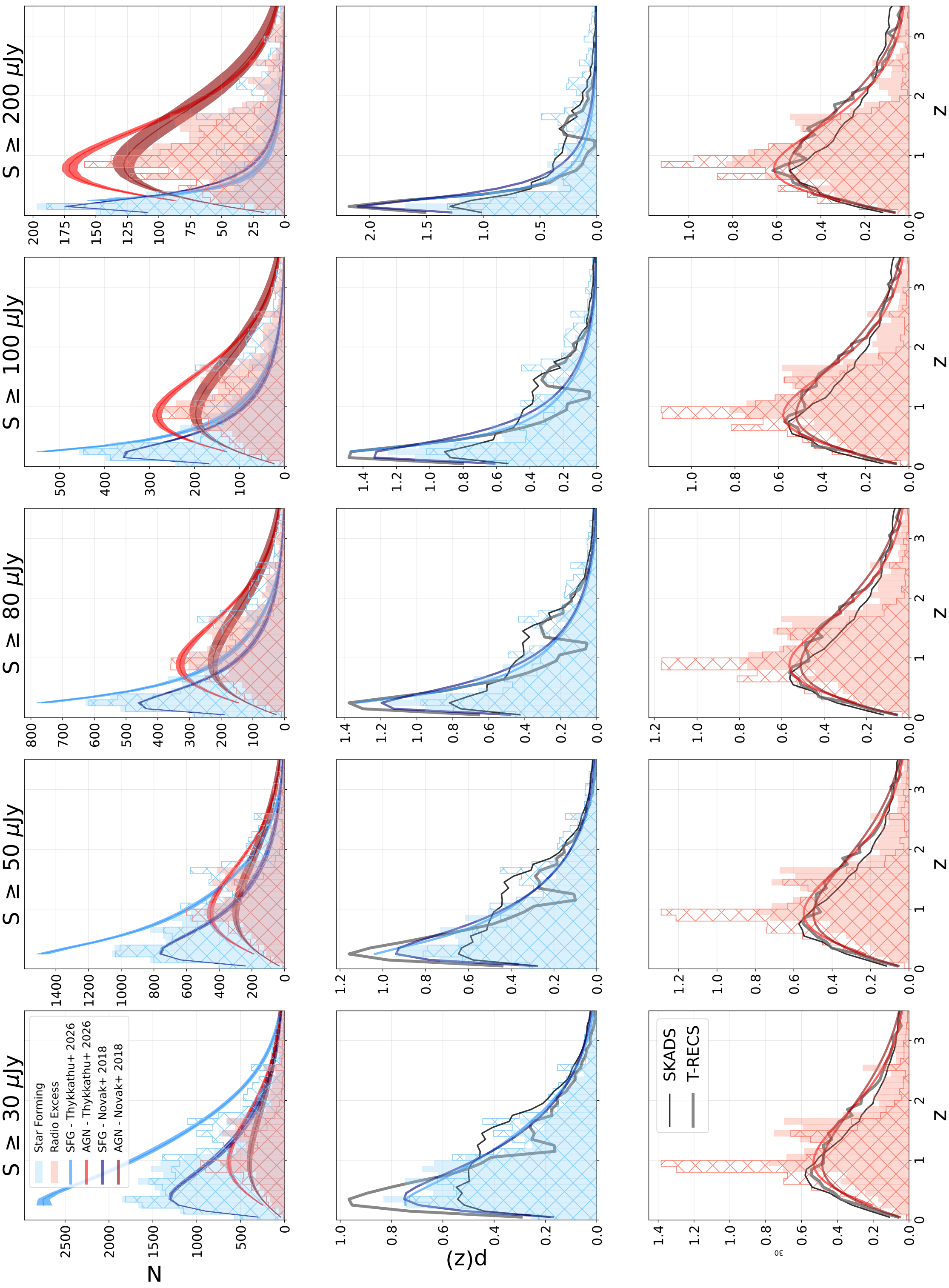

Figure 13. Upper panel: Comparison of the observed redshift distribution of sources identified as star-forming (blue histograms) and radio-excess (red histograms) using the selection outlined in Section 5.2 for the non-evolving (filled) and evolving (hatched) REX relations. This is compared to the predicted redshift distribution from the luminosity functions of Novak et al. (2018) (blue/red for SFGs and AGN) and Thykkathu et al. (2026) (navy/dark red). The middle and lower panels show the normalised redshift distribution ($p(z)$) from the star forming sources and radio excess sources respectively. In the middle panel we also indicate the fraction of star forming sources from the non-evolving and evolving (indicated by (e)) models. Additionally, the two lower panels show the distributions from the simulations from SKADS (Wilman et al. 2008, black) and T-RECS (Bonaldi et al. 2023, grey). The panels from left to right indicate 1.4 GHz flux limits applied of 30 μJy, 50 μJy, 80 μJy, 100 μJy and 200 μJy respectively.

in this work and both simulations peak at a similar redshift and show similar behaviour.

Overall, the redshift distributions suggest that the simple radio-excess criterion based on Cook et al. (2024) and with an evolving relation adopted in this work show broad agreement with the predicted distributions from simulations and measured luminosity functions. However, for individual sources it is imperative that more comprehensive diagnostics and SED fitting (including the far-IR) are used to accurately characterise sources. This will be addressed in future work (Jackson et al. in prep) which will provide much more comprehensive host properties for our radio sample to characterise sources and to allow for galaxy evolution studies, for different sub-populations.

We note that whilst this broad source classification does allow the potential to study how the AGN fraction varies as a function of host galaxy stellar mass for the near-IR sample of galaxies and to make comparisons with works such as Kondapally et al. (2025), the simplicity in the source characterization using the REX method outlined above will lead to some misclassification of sources. Whilst misclassification is a less significant issue when the broad redshift distributions are being considered, it can become more important when the AGN fractions are small. This can lead to a flattening of the distribution of AGN fraction as a function of mass (which is expected to be $\propto M^{2.5}$ for LERGs and $\propto M^{1.5}$ for HERGs, see e.g. Best et al. 2005; Janssen et al. 2012; Kondapally et al. 2025), especially when the lowest mass is considered. Initial investigations by the authors identified some of these challenges and also noted differences in the relationships between the three fields studied, such as whether there was any potential redshift evolution. Such differences may arise from differences in the REX criterion applied, which uses an average relation from the combined three fields. However, given the differences in the quality of multi-wavelength data between the fields (with COSMOS having the deepest K_s data) and differences in their redshift estimation, such relations may vary. As such, we leave this comparison to future work where more comprehensive source classification has been performed and where AGN have been sub-categorised into LERGs and HERGs, for more appropriate comparison to the work of Kondapally et al. (2025).

6 CONCLUSIONS

In this work we have presented a cross-matched catalogue for radio sources in the MIGHTEE Data Release 1 catalogues of Hale et al. (2025) over a total of 7.5 sq deg across the CDFS-DEEP (1.5 sq deg), COSMOS (1.7 sq deg) and XMM-LSS (4.3 sq deg) fields through matching to K_s band catalogues from the UltraVISTA (McCracken et al. 2012) and VIDEO (Jarvis et al. 2013) surveys. We adopted a combined method of using statistical cross matching, through the likelihood ratio (LR) technique, and visual inspection. This combined approach minimises the need for visual inspection, allowing sources which are compact and with a high likelihood host match to be adopted, without the need for visual inspection. This is the case for $\sim$ 80 per cent of sources. Instead for larger sources or where the best LR counterpart is not deemed a secure enough match, these were sent for visual inspection using the Zooniverse platform and the input from members of the MIGHTEE consortium.

Combined, this process provided host galaxy counterparts and redshifts (within the multi-wavelength areas discussed) for $\sim$91 per cent of sources in the CDFS-DEEP field, increasing to $\sim$92 per cent in the XMM-LSS field and $\sim$95 per cent of sources in COSMOS. Of these, close to 50 per cent of sources in the COSMOS field have spectroscopic counterparts, reducing to 30 per cent and 20 per cent of sources with spectroscopic counterparts in the XMM-LSS and CDFS-DEEP fields respectively. These fractions in CDFS-DEEP and XMM-LSS will be increased by the targetted search of spectra for MIGHTEE sources in the ORCHIDSS (Duncan et al. 2023) survey. For those sources with counterparts, we compare the luminosity distribution to that of the MIGHTEE-ES counterpart catalogue of Whittam et al. (2024), demonstrating that due to the areal coverage, we have an order of magnitude more sources with counterparts and for those sources with $L_{1.4} \sim 10^{23}$ WHz^{-1} we are able to trace these over 1-2 Gyr further in lookback time due to increased in sensitivity of the data. The sky source density is additionally a factor of 2 higher compared to the LOFAR deep fields DR1 (Kondapally et al. 2021), albeit with MIGHTEE-DR1 covering a smaller area and so having fewer total source numbers. Combined, this will allow MIGHTEE-DR1 to examine the properties of sources with lower radio luminosities, which are crucial to understand galaxy evolution in less extreme sources.

Finally, we use star formation rates predicted by Cook et al. (2024) based on the radio luminosity (using both an evolving and non-evolving model) compared to SFRs from SED fitting in Stylianou et al. (in prep), to define a radio-excess population (AGN analogue) and star forming sample of galaxies. Investigating the redshift distributions of these AGN and SFG analogues, we compare this to both simulated radio continuum catalogues (T-RECS and SKADS) and the inferred redshift distributions from luminosity functions from VLA 3 GHz COSMOS data (Novak et al. 2018) and from MIGHTEE-ES data (Thykkathu et al. 2026). We find broad agreement in the redshift distributions with our radio excess and star forming samples, to that from the luminosity functions, though find the distribution of MIGHTEE-DR1 galaxies at the faintest flux densities are lower than predicted by MIGHTEE-ES data. As this may be a result of source confusion in the MIGHTEE-ES data and potential source blending, this emphasises that high resolution imaging are essential to constrain the faint end slope in the luminosity functions of SFGs and to accurately predict the expected distribution of the faintest sources within radio surveys. Whilst we have demonstrated that a radio excess cut using the host properties of the radio sources from optical to near-IR catalogues does allow us to broadly replicate expectations for the redshift distributions of AGN and SFGs, exact source classifications require a number of diagnostic from across the electromagnetic spectra (e.g. the identification of X-ray AGN) and using a combination of SED and spectral information. For MIGHTEE sources, improved SED modelling for AGN components and which utilise far-IR data will be essential.

ACKNOWLEDGEMENTS

We thank I. Smail for helpful comments on the draft manuscript. CLH and PNB acknowledges support from the Science and Technology Facilities Council (STFC) through grant ST/Y000951/1 and from the Leverhulme Trust through an Early Career Research Fellowship. CLH, MJJ and IHW also acknowledge support from the Oxford Hintze Centre for Astrophysical Surveys which is funded through generous support from the Hintze Family Charitable Foundation. MJJ, AAV, HP, SLJ, NS, MNT, RGV acknowledge the support from a UKRI Frontiers Research Grant [EP/X026639/1], which was selected by the ERC. SLJ also acknowledges support from the STFC consolidated grants [ST/S000488/1] and [ST/W000903/1]. MNT has been supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programmes (grant agreement no. 101018897 CosmicExplorer). MJH thanks the UK STFC for support [ST/V000624/1, ST/Y001249/1]. DJBS acknowledges support from the United Kingdom's Science and Technology Facilities Council (STFC) via grant ST/Y001028/1, and from the Leverhulme Trust via Research Project Grant RPG-2025-078. JD, LS, KKC and NN acknowledge partial research support by the National Research Foundation of South Africa (Ref Number CSUR240426216203) and an Africa-Oxford Catalyst Collaboration Grant (AfOx-290). L.V.M. acknowledges financial support from the grant CEX2021-001131-S funded by MICIU/AEI/ 10.13039/501100011033, from the grant PID2021-123930OB-C21 and PID2024-155817OB-I00 funded by MICIU/AEI/ 10.13039/501100011033 and by ERDF/EU. LM, TFR & MV acknowledge financial support from the Inter-University Institute for Data Intensive Astronomy (IDIA), a partnership of the University of Cape Town, the University of Pretoria and the University of the Western Cape. LM & MV also acknowledge support from the South African Department of Science and Innovation's National Research Foundation under the ISARP RADIOMAP Joint Research Scheme (DSI-NRF Grant Number 150551) and the CPRR Programme (DSI-NRF Grant Number SRUG2204254729 and SRUG22031677). The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science, Technology and Innovation. We acknowledge the use of the ilifu cloud computing facility – `www.ilifu.ac.za`, a partnership between the University of Cape Town, the University of the Western Cape, Stellenbosch University, Sol Plaatje University and the Cape Peninsula University of Technology. The ilifu facility is supported by contributions from the Inter-University Institute for Data Intensive Astronomy (IDIA), the Computational Biology division at UCT and the Data Intensive Research Initiative of South Africa (DIRISA).

DATA AVAILABILITY

The catalogues produced in this work will be released alongside the publication of this article.

REFERENCES

Ahumada R., et al., 2020, ApJS, 249, 3
Aihara H., et al., 2022, PASJ, 74, 247
Alarcon A., et al., 2021a, MNRAS, 501, 6103
Alarcon A., et al., 2021b, MNRAS, 501, 6103
Algera H. S. B., et al., 2020, ApJ, 903, 139
Almosallam I. A., Lindsay S. N., Jarvis M. J., Roberts S. J., 2016a, MNRAS, 455, 2387
Almosallam I. A., Jarvis M. J., Roberts S. J., 2016b, MNRAS, 462, 726
Arnaudova M. I., et al., 2025, MNRAS, 542, 2245
Arnouts S., Ilbert O., 2011, LePHARE: Photometric Analysis for Redshift Estimate, Astrophysics Source Code Library, record ascl:1108.009
Banfield J. K., et al., 2015, MNRAS, 453, 2326
Bell E. F., 2003, ApJ, 586, 794
Benson A. J., 2010, Phys. Rep., 495, 33
Bertin E., Arnouts S., 1996, A&AS, 117, 393
Best P. N., Heckman T. M., 2012, MNRAS, 421, 1569
Best P. N., Kauffmann G., Heckman T. M., Brinchmann J., Charlot S., Ivezić Ž., White S. D. M., 2005, MNRAS, 362, 25
Best P. N., et al., 2023, MNRAS, 523, 1729
Bonaldi A., Bonato M., Galluzzi V., Harrison I., Massardi M., Kay S., De Zotti G., Brown M. L., 2019, MNRAS, 482, 2
Bonaldi A., Hartley P., Ronconi T., De Zotti G., Bonato M., 2023, MNRAS, 524, 993
Bradshaw E. J., et al., 2013, MNRAS, 433, 194
Briggs D. S., 1995, PhD thesis, New Mexico Institute of Mining and Technology
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Buchner J., et al., 2024, A&A, 692, A161
Calistro Rivera G., Lusso E., Hennawi J. F., Hogg D. W., 2016, ApJ, 833, 98
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Cappelluti N., et al., 2009, A&A, 497, 635
Chabrier G., 2003, PASP, 115, 763
Chen C. T. J., et al., 2018, MNRAS, 478, 2132
Chen S., Laor A., Behar E., Baldi R. D., Gelfand J. D., 2023, MNRAS, 525, 164
Cirasuolo M., et al., 2020, The Messenger, 180, 10
Cochrane R. K., et al., 2023, MNRAS, 523, 6082
Coil A. L., et al., 2011, ApJ, 741, 8
Condon J. J., 1974, ApJ, 188, 279
Condon J. J., 1992, ARA&A, 30, 575
Cook R. H. W., et al., 2024, MNRAS, 531, 708
Cool R. J., et al., 2013, ApJ, 767, 118
Cooper M. C., et al., 2012, MNRAS, 425, 2116
DESI Collaboration et al., 2025, arXiv e-prints, p. arXiv:2503.14745
Dahlen T., et al., 2013, ApJ, 775, 93
Damjanov I., Zahid H. J., Geller M. J., Fabricant D. G., Hwang H. S., 2018, ApJS, 234, 21
Das S., et al., 2024, MNRAS, 531, 977
Davies L. J. M., et al., 2015, MNRAS, 447, 1014
Davies L. J. M., et al., 2017, MNRAS, 466, 2312
Davies L. J. M., et al., 2018, MNRAS, 480, 768
Davies L. J. M., et al., 2025, MNRAS, 544, 3005
Delhaize J., et al., 2017, A&A, 602, A4
Delvecchio I., et al., 2021, A&A, 647, A123
Drake A. B., et al., 2024, MNRAS, 534, 1107
Driver S. P., et al., 2019, The Messenger, 175, 46
Driver S. P., et al., 2022, MNRAS, 513, 439
Duncan K. J., Jarvis M. J., Brown M. J. I., Röttgering H. J. A., 2018, MNRAS, 477, 5177
Duncan K. J., et al., 2021, A&A, 648, A4
Duncan K., et al., 2023, The Messenger, 190, 25
Eales S., et al., 2009, ApJ, 707, 1779
Euclid Collaboration et al., 2025, A&A, 697, A1
Fleuren S., et al., 2012, MNRAS, 423, 2407
Garilli B., et al., 2014, A&A, 562, A23
Garn T., Green D. A., Riley J. M., Alexander P., 2009, MNRAS, 397, 1101
Gupta Y., et al., 2017, Current Science, 113, 707
Guzzo L., et al., 2014, A&A, 566, A108
Hale C. L., et al., 2019, A&A, 622, A4
Hale C. L., et al., 2023, MNRAS, 520, 2668
Hale C. L., et al., 2025, MNRAS, 536, 2187
Hales C. A., et al., 2014, MNRAS, 441, 2555
Hardcastle M. J., Croston J. H., 2020, New Astron. Rev., 88, 101539
Hardcastle M. J., Evans D. A., Croston J. H., 2007, MNRAS, 376, 1849
Hardcastle M. J., et al., 2023, A&A, 678, A151
Hasinger G., et al., 2018, ApJ, 858, 77
Hatfield P. W., Jarvis M. J., Adams N., Bowler R. A. A., Häußler B., Duncan K. J., 2022, MNRAS, 513, 3719
Heckman T. M., Best P. N., 2014, ARA&A, 52, 589
Heywood I., 2020, oxkat: Semi-automated imaging of MeerKAT observations, Astrophysics Source Code Library, record ascl:2009.003
Heywood I., Hale C. L., Jarvis M. J., Makhathini S., Peters J. A., Sebokolodi M. L. L., Smirnov O. M., 2020, MNRAS, 496, 3469
Heywood I., et al., 2021, MNRAS,
Hoaglin D. C., Mosteller F., Tukey J. W., 1983, Understanding robust and exploratory data anlysis
Hopkins A., et al., 2025, Publ. Astron. Soc. Australia, 42, e071
Hotan A. W., et al., 2021, Publ. Astron. Soc. Australia, 38, e009
Hsu L.-T., et al., 2014, ApJ, 796, 60
Ilbert O., et al., 2006, A&A, 457, 841

Ilbert O., et al., 2009, ApJ, 690, 1236
Ivezić Ž., et al., 2019, ApJ, 873, 111
Iwasawa K., et al., 2020, A&A, 639, A51
Jackson C. L., Matthews J. H., Whittam I. H., Jarvis M. J., Temple M. J., Rankine A. L., Hewett P. C., 2026, MNRAS, 546, stag065
Janssen R. M. J., Röttgering H. J. A., Best P. N., Brinchmann J., 2012, A&A, 541, A62
Jarvis M. J., et al., 2013, MNRAS, 428, 1281
Jarvis M., et al., 2016, in MeerKAT Science: On the Pathway to the SKA. p. 6 (arXiv:1709.01901), doi:10.22323/1.277.0006
Jonas J. L., 2009, IEEE Proceedings, 97, 1522
Jonas J., MeerKAT Team 2016, in MeerKAT Science: On the Pathway to the SKA. p. 1, doi:10.22323/1.277.0001
Kauffmann G., Charlot S., 1998, MNRAS, 297, L23
Khostovan A. A., et al., 2025, arXiv e-prints, p. arXiv:2503.00120
Kondapally R., et al., 2021, A&A, 648, A3
Kondapally R., et al., 2022, MNRAS, 513, 3742
Kondapally R., et al., 2025, MNRAS, 536, 554
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Lacy M., et al., 2020, PASP, 132, 035001
Lacy M., et al., 2021, MNRAS, 501, 892
Lal D. V., Taylor R., Sekhar S., Ishwara-Chandra C., Dutta S., Kolwa S., 2025, ApJ, 991, 9
Laor A., Behar E., 2008, MNRAS, 390, 847
Le Fevre O., et al., 2013, arXiv e-prints, p. arXiv:1307.0545
Lidman C., et al., 2020, MNRAS, 496, 19
Lilly S. J., et al., 2007, ApJS, 172, 70
Lintott C., et al., 2012, in American Astronomical Society Meeting Abstracts #219. p. 125.07
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Maltby D. T., et al., 2016, MNRAS, 459, L114
Mao M. Y., et al., 2012, MNRAS, 426, 3334
Martin D. C., et al., 2005, ApJ, 619, L1
Matthews A. M., Condon J. J., Cotton W. D., Mauch T., 2021, ApJ, 914, 126
Mauch T., Sadler E. M., 2007a, MNRAS, 375, 931
Mauch T., Sadler E. M., 2007b, MNRAS, 375, 931
Mauduit J. C., et al., 2012, PASP, 124, 714
Mazzolari G., et al., 2026, arXiv e-prints, p. arXiv:2603.10176
McAlpine K., Smith D. J. B., Jarvis M. J., Bonfield D. G., Fleuren S., 2012, MNRAS, 423, 132
McCracken H. J., et al., 2012, A&A, 544, A156
McLure R. J., et al., 2013, MNRAS, 428, 1088
Mingo B., Hardcastle M. J., Croston J. H., Dicken D., Evans D. A., Morganti R., Tadhunter C., 2014, MNRAS, 440, 269
Mohan N., Rafferty D., 2015, PyBDSF: Python Blob Detection and Source Finder, Astrophysics Source Code Library, record ascl:1502.007 (ascl:1502.007)
Moncelsi L., et al., 2011, ApJ, 727, 83
Morrissey P., et al., 2007, ApJS, 173, 682
Murphy E. J., 2009, ApJ, 706, 482
Njeri A., et al., 2026, MNRAS, 546, stag097
Novak M., et al., 2017, A&A, 602, A5
Novak M., Smolčić V., Schinnerer E., Zamorani G., Delvecchio I., Bondi M., Delhaize J., 2018, A&A, 614, A47
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Patel H., Clements D. L., Rowan-Robinson M., Vaccari M., 2011, MNRAS, 415, 1738
Pracy M. B., et al., 2016, MNRAS, 460, 2
Rhee J., et al., 2023, MNRAS, 518, 4646
Sabater J., et al., 2021, A&A, 648, A2
Sanders D. B., et al., 2007, ApJS, 172, 86
Sawicki M., et al., 2019, MNRAS, 489, 5202
Schinnerer E., et al., 2010, ApJS, 188, 384
Scodeggio M., et al., 2018, A&A, 609, A84
Scoville N., et al., 2007, ApJS, 172, 1
Shimwell T. W., et al., 2026, A&A, 707, A198
Shirley R., et al., 2019, MNRAS, 490, 634
Smith D. J. B., et al., 2021, A&A, 648, A6
Smolčić V., et al., 2017a, A&A, 602, A1
Smolčić V., et al., 2017b, A&A, 602, A2
Smolčić V., et al., 2017c, A&A, 602, A6
Sutherland W., Saunders W., 1992, MNRAS, 259, 413
Tasse C., Best P. N., Röttgering H., Le Borgne D., 2008, A&A, 490, 893
Tasse C., et al., 2021, A&A, 648, A1
Thykkathu N. J., et al., 2026, arXiv e-prints, p. arXiv:2601.14913
Vaccari M., 2022, in Zenodo dataset. Zenodo, p. 6368348, doi:10.5281/zenodo.6368347
Vaccari M., 2026, arXiv e-prints, p. arXiv:2605.07998
Vaccari M., et al., 2016, in The 4th Annual Conference on High Energy Astrophysics in Southern Africa (HEASA 2016). p. 26 (arXiv:1704.01495), doi:10.22323/1.275.0026
Varadaraj R. G., Bowler R. A. A., Jarvis M. J., Adams N. J., Häußler B., 2023, MNRAS, 524, 4586
Varadaraj R. G., et al., 2026, MNRAS, 547, stag473
Vardoulaki E., et al., 2021, A&A, 648, A102
Vernstrom T., Scott D., Wall J. V., Condon J. J., Cotton W. D., Kellermann K. I., Perley R. A., 2016, MNRAS, 462, 2934
White S. V., Jarvis M. J., Häußler B., Maddox N., 2015, MNRAS, 448, 2665
White S. V., et al., 2025, MNRAS, 541, 3636
Whiting M., Humphreys B., 2012, Publ. Astron. Soc. Australia, 29, 371
Whittam I. H., Prescott M., McAlpine K., Jarvis M. J., Heywood I., 2018, MNRAS, 480, 358
Whittam I. H., et al., 2022, MNRAS, 516, 245
Whittam I. H., et al., 2024, MNRAS, 527, 3231
Whittam I. H., et al., 2025, MNRAS, 543, 507
Williams W. L., et al., 2019, A&A, 622, A2
Wilman R. J., et al., 2008, MNRAS, 388, 1335
Wong O. I., et al., 2025, MNRAS, 536, 3488
Zou F., et al., 2022, ApJS, 262, 15
de Jong R. S., et al., 2019, The Messenger, 175, 3
de Jong J. M. G. H. J., et al., 2024, A&A, 689, A80
van Haarlem M. P., et al., 2013, A&A, 556, A2
van der Vlugt D., et al., 2021, ApJ, 907, 5

APPENDIX A: CROSS-MATCHED CATALOGUE COLUMNS

We present the first ten lines of the catalogues released with this work in Table A1. Below is a descriptions of the columns listed:

- **`Source_Name`** - A source name for the object given by the RA and Dec in the form JHHMMSS.SS±DDMMSS.S with a prefix of MGTDR1. The RA/Dec used for the source name is taken as the host position where available and, if unavailable, the radio position instead.
- **`RA_radio`** - The RA of the radio centre in degrees. This is taken as the position from the PyBDSF catalogue, or the flux weighted centroid for objects matched within the zoo.
- **`Dec_radio`** - The Dec of the radio centre in degrees (as above)
- **`RA_host`** - The RA (in degrees) of the cross-matched host galaxy. If no host was assigned (either due to being in a masked region or as there was no consensus) this is denoted as -99.
- **`Dec_host`** - The Dec (in degrees) of the cross-matched host galaxy (or -99 if no host, as above)
- **`RA_host_newcat`** - The RA (in degrees) of the host galaxy (if available) in the updated catalogues of Stylianou et al. (in prep) matched within 1 arcsec, see Section 4.3.
- **`Dec_host_newcat`** - The Dec (in degrees) of the host galaxy (if available) in the updated catalogues of Stylianou et al. (in prep) matched within 1 arcsec, see Section 4.3
- **`Flux`** - The MIGHTEE integrated flux density of the source taken

from PyBDSF, or the summed flux density of the objects combined in the zoo
- **Flux_err** - The error in the MIGHTEE integrated flux density of the source from PyBDSF, or through the propagation of uncertainties from PyBDSF of the objects combined together
- **Peak_flux**- The MIGHTEE peak flux density of the source taken from PyBDSF or the maximum peak flux density of the objects combined in the zoo
- **Peak_flux_err** - The MIGHTEE peak flux density error of the source taken from PyBDSF or the corresponding error associated with the maximum peak flux density of the objects combined in the zoo
- **match_method** - Method used to match sources given by either: 'zoo' for sources matched in the MIGHTEE zoo project; 'LR' for sources matched using the Likelihood ratio method; 'gaussian' for sources split into their individual Gaussian components. Sources which do not have a host are assigned a value of 'NoMatch' but were all sent to the zoo for processing and those which were within masked regions are denoted as 'None'. For those sources which were changed during the final checks on the catalogue (see Section 4.2), these were flagged as 'Post'.
- **Size** - Size of the source (in arcsec) from the zoo project as discussed in Williams et al. (2019) or if the source is from the original PyBDSF catalogue, then the 'Maj' column from PyBDSF is adopted
- **Art_prob** - Probability of being an artefact from the MIGHTEE zoo project
- **Blend_prob** - Probability of being a blended source from the MIGHTEE zoo project
- **Zoom_prob** - Probability of being too zoomed in from the MIGHTEE zoo project
- **Hostbroken_prob** - Probability of having a multi-wavelength host which was split into multiple sources from the MIGHTEE zoo project
- **Imagemissing_prob** - Probability of having a multi-wavelength image missing when looked at in the MIGHTEE zoo project
- **Badclick** - Clicks to a host which were not in the catalogue cross-matched to in MIGHTEE zoo
- **z_BEST_peak** - Best photometric redshift from the catalogues of Stylianou+ in prep, taken as the peak in the redshift pdf (using galaxy templates where template fitting alone)
- **z_phot_hpd68_low** - Lower 68th percentile around z_BEST_peak
-**z_phot_hpd68_high** - Upper 68th percentile around z_BEST_peak
- **z_phot_percentile_50** - 50th percentile of the redshift pdf of Stylianou+ in prep
- **z_phot_percentile_16** - 16th percentile of the redshift pdf of Stylianou+ in prep
- **z_phot_percentile_84** - 84th percentile of the redshift pdf of Stylianou+ in prep
- **z_pau** - Redshift from the PAU catalogue Alarcon et al. (2021b) (COSMOS only)
- **z_phot** - Best photometric redshift. For CDFS-DEEP/XMM-LSS this is the same as z_BEST_peak, for COSMOS it is z_BEST_peak else a PAU redshift was available, in which case it is z_pau
- **z_spec** - Spectroscopic redshift, if available
- **Origin_z_spec** - Flag to describe the origin of the spectroscopic redshift, where available. These redshifts were all taken from the compilation of redshifts from Vaccari (2022)[12]. The possible labels are as follows: ACES-120604 (Cooper et al. 2012); BLAST-SPECZ (Eales et al. 2009; Moncelsi et al. 2011); COSMOS-V1.0 (Khostovan et al. 2025); DDFDATA-V1 (Zou et al. 2022); DEIMOS-10K (Hasinger et al. 2018); DESI-DR1 (DESI Collaboration et al. 2025); GAMA-DR4 (Driver et al. 2022); GAMA-G10-DR1 (Davies et al. 2015); HSU14 (Hsu et al. 2014); ICL-WHT (Patel et al. 2011); MAO12 (Mao et al. 2012); the NASA/IPAC Extragalactic Database (NED; the identifier includes a date label of YYYYMMDD for reference); OzDES-DR2 (Lidman et al. 2020); PRIMUS-DR1 (Coil et al. 2011; Cool et al. 2013); SDSS-DR16 (Ahumada et al. 2020); UDS-101018 and UDSz-140325 compilations from Alamani et al. of 10/10/2018 and 14/03/2025 respectively from UDS using data from Bradshaw et al. (2013); McLure et al. (2013) and Maltby et al. (2016) ; VIPERS-PDR2 (Guzzo et al. 2014; Garilli et al. 2014; Scodeggio et al. 2018); VVDS-02HR-DEEP-UDEEP (Le Fevre et al. 2013); VVDS-CDFS-DEEP (Le Fevre et al. 2013); hCOSMOS (Damjanov et al. 2018); zCOSMOS-BRIGHT-DR3 (Lilly et al. 2007).
- **z_best_final** - The best redshift available for the source from either z_phot or, if it is available, z_spec.
- **Origin_z_best_final** - Flag to attribute where the redshift is taken from. This is given as one of: None; Specz; Pau_z; Photoz
- **EffFreq** - Effective frequency at the source position in MHz
- **L1400** - 1.4 GHz scaled luminosity in WHz^{-1} (assuming a spectral index, $\alpha = 0.7$). Given as -99 if no redshift was available for the source
- **Mask_2025_08** - Flag to indicate whether a source was within the cross-matching area (given by a value of 1) or was within the masked region and thus not cross-matched (given by a value of 0) when the MIGHTEE zoo was started
- **Mask_Matching** - Flag to indicate whether a source was within the masked area of the catalogues of Stylianou et al. (in prep). Sources in the masked region are denoted by 0 and those in the cross-match region are given a value of 1
- **Mask_Final** - Combined flag to indicate whether a source was within the area in which cross-matching took place (given by a value of 1 - taken as sources with both Mask_2025_08 and Mask_Matching ==1) or was within the masked region and thus not cross-matched (given by a value of 0)
- **N_in_src** - Number of initial PyBDSF sources which have been combined together in this final catalogue

[12] using the compilation from 20th March 2025, see also Vaccari (2026).

Source_Name	RA_radio *deg*	Dec_radio *deg*	RA_host *deg*	Dec_host *deg*	RA_host_newcat *deg*	DEC_host_newcat *deg*	Flux *Jy*	Flux_err *Jy*
MGTDR1J095852.80+022603.0	149.719936	2.434258	149.720009	2.434193	149.720008	2.434201	0.000020	0.000005
MGTDR1J095852.80+023833.6	149.720004	2.642668	-99	-99	-99	-99	0.000034	0.000006
MGTDR1J095852.83+021853.2	149.720004	2.314844	149.720151	2.314803	149.720149	2.314807	0.000034	0.000005
MGTDR1J095852.77+021713.3	149.720010	2.286729	149.719886	2.287034	149.719893	2.287033	0.000017	0.000006
MGTDR1J095852.85+021933.6	149.720129	2.326008	149.720212	2.326001	149.720218	2.325999	0.000038	0.000006
MGTDR1J095852.84+012429.4	149.720175	1.408176	-99	-99	-99	-99	0.000184	0.000012
MGTDR1J095852.89+023004.3	149.720244	2.500956	149.720405	2.501200	149.720413	2.501209	0.000024	0.000006
MGTDR1J095852.87+015305.0	149.720251	1.884682	149.720329	1.884732	149.720338	1.884725	0.000043	0.000009
MGTDR1J095852.87+025917.6	149.720307	2.988232	-99	-99	-99	-99	0.000044	0.000010
MGTDR1J095852.90+024237.0	149.720341	2.710303	149.720436	2.710278	149.720448	2.710277	0.000054	0.000007

Peak_flux *Jy/beam*	Peak_flux_err *Jy/beam*	match_method	Size *(arcsec)*	Art_prob	Blend_prob	Zoom_prob	Hostbroken_prob	Imagemissing_prob	Badclick
0.000022	0.000003	LR	5.071870	-99	-99	-99	-99	-99	-99
0.000033	0.000003	None	5.609934	-99	-99	-99	-99	-99	-99
0.000028	0.000003	LR	6.118757	-99	-99	-99	-99	-99	-99
0.000014	0.000003	LR	6.132972	-99	-99	-99	-99	-99	-99
0.000031	0.000003	LR	5.893473	-99	-99	-99	-99	-99	-99
0.000170	0.000007	None	5.507381	-99	-99	-99	-99	-99	-99
0.000019	0.000003	LR	6.423954	-99	-99	-99	-99	-99	-99
0.000019	0.000003	zoo	9.601028	0	0.200000	0	0	0	0
0.000043	0.000006	None	5.706223	-99	-99	-99	-99	-99	-99
0.000050	0.000004	LR	5.559336	-99	-99	-99	-99	-99	-99

Z_BEST_peak	z_hpd68_low	z_hpd68_high	z_p50	z_p16	z_p84	z_pau	z_phot	z_spec	Origin_z_spec
0.118013	0.095011	0.139015	0.117116	0.094519	0.138905	-99	0.118013	0.124988	NED-20250320
-99	-99	-99	-99	-99	-99	-99	-99	-99	-99
1.093121	1.036115	1.145127	1.087098	1.021371	1.134961	1.154000	1.154000	-99	-
0.654073	0.619069	0.690077	0.655507	0.620813	0.691617	0.762000	0.762000	-99	-
2.294255	1.833204	2.766307	2.304621	1.842710	2.776072	-99	2.294255	-99	-
-99	-99	-99	-99	-99	-99	-99	-99	-99	-99
1.818202	1.754195	1.859207	1.799471	1.727503	1.845561	-99	1.818202	0.485916	NED-20250320
1.490166	1.219135	1.782198	1.493301	1.198593	1.762957	-99	1.490166	-99	-
-99	-99	-99	-99	-99	-99	-99	-99	-99	-99
0.870097	0.836093	0.899100	0.865833	0.831762	0.895630	0.889000	0.889000	0.870000	NED-20250320

z_best_final	Origin_z_best_final	EffFreq *MHz*	L1400 WHz^{-1}	Mask_2025_08	Mask_Matching	Mask_final	N_in_src
0.124988	Specz	1230.390000	7.359214e+20	1	1	1	1
-99	None	1210.980000	-99	0	0	0	1
1.154000	Pau_z	1235.130000	1.819307e+23	1	1	1	1
0.762000	Pau_z	1236.200000	3.602411e+22	1	1	1	1
2.294255	Photoz	1234.660000	9.909803e+23	1	1	1	1
-99	None	1210.690000	-99	0	0	0	1
0.485916	Specz	1225.290000	1.716798e+22	1	1	1	1
1.490166	Photoz	1234.580000	4.188902e+23	1	1	1	1
-99	None	1179.160000	-99	0	0	0	1
0.870000	Specz	1207.250000	1.482078e+23	1	1	1	1

Table A1. Example 10 rows from the catalogue released alongside this work. Explanations of each column are given in Appendix A.

[1] *Institute for Astronomy, University of Edinburgh, Royal Observatory Edinburgh, Blackford Hill, Edinburgh, EH9 3HJ, UK*
[2] *Astrophysics, Department of Physics, Denys Wilkinson Building, University of Oxford, Keble Road, Oxford, OX1 3RH, UK*
[3] *Department of Physics and Astronomy, University of the Western Cape, Robert Sobukwe Road, 7535 Bellville, Cape Town, South Africa*
[4] *Department of Physics, Astronomy and Mathematics, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK*
[5] *Inter-University Institute for Data Intensive Astronomy, Department of Astronomy, University of Cape Town, 7701 Rondebosch, Cape Town, South Africa*
[6] *INAF - Istituto di Radioastronomia, via Gobetti 101, 40129 Bologna, Italy*
[7] *Department of Astronomy, University of Cape Town, 7701 Rondebosch, Cape Town, South Africa*
[8] *Dipartimento di Fisica e Astronomia, Università di Bologna, Via P. Gobetti 93/2, 40129 Bologna, Italy*
[9] *UNISA Centre for Astrophysics and Space Sciences, College of Science, Engineering and Technology, University of South Africa, Cnr Christian de Wet Rd and Pioneer Avenue,Florida Park, 1709 Roodepoort, South Africa*
[10] *Green Bank Observatory, P.O. Box 2, Green Bank, WV 24944*
[11] *Victoria University of Wellington, Wellington, New Zealand*
[12] *South African Astronomical Observatory, PO Box 9, Observatory 7935, South Africa*
[13] *Institute for Fundamental Physics of the Universe (IFPU), Via Beirut 2, I-34151 Trieste, Italy*
[14] *SISSA - International School for Advanced Studies, Via Bonomea 265, 34136 Trieste, Italy*
[15] *INAF - Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, I-34131 Trieste, Italy*
[16] *INFN – National Institute for Nuclear Physics, Via Valerio 2, I-34127 Trieste, Italy*
[17] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[18] *Centre for Space Research, North-West University, Potchefstroom 2520, South Africa*
[19] *National Institute for Theoretical and Computational Sciences (NITheCS), Potchefstroom 2520, South Africa*
[20] *INAF, Osservatorio Astrofisico di Catania, Via Santa Sofia 78, 95123 Catania, Italy*
[21] *University College London, Department of Physics and Astronomy, Gower Street, London, WC1E 6BT, UK*
[22] *Max Planck Institute for Radio Astronomy, Auf dem Hügel 69, 53121 Bonn, Germany*
[23] *Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, School of Natural Sciences, The University of Manchester, Manchester, M13 9PL, UK*
[24] *Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía s/n, 18008 Granada, Spain*
[25] *National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China*
[26] *Guizhou Radio Astronomical Observatory, Guizhou University, Guiyang 550000, China*
[27] *INAF-Osservatorio di Astrofisica e Scienza dello Spazio, Via Gobetti 93/3, 40129, Bologna, Italy*
[28] *Department of Physics, University of Antananarivo, PO Box 906, Antananarivo 101, Madagascar*
[29] *IAASARS, National Observatory Athens, Lofos Nymfon, 11852, Athens, Greece*
[30] *Thüringer Landessternwarte, Sternwarte 5, 07778 Tautenburg, Germany*
[31] *SKA Observatory, Jodrell Bank, Lower Whitington, Macclesfield, SK11 9FT, UK*
[32] *The Observatories of the Carnegie Institution for Science, 813 Santa Barbara Street, Pasadena, CA 91101, USA*

This paper has been typeset from a TEX/LATEX file prepared by the author.