
The Properties of Hitherto Unidentified Radio Sources in the MIGHTEE Survey

*Submitted to the University of Hertfordshire in partial fulfilment of the requirement of
the degree of Master of Science by Research*

Author:

Monika GELAZIUTE

Supervised by:

Dr. Daniel J.B. SMITH

Dr. Alyssa DRAKE

Dr. Carolyn DEVEREUX

Centre for Astrophysics Research
Department of Physics, Astronomy and Mathematics
University of Hertfordshire
September 2022

Abstract

Radio astronomy is a pioneering branch of astronomy research - new radio surveys have gathered a lot of attention in the last few years, due to the huge expansion in capabilities of new radio facilities, with new observatories such as MeerKAT and LOFAR surveying wide areas to large depths at great speed. But radio sources alone are not especially informative, e.g. no redshift information can be obtained from radio continuum data, and therefore no clear route to obtaining information about the physics in the sources. This work investigates the properties of the no-ID radio sources taken from the MIGHTEE catalogue produced by Prescott et al. (submitted), specifically in the COSMOS field. The catalogue we use for the NIR data is described by McCracken et al. (2012) the UltraVISTA survey with deep and ultra-deep coverage of the COSMOS field. Combining the data from these two catalogues we are able to determine that the 114 no-ID sources have 141.8 ± 31.4 excess galaxies when the separation is up to $6''$. We find that these sources have a flux density distribution of 0.007 - 1.840 mJy, K_S magnitude distribution of mainly $20 < m_K < 24$ and a redshift distribution of $0 < z < 2$. The magnitude distribution includes only the sources we can detect, thus the rest must be fainter. We therefore suspect a bimodal magnitude distribution for the no-ID counterparts. Some fraction of the sources also have unknown redshifts because we have not been able to detect any counterparts even in the UltraVISTA data reaching $m_K < 26$. Our results suggest that data such as from VIDEO are as good for cross-identifying MIGHTEE sources as UltraVISTA data, considering there is no significant difference between the source population when $m_K < 24$ and $m_K < 26$, which may benefit MIGHTEE investigations in the XMM-LSS, ELAIS-S1 and ECDFS fields. To further our investigation, we look at how many of the 114 sources would have at least one possible multi-wavelength counterpart using part of the Likelihood Ratio technique. We find that 60.3 ± 16.1 per cent or around 68 of the no-ID sources have at least one counterpart detected in the UltraVISTA data. We also find that at search radius of $10''$, there are still three sources with no possible counterparts. We agree with other works in the literature that combining different methods of cross-identification may be the most efficient and rewarding approach for radio-continuum surveys.

Declaration

I declare that no part of this work is being submitted concurrently for another award of the University or any other awarding body or institution. This thesis contains a substantial body of work that has not previously been submitted successfully for an award of the University or any other awarding body or institution.

Except where indicated otherwise in the submission, the submission is my own work and has not previously been submitted successfully for any award.

Acknowledgements

I would like to express my warmest gratitude to my supervisor Dr. Daniel J. B. Smith for offering me his exceptional tutelage not only with this project but also for all my years at the University of Hertfordshire. I have been lucky to work with him for four years and when times were exceptionally hard I always knew I could count on him. I would also like to extend my thanks to my second supervisors Dr. Alyssa Drake, who, throughout this year, has helped me on various occasions and motivated me to keep going, as well as Dr. Carolyn Devereux, who has always offered her guidance. I would like to thank the University of Hertfordshire and the Centre for Astrophysics Research for giving me the opportunity to pursue a higher degree in Astrophysics and for being so hospitable throughout my studies here. I also want to show appreciation to the whole MIGHTEE team and everyone who worked on the MIGHTEE project, especially Dr. Imogen Whittam, who's been very helpful and offered advice. Last but certainly not the least, I would like to thank my family and friends for their continuous and endless support through my studies here, as it was their kind words of encouragement that kept me going in difficult times.

Contents

Abstract	i
Acknowledgements	iii
Contents	iv
List of Figures	vi
1 Introduction	1
1.1 Radio Astronomy	3
1.1.1 MeerKAT Telescope	5
1.2 Synchrotron Radiation	6
1.2.1 Active Galactic Nuclei	7
1.2.2 Supernova Remnants	10
1.3 Near-Infrared Astronomy	11
1.4 Multiwavelength Science	13
2 Data	15
2.1 MIGHTEE Survey	15
2.2 COSMOS Field	16
2.3 The Source Catalogues	16
2.3.1 MIGHTEE Early Science Data	17
2.3.2 Multiwavelength Catalogue	17
2.3.3 Photometric Redshifts	18
2.3.4 Cross-Matched Catalogue	18
2.3.5 K_S Band Data	25
3 Properties of the no-ID Radio Sources	28
3.1 Visual Inspection	28
3.2 Flux Density Distribution in the Radio Catalogue	29
3.3 The Multi-Wavelength Counterparts	30
3.3.1 Finding the Number of Excess K_S Band Sources	35
3.3.2 Magnitude Distribution of the no-ID Sources	35
3.3.3 Finding the Number of Excess z sources	37
3.3.4 Redshift Distribution of the no-ID Sources	39
3.3.5 Total Number of Excess Sources Detected	39
3.4 Fraction of Excess K Source Population Detected	40

3.5 Truly Unidentified Sources	43
4 Discussion	45
5 Conclusions	52
A Propagation of Uncertainty	56
A.1 The Uncertainty of the Number of Excess Galaxies	56
A.2 The uncertainty on the Q_0 value	58
Bibliography	59

List of Figures

1.1	The observed radio/FIR spectrum of M82 (depicted with the solid line), with the top x axis showing the wavelength in [cm], and the bottom x axis showing the frequency in [GHz]; the y axis here shows the flux density in [Jy]. The spectrum obtained is a sum of three components: free-free (dashed line), synchrotron (dot-dash line) and dust (dotted line) emission.	4
1.2	An image of MeerKAT receivers, which are part of the MeerKAT telescope based in South Africa. This image was taken by SARA0.	6
1.3	A giant radio galaxy found by Delhaize et al. (2021). This image was made by overlaying MIGHTEE data (in yellow) over a HSC g , r , and i bands composite optical image. Magenta contour line indicates the lowest contour level. The host galaxy may be seen in the inset, which depicts an enlarged view of the core region.	7
1.4	AGN unification scheme, adapted from Thorne et al. (submitted). This image depicts how changing the viewing angle can lead to different classifications of detected radio sources with an active nucleus, despite coming from the same underlying physical structure. It also underlines the type of emission that can be expected depending on the orientation of the source relative to the observer.	9
1.5	Example images of FRI and FR II radio sources. The left side image shows an example of the FRI CenA, while the right hand side shows 3C285, an FR II. Image adapted from (Baldi et al., 2019).	10
1.6	Displayed in this Murchison Widefield Array (MWA) radio image is a fresh perspective of the Milky Way galaxy. Small spherical bubbles seen mostly in the middle section of this image represent visible supernova remnants. Additionally, blue sections in the image are indicative of areas where massive stars are currently forming. The bright white area at the center of the image conceals the supermassive black hole situated at the core of our galaxy. The colours in this image represent different radio frequencies, with red and green showing the lowest and middle frequencies respectively, while blue represent the highest frequencies. Image credit: Natasha Hurley-Walker, ICRAR & Curtin / GLEAM Team.	11
1.7	A selection of filter transmission curves (shown in colour, from the horizontal axis), overlaid with an example spectral energy distribution (SED) of a galaxy at $z = 0$ and $z = 1$, taken from Smith et al. (2021). The SED shows that the near-infrared filters (J and K -bands with effective wavelengths of 1.2 and 2.2 μm , respectively) sample the old stellar population of galaxies at these redshifts.	12
2.1	The MIGHTEE COSMOS Early Science image in radio wavelength at 1.3 GHz (Heywood et al., 2021) with all 123 unmatched radio sources circled in green. The circles have a radius of 14.4". This image covers 1.6 deg^2 , however our data covers only 0.8 deg^2 , hence why there are no circled sources around the edge of the image. In this image north is up and east is left.	19

2.2	A figure illustrating sources that are all marked with different flags. Going from left to right for each column the flags are 100 - single comp., 101 - no match, 102 - too confused to split, 103 - junk/artefact, 105 - unmatched lobe, 112 - split flux and 120 - multiple comp. respectively. Flags are also indicated along the top row.	22
2.3	The first 48 sources out of the 123 with no matched counterparts. All these cutouts are $30'' \times 30''$ in size and have the radio contours from the MIGHTEE Early Science Data overlaid on top of the UltraVISTA K_s band image. The positions of the no-ID radio sources are also highlighted with red markers. The radio contours are equally log spaced with 12 levels ranging from $5.5 \mu\text{Jy}$ to 5.5 mJy , with $5.5 \mu\text{Jy}$ being the 1σ noise level of the image. North is up and east is left in all of these images.	23
2.4	The second assortment of 48 sources out of the 123 that have no identified counterparts. All these cutouts are $30'' \times 30''$ in size and have the radio contours from the MIGHTEE Early Science Data overlaid on top of the UltraVISTA K_s band image. The positions of the no-ID radio sources are shown with red markers. The radio contours are equally log spaced with 12 levels ranging from $5.5 \mu\text{Jy}$ to 5.5 mJy , with $5.5 \mu\text{Jy}$ being the 1σ noise level of the image. North is up and east is left in all of these images.	24
2.5	The third collection the last 27 sources out of the 123 with no matched counterparts. All these cutouts are $30'' \times 30''$ in size and have the radio contours from the MIGHTEE Early Science Data overlaid on top of the UltraVISTA K_s band image. The red markers also show the position of the no-ID radio sources. The radio contours are equally log spaced with 12 levels ranging from $5.5 \mu\text{Jy}$ to 5.5 mJy , with $5.5 \mu\text{Jy}$ being the 1σ noise level of the image. North is up and east is left in all of these images.	25
2.6	These plots present how we have measured the limiting magnitude. By applying a line of best fit to the K_s magnitude distribution, we were able to locate where the 50 per cent completeness limit lies. The value achieved is $m_K = 25.91$. . .	26
2.7	A histogram with both the full multi-wavelength catalogue of m_K distribution containing all possible cross-matches, and the magnitude distribution for those sources identified as counterparts of MIGHTEE sources in the COSMOS field.	27
3.1	Histograms for flux density ranges in all the subgroups of radio sources. Each panel contains flux density information of the whole radio catalogue (in blue) against each flag class (in magenta).	31
3.2	The m_K distribution in the whole UltraVISTA catalogue compared to the K_s band sources within $10''$ of a no-ID radio source. The full catalogue is shown in blue, whereas the sources in the search area are shown in magenta. The normalised catalogue is shown as a black line. It is clearly visible there are excess sources as the magenta histogram extends over the black line in most bins.	32
3.3	MOC plots of the optical catalogue (in purple), the search zone around no-ID sources (in green) and the intersection of these two areas (in yellow). Important to note here that some of the no-ID sources fall in the no optical coverage zone. These circular areas with no data are caused by very bright sources that would outshine the sources around them, this way making them invisible.	34

3.4	The K_S magnitude distribution near no-ID radio sources. The black represents the distribution of the background source population in the search area of $10''$, whereas the purple shows the actual K_S sources found in the search area. The excess sources can then be found by taking the black histogram away from the purple. This is shown in blue. It is visible that in most bins the excess is positive.	36
3.5	The background source population and the K_S sources found within the search area together with their respective uncertainties, shown as the gray and the purple shaded regions respectively. This allows us to see if the excess we found is significant.	36
3.6	A histogram of the excess K_S band sources around the no-ID radio sources within the $10''$ search area. All statistical excess is notably in the $20 < m_K < 24$ section with no significant excess beyond $m_K > 24$.	37
3.7	A histogram representing the redshift distribution of the K_S band sources. The black colour depicts the background source population, while the purple shows those sources that we found in our defined search area of $10''$. Blue shows the result we get after subtracting the background from the number of K_S sources in the search zone.	38
3.8	A plot representing the background source population in the search area together with the population of the K_S sources that are in this search area. In this plot, the uncertainties are also included to check if our results are of significant value.	38
3.9	The number of excess z sources around the no-ID MIGHTEE sources within the $10''$ search area. The majority of the population can be found within the $0 < z < 2$ values.	39
3.10	The number of excess sources detected around the no-ID MIGHTEE sources. The error bars derivation can be found in A.	40
3.11	The fraction of blanks against the search radius in the random and the no-ID catalogues. The random catalogue is shown in orange and the purple represents the no-ID radio sources. The shaded regions indicate their respective uncertainties.	42
3.12	The trend of Q_0 as a function of search radius. The peak of this plot is our best estimate for Q_0 , which peaks at just over $3''$.	43
4.1	In the upper left is the magnitude distributions of cross-matched K_S counterparts (in blue) and of the excess galaxies found around the no-ID sources in the COSMOS field (in magenta). The vertical black bars in the magenta plot represent the error bars for the excess galaxy number. In the upper right is the zoomed-in version of the same plot to inspect the excess sources distribution more closely. The lower plots are the same as the upper ones, but instead showing the redshift information.	48

Chapter 1

Introduction

With new radio surveys coming out in the last 5 years (such as Williams et al., 2019; Intema et al., 2017; Jarvis et al., 2017), and new radio telescopes being built (such as Square Kilometer Array (SKA) - Dewdney et al., 2009a; Weltman et al., 2020), radio astronomy is entering a new era of greatly increased capabilities, such as much better sensitivity, depth and spatial resolution of data (Smith et al. 2021). SKA Precursor and Pathfinder surveys are leading this field into far more sensitive and better observations. Radio observations serve as a uniquely powerful tool in astronomy, since they can provide a view of the universe that is unobscured by the cosmic dust and can also detect neutral HI gas that is invisible in optical wavelengths (Varga et al., 2012; Kondapally et al., 2021). Radio emission can be caused by the cores of Active Galactic Nuclei (AGN; White et al. 2015, Whittam et al. 2016, Heywood et al. 2021) and their jet-driven lobes (Laing et al. 2011, Fanaroff et al. 2021). Emitted in the form of synchrotron radiation (which is generated by relativistic electrons spiraling through magnetic fields) or thermal emission at higher radio frequencies, it can also be used to detect star formation in regular galaxies (Condon 1992, Jarvis et al. 2010, Murphy et al. 2017, Delvecchio et al. 2021). On bigger Mpc scales, radio observations can reveal diffuse radio haloes that trace hot gas in galaxy clusters (van Weeren et al. 2019).

While observations at long radio wavelengths reveal some insightful features about the galaxies we would not otherwise be able to measure, observations of the same sources at other wavelengths can help us determine other useful characteristics of these extragalactic sources, such as

their optical luminosities, stellar masses and most importantly - redshifts (Prescott et al., submitted). One of the most recent and state-of-the-art surveys based in the southern hemisphere is called MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE; more about this survey can be found in Sec. 2.1). However, there are about 2% of the MIGHTEE sources that do not have any identified multi-wavelength counterparts. These sources are most likely a mixture of high redshift and very dusty sources, suggesting that they could be very distant (i.e., high redshift) AGN or star forming (SF) galaxies with high star formation rates (SFRs). In both cases, those are really extreme sources, therefore hard to be explained by models or constraints of the universe. In this project we will investigate these MIGHTEE sources in detail, both individually and all-together using statistical techniques. We intend to shed some light on what these sources might be and what their properties are.

The reason we have chosen this project is because it gives us an invaluable chance to explore one of the deepest and best covered regions available to us, this way giving us an opportunity to determine the properties of the faint radio source population, all while working with radio sources that will be regularly monitored by the SKA. Previous studies have also investigated the properties of optically faint or undetected radio sources. Varga et al. (2012) used stacking techniques to reveal a strongly reddened, faint active galactic nucleus population. Their goal was to find optical emission from isolated, unresolved radio sources of the VLA FIRST survey (Very Large Array Faint Images of the Radio Sky at Twenty-cm; Becker et al., 1994) that have no known optical counterparts in the Sloan Digital Sky Survey (SDSS) Stripe 82 co-added data set (Varga et al., 2012). As they point out, there has been discussion about the nature of the optically undetected fraction for a long time. Infrared-faint radio sources (IFRSs) have also been investigated by Maini et al. (2016); Simpson (2017); Singh et al. (2017) and Orenstein et al. (2019). The goal of these projects was to study the nature of the IFRSs and their redshift distribution. As they conclude, all these sources have high redshift, and majority of them contain AGN or are starburst galaxies. Orenstein et al. (2019) investigate the redshift distribution of IFRSs and their findings suggest that IFRSs might be a subset of a larger class of high- z radio galaxies (HzRGs).

1.1 Radio Astronomy

Astronomers have used radio waves to explore the universe for almost 100 years now. It is as a result of radio astronomy that we have discovered cosmic microwave background (CMB) radiation (Penzias and Wilson, 1965; Dicke et al., 1965). Karl Jansky first reported detecting radio radiation coming from the Milky Way in 1933 (Jansky, 1933). Radio astronomy started gaining real traction in the middle of the twentieth century, when aperture synthesis and radio interferometry were developed (Djorgovski et al., 2013). This led to a plethora of new surveys being released, such as the 3C and 4C Cambridge surveys (Edge et al., 1959; Gower, 1966). Different sources quote varying ranges for radio frequency, which might be due to the rapid development of radio astronomy. Longair (2011) proposes a range of $3 \text{ MHz} \leq \nu \leq 30 \text{ GHz}$ and $100 \text{ m} \leq \lambda \leq 1 \text{ cm}$ for the radio waveband. Blind surveys have also become essential. The view of the radio sky is dependent on two things - sensitivity and frequency, and both have a strong influence on the type of source detected. For example, at high frequency it is possible to see flat spectrum sources that may not be detected at low frequencies (de Zotti et al., 2010), and conversely, at low frequencies steep spectrum sources may be visible, which cannot be detected at high frequencies. Both low and high frequency observations are therefore very valuable.

One of the main benefits that radio astronomy has is that it can see objects unobscured by dust. This makes radio observations an incredibly powerful tool, as we can observe galaxies in the distant universe and with a high redshift due to the large luminosity of radio sources (Seymour et al., 2007). This offers a unique window into an early universe and clues as to how galaxies formed. Due to the nature of radio waves, the dusty interstellar medium is practically transparent, as the dust grains are much smaller than the size of a radio wave. The size of a large dust grain, for comparison, is $1 \mu\text{m}$ (Mathis et al., 1977). This allows radio waves to pass through dusty environments undisturbed.

Radio spectra vary as a power law in frequency, such that $S_\nu \propto \nu^\alpha$, with S being the flux density, ν being the frequency, and α being the spectral index, which is typically -0.7 for normal star-forming galaxies. Some authors also choose to use the opposite sign convention for the spectral index. An example of a typical radio spectral energy distribution (SED) can be seen in Figure 1.1. This is an observed radio/FIR spectrum of M82, taken from Condon (1992). It is worth

noting that at frequencies near 1 GHz, synchrotron emission is predominant compared to free-free emission in normal galaxies, as is visible from this figure.

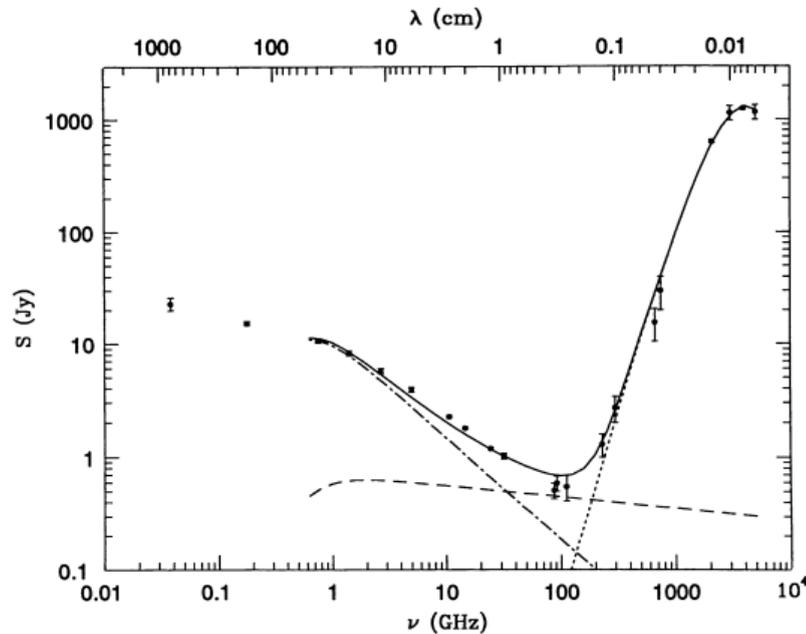


FIGURE 1.1: The observed radio/FIR spectrum of M82 (depicted with the solid line), with the top x axis showing the wavelength in [cm], and the bottom x axis showing the frequency in [GHz]; the y axis here shows the flux density in [Jy]. The spectrum obtained is a sum of three components: free-free (dashed line), synchrotron (dot-dash line) and dust (dotted line) emission.

Radio telescopes have large specialised antennas and receivers to collect emission coming from the sky. Radio telescopes can operate on their own or combined with multiple other radio telescopes - a technique called radio interferometry (Thompson et al., 2017). Using multiple telescopes together works in a similar way to using an optical telescope with a much bigger aperture, as the resolution is defined by the distance between the telescopes (Lazio, 2002). Using radio interferometry allows for a much higher angular resolution than using just one telescope (Monnier and Allen, 2013). Since radio wavelengths are the longest ones on the EM spectrum, and angular resolution depends on the aperture in proportion to the wavelength observed ($\propto \lambda/D$), radio telescopes need to be much larger than optical telescopes to achieve the same spatial resolution images (Thompson et al., 2017).

Interferometry makes use of aperture synthesis - a process that increases the resolution of the telescope to the size of an angular resolution that would be achieved if one had a telescope as large as the distance between two antennas that are furthest away from each other (Burke and Graham-Smith, 2009). It does so by superposing incoming signals from different antennas, as

waves with the same phases will become bigger at peaks, and waves with the opposite phases will cancel each other out (a process called constructive/destructive wave interference; Kipnis, 1991). Each telescope is connected via some type of transmission line, i.e. optical fiber or coaxial cable. This allows for an increase in total signal amassed, as well as for aperture synthesis. To achieve the best quality data, many different separations between antennas are necessary (Swenson and Mathur, 1968). These separations are called baselines. In order to map out the image (as interferometers do not produce optical images as optical telescopes do), Fourier transforms of the brightness distribution of the object are used (Burke and Graham-Smith, 2009).

Radio observations from Earth are limited due to numerous different factors. At longer wavelengths or short frequencies, the ionosphere does not transmit all waves, since it reflects those waves that have lower frequencies than its characteristic plasma frequency (Bougeret et al., 2008). Most radio observatories are set up at very dry and high locations, as water vapour interferes with higher frequency radio emission (such as the Gigahertz range; Hafez et al., 2019). Terrestrial devices emitting radio frequencies can interfere with observations as well, which is another reason for radio observatories to be built in remote locations (Condon and Ransom, 2016).

1.1.1 MeerKAT Telescope

MeerKAT began operating in 2018 and consists of 64 interlinked radio antennas, all located in the Meerkat National Park, which is a semi-desert region in Southern Africa (Mauch et al., 2020). Each antenna is rigged with cryogenic receivers and the dishes are 13.5 m in diameter. The minimum baseline for MeerKAT is 29 metres and the maximum is 8 kilometers. There are three frequency bands, together ranging from 0.58 GHz to 14.5 GHz (Jonas and MeerKAT Team, 2016). The 64 dishes are distributed in a way that 70 per cent of the dishes are in the inner component, having a mean dispersion of 300 metres; and the outer component with the remaining 30 per cent of the dishes has them allocated so that the mean dispersion is 2 500 metres. All the dishes are separated such that they have a Gaussian distribution (Mauch et al., 2020). MeerKAT offers a broad variety of observation modes that include polarisation and spectral line imaging, deep continuum, pulsar timing and transient searches (Jonas and MeerKAT Team, 2016).

Figure 1.2 shows some of the installed and operating MeerKAT antennas. This image was taken by and belongs to South African Radio Astronomy Observatory (SARAO).



FIGURE 1.2: An image of MeerKAT receivers, which are part of the MeerKAT telescope based in South Africa. This image was taken by SARAO.

1.2 Synchrotron Radiation

There are a lot of different sources in the universe that can cause radio detections. These objects tend to be some of the most energetic physical processes that we are aware of. Some examples that are relevant to this project and emit radio waves are radio galaxies (Jones et al., 2004), active galactic nuclei (AGN; Radcliffe et al., 2021) and supernova remnants (SNRs; Dubner and Giacani, 2015). With current sensitivity of data we can also detect fainter objects, such as star-forming galaxies and radio-quiet AGN (Best and Heckman, 2012; Whittam, 2018). The process responsible for most radio emission at GHz frequencies is synchrotron radiation (Longair, 2011). This non-thermal mechanism dominates much of high energy astrophysics. Some more examples of radio sources are given in the next section.

Synchrotron radiation is a result of high-energy (and often ultra-relativistic) electrons gyrating in a magnetic field and radiating away (Elder et al., 1947). The radiation highly relativistic electrons release is focused in the direction of their immediate motion. As a result, an observer notices the radiation from electrons whose orbital motion is on or near the plane where the observer is located (Thompson et al., 2017).

1.2.1 Active Galactic Nuclei

Radio galaxies are known as galaxies with very bright and powerful radio sources. These galaxies are 1 000 to 100 million times more bright in radio than Milky Way (Longair, 2011). Counter-intuitively, it is the objects in the galaxy that cause the radio emission and not the galaxy as a whole. According to Kauffmann et al. (2003), the host galaxies are found to fall within these three categories: disturbed/interacting galaxies, single disc galaxies and single blue spheroidal/amorphous galaxies. The radio sources are stemming from the active galactic nuclei, which reside in the centre of a host galaxy and emit powerful relativistic material in the form of jets (Shields, 1999). These jets can frequently extend far out of the host galaxy and reach sizes of over 2 Mpc, just like in the two giant radio galaxies recently found in the COSMOS field (Delhaize et al., 2021). One of these galaxies is shown in Figure 1.3. This was made by overlaying MIGHTEE data (yellow contours) on a Hyper Suprime-Cam (HSC; Aihara et al., 2019) g , r , and i bands composite optical image. The lowest contour level is accentuated in magenta to display the full size of the galaxy. The jets and the radio lobes are clearly visible here.

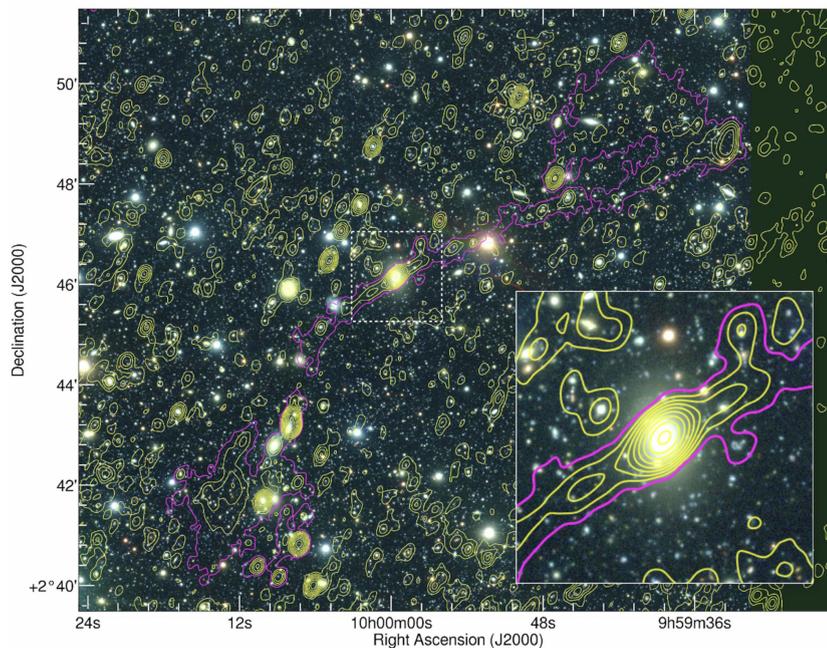
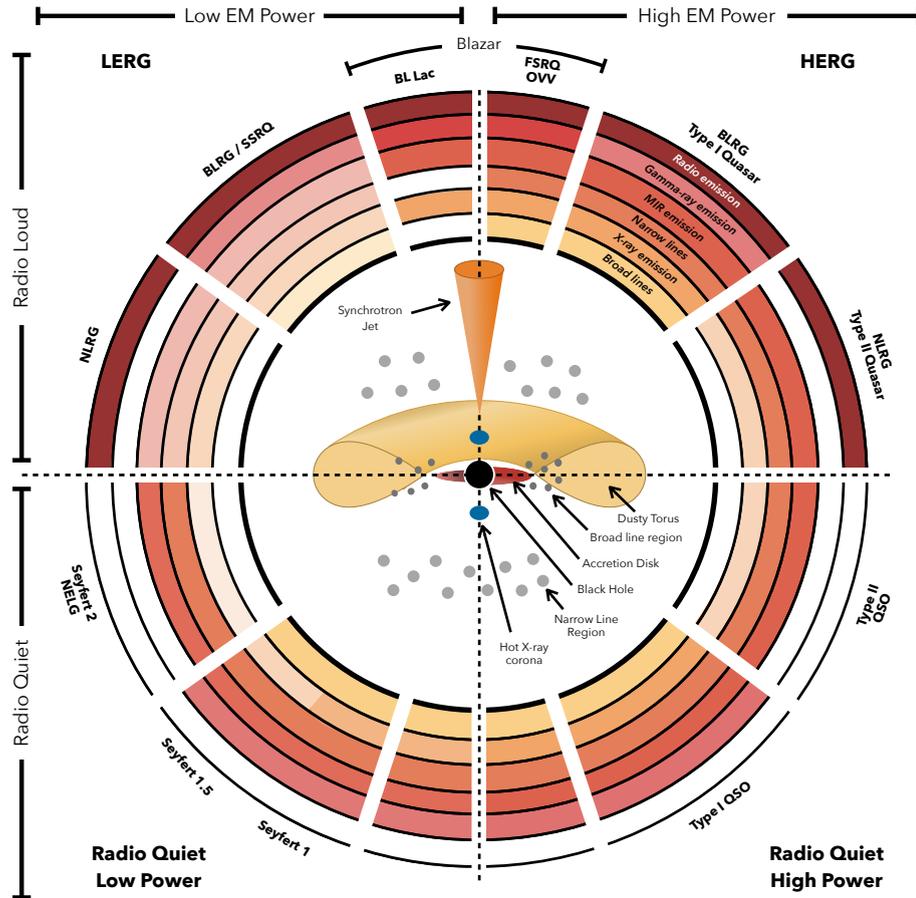


FIGURE 1.3: A giant radio galaxy found by Delhaize et al. (2021). This image was made by overlaying MIGHTEE data (in yellow) over a HSC g , r , and i bands composite optical image. Magenta contour line indicates the lowest contour level. The host galaxy may be seen in the inset, which depicts an enlarged view of the core region.

The reason AGN are an important topic is because they have an impact on the star formation rate in galaxies, and consequently the evolution of galaxies (Harrison, 2017; Scholtz et al., 2018), as they are thought to affect the properties of the intergalactic medium (Fabian, 2012; Shin et al., 2019). AGN are most likely powered by material accretion onto a super-massive black hole (size of $10^6 - 10^{10} M_{\odot}$) that, it is believed, resides in the centre of every galaxy (Lynden-Bell, 1969; Marconi and Hunt, 2003; Harrison, 2017). AGN are exceptionally helpful when studying distant objects and their evolution - as one of the objects we can detect the farthest into the universe, and one of the brightest sources we are aware of (Yu and Tremaine, 2002; Marconi et al., 2004), AGN shed light on how these objects appeared in the early universe, which allows us to use their evolution in the luminosity function as a constraint for theoretical models of our universe. The most luminous AGNs are classed as quasars. Their luminosity can reach 100 000 times the luminosity of the Milky Way (Frank et al., 2002).

There are many different observational manifestations of AGN activity in distant galaxies; unified models of AGN (Antonucci, 1993; Urry and Padovani, 1995; Netzer, 2015) rely on the idea that the same underlying physical structure, shown in Figure 1.4 (taken from Thorne et al., submitted), can explain a range of sources containing an active nucleus. Some of those are further divided up into additional categories. Some examples include Type I and Type II (Ramos Almeida et al., 2011), Radio-Loud and Radio-Quiet (Wilson and Colbert, 1995), High/Low-excitation Radio Galaxies (HERGs and LERGs; Smolčić and Riechers, 2011). Type I and Type II AGN are characterised by broad and narrow emission lines respectively. The broad lines are seen when looking down the long axis of the dusty torus, while the narrow lines are detected by observing gas clouds at larger distances (and therefore with narrower velocity profiles) that have been ionized by the strong radiation coming from the AGN (Netzer, 2015; Joh et al., 2021).

Radio-Loud (RL) and Radio-Quiet (RQ) AGN are differentiated by the presence or absence of large-scale radio jets and lobes (Wilson and Colbert, 1995), with RQ AGN being typically 1 000 fainter than RL AGN (Panessa et al., 2019). The main distinction between LERGs and HERGs is that the former lack strong emission lines in their optical spectra (Smolčić and Riechers, 2011). Other observable features of AGN depend on several factors, including the mass of the core black hole and the rate of gas accretion onto it, the orientation of the accretion disc, the degree of dust obscuration of the nucleus, and the existence or absence of jets (Burke and Graham-Smith, 2009).



J. E. Thorne

FIGURE 1.4: AGN unification scheme, adapted from Thorne et al. (submitted). This image depicts how changing the viewing angle can lead to different classifications of detected radio sources with an active nucleus, despite coming from the same underlying physical structure. It also underlines the type of emission that can be expected depending on the orientation of the source relative to the observer.

Unified models do not normally explain the Fanaroff–Riley Class I and Class II (FRI and FRII; Fanaroff and Riley, 1974) classes (examples of FRI and FRII sources can be seen in Figure 1.5). In FRIs, the jets are brightest near the nucleus, while in FRIIs, the peak brightness is at the very edge of the radio source (Burns et al., 1983; Alexander and Leahy, 1987).

Crucial to this work is how these different types of source appear in the near-infrared. While for radio galaxies and Seyferts the principal feature that dominates the K band is the star light (recall that these are some of the most massive galaxies in the Universe), for QSO-like sight-lines (i.e. close to face-on) the brightness is increased at all wavelengths by the emission from the accretion disk. Together with the reduced influence of dust relative to shorter wavelengths, this underscores the fact that K band is an excellent wavelength to choose, when searching for

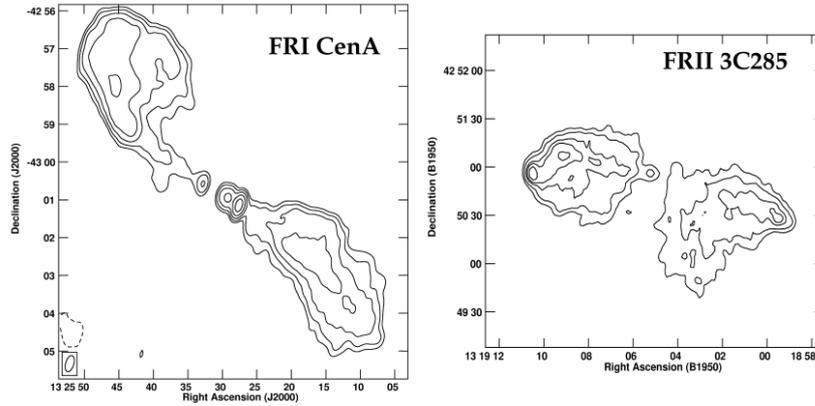


FIGURE 1.5: Example images of FRI and FR II radio sources. The left side image shows an example of the FRI CenA, while the right hand side shows 3C285, an FR II. Image adapted from (Baldi et al., 2019).

the counterparts of radio sources.

1.2.2 Supernova Remnants

Supernovae are highly-energetic events in which a star with a critical mass of $\geq 8 \pm 1 M_{\odot}$ (Smartt, 2009) explodes and its envelope is ejected at great velocity. The explosion is a very violent and luminous occurrence, with as much as $\sim 10^{46}$ J energy (Smartt, 2009; Janka, 2012) released into space. The optical luminosity it reaches can be as high as the light of a small galaxy. This can be told when examining the absolute magnitude distributions of the galaxies and some of the brightest supernovae: Type Ia supernovae reach a mean absolute magnitude of $M_R = -19.02$ (Papadogiannakis et al., 2019), while the absolute magnitude distribution for galaxies in R band is $-24 < M_R < -17$ (Fernández Lorenzo et al., 2012). This goes to show that the absolute magnitude of SNe falls within the absolute magnitude distribution of galaxies.

The expelled outer layers of the supernova can be seen for around 100 000 years after the explosion and are referred to as supernova remnants (Longair, 2011). The size of an SNR can reach tens of parsecs (Reich, 2002). A good example of how SNR show up in radio wavelengths can be seen in Figure 1.6 (Image credit: Natasha Hurley-Walker, ICRAR & Curtin / GLEAM Team). This image is a composition of data that was included by GaLactic and Extragalactic All-sky MWA survey (GLEAM) from the Murchison Widefield Array (MWA) in their 2019 data release. It contains 27 new SNRs, along with information about the star forming regions in this field, and the supermassive black hole in the centre of our galaxy (Hurley-Walker et al., 2019).

The SNRs can be seen as small spherical bubbles in this image, dotted around mainly in the middle section. The wide frequency range that GLEAM uses is what allowed them to achieve these results.

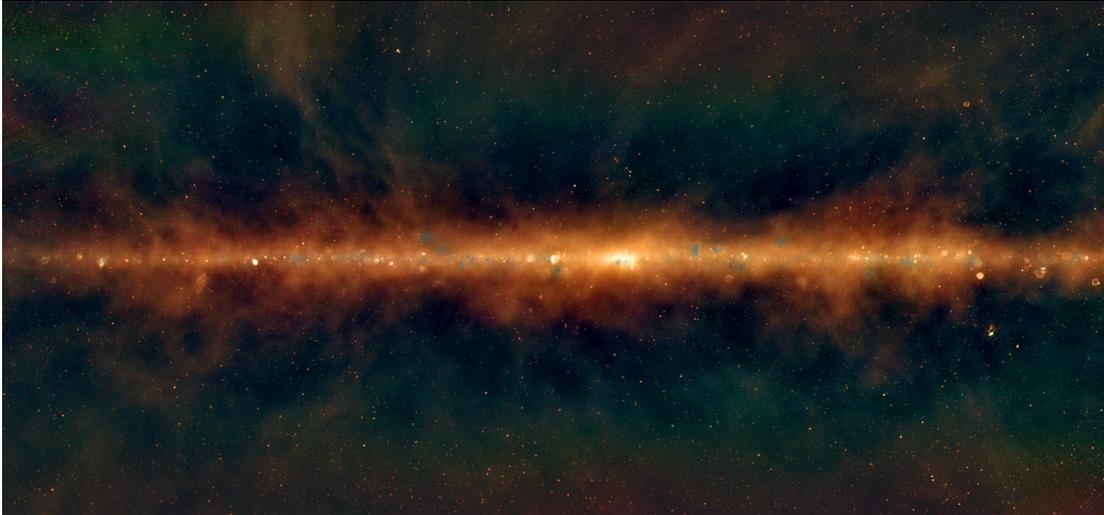


FIGURE 1.6: Displayed in this Murchison Widefield Array (MWA) radio image is a fresh perspective of the Milky Way galaxy. Small spherical bubbles seen mostly in the middle section of this image represent visible supernova remnants. Additionally, blue sections in the image are indicative of areas where massive stars are currently forming. The bright white area at the center of the image conceals the supermassive black hole situated at the core of our galaxy. The colours in this image represent different radio frequencies, with red and green showing the lowest and middle frequencies respectively, while blue represent the highest frequencies.

Image credit: Natasha Hurley-Walker, ICRAR & Curtin / GLEAM Team.

SNRs are very powerful sources of high-energy electrons (Reynolds, 2008). The energy released by these events can be measured and analysed, and is much higher than the required minimum energy for radio emission to be present (Longair, 2011). SNRs are a good means of tracing star formation, since the most massive stars that produce supernovae at the end of their lives are short-lived (in astrophysical terms), increased rates of supernovae are directly associated with a young stellar population (i.e. with star formation). In addition, SN explosions can trigger further star formation by sweeping up and compressing gas in the surrounding medium (Nagakura et al., 2009; Oser et al., 2010).

1.3 Near-Infrared Astronomy

The infrared wave-band is frequently divided into subdivisions, such as far-infrared (FIR), infrared (IR) and near-infrared (NIR). The main discrepancy between these bands is due to the

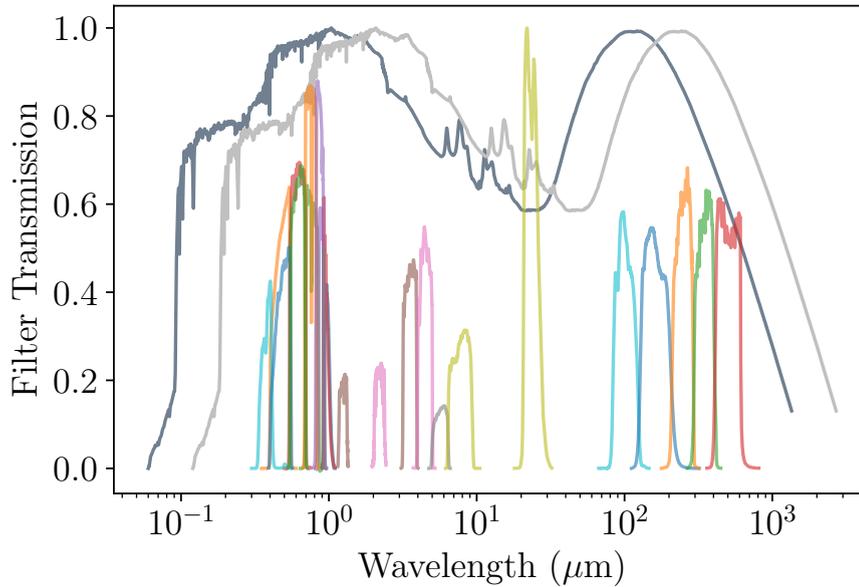


FIGURE 1.7: A selection of filter transmission curves (shown in colour, from the horizontal axis), overlaid with an example spectral energy distribution (SED) of a galaxy at $z = 0$ and $z = 1$, taken from Smith et al. (2021). The SED shows that the near-infrared filters (J and K -bands with effective wavelengths of 1.2 and $2.2 \mu\text{m}$, respectively) sample the old stellar population of galaxies at these redshifts.

historical baggage. Observations at wavelengths of $3 \mu\text{m}$ and less are usually near-infrared, meanwhile those produced at longer wavelengths are made in the thermal infrared waveband (IR and FIR) (Longair, 2011). Franceschini et al. (2008) quotes the NIR range as $1.00 - 10.00 \mu\text{m}$; McCracken et al. (2012) suggests this range to be $1.00 - 2.5 \mu\text{m}$. The structure of our Galaxy is clearly visible in images taken in the near-infrared waveband because it has greatly decreased interstellar extinction by interstellar dust particles relative to optical wavelengths (Hsu et al., 2019). Thus, NIR astronomy has a similar advantage to that of radio astronomy - a view of the sky that is less hindered by dust. NIR frequencies cover Y , J , H and K bands (Rosslowe and Crowther, 2017). K' ($\lambda_{eff} = 2.120 \mu\text{m}$), K_S ($\lambda_{eff} = 2.150 \mu\text{m}$) and K ($\lambda_{eff} = 2.200 \mu\text{m}$) are some of the subcategories for the K band (Tokunaga et al., 2002). Near-infrared astronomy allows us to see cooler redder stars with less influence of the obscuration of dust (see Figure 1.7 for an illustrative example). Red dwarfs and large red giant stars are predominant in this waveband.

Nonetheless, NIR observations are a challenging task. McCracken et al. (2012) states three reasons as to why that is the case, the first one being the extreme brightness of the sky background

in the NIR wavebands. A multiple of short exposures have to be used to avoid detector saturation on the sky, which can reach $15\text{mag}/\text{arcsec}^2$ brightness in AB magnitudes in the K_S band - . In return, this can be very expensive, as it requires more computational effort. Another reason is due to the fact that the sky background also varies with time, fluctuating to magnitudes that are brighter than the sources being observed. This means that the sky background has to be meticulously subtracted from each exposure before scientific analysis can be completed. The final reason has to do with high-priced detectors that need to be used instead of the regular silicon CCDs, since these are ineffective when it comes to NIR observations.

The data used in this project contains K_S band data taken from the UltraVISTA catalogue. More about this catalogue can be found in Section 2.3.5. The K_S waveband covers $2.00 - 2.31 \mu\text{m}$, as stated by Skrutskie et al. (2006). As well as the benefits of being less impacted by dust obscuration than shorter wavelengths, Simpson (2017) suggest that another reason for using K band imaging to identify the counterparts of radio galaxies is the ' $K - z$ ' relation (Willott et al., 2003). Since radio galaxies are typically the most massive galaxies in the universe, with similar absolute magnitudes irrespective of the redshift, we should expect an approximately linear increase in apparent magnitude with redshift; this is the $K - z$ relation.

1.4 Multiwavelength Science

The ability to identify radio source locations with great accuracy, so it would then be possible to match them with objects discovered in the optical and other regions of the electromagnetic spectrum, is crucial for advancing astronomy. This allows us to achieve the most important piece of information about astronomical sources, i.e. redshift. Without redshift we would not be able to determine distance, which is key for characterisation of sources. Redshifts can either be photometric or spectroscopic (Cimatti et al., 2002). Photometric redshifts are those obtained from using photometry and are less favoured over the spectroscopic redshifts due to being less accurate (it can have errors up to $\delta z = 0.5$ due to the wide wavelength ranges of the photometric filters and the required suppositions regarding the makeup of the spectrum at the light source; Bolzonella et al., 2000). Spectroscopic redshifts are obtained by using spectroscopic information of the source and comparing it to the laboratory spectra, as the atom absorption and emission lines are well-known and distinct (Brand, 1995).

The ability to assess variables like intensity, polarisation, and frequency spectrum with comparable angular resolution in both the radio and optical wavelengths is also essential (Thompson et al., 2017). Looking at a source in different EM frequencies can reveal different parts of it (Scoville et al., 2007). For example, UV band reveals very hot stars that may not be visible in the optical or near-infrared (Ortiz and Guerrero, 2016), and gamma-ray emission can detect accretion disks around black holes (Aharonian, 2004; Barkov et al., 2012).

While very useful on its own, no one wavelength can offer all the information about a single source and that is why multiwavelength science has proved to be indispensable in astronomy. By being able to measure different parameters in each waveband and then combine that information together, we can get a full picture of what is transpiring in the observed source (Middleton et al., 2017).

Cowie et al. (1990) says that by using a infrared-optical information of radio sources as a function of redshift, we can attain details of how elliptical galaxies formed, the evolution of galactic discs and the history of star formation in the universe as a whole. This statement is strengthened by the likes of (Varga et al., 2012; Patil et al., 2019; Smith et al., 2021, Prescott et al., submitted) and many more. It is clear from this that investigating the near-infrared/optical properties of previously unidentified radio sources can shed light on populations of sources that are otherwise unknown and further our knowledge of how the universe formed. Throughout this work, the cosmological parameters have adopted values of $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and $H_0 = 70 \text{ kms}^{-1} \text{ Mpc}^{-1}$.

Chapter 2

Data

2.1 MIGHTEE Survey

The data used in this project are mainly taken from the MeerKAT Large Survey Project called MIGHTEE (MeerKAT International GHz Tiered Extragalactic Exploration; Jarvis et al., 2017). Its main objective is to probe the cosmic evolution of galaxies and active galactic nuclei (AGN) by taking deep images of the sky. This survey covers over 20 deg^2 of sky visible from the southern hemisphere, at a resolution of $\sim 6 \text{ arcsec}$, with sensitivity averaging at $\sim 1 \mu\text{Jy}$ across the bandwidth of 900-1670 MHz, totaling in ~ 1000 hours of observation time (Jarvis et al., 2017). MIGHTEE surveys four extragalactic deep fields with some of the best multi-wavelength data in existence: XMM Large Scale Structure Survey (XMM-LSS; Pierre et al., 2004), The Cosmic Evolution Survey (COSMOS; Scoville et al., 2007), The European Large-Area ISO Survey (ELAIS-S1; Oliver et al., 2000) and Extended Chandra Deep Field-South Survey (ECDFS; Lehmer et al., 2005). As is suggested in the name, this survey uses data collected by the South African telescope MeerKAT (Meer-Karoo Array Telescope; Jonas, 2009), a precursor to the Square Kilometer Array (SKA), which will be fully operational by the end of this decade (Dewdney et al., 2009b).

MIGHTEE is designed as a tiered radio continuum deep imaging galaxy evolution survey (Maddox et al., 2021). The survey will help us understand the role of active galactic nuclei (AGN) and star formation in galaxy formation and evolution, and their dependence on stellar mass and environment over the cosmic time. It will also reveal how the neutral hydrogen progresses into

stars and how it fuels the AGN activity, as well as shed some light on how the cosmic magnetic fields emerge and evolve in galaxies and clusters, and what their properties are (Jarvis et al., 2017). With the unprecedented depths it will reach, MIGHTEE will be able to detect nearly every active radio galaxy to the very edge of the universe in every field (Heywood et al., 2021).

2.2 COSMOS Field

In this project, we have used the MIGHTEE data covering 1.6 square degrees of the COSMOS field (Heywood et al., 2021). These data were collected using a single deep pointing in the region. COSMOS is a deep, multi-wavelength survey centered at (J2000) RA = +150.119 and DEC = +2.206. Its target is to measure the evolution of galaxies on scales from a few kpc to 10s of Mpc. This field has been covered by all the accessible wavelengths from radio to X-ray with most of the major ground and space-based telescopes, such as VLA, ESO-VLT, Subaru and Hubble, Chandra, XMM etc. (Scoville et al., 2007; Elvis et al., 2009a). Some examples of the coverage in each EM spectrum region are as follows: Radio - Carilli et al. (2008), Smolčić et al. (2009); Sub-millimeter - Casey et al. (2013); far-IR - Lutz et al. (2011), Oliver et al. (2012); mid-IR - Fu et al. (2010); NIR - McCracken et al. (2012), Emerson and Sutherland (2010); Optical - Taniguchi et al. (2015); UV - Zamojski et al. (2007), Scoville et al. (2007); X-ray - Hasinger et al. (2007), Elvis et al. (2009b).

2.3 The Source Catalogues

The cross-matched catalogue that we are using was produced by using the 2021 PYBDSF (Python Blob Detection and Source Finder; Mohan and Rafferty, 2015) radio source catalogue, the production of which is described in detail in Heywood et al. (2021), and linking it with the multiwavelength catalogue for the COSMOS field, which was produced by Adams et al. (2020). All the sources in the catalogue were cross-matched by visual inspection carried out by a team of professionals in the field (Prescott et al., submitted). More about this can be found in Section 2.3.4.

2.3.1 MIGHTEE Early Science Data

The MIGHTEE Early Science release used MeerKAT to observe the COSMOS field for 25 hours, with each source observed for 17.45 hours. Two images were produced by first flagging and then calibrating the raw data. The ‘high’ resolution image has a resolution of 5” with a sensitivity of $5.5 \mu\text{Jy beam}^{-1}$, whereas the second image has a lower resolution of 8.6”, but a higher sensitivity of $1.7 \mu\text{Jy beam}^{-1}$. These properties are dictated by the maximum baseline and the collecting area of the telescope. The latter catalogue is what is being used for the cross-match catalogue described in Section 2.3.4. All the sources in this catalogue have a peak brightness that surpasses the local background noise by 5σ . Prescott et al. (submitted) has also exploited the observations made by VLA-COSMOS 3 GHz Large Project by Smolčić, V. et al. (2017). A much higher resolution of 0.78” with sensitivity of $2.3 \mu\text{Jy beam}^{-1}$, which corresponds to $3.7 \mu\text{Jy beam}^{-1}$ in 1.5 GHz frequency, has enabled the team to identify the cross-matches with better precision due to higher positional accuracy. Meanwhile MIGHTEE offers the possibility to expose faint emission from radio sources that are diffuse and extended. Together these two surveys complement each other and have allowed for a higher accuracy when it came to cross-matching sources with their multiwavelength counterparts (Prescott et al., submitted).

2.3.2 Multiwavelength Catalogue

The multiwavelength catalogue that was used for making the cross-matching was produced by Adams et al. 2020 (see also Adams et al. 2021). It combines data from many different surveys and consists of 13-band photometry. Optical images were taken from the HyperSuprimeCam Strategic Survey Programme (HSC SSP) and the Canada–France–Hawaii Telescope Legacy Survey (CFHTLS; Aihara et al., 2017a,b). The latter survey covers the u*-band, whereas the former one covers the *grizy* bands. The near-infrared data were collected from the UltraVISTA survey for the COSMOS field (McCracken et al., 2012). The infrared data were obtained from Spitzer Extended Deep Survey (Ashby et al., 2013), specifically the 3.6 and 4.5 micron bands. The image processing and catalogue creation are described in Bowler et al. (2020), but in short it was made by using the fourth data release (DR4) of the UltraVISTA survey, which contains deep data in *YJHK_s* bands. More about UltraVISTA and which band we used in this project can

be found in Section 2.3.5. All the objects have signal to noise ratio (SNR) ≥ 5 . This was computed using images with the background subtracted and with circular apertures of 1.8'' diameter. The common pixel scale for COSMOS field was $0.15'' \text{ pix}^{-1}$. The catalogues themselves were produced by stacking the $J+H$ and $H+K_s$ data and utilising them as the detection images, and then using ‘dual-image’ mode in SExtractor (Bertin, E. and Arnouts, S., 1996). This software is used to detect, measure, deblend and classify astronomical sources in the most optimal way.

2.3.3 Photometric Redshifts

The photometric redshifts for this multiwavelength catalogue were produced by using two different techniques, employing a hierarchical Bayesian approach. The first technique is to use the LEPHARE Spectral Energy Distribution (SED) fitting code (Arnouts et al., 1999; Ilbert, O. et al., 2006), which is a traditional template fitting technique. Together with this, a machine learning algorithm that uses the performance of Gaussian process photometric redshift estimation (Almosallam et al. 2016a, 2016b) is adopted. By using a hierarchical Bayesian combination of these two techniques, the results outperform using just one of these techniques on its own, as outlined by Duncan et al. (2018). To determine the reliability of the photometric redshifts, Hatfield et al. (2022) compared the photometric and spectroscopic redshifts for a subset of sources with both, finding that around 5 per cent of these objects had $|z_{\text{spec}} - z_{\text{phot}}| / (1 + z_{\text{spec}}) > 0.15$. Although this value does give us some impression about the performance of the photo-z method in general, it is unclear whether the same figure is appropriate for the photometric redshifts of the no-ID MIGHTEE sources that we have been able to statistically study, since the no-ID counterparts are fainter than the average source in the Adams et al. (2020) catalogue. However, this is the best we can do with the current data.

2.3.4 Cross-Matched Catalogue

The catalogue with the multiwavelength counterparts contains information for 6 263 radio sources in the central part of the MIGHTEE Early Science Data in the COSMOS field, more specifically it covers the inner $\sim 0.8 \text{ deg}^2$ circle. The reason the area is less than 1.6 deg^2 is because not all wavelengths cover the entire area of 1.6 deg^2 . For example, the optical data only

cover 1 deg^2 (Adams et al., 2020). Another reason is due to the primary beam gain drop as it approaches the edges of the circle region. Prescott et al. (submitted) chooses to restrict the area to where the gain drops to 0.5, this way reducing the sources from 9 896 radio components to 6 070, as well as excluding the sources that fall within the masked out area within the NIR data. The COSMOS field can be seen in Figure 2.1. The sources with no multiwavelength counterparts are circled in green, where the radius of the circle is $14.4''$. As is evident in this figure, there are no unmatched sources in the edges of the field, as outlined above.

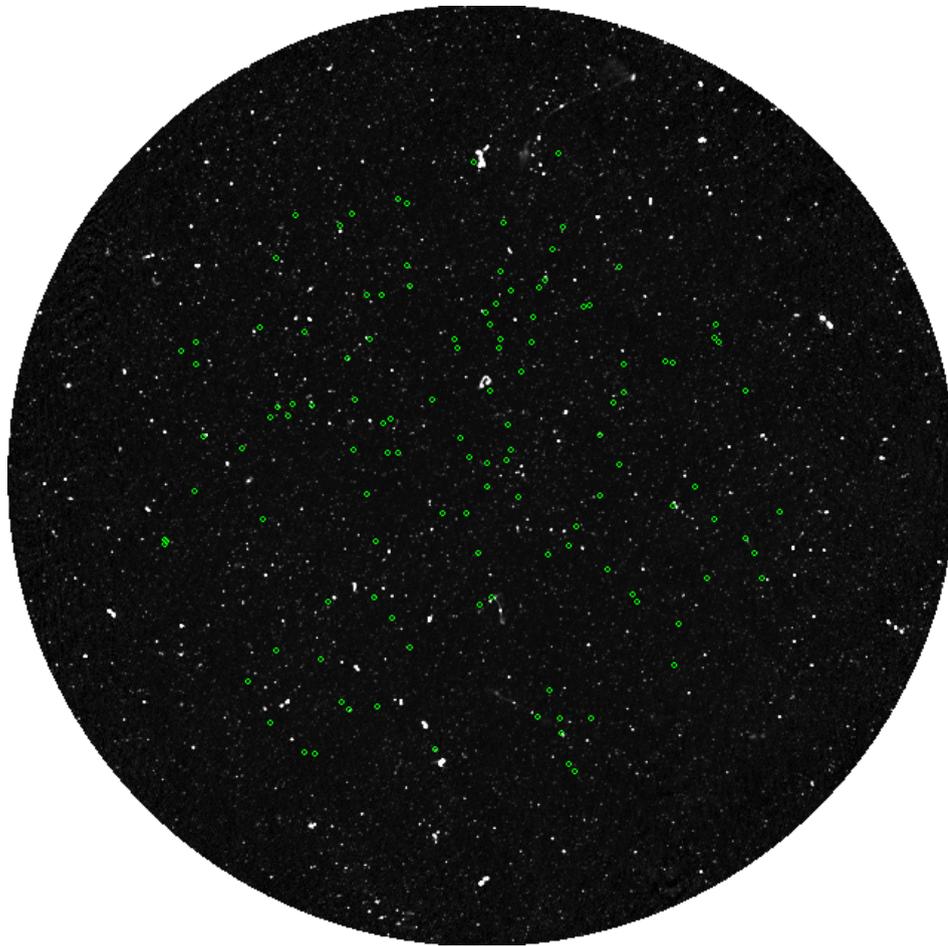


FIGURE 2.1: The MIGHTEE COSMOS Early Science image in radio wavelength at 1.3 GHz (Heywood et al., 2021) with all 123 unmatched radio sources circled in green. The circles have a radius of $14.4''$. This image covers 1.6 deg^2 , however our data covers only 0.8 deg^2 , hence why there are no circled sources around the edge of the image. In this image north is up and east is left.

The process to cross-match the sources from the PYBDSF radio catalogue to their optical counterparts has been carried out by visual inspection. To do this, Prescott et al. (submitted) overlaid radio contours from MIGHTEE, as well as 3 GHz data from VLA-COSMOS observations by Smolčić, V. et al. (2017), on top of a K_s band image from UltraVISTA. They also map the

positions of already known sources from their multiwavelength catalogue on the overlays. The VLA observations allow for an easier identification of a counterpart, due to its high resolution (0.78"). Meanwhile, better sensitivity of the MeerKAT images exposes the diffuse radio sources more effectively. To further the visual classification, two different sizes of overlays are produced - one with a size of 0.5' \times 0.5' and a 3' \times 3'. The larger overlays facilitate the discovery and classification of extended sources, while the smaller one allows for the stringent selection of a counterpart in more crowded regions of galaxies. All the overlays were randomly divided into subsamples of 100 sources and then inspected by three different classifiers from a team of six people to ensure a good quality inspection and identification with independent agreement across the team.

When checking the overlays, they are being classified into four different categories: single-component, multiple-component, no visible optical counterpart and confused source (Prescott et al., submitted). In case a mismatch occurred, the source was revisited and re-classified. This way, 5 240 out of 6 071 sources were classified to have optical counterparts, resulting in an 86 per cent success rate. The following describes all the classifications in the catalogue we used in more detail.

The sources are all flagged for different properties and there are seven flags in total. Those are denoted as follows: 100 - single comp., 101 - no optical match, 102 - too confused to split, 103 - junk/artefact, 105 - unmatched lobe, 112 - split flux and 120 - multiple comp. There are 5 039, 123, 715, 12, 18, 292 and 64 sources in each of these categories respectively. An example MIGHTEE cutout image of sources with all the different flags can be seen in Figure 2.2. This Figure illustrates five sources for each flag, where every column represents a different flag title. Sources flagged with 101, 102 and 103 have no multiwavelength counterparts identified. In this project we chose to focus on sources flagged with 101. There are 123 sources that have been flagged this way. The information for these, such as their IDs, coordinates and flux densities can all be found in Table 2.1.

Figures 2.3, 2.4 and 2.5 were generated in order for us to visually inspect whether there are visible K_s band sources in the images, while also looking at the radio contours to see what our radio data distribution looks like on top of the optical data, very much like Prescott et al. (submitted) has done to identify cross-matches. As is visible from these figures, we can notice

TABLE 2.1: The table for all the sources with no multiwavelength counterparts containing information about their IDs, coordinates in the sky, and their flux densities, denoted as S .

ID	RA (deg)	Dec (deg)	S (mJy)	ID	RA (deg)	Dec (deg)	S (mJy)
COSMOS05403	150.5985	2.1004	0.05355	COSMOS05813	150.1096	2.1879	0.18558
COSMOS05404	150.5963	2.104	0.03236	COSMOS05815	150.1091	2.2239	0.03837
COSMOS05405	150.5973	2.107	0.04779	COSMOS05820	150.1057	2.3327	0.02817
COSMOS05414	150.5735	2.3952	0.05459	COSMOS05821	150.1048	2.4353	0.04996
COSMOS05419	150.5524	2.1802	0.03594	COSMOS05825	150.103	2.0186	0.00738
COSMOS05420	150.552	2.4086	0.09004	COSMOS05834	150.0951	2.4662	0.06922
COSMOS05421	150.5516	2.3747	0.05846	COSMOS05839	150.0923	2.3984	0.03963
COSMOS05423	150.5397	2.2631	0.0679	COSMOS05842	150.0904	2.4128	0.03095
COSMOS05455	150.4809	2.2465	0.03574	COSMOS05844	150.0886	2.5154	0.04722
COSMOS05464	150.472	1.8901	0.10696	COSMOS05846	150.0838	2.5911	0.02312
COSMOS05472	150.4538	2.4304	0.00963	COSMOS05850	150.0799	2.227	0.27302
COSMOS05475	150.4493	2.1376	0.05129	COSMOS05854	150.0779	2.2819	0.06902
COSMOS05483	150.4385	2.2933	0.03422	COSMOS05857	150.0736	2.4873	0.02997
COSMOS05484	150.4382	1.8261	0.04287	COSMOS05859	150.0725	2.2431	0.03382
COSMOS05493	150.4283	1.9382	0.03377	COSMOS05874	150.0631	2.1719	0.03414
COSMOS05495	150.4264	2.3098	0.19299	COSMOS05881	150.0576	2.3637	0.03574
COSMOS05496	150.4293	2.5351	0.03436	COSMOS05899	150.0414	2.4082	0.0582
COSMOS05512	150.4112	2.2962	0.02729	COSMOS05902	150.0397	2.447	0.02162
COSMOS05515	150.4053	2.3127	0.06228	COSMOS05905	150.0333	1.8363	0.35185
COSMOS05519	150.4004	2.6019	0.04523	COSMOS05915	150.03	2.4915	0.09866
COSMOS05528	150.387	1.7827	0.08391	COSMOS05921	150.0217	2.5016	0.03791
COSMOS05529	150.3867	2.4231	0.03965	COSMOS05932	150.0173	2.0841	0.01875
COSMOS05540	150.3749	2.3099	0.03113	COSMOS05936	150.0149	1.8771	0.0392
COSMOS05547	150.3713	1.779	0.04986	COSMOS05943	150.011	2.5491	0.0318
COSMOS05551	150.3618	1.9242	0.03449	COSMOS05954	150.0013	2.6964	0.05866
COSMOS05566	150.3513	2.0105	0.03627	COSMOS05955	149.997	1.8104	1.83989
COSMOS05581	150.3325	2.5852	0.03115	COSMOS05956	149.9986	1.8345	0.03246
COSMOS05584	150.3293	1.8596	0.08674	COSMOS05961	149.9948	2.5825	0.03386
COSMOS05592	150.3213	2.3839	0.39421	COSMOS05973	149.9856	2.0961	0.02869
COSMOS05593	150.3188	1.8473	1.1933	COSMOS05974	149.9851	1.764	0.04623
COSMOS05597	150.315	2.6043	0.04791	COSMOS05982	149.9756	1.7525	0.0528
COSMOS05600	150.3113	2.2433	0.04447	COSMOS05983	149.9735	2.1258	0.0249
COSMOS05605	150.3104	2.3191	0.04642	COSMOS05991	149.962	2.4618	0.02126
COSMOS05618	150.2916	2.4794	0.03378	COSMOS06003	149.9538	2.4642	0.02602
COSMOS05620	150.2922	2.1761	0.06407	COSMOS06004	149.9521	1.8344	0.03602
COSMOS05627	150.287	2.4122	0.037	COSMOS06013	149.9383	2.1748	0.04497
COSMOS05631	150.2814	2.0185	0.0101	COSMOS06014	149.9382	2.2663	0.05444
COSMOS05632	150.2771	1.8517	0.09379	COSMOS06032	149.9264	2.0602	0.01984
COSMOS05633	150.2783	2.1048	0.05876	COSMOS06043	149.9181	2.3162	0.03046
COSMOS05636	150.2687	2.4797	0.0327	COSMOS06062	149.9094	2.222	0.03237
COSMOS05642	150.268	2.2847	0.03215	COSMOS06063	149.9084	2.5233	0.0369
COSMOS05650	150.2613	2.2387	0.02074	COSMOS06068	149.9031	2.374	0.02374
COSMOS05655	150.2555	2.2908	0.0305	COSMOS06073	149.9013	2.3307	0.0138
COSMOS05662	150.2538	1.9876	0.02907	COSMOS06079	149.8889	2.0228	0.02226
COSMOS05669	150.2451	2.6271	0.03252	COSMOS06085	149.8826	2.0127	0.04753
COSMOS05671	150.2442	2.2383	0.01223	COSMOS06119	149.839	2.3777	0.07238
COSMOS05679	150.2278	2.4922	0.05245	COSMOS06126	149.8282	2.1569	0.04275
COSMOS05685	150.2323	2.6184	0.11238	COSMOS06127	149.8288	2.3755	0.0652
COSMOS05690	150.2311	2.5247	0.02438	COSMOS06129	149.8266	1.9151	0.03151
COSMOS05693	150.2262	1.9425	0.02921	COSMOS06137	149.8195	1.9775	0.02699
COSMOS05715	150.1929	2.32	0.03306	COSMOS06155	149.7944	2.1873	0.03015
COSMOS05717	150.1892	1.7865	0.47648	COSMOS06168	149.7759	2.0471	0.05448
COSMOS05729	150.1777	2.1472	0.03809	COSMOS06173	149.766	2.1372	0.02926
COSMOS05754	150.159	2.4113	0.02521	COSMOS06175	149.7646	2.4145	0.03393
COSMOS05761	150.1545	2.3978	0.02758	COSMOS06177	149.7617	2.4348	0.05725
COSMOS05768	150.1509	2.262	0.02593	COSMOS06179	149.7582	2.4081	0.04799
COSMOS05778	150.1403	2.1471	0.14225	COSMOS06206	149.7178	2.3338	0.09679
COSMOS05783	150.1365	2.2323	0.06717	COSMOS06208	149.718	2.109	0.02866
COSMOS05791	150.1289	2.6833	0.02989	COSMOS06222	149.7047	2.0868	0.07345
COSMOS05799	150.1226	2.0864	0.03326	COSMOS06234	149.6922	2.0465	0.03968
COSMOS05803	150.1212	2.0069	0.02763	COSMOS06247	149.665	2.1478	0.11415
COSMOS05810	150.1112	2.4535	0.04145				

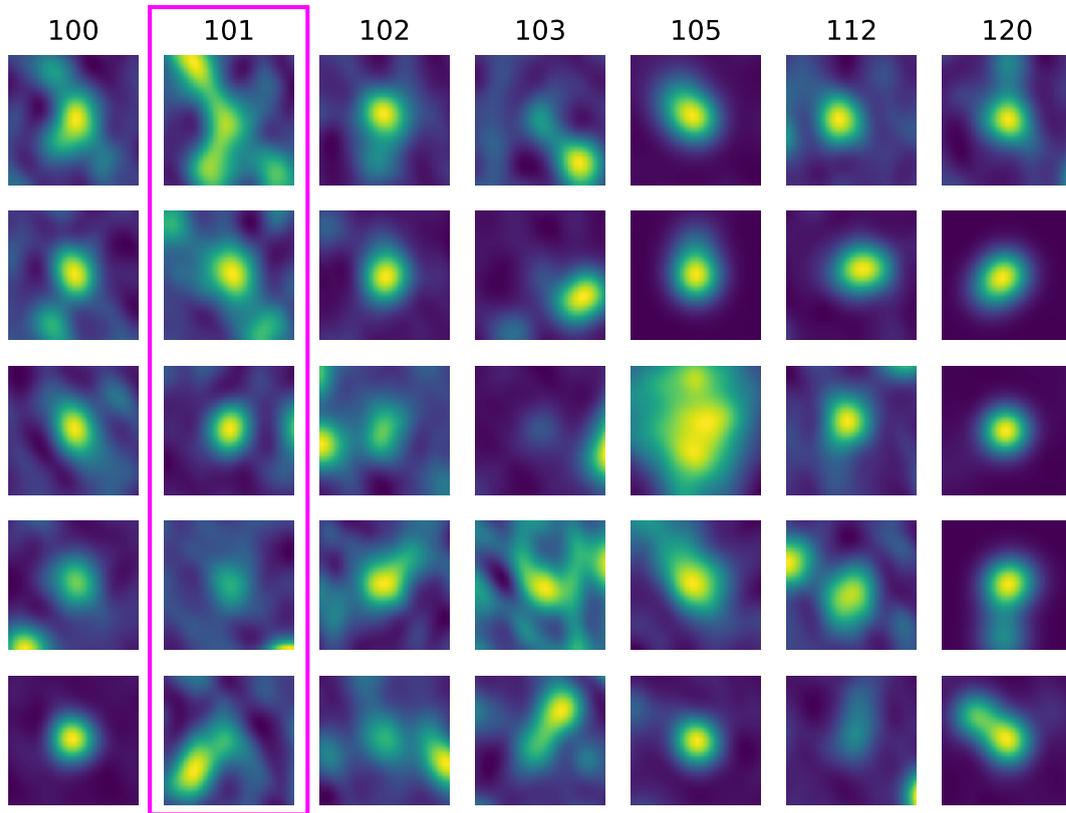


FIGURE 2.2: A figure illustrating sources that are all marked with different flags. Going from left to right for each column the flags are 100 - single comp., 101 - no match, 102 - too confused to split, 103 - junk/artefact, 105 - unmatched lobe, 112 - split flux and 120 - multiple comp. respectively. Flags are also indicated along the top row.

that there are indeed optical sources that fall perfectly in the middle of the cutouts. These cutouts are centered at the coordinates of the no-ID sources ($F=101$), and are all $30'' \times 30''$ in size to see better what is present in the area surrounding these sources. All of them are orientated with north being up and east being left. The exact position of the no-ID radio source is accentuated with red markers.

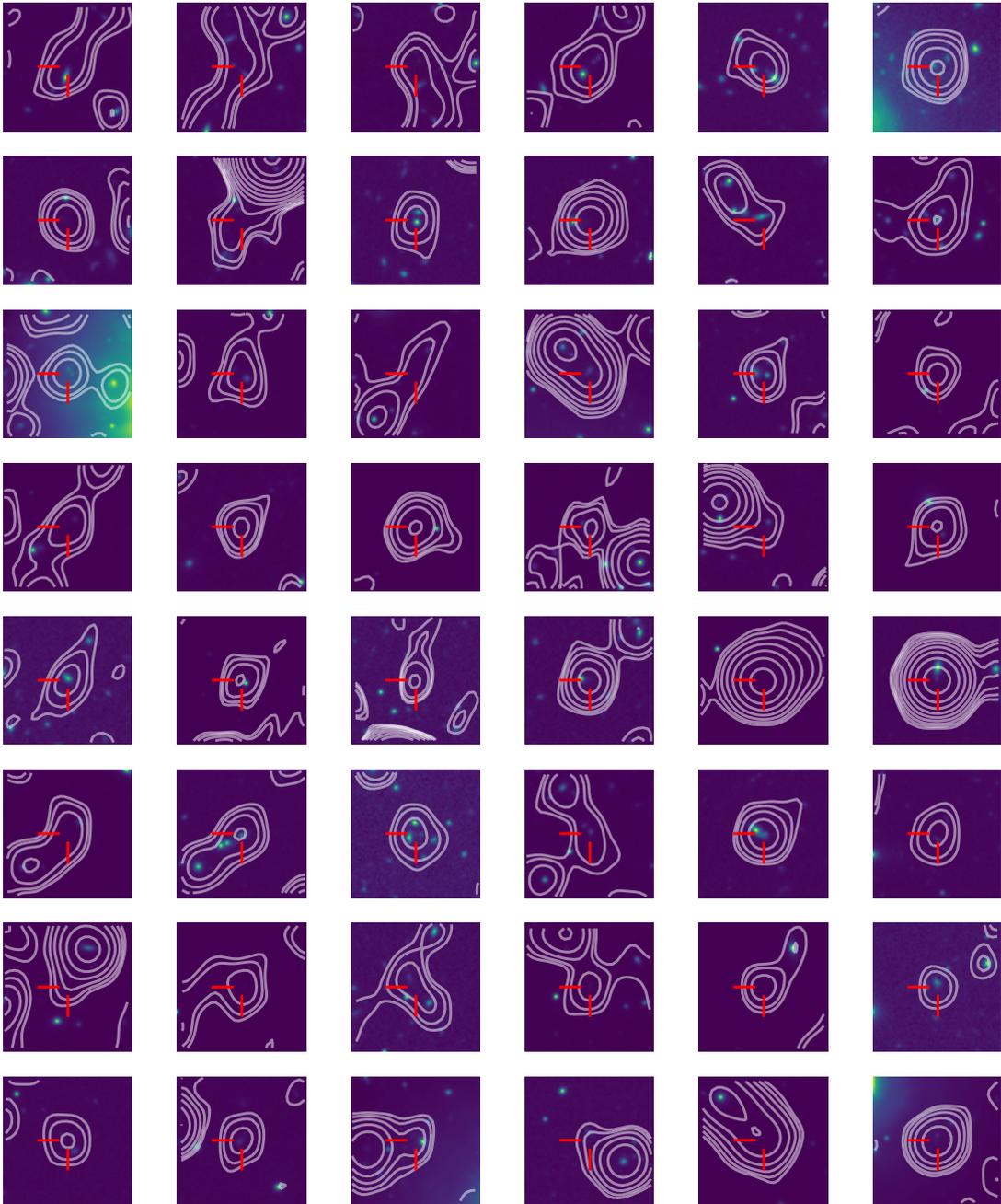


FIGURE 2.3: The first 48 sources out of the 123 with no matched counterparts. All these cutouts are $30'' \times 30''$ in size and have the radio contours from the MIGHTEE Early Science Data overlaid on top of the UltraVISTA K_s band image. The positions of the no-ID radio sources are also highlighted with red markers. The radio contours are equally log spaced with 12 levels ranging from $5.5 \mu\text{Jy}$ to 5.5 mJy , with $5.5 \mu\text{Jy}$ being the 1σ noise level of the image. North is up and east is left in all of these images.

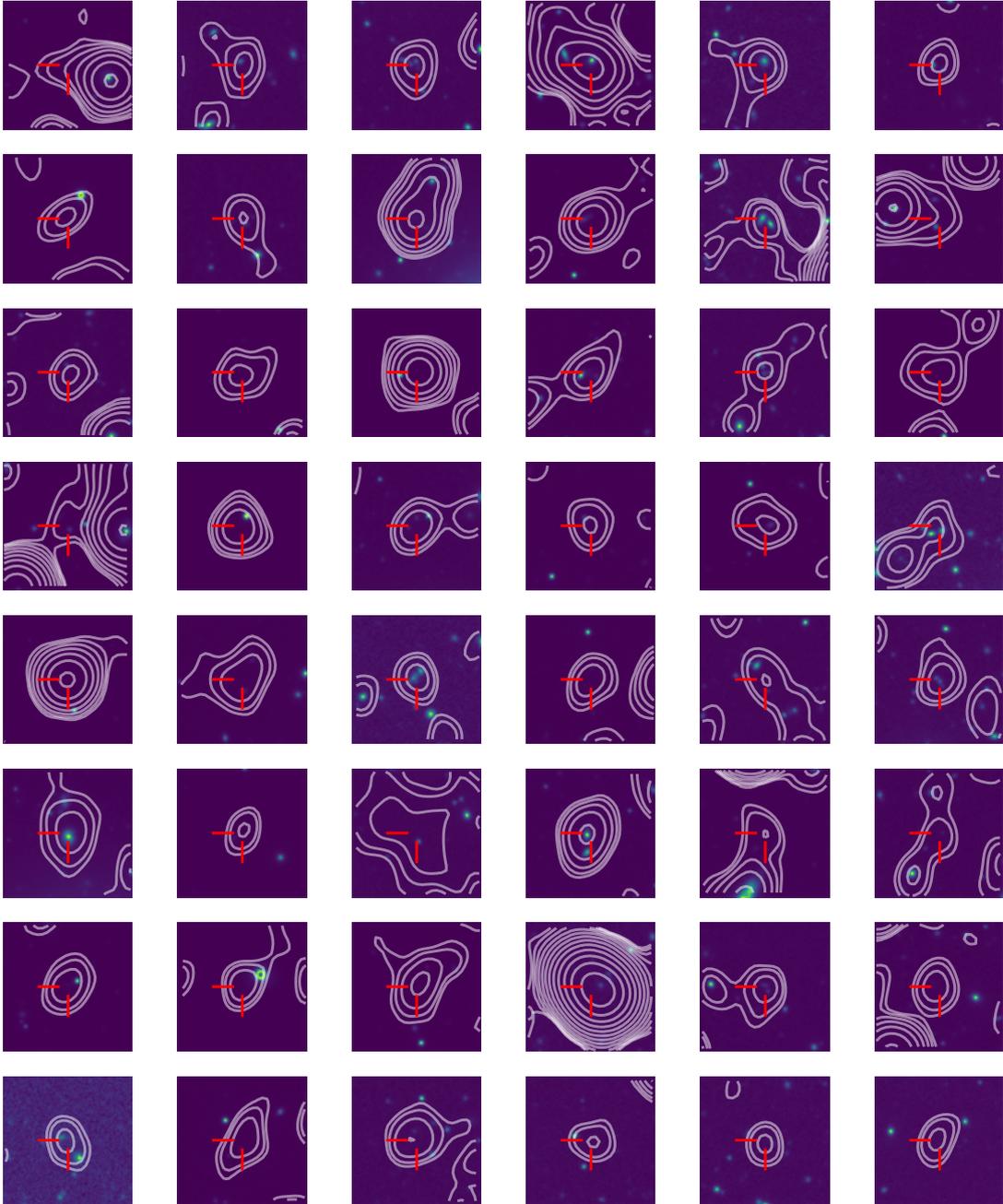


FIGURE 2.4: The second assortment of 48 sources out of the 123 that have no identified counterparts. All these cutouts are $30'' \times 30''$ in size and have the radio contours from the MIGHTEE Early Science Data overlaid on top of the UltraVISTA K_s band image. The positions of the no-ID radio sources are shown with red markers. The radio contours are equally log spaced with 12 levels ranging from $5.5 \mu\text{Jy}$ to 5.5mJy , with $5.5 \mu\text{Jy}$ being the 1σ noise level of the image.

North is up and east is left in all of these images.

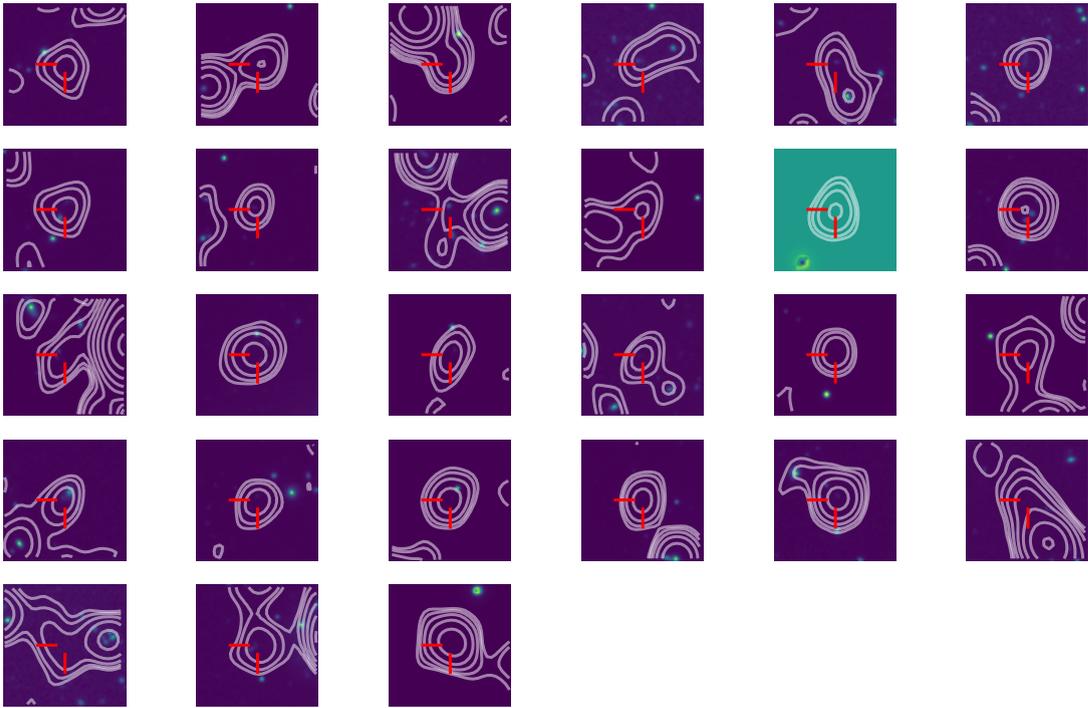


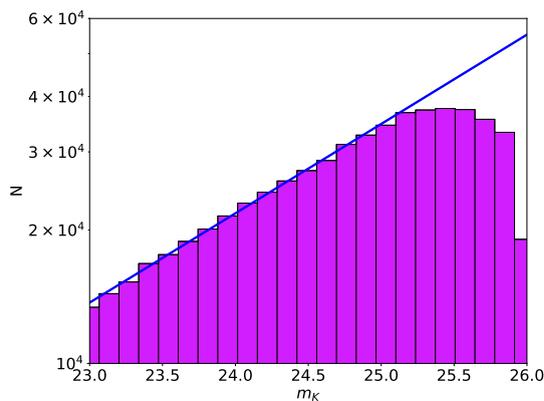
FIGURE 2.5: The third collection the last 27 sources out of the 123 with no matched counterparts. All these cutouts are $30'' \times 30''$ in size and have the radio contours from the MIGHTEE Early Science Data overlaid on top of the UltraVISTA K_S band image. The red markers also show the position of the no-ID radio sources. The radio contours are equally log spaced with 12 levels ranging from $5.5 \mu\text{Jy}$ to 5.5 mJy , with $5.5 \mu\text{Jy}$ being the 1σ noise level of the image. North is up and east is left in all of these images.

2.3.5 K_S Band Data

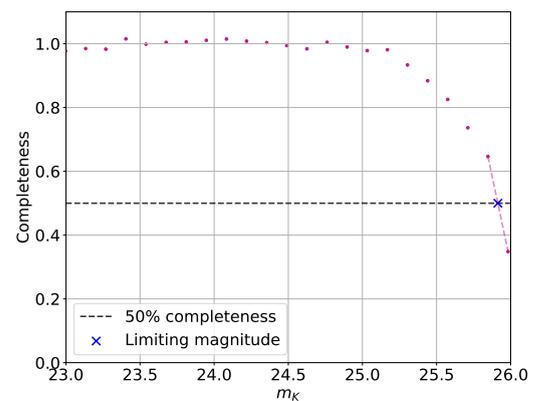
The data that we used to find multiwavelength counterparts to the no-ID sources are taken from the fourth data release of the UltraVISTA survey (McCracken et al., 2012; Moneti et al., 2019). Specifically, we have used the K_S band data to look for the multiwavelength counterparts for the radio sources in question. UltraVISTA data encompasses the central region of the COSMOS field and is an ultra-deep near-infrared imaging survey. The 5σ limiting magnitude for the K_S band is 24.9 for the ultra-deep regions, and 24.5 for the deep regions. These are AB magnitudes. The reason for these two different measurements is due to how the data were collected - there are four deep and four ultra-deep stripes that cover the field of $1.5 \text{ deg} \times 1.2 \text{ deg}$.

The 5σ limiting magnitude is what Prescott et al. (submitted) have used as a cut-off when looking for K_S band counterparts for the no-ID radio sources. However, to be more specific, we have measured the 50 per cent magnitude completeness limit, which can be seen in Figure 2.6. We achieved this by plotting a line of best fit over a histogram of the K_S magnitudes.

This line serves as a guide of the *true population*, meanwhile the histogram itself represents the observed population of magnitudes. By dividing the latter by the former, we can estimate the completeness of the magnitude distribution that will vary from one to zero, or in other words - from complete to empty coverage. In Figure 2.6(a), the K_S magnitude distribution is shown as the purple histogram. Here the magnitude distribution is limited to where $m_K = 26$, as we already know that the 5σ limiting magnitude is 24.9. The blue line going across the bin edges represents the line of best fit. This was made with the consideration that this line should be as close to the straight slope (in log space) the bins make as possible. The plot in Figure 2.6(b) was achieved by plotting the bin centres values from the figure to the left against the completeness values that were calculated by using the method described above. This is represented by the purple scatter points in this plot. As expected, this line stays relatively constant at around the value of one (or 100 per cent), and then rapidly declines as it starts approaching the limiting magnitude value. This is because the fainter magnitudes are harder to detect, so the catalogue becomes increasingly incomplete as we approach fainter magnitudes. This value is denoted as a blue cross in this figure. The value we found for the 50 per cent completeness limit is $m_K = 25.91$, and it applies on average over the whole area. It is important to note that our results may become increasingly unreliable as we approach fainter magnitudes (of $m_K > 24.5$, the 5σ point source limiting magnitude of the deep stripe) because of the different coverage in the UltraVISTA field.



(A) A histogram of the m_k band in the UltraVISTA catalogue, limited to where $m_k=26$. N here denotes the number of elements in each bin.



(B) This plot shows the 50 per cent completeness limit, denoted as a blue cross and located where the line of 50 per cent completeness meets the line of observed completeness distribution.

FIGURE 2.6: These plots present how we have measured the limiting magnitude. By applying a line of best fit to the K_S magnitude distribution, we were able to locate where the 50 per cent completeness limit lies. The value achieved is $m_K = 25.91$.

Figure 2.7 shows the K_S magnitude distribution for cross-matched sources compared with the magnitude distribution for the whole UltraVISTA K_S band catalogue. This includes both the single component sources (Flag = 100), as well as the multiple component ones (Flag = 120). The magnitude axis, as in Figure 2.6, has been limited to $m_K = 26$. It is immediately obvious that the magnitude distribution for the cross matched counterparts peaks at $m_K \sim 20$, rather than increasing towards the faintest sources in the catalogues. In Figure 2.6b we showed that incompleteness in the UltraVISTA catalogue becomes significant only at $m_K > 24.5$, meaning that incompleteness can not explain this turn-down in the magnitude distribution of radio source counterparts that we observe. This turn-down is perhaps the result of the fact that the counterparts of radio sources are likely to be the most massive or most star-forming galaxies at any given epoch, therefore we should expect them to be brighter than the population in general at K_S band. However, we speculate that the requirement for unanimity amongst the team of visual classifiers may also contribute, since it is more difficult to achieve consensus when there are a larger number of background galaxies of equal or brighter apparent magnitude.

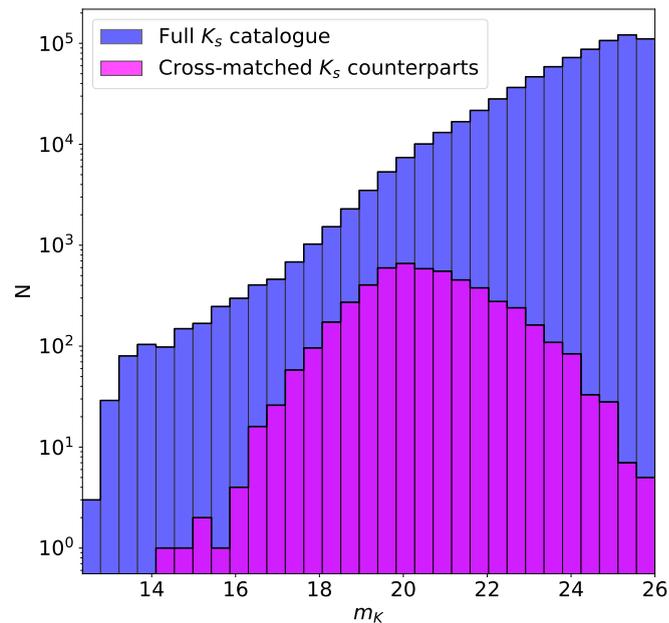


FIGURE 2.7: A histogram with both the full multi-wavelength catalogue of m_K distribution containing all possible cross-matches, and the magnitude distribution for those sources identified as counterparts of MIGHTEE sources in the COSMOS field.

Chapter 3

Properties of the no-ID Radio Sources

3.1 Visual Inspection

To start our investigation of the data, we visually inspected the radio sources that do not have any multi-wavelength counterparts assigned to them. In order to do this, we took a similar approach as already described in Section 2.3.4, and made cutout images of the radio data centered on the positions of no-ID sources. In Figures 2.3, 2.4 and 2.5, one can discern the presence of a radio source by looking at the shape of the radio contours. Since all these cutouts are centered at the no-ID radio sources, we can easily check if they are spurious by inspecting the contour shape in the middle of cutouts. We assume the sources are real since the flux distribution shown in Figure 3.1c is clearly well away from the faint end, where all the false positives and spurious sources would be expected if they existed. Therefore, we expect to see contour lines that are encompassing the centre of the cutout. The shape will naturally depend on the complexity of the source. After visual inspection, ~ 54 per cent of the sample appear morphologically simple (i.e. Gaussian shape), while the remaining ~ 46 per cent seem complex. For example, a morphologically simple source will have circular contour lines surrounding it, as can be seen in Figure 2.3, first cutout in the second row counting from the top, or the very last cutout at the bottom row. An extended or more complex source will still have contour lines around it, but these may not be as trivial to discern as a separate source. A good illustration of this can be seen in Figure 2.3, first cutout in the second to last row, in which an extension is clearly visible at the centre of the cutout from a nearby source to the north-west of the image. The complex

sources are the ones best identified by visual inspection, while the simple ones can be done automatically (as demonstrated by Williams et al., 2019; Kondapally et al., 2021). Figures 2.3, 2.4 and 2.5 have also allowed us to inspect what the the K_S band source distribution looks like around the no-ID radio sources, offering a feel for the problem.

3.2 Flux Density Distribution in the Radio Catalogue

To gather more information about our no-ID sources, we use the flux densities from the cross-matched catalogue, described in Section 2.3.4, and plot them as a histogram to demonstrate the range of their values. We did this to find out what is the range of the flux densities in all the sources compared to the sources with no multi-wavelength counterparts. Looking at the flux densities of the sources will help us determine what the sources could possibly be. The brighter, the more likely that they are radio-loud AGN, or, if the sources are dim, they are more likely to be young, star-forming galaxies or radio-quiet AGN. Some other works that investigate the properties of radio sources as a function of flux density are described in Mauch and Sadler (2007), Best and Heckman (2012), and Best et al., (in preparation). Their findings all agree that the radio-loud AGN are the brightest in the population, meanwhile at the faint end is a mix of radio-quiet AGN and star-forming galaxies (SFGs). Mauch and Sadler (2007) estimate that all the AGN in their investigated surveys that have a redshift of about $z \approx 0.8$, dominate the radio source population when the flux density is above 10 mJy, at frequency of 1.4 GHz. Below that limit, there is an increase of nearby SFGs. Best et al. (in preparation) suggest that at flux densities of above a few mJy, jet-driven AGN are dominating the radio sky. In their investigation, at 150MHz frequencies and $\sim 100 \mu\text{Jy}$, 90 per cent of the population are star forming galaxies and the rest are radio-quiet AGN. The details of the ratio of SFGs to radio-quiet AGN is a hot topic, as is evident by this research.

We make this comparison for sources flagged in each of the different classes, in order to compare how the distribution varies. However, the most useful plot is the one containing the flux density distribution for the no-ID sources (or F=101). Figure 3.1 contains all seven graphs that show flux density distributions against the number of sources in the catalogue (which is equal to 6263) for all differently flagged sources. In all these plots, blue represents the distribution of all sources, meanwhile we are showcasing the flux density distribution for each flag class in

magenta. The full flux density distribution range is approximately 10^{-6} to 10^{-1} Jy. Meanwhile the sources with no multiwavelength counterparts (F=101) have a flux density range of 0.007 - 1.840 mJy. For comparison, the flux density range for single component sources, which is shown in Figure 3.1a, is between 1.42×10^{-3} and 26.07 mJy. The flux density distribution range of the no-ID sources falls within the distribution of the one component sources. It is also clear that the flux density distribution of multiple component sources is shifted towards the right end of the spectrum, meaning that sources flagged with F=120 have higher flux density values, as is expected, since multiple component sources are typically AGN dominated by radio jets or radio jets with lobes (Vardoulaki et al., 2019).

92.7 per cent of the no-ID radio sources (or 121 out of 123) have flux densities of $f \leq 129 \mu\text{Jy}$. 3.3 per cent have flux densities of $f \leq 12.9 \mu\text{Jy}$. To compare with LOFAR numbers by Best et al. (in preparation), we have scaled the 1.28 GHz frequencies used in the MIGHTEE radio catalogue to 150 MHz used in LOFAR, assuming the spectral index of 0.7. $100 \mu\text{Jy}$ at 150 MHz becomes $12.9 \mu\text{Jy}$ at 1.28 GHz, while 1 mJy at 150MHz becomes $129 \mu\text{Jy}$ at 1.28 GHz. This indicates that the hitherto unidentified sources in the MIGHTEE catalogue might be dominated by star-forming galaxies, with some fraction of radio-quiet AGN as well.

3.3 The Multi-Wavelength Counterparts

To start our analysis, we use the UltraVISTA fourth data release (McCracken et al., 2012; Moneti et al., 2019). We choose to filter the K_S catalogue we are using by adopting the previously mentioned magnitude corresponding to 50 per cent completeness. To avoid considering sources with redshifts that are highly likely to be incorrect, we also limit our study to include only sources with $z < 7$. The reason for that is because there are very few sources likely to be bright enough to be detected in the UltraVISTA data with a true redshift > 7 . Since photometric redshifts have significant scatter, the objects at more modest redshifts scattered up to beyond $z > 7$ are likely to overwhelmingly outnumber the few (if any) with true redshifts > 7 . Applying this mask to the catalogue (where $m_K \leq 25.91$ and $z \leq 7$) reduces the number of sources in the catalogue from 995 049 to 753 516.

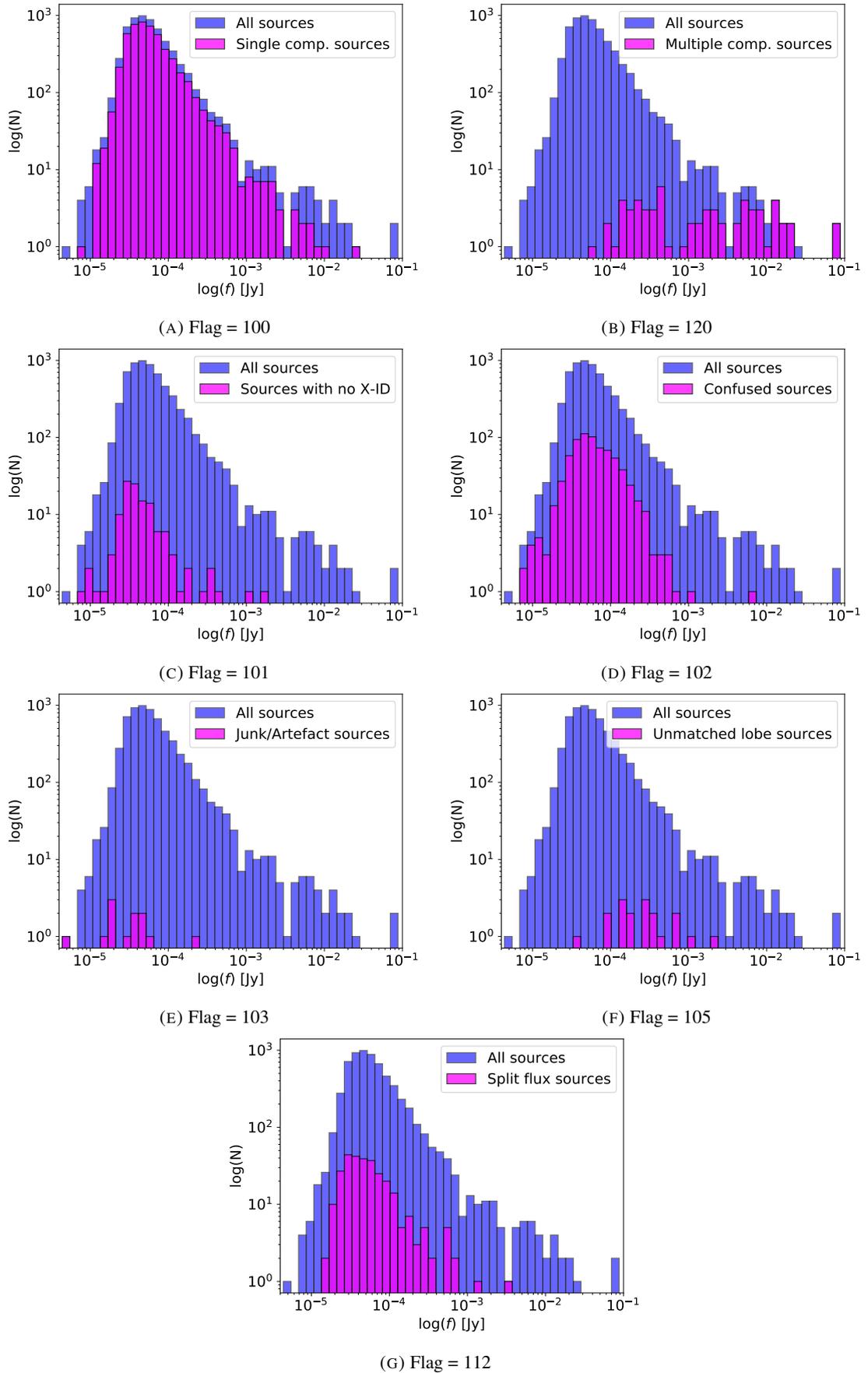


FIGURE 3.1: Histograms for flux density ranges in all the subgroups of radio sources. Each panel contains flux density information of the whole radio catalogue (in blue) against each flag class (in magenta).

After this, we set a search radius variable (r_{search}), within which we will search for possible counterparts. In this work, we are using various values for this measurement, ranging from 1'' to 50''. By establishing a search area, we can then look for the K_S band sources that fall within this space and treat them as potential multi-wavelength counterparts. To see if looking for counterparts for these no-ID sources is even plausible, in Figure 3.2, we plot a distribution of all the K_S band sources in the UltraVISTA DR4, and overlay the histogram of the sources that fall within our set search area. In this case, the search radius is set to 10''. Then we re-normalise the magnitude distribution of the whole field to the area of within 10'' of a no-ID radio source. This enables us to see if there is an excess of K_S band sources near the no-ID radio positions, relative to the background distribution. This lets us deduce that there is indeed an excess of K_S sources that could be unidentified cross-matches for our no-ID radio catalogue. This can be seen in Figure 3.2. Here, the full catalogue is represented by the blue, while the K_S sources that are in our search area are in magenta. The re-normalised K_S band magnitude distribution is shown in a solid black line. In this graph it is visible that there is an excess of sources, as the magenta histogram extends over the black line in all but four bins, and the three bins out of these are at the faint end of the distribution.

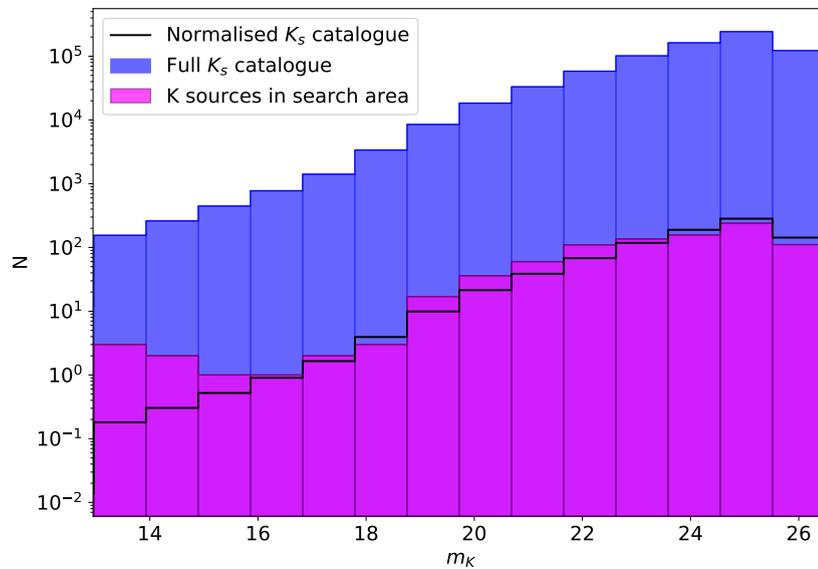


FIGURE 3.2: The m_K distribution in the whole UltraVISTA catalogue compared to the K_S band sources within 10'' of a no-ID radio source. The full catalogue is shown in blue, whereas the sources in the search area are shown in magenta. The normalised catalogue is shown as a black line. It is clearly visible there are excess sources as the magenta histogram extends over the black line in most bins.

To calculate the number of excess galaxies, we first estimate the average number of sources per

unit area in the full K_S band catalogue, and subtract it from the sources in the search area. In order to perform this calculation we need to know the exact sizes of the optical catalogue and the search area. However, this is not as trivial as just counting πr_{search}^2 and multiplying by the number of sources, as some of the no-ID sources are very close together, meaning that the search areas around them would overlap. We need a way to count the overlapping regions only once. To do this, we employ MOCPy (Multi-Order Coverage) package in Python that is designed to define unusual sky regions (Fernique et al., 2014). MOCPy allows to find the UNION between all the search area circles, so if we got two circles that overlap, the overlapping area would only be counted once.

Using this package also allows us to find an INTERSECTION between the area covered by the optical and the radio data. As one can discern from Figure 3.3 (which shows these two areas plotted together, as well as their intersection area), there are areas of radio data with no K_S band coverage. This is most probably due to extremely bright sources making light from faint galaxies surrounding them undetectable. This leaves these circular patterns across the field. It is important to note the radio sources that fall within the areas of no optical coverage, as it would, of course, not be possible to find any K_S band counterparts for them in these areas. There are nine sources in total that fall in the regions of no optical data coverage. This leaves us with 114 no-ID sources we can work with. While being able to deduce this visually from Figure 3.3, we still employ a MOCPy function that allows us to look for the radio sources that the optical coverage contains. Using this, we also get 114 sources in total. In Figure 3.3, the purple represents the optical catalogue, the green embodies the radio catalogue, and the yellow shows the near no-ID radio sources where we can find potential counterparts.

To be able to plot these fields, we have to have the coordinates of the sources in them, as well as choose an order of maximum depth, which defines the resolution of the HEALPix (Hierarchical Equal Area isoLatitude Pixelisation ; Górski et al., 2005; Calabretta and Roukema, 2007) cell. For the radio field (which also doubles as our search area) we use the FROM CONE function and feed it the coordinates of the no-ID sources, as well as the radius angle for the cone, which in this case is equal to the search radius, and we choose the order of the resolution to be 25. The way we narrowed it down was by using a simple calculation and a trial and error method: as there are 114 sources, if they were not overlapping, the total area would be $114\pi r_{search}^2 = 114\pi \times (10/3600)^2 = 0.00276 \text{ deg}^2$. The lower bound, for comparison, with three overlapping

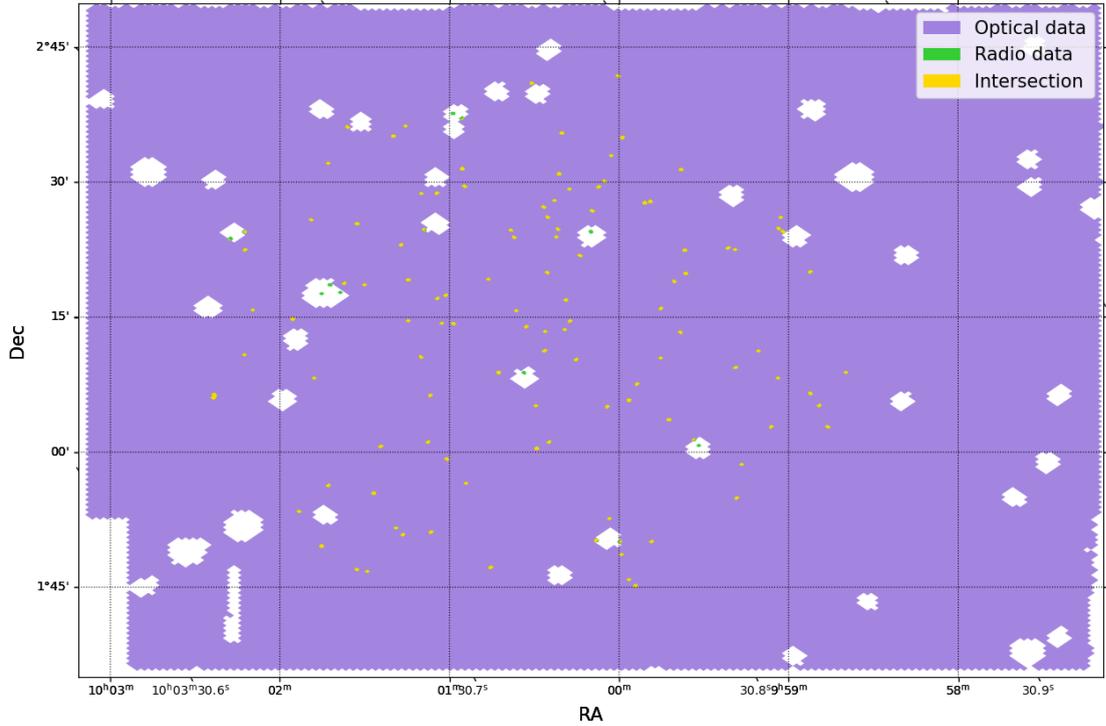


FIGURE 3.3: MOC plots of the optical catalogue (in purple), the search zone around no-ID sources (in green) and the intersection of these two areas (in yellow). Important to note here that some of the no-ID sources fall in the no optical coverage zone. These circular areas with no data are caused by very bright sources that would outshine the sources around them, this way making them invisible.

regions would be $111\pi r_{search}^2 = 114\pi \times (10/3600)^2 = 0.00269 \text{ deg}^2$. Having performed this calculation, we now know that the area we compute by using MOCpy should be somewhat lower than the first value and higher than the second value, as there are only three sources that have their search areas overlapping. Using the order of 25 we achieve a value of 0.00272 deg^2 . To plot the optical field we apply the FROM SKY COORDS function and just feed the coordinates of the whole K_S catalogue, as well as choosing the 13th order to achieve a 1.762 deg^2 optical area. This is consistent with Moneti et al. (2019), where the optical area they quote is $1.5 \text{ deg} \times 1.2 \text{ deg}$. Our choice of order is also based on some calculations of the cell size - it is important not to choose an order that would be too small, otherwise the MOC map becomes fragmented as gaps appear between cells, leading to miscalculations of the area size. 13th order gives a cell size of $25.8''$, which is small enough to display detail of the area but not too small to start to fragment.

To convert the area covered by the MOCpy region to a numerical value in square degrees we use the MOCpy SKY FRACTION function and multiply it by 41 253, which is the number of square degrees in the whole sky.

3.3.1 Finding the Number of Excess K_S Band Sources

Now that we have our radio and optical fields measured, we can calculate $n_{perArea}$ by calculating how many elements there are in each bin (n) when we plot a histogram for the distribution of m_K for the whole catalogue, then dividing n by the optical area and multiplying the result by the search area, this way scaling the K_S band catalogue to the search area and finding out the distribution of the background sources in this field. We also plot the magnitude distribution of the K_S band sources in the search zone, just like we did in Figure 3.2. Following the method of Smith et al. (2017), by taking the background distribution away from the K_S sources in the search area, we can find the exact number of excess sources in our search field. The plot just described can be seen in Figure 3.4. Here, the black histogram represents the background sources, the purple shows the K_S sources that fall within the search area and the blue histogram shows the result of taking the former away from the latter, leaving us with only the excess galaxies. In this figure we can see that most bins, except the ones at the faint end of the magnitude, contain some number of excess sources. Even with the negative excess in some bins, the overall result is still positive.

The components in this plot all have their respective uncertainties. To determine whether the excess is significant or not, in Figure 3.5 we show the same plot but with Poisson error bars included. More on how the uncertainties for this plot were calculated can be found in Appendix A. In Figure 3.5, the black line represents the background sources with the error bars shown as the shaded regions, while the purple shows the K_S sources in the search area. This graph is indicative that our results are significant, as the error bars for the K_S sources do not overlap with the black histogram.

3.3.2 Magnitude Distribution of the no-ID Sources

Looking at Figure 3.6 we can deduce the magnitude distribution of the no-ID sources. Knowing the magnitude distribution when linked with the redshift information allows us to characterise the luminosity distribution of the statistically detected counterparts. In our case, the K_S magnitude distribution for the possible multi-wavelength counterparts for the unidentified sources is roughly from $m_K = 20$ to $m_K = 24$. This means that the counterparts of the sources are faint

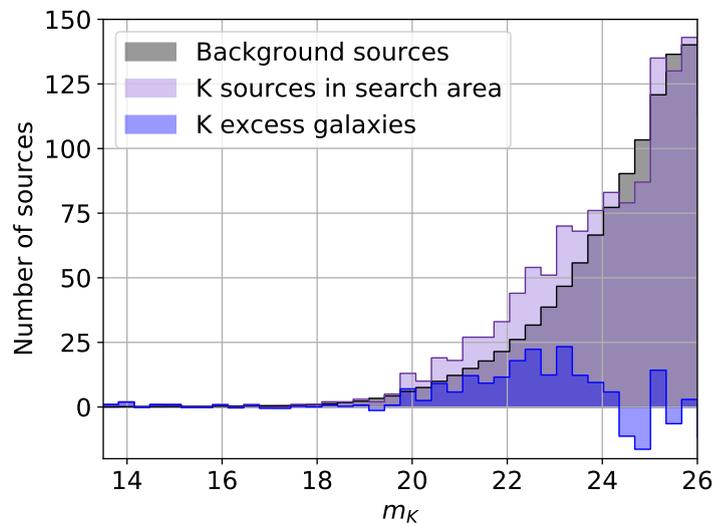


FIGURE 3.4: The K_S magnitude distribution near no-ID radio sources. The black represents the distribution of the background source population in the search area of $10''$, whereas the purple shows the actual K_S sources found in the search area. The excess sources can then be found by taking the black histogram away from the purple. This is shown in blue. It is visible that in most bins the excess is positive.

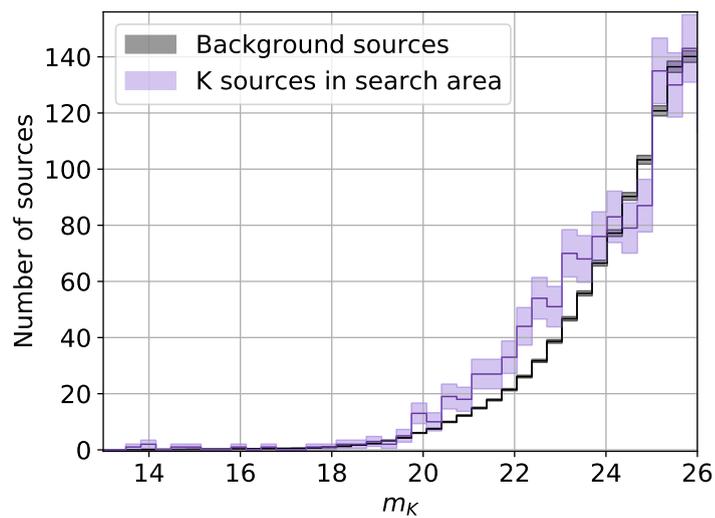


FIGURE 3.5: The background source population and the K_S sources found within the search area together with their respective uncertainties, shown as the gray and the purple shaded regions respectively. This allows us to see if the excess we found is significant.

K_S band sources. Most notably, there is no significant excess beyond $m_K > 24$, which means that $K_S < 24$ data should be sufficient for detecting all counterparts of unidentified MIGHTEE sources. This is a useful find, since it means that for further MIGHTEE investigations in the other fields (such as ELAIS-S1, XMM-LSS, ECDFS), VIDEO data (Jarvis et al., 2013) could be used to find the IDs of the sources that have not been identified, as deep data such as in UltraVISTA is not needed, and VIDEO is a much wider area covering all of MIGHTEE DR1, except for the COSMOS field. However, as will be discussed further in Section 3.4, we only detect ~ 60 per cent of the possible counterparts, which leave the remaining ~ 40 per cent with $m_{K_S} > 26$. To find these sources, one way to go would be to employ pointed observations, with small areas and very deep data collection (e.g. using JWST, or perhaps far-infrared/sub-millimetre interferometry with ALMA, given the likely dusty and high-redshift nature of these sources).

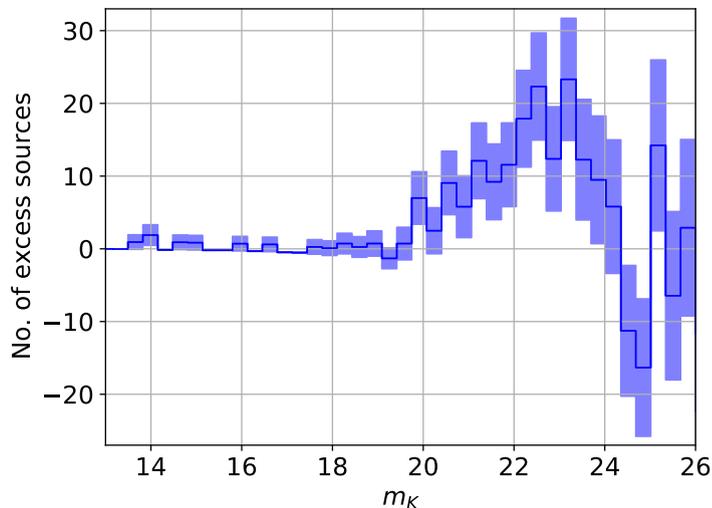


FIGURE 3.6: A histogram of the excess K_S band sources around the no-ID radio sources within the $10''$ search area. All statistical excess is notably in the $20 < m_K < 24$ section with no significant excess beyond $m_K > 24$.

3.3.3 Finding the Number of Excess z sources

We also make the same plot as in Figure 3.4 but this time using a redshift distribution for all the components. As already mentioned in Chapter 1, redshift information is of the utmost importance when analysing astronomical source properties. It is also a good sanity check for our calculations since if our analyses are correct, the number of excess galaxies from Figure 3.4 and this plot, which is shown in Figure 3.7 should be the same. Figure 3.7 shows the respective plot of the redshift distribution between the background sources and the sources around no-ID radio

positions, as well as the excess of sources it results in. In this figure, black represents the background source count, purple - the sources in our search area, and blue - the excess sources. It is noticeable that there are more negative bins compared to Figure 3.4, however the majority of them have significance values ranging between 0.09 - 2.05σ and only one bin reaches over three σ . Figure 3.8 shows the same information about the background and the search area sources but this time including the error bars, which are represented as the shaded regions.

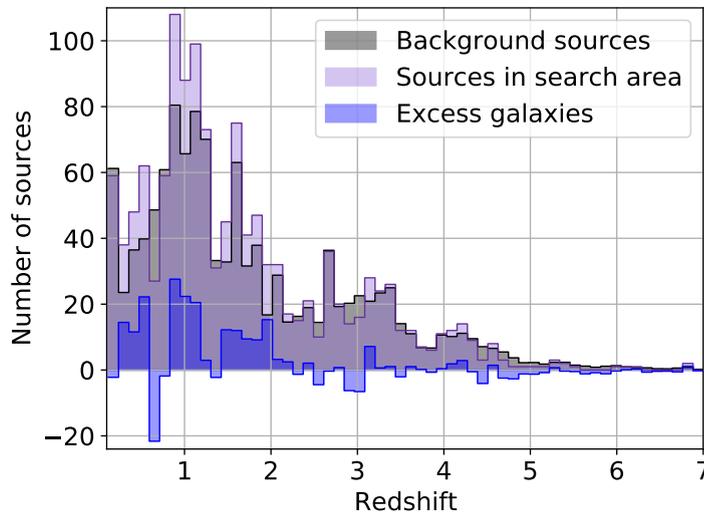


FIGURE 3.7: A histogram representing the redshift distribution of the K_S band sources. The black colour depicts the background source population, while the purple shows those sources that we found in our defined search area of $10''$. Blue shows the result we get after subtracting the background from the number of K_S sources in the search zone.

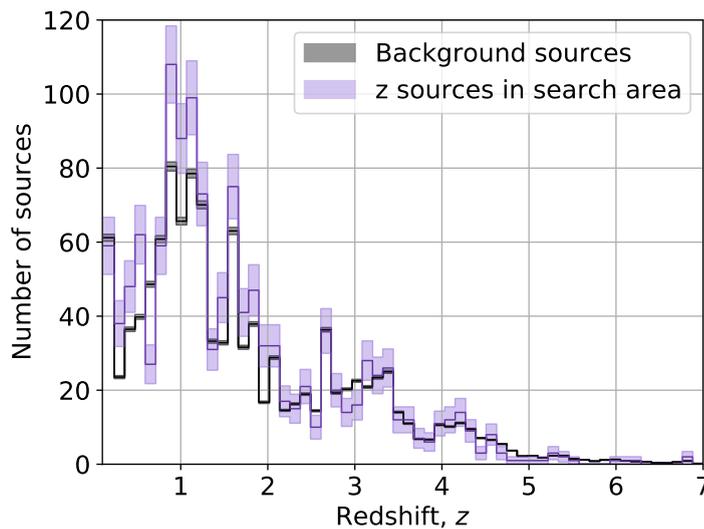


FIGURE 3.8: A plot representing the background source population in the search area together with the population of the K_S sources that are in this search area. In this plot, the uncertainties are also included to check if our results are of significant value.

3.3.4 Redshift Distribution of the no-ID Sources

Knowing the redshift distribution is crucial for understanding the characteristics of astronomical sources as without them physical properties such as distance cannot be determined. Looking at Figure 3.9 we can deduce that the redshift distribution for the no-ID sources is mostly in the range of $0 < z \leq 2$. This means that our sources have comparatively low redshifts, however they fall within the range where most of the background sources lie. As well as this, we also find that out of the 1303 K_S sources in the search area of $10''$, 92 per cent have only the photometric redshifts, and just 8 per cent contain spectroscopic redshifts. We do not identify the particular excess sources, and therefore cannot determine whether they have photometric or spectroscopic redshifts, however, knowing that the majority of the K_S sources contain only photometric information already offers some knowledge about our population.

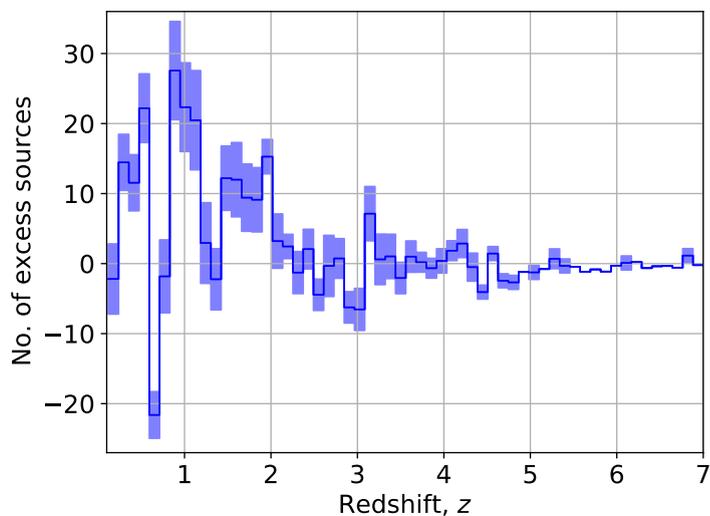


FIGURE 3.9: The number of excess z sources around the no-ID MIGHTEE sources within the $10''$ search area. The majority of the population can be found within the $0 < z < 2$ values.

3.3.5 Total Number of Excess Sources Detected

Our results emerge from summing the blue histogram in either Figure 3.4 or 3.7. As an equation, it can be written as follows:

$$n_{\text{excess}} = \sum_{i=1}^{i=b} (n_i^{\text{search}} - n_i^{\text{bg}}), \quad (3.1)$$

where n_{excess} represents the number of excess galaxies, n_i^{search} shows the number of sources in the search area within the i -th bin, n_i^{bg} is $n_{perArea}$, and b represents the number of bins. When the search radius is set to $10''$, the resulting value is 137.5 ± 49.7 , meaning that the excess of K_S -band galaxies has been detected with a statistical significance of 2.8σ . To see how this number of excess galaxies and the statistical significance varies with different search radii, in Figure 3.10 we show the number of excess sources found increases with the search radii. We can also notice that the error bars increase with the search radii. An interesting feature of the plot to notice is that it starts to plateau at around $6''$ and then starts to rise again at around $14''$. $6''$ is the point where we detect the largest amount of excess sources with the highest significance too - 141.8 ± 31.4 , resulting in statistical significance of 4.5σ . The trend of the plot may be explained by the fact that at a certain radius we will have detected all the multi-wavelength counterparts there are for the no-ID sources. As we approach larger separations we will start to detect neighbouring galaxies rather than the counterparts, since galaxies preferentially live in groups or clusters (Press and Schechter, 1974; Peebles, 1980; Springel et al., 2005). Hence the steep increase of excess galaxies at longer search radii.

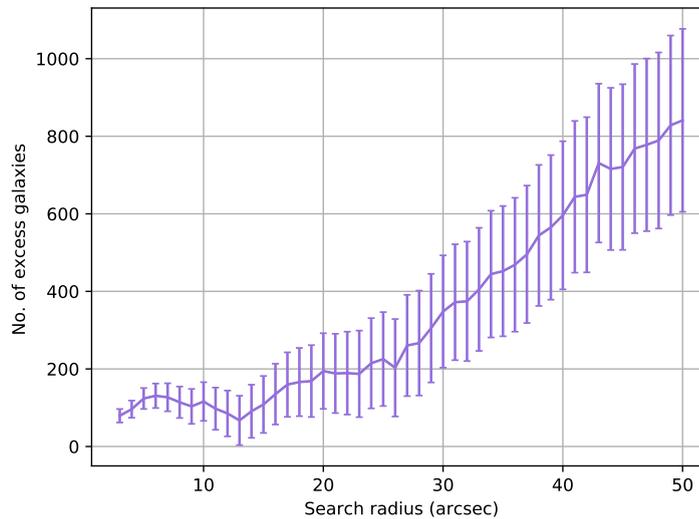


FIGURE 3.10: The number of excess sources detected around the no-ID MIGHTEE sources. The error bars derivation can be found in A.

3.4 Fraction of Excess K Source Population Detected

So far we have looked at this problem as a whole, and we have been carrying out the statistical analysis for all the sources together. Now, to gather more information, we carry on to look at

these sources more individually. By using a part of the likelihood ratio method as described by Fleuren et al. (2012), we calculate the parameter quoted as Q_0 . The likelihood ratio method (LR; Sutherland and Saunders, 1992) relies on estimating the ratio of the probability that two sources are associated to the probability that they are unrelated. In the LR formalism, one of the key tasks is to estimate the parameter Q_0 - the fraction of sources which have at least one counterpart detected in the data being searched. This is clearly very useful for the current investigation, since now that we have detected an excess of K_S band sources around the MIGHTEE no-ID positions, it enables us to quantify whether e.g. every no-ID radio source has a counterpart detected but not identified ($Q_0 = 1$) or more likely what fraction of the no-ID radio sources have at least one K_S band detected counterpart. Similar work has been done by Smith et al. (2017), where the method from Fleuren et al. (2012) has been carried out in order to calculate how many sub-millimetre galaxies have at least one potential counterpart in the Ultra-Deep Survey (UDS; Hartley et al., 2013) field. Since it is hard to calculate Q_0 directly due to a high probability of detecting counterpart sources by chance association (i.e. the unrelated background sources), it is more convenient to calculate $1 - Q_0$, or the number of truly blank sources with no possible counterparts. According to Fleuren et al. (2012), to find this value we use the formula below:

$$1 - Q_0 = \frac{\bar{S}}{\bar{R}} = \frac{N_S^B}{N_S} / \frac{N_R^B}{N_R} = \frac{N_S^B N_R}{N_S N_R^B}. \quad (3.2)$$

In this formula, \bar{S} stands for the fraction of no-ID sources with no possible counterpart (or blank no-ID sources), and \bar{R} stands for the fraction of blank random sources. \bar{R} is something we need to evaluate by generating a catalogue of 1000 random sources in the area that covers the optical catalogue field. We choose to do this with 1000 random positions (instead of 114 to match the no-ID catalogue) to increase our numerical accuracy while not compromising the computational time it takes. N_S , N_S^B , N_R and N_R^B stand for the total number of no-ID sources, the total number of blank no-ID sources, the total number of random radio sources and the total number of blank number of random radio sources. Technically, this equation will only have two variables, as all the other values are constants. N_R and N_R^B will vary every time we generate a new random source catalogue. The reason for variability in N_R is the same reason why we only have 114 sources out of 123 we started with. We need to filter out those random sources that fall within the *holes* in the area covered by UltraVISTA catalogue. All the values are calculated as a function of search

radius to see how the answer varies with the search area and find the best estimate for Q_0 . As an example, one generation of the random catalogue produced values of $N_S = 114$, $N_R = 977$, $N_S^B = 7$ and $N_R^B = 70$, when the search radius was set equal to $5''$, giving $(1 - Q_0) = 0.857$.

The results can be seen in Figure 3.11. In this plot, the random catalogue is represented by the orange colour, meanwhile the no-ID radio sources are in purple. The shaded regions indicate their respective uncertainties. These were calculated using Poisson statistics and error propagation. The shapes both trends make seem to be acting as predicted, as it is anticipated that at first the fraction of blanks in both cases would be close to one, since the search radius is very small. When it reaches just over $1''$ for the random catalogue, and even before the $1''$ line for the no-ID catalogue, the lines drastically drop and by $10''$ both lines are trending towards zero. This is because with large search radius, it is virtually guaranteed to come across a galaxy due to the depth of the UltraVISTA data and hence the large source density.

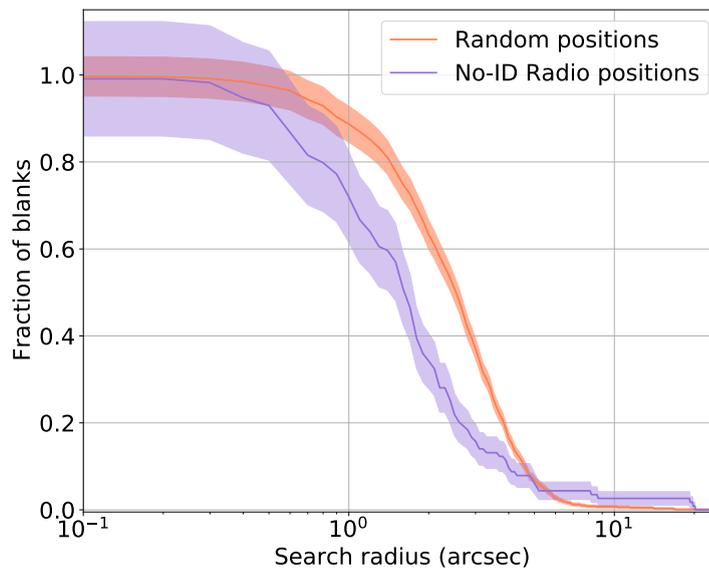


FIGURE 3.11: The fraction of blanks against the search radius in the random and the no-ID catalogues. The random catalogue is shown in orange and the purple represents the no-ID radio sources. The shaded regions indicate their respective uncertainties.

To obtain our value for $1 - Q_0$ we need to divide the purple line in Figure 3.11 by the orange line. Then, we just need to rearrange Equation 3.2 for Q_0 , which can be seen below:

$$Q_0 = 1 - \frac{\bar{S}}{\bar{R}}. \quad (3.3)$$

Now, we can plot our results and determine the best estimate of Q_0 . Including uncertainties, it is shown in Figure 3.12. It can be seen that our best estimate for Q_0 is at the peak of the graph, more specifically at a point 0.603 ± 0.161 . The value of the search radius at this point is $3.1''$. In other words, according to our analysis, 60.3 ± 16.1 per cent of our no-ID sources have at least one plausible multiwavelength counterpart.

Another way to determine Q_0 would be to perform Monte Carlo simulations. This way reduces the amount of error propagation needed, however is longer to compute, and due to the difficulties associated with doing realisations of the distribution of true counterparts, it is simpler to do the analytic error propagation (which we do in appendix A). Hence our best estimate of Q_0 is 60.3 ± 16.1 per cent.

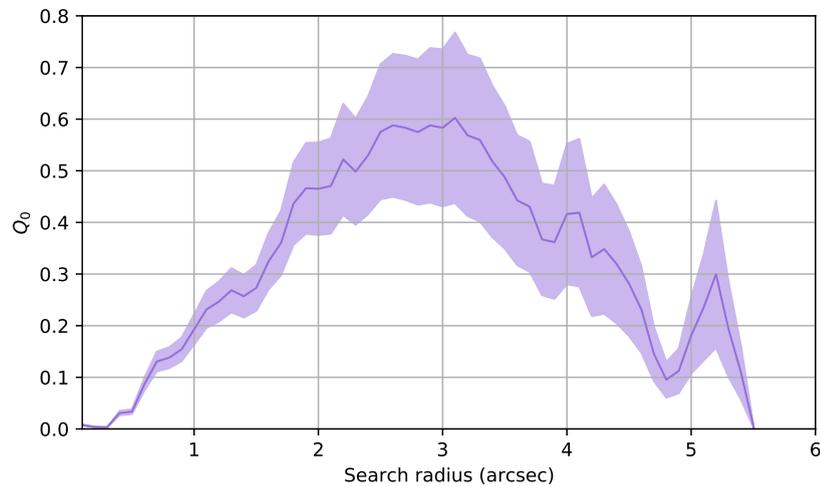


FIGURE 3.12: The trend of Q_0 as a function of search radius. The peak of this plot is our best estimate for Q_0 , which peaks at just over $3''$.

3.5 Truly Unidentified Sources

While we are able to statistically find the multi-wavelength counterparts to most sources, there are still some that fall under the truly unidentified source category. These are the sources that cannot be detected by visual inspection or the LR method. Since we found Q_0 , we can deduce that the $100 - (60.3 \pm 16.1) = 39.7 \pm 16.1$ per cent of the no-ID sources have no multiwavelength counterparts in the UltraVISTA catalogue, meaning that they must be fainter than $m_K = 26$. This suggests a bimodal magnitude distribution for the K_S band counterparts of the no-ID MIGHTEE sources. There are also sources that are classed as ‘blank’, as they have no sources around them

even with a $10''$ radius. We have constructed a table that displays how many sources there are detected for when $r_{search} < 2''$, $r_{search} < 5''$ and $r_{search} < 10''$. The data can be seen in Table 3.1. The numbers presented are summed cumulatively from top to bottom. We notice that there are 39 unidentified sources when $r_{search} < 2''$, however only seven when $r_{search} < 7''$. At $r_{search} < 10''$, we still find three blank sources. Given the SNR of these three blanks, and the fact that we expect > 99.9 per cent of true counterparts to be located at $< 10''$ from the source location, we can be sure that at least some of our no-ID radio sources are truly blank.

$N_{possible}$	$N_{sources}(< 2'')$	$N_{sources}(< 5'')$	$N_{sources}(< 10'')$
$n = 0$	39	7	3
$n \leq 1$	89	15	5
$n \leq 2$	110	33	8
$n \leq 3$	114	51	10
$n \leq 4$	114	72	11
$n \leq 5$	114	92	12
$n \leq 6$	114	105	13
$n \leq 7$	114	111	20
$n \leq 8$	114	113	26
$n \leq 9$	114	114	37
$n \leq 10$	114	114	45
$n \leq 11$	114	114	57
$n \leq 12$	114	114	61
$n \leq 13$	114	114	69
$n \leq 14$	114	114	81
$n \leq 15$	114	114	98
$n \leq 16$	114	114	102
$n \leq 17$	114	114	106
$n \leq 18$	114	114	106
$n \leq 19$	114	114	108
$n \leq 20$	114	114	110
$n \leq 21$	114	114	112
$n \leq 22$	114	114	114

TABLE 3.1: The cumulative table showcasing how many possible counterparts MIGHTEE no-ID sources have at three different search radii.

Chapter 4

Discussion

The purpose of this research is to investigate the properties of the no-ID radio sources that are located in the COSMOS field and are documented in the MIGHTEE catalogue made by Prescott et al. (submitted). There are 123 of these sources in total, which reduce to 114 after excluding the sources that are not covered by the UltraVISTA K_S band catalogue (McCracken et al., 2012). This project closely follows the work completed by Smith et al. (2017), in which a complete distribution of redshifts for sub-millimetre galaxies in the Sub-millimetre Common-User Bolometer Array (SCUBA; Holland et al., 1999) Cosmology Legacy Survey (CLS; Geach et al., 2016) UKIDSS Ultra-Deep Survey (UDS) field has been found using some of the techniques that we have adopted in this assignment.

The first result we achieved was finding the flux density distribution of the no-ID radio sources, which we find to be 0.007 - 1.840 mJy. This falls on the lower side of the flux density distribution for all the sources in the catalogue, however that is where the majority of them lie. We estimate, having compared our work with the radio source demographics as a function of flux density in Best et al. (submitted), that the majority of the no-ID radio sources will be star-forming galaxies, with a lower fraction of radio-quiet AGN too. This is because 92.7 per cent of the sources have a flux density that is lower than $129 \mu\text{Jy}$ - that is the point where Best et al. (submitted) indicate that the source population becomes dominated by SFGs. The histograms for the no-ID source flux densities can be seen in Figure 3.1(c), where it is visible that the no-ID source distribution peaks around the value of $40 \mu\text{Jy}$. The full catalogue distribution peaks at around

$50\mu\text{Jy}$. Notable characteristics of the no-ID source distribution is that none of the sources are so faint they could be spurious sources, but they do all reside at the faint end of the spectrum.

To estimate the typical physical (as opposed to apparent) properties of the no-ID radio sources, we proceed as follows. We take their average redshift ($z = 1.5$ from Figure 3.9) along with a typical flux density ($40\mu\text{Jy}$; Figure 3.1c), and use equation 1 from Read et al. (2018) to determine the mean radio luminosity of 4.5×10^{22} W. We would like to convert this to an SFR, so we first convert to 150 MHz assuming a standard power law relation and a spectral index of -0.7 to obtain a typical 150 MHz luminosity of 2.1×10^{23} W. This allows us to then use the mass-independent relationship between SFR and 150 MHz luminosity from Smith et al. (2021), obtaining an SFR of $12.8 M_{\odot} \text{ yr}^{-1}$. By assuming that the typical no-ID source is a star forming galaxy, we can use the main sequence relation from Schreiber et al. (2015) to estimate a typical stellar mass for a galaxy with this SFR, obtaining $10^{10} M_{\odot}$. To see what range of K -band magnitudes would be appropriate for such a massive galaxy, we turn to the LOFAR deep fields catalogue from Duncan et al. (2021), and find that galaxies around $z = 1.5$ with similar masses have a range of K magnitudes between $22 < m_K < 24$. That this range is fully consistent with the excess which we attribute to the no-ID MIGHTEE sources in Figure 3.6, suggests that these galaxies are consistent with being normal main-sequence star-forming galaxies.

Similar work has been done by Whittam et al. (2022), where the demographics of the radio source population as a function of flux density have been investigated. Around 65 per cent of the MIGHTEE radio sources are SFGs or probable SFGs at flux densities of $S > 19.5\mu\text{Jy}$ at 1.4 GHz frequencies. Comparatively, 120 out of 123 of the no-ID radio sources are above the mentioned flux density, when scaled from 1.28 GHz to 1.4 GHz for comparison purposes.

We have used the UltraVISTA catalogue (McCracken et al., 2012) for the cross-matching element of this project. By locating the K_S sources from the UltraVISTA data that fall within a set search radius, we were able to statistically study multi-wavelength counterparts to our no-ID radio sources. When the search radius is $10''$, the number of excess galaxies we find is 137.5 ± 49.7 . The majority of the excess sources are found between magnitudes 20 and 24, as can be seen in Figure 3.4. There is no statistically significant excess at $m_K > 24$, which is very useful to note, as this means that nearly all detectable counterparts for the no-ID sources that are visible

from the ground are observable in $m_K < 24$ data. This means that we could use large area surveys (such as VIDEO Jarvis et al. (2013), which overlays with all the other MIGHTEE fields) to do the cross-identification of no-ID sources in the COSMOS field and we would be almost as effective as using UltraVISTA, since the magnitude distribution for counterparts found by visual inspection has only ~ 2 per cent of counterparts with $24 < m_K < 26$. Of course in neither case are we able to detect the 40% of the population that is fainter than $K_S = 26$ without recourse to additional observations.

The redshift distribution of the excess galaxies is predominantly $0 < z \leq 2$, as can be seen in Figure 4.1, which demonstrates the magnitude and redshift distributions of excess sources we found around no-ID sources against the magnitude and redshift distributions of the cross-matched K_S counterparts. This can indicate that the reason these counterparts were not detected is due to how faint they are, even if they have a comparably low redshift. As illustrated in Figure 2.7, the number of cross-matched K_S counterparts drops considerably when approaching the faint end of the spectrum. It is useful to compare this figure with Figure 4.1 - the magnitude distribution of the no-ID counterparts does not have any significant excess in the bins beyond $m_K = 24$, as the significance of the bins seen at the very right are less than 1σ away from zero. This is compatible with the data shown in Figure 2.7, as well as Figure 3.4, since we find the counterparts that were left out in the visual inspection. It is fair to assume that the reason for this might just be due to the sources being so faint that visual inspection becomes harder and less reliable. Prescott et al. (submitted) were able to achieve 86 per cent completeness for the MIGHTEE catalogue. 123 of the unidentified sources are the ones not included in these 86 per cent and falling under the flag called “no match”, and are the sources we are investigating.

While visual inspection is often claimed to be highly trustworthy and has a high completeness percentage (see e.g. Mauch and Sadler, 2007), it is a time-consuming method. Another method to cross-match astronomical sources is the method of the nearest neighbour. This method is fast, yet compared to visual inspection yields a lower completeness model because it is unreliable in crowded fields or for sources that have extended features (Ivezić et al., 2002; Sadler et al., 2002; Best et al., 2005). Best et al. (2005) achieve the best result out of these three cases, with 95 per cent completeness and a 98.9 per cent reliability rate, since they use two surveys with different resolutions to identify radio sources. Another method is employing a likelihood ratio already mentioned in Section 3.4 and described in Sutherland and Saunders (1992), as well as

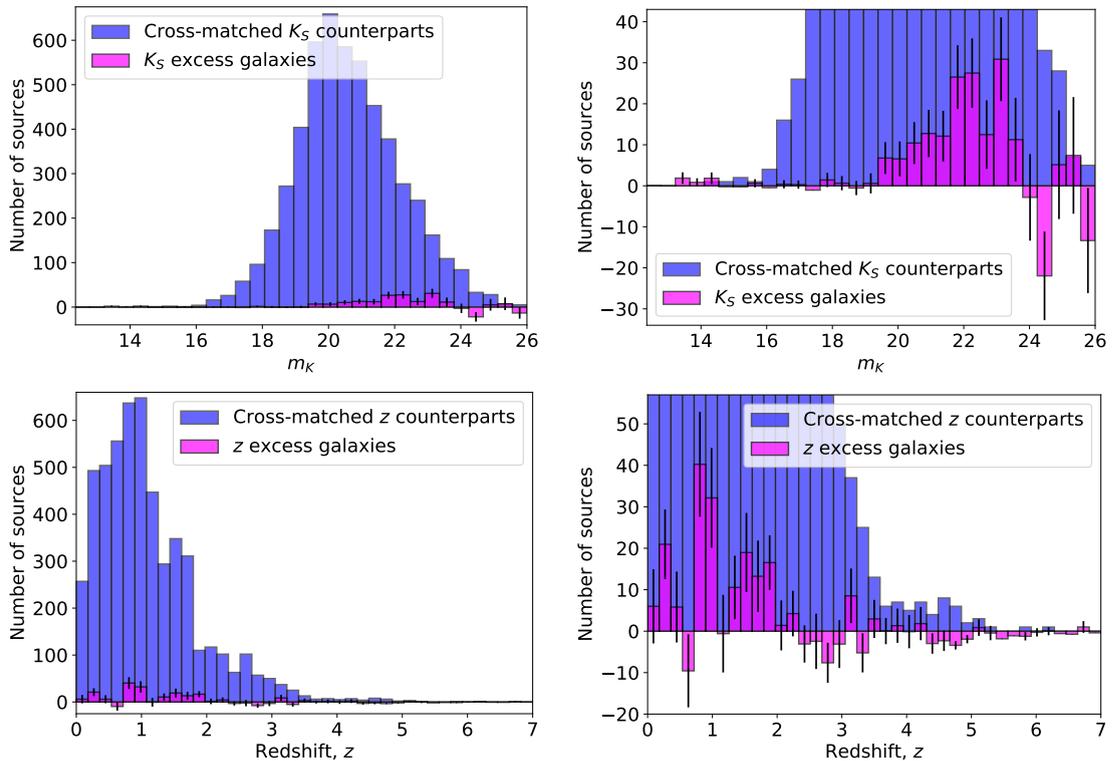


FIGURE 4.1: In the upper left is the magnitude distributions of cross-matched K_S counterparts (in blue) and of the excess galaxies found around the no-ID sources in the COSMOS field (in magenta). The vertical black bars in the magenta plot represent the error bars for the excess galaxy number. In the upper right is the zoomed-in version of the same plot to inspect the excess sources distribution more closely. The lower plots are the same as the upper ones, but instead showing the redshift information.

Fleuren et al. (2012). This is a frequently employed statistical method to find the counterparts of sources detected at various wavelengths (e.g. Ciliegi et al., 2003; Smith et al., 2011; McAlpine et al., 2012; Fleuren et al., 2012; Smith et al., 2017; Kondapally et al., 2021). It performs well in crowded fields but, like the nearest neighbour method, is less trustworthy when it comes to complex morphologies of radio sources. In our case we have statistically identified that a counterpart exists within the catalogue for 60.3 ± 16.1 per cent of the no-ID sources (parameterised by Q_0 in the LR formalism). We determine our best estimate for Q_0 by looking at Figure 3.12 and finding where the trend reaches its peak, which is at $3.1''$. The analytical error propagation for Q_0 is described in A.2. Hypothetically, we could also find Q_0 by using Monte Carlo simulation, however due to reasons mentioned in Section 3.4, we refrain from using this method.

It is useful to compare our work with others, although it is important to mention here that while we were looking at sources that have not been identified by visual inspection, these works were employing the LR method for all the available radio sources in their inspected sky regions. By

adopting the LR method, Ciliegi et al. (2003) identify 92.1 per cent of the optical counterparts for the radio sources in the Lockman Hole with 94.6 per cent reliability. It is difficult to make meaningful comparisons, given that every survey has a different combination of radio survey sensitivity and magnitude limit in the ancillary data but nevertheless it is clear that it is possible to find a large fraction of counterparts to radio survey data using the LR method. While looking at these numbers it is useful to note that these investigations used less sensitive data compared to MIGHTEE and the UltraVISTA data. Another more recent example of using likelihood ratios is given in Kondapally et al. (2021). The Low Frequency Array (LOFAR; van Haarlem et al., 2013) Two Metre Sky Survey (LoTSS; Shimwell et al., 2017, 2019) is a great model for comparison, as it is a pioneering survey in the northern hemisphere, with very sensitive and deep data (Shimwell et al., 2017). In the LoTSS deep fields, Kondapally et al. (2021) use the likelihood ratio to identify mostly those sources that have lower flux densities. The identification rate for the LoTSS sources turns out to be up to 97 per cent. Their completeness value for the compact sources is > 99.7 per cent with the same reliability measurement. Kondapally et al. (2021) use an array of different approaches to identify an astronomical source depending on what their properties are. To complete this they use a decision tree to sort the sources based on whether they are complex or need deblending or need visual inspection. Then it is decided what the next step is to process the source for the best possible classification. Such an approach proves to be effective when dealing with large amounts of data and sources that differ in their complexity. Prescott et al. (submitted) also performs the LR method to compare how successful it is against the visual inspection and finds that it under-performs compared to visual classification. However, visual classification is a biased approach that depends on the expertise of the investigators who are carrying out the cross-identification, and there is no absolute certainty with these results. His results show that only ~ 60 per cent of the catalogue can be successfully identified with a reliability of 95 per cent. This exemplifies that visual inspection is still a vital method when it comes to identifying radio sources.

Having conducted the LR method, Prescott et al. (submitted) recognise that using the likelihood ratio in combination with visual inspection could result in a faster classification process, and the results from our work confirm this and indicate that using these two methods combined could help identify those sources that have been missed by visual inspection, or that deeper near-IR imaging will not be the most beneficial multi-wavelength data set for improving counterpart

identification. This approach is what will be employed for the future classification of XMMLSS, E-CDFS and ELAIS-S1 fields (Prescott et al., submitted). For investigation purposes, we have also calculated the Q_0 parameter for the whole MIGHTEE catalogue and our best estimate is $95 \pm 0.5 \%$ at $3.1''$. Prescott et al. (submitted) does not quote the Q_0 value but the completeness and the reliability measures instead, which are 58.4% and 92.9% respectively for the ‘low’ resolution image. Our value for Q_0 indicates that the LR method could be of great use in cross-matching of radio sources.

Since $Q_0 \neq 1$, there are also radio sources that have been truly undetected in the UltraVISTA data. Reasons for that may vary, but it could be that these sources are too obscured by dust, very faint or have a very high redshift. Another possible reason is due to them being extended radio sources with jets or lobed sources. This can mean that, for example, a lobe gets detected as a radio source but we would not be able to find multiwavelength counterparts to it because we would only be able to find them for the host galaxy, from which the jets are extending. Sometimes the host galaxy can go undetected and only the jets or the lobes are exposed. It could also be that the sources are simply false positive, however after having visually inspected these sources, as well as having performed analysis on them, we conclude that the reason for non-detection is not because any of these sources are spurious. The false-positivity is also only a problem when the radio sources are so faint, they reside very near the faint limit of the survey. The number of truly unidentified sources in this catalogue is 39.3 ± 16.1 per cent out of the 123 sources or 0.79 per cent of the whole MIGHTEE catalogue. These sources are fainter than $m_K = 26$. At $10''$ search radius, there are still three sources that are ‘blanks’ (i.e. they have no possible multiwavelength counterparts). It is imperative to mention here that this percentage is calculated by using those sources that are covered by the UltraVISTA K_S band catalogue, i.e. the sources that are positioned in the white gaps seen in Figure 3.3 are excluded. There are 142 of them in total and they are not possible to identify using the means of this investigation, since they are not covered by the data we are using. Kondapally et al. (2021) states that after the cross-matching efforts, there are still ~ 2.4 - 3.1 per cent of sources, across the three sky fields they have investigated, that have not got any multiwavelength counterparts assigned to them. This, compared to our work, shows that the number of truly unidentified sources in the MIGHTEE catalogue for COSMOS field is in a sensible ballpark. It also shows that additional analysis for primarily unidentified sources may lead to a lower percentage of the truly unidentified sources.

The next sensible steps to make would be to repeat this analysis but this time using different properties of the data, e.g. the colour of the galaxies instead of their magnitudes. Colour is a useful characteristic, since it can be mapped onto galaxy properties more informatively than magnitude (the colour of the galaxy can describe its age, for example, or the amount of dust it obtains, as well as give information about star formation). This is done in aforementioned Kondapally et al. (2021), where they use 16 different colour bins to conduct the LR method, while using the magnitude information as well. Another useful thing to check would be to carry out the full likelihood ratio technique rather than just calculating the Q_0 parameter and to do this for the identified sources of this catalogue. This would be useful to do for ourselves as we could see what fraction of the sources we believe are credible. Also, we could repeat the analysis in multiple dimensions simultaneously, as this could result in getting the luminosity distribution (for that we would need the analysis to be a function of redshift and magnitude). We could achieve this by making a 2D histogram with K_S magnitudes on one axis and redshift on the other, and then subtracting off the background the same way we have already done earlier. The cost of this simultaneous approach, however, is that the number of galaxies per bin is lower, so our small sample of galaxies is most likely not enough for any of the approaches suggested above. It may be useful to attempt to account for the varying coverage of the UltraVISTA map (the deep and the ultra-deep stripes of data). The last suggestion would be to obtain higher resolution radio data. This would allow us to identify the radio positions to higher resolution, which in turn would let us find out where the counterparts are located with higher precision. If we assume that, because these sources are all faint they are likely to be single-component star forming galaxies and radio-quiet AGN, they should all remain unresolved in higher resolution radio data, allowing unambiguous identification of their counterparts. This should allow us to cross-identify 60.3 percent of our sample, given that our Q_0 estimate shows that this percentage of the radio sources have detected (but not identified) counterparts in the existing UltraVISTA K-band imaging. To do this we could perhaps employ the Atacama Large Millimeter Array (ALMA; Brisbin et al., 2017) survey of the COSMOS field, which has a resolutions of $\sim 1''$ (Brisbin et al., 2017). The ALMA data exist over only 0.72 degrees squared of the COSMOS field. It would be beneficial to see how useful the ALMA data are for cross-identifying radio sources. We suggest this could be a useful avenue for the future investigations.

Chapter 5

Conclusions

This thesis describes analysis of unidentified radio sources taken from the MIGHTEE survey catalogue, produced by Prescott et al. (submitted). This project is chosen due to a recent boom of incredibly sensitive radio data due to the increased capabilities of radio astronomy. With SKA construction being underway, working with its precursor MeerKAT is a great opportunity to take a peek at a part of the sky that will be regularly monitored by what will be the most sensitive radio instrument so far. In this work we first set out to understand the problem by learning about radio and multi-wavelength science, and how crucial to analysis it is to be able to detect counterparts to radio sources in other wavelengths, as arguably the most important parameter we can obtain is redshift. Radio sources are usually associated with very energetic processes - at these frequencies a majority of radio detections are dominated by synchrotron radiation due to AGN/SF activity.

Our main objective for this project was to find out what are the properties of no-ID sources in the MIGHTEE survey, or more specifically we looked at 114 no-ID sources and employed the UltraVISTA survey described by McCracken et al. (2012) to statistically study the properties of the K_S magnitude counterparts to them. Since Prescott et al. (submitted) used the means of visual inspection to carry out the cross-identification of all the 6 263 sources in the catalogue, we have chosen to use different methods to see if we can find the counterparts to these unidentified sources in a less subjective manner. First, we have looked at the population of no-ID sources as a whole and found out what is their flux density, magnitude and redshift distributions. The no-ID flux density distribution differs from the corresponding distribution for those MIGHTEE

sources that have IDs. This places the no-ID MIGHTEE radio sources in the flux density region where most star-forming galaxies are found, as well as some radio-quiet AGN, as compared with Best et al. (in preparation) and Whittam et al. (2022). Another intriguing find was the m_K distribution of the no-ID sources, which is mainly found at $20 < m_K < 24$. This means that all these sources are on the dim end of the spectrum, with no significant excess at $24 < m_K < 26$. Since we only find ~ 60 per cent of the possible counterparts (as will be discussed further), the rest of the population must be fainter than $m_K = 26$, which could offer a good explanation as to why they have not been identified through visual inspection. When using this approach, it becomes increasingly more difficult to reach a consensus with dim sources, as they are by definition harder to distinguish from the unrelated galaxy population and a classification using the method of Prescott et al. (submitted) requires an agreement between the inspection board. A further interesting find was the range of redshifts for our no-ID sources. We have calculated this to be in the range of $0 < z < 2$. There is some possible excess in a few bins at higher redshifts, however these are at most of 0.75σ significance and are therefore excluded from the quoted distribution. We were able to calculate how many total counterparts we can find by setting a search radius and looking for $m_K < 26$ sources around our no-ID sources. The results for this can be seen in Figure 3.10, to around 10-14" mark. After that we can see the trend increase drastically, as we start including the neighbouring and background galaxies rather than possible counterparts. We discover that at a 6" search radius - the point where we detect the largest amount of excess galaxies, the number of excess galaxies we find is 141.8 ± 31.4 , proving our results to be significant at a 4.5σ level. It is hard to determine exactly where the plot in this figure starts to plateau (this indicates that all of the possible counterparts are included, and that going out further will not result in finding any more of them).

The next step was to examine these no-ID radio sources more individually, as having the number of excess galaxies that applies to the whole no-ID population does not give a clear picture of how these excess galaxies are distributed amongst the no-ID sources. In order to do this, we have followed the steps of Smith et al. (2017) and used the method of Fleuren et al. (2012) to find the parameter Q_0 that determines what percentage of our 114 sources have at least one multiwavelength counterpart. This is a well established approach, as it is used by the likes of Smith et al. (2011); Fleuren et al. (2012); McAlpine et al. (2012); Smith et al. (2017); Kondapally et al. (2021), as well as by Prescott et al. (submitted), who use this approach to test how reliable

it would be for the catalogue we are using. We find that our Q_0 value is 60.3 ± 16.1 per cent. We can calculate this value in two ways - one is to make one random catalogue and find the uncertainty analytically, which attained a value of 60.3 ± 16.1 per cent. This has allowed us to see how Q_0 varies with different search radii. Therefore, we find that around 60 per cent of the 114 sources have at least one associated multi-wavelength counterpart. Compared to other works, our value for Q_0 parameter is quite small, however it is essential to mention that we are working with sources that have not been identified during the main cross-identification process, which means these sources are more difficult to identify. The works we are comparing with have used the LR method for all the available sources rather than a specific cohort as we have. When we calculate Q_0 for the whole catalogue, we get a value of 95 ± 0.5 per cent, which is close to the value quoted by Kondapally et al. (2021). This value is also close to the percentage of sources in the MIGHTEE catalogue that have cross-IDs.

The following action was to see how many truly unidentified sources there are. There are sources which we have not been able to identify simply because they are not covered by the UltraVISTA data (most likely because they fall in areas that are outshone by objects of stellar nature), of which there were nine out of 123. Then there are objects which are truly ‘blank’, meaning that we could not detect any possible counterparts around them, even though they fall within the “good” parts of the map. Even with a search radius of $10''$, we cannot detect any counterparts for three sources. There are also ~ 40 per cent of the sources that are fainter than $m_K = 26$, and therefore are truly unidentified. This percentage comes from our Q_0 calculations. All of this indicates that some sources, even with one of the deepest and most sensitive data sets available, are still really hard to identify. These might be very obscured, faint or high-redshift objects, as is indicated by previous works on IR-faint radio sources by Maini et al. (2016); Singh et al. (2017); Simpson (2017) and Orenstein et al. (2019).

We recommend to continue this analysis by performing it for a different parameter, e.g. colour, as it would offer more information about the sources (for instance blue indicates young, star forming galaxies and red colour suggests dusty and old galaxies). Although this recommendation is for future work in this area, as a larger sample would be required to carry this out, other works agree that this is a useful approach. Along with this, for future research in this area it might be useful to perform a simultaneous analysis approach, which could lead to getting such

properties as the luminosity distribution. However, for this we would also need a larger sample of galaxies.

Appendix A

Propagation of Uncertainty

A.1 The Uncertainty of the Number of Excess Galaxies

First of all, we need to note that we treat all the numerical values with Poisson statistics, meaning that all the counted values will have uncertainties of \sqrt{n} , where n is the count number. For simplicity, we display all the values and their accompanied uncertainties in Table A.1. In this table, n_{kar} and n_{zar} stands for the number of K_S and z sources respectively that are in the search area around the no-ID radio sources. n_k and n_z stand for the number of sources in the whole K_S catalogue, and A_{opt} and A_{rad} stand for the optical catalogue area and the search area respectively.

TABLE A.1

Value	Uncertainty
n_{kar}	$\sqrt{n_{kar}}$
n_{zar}	$\sqrt{n_{zar}}$
n_k	$\sqrt{n_k}$
n_z	$\sqrt{n_z}$
A_{opt}	$0.01 \times A_{opt}$
A_{rad}	$0.01 \times A_{rad}$

Now, to find the uncertainty on the number of excess galaxies we have to consider both the background source count value and the number of K_S sources in the search area (n_{kar}). Both these values will have their own uncertainties. To find the uncertainty of n_{kar} we can refer to Table A.1. To find the uncertainty of the background sources, which we earlier already referred to as $n_{perArea}$ in Section 3.3.1, we have to carry out error propagation, since this value

is composed of three different variables with their own associated uncertainties. For simplicity, we will now call this value n_{pa} . The equation for n_{pa} in terms of our named variables is given below:

$$n_{pa} = \frac{n}{A_{opt}} \times A_{rad}. \quad (\text{A.1})$$

The calculation for this value would be carried out in the same way irrespective of whether it is done for m_k or z sources, hence why the computation is given as a general solution for both these versions, where n can stand for both n_k and n_z . The resulting value would also be the same, since by definition $n_k = n_z$. The uncertainty on n_{pa} would then be

$$\delta n_{pa} = \sqrt{\left(\frac{\partial n_{pa}}{\partial n} \times \delta n\right)^2 + \left(\frac{\partial n_{pa}}{\partial A_{opt}} \times \delta A_{opt}\right)^2 + \left(\frac{\partial n_{pa}}{\partial A_{rad}} \times \delta A_{rad}\right)^2}. \quad (\text{A.2})$$

The partial derivatives that the equation above contains can be solved like this:

$$\frac{\partial n_{pa}}{\partial n} = \frac{A_{rad}}{A_{opt}}, \quad \frac{\partial n_{pa}}{\partial A_{opt}} = -\frac{n \times A_{rad}}{A_{opt}^2}, \quad \frac{\partial n_{pa}}{\partial A_{rad}} = \frac{n}{A_{opt}}. \quad (\text{A.3})$$

Combining Equations A.2 and A.3 yields

$$\delta n_{pa} = \sqrt{\left(\frac{A_{rad}}{A_{opt}} \sqrt{n}\right)^2 + \left(-\frac{n \times A_{rad}}{A_{opt}^2} \times 0.01 A_{opt}\right)^2 + \left(\frac{n}{A_{opt}} \times 0.01 A_{rad}\right)^2} \quad (\text{A.4})$$

$$= \sqrt{\left(\frac{A_{rad} \sqrt{n}}{A_{opt}}\right)^2 + \left(\frac{0.01 n A_{rad}}{A_{opt}}\right)^2 + \left(\frac{0.01 n A_{rad}}{A_{opt}}\right)^2} \quad (\text{A.5})$$

$$= \sqrt{\left(\frac{A_{rad} \sqrt{n}}{A_{opt}}\right)^2 + 2 \left(\frac{0.01 n A_{rad}}{A_{opt}}\right)^2}. \quad (\text{A.6})$$

Since all these values are known to us, we can just plug them in and see what the result is:

$$\delta n_{pa} = \sqrt{\left(\frac{0.00272\sqrt{753516}}{1.762}\right)^2 + 2\left(\frac{0.01 \times 753516 \times 0.00272}{1.762}\right)^2} = \sqrt{272.53} = 16.51 \quad (\text{A.7})$$

Having found the error on n_{pa} we can now carry on to find the uncertainty of the number of excess galaxies. To reiterate, we find the excess number by taking away the number of background sources from the number of K_S sources found within the search area:

$$n_{excess} = n_{kar} - n_{pa} \quad (\text{A.8})$$

Therefore, the uncertainty on n_{excess} will simply be

$$\delta n_{excess} = \sqrt{(\delta n_{pa})^2 + (\delta n_{kar})^2} = \sqrt{(16.51)^2 + 1303} = 39.69. \quad (\text{A.9})$$

A.2 The uncertainty on the Q_0 value

To find the uncertainty of Q_0 we have to carry out error propagation, since this is not a trivial equation and every value in it has an uncertainty of their own. The main formula for more complex error propagation is given in Equation A.10, where we are using an example equation of $c = ab$ and both a and b have uncertainties.

$$\delta c = \sqrt{\left(\frac{\partial c}{\partial a} \times \delta a\right)^2 + \left(\frac{\partial c}{\partial b} \times \delta b\right)^2} \quad (\text{A.10})$$

In terms used in the formula to find Q_0 , the equation for δQ_0 is given in Equation A.11. To find the uncertainties of N_S, N_S^B, N_R and N_R^B we are using Poisson statistics, i.e. applying the rule of $\sigma = \sqrt{\mu}$, where σ stands for the uncertainty, and μ is the number of detections. In our case, $\delta N_S = \sqrt{N_S}$, $\delta N_S^B = \sqrt{N_S^B}$, $\delta N_R = \sqrt{N_R}$ and $\delta N_R^B = \sqrt{N_R^B}$.

$$\delta Q_0 = \sqrt{\left(\frac{\partial Q_0}{\partial N_S} \times \delta N_S\right)^2 + \left(\frac{\partial Q_0}{\partial N_S^B} \times \delta N_S^B\right)^2 + \left(\frac{\partial Q_0}{\partial N_R} \times \delta N_R\right)^2 + \left(\frac{\partial Q_0}{\partial N_R^B} \times \delta N_R^B\right)^2} \quad (\text{A.11})$$

Next, we have to find all the partial derivatives in this equation. The computation of that is shown Equation A.12. Then, combining Eq. A.11 with Eq. A.12 yields Equation A.14.

$$\frac{\partial Q_0}{\partial N_S} = -\frac{N_S^B N_R}{N_S^2 N_R^B}, \quad \frac{\partial Q_0}{\partial N_S^B} = \frac{N_R}{N_S N_R^B}, \quad \frac{\partial Q_0}{\partial N_R} = \frac{N_S^B}{N_S N_R^B}, \quad \frac{\partial Q_0}{\partial N_R^B} = -\frac{N_S^B N_R}{N_S (N_R^B)^2} \quad (\text{A.12})$$

$$\delta Q_0 = \sqrt{\left(-\frac{N_S^B N_R}{N_S^2 N_R^B} \delta N_S\right)^2 + \left(\frac{N_R}{N_S N_R^B} \delta N_S^B\right)^2 + \left(\frac{N_S^B}{N_S N_R^B} \delta N_R\right)^2 + \left(-\frac{N_S^B N_R}{N_S (N_R^B)^2} \delta N_R^B\right)^2} \quad (\text{A.13})$$

$$\delta Q_0 = \frac{1}{N_S N_R^B} \sqrt{\left(\frac{N_S^B N_R}{N_S} \delta N_S\right)^2 + \left(N_R \delta N_S^B\right)^2 + \left(N_S^B \delta N_R\right)^2 + \left(\frac{N_S^B N_R}{N_R^B} \delta N_R^B\right)^2} \quad (\text{A.14})$$

Bibliography

Adams, N.J., Bowler, R.A.A., Jarvis, M.J., et al., 2020. The rest-frame UV luminosity function at $z = 4$: a significant contribution of AGNs to the bright end of the galaxy population. *MNRAS*, 494(2):1771.

Adams, N.J., Bowler, R.A.A., Jarvis, M.J., et al., 2021. Evolution of the galaxy stellar mass function: evidence for an increasing M^* from $z = 2$ to the present day. *Monthly Notices of the Royal Astronomical Society*, 506(4):4933.

Aharonian, F.A., 2004. *Very high energy cosmic gamma radiation : a crucial window on the extreme Universe*.

Aihara, H., AlSayyad, Y., Ando, M., et al., 2019. Second data release of the Hyper Suprime-Cam Subaru Strategic Program. *Publications of the Astronomical Society of Japan*, 71(6). 114.

Aihara, H., Arimoto, N., Armstrong, R., et al., 2017a. The Hyper Suprime-Cam SSP Survey: Overview and survey design. *Publications of the Astronomical Society of Japan*, 70(SP1). S4.

Aihara, H., Armstrong, R., Bickerton, S., et al., 2017b. First data release of the Hyper Suprime-Cam Subaru Strategic Program. *Publications of the Astronomical Society of Japan*, 70(SP1). S8.

Alexander, P. and Leahy, J.P., 1987. Ageing and speeds in a representative sample of 21 classical double radio sources. *MNRAS*, 225:1.

Almosallam, I.A., Jarvis, M.J., and Roberts, S.J., 2016b. GPZ: non-stationary sparse Gaussian processes for heteroscedastic uncertainty estimation in photometric redshifts. *MNRAS*, 462(1):726.

- Almosallam, I.A., Lindsay, S.N., Jarvis, M.J., et al., 2016a. A sparse Gaussian process framework for photometric redshift estimation. *Monthly Notices of the Royal Astronomical Society*, 455(3):2387.
- Antonucci, R., 1993. Unified models for active galactic nuclei and quasars. *Annual Review of Astronomy and Astrophysics*, 31:473.
- Arnouts, S., Cristiani, S., Moscardini, L., et al., 1999. Measuring and modelling the redshift evolution of clustering: the Hubble Deep Field North. *Monthly Notices of the Royal Astronomical Society*, 310(2):540.
- Ashby, M.L.N., Willner, S.P., Fazio, G.G., et al., 2013. SEDS: The Spitzer Extended Deep Survey. Survey Design, Photometry, and Deep IRAC Source Counts. *ApJ*, 769(1):80.
- Baldi, R.D., Torresi, E., Migliori, G., et al., 2019. The High Energy View of FR0 Radio Galaxies. *Galaxies*, 7(3):76.
- Barkov, M.V., Khangulyan, D.V., and Popov, S.B., 2012. Jets and gamma-ray emission from isolated accreting black holes. *MNRAS*, 427(1):589.
- Becker, R.H., White, R.L., and Helfand, D.J., 1994. The VLA's FIRST Survey. In D.R. Crabtree, R.J. Hanisch, and J. Barnes, editors, *Astronomical Data Analysis Software and Systems III*, volume 61 of *Astronomical Society of the Pacific Conference Series*, page 165.
- Bertin, E. and Arnouts, S., 1996. SExtractor: Software for source extraction. *Astron. Astrophys. Suppl. Ser.*, 117(2):393.
- Best, P.N. and Heckman, T.M., 2012. On the fundamental dichotomy in the local radio-AGN population: accretion, evolution and host galaxy properties. *MNRAS*, 421(2):1569.
- Best, P.N., Kauffmann, G., Heckman, T.M., et al., 2005. A sample of radio-loud active galactic nuclei in the Sloan Digital Sky Survey. *MNRAS*, 362(1):9.
- Bolzonella, M., Miralles, J.M., and Pello', R., 2000. Photometric redshifts based on standard sed fitting procedures.
- Bougeret, J., Goetz, K., Kaiser, M., et al., 2008. S/waves: The radio and plasma wave investigation on the stereo mission. *Space Science Reviews*, 136(1-4):487. Funding Information:

- At the Observatoire de Paris, this work was performed with the support of both CNES and CNRS. We wish to thank Jean-Yves Prado for his support concerning the CNES tracking antenna facility. At Minnesota and Berkeley, this work was performed with the support of NASA contract: NAS5-03076. I. Zouganelis acknowledges the European Marie Curie Fellowship SOPHYSM for financial support.
- Bowler, R.A.A., Jarvis, M.J., Dunlop, J.S., et al., 2020. A lack of evolution in the very bright end of the galaxy luminosity function from $z \approx 8$ to 10. *Monthly Notices of the Royal Astronomical Society*, 493(2):2059.
- Brand, J., 1995. *Lines of Light: The Sources of Dispersive Spectroscopy, 1800-1930*. Gordon & Breach Publ.
- Brisbin, D., Miettinen, O., Aravena, M., et al., 2017. An ALMA survey of submillimeter galaxies in the COSMOS field: Multiwavelength counterparts and redshift distribution. *A&A*, 608:A15.
- Burke, B.F. and Graham-Smith, F., 2009. *An Introduction to Radio Astronomy*.
- Burns, J.O., Feigelson, E.D., and Schreier, E.J., 1983. The inner radio structure of Centaurus A : clues to the origin of the jet X-ray emission. *ApJ*, 273:128.
- Calabretta, M.R. and Roukema, B.F., 2007. Mapping on the HEALPix grid. *MNRAS*, 381(2):865.
- Carilli, C.L., Lee, N., Capak, P., et al., 2008. Star Formation Rates in Lyman Break Galaxies: Radio Stacking of LBGs in the COSMOS Field and the Sub- μ Jy Radio Source Population. *ApJ*, 689(2):883.
- Casey, C.M., Chen, C.C., Cowie, L.L., et al., 2013. Characterization of SCUBA-2 450 μ m and 850 μ m selected galaxies in the COSMOS field. *MNRAS*, 436(3):1919.
- Ciliegi, P., Zamorani, G., Hasinger, G., et al., 2003. A deep VLA survey at 6 cm in the Lockman Hole. *A&A*, 398:901.
- Cimatti, A., Mignoli, M., Daddi, E., et al., 2002. The k20 survey. *Astronomy & Astrophysics*, 392(2):395.

- Condon, J.J., 1992. Radio emission from normal galaxies. *Annual Review of Astronomy and Astrophysics*, 30:575.
- Condon, J.J. and Ransom, S.M., 2016. *Essential Radio Astronomy*. Princeton University Press, sch - school edition edition.
- Cowie, L.L., Gardner, J.P., Lilly, S.J., et al., 1990. A K Band Deep Galaxy Survey. *ApJ*, 360:L1.
- de Zotti, G., Massardi, M., Negrello, M., et al., 2010. Radio and millimeter continuum surveys and their astrophysical implications. , 18(1-2):1.
- Delhaize, J., Heywood, I., Prescott, M., et al., 2021. MIGHTEE: are giant radio galaxies more common than we thought? *MNRAS*, 501(3):3833.
- Delvecchio, I., Daddi, E., Sargent, M.T., et al., 2021. The infrared-radio correlation of star-forming galaxies is strongly M_* -dependent but nearly redshift-invariant since $z \sim 4$. *A&A*, 647:A123.
- Dewdney, P.E., Hall, P.J., Schilizzi, R.T., et al., 2009a. The Square Kilometre Array. *IEEE Proceedings*, 97(8):1482.
- Dewdney, P.E., Hall, P.J., Schilizzi, R.T., et al., 2009b. The Square Kilometre Array. *IEEE Proceedings*, 97(8):1482.
- Dicke, R.H., Peebles, P.J.E., Roll, P.G., et al., 1965. Cosmic Black-Body Radiation. *ApJ*, 142:414.
- Djorgovski, S.G., Mahabal, A., Drake, A., et al., 2013. *Sky Surveys*, pages 223–281. Springer Netherlands, Dordrecht.
- Dubner, G. and Giacani, E., 2015. Radio emission from supernova remnants. *The Astronomy and Astrophysics Review*, 23(1).
- Duncan, K.J., Jarvis, M.J., Brown, M.J.I., et al., 2018. Photometric redshifts for the next generation of deep radio continuum surveys – II. Gaussian processes and hybrid estimates. *Monthly Notices of the Royal Astronomical Society*, 477(4):5177.
- Duncan, K.J., Kondapally, R., Brown, M.J.I., et al., 2021. The LOFAR Two-meter Sky Survey: Deep Fields Data Release 1. IV. Photometric redshifts and stellar masses. *A&A*, 648:A4.

- Edge, D.O., Shakeshaft, J.R., McAdam, W.B., et al., 1959. A survey of radio sources at a frequency of 159 Mc/s. , 68:37.
- Elder, F.R., Gurewitsch, A.M., Langmuir, R.V., et al., 1947. Radiation from Electrons in a Synchrotron. *Physical Review*, 71(11):829.
- Elvis, M., Civano, F., Vignali, C., et al., 2009a. The Chandra COSMOS Survey. I. Overview and Point Source Catalog. *ApJ*, 184(1):158.
- Elvis, M., Civano, F., Vignali, C., et al., 2009b. The Chandra COSMOS Survey. I. Overview and Point Source Catalog. *ApJ*, 184(1):158.
- Emerson, J. and Sutherland, W., 2010. The Visible and Infrared Survey Telescope for Astronomy (VISTA): Looking Back at Commissioning. *The Messenger*, 139:2.
- Fabian, A.C., 2012. Observational Evidence of Active Galactic Nuclei Feedback. *Annual Review of Astronomy and Astrophysics*, 50:455.
- Fanaroff, B., Lal, D.V., Venturi, T., et al., 2021. A new look at old friends – i. imaging classical radio galaxies with ugmrt and meerkat. *Monthly Notices of the Royal Astronomical Society*, 505(4):6003–6016.
- Fanaroff, B.L. and Riley, J.M., 1974. The morphology of extragalactic radio sources of high and low luminosity. *MNRAS*, 167:31P.
- Fernández Lorenzo, M., Sulentic, J., Verdes-Montenegro, L., et al., 2012. The AMIGA sample of isolated galaxies. X. A first look at isolated galaxy colors. *A&A*, 540:A47.
- Fernique, P., Boch, T., Donaldson, T., et al., 2014. MOC - HEALPix Multi-Order Coverage map Version 1.0. IVOA Recommendation 02 June 2014.
- Fleuren, S., Sutherland, W., Dunne, L., et al., 2012. Herschel-ATLAS: VISTA VIKING near-infrared counterparts in the Phase 1 GAMA 9-h data. *MNRAS*, 423(3):2407.
- Franceschini, A., Rodighiero, G., and Vaccari, M., 2008. Extragalactic optical-infrared background radiation, its time evolution and the cosmic photon-photon opacity. *A&A*, 487(3):837.
- Frank, J., King, A., and Raine, D.J., 2002. *Accretion Power in Astrophysics: Third Edition*.

- Fu, H., Yan, L., Scoville, N.Z., et al., 2010. Decomposing Star Formation and Active Galactic Nucleus with Spitzer Mid-infrared Spectra: Luminosity Functions and Co-evolution. *ApJ*, 722(1):653.
- Geach, J.E., Dunlop, J.S., Halpern, M., et al., 2016. The SCUBA-2 Cosmology Legacy Survey: 850 m maps, catalogues and number counts. *Monthly Notices of the Royal Astronomical Society*, 465(2):1789.
- Górski, K.M., Hivon, E., Banday, A.J., et al., 2005. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *ApJ*, 622(2):759.
- Gower, J.F.R., 1966. The source counts from the 4C survey. *MNRAS*, 133:151.
- Hafez, Y.A., Albaqami, F.H., Almutairi, A.Z., et al., 2019. Pre-correction of water vapour emission for radio astronomy observations at al-uyayna observatory. *Journal of Physics: Conference Series*, 1269(1):012006.
- Harrison, C.M., 2017. Impact of supermassive black hole growth on star formation. *Nature Astronomy*, 1:0165.
- Hartley, W.G., Almaini, O., Mortlock, A., et al., 2013. Studying the emergence of the red sequence through galaxy clustering: host halo masses at z $gt; 2$. *Monthly Notices of the Royal Astronomical Society*, 431(4):3045.
- Hasinger, G., Cappelluti, N., Brunner, H., et al., 2007. The XMM-Newton Wide-Field Survey in the COSMOS Field. I. Survey Description. *ApJ*, 172(1):29.
- Hatfield, P.W., Jarvis, M.J., Adams, N., et al., 2022. Hybrid photometric redshifts for sources in the COSMOS and XMM-LSS fields. *MNRAS*, 513(3):3719.
- Heywood, I., Jarvis, M.J., Hale, C.L., et al., 2021. MIGHTEE: Total intensity radio continuum imaging and the COSMOS / XMM-LSS Early Science fields. *MNRAS*.
- Holland, W.S., Robson, E.I., Gear, W.K., et al., 1999. SCUBA: a common-user submillimetre camera operating on the James Clerk Maxwell Telescope. *Monthly Notices of the Royal Astronomical Society*, 303(4):659.

- Hsu, L.T., Lin, L., Dickinson, M., et al., 2019. Near-infrared Survey and Photometric Redshifts in the Extended GOODS-North Field. *ApJ*, 871(2):233.
- Hurley-Walker, N., Filipović, M.D., Gaensler, B.M., et al., 2019. New candidate radio supernova remnants detected in the gleam survey over $345^\circ < l < 60^\circ$, $180^\circ < l < 240^\circ$. *Publications of the Astronomical Society of Australia*, 36:E045.
- Ilbert, O., Arnouts, S., McCracken, H. J., et al., 2006. Accurate photometric redshifts for the cfht legacy survey calibrated using the vimos vlt deep survey. *A&A*, 457(3):841.
- Intema, H.T., Jagannathan, P., Mooley, K.P., et al., 2017. The GMRT 150 MHz all-sky radio survey. First alternative data release TGSS ADR1. *A&A*, 598:A78.
- Ivezić, Ž., Menou, K., Knapp, G.R., et al., 2002. Optical and radio properties of extragalactic sources observed by the FIRST survey and the sloan digital sky survey. *The Astronomical Journal*, 124(5):2364.
- Janka, H.T., 2012. Explosion Mechanisms of Core-Collapse Supernovae. *Annual Review of Nuclear and Particle Science*, 62(1):407.
- Jansky, K.G., 1933. Radio Waves from Outside the Solar System. *Nature*, 132(3323):66.
- Jarvis, M.J., Bonfield, D.G., Bruce, V.A., et al., 2013. The VISTA Deep Extragalactic Observations (VIDEO) survey. *MNRAS*, 428(2):1281.
- Jarvis, M.J., Smith, D.J.B., Bonfield, D.G., et al., 2010. Herschel-ATLAS: the far-infrared-radio correlation at $z \lesssim 0.5$. *MNRAS*, 409(1):92.
- Jarvis, M.J., Taylor, A.R., Agudo, I., et al., 2017. The meerkat international ghz tiered extragalactic exploration (mightee) survey.
- Joh, K., Nagao, T., Wada, K., et al., 2021. Do gas clouds in narrow-line regions of Seyfert galaxies come from their nuclei? , 73(4):1152.
- Jonas, J. and MeerKAT Team, 2016. The MeerKAT Radio Telescope. In *MeerKAT Science: On the Pathway to the SKA*, page 1.
- Jonas, J.L., 2009. MeerKAT - The South African Array With Composite Dishes and Wide-Band Single Pixel Feeds. *IEEE Proceedings*, 97(8):1522.

- Jones, M., Adams, D., Lambourne, R., et al., 2004. *An Introduction to Galaxies and Cosmology*. S 282. Open University.
- Kauffmann, G., Heckman, T.M., Tremonti, C., et al., 2003. The host galaxies of active galactic nuclei. *MNRAS*, 346(4):1055.
- Kipnis, N., 1991. *History of the Principle of Interference of Light*. Historical studies. Springer.
- Kondapally, R., Best, P.N., Hardcastle, M.J., et al., 2021. The LOFAR Two-meter Sky Survey: Deep Fields Data Release 1. III. Host-galaxy identifications and value added catalogues. *A&A*, 648:A3.
- Laing, R.A., Guidetti, D., Bridle, A.H., et al., 2011. Deep imaging of Fanaroff–Riley Class I radio galaxies with lobes. *Monthly Notices of the Royal Astronomical Society*, 417(4):2789.
- Lazio, T.W., 2002. Radio-astronomy interferometry. In R.A. Meyers, editor, *Encyclopedia of Physical Science and Technology (Third Edition)*, pages 675–686. Academic Press, New York, third edition edition.
- Lehmer, B.D., Brandt, W.N., Alexander, D.M., et al., 2005. The Extended Chandra Deep Field-South Survey: Chandra Point-Source Catalogs. *ApJ*, 161(1):21.
- Longair, M.S., 2011. *High Energy Astrophysics*.
- Lutz, D., Poglitsch, A., Altieri, B., et al., 2011. PACS Evolutionary Probe (PEP) - A Herschel key program. *A&A*, 532:A90.
- Lynden-Bell, D., 1969. Galactic Nuclei as Collapsed Old Quasars. *Nature*, 223(5207):690.
- Maddox, N., Frank, B.S., Ponomareva, A.A., et al., 2021. Mightee-hi: The hi emission project of the meerkat mightee survey. *Astronomy Astrophysics*, 646:A35.
- Maini, A., Prandoni, I., Norris, R.P., et al., 2016. Infrared-faint radio sources in the SERVS deep fields. Pinpointing AGNs at high redshift. *A&A*, 596:A80.
- Marconi, A. and Hunt, L.K., 2003. The relation between black hole mass, bulge mass, and near-infrared luminosity. *The Astrophysical Journal*, 589(1):L21.
- Marconi, A., Risaliti, G., Gilli, R., et al., 2004. Local supermassive black holes, relics of active galactic nuclei and the X-ray background. *MNRAS*, 351(1):169.

- Mathis, J.S., Rumpl, W., and Nordsieck, K.H., 1977. The size distribution of interstellar grains. *ApJ*, 217:425.
- Mauch, T., Cotton, W.D., Condon, J.J., et al., 2020. The 1.28 GHz MeerKAT DEEP2 image. *The Astrophysical Journal*, 888(2):61.
- Mauch, T. and Sadler, E.M., 2007. Radio sources in the 6dFGS: local luminosity functions at 1.4 GHz for star-forming galaxies and radio-loud AGN. *MNRAS*, 375(3):931.
- McAlpine, K., Smith, D.J.B., Jarvis, M.J., et al., 2012. The likelihood ratio as a tool for radio continuum surveys with Square Kilometre Array precursor telescopes. *MNRAS*, 423(1):132.
- McCracken, H.J., Milvang-Jensen, B., Dunlop, J., et al., 2012. UltraVISTA: a new ultra-deep near-infrared survey in COSMOS. *A&A*, 544:A156.
- Middleton, M.J., Casella, P., Gandhi, P., et al., 2017. Paving the way to simultaneous multi-wavelength astronomy. , 79:26.
- Mohan, N. and Rafferty, D., 2015. PyBDSF: Python Blob Detection and Source Finder.
- Moneti, A., McCracken, H.J., Hudelot, P., et al., 2019. The fourth ultravista data release.
- Monnier, J.D. and Allen, R.J., 2013. *Radio and Optical Interferometry: Basic Observing Techniques and Data Analysis*, pages 325–373. Springer Netherlands, Dordrecht.
- Murphy, E.J., Momjian, E., Condon, J.J., et al., 2017. The GOODS-N Jansky VLA 10 GHz Pilot Survey: Sizes of Star-forming μ JY Radio Sources. *ApJ*, 839(1):35.
- Nagakura, T., Hosokawa, T., and Omukai, K., 2009. Star formation triggered by supernova explosions in young galaxies. *MNRAS*, 399(4):2183.
- Netzer, H., 2015. Revisiting the Unified Model of Active Galactic Nuclei. *Annual Review of Astronomy and Astrophysics*, 53:365.
- Oliver, S., Rowan-Robinson, M., Alexander, D.M., et al., 2000. The European Large Area ISO Survey - I. Goals, definition and observations. *MNRAS*, 316(4):749.
- Oliver, S.J., Bock, J., Altieri, B., et al., 2012. The Herschel Multi-tiered Extragalactic Survey: HerMES. *MNRAS*, 424(3):1614.

- Orenstein, B.J., Collier, J.D., and Norris, R.P., 2019. The redshift distribution of infrared-faint radio sources. *MNRAS*, 484(1):1021.
- Ortiz, R. and Guerrero, M.A., 2016. Ultraviolet emission from main-sequence companions of AGB stars. *MNRAS*, 461(3):3036.
- Oser, L., Ostriker, J.P., Naab, T., et al., 2010. The Two Phases of Galaxy Formation. *ApJ*, 725(2):2312.
- Panessa, F., Baldi, R.D., Laor, A., et al., 2019. The origin of radio emission from radio-quiet agn.
- Papadogiannakis, S., Goobar, A., Amanullah, R., et al., 2019. R-band light-curve properties of Type Ia supernovae from the (intermediate) Palomar Transient Factory. *MNRAS*, 483(4):5045.
- Patil, P., Nyland, K., Lacy, M., et al., 2019. Multiband Optical and Near-Infrared Properties of Faint Submillimeter Galaxies with Serendipitous ALMA Detections. *ApJ*, 871(1):109.
- Peebles, P.J.E., 1980. *The large-scale structure of the universe*.
- Penzias, A.A. and Wilson, R.W., 1965. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *ApJ*, 142:419.
- Pierre, M., Valtchanov, I., Altieri, B., et al., 2004. The XMM-LSS survey. Survey design and first results. , 2004(9):011.
- Press, W.H. and Schechter, P., 1974. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425.
- Radcliffe, J.F., Barthel, P.D., Garrett, M.A., et al., 2021. The radio emission from active galactic nuclei. *A&A*, 649:L9.
- Ramos Almeida, C., Levenson, N.A., Alonso-Herrero, A., et al., 2011. Testing the Unification Model for Active Galactic Nuclei in the Infrared: Are the Obscuring Tori of Type 1 and 2 Seyferts Different? *ApJ*, 731(2):92.
- Read, S.C., Smith, D.J.B., Gürkan, G., et al., 2018. The Far-Infrared Radio Correlation at low radio frequency with LOFAR/H-ATLAS. *MNRAS*, 480(4):5625.
- Reich, W., 2002. Radio observations of supernova remnants.

- Reynolds, S.P., 2008. Supernova remnants at high energy. *Annual Review of Astronomy and Astrophysics*, 46:89.
- Rosslowe, C.K. and Crowther, P.A., 2017. A deep near-infrared spectroscopic survey of the Scutum-Crux arm for Wolf–Rayet stars. *Monthly Notices of the Royal Astronomical Society*, 473(3):2853.
- Sadler, E.M., Jackson, C.A., Cannon, R.D., et al., 2002. Radio sources in the 2dF Galaxy Redshift Survey – II. Local radio luminosity functions for AGN and star-forming galaxies at 1.4 GHz. *Monthly Notices of the Royal Astronomical Society*, 329(1):227.
- Scholtz, J., Alexander, D.M., Harrison, C.M., et al., 2018. Identifying the subtle signatures of feedback from distant AGN using ALMA observations and the EAGLE hydrodynamical simulations. *MNRAS*, 475(1):1288.
- Schreiber, C., Pannella, M., Elbaz, D., et al., 2015. The Herschel view of the dominant mode of galaxy growth from $z = 4$ to the present day. *A&A*, 575:A74.
- Scoville, N., Aussel, H., Brusa, M., et al., 2007. The Cosmic Evolution Survey (COSMOS): Overview. *ApJ*, 172(1):1.
- Seymour, N., Stern, D., De Breuck, C., et al., 2007. The Massive Hosts of Radio Galaxies across Cosmic Time. *ApJ*, 171(2):353.
- Shields, G.A., 1999. A Brief History of Active Galactic Nuclei. *PASP*, 111(760):661.
- Shimwell, T.W., Röttgering, H.J.A., Best, P.N., et al., 2017. The LOFAR Two-metre Sky Survey. I. Survey description and preliminary data release. *A&A*, 598:A104.
- Shimwell, T.W., Tasse, C., Hardcastle, M.J., et al., 2019. The LOFAR Two-metre Sky Survey. II. First data release. *A&A*, 622:A1.
- Shin, J., Woo, J.H., Chung, A., et al., 2019. Positive and negative feedback of AGN outflows in NGC 5728. *The Astrophysical Journal*, 881(2):147.
- Simpson, C., 2017. Extragalactic radio surveys in the pre-Square Kilometre Array era. *Royal Society Open Science*, 4(7):170522.

- Singh, V., Wadadekar, Y., Ishwara-Chandra, C.H., et al., 2017. On the nature of infrared-faint radio sources in the Subaru X-ray Deep and Very Large Array-VIMOS VLT Deep Survey fields. *MNRAS*, 470(4):4956.
- Skrutskie, M.F., Cutri, R.M., Stiening, R., et al., 2006. The Two Micron All Sky Survey (2MASS). *AJ*, 131(2):1163.
- Smartt, S.J., 2009. Progenitors of Core-Collapse Supernovae. *Annual Review of Astronomy and Astrophysics*, 47(1):63.
- Smith, D.J.B., Dunne, L., Maddox, S.J., et al., 2011. Herschel-ATLAS: counterparts from the ultraviolet-near-infrared in the science demonstration phase catalogue. *MNRAS*, 416(2):857.
- Smith, D.J.B., Haskell, P., Gürkan, G., et al., 2021. The LOFAR Two-metre Sky Survey Deep Fields. The star-formation rate-radio luminosity relation at low frequencies. *A&A*, 648:A6.
- Smith, D.J.B., Hayward, C.C., Jarvis, M.J., et al., 2017. A complete distribution of redshifts for submillimetre galaxies in the SCUBA-2 Cosmology Legacy Survey UDS field. *MNRAS*, 471(2):2453.
- Smolčić, V. and Riechers, D.A., 2011. The Molecular Gas Content of $z \lesssim 0.1$ Radio Galaxies: Linking the Active Galactic Nucleus Accretion Mode to Host Galaxy Properties. *ApJ*, 730(2):64.
- Smolčić, V., Zamorani, G., Schinnerer, E., et al., 2009. A Radio View of the Sky: the Cosmic History of Star-Forming and AGN Galaxies. In W. Wang, Z. Yang, Z. Luo, and Z. Chen, editors, *The Starburst-AGN Connection*, volume 408 of *Astronomical Society of the Pacific Conference Series*, page 116.
- Smolčić, V., Novak, M., Bondi, M., et al., 2017. The vla-cosmos 3 ghz large project: Continuum data and source catalog release. *A&A*, 602:A1.
- Springel, V., White, S.D.M., Jenkins, A., et al., 2005. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435(7042):629.
- Sutherland, W. and Saunders, W., 1992. On the likelihood ratio for source identification. *MNRAS*, 259:413.

- Swenson, G. W., J. and Mathur, N.C., 1968. The Interferometer in Radio Astronomy. *IEEE Proceedings*, 56(12):2114.
- Taniguchi, Y., Kajisawa, M., Kobayashi, M.A.R., et al., 2015. The Subaru COSMOS 20: Subaru optical imaging of the HST COSMOS field with 20 filters*. , 67(6):104.
- Thompson, A.R., Moran, J.M., and Swenson, George W., J., 2017. *Interferometry and Synthesis in Radio Astronomy, 3rd Edition*.
- Tokunaga, A.T., Simons, D.A., and Vacca, W.D., 2002. The Mauna Kea Observatories Near-Infrared Filter Set. II. Specifications for a New JHKL'M' Filter Set for Infrared Astronomy. *PASP*, 114(792):180.
- Urry, C.M. and Padovani, P., 1995. Unified Schemes for Radio-Loud Active Galactic Nuclei. *PASP*, 107:803.
- van Haarlem, M.P., Wise, M.W., Gunst, A.W., et al., 2013. LOFAR: The LOw-Frequency ARray. *A&A*, 556:A2.
- van Weeren, R.J., de Gasperin, F., Akamatsu, H., et al., 2019. Diffuse Radio Emission from Galaxy Clusters. , 215(1):16.
- Vardoulaki, E., Jiménez Andrade, E.F., Karim, A., et al., 2019. A closer look at the deep radio sky: Multi-component radio sources at 3 GHz VLA-COSMOS. *A&A*, 627:A142.
- Varga, J., Csabai, I., and Dobos, L., 2012. Revealing a strongly reddened, faint active galactic nucleus population by stacking deep co-added images. *MNRAS*, 426(2):833.
- Weltman, A., Bull, P., Camera, S., et al., 2020. Fundamental physics with the Square Kilometre Array. , 37:e002.
- White, S.V., Jarvis, M.J., Häußler, B., et al., 2015. Radio-quiet quasars in the VIDEO survey: evidence for AGN-powered radio emission at S1.4 GHz
lt; 1 mJy. *Monthly Notices of the Royal Astronomical Society*, 448(3):2665.
- Whittam, I.H., 2018. Understanding mechanical feedback from hergs and lergs. *Proceedings of the International Astronomical Union*, 14(A30):86–89.

- Whittam, I.H., Jarvis, M.J., Hale, C.L., et al., 2022. MIGHTEE: the nature of the radio-loud AGN population. *MNRAS*, 516(1):245.
- Whittam, I.H., Riley, J.M., Green, D.A., et al., 2016. The faint source population at 15.7 GHz – III. A high-frequency study of HERGs and LERGs. *Monthly Notices of the Royal Astronomical Society*, 462(2):2122.
- Williams, W.L., Hardcastle, M.J., Best, P.N., et al., 2019. The LOFAR Two-metre Sky Survey. III. First data release: Optical/infrared identifications and value-added catalogue. *A&A*, 622:A2.
- Willott, C.J., Rawlings, S., Jarvis, M.J., et al., 2003. Near-infrared imaging and the K—z relation for radio galaxies in the 7C Redshift Survey. *Monthly Notices of the Royal Astronomical Society*, 339(1):173.
- Wilson, A.S. and Colbert, E.J.M., 1995. The Difference between Radio-loud and Radio-quiet Active Galaxies. *ApJ*, 438:62.
- Yu, Q. and Tremaine, S., 2002. Observational constraints on growth of massive black holes. *MNRAS*, 335(4):965.
- Zamojski, M.A., Schiminovich, D., Rich, R.M., et al., 2007. Deep GALEX Imaging of the COSMOS HST Field: A First Look at the Morphology of $z \sim 0.7$ Star-forming Galaxies. *ApJ*, 172(1):468.