Studying Galaxy Evolution with Modern Radio Surveys

Catherine Hale

Merton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Hilary 2019

For Katherine, who is always in my thoughts

The biggest thanks I need to convey is to my supervisor, Matt Jarvis. My DPhil would not have been such a great experience without him, especially as his confidence in me has been unwavering (even when I wavered). I will never be able to thank him enough for his help, guidance and positivity. His support has helped me develop into a better scientist and provided me with so many opportunities which made the DPhil experience all the more fun.

To my family, I would never have been able to finish this DPhil without your love and support. Firstly to my parents, Sandra and David who have both been incredible in being there for me through every step of my studies. I owe a debt of gratitude you. To my sister, Rebecca, you have always been both a best friend and a sister and you have never failed to make me smile. Thank you for always being such a fun loving person and always being at the other end of the phone to sing me through any situation. Finally to my Grandad, Bob, thank you for always telling me that you are proud of me. I love you all so much. And to Jeremy, Derryn, Kyra and Elyse. Thank you for being so amazing both whilst I was in Perth and throughout my life, you are all family to me.

Special mention also needs to go to Sam for putting up with me as an office mate for all four years; Sergio for providing many opportunities to dance in the office; Naomi for her words of wisdom, support and cheese jokes; Paul for always rallying us for lunch earlier each day; Joe for all the chats over tea and cat pictures and finally Rebecca B for all the chats and reassurance! And to the many others who have helped along the way, including: Aprajita, Ashling, Ben, Courtenay, Greg, Holly, Josie, Katherine, Leah, Lucy, Liz, Kit, Marius, Nathan, Neetu, Nicole, Peter, Ravin, Rebecca E., Will and Zahra. Thank you for making my Oxford experience so great.

Thanks must also go to those people I have collaborated with throughout my DPhil. To Ian and Wendy, thanks for all your help and wisdom. To Aaron, Luke and Simon, thanks for helping me have the opportunity to spend 3 months at ICRAR, I had such a great experience there.

Finally to Robert, thank you for always believing in me. You have always been there and I wouldn't have been nearly as happy throughout these years without you.

Declaration - *Catherine Hale, March 2019*

I declare the work that is presented in this Thesis is my own work, unless otherwise stated, and has been completed during my time studying for a DPhil in Astrophysics at the University of Oxford, commencing in October 2015. My supervisor throughout the duration of my studies has been Prof. Matt Jarvis and the work presented in this thesis has not been submitted for any other degree. At the time of submission, I have three papers, their relation to this thesis is described below:

- The work that is presented in Chapter 2 is based on the work: *The clustering and bias of radio-selected AGN and star-forming galaxies in the COSMOS field* (Hale et al., 2018). For this paper, I am the lead author, my co-authors are: M. J. Jarvis, I. Delvecchio, P. W. Hatfield, M. Novak, V. Smolčić and G. Zamorani.

- The work presented in Chapter 3 has currently not been submitted to any journal. However the work may be developed into a paper in the future.

- The work presented in Chapter 4 is based on my paper: *Radio Source Extraction with ProFound* (Hale et al., 2019a). I am lead author on this paper and the work was completed as part of a Long Term Attachment (LTA) funded by the Science and Technologies Facilities Council (STFC) to Perth, Australia. The co-authors are: A. Robotham, L. Davies, M. Jarvis, S. Driver, I. Heywood.

- The work presented in Chapter 5 is based on the work from: *LOFAR Observations of the XMM-LSS Field* (Hale et al., 2019b). I am lead author on this paper, the co-authors are: W. Williams, M. Jarvis, M. Hardcastle, L. Morabito, T. Shimwell, C. Tasse, P. Best, J. Harwood, I. Heywood, I. Prandoni, H. Röttgering, J. Sabater, D. Smith, R. van Weeren. In this work, T. Shimwell ran the PRE-FACTOR pipeline for the data and W. Williams helped me with the data reduction pipelines.

This Thesis was created using LaTeX using the ociamthesis.cls class[1] and `plainnat` bibliography style file. Except for some work in Chapter 4 where I used R, my work made use of `Python` and many packages including `numpy` (van der Walt et al., 2011), `astropy` (Astropy Collaboration et al., 2013), `matplotlib` (Hunter, 2007) and `scipy` (Jones et al., 2001). As well as this, packages such as TOPCAT (Taylor, 2011) and `ds9` (Joye & Mandel, 2003) have been incredibly helpful. For this work, I have also made use of the ADS[2] and arXiv[3].

[1] `https://www.maths.ox.ac.uk/members/it/faqs/latex/thesis-class`
[2] `https://ui.adsabs.harvard.edu/`
[3] `https://arxiv.org/archive/astro-ph`

Abstract

Modern radio surveys are transforming our view of the extragalactic sky, observing both Star Forming Galaxies (SFGs) and Active Galactic Nuclei (AGN) over large fractions of the history of the Universe. This allows us to investigate their evolution. In Chapter 2, I investigate how the relationship between galaxies and their dark matter environments evolves, through studying their angular clustering. Using the VLA 3 GHz COSMOS Survey (Smolčić et al., 2017a,b), I find that AGN reside in more massive haloes than SFGs and that radiatively inefficient AGN tend to reside in more massive haloes than their efficient counterparts. I investigate this further, in Chapter 3, by parameterising the Conditional Luminosity Function (CLF), which links environment to the luminosity distribution of these sources. This is novel at radio frequencies. From this, I study how the typical properties of galaxies vary with environment (or luminosity), again concluding that inefficient AGN reside in the most massive haloes, but that the halo mass supporting AGN and SFGs is smaller at higher redshifts. In Chapter 4, I investigate whether the multi-wavelength source extraction software ProFound (Robotham et al., 2018) can accurately quantify the flux densities of radio sources, irrespective of morphology. This is especially important for modern telescopes with a combination of long and short baselines. I find that ProFound is able to outperform commonly used software (PyBDSF and AEGEAN) when complex morphologies are considered and perform competitively when modelling sources with smooth morphologies. Finally, in Chapter 5, I present the deepest low frequency (144 MHz) observations of the XMM-LSS field, using LOFAR. This provides a low frequency catalogue which can be used in conjunction with higher frequency radio surveys. Combining the work of this Thesis will enable greater understanding of the relationship between galaxies and their environments, using future deep extragalactic radio surveys and this is discussed in Chapter 6.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Since the Big Bang, our Universe has been evolving with matter interacting to form different elements, molecules and importantly to form the stars and galaxies observed today. Whilst these interactions have predominantly occurred due to the attraction of matter under the influence of gravity, the Universe has nonetheless been expanding and continues to do so. Hubble (1929) first noted this expansion of the Universe, observing that the velocities of galaxies that were further from the Earth are moving away faster than those closer to the Earth. He described this relation between velocities (v) of local galaxies and their distances (d) by the proportionality law, now known as Hubble's law:

$$v = H_0 d, \tag{1.1}$$

where H_0 is the proportionality constant, known as Hubble's constant. Hubble first determined this constant to have a value of $\sim$500 kms^{-1} Mpc^{-1} (Hubble, 1929).

With this expansion, light from objects moving away from us (the observer) are described to be 'red'-shifted. This is due to the fact that with the expanding Universe, photons emitted at one wavelength become stretched as space expands on their journey to an observer and so are shifted towards redder colours, the term redshift (z)

is used to quantify this shift (see e.g. Carroll & Ostlie, 2006; Peacock, 1999; Peebles, 1980; Hogg, 1999). By observing spectral lines of known emission wavelength, the redshift can be calculated using:

$$z = \frac{\lambda_{obs} - \lambda_{em}}{\lambda_{em}} = \frac{\nu_{em} - \nu_{obs}}{\nu_{obs}}, \tag{1.2}$$

where λ_{obs} (ν_{obs}) corresponds to the observed wavelength (frequency) for the spectral line and λ_{em} (ν_{em}) the corresponding wavelength (frequency) that the spectral line was emitted at.

Whilst Hubble's law holds locally (see e.g. Freedman et al., 2001; Peacock, 1999; Carroll & Ostlie, 2006; Riess et al., 2016), deviations are known at higher redshifts (Kristian et al., 1978; Riess et al., 1998; Schmidt et al., 1998). Without accurate measurements of distances, the idea of where galaxies are relative to the age of the Universe and hence the evolution of galaxies across the history of the Universe cannot be understood. However, it is not just with redshifts and Hubble's constant that the distances to galaxies can be calculated. Crucial is an understanding of the different components that constitute our Universe and how these have evolved. This is investigated through the study of cosmology.

1.1 Cosmology

Despite our observations of galaxies across the electromagnetic spectrum being a result of the baryons that constitute 'ordinary matter', this turns out to only contribute a small fraction of the current 'make-up' of the Universe (∼4-5%; e.g. Spergel et al., 2003; Planck Collaboration et al., 2018) and instead there are other contributions to the Universe. The principles of cosmology discussed in this section are covered by many texts relating to cosmology (see e.g. Peacock, 1999; Dodelson, 2003) and an overview is given here. The contributions to the total fractional energy density in the

Universe can be given as:

$$1 = \Omega_M(z) + \Omega_R(z) + \Omega_k(z) + \Omega_\Lambda(z), \tag{1.3}$$

where the fractional energy density for a component (i), Ω_i, is given by $\frac{\rho_i}{\rho_c}$. Here ρ_i corresponds to the density for that component and the critical density, ρ_c is given as:

$$\rho_c = \frac{3H^2}{8\pi G}, \tag{1.4}$$

where H quantifies the growth rate of the Universe, given by $\left(\frac{\dot{a}}{a}\right)$ where a is the scale factor of the Universe. The current value of this growth rate, $\left(\frac{\dot{a}}{a}\right)_0$ is known as the Hubble Constant, H_0, as discussed earlier. The components in Equation 1.3 relate to matter (Ω_M), radiation (Ω_R), curvature (Ω_k) (i.e. whether the Universe is flat, or not) and the cosmological constant (Ω_Λ) and their relative contributions have varied across the history of the Universe. Whilst the density in the cosmological constant has no redshift dependence, $\rho_M(z) \propto (1+z)^3$ and $\rho_R(z) \propto (1+z)^4$. For a flat Universe, radiation dominated at early times, whilst since then matter and subsequently dark energy have dominated. This can be explained through the redshift dependence of each component. At large redshifts, radiation will dominate due to its $(1+z)^4$ dependence, whilst at late times, as is currently the case, the cosmological constant will dominate.

Current measurements of these cosmological parameters are made through a variety of techniques such as the lensing of galaxies (e.g. Benjamin et al., 2007; DES Collaboration et al., 2018a) and through Type 1a Supernovae distance measurements (e.g. Rest et al., 2014; DES Collaboration et al., 2018b). One of the most precise methods, though, for determining the cosmological parameters originates from measuring the power spectrum of the Cosmic Microwave Background (CMB, e.g. Spergel et al., 2003; Planck Collaboration et al., 2018). The CMB is the radiation left over from the Big Bang (Dicke et al., 1965), discovered by Penzias & Wilson (1965). These

studies suggest we live in a flat universe (i.e. angles in a triangle sum exactly to 180°) with little radiation contribution ($\Omega_{R,0} \sim 0$) and with ∼30% of energy density in matter ($\Omega_{M,0} \sim 0.3$) and the corresponding ∼70% of energy density is in the cosmological constant ($\Omega_{\Lambda,0} \sim 0.7$). This contribution from the cosmological constant (or dark energy) is thought to be driving the acceleration of the Universe, which has been observed through studying the distances to Type Ia supernovae (Riess et al., 1998; Schmidt et al., 1998).

Whilst 30% of the energy density today is within matter, as I stated earlier, only ∼5% of this contribution is from ordinary matter. The remaining ∼25% is thought to originate from dark matter (see e.g. Spergel et al., 2003; Planck Collaboration et al., 2014, 2018). Whilst dark matter cannot be seen, hence the term dark, its presence was first inferred by Zwicky (1933) whilst investigating the Coma cluster using its velocity dispersion. Zwicky observed that the mass of the cluster must be larger than the mass inferred from the light emitted by the galaxy. Around 50 years later, further evidence for dark matter was found due to its effect on the dynamics of spiral galaxies (e.g. Rubin et al., 1980, 1985; Kent, 1986, 1987, 1988).

With the combination of both knowledge of redshift for a source, as well as the cosmological parameters, it is possible to determine the distances to galaxies in the Hubble flow. In a flat Universe, the distance (D) is given by (see e.g. Hogg, 1999):

$$D = \frac{c}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_{M,0}(1+z')^3 + \Omega_{k,0}(1+z')^2 + \Omega_{\Lambda,0}}}. \tag{1.5}$$

As mentioned earlier, whilst Hubble first measured the proportionality constant between the distances and velocities of galaxies, H_0∼500 kms^{-1} Mpc^{-1}, current measurements place this value at ∼70 kms^{-1} Mpc^{-1} (Spergel et al., 2003; Addison et al., 2013; Planck Collaboration et al., 2018). However there are tensions between measurements of the Hubble constant using distance ladder measurements (Riess et al., 2009) and those from the CMB (Spergel et al., 2003; Planck Collaboration et al.,

2018). For the purpose of astronomy, though, it is often useful to instead describe distances in terms of the luminosity distance (D_L). This is useful as it can relate the intrinsic luminosities (L) that galaxies emit to the observed fluxes that we receive as an observer (S). This relation is given by:

$$S = \frac{L}{4\pi D_L^2}, \tag{1.6}$$

where this distance, D_L, can be related to the distance in Equation 1.5, and hence the cosmological parameters that describe the universe by the relation $D_L = (1+z)D$ (Hogg, 1999; Carroll et al., 1992). This additional $(1+z)$ factor arises from the change in arrival rate of photons as the Universe expands.

For this Thesis, I will assume a standard cosmology of a flat, radiation-less Universe (i.e. $\Omega_{k,0} = 0$ and $\Omega_{R,0} = 0$) where the contribution of matter to the total energy density is 30% ($\Omega_{M,0} = 0.3$) and the majority of energy density is in dark energy ($\Omega_{\Lambda,0} = 0.7$). I use a value for Hubble's constant (H_0) of 70 kms^{-1} Mpc^{-1}, or $100h$ kms^{-1} Mpc^{-1}, where $h = 0.7$. These are standard values that are often used in the literature (e.g. Spergel et al., 2003; Mauch & Sadler, 2007; Lindsay et al., 2014b; Pracy et al., 2016; Smolčić et al., 2017b). The value of σ_8, which relates to the amplitude of the power spectrum at 8 Mpc h^{-1} (see e.g. Primack, 2005; Planck Collaboration et al., 2014), is taken as 0.8, consistent with Spergel et al. (2003); Planck Collaboration et al. (2016).

1.2 Large Scale Structure

Knowledge of the cosmological parameters that describe our Universe and hence the distances to galaxies has allowed greater understanding of the distribution of matter throughout the Universe. Large galaxy surveys allow an image of the spatial locations of galaxies within the Universe to be built up, with surveys such as the 2dF Galaxy

Redshift Survey (2dFGRS; Colless et al., 2001) and the Sloan Digital Sky Survey (SDSS; Tegmark et al., 2004). These large surveys have shown how galaxies are found in large groups, along filaments as well as in voids (see e.g. von Benda-Beckmann & Müller, 2008; Wen et al., 2012; Pan et al., 2012; Zhang et al., 2013).

However, the distribution of baryonic matter in the Universe is still challenging to model with large cosmological simulations. The interaction between baryons is governed by not only gravity but also the interaction between particles such as in forming different elements and molecules. Also challenging is the lack of mass and spatial resolution. For simulations such as EAGLE (Schaye et al., 2015), HORIZON-AGN (Dubois et al., 2014) and ILLUSTRIUS (Vogelsberger et al., 2014), these have a mass resolution O($\gtrsim 10^5 - 10^6 M_\odot$) with corresponding spatial resolution O($\gtrsim$0.1 kpc). This also means there will be missing physics within these simulations. For dark matter, on the other hand, the large scale distribution is thought to be well understood in relation to the assumed cosmology from using cosmological simulations (e.g. Navarro et al., 1996, 1997), because dark matter is purely governed by gravity.

In regions of over-densities, this dark matter will clump together to form dark matter haloes. These haloes will form where the density of the dark matter is greater than a critical value which ensures gravitational collapse (see Figure 8 of Peacock, 2003). It is in these haloes where galaxies are likely to form. Therefore galaxies are known to be biased tracers of dark matter (e.g. Kaiser, 1984; Peacock, 1999; Cooray & Sheth, 2002; Mo et al., 2010), meaning that they are more clustered than dark matter. Using functional forms for the distribution of haloes, it is possible to find a description of bias, b, related to halo mass (see e.g. Mo & White, 1996; Sheth & Tormen, 1999; Peacock, 2003; Tinker et al., 2010; Desjacques et al., 2018, and `http://www.damtp.cam.ac.uk/user/examples/3R2La.pdf`), a summary of this is given here. One such method that has been well discussed in the literature is from Mo & White (1996). In their work, Mo & White (1996) used the mass function

formalism of Press & Schechter (1974) given by:

$$n_M \propto \nu \exp(-\nu^2/2), \tag{1.7}$$

(see e.g. Press & Schechter, 1974; Peacock, 2003; Desjacques et al., 2018) where n_M is the number density of haloes and ν is a parameter given by $\delta_c/\sigma(M)$. δ_c is the critical density for which a density above (or equal) to this defines a halo having formed. $\sigma(M)$ is the variance of the density field within a given mass. It is using this mass function description that an analytic form for galaxy bias can be described. In this Press-Schechter formalism, Mo & White (1996) found that the galaxy bias could be rewritten as:

$$b(\nu) = 1 + \frac{\nu^2 - 1}{\delta_c}, \tag{1.8}$$

see e.g. Peacock (2003); Desjacques et al. (2018). However, as the Press & Schechter (1974) formalism for n_M was found to not be an accurate representation of the dark matter simulations on some scales (see Figure 2 of Sheth & Tormen, 1999), a modified formalism by Sheth & Tormen (1999) provided a different analytical model for the galaxy bias, given by:

$$b(\nu) = 1 + \frac{a\nu - 1}{\delta_c} + \frac{2p/\delta_c}{1 + (a\nu)^p}, \tag{1.9}$$

where a and p are constants with values 0.707 and 0.3 respectively (see Sheth & Tormen, 1999; Desjacques et al., 2018). These analytic relations between halo mass, M_h, (which can be calculated from ν) and b therefore mean that through studying the bias of galaxies observed in large galaxy surveys, it is therefore possible to determine the average mass of haloes that these galaxies reside in. Models for the bias are continuing to be developed, such as in Tinker et al. (2010), which will be used for the work in Chapters 2 and 3.

Measuring the clustering of galaxies, by determining the two-point correlation function (TPCF), is one way to determine the bias of galaxies. The TPCF quantifies the excess number of galaxies compared to that of a random distribution of galaxies at different separation scales. This can be described by using the spatial distribution of sources (the spatial correlation function, $\xi(r)$) as well as the projected angular spacing of galaxies (the angular correlation function, $\omega(\theta)$) (e.g. Peebles, 1980; Peacock, 1999; Coil, 2013). The determination of the spatial correlation function, $\xi(r)$, is essential to determine bias, their relation is given by:

$$b^2(z) = \frac{\xi_{gal}(r, z)}{\xi_{DM}(r, z)}, \tag{1.10}$$

where $\xi_{DM}(r, z)$ is the spatial correlation function of dark matter (see e.g. Mo & White, 1996; Peebles, 1980; Peacock & Smith, 2000). In this, I note that bias evolves with redshift, as does the spatial clustering of both galaxies and dark matter. This will be important in both the work of Chapters 2 and 3.

1.2.1 The Angular Two-Point Correlation Function, $\omega(\theta)$

Whilst the spatial correlation function, as described above, is essential to understand galaxy bias, it requires knowledge of accurate redshifts for its determination. However, the angular correlation function, $\omega(\theta)$, uses the projected angular spacings between galaxies observed over the celestial sphere to describe their clustering (e.g. Peebles, 1980; Blake & Wall, 2002; Overzier et al., 2003; Wilman et al., 2003; Lindsay et al., 2014b; Hatfield et al., 2016; Chen & Schwarz, 2016; Magliocchetti et al., 2017). This does not require redshift information. It can be defined (as in Peebles, 1974; Jones et al., 2004; Coil, 2013) through:

$$\delta P = \sigma[1 + \omega(\theta)]\delta\Omega, \tag{1.11}$$

which relates the probability of observing galaxy pairs within certain angular separa-

tions (δP) to the solid angle element being observed ($\delta\Omega$), for a given surface density (σ). For radio surveys, having accurate redshift information for all sources is uncommon and hence the angular correlation function is often relied upon (Blake & Wall, 2002; Wilman et al., 2003; Chen & Schwarz, 2016).

To estimate the angular clustering, in practice, the number of pairs of galaxies within certain angular separations are compared to what would be expected from a completely random distribution of sources. Comparing pairs of galaxies within different angular separations in this way is known as using an estimator (e.g. Landy & Szalay, 1993; Hamilton, 1993). This random catalogue should be realistic and take into account any regions where sources cannot be or are less likely to be observed due to the noise properties or masked regions. Naively, this excess probability would be defined by Equation 1.12, where $\overline{DD}(\theta)$ and $\overline{RR}(\theta)$ are the number of pairs of sources within a certain angular separation for the data and random catalogues respectively.

$$\omega(\theta) = \frac{\overline{DD}(\theta)}{\overline{RR}(\theta)} - 1 \tag{1.12}$$

These need to be normalised over all angles to take into account differences in the number of sources for each catalogue. However, estimators of $\omega(\theta)$ were investigated in Landy & Szalay (1993) who determined that the estimator given in Equation 1.13 was preferred, as its errors were more similar to the expected Poissonian errors. This has been used frequently in the literature (see e.g. Blake & Wall, 2002; Blake et al., 2004; Lindsay et al., 2014a,b; Hatfield et al., 2016; Retana-Montenegro & Röttgering, 2017).

$$\omega(\theta) = \frac{\overline{DD}(\theta)}{\overline{RR}(\theta)} - \frac{2\overline{DR}(\theta)}{\overline{RR}(\theta)} + 1. \tag{1.13}$$

Over a large angular range, $\omega(\theta)$ is observed to behave as a power law (Davis & Peebles, 1983; Davis et al., 1988; Roche & Eales, 1999; Norberg et al., 2002), with a

slope similar to -0.8. As galaxies are a biased tracer of dark matter, this power law behaviour is expected to be adhered to over the largest scales. On these large scales, this corresponds to the clustering of galaxies within different dark matter haloes. Clustering on these scales is often known as the '2-halo' clustering term (Cooray & Sheth, 2002; Yang et al., 2003; Zheng et al., 2007; Hatfield et al., 2016). It is behaviour on these large scales that is related to the determination of bias in Equation 1.10 (see e.g. Peebles, 1980; Peacock & Smith, 2000; Hatfield et al., 2016).

However on the smaller angular scales, if a survey is sensitive enough and has sufficient angular resolution, then the '1-halo term' clustering may also be observed. This '1-halo term' relates to the clustering of galaxies from within the same dark matter halo (see e.g. Cooray & Sheth, 2002; Yang et al., 2003). This describes the clustering between that of the brighter 'central' galaxies and 'satellite' galaxies from within the same dark matter halo (see e.g. Zehavi et al., 2004; Blake et al., 2008; Hatfield et al., 2016). These two terms are illustrated by Figure 1.1, taken from Yang et al. (2003), for the spatial correlation function. This excess of clustering on small scales due to the '1-halo term' is yet to be observed in radio surveys, due to the sensitivity requirements necessary (but see Chapter 6). However, an excess of clustering in radio surveys on these small scales has been observed in work by, for example, Blake & Wall (2002) who measured a slope of -3.4 below angles of $\sim 0.1^{\circ}$ in NVSS (Condon et al., 1998). This clustering however arises from multi-component objects. For example, double lobed objects (such as FRII type sources, see Section 1.4) that have not been identified as arising from the same source, will cause an excess of clustering on these small scales. Not properly identifying these multi component objects may therefore lead to confusion when attempting to measure the '1-halo' clustering term from radio continuum surveys in the future.

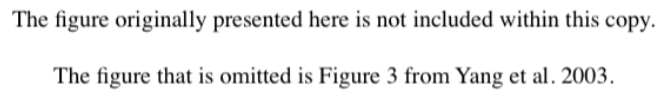


Figure 1.1: Figure 3 from Yang et al. (2003) of the spatial clustering of galaxies. This indicates the two clustering contributions: the 1-halo clustering term from the clustering of galaxies from within the same dark matter halo and the 2-halo term indicates the clustering of galaxies from within different dark matter haloes.

1.2.2 Relating Angular Clustering to Bias

Whilst $\omega(\theta)$ is typically calculated in radio surveys, due to the lack of accurate redshift information from these surveys (see Section 1.4), it is $\xi(r)$ that is essential to calculate the bias and to relate galaxies to their halo masses (as in Equation 1.10). This means that redshift information is required. However, by using Limber inversion (see e.g. Limber, 1953; Peebles, 1980; Overzier et al., 2003; Lindsay et al., 2014a,b) spectroscopic redshifts are not essential and instead the distribution of photometric redshifts can be relied upon. The spatial correlation function is also observed to have a power law distribution, $\xi(r) = r_0 = (r_0/r)^{\gamma}$ (Peebles, 1980; Saunders et al., 1992; Peacock, 1999; Zehavi et al., 2002; Hawkins et al., 2003; Carroll & Ostlie, 2006), where r_0 is known as the clustering length. The bias for a population of galaxies can then be determined by comparing this to the dark matter $\xi_{DM}(r)$ which leads to a definition

for b given by:

$$b(z) = \left(\frac{r_0(z)}{8}\right)^{\gamma/2} \frac{J_2^{1/2}}{\sigma_8 D_g(z)/D_g(0)} \tag{1.14}$$

as discussed in Koutoulidis et al. (2013) and used in Lindsay et al. (2014a,b). Here $D_g(z)$ is the cosmological growth factor (see e.g. Carroll et al., 1992; Hamilton, 2001) and $J_2 = 72/[(3-\gamma)(4-\gamma)(6-\gamma)2^\gamma]$. Limber inversion (Limber, 1953) can then be used to determine the clustering length r_0, through relating the amplitude (A) and slope $(1-\gamma)$ of the angular two-point correlation function to the redshift distribution of the sources observed, $N(z)$. r_0 can be determined through:

$$A = {r_0}^\gamma H_\gamma \left(\frac{H_0}{c}\right) \frac{\int_0^\infty N^2(z)(1+z)^{\gamma-(3+\epsilon)} D^{1-\gamma}(z) E(z) dz}{\left[\int_0^\infty N(z) dz\right]^2}, \tag{1.15}$$

(as in Overzier et al., 2003; Wilman et al., 2003; Lindsay et al., 2014b). The other terms in Equation 1.15 are all related to the cosmology being assumed or to the slope of $\omega(\theta)$. H_γ is defined in 1.16 and is related to the projection of the correlation function (see Koutoulidis et al., 2013) and $D(z)$ is defined in Equation 1.5. Finally, $E(z)$ is related to the assumed cosmology and is given by Equation 1.17. ϵ is chosen as $\gamma - 3$, which ensures that comoving clustering is being considered (Rengelink, 1999; Lindsay et al., 2014b).

$$H_\gamma = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{\gamma-1}{2})}{\Gamma(\frac{\gamma}{2})}, \tag{1.16}$$

$$E(z) = [\Omega_{m,0}(1+z)^3 + \Omega_{k,0}(1+z)^2 + \Omega_{\Lambda,0}]^{\frac{1}{2}}. \tag{1.17}$$

Using Equation 1.15 in order to determine r_0 and hence b will be crucial to the work in Chapters 2 and 3 in order to investigate the relationship between different galaxy populations and their dark matter haloes.

1.2.3 The Relationship Between Galaxies and Dark Matter Haloes

As galaxies typically form in dark matter haloes, much work has been done to investigate the relationship between galaxy bias and their properties. This includes studying the relationship between bias and galaxy luminosities and colours (e.g. Zehavi et al., 2005; Shen et al., 2009; Skibba et al., 2014; Dolley et al., 2014) as well as star formation rate and stellar masses (e.g. Hatfield et al., 2016; Tinker et al., 2017; Cochrane et al., 2018). These have all suggested that clustering increases with both stellar mass and star formation rates and that redder galaxies are much more clustered than bluer ones. Work by Mandelbaum et al. (2009) has also suggested that galaxies hosting radio AGN are typically in more massive haloes to those in which a radio AGN is not present. These all suggest that the environments in which these galaxy populations reside are related to their observable properties, suggesting these dark matter environments influence galaxies. I investigate this in Chapters 2 and 3.

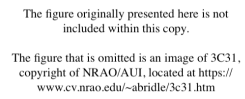
1.3 Active Galactic Nuclei

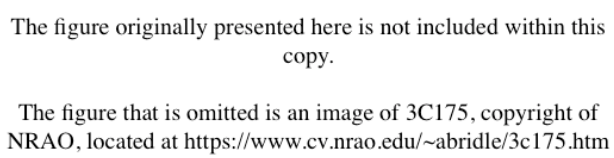
Bulge dominated galaxies are believed to have a supermassive black hole (SMBH) at their centre (Rees, 1998) which is also thought to affect their properties. This influence has been inferred from tight correlations between black hole mass (M_{BH}) and galaxy bulge mass (Magorrian et al., 1998) or stellar velocity dispersion (Ferrarese & Merritt, 2000), as well as the fact that stellar mass growth appears to trace black hole growth (Shankar et al., 2009; Yang et al., 2018). The formation of such large black holes so early in the age of the Universe (e.g. McLure & Jarvis, 2002; Netzer et al., 2007; Momjian et al., 2014) is one part of astronomy which is still not well understood and is being investigated using cosmological simulations (see e.g. Taylor & Kobayashi, 2014; Regan et al., 2018).

If a supermassive black hole within a galaxy is accreting large amounts of material, it is said to have an Active Galactic Nucleus (AGN). For many years, the typical picture of an AGN involves a black hole surrounded by an accretion disk, which provides the source for material to accrete onto the object, which is then surrounded by a dusty torus and clouds of gas orbiting the black hole (Antonucci, 1993; Urry & Padovani, 1995). This accretion disk is geometrically thin, but optically thick, which means that radiation is not easily able to escape the disk (Carroll & Ostlie, 2006) whilst the torus is geometrically thick (see e.g. Urry & Padovani, 1995; Heckman & Best, 2014). This large infall of material heats up as it accretes onto the black hole which may produce jets that can penetrate to much greater scales than the size of the galaxies which they inhabit.

Within this model, the observed properties of AGN are determined based on the orientation of the observer to the AGN. This orientation effect is useful in explaining why some AGN have narrow emission lines (Type II AGN) where the line of sight is obscured by the torus. For Type I AGN, the observer is oriented to be able to see more directly onto the black hole, allowing observations of broad emission lines and emission from the accretion disk which, for example, can be seen in the spectra of Quasars (Peterson, 1997; Vanden Berk et al., 2001; Carroll & Ostlie, 2006). AGN can also be defined as either Radio-loud or Radio-quiet based on the ratio of their radio to optical luminosities (see e.g. Kellermann et al., 1989; Becker et al., 1997).

Whilst AGN can be classified by the emission line properties observed, a morphological classification for radio AGN (that have jets) has been used extensively in the literature. This classification describes sources as Fanaroff-Riley Type I or II (FRI/FRII) sources (Fanaroff & Riley, 1974). FRII sources have powerful jets that have hot spots where they impact with the external inter-galactic medium (see the right hand panel of Figure 1.2). FRI sources, on the other hand, are identifiable by their less powerful, typically puffy jets that are brightest near the centre of the black





An example FRI source An example FRII source

Figure 1.2: The morphologies of Fanaroff-Riley Type I and II AGN (FRI/FRII). The left hand panel, copyright of NRAO/AUI, shows an AGN with an FRI morphology (3C31) whilst the right hand panel, copyright of NRAO, shows an FRII morphology source (3C175).

hole. An example of an FRI type source is shown in the left hand panel of Figure 1.2.

In recent years however, other factors such as the relative accretion rate of an AGN have been found to be important (see Hardcastle et al., 2007; Heckman & Best, 2014) and may reflect differences in how matter is accreted onto the black hole. In this definition, AGN are typically split into two types: radiative- and jet-mode (for a review see Heckman & Best, 2014). Radiative-mode AGN are believed to be efficient at accreting material and have luminosities of $L \gtrsim 0.01 L_{\rm Edd}$, where $L_{\rm Edd}$ is the Eddington luminosity (Peacock, 1999; Mo et al., 2010), which is the luminosity where, for spherically symmetric accretion, the pressure from gravitational attraction of the material onto the black hole is balanced exactly by the radiation pressure being emitted from material accreting onto the black hole. It is given by:

$$L_{\rm Edd} = \frac{4\pi G M_{BH} m_p c}{\sigma_T}, \tag{1.18}$$

where G is the gravitational constant, m_p the proton mass, c, the speed of light in a vacuum and σ_T is the Thompson scattering cross section. M_{BH} is the mass of the accreting black hole.

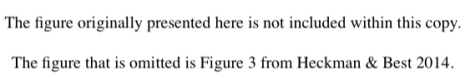


Figure 1.3: Figure 3 from Heckman & Best (2014) which shows a comparison between the two accretion mechanisms that are believed to exist for AGN. The left hand panel shows the accretion mechanism for Radiative-mode AGN, which accrete in the typical unified model approach (Antonucci, 1993; Urry & Padovani, 1995) and the right hand model shows the advection dominated accretion mode of Jet-mode AGN.

For radiative accretion mode, the AGN is consistent with the typical model for AGN (Antonucci, 1993; Urry & Padovani, 1995). This is shown in the left hand of Figure 3 from Heckman & Best (2014), which is shown here in Figure 1.3. On the other hand, the jet-mode AGN are believed to be inefficient, with advection dominated accretion of matter onto the black hole with an accretion rate of $L \lesssim 0.01 L_{\rm Edd}$ (Heckman & Best, 2014). Advection dominated accretion flows describe the slow accretion of matter within optically thin disks but where the cooling time of the gas is long, such that heat is not dissipated quickly and is instead transported towards the black hole (see e.g. Narayan et al., 1998; Longair, 2011; Heckman & Best, 2014). This accretion mode is illustrated in the right hand panel of Figure 1.3.

As they accrete matter via different mechanisms, this also means that there exist differences in their spectral properties (Best & Heckman, 2012; Whittam et al., 2016, 2018; Williams et al., 2018). Work investigating their number density evolution has also shown that the two populations evolve differently to one another (e.g. Best & Heckman, 2012; Pracy et al., 2016). These are therefore thought to be two differ-

ent classes of object, however in practice there may be a continuous distribution of accretion rates, not a clear cut dichotomy (see Whittam et al., 2018). An overview of different AGN types and how they relate to the AGN mode is given in Figure 4 of Heckman & Best (2014). Importantly, for the work presented in this Thesis, I note that radio loud jet mode AGN are known as Low Excitation Radio Galaxies (LERGs), whereas radiative mode radio loud AGN are known as High Excitation Radio Galaxies (HERGs).

It is thought that AGN have a large influence on the properties of the host galaxy. As illustrated in the review by Silk & Mamon (2012) (shown in Figure 1.4) and described in Benson et al. (2003), the observed luminosity distribution of galaxies is lower at high and low luminosities compared to what is predicted from converting the mass function predicted with Cold Dark Matter theory to a galaxy luminosity function. These are often attempted to be explained and compensated for in simulations (see e.g. Benson et al., 2003; Bower et al., 2012; Naab & Ostriker, 2017; Sijacki et al., 2007; Regan et al., 2018) by considering supernova feedback at low luminosities, whilst at high luminosities much higher energies are needed to limit the growth of galaxies. As such, AGN are thought to have large scale influence on galaxies limiting the size of their growth in a 'feedback' mechanism (see reviews in Fabian, 2012; Morganti, 2017) by heating the surroundings and stopping the formation of stars. This is because the energies needed are only thought to be possible with processes related to supermassive black holes.

1.4 Radio Emission from Galaxies

When considering radio continuum emission from galaxies, the observed flux comes from a combination of processes (these are discussed in e.g. Carroll & Ostlie, 2006; Longair, 2011; Burke & Graham-Smith, 2014). The emission from the galaxy M82,

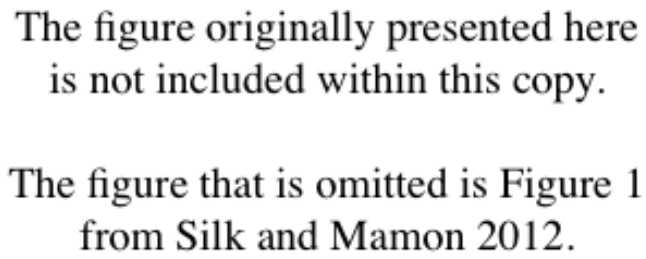


Figure 1.4: Figure 1 from Silk & Mamon (2012) indicating how the observed galaxy luminosity function differs to what is predicted by Cold Dark Matter (CDM) cosmology. This difference is typically explained by supernova feedback at low luminosities and AGN feedback at high luminosities.

presented in Condon (1992) and shown here in Figure 1.5, has three contributions to the observed spectrum. At the highest frequencies (moving into the IR) emission from thermal dust (dotted line in Figure 1.5) has the largest contribution. At slightly lower frequencies, the spectrum is dominated by free-free emission (dashed line in Figure 1.5), where radiation is emitted when free electrons interact with ions and are caused to accelerate and change direction. Finally, at the lowest radio frequencies, synchrotron emission dominates (dot-dashed line in Figure 1.5). For the radio surveys used in this Thesis, synchrotron emission is the dominant mechanism by which radio emission is generated, however there may be contributions from free-free emission when high frequency radio surveys are used.

Synchrotron emission is generated when electrons spiral in magnetic fields. Assuming a power law distribution for the initial energy of an electron population, the synchrotron spectrum generated behaves as a power law distribution, where the flux density of the emission is given by S_ν and the frequency being observed as ν:

$$S_\nu \propto \nu^{-\alpha} \tag{1.19}$$

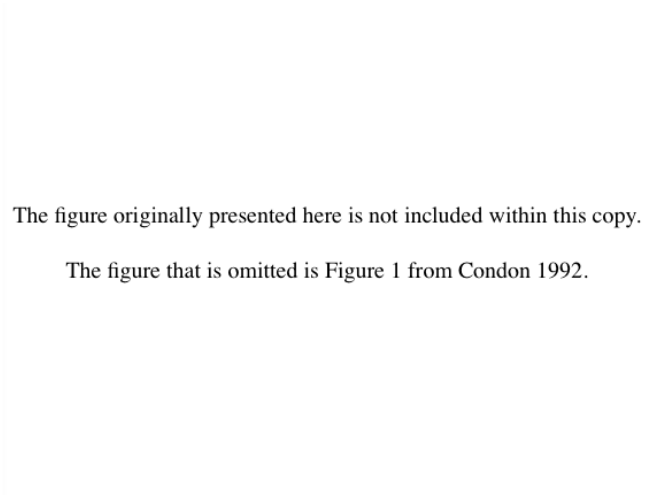


Figure 1.5: Figure 1 from Condon (1992) showing the spectrum of M82 from the radio to the infrared. Individual contributions to the spectrum from synchrotron (dot-dashed), free-free (dashed) and thermal dust (dotted) emission is shown as well as the total contribution (solid).

where α is known as the spectral index of the synchrotron power law distribution and is typically found to have an average value of $\sim 0.7 - 0.8$ (e.g. Condon, 1992; Intema et al., 2017; Smolčić et al., 2017a). For the remainder of this Thesis, unless otherwise stated, a value of $\alpha = 0.7$ is adopted. As can be inferred from Equation 1.19, a redshift in the frequency will not alter the shape of the underlying power law distribution. This means a redshift cannot be inferred from radio continuum observations. As such, redshifts are only available for radio sources through cross matching to multi-wavelength counterparts (e.g. McAlpine et al., 2012; Banfield et al., 2015; Williams et al., 2019).

1.4.1 Galaxy Populations Observed in Radio Surveys

In the synchrotron regime, two main populations of sources are observed and compared, these are AGN and Star Forming Galaxies (SFGs) (e.g. Mauch & Sadler, 2007; Prescott et al., 2016; Smolčić et al., 2017b). For AGN, the source of the magnetic fields necessary for synchrotron emission is thought to be related to the high energy jets,

whilst for SFGs it is the magnetic field in supernova remnants that are responsible for the synchrotron emission. For AGN, these can be further classified into different sources as discussed in Section 1.3. SFGs and AGN are therefore modelled separately in extragalactic simulations (Wilman et al., 2008; Bonaldi et al., 2019). For example, Wilman et al. (2008) produced semi-empirical simulations of the radio skies at five radio frequencies, with a source distribution as a function of flux density at 1.4 GHz reproduced here in Figure 1.6, which can be seen for the different populations[1]. These simulated extragalactic skies will be relied upon throughout this Thesis.

The AGN population in Wilman et al. (2008) arises from the combination of Fanaroff Riley Type I and Type II sources (FRI and FRII) (Fanaroff & Riley, 1974) and Radio Quiet Quasars (RQQ). For the SFGs, this is a combination of the normal SFGs (denoted by SF in Figure 1.6) and the bright starburst (SB) galaxies. These classifications highlight the fact that AGN are typically the dominant sources at higher fluxes ($S_{1.4\ \mathrm{GHz}} \gtrsim 0.1$ mJy) whereas SFGs dominate at fainter regimes. Observing large populations of SFGs therefore relies on having sufficiently deep observations.

1.4.2 Use of Radio Emission

Despite the lack of redshifts available from continuum observations and the need for multi-wavelength surveys to obtain these, there are many reasons why radio surveys are advantageous to study galaxy populations. The first of these, which has already been touched upon, is the ability to observe both AGN and SFG populations. These allow investigations into both the evolution of AGN and SFGs (e.g. Best & Heckman, 2012; Prescott et al., 2016; Pracy et al., 2016; Smolčić et al., 2017; Novak et al., 2017). For AGN, jets that may not be observable at other wavelengths are often bright and visible at radio frequencies. Furthermore, radio emission is important as it is not attenuated by dust. This is because radio wavelengths are much larger than the size

[1]In Figure 1.6 the source counts are Euclidean normalised, meaning they have a $S^{2.5}$ factor when presented, see the note in Heywood et al. (2013).

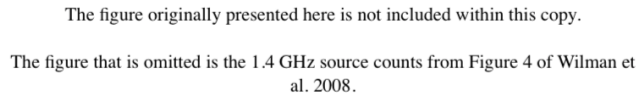


Figure 1.6: Source counts from Figure 4 of Wilman et al. (2008) for the different extragalactic source populations in the simulated skies at 1.4 GHz.

of dust grains and so the radio photons are unabsorbed by dust as they travel to the observer (Longair, 2011; Burke & Graham-Smith, 2014). Work has also shown that radio observations of SFGs are known to be correlated to the far-IR (see e.g. Bell, 2003; Appleton et al., 2004; Jarvis et al., 2010; Read et al., 2018) suggesting they are both related to stellar processes, with the far-IR tracing the reprocessed dust emission resulting from star formation, whilst radio emission traces supernova remnants which are correlated with star formation. This suggests that radio emission from SFGs can be used for unbiased Star Formation Rate (SFR) estimates (see e.g. Bell, 2003; Garn et al., 2009; Davies et al., 2017; Gürkan et al., 2018).

1.4.3 Radio Luminosities

Although flux densities of radio sources are often discussed (de Zotti et al., 2010), it is more useful to compare the luminosities of sources (Mauch & Sadler, 2007; Best et al., 2014; Pracy et al., 2016; Smolčić et al., 2017). I therefore re-derive the

equation for calculating a spectral luminosity accounting for k-corrections (see e.g. Hogg, 1999; Novak et al., 2017). This means taking into account the fact that a flux density observed at the earth at one frequency corresponds to a different rest frame frequency. For the relevant frequencies involved, I define ν_{em} as the frequency which the radiation is emitted at in the rest frame, ν_{obs} as the frequency of this photon when it reaches the observer (see Equation 1.2) and $\nu_{obs'}$ as the frequency at which observations are taken. In terms of cosmological distances, we can relate the spectral luminosity emitted ($L_{\nu_{em}}$) to the observed flux (that the photon is redshifted to), taking into account the luminosity distance D_L:

$$L_{\nu_{em}} = \frac{S_{\nu_{obs}} 4\pi {D_L}^2}{1+z}, \tag{1.20}$$

see e.g. Hogg (1999). In this equation, the $(1+z)$ factor arises due to bandwidth stretching, this is because the frequency bandwidth over which the radiation was emitted will be stretched with the expansion of the Universe. Using Equation 1.19, the flux observed at the emitted frequency is given by:

$$L_{\nu_{em}} = S_{\nu_{em}} \left(\frac{\nu_{em}}{\nu_{obs}}\right)^{\alpha} \frac{4\pi {D_L}^2}{1+z}. \tag{1.21}$$

Using the definition of redshift, Equation 1.2, this leads to:

$$L_{\nu_{em}} = S_{\nu_{em}} (1+z)^{\alpha} \frac{4\pi {D_L}^2}{1+z}. \tag{1.22}$$

If the observations are at a different frequency to ν_{em} then Equation 1.19 can be used to relate $S_{\nu_{em}}$ to $S_{\nu_{obs'}}$. Equation 1.22 then becomes:

$$L_{\nu_{em}} = \left(\frac{\nu_{obs'}}{\nu_{em}}\right)^{\alpha} S_{\nu_{obs'}} (1+z)^{\alpha-1} 4\pi {D_L}^2. \tag{1.23}$$

1.5 Introduction to Radio Observations

As radio continuum observations are being used for the work in this Thesis, a brief overview of the techniques necessary to produce these images is described here. These concepts are more thoroughly explained in, for example, books such as Taylor et al. (1999); Wilson et al. (2009) and online tutorials such as `https://casaguides.nrao.edu/index.php/Karl_G._Jansky_VLA_Tutorials`. Unlike optical and near-IR surveys that make use of single dishes or detectors for their observations, radio continuum observations often rely on interferometry between receivers in order to construct images of the sky. The advantage of combining the data received at multiple telescope antennas is the improved resolution that can be obtained compared to single dish detectors. However, with the use of interferometry, this will inherently limit the quality of images due to incompleteness in the observations of some spatial scales, due to the distribution of the antennas. This will be discussed in Section 1.5.2.

1.5.1 Single Dish Observations

For single dish radio observations, all spatial scales are observed with the dish, meaning that no observed intensity is lost from the source. However, as radio wavelengths are long, this means that without a large aperture dish the resolution of observations is limited to comparatively poor resolution. This resolution is defined by both the wavelength (λ; or frequency, ν) of the observations and the diameter of the dish (D):

$$\theta_{res} \sim \frac{\lambda}{D} \sim \frac{c}{\nu D} \tag{1.24}$$

In the optical, λ is small and so dishes can be O($\sim$m) to get high resolution (sub-arcsecond) images. Contrary to this, in the radio a typical observing frequency of 1.4 GHz would require a dish of diameter O($\sim$km) in order to obtain 10" resolution. For the largest single dish radio telescope, FAST (Nan et al., 2011), its 500m diameter

corresponds to ∼1.5' resolution at a frequency of 1.4 GHz. This is therefore limiting for deep extragalactic continuum surveys, which would reach confusion for relatively shallow observations. The confusion limit of a survey defines the depth at which, due to a combination of the resolution and depth of the observations and the source count distribution of extragalactic sources, multiple sources will be observed within the same beam area. The background emission will be elevated due to the combination of these faint background sources (see e.g. Wilson et al., 2009; Condon et al., 2012; Zwart et al., 2015a).

1.5.2 Interferometers

With interferometry, on the other hand, the combination of antennas allows the resolution of observations to be improved (see e.g. Wilson et al., 2009; Burke & Graham-Smith, 2014) as the important quantity is not the dish diameter, D, but the maximum baseline (B) between antennas in the array:

$$\theta_{res} \sim \frac{\lambda}{B} \sim \frac{c}{\nu B}. \tag{1.25}$$

For the VLA[2], in its most extended configuration, there is a maximum baseline of ∼36 km. This corresponds to ∼arcsecond resolution at 1.4 GHz, which allows high resolution observations of extragalactic sources and reduces the flux limit at which confusion will be important.

Whilst this increase in resolution highlights the advantages of using interferometry, the spatial information of the source can be limited. As data between pairs of antennas is recorded (the visibilities), only information on the spatial scales corresponding to the distances between pairs of antennas can be recorded. This means that spatial information on the largest angular scales is lost from not having dishes

[2] `https://science.nrao.edu/facilities/vla`

Figure 1.7: Figure 4 from Heywood et al. (2016) which indicates how information on various spatial scales can be lost due to the configuration of antennas in the interferometer. This is shown for A-configuration (panel 1; Hodge et al., 2011), B-configuration (panel 2; Becker et al., 1995), D-configuration (panel 3; Condon et al., 1998) and C&B configuration (panel 4; Heywood et al., 2016).

close enough together, whilst a lack of long baselines limits the resolution of images and observations of the smallest angular scales. This is highlighted in Figure 1.7, taken from Heywood et al. (2016). In this image the most extended configuration of the VLA (A-configuration, first panel; Hodge et al., 2011) is used to obtain high resolution imaging, but lacks an ability to observe extended, large scale emission. With the most compact configuration (D, third panel; Condon et al., 1998), on the other hand, the large scale extended emission is clearly detected, but the image lacks the high resolution shown in the first panel.

1.5.3 *uv*-coverage

As discrete measurements are made using an interferometer, related to the interferometer configuration, the visibilities that are recorded are related to the Fourier transform of the sky brightness distribution sampled at discrete points, related to the sampling function. This sampling function is often represented as a *uv*-coverage plot. Here u and v are often given in values of kλ and correspond to projected spatial differences between antennas. With a snapshot image of a field, this would only

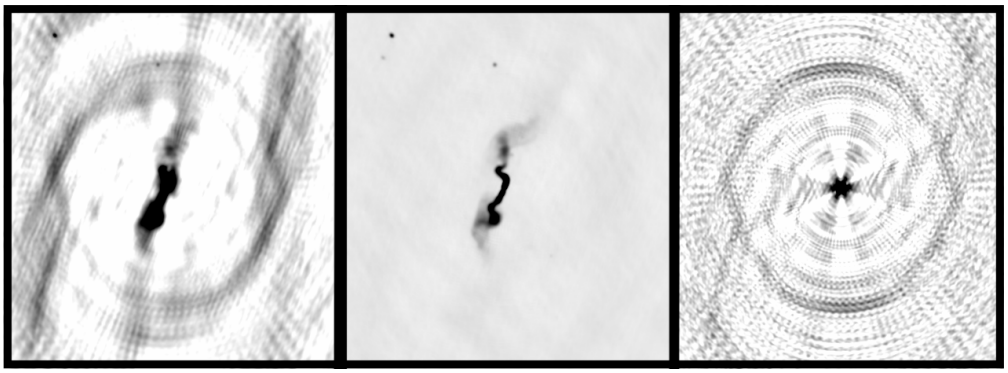


Figure 1.8: Image highlighting the difference between a dirty and clean image for 3C31. This is from public VLA data that I reduced after a tutorial by I. Heywood (Oxford), and was first published in Laing & Bridle (2002). The left hand panel shows the dirty image, whilst the central panel shows the cleaned image. The right hand panel shows the dirty beam/PSF of the observations which arise from the Fourier transform of the *uv*-coverage. These are all set to a flux density per beam scale of -5×10^{-3} to 5×10^{-2} Jy/beam.

have discrete points on the *uv*-plane, corresponding to the instantaneous baselines produced by the array configuration. This would be a sparsely covered *uv*-plane however, by observing for longer periods, the rotation of the Earth allows more of the *uv*-plane to be filled in as the change in projected baselines fill out the *uv*-plane with elliptical shapes (see e.g. Thompson, 1999; Wilson et al., 2009; Burke & Graham-Smith, 2014). The Fourier Transform (FT) of this *uv*-coverage is known as the dirty beam and is a representation of the point spread function in the dirty image. The corresponding initial image (dirty image), which is the FT of the visibilities, will therefore have artefacts surrounding sources due to this incompleteness (see Figure 1.8). Image quality is improved by either better *uv* sampling, or through cleaning the dirty image.

1.5.4 Cleaning the Dirty Radio Image

In order to truly understand what the underlying sky distribution of sources in a target observation actually are, it is important to go from the dirty image to the

calibrated, cleaned image of the field (see Figure 1.8). To obtain the cleaned image the data is calibrated to account for instrumental gains as well as trying to remove the artefacts in the image arising from the dirty beam. The method to do this is discussed in many texts, including Taylor et al. (1999); Wilson et al. (2009); Burke & Graham-Smith (2014) and I give an overview in this section. For radio observations taken with an interferometer, the measurements are well modelled via the Radio Interferometry Measurement Equation (RIME, see e.g. Hamaker et al., 1996; Smirnov, 2011). A simple RIME describing the calibration at the phase centre is:

$$V_{pq} = G_p X_{pq} G_q{}^*. \tag{1.26}$$

where V_{pq} are the observed visibilities measured for antennas p and q and the G terms define the gains that arise from factors such as the instrumental gains as well as the delays that are needed to ensure coherence along the wavefront that is being measured at the different antennas. X_{pq} is the quantity to be predicted as it describes the true sky distribution.

In order to go from the dirty image to the best estimate for the true sky distribution, several steps need to be taken. Firstly, the data must be flagged in order to remove bad data that may arise from radio frequency interference (RFI) or from specific antennas that may have had issues during the observations. As astronomical emission should typically be smooth (excluding spectral lines) in time and frequency, large spikes in the amplitudes across the bandwidth can be used to identify such regions. Once this is done, a primary calibrator is used to calculate the response of the antennas and any gains within the instrument. This primary calibrator is a source that is bright, with a known source model. This not only allows conversion between measured voltages to the required flux scale but also can be used to correct for any frequency response of the detectors (known as bandpass calibration). Secondary calibrator sources within the field (or near the field) can also be used to track drifts

in the gains across the observations. Once any gains from the calibrator have been determined and applied to the target data, the image is CLEAN-ed.

CLEAN was first developed by Högbom (1974) and is an iterative process in which the effects of the dirty beam can be reduced within the image. This is through building up a sky model of point source components, which are convolved with the dirty beam and subsequently removed from the image. By beginning with the brightest source and removing point source components at the source location but with a peak flux $\sim$10% of the observed peak flux, these components can be iteratively removed (and this can be continued for other bright sources) until some threshold criterion is met. The CLEAN-ed version of the image is then the combination of these point source components convolved with the clean beam (a smooth model of the dirty beam, typically a 2D Gaussian) added back into the residual image. After this process, further improvement of the image can be made through self calibration (see e.g. Pearson & Readhead, 1984; Taylor et al., 1999; Burke & Graham-Smith, 2014). Self-calibration involves using a sky model to try and further account for any amplitude and phase errors using cycles of calibration for phase only and then amplitude and phase gains. By using the sky model (which is thought to represent the true emission) to minimise differences between the data and the sky model, this process helps to reduce artefacts that are typically observed around bright sources. Finally, for observations with large fields of view, it may also be important to calibrate for direction dependent effects, which can arise from the ionosphere (at low frequencies) as well as changing beam patterns (see e.g. van Weeren et al., 2016; Tasse et al., 2018; Shimwell et al., 2019). This often involves splitting up the field of view into different directions and doing further rounds of calibration to reduce amplitude and phase errors. With a combination of all these processes, this should minimise the amount of artefacts that are observed across the image.

It is with these cleaned and high fidelity images that it is possible to produce the science both described within this introduction as well as the science that I have performed that is be described within this Thesis.

1.6 Radio Continuum Surveys

Due to the advantages of using radio continuum images to study galaxies, many large surveys of the extragalactic sky at radio frequencies have been performed since the 1950s. In this Chapter, I shall give a brief overview of some of the radio surveys that paved the way in defining extragalactic catalogues of the radio sky as well as those which have been made recently, or are soon to be available. This will include precursor telescopes to the Square Kilometre Array (SKA). Many of these surveys are discussed within the review by Norris (2017).

Some of the most well-known radio surveys are the Cambridge Surveys starting with the first Cambridge Survey (Shakeshaft, 1957). Commencing in the 1950s, the Cambridge Surveys focussed on making detailed observations of the brightest radio sources using telescopes in Cambridge. One of the most notable and well known of these surveys is the 3C Survey (Edge et al., 1959, at 159 MHz). Since the 3C Survey, these observations have not only been revised (e.g. 3CR; Bennett, 1962) but other Cambridge Surveys have succeeded this at different frequencies, declinations and sensitivities: 4C (at 178 MHz Pilkington & Scott, 1965; Gower et al., 1967), 5C (at 408 and 1407 MHz; Pearson, 1975), 6C (at 151MHz, this was released in six papers, including Baldwin et al., 1985; Hales et al., 1993), 7C (at 151 MHz, in several papers including McGilchrist et al., 1990; Riley et al., 1999), 8C (at 38 MHz Rees, 1990), 9C (Waldram et al., 2003, at 15 GHz;) and 10C (at 15.7 GHz; AMI Consortium et al., 2011).

More recently, the Very Large Array (VLA; `https://science.nrao.edu/facilities/`

vla) has been instrumental in improving observations of extragalactic radio surveys. Based in New Mexico, United States, the VLA consists of 28 radio dishes on moveable tracks that can be placed in 4 different configurations (A, B, C and D) and can observe in 10 frequency bands from 74 MHz to 43 GHz. As shown in Figure 1.7, A-configuration is the most extended, which allows the highest resolution observations, whilst D-configuration is the most compact and sensitive to larger angular features. This array has been used to produce some of the most well used large area radio surveys. The Faint Images of the Radio Sky at Twenty centimetres Survey (FIRST; Becker et al., 1995; Helfand et al., 2015) is one of the most well known of the surveys produced with the VLA. Its large area coverage (10,000 deg^2), high resolution (5"; in B configuration) and deep (0.15 mJy/beam rms) observations at 1.4 GHz have led the way for a large number of scientific papers and research. This includes investigating the clustering of radio sources (Cress et al., 1996; Overzier et al., 2003); looking at large population statistics such as luminosity functions and their evolution (e.g. Best & Heckman, 2012; Best et al., 2014; Pracy et al., 2016). The NRAO VLA Sky Survey (NVSS; Condon et al., 1998) is another large survey observed with the VLA. This also observed at 1.4 GHz but with a 45" beam, however it was more sensitive to extended emission than FIRST. NVSS observed to a flux limit of ~ 0.45 mJy/beam rms but observed down to Declinations of -40°, compared to -10° for FIRST. Finally, at low frequencies, the VLA Low-Frequency Sky Survey (VLSS; Cohen et al., 2007) conducted observations above a Declination of -30° at 74 MHz. This was at 80" resolution with an rms of 0.1 Jy/beam.

Complementary to these, the Sydney University Molonglo Sky Survey (SUMSS; Bock et al., 1999), is an 843 MHz survey using the Molonglo radio facility to produce 43" resolution images to 1 mJy/beam. This helped improve sky models of the southern skies, not possible with the northern hemisphere telescopes.

1.6.1 Current Radio Continuum Surveys

With current radio surveys, that are not acting as precursors or pathfinders to the SKA, there are a large range of extragalactic surveys. These are important in both observing over wider areas as well as targeted observations to produce highly sensitive observations. These are producing high fidelity images of thousands of radio galaxies.

In terms of large area surveys, the VLA Sky Survey (VLASS, `https://science.nrao.edu/science/surveys/vlass`; Myers & VLASS Survey Team, 2018) is performing observations at 2.5" resolution over the northern sky eventually to ∼70 μJy/beam rms at ∼3 GHz. At lower frequencies, another very notable survey is the TIFR GMRT Sky Survey alternative data release (TGSS-ADR; Intema et al., 2017). This was conducted using the Giant Metre Radio Telescope (GMRT; Swarup et al., 1991, `http://www.gmrt.ncra.tifr.res.in/`), in India, and has recently produced large sky observations of the northern sky down to -53° at 150 MHz. It reaches a flux limit of ∼5 mJy/beam rms at ∼25" resolution. This was a revolutionary survey at these frequencies, producing high resolution imaging, as well as providing a sky model which was essential for the LOFAR surveys (see Section 1.6.1.1 and Chapter 5).

Instead, if smaller area observations are desired, many of the most sensitive observations at radio frequency have been conducted with the VLA. This includes 3 GHz observations over the 2 deg^2 of the COSMOS field (VLA-COSMOS; Smolčić et al., 2017a) at 0.75" resolution to ∼2.3 μJy/beam rms. These data are crucial for the work of Chapters 2 and 3. Current observations are also underway with the VLA to produce the deepest image of the radio sky at 1.4 GHz by the CHILES Con-Pol Survey (`http://chilesconpol.com`). This will observe over of 0.2 deg^2 of the COSMOS field, eventually reaching ∼400 nJy depth at 4" resolution, making it one of the deepest radio surveys. However, it is limited compared to other surveys by its small area and confusion. Finally, in the UK, the enhanced Multi Element Remotely

Linked Interferometry Network (e-MERLIN; Thomasson, 1986; Garrington et al., 2004, http://www.e-merlin.ac.uk/) is one facility aiming to produce high quality imaging of AGN and SFGs to high redshift, over small areas. The eMERGE Survey (http://www.e-merlin.ac.uk/legacy/projects/emerge.html) aims to survey at both 1.4 and 5.5 GHz to incredibly high (sub arcsecond) resolution to sub-μJy depth, which will be one of the highest angular resolution surveys (but see Section 1.6.1.1).

1.6.1.1 SKA Pathfinder/Precursor Surveys

At low frequencies, there are two radio facilities that are important in preparing for the SKA. Both these telescopes consist of dipole antennas and are able to observe large areas of the sky in individual pointings due to the large field of view of radio observations at these long wavelengths. In Europe, the LOw Frequency ARray (LOFAR, http://www.lofar.org/; van Haarlem et al., 2013) consists of ∼40 stations of dipole interferometers across the Netherlands and Europe with baselines of O(∼km) to O(∼100 km) and observes at two frequency ranges. The low band antenna (LBA) operates at 30-80 MHz and the high band antenna (HBA) operates at 120-220 MHz. LOFAR's HBA has recently been shown to be able to image large areas of the sky to high sensitivity (∼70-100 μJy/beam rms) and resolution (∼6") at ∼150 MHz over specific fields (Williams et al., 2016) and over larger areas (Hardcastle et al., 2016; Shimwell et al., 2017, 2019). This telescope is used in the work of Chapter 5. In the future, the deeper tiered LOFAR surveys will become available (Röttgering, 2010; van Haarlem et al., 2013) and the international stations of LOFAR will be combined with the Dutch array and will produce sub-arcsecond resolution of some of the bright sources observed in the LOFAR Two-metre Sky Survey (LoTSS; Shimwell et al., 2019). In Western Australia, on the other hand, the low frequency SKA precursor is the Murchison Widefield Array (MWA, http://www.mwatelescope.org/; Tingay et al., 2013). This has recently released a large, southern sky survey: the GaLactic

and Extra-galactic All-sky MWA Survey GLEAM (GLEAM; Hurley-Walker et al., 2017). This differs to LOFAR in its poorer resolution ($\sim$2') and sensitivity, with fluxes for the faintest sources of O($\sim$10 mJy). However, its short observations has provided observations of thousands of sources across the whole southern sky with multi-frequency measurements from $\sim$70-230 MHz. This has provided spectral index information for these sources across the bandwidth of the receivers.

At higher frequencies ($\sim$GHz) the SKA precursor is located in South Africa. The Meer Karoo Array Telescope (MeerKAT; Jonas, 2009; Jonas & MeerKAT Team, 2016), consists of 64 dishes on the site which will be developed into the SKA. The radio continuum survey using MeerKAT, the MeerKAT International Giga Hertz Tiered Extragalactic Exploration (MIGHTEE Jarvis et al., 2017) survey will observe a total of $\sim$20 deg^2 over four fields that have a wealth of ancillary information (XMM-LSS, COSMOS, ELAIS-S1 and E-CDFS) to μJy depth. Observing over these fields ensures the necessary ancillary information to constrain redshifts, stellar masses and star formation rates for the host sources. In Australia, the Australian SKA Pathfinder (ASKAP; `https://www.atnf.csiro.au/projects/askap/index.html`) instead consists of 36 dishes in the Western Australian Desert. Its radio continuum survey, the Evolutionary Map of the Universe (EMU Norris et al., 2011), will make observations of the entire southern sky to $\sim$10μJy, providing complementary high frequency observations to GLEAM (Hurley-Walker et al., 2017) and improving upon southern sky models.

1.6.2 Future Radio Surveys with the SKA

In the future, radio surveys will be revolutionised by the Square Kilometre Array (SKA; `https://www.skatelescope.org/`). The SKA plans to incorporate both high frequency dishes across South Africa and low frequency arrays across Western Australia over baselines of 100s of kilometres, as well as in dense core regions. These combinations of baselines will allow the telescope to both produce high resolution

observations as well as have sensitivity to extended emission. The continuum surveys have many aims, described in Prandoni & Seymour (2015), which includes studying star formation and AGN activity across the age of the universe and determining their importance towards advancing and influencing the evolution of galaxies.

In this Thesis, I discuss the evolution in the clustering of different radio populations in Chapter 2, making use of the VLA 3 GHz COSMOS Survey (Smolčić et al., 2017a). I then build upon this work in Chapter 3 by combining the clustering of radio sources with their luminosity functions in order to constrain the Conditional Luminosity Function. In Chapter 4, I show how the multi-frequency source extraction package ProFound can accurately obtain catalogues of sources in radio continuum images, before finally presenting observations of the XMM-LSS field with LOFAR (van Haarlem et al., 2013) in Chapter 5. I will finally draw conclusions and discuss future work in Chapter 6.

Chapter 2

Evolution in the Clustering of SFGs and AGN

As described in Chapter 1, one way in which it is possible to study the dark matter environments that galaxies reside in is through measuring their clustering. This is due to relationships between bias and halo mass (Mo & White, 1996; Sheth & Tormen, 1999; Tinker et al., 2010). With deep radio surveys, where both AGN and SFGs can be observed due to their synchrotron emission (see Section 1.4), the haloes in which these populations inhabit can be compared. This allows the influence of environment on the properties of galaxies to be studied. Whilst this is beginning to be investigated using radio continuum observations (Magliocchetti et al., 2017), the differences in the redshift distributions of SFGs and AGN means it is often challenging to directly compare their bias (as bias is known to evolve with z). It is therefore essential to use more sensitive observations where large populations of these faint SFGs can be observed. This makes the investigation of the evolution of bias (with redshift) easier, because the clustering of SFGs and AGN can be compared using similar redshift ranges.

The VLA 3 GHz COSMOS Survey (Smolčić et al., 2017a) is one such survey that

has observed large populations of both SFGs and AGN. Covering the $\sim$2 deg^2 of the COSMOS field, this survey both has a significant area to investigate the large scale clustering (which is essential for studying bias) and is also incredibly sensitive ($\sim$ 2.3 μJy/beam rms). Importantly though, the wealth of excellent ancillary data across the field allowed the classification of the observed radio sources into AGN and SFGs (Smolčić et al., 2017b). These factors combined means that it is one of the best radio catalogues available to study the clustering of these two radio populations.

Furthermore, the work by Smolčić et al. (2017b) also classified the observed AGN based on their accretion efficiency, allowing a large number of H/LERG analogues to be determined. Whilst in the past, classification of AGN by source morphology has commonly been used (see Section 1.3), the H/LERG separation is thought to be physically motivated, with the separation based on how the AGN accretes. These two populations are therefore thought to reside within different environments (Hardcastle, 2004; Gendre et al., 2013; Heckman & Best, 2014). It is possible that these differences are influenced by (or related to) the dark matter haloes they reside in. In addition, the H/LERG separation is useful for modern, sensitive, large area radio continuum surveys where large numbers of radio sources are observed over areas with multi-wavelength coverage. This is because accurate morphological classifications rely on being able to observe jets and then rely on visual inspection (Banfield et al., 2015; Williams et al., 2019), which is time-consuming for such large numbers of sources, or classification algorithms (Ma et al., 2018; Wu et al., 2019). Whilst studying the clustering for FRI/II sources is important, as the two populations impact their environments differently, to investigate the relationship between environment and how the AGN are fuelled, studying the clustering of H/LERGs is more relevant. Therefore where multi-wavelength data is available, it is physically motivated to classify radio AGN based on their accretion mode and compare the clustering of these two populations. Such a comparison of the clustering for H/LERGs using radio surveys has not

been previously carried out.

In this Chapter I present work from Hale et al. (2018) where I investigated the clustering of different galaxy populations within the VLA 3 GHz COSMOS Survey (Smolčić et al., 2017a). In this work, I not only investigate how the clustering of SFGs and AGN compare and evolve with redshift but also, for the first time in the radio, investigated how the clustering of AGN may be related to accretion efficiency.

2.1 Data

The data used for the investigations in this Chapter are from the VLA 3 GHz COSMOS Survey (Smolčić et al., 2017a). This provided a catalogue of 10,830 radio AGN and SFGs (above 5σ) over the $\sim$2 deg^2 of the COSMOS field. The observations were made over a total of 384 hours at S-band (2-4 GHz) using the VLA and have an angular resolution of 0.75". The long duration of observations as well as the tiling strategy mean that these observations are incredibly sensitive, reaching an average rms of $\sim$2.3 μJy/beam. Whilst 10,830 sources above 5σ were observed over the field, for this work I made use of the 5.5σ catalogue that was also provided. This is because, as noted by Smolčić et al. (2017a), false detections of sources in the catalogue at the level of $\sim$3% could be reduced to $\sim$0.4% using the 5.5σ catalogue. Although going from 5σ to 5.5σ is only a small ($\sim$2.5%) reduction in the percentage of false detections, I wanted to ensure a clean catalogue of sources was used. By using the 5.5σ catalogue, $8,928$ sources were available over the observed $\sim$2 deg^2. This corresponded to $\sim$ 86% of the 5σ catalogue. This was still a large number of sources, sufficient to quantify the angular clustering whilst ensuring clean samples of the different populations were compared.

Whilst these radio observations are incredibly sensitive compared to other surveys, in order to evaluate the bias, knowledge of redshifts are essential (see Section 1.2).

Fortunately, by design, these observations were chosen to overlap with the well studied COSMOS field. This has a wealth of multi-wavelength observations (e.g. Hasinger et al., 2007; Schinnerer et al., 2010; Oliver et al., 2012; McCracken et al., 2012). Multi-wavelength observations were used, in a complementary paper (Smolčić et al., 2017b) to the main survey paper, in order to provide redshifts for these sources, as well as determine their source type. As described in Section 1.4, the observed radio sources are a combination of SFGs and AGN. Using a variety of methods, ∼86% of the 5.5σ radio sources had a redshift identified. These were subsequently classified as a SFG or an AGN, whilst further classifying AGN by their accretion efficiency.

From their definition, radiatively efficient AGN were described as High to moderate Luminosity AGN (HLAGN), whilst radiatively inefficient were classified as Moderate to low Luminosity AGN (MLAGN). Smolčić et al. (2017b) considered these to be equivalent to high redshift HERGs and LERGs (described in Section 1.3). The methods used to classify these different populations were through a variety of techniques. These are described thoroughly in Smolčić et al. (2017b), but are summarised here:

- *X-Ray Luminosities:* As AGN are known to be bright in the X-ray (Brandt & Alexander, 2015), the X-ray luminosities of these sources (for those with counterparts) was one method which was used to determine whether an AGN was present. Smolčić et al. (2017b) defined X-ray AGN as those with $L_X > 10^{42}$ erg s^{-1} as had been used in previous work of Szokoly et al. (2004). Any AGN that satisfied this criteria were classified as a HLAGN.

- *SED fitting:* Spectral Energy Distribution (SED) fitting was also performed to see if there is evidence of an AGN being present. This was through modelling the spectra with and without an AGN component as in Delvecchio et al. (2017). The presence of an AGN was decided upon if the reduced χ^2 using the AGN model was smaller than when an AGN component was not included.

- *mid IR-colours:* The ratios between colours in the mid-IR were used in order to identify whether there was a torus surrounding an AGN present. This method has been previously used in Donley et al. (2012) and uses observations in the range of 3.6 μm to 8 μm. It is based on the ratio between the fluxes at 3.6 μm to 5.8 μm and the ratio between fluxes at 4.5 μm and 8.0 μm.

- *Rest Frame Colours:* The dust corrected near-UV to r-band colours were used to identify those galaxies where star formation dominated the emission. This followed the work of Ilbert et al. (2010) and exploited the fact that UV emission is present when star formation occurs.

- *Star Formation Rates vs. Radio Luminosities:* Finally, a ratio of radio luminosity to SFR measured from the IR was also used to identify quiescent and star forming galaxies. This is because the 1.4 GHz luminosities of SFGs are known to be correlated with SFRs (e.g. Bell, 2003; Garn et al., 2009; Davies et al., 2017). Those which were dominated by IR star formation rates were classified as SFGs, whilst those where the radio emission dominates are classified as quiescent. The cut off luminosity used was allowed to evolve with redshift and given as: $\log_{10}(L_{1.4\mathrm{GHz}}/\mathrm{WHz}^{-1}) = 21.984 \times (1+z)^{0.013}$.

However, Smolčić et al. (2017b) noted that some sources satisfied criteria suggesting that they were both MLAGN and SFGs. In order to cleanly investigate the clustering of these populations, I therefore made further cuts on the source populations I considered. The populations that I use for the remainder of this Chapter are given by:

- *AGN:* This is defined as the combination of the full HLAGN and MLAGN populations.

- *SFGs:* These are defined as 'clean' SFGs, which are those SFGs for which there

was no possible contamination by the MLAGN population (see Smolčić et al., 2017b).

- *HLAGN:* This is the full HLAGN population.
- *MLAGN*: For the purpose of this Chapter these are the 'clean' MLAGN sample where there was no possible contamination from SFGs. This cut, however, was not made when considering the full AGN population. This is because it is likely that given they satisfy both criteria that they have both star formation and AGN activity within the source. However, to ensure a clean sample when directly comparing clustering based on accretion mode this cut was imposed.

I also removed the sources at the very edge of the field (where the noise was higher), excluding all sources that were larger than 1 deg from the centre of the radio image. This resulted in a final sample of 8,887 unclassified radio sources. For the sources categorised by type, the number of sources available to calculate the clustering are further reduced. This is because multi-wavelength counterparts are not available in all regions of the radio images due to, for example, masking around bright stars. Hence a mask from Laigle et al. (2016) was used[1], the impact of which can be seen in Figure 2.1. For the sources with a multi-wavelength classification, a final catalogue of 3,704 (clean) SFGs and 2,937 AGN was used for this work. These AGN were further split into 1456 HLAGN (radiatively efficient) and 726 (clean) MLAGN (radiatively inefficient). The luminosity-redshift distributions of these four populations can be seen in Figure 2.2.

[1] using the package `pyregion` (`https://pyregion.readthedocs.io/en/latest/index.html`) to apply the mask

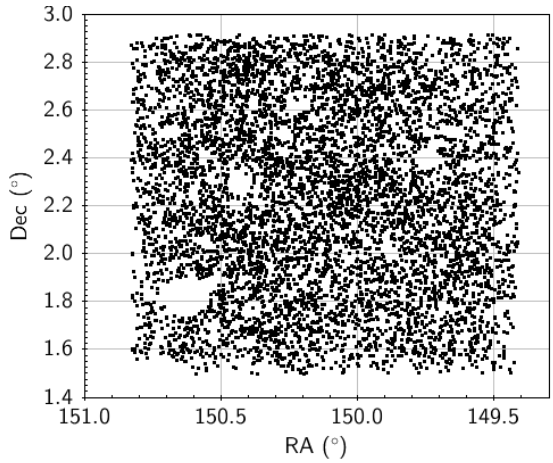
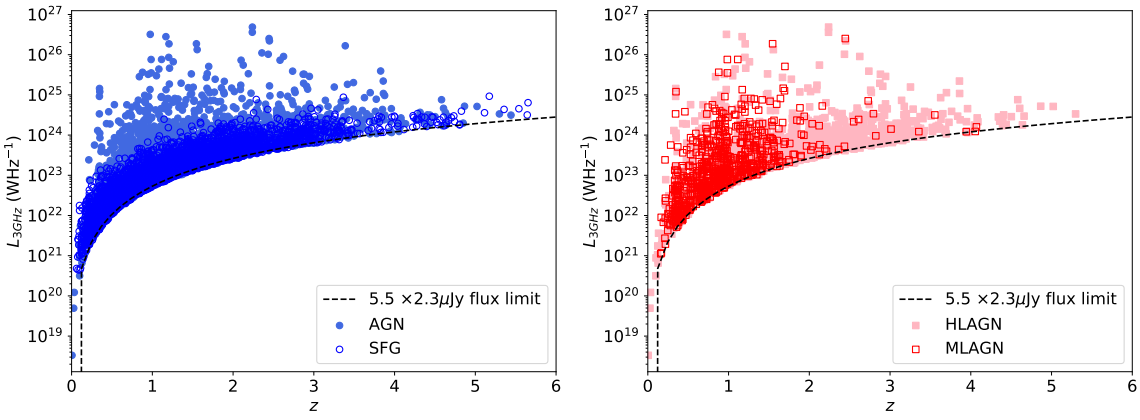


Figure 2.1: The locations of the radio sources used in this work after the mask from Laigle et al. (2016), defined using the optical and near-infrared data, has been applied.



(a) (b)

Figure 2.2: The luminosity distribution of (a) AGN (light blue filled circles) and SFGs (dark blue open circles) and (b) HLAGN (light red filled squares) and MLAGN (open red squares) with redshift. Also shown (black dashed line) is the luminosity limit for the flux limit of 5.5 × 2.3 μJy for a source with a spectral index, $\alpha = 0.7$.

2.2 The Angular Two-Point Correlation Function, $\omega(\theta)$

The angular Two-Point Correlation Function (TPCF) is one way in which the large scale structure of galaxies in the Universe can be understood. As described in Section 1.2, the angular TPCF, $\omega(\theta)$, quantifies the excess probability that galaxies are clustered compared to if they were randomly distributed across the sky. In this work, I made use of the estimator from Landy & Szalay (1993), as described in Section 1.2 and given in Equation 1.13. This uses pair counts of galaxies at given angular separations for both the observed and random catalogue to determine this excess clustering in the data.

As pairs of galaxies are counted, the uncertainties using this estimator would naively be Poissonian. However, this will underestimate the actual errors, as the counts in different bins are correlated. In order to better estimate the uncertainties, bootstrap resampling (e.g. Ling et al., 1986) is used. This has previously been used in clustering work such as by Cress et al. (1996); Gilli et al. (2009) and Lindsay et al. (2014b). When comparing the results using bootstrap errors to those from Poissonian errors (given in Equation 2.1, see Lindsay et al., 2014b), Cress et al. (1996) and Gilli et al. (2009) both found this increased uncertainties by a factor of $\sim$2.

$$\Delta\omega = \frac{1 + \omega(\theta)}{\sqrt{DD(\theta)}}. \tag{2.1}$$

To bootstrap resample the errors, a number of points (n) in the data set are replaced with random points from a copy of the data set. $\omega(\theta)$ is then recalculated using this new data set. Using this method means that in each calculation of $\omega(\theta)$, a certain source could be included in the sample used to calculate $\omega(\theta)$ multiple times, once, or not at all. This process can then be repeated a total of N times. The different

values of $\omega(\theta)$ can then be used to calculate the uncertainties as:

$$\Delta\omega_{bootstrap} = \left[\sum_{i=1}^{N}[\omega_i - \bar{\omega}]^2/(N-1)\right]^{\frac{1}{2}} \tag{2.2}$$

where $\bar{\omega}$ is the mean value of ω, given by $\sum_{i=1}^{N}\omega_i/N$. For each sample, approximately $\sim \frac{1}{3}$ of the data points were replaced (i.e. $n \sim \frac{1}{3}$) when calculating the bootstrapped errors. This should be a compromise between removing sufficient data points to accurately estimate the error but not removing too many as to make the data set appear too dissimilar from reality. This is repeated 100 times (N=100) and uncertainties on the correlation function are calculated using the 16^{th} and 84^{th} percentiles of the bootstrap samples.

2.2.1 Random Catalogue

To calculate $\omega(\theta)$ (Equation 1.13) it is necessary to have a complementary random catalogue. This random catalogue should represent a realistic population of sources that could be observed within a galaxy survey, but without any large scale structure influencing galaxy positions. If noise was uniform across the image, then random positions generated across the celestial sphere would be sufficient to generate the random catalogue. This would involve generating positions uniformly sampled in RA and cos(Dec).

However, the radio image used for this work has non-uniform noise across the field of view and so sources are unable to be detected uniformly across the map. This non-uniform noise is a result of large artefacts around bright sources and due to the fall off in sensitivity of the primary beam. Sources that are faint will be most affected by the non-uniform detection across the image, whilst bright sources will be able to be more uniformly detected. Therefore, when generating a random catalogue, it is necessary to determine whether a source would be detectable given the rms at the

random position generated. This should also take into account any masked regions (see Figure 2.1).

Therefore, in order to devise a random catalogue, I made use of realistic simulated extragalactic catalogues of the radio sky. Using the flux densities in the Square Kilometre Array (SKA) Design Study Simulated Skies simulations (S^3; Wilman et al., 2008, 2010, hereafter S^3), I constructed random catalogues through populating the image and determining whether a source was detectable. The flux density distributions of these simulated catalogues were discussed in Section 1.4 (see Figure 1.6). These mock catalogues contain flux information at multiple frequency bands. For the data used here, at 3 GHz, S^3 does not have flux density measurements, therefore I scaled the fluxes in S^3 from 1.4 GHz to 3 GHz using Equation 1.19.

The method I used to determine the random catalogue is illustrated in Figure 2.3 and is described below. It is essential that there are more random sources than data in order to ensure that the errors that arise in $\omega(\theta)$ are dominated by the real observations. Therefore each random catalogue had 20 times more sources than within the observed data set that was being investigated. The method used is:

1. A random position is selected uniformly in RA and cos(Dec) and the corresponding rms value at this location is found.

2. This position is checked to determine whether it is within the masked region or not. If it is in the masked region, then a new location is generated and the process is started again from Step 1.

3. As noise would be added onto any real measurement of a sources, a noise term is also included. This is determined by sampling from a Gaussian distribution centred on zero and with a width corresponding to the rms at that position.

4. A flux and redshift randomly chosen from the S^3 catalogue is assigned to the random source.

5. If the combination of the flux from S^3 (Step 4) plus the random noise (Step 3) is greater than N times the rms at the location then it can be included in the random sample. If this threshold is not met then the source is not included in the random catalogue and the process is repeated from Step (1). The value of N is determined by the flux limit criterion used in the catalogue of the survey, taken here as 5.5σ (i.e. N=5.5).

6. This processes is repeated until there are the desired number of sources in the random catalogue ($20 \times N_{data}$).

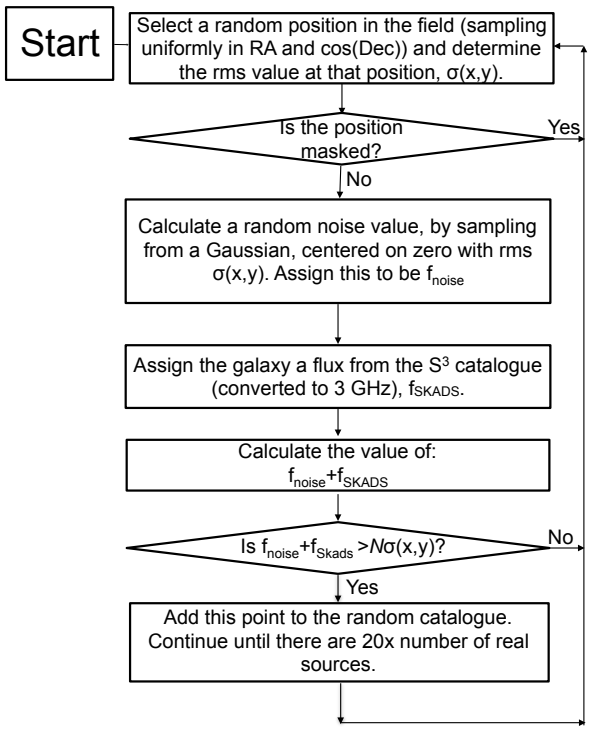


Figure 2.3: Flowchart illustrating how the random catalogues to calculate $\omega(\theta)$, were generated. In this, $\sigma(x, y)$ corresponds to the rms at the position x, y.

The faint sources, which are more numerous, are also most affected by this sampling method. Because of this, the underlying source count distribution will be important for creating an accurate random catalogue, especially when AGN and SFGs are

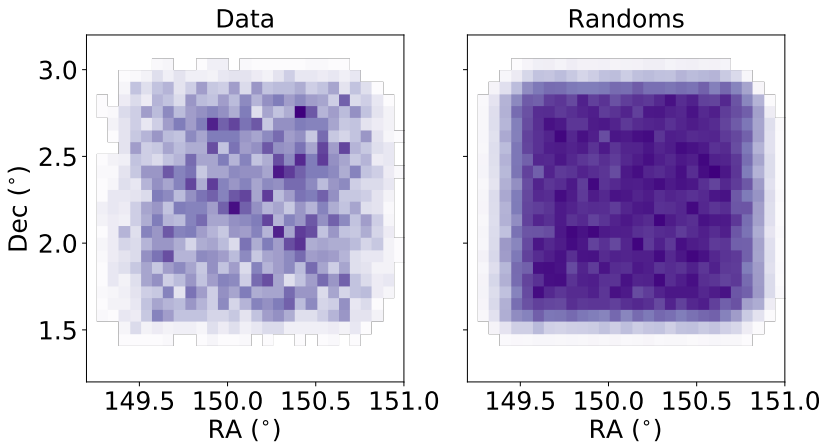


Figure 2.4: The density of all radio sources in the 3 GHz VLA COSMOS Survey (left) compared to random sources (right).

considered separately. The source counts at 1.4 GHz for S^3 can be seen in Figure 1.6. The SFGs are typically fainter than AGN and so SFGs will be more affected by the non-uniform noise in the image, and thus be more affected by this sampling method. Therefore, it is essential that when the clustering of SFGs is compared to AGN that the respective SFG and AGN populations from S^3 are used in Step (4), above. The distribution of random sources compared to data generated for all sources in the field can be seen in Figure 2.4. This shows that the fall off in sensitivity towards the edge of the radio image is being accounted for.

2.2.2 Fitting and the Integral Constraint

To fit the TPCF, methods similar to Blake & Wall (2002); Overzier et al. (2003); Wilman et al. (2003); Lindsay et al. (2014b) and Magliocchetti et al. (2017) were used. For this, I assumed a power law distribution of $\omega(\theta)$ on the large angular scales (where the '2 halo-term' dominates). On the smallest angular scales, sensitive surveys will be dominated by the '1-halo term' (see Section 1.2) where the clustering of galaxies from within the same dark matter halo dominates. Whilst I mentioned in Section 1.2 that in Blake & Wall (2002) a small scale clustering excess was observed

due to multi-component objects, observing a slope of $\omega \propto \theta^{-3.4}$, this is not expected here. This is because the catalogue used from Smolčić et al. (2017a,b) was cross-matched to multi-wavelength data in order to establish host galaxy counterparts and redshifts for these sources. It is therefore unlikely that there will be any additional excess in small scale clustering. Therefore, I assume a single power law distribution, such that $\omega(\theta) = A\theta^{-0.8}$. This is similar to the slopes found in Peebles (1974); Roche & Eales (1999); Norberg et al. (2002) and Lindsay et al. (2014b).

However it is not accurate to just fit $\omega(\theta)$ by minimising the χ^2 value using this power law functional form. This is because the value for $\omega(\theta)$ will be biased low at the largest angular separations. This is due to the fact that within a finite field size, then given the amount of galaxies observed within the survey the number of galaxy pairs at large angular separations is small and so $\omega(\theta)$ is underestimated. To account for this, I followed the work of Lindsay et al. (2014b) and incorporated an integral constraint when fitting the power law distribution of $\omega(\theta)$ (see Groth & Peebles, 1977). The integral constraint, σ^2_{IC} is defined in Groth & Peebles (1977) and can be approximated using an estimator as:

$$\sigma^2 = \frac{\sum RR(\theta)\omega_{true}(\theta)}{\sum RR(\theta)} \tag{2.3}$$

(see Roche & Eales, 1999; Lindsay et al., 2014b; Hatfield et al., 2016). From this, the true TPCF that would be observed if there was no limited field size, ω_{true}, is described by the power law form mentioned above. It can be related to the observed clustering with a model, ω_{model}, which has the value of ω_{true} with the integral constraint, σ^2_{IC}, subtracted. As σ^2_{IC} is a single number, it will have a much larger effect at large angles, where the value of $\omega(\theta)$ is smaller. This model is then fit to the data by minimizing χ^2 over the range $\theta \sim 10^{-3} - 0.5^{\circ}$, using angular bins uniformly separated in log space.

2.2.3 The Clustering Length and Bias

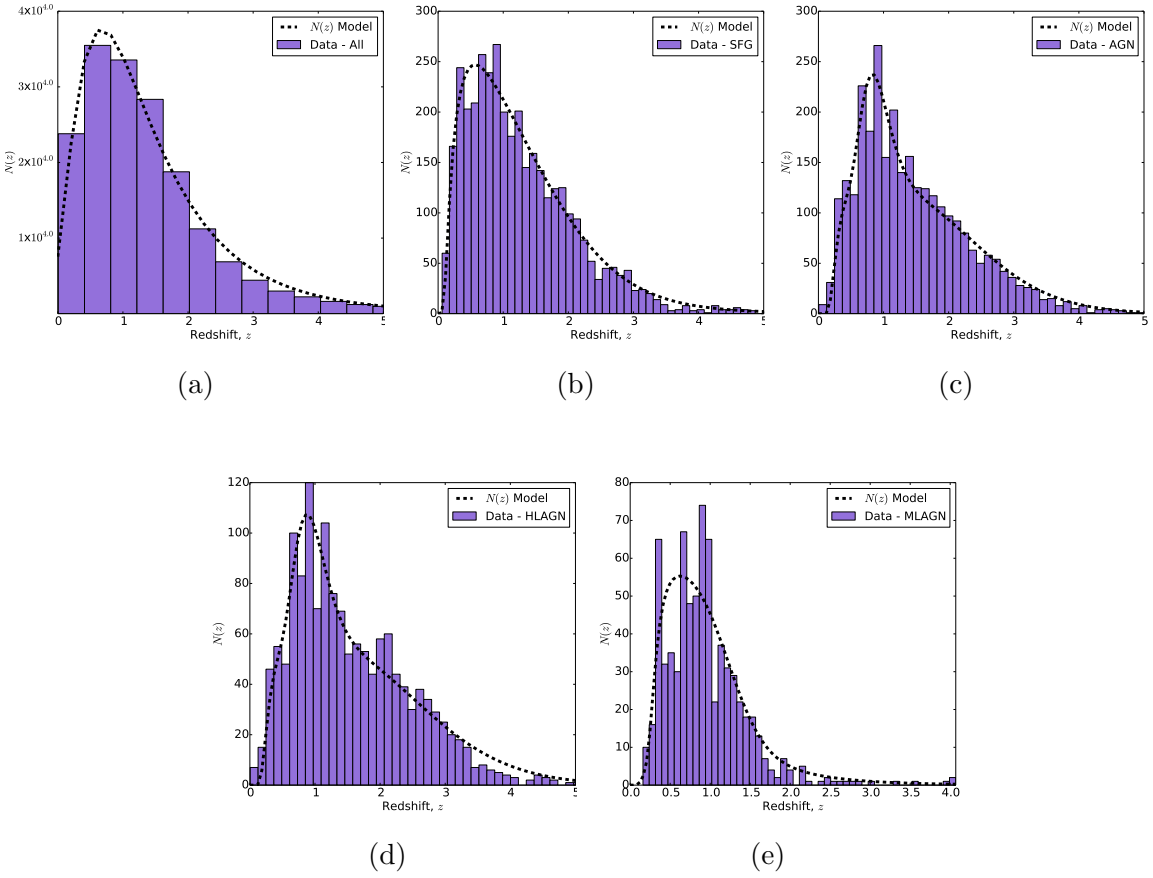
Bias, b, is the important quantity for constraining the haloes that different populations reside in, because of the relations between b and halo mass (Mo & White, 1996; Sheth & Tormen, 1999; Tinker et al., 2010). As described in Section 1.2.2, the bias can be determined from $\omega(\theta)$ using Limber inversion (Limber, 1953) to calculate a value of r_0, before relating this to the bias as in Equation 1.14. However the definition of r_0 in Equation 1.15 requires knowledge of the redshift distribution, $N(z)$. Whilst recent work in Allevato et al. (2016) used the Probability Density Functions (PDFs) of photometric redshifts to directly measure the spatial clustering, I instead used a smoothed representation of the redshift distribution. This functional form is not physically motivated, but was chosen to reflect the observed distributions. A histogram of the redshift distribution for each source population and the modelled redshift distributions can be seen in Figure 2.5. However when the clustering of the full radio population is considered, redshifts are not always available. As mentioned, only $\sim 86\%$ of these sources had redshifts and so the redshift distribution associated from the random catalogue is used to calculate r_0 and is modelled in Figure 2.5(a).

Using Equation 1.15, I then calculated r_0 and its associated uncertainties. By randomly generating a value of A and calculating its associated probability calculated from its χ^2 value, a sample of r_0 values was generated. The value of r_0 was then quantified by the median of this sample and the uncertainties from the 16^{th} and 84^{th} percentiles. By propagating the r_0 values of the sample through Equation 1.14 and evaluating the bias at the median redshift of the sample, the PDF for the bias and its quoted value and uncertainties were also determined.

If spectroscopic redshifts were available for the sources, then the real (or redshift) space correlation function would provide a more direct measurement of r_0 and b. This would alleviate any reliance on Limber Inversion (Equation 1.15). Limber inversion makes assumptions and approximations in order to obtain the Equation 1.15 (see

e.g. Peebles, 1980) including assuming that the sources are only observed over small angles and assumptions about accurate knowledge of the selection function of galaxies. Directly measuring spatial correlation functions on the other hand removes any reliance on modelling $N(z)$ and needing to observe over small angles only (although as the area of this field is $\sim$2 deg^2 and so this is not an issue here). Therefore with future surveys such as with WEAVE-LOFAR (Smith et al., 2016, ; and see Chapter 6) this will allow r_0 and b to be calculated more directly.

For the sources that were classified into SFGs and AGN, I also compared whether the redshift distribution from the random catalogue provided similar values of r_0 to when the model of the true redshift distribution was used. Both of the $N(z)$ distributions produced similar results, as such using the random catalogue for the $N(z)$ of the full AGN population should still give an appropriate value of r_0. In this work, I compared the clustering for the full population of the classified samples (e.g. SFGs and AGN) as well as a high and low redshift sub-sample. When calculating r_0 for these sub-populations, I made use of the same $N(z)$ model that was generated for the full redshift population and simply restricted the redshift range to reflect the redshifts being considered. The two redshift samples investigated were a low ($z \leq 1$) and a high redshift sample ($z > 1$).



(a) (b) (c)

(d) (e)

Figure 2.5: The observed redshift distribution (purple histogram) and the corresponding redshift model (black dashed line) for (a) All galaxies, (b) SFGs, (c) AGN, (d) HLAGN and (e) MLAGN.

2.3 Results

The results from investigating the clustering of radio sources using the VLA 3 GHz COSMOS Survey are presented in Table 2.1. For each investigation I recorded in Table 2.1 which population was being investigated (e.g. SFG or AGN) and the redshift range that they are being considered over. The fitting parameters of $\omega(\theta)$ as well as the values of the clustering length, r_0, and bias, b, that are determined using Equations 1.15 and 1.14 are also included. The associated fit $\omega(\theta)$ and the r_0 distributions are presented in Figures 2.6 - 2.10.

Galaxy Type	Redshift Range	N	Median redshift (z_{med})	Median Luminosity WHz^{-1}	$\log_{10}(A)$ Best	$\log_{10}(A)$ Median	Best r_0 (Mpc h^{-1})	Median r_0 (Mpc h^{-1})	$b(z_{med})$
All	All	8887	1.16	-	-2.83	$-2.8^{+0.1}_{-0.1}$	7.0	$7.0^{+0.3}_{-0.4}$	$2.7^{+0.1}_{-0.1}$
SFG	All	3704	1.07	1.28×10^{23}	-2.88	$-2.9^{+0.1}_{-0.1}$	6.1	$6.1^{+0.6}_{-0.7}$	$2.3^{+0.2}_{-0.2}$
AGN	All	2937	1.24	3.22×10^{23}	-2.57	$-2.6^{+0.1}_{-0.1}$	9.6	$9.6^{+0.7}_{-0.7}$	$3.6^{+0.2}_{-0.2}$
HLAGN	All	1456	1.35	3.48×10^{23}	-2.59	$-2.6^{+0.1}_{-0.1}$	9.8	$9.8^{+1.3}_{-1.4}$	$3.9^{+0.5}_{-0.5}$
MLAGN	All	726	0.88	1.28×10^{23}	-2.33	$-2.4^{+0.1}_{-0.2}$	9.5	$9.5^{+1.5}_{-1.8}$	$3.1^{+0.4}_{-0.5}$
SFG	z<1	1756	0.62	3.81×10^{22}	-2.56	$-2.6^{+0.1}_{-0.1}$	5.0	$5.0^{+0.5}_{-0.6}$	$1.5^{+0.1}_{-0.2}$
AGN	z<1	1126	0.70	6.85×10^{22}	-2.25	$-2.3^{+0.1}_{-0.1}$	6.9	$6.9^{+0.6}_{-0.7}$	$2.1^{+0.2}_{-0.2}$
HLAGN	z<1	488	0.69	6.38×10^{22}	-2.40	$-2.9^{+0.5}_{-1.9}$	5.7	$5.8^{+1.4}_{-1.8}$	$1.8^{+0.4}_{-0.5}$
MLAGN	z<1	485	0.70	6.82×10^{22}	-2.00	$-2.0^{+0.1}_{-0.1}$	9.7	$9.7^{+1.2}_{-1.3}$	$2.9^{+0.3}_{-0.3}$
AGN	z $\geq$ 1	1811	1.77	5.63×10^{23}	-2.58	$-2.6^{+0.1}_{-0.1}$	7.3	$7.3^{+0.9}_{-0.9}$	$3.5^{+0.4}_{-0.4}$
HLAGN	z $\geq$ 1	968	1.85	5.60×10^{23}	-2.68	$-3.1^{+0.4}_{-1.7}$	6.5	$6.5^{+1.5}_{-1.9}$	$3.2^{+0.7}_{-0.9}$
MLAGN	z $\geq$ 1	241	1.31	3.30×10^{23}	-2.02	$-2.5^{+0.5}_{-2.1}$	11.3	$11.3^{+2.5}_{-3.0}$	$4.3^{+0.9}_{-1.1}$

Table 2.1: Parameters describing the fits to the TPCF analyses: galaxy type; redshift range; number of sources (N); median redshift (z_{med}); median luminosity; best fit A; median A; best fit r_0, median r_0; best fit bias (b) and median b. The values of the bias have been evaluated at the median redshift of the sample being investigated.

2.3.1 Clustering of the Full Radio Source Population

The fit to $\omega(\theta)$, from investigating the clustering of the full sample of 8,887 radio sources, is shown in Figure 2.6(a). For these, I measure a clustering amplitude of $A = 10^{-2.8\pm0.1}$, which corresponds to a clustering length of $r_0 = 7.0^{+0.3}_{-0.4}$ Mpc h^{-1}, once the redshift distribution is taken into account. The distribution for these r_0 values is shown in Figure 2.6(b) and corresponds to a bias of $b = 2.7^{+0.1}_{-0.1}$ at the median redshift for the random sample, z_{med}=1.16. This catalogue is a combination of both AGN and SFGs and so r_0 will be an average of what is observed for both the AGN and SFG subsamples. Whilst the clustering length of full radio populations have been calculated in the literature previously (e.g. Overzier et al., 2003; Wilman et al., 2003; Lindsay et al., 2014b), these have typically been dominated by AGN with only a small fraction of SFGs. For this work, the numbers of sources that have been classified (see Table 2.1) suggest that SFGs are the dominant populations, constituting ∼55% of the classified sources. This is similar to the ∼60% of SFGs that is predicted using S^3.

2.3.2 Clustering by Source Type

SFGs

The measured angular two-point correlation functions for SFGs and the associated r_0 distribution can be seen in Figure 2.7. This shows the TPCF for all SFGs in (a) and for the low redshift sample of SFGs ($z \leq 1$) in (b). A high redshift sample ($z > 1$) is not presented as the clustering was poorly constrained. For the full redshift sample of SFGs, a clustering amplitude distribution of $A = 10^{-2.9^{+0.1}_{-0.1}}$ is measured. Through sampling A to find r_0, this resulted in a measured clustering length of $r_0 = 6.1^{+0.6}_{-0.7}\,\mathrm{Mpc}\,h^{-1}$, at a median redshift of z_{med}=1.07. This corresponds to a bias, $b = 2.3^{+0.2}_{-0.2}$. For the SFGs at low redshift (with $z \leq 1$), on the other hand, a larger amplitude is measured, $A = 10^{-2.6^{+0.1}_{-0.1}}$. This however corresponds to a smaller

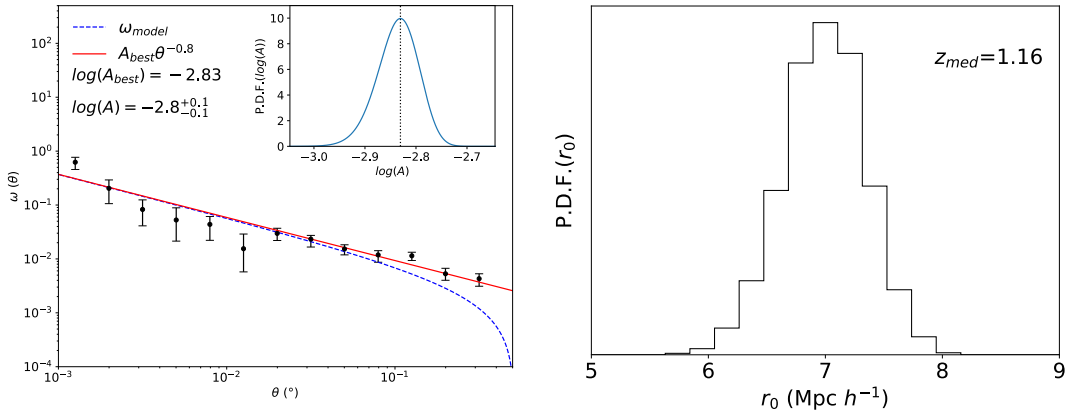
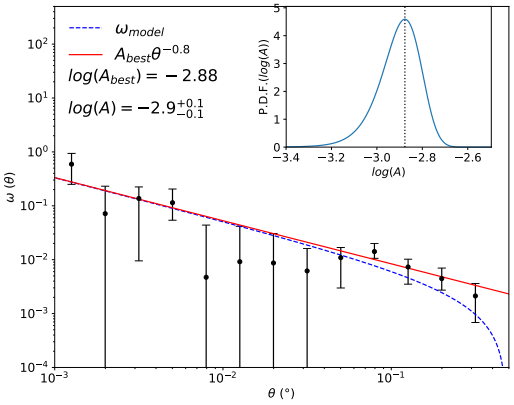


Figure 2.6: The measured TPCF and best fit for all COSMOS radio sources. The left panel shows the TPCF, the red solid line represents the best fit power law to $\omega(\theta)$ and the dashed blue line is the same as the red line but with the integral constraint, σ^2_{IC} subtracted from it. The inset in the top right corner shows the probability density function (PDF) for the value of A to fit the TPCF, with the dashed line showing the best fit value. The right panel shows the corresponding PDF for r_0.

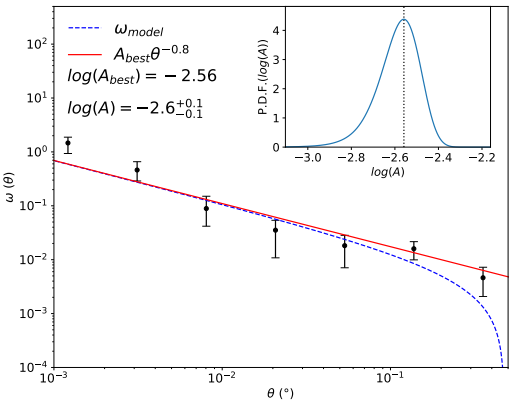
measured clustering length, $r_0 = 5.0^{+0.5}_{-0.6}\,\mathrm{Mpc}\,h^{-1}$ when the redshift distribution of SFGs (Figure 2.5b) is restricted to have $z \leq 1$. This corresponds to a bias of $1.5^{+0.1}_{-0.2}$, given the median redshift for this population is found to be $z_{med} = 0.62$. I show a comparison between the distribution of r_0 for the two samples in Figure 2.7(c).

AGN

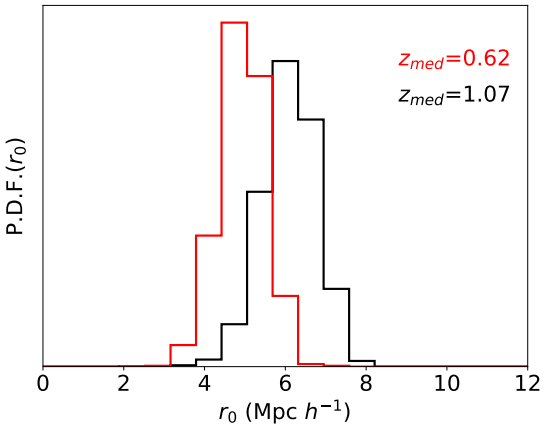
Figure 2.8 shows the angular two-point correlation functions for all AGN (a) and the low ($z \leq 1$) and high redshift AGN ($z > 1$) samples are shown in panels (b) and (c) respectively. For the full redshift sample of AGN, I measure a clustering amplitude of $A = 10^{-2.6^{+0.1}_{-0.1}}$, which relates to a clustering length of $r_0 = 9.6^{+0.7}_{-0.7}\,\mathrm{Mpc}\,h^{-1}$ at $z_{med} = 1.24$, corresponding to a bias of $b = 3.6^{+0.2}_{-0.2}$. In comparison, at low redshifts a larger clustering amplitude is found ($A = 10^{-2.3^{+0.1}_{-0.1}}$) but when the redshift distribution is accounted for this corresponds to a smaller value of r_0 than for the full redshift distribution, with a value of $r_0 = 6.9^{+0.6}_{-0.7}\,\mathrm{Mpc}\,h^{-1}$ at a lower redshift of $z_{med} = 0.70$,



(a) All SFGs

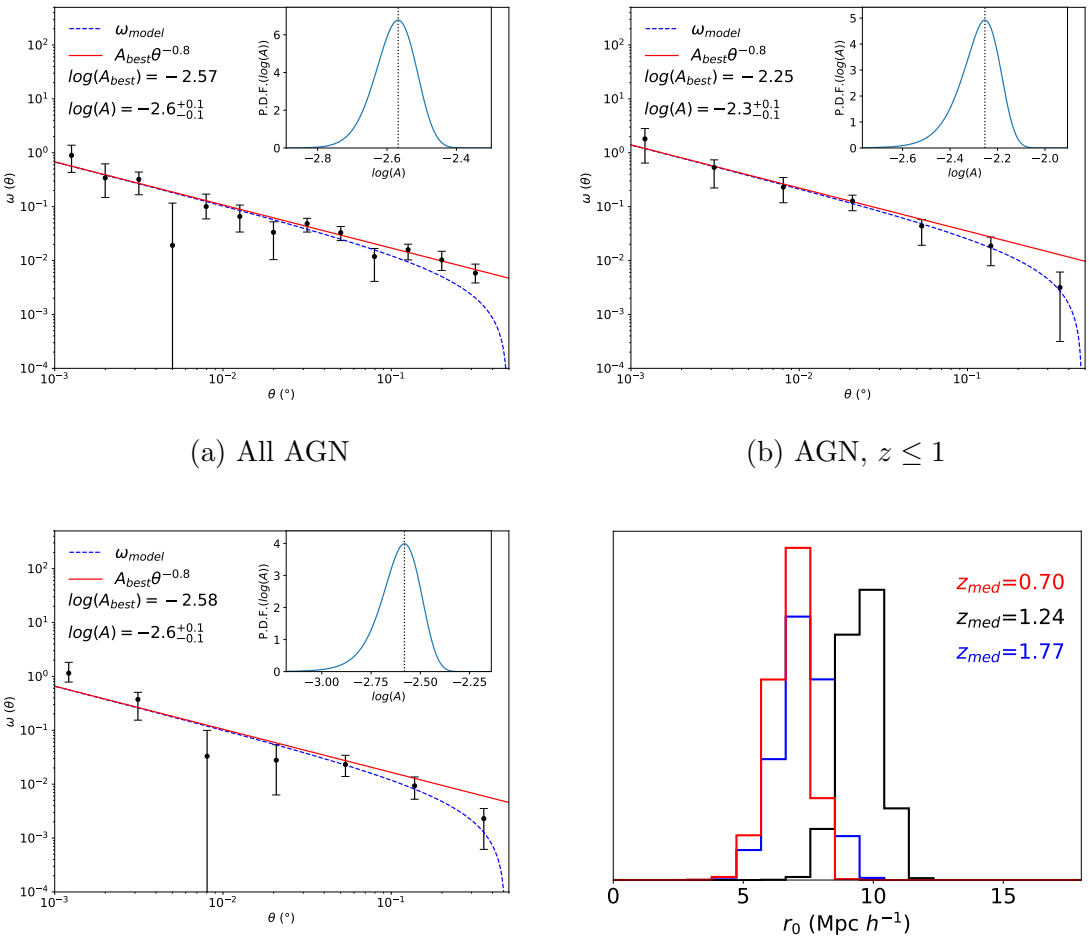


(b) SFGs, $z \leq 1$



(c) SFGs r_0

Figure 2.7: The measured TPCF and best fit for all SFGs (top left panel) and SFGs with $z \leq 1$ (top right panel). The red solid line represents the power law best fit to $\omega(\theta)$ and the dashed blue line is the same as the red line but with the integral constraint, σ^2_{IC} subtracted from it. The inset in the top right corner shows the probability density function for the value of A to fit the TPCF, with the dashed line showing the best fit value. The corresponding r_0 distributions are given in the bottom right panel for all z (black) and $z \leq 1$ (red).



(a) All AGN

(b) AGN, $z \leq 1$

(c) AGN, $z > 1$

(d) AGN r_0

Figure 2.8: The measured TPCF and best fits for all AGN (top left panel), AGN with $z \leq 1$ (top right panel) and AGN with $z \geq 1$ (bottom left panel). The red solid line represents the power law best fit to $\omega(\theta)$ and the dashed blue line is the same as the red line but with the integral constraint, σ^2_{IC} subtracted from it. The inset in the top right corner shows the probability density function for the value of A to fit the TPCF, with the dashed line showing the best fit value. The corresponding r_0 distributions are given in the bottom right panel for all z (black), $z \leq 1$ (red) and $z > 1$ (blue).

which corresponded to a value of $b = 2.1^{+0.2}_{-0.2}$. Finally, at high redshifts ($z > 1$) a lower value of r_0 is determined compared to when all AGN were considered, with $r_0 = 7.3^{+0.9}_{-0.9}$ at $z_{med} = 1.77$. However, given the differences in the redshift distributions, this corresponds to a more similar value of bias to the full z population, $b = 3.5^{+0.4}_{-0.4}$. A comparison of the r_0 distribution for these different samples is shown in Figure 2.8(d).

Finally, for HLAGN and MLAGN (described in Section 2.1) the angular correlation functions and fits to $\omega(\theta)$ are presented in Figure 2.9 (HLAGN) and Figure 2.10 (MLAGN). Again, for both of these, panels (a)-(c) show $\omega(\theta)$ for the full, low redshift ($z \leq 1$) and high redshift ($z > 1$) samples respectively. For the HLAGN, I measure $A = 10^{-2.6^{+0.1}_{-0.1}}$ for the full HLAGN sample, $A = 10^{-2.9^{+0.5}_{-1.9}}$ for the $z \leq 1$ sample and $A = 10^{-3.1^{+0.4}_{-1.7}}$ for the $z > 1$ sample. For the MLAGN, instead clustering amplitudes of $A = 10^{-2.4^{+0.1}_{-0.2}}$ for the full z sample, $A = 10^{-2.0^{+0.1}_{-0.1}}$ ($z \leq 1$) and $A = 10^{-2.5^{+0.5}_{-2.1}}$ ($z > 1$) were measured. Taking into account the redshift distributions in Figure 2.5, I measured a clustering length of $r_0 = 9.8^{+1.3}_{-1.4}\,\mathrm{Mpc}\,h^{-1}$ ($b = 3.9^{+0.5}_{-0.5}$) for the full HLAGN population ($z_{med} = 1.35$) whilst a similar clustering length for MLAGN is measured, $r_0 = 9.5^{+1.5}_{-1.8}\,\mathrm{Mpc}\,h^{-1}$ ($b = 3.1^{+0.4}_{-0.5}$). However, the median redshift for the MLAGN population is biased towards lower redshifts, with a corresponding median redshift over the full population of z_{med}=0.88. For the low redshift subsamples ($z \leq 1$) the median redshifts for the two populations are much more comparable, with $z_{med} \sim$0.7. At these redshifts, the measured clustering length for MLAGN is $r_0 = 9.7^{+1.2}_{-1.3}\,\mathrm{Mpc}\,h^{-1}$ ($b = 2.9^{+0.3}_{-0.3}$) whilst for HLAGN a smaller clustering length is measured, $r_0 = 5.8^{+1.4}_{-1.8}\,\mathrm{Mpc}\,h^{-1}$ ($b = 1.8^{+0.4}_{-0.5}$). This suggests the MLAGN are much more clustered than the efficiently accreting HLAGN. Restricting to higher redshifts only, again the HLAGN population are biased to much larger redshifts. These HLAGN are at a median redshift of 1.85 and a clustering length of $r_0 = 6.5^{+1.5}_{-1.9}\,\mathrm{Mpc}\,h^{-1}$ ($b = 3.2^{+0.7}_{-0.9}$) is measured. The MLAGN are instead at a median redshift of $z = 1.31$, which is much more similar to the median redshift of the full

HLAGN population. For these high redshift MLAGN, a large clustering length is measured, $r_0 = 11.3^{+2.5}_{-3.0}\,\mathrm{Mpc}\,h^{-1}$ ($b = 4.3^{+0.9}_{-1.1}$). A comparison of the distribution of r_0 are shown in Figure 2.9(d) for HLAGN and Figure 2.10(d) for MLAGN.

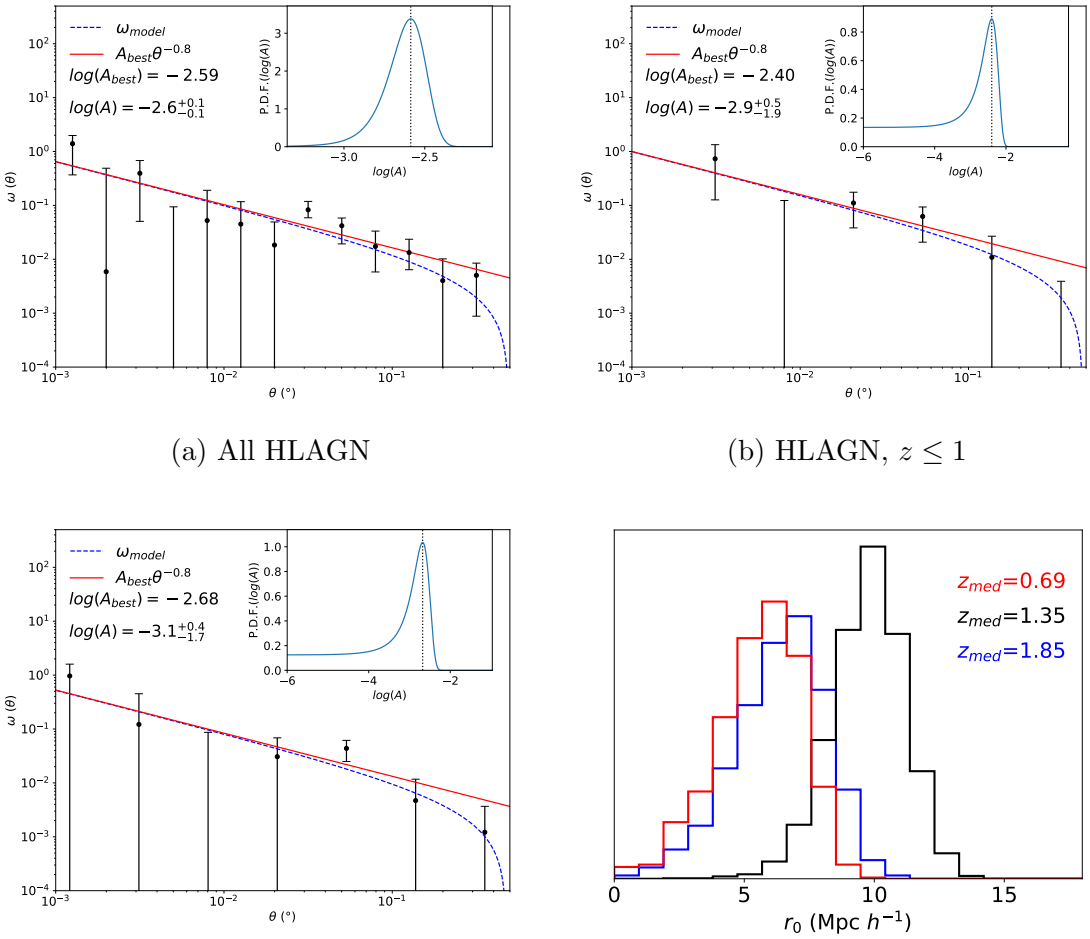
2.4 Discussion

Comparing the measured clustering amplitudes is challenging, because the redshift distributions may be skewed high or low for different galaxy populations. Therefore, in this discussion I compare the measured r_0 and b values which account for the redshift distribution of each sample. These are essential to compare the dark matter haloes that galaxies reside in.

2.4.1 The Clustering of the Full Radio Source Population

In Figure 2.11(a), the comparison of the clustering length for the full radio source population is compared to previous work from Overzier et al. (2003); Magliocchetti et al. (2004); Gilli et al. (2005); Lindsay et al. (2014b) and Magliocchetti et al. (2017). Also shown in Figure 2.11 are the $r_0(z)$ models for the different galaxy populations in the extragalactic simulated skies catalogue of Wilman et al. (2008) for FRI/FRII sources, radio quiet quasars (RQQ), starburst galaxies (SB) and normal star forming galaxies (SFG).

As described in Section 2.3.1, the population of sources observed in the 3 GHz VLA COSMOS survey are a combination of both AGN and SFGs, and are dominated by the SFG populations. For previous surveys investigating the clustering of radio galaxies, the samples used are dominated by AGN. For example, Magliocchetti et al. (2017) also investigated $\omega(\theta)$ for radio galaxies in the COSMOS field. They used a flux limit of 0.15 mJy and identified 250 SFGs compared to 640 AGN (corresponding to a SFG percentage of $\sim 30\%$). Whilst in the work of Lindsay et al. (2014b), a 1 mJy



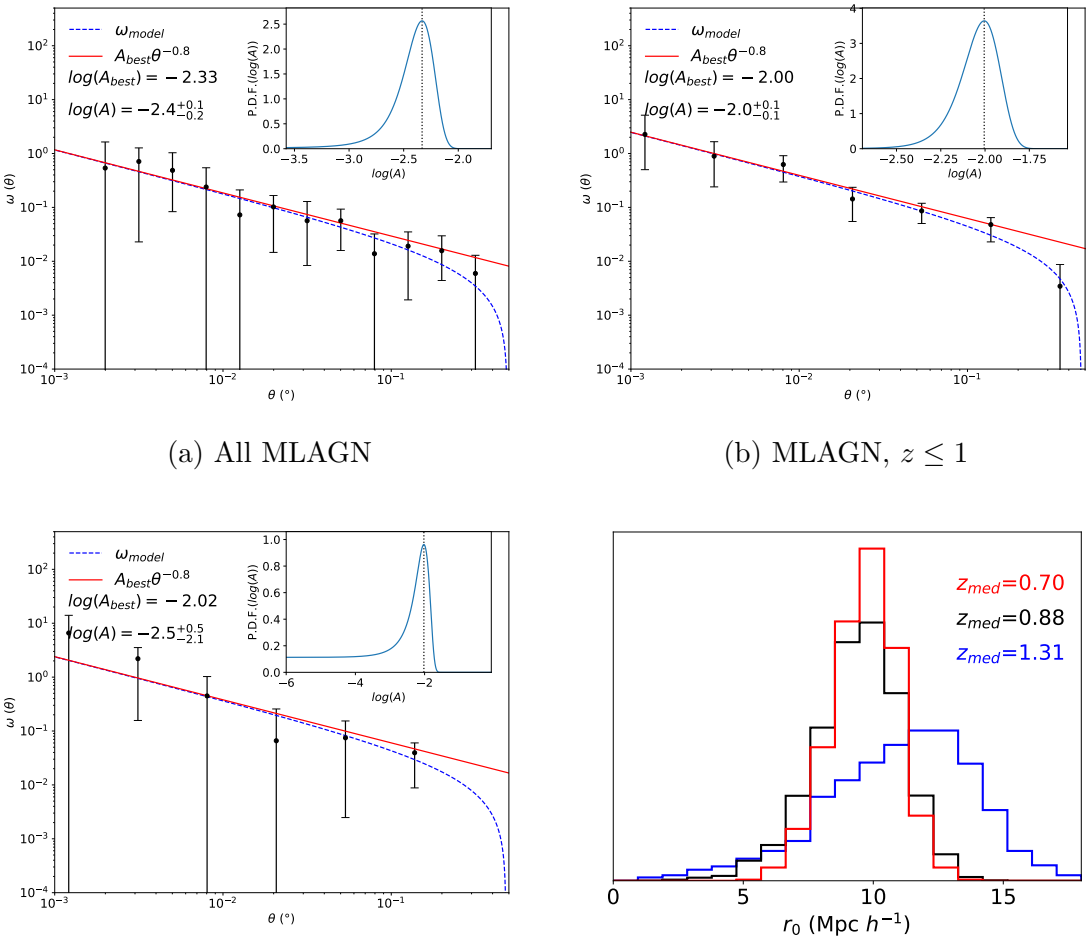
(a) All HLAGN

(b) HLAGN, $z \leq 1$

(c) HLAGN, $z > 1$

(d) HLAGN r_0

Figure 2.9: The measured TPCF and best fits for All HLAGN (top left panel), HLAGN with $z \leq 1$ (top right panel) and HLAGN with $z \geq 1$ (bottom left panel). The red solid line represents the power law best fit to $\omega(\theta)$ and the dashed blue line is the same as the red line but with the integral constraint, σ^2_{IC} subtracted from it. The inset in the top right corner shows the probability density function for the value of A to fit the TPCF, with the dashed line showing the best fit value. The corresponding r_0 distributions are given in the bottom right panel for all z (black), $z \leq 1$ (red) and $z > 1$ (blue).



(a) All MLAGN

(b) MLAGN, $z \leq 1$

(c) MLAGN, $z > 1$

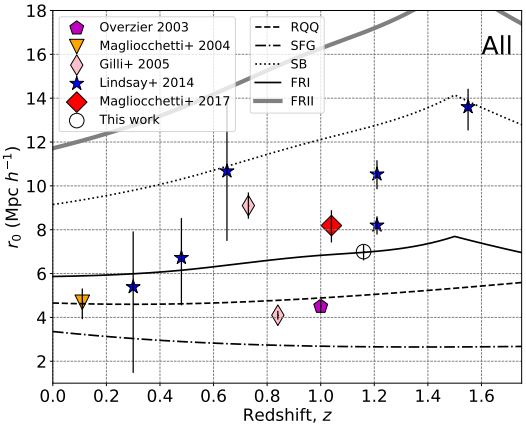
(d) MLAGN r_0

Figure 2.10: The measured TPCF and best fits for All MLAGN (top left panel), MLAGN with $z \leq 1$ (top right panel) and MLAGN with $z \geq 1$ (bottom left panel). The red solid line represents the power law best fit to $\omega(\theta)$ and the dashed blue line is the same as the red line but with the integral constraint, σ^2_{IC} subtracted from it. The inset in the top right corner shows the probability density function for the value of A to fit the TPCF, with the dashed line showing the best fit value. The corresponding r_0 distributions are given in the bottom right panel for all z (black), $z \leq 1$ (red) and $z > 1$ (blue).

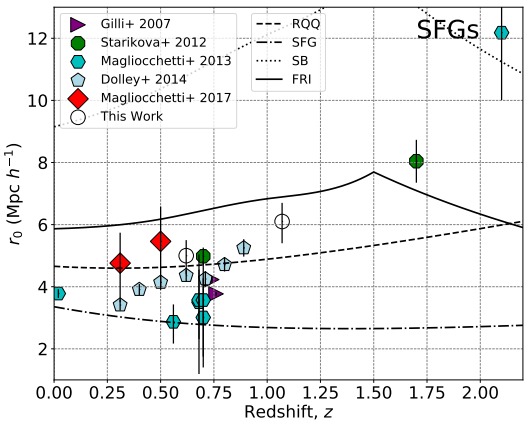
flux limit was used and hence are also dominated by AGN. The corresponding flux limit for this work at 1.4 GHz is instead ∼22 μJy. The work by Magliocchetti et al. (2017) is most comparable to this work, however they find a value of $r_0 = 8.19^{+0.70}_{-0.77}$ Mpch^{-1} at $z \sim 1.04$. This is larger than measured for this work ($7.0^{+0.3}_{-0.4}$ Mpc h^{-1} at z_{med}=1.16) whilst also being at a lower median redshift (where r_0 should be smaller). This decline in r_0 at fainter flux limits has been investigated in previous work (Wilman et al., 2003) and is likely to reflect the fact that with fainter flux limits, lower mass haloes can be traced. However, as there is also a different mixture of populations, comparing these clustering lengths to gain physical insights is challenging. As such, in the following, I move on to discussing the clustering of sources for the AGN and SFG populations separately.

2.4.2 The Clustering of SFGs and AGN

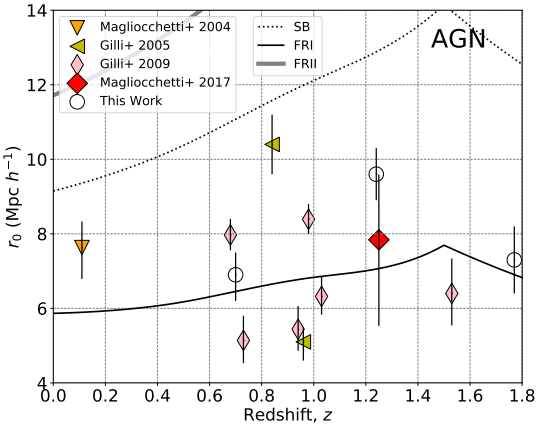
For SFGs, the measured value of r_0 is again compared to previous measurements from Gilli et al. (2007); Starikova et al. (2012); Magliocchetti et al. (2013); Dolley et al. (2014) and Magliocchetti et al. (2017) in Figure 2.11(b). As the median redshift for the full sample (1.07) is much larger that for the SFGs in Magliocchetti et al. (2017) ($z \sim 0.5$), the low redshift sample ($z \leq 1$) in this work ($z_{med} \sim 0.6$) is compared to that of Magliocchetti et al. (2017). For these SFGs I measured a clustering length of $r_0 = 5.0^{+0.5}_{-0.6}\,\mathrm{Mpc}\,h^{-1}$ at $z_{med} = 0.62$ which is comparable to that found in Magliocchetti et al. (2017), $r_0 = 5.46^{+1.12}_{-2.10}$ Mpc h^{-1}. This is also similar to the work of Starikova et al. (2012), who found $r_0 = 5.0^{+0.3}_{-0.3}\,\mathrm{Mpc}\,h^{-1}$ at $z_{med} = 0.7$ using star forming galaxies observed with Spitzer. The value of r_0 for low redshift SFGs is also similar to (but slightly higher than) those measured by Dolley et al. (2014) who investigated the evolution of the clustering of mid-IR selected SFGs. From the two r_0 measurements for SFGs in this work I find that, over the redshift range considered, the clustering length of SFGs increase. This increase follows a similar evolution in r_0



(a)



(b)



(c)

Figure 2.11: Comparison of the r_0 values calculated here (open circles) to previous studies for: All galaxies (top); SFGs (bottom left) and AGN (bottom right). These are for for all z, $z \leq 1$ and $z \geq 1$ where appropriate. Previous work is from: Overzier et al. (2003) (magenta pentagon); Magliocchetti et al. (2004) (orange downwards triangles); Gilli et al. (2005) (yellow left triangles); Gilli et al. (2007) (purple right triangles); Gilli et al. (2009) (pink thin diamonds; Starikova et al. (2012) (green octagon); Magliocchetti et al. (2013) (cyan hexagons); Lindsay et al. (2014b) (blue stars); Dolley et al. (2014) (pale blue pentagons) and Magliocchetti et al. (2017) (red diamonds). This is adapted from Figure 4 in Magliocchetti et al. (2017). For all studies shown, the redshift presented are the median redshift of the sample, however the samples will span a large range in redshifts, which is not shown. The lines show the evolution of r_0 as expected from Wilman et al. (2008) for: Radio Quiet Quasars (RQQ, dashed); normal Star Forming Galaxies (SFG, dot-dashed); StarBursts (SB, dotted); Fanaroff-Riley Type 1 AGN (FRI, solid black) and Fanaroff-Riley Type 2 AGN (FRII, solid grey). The decline in r_0 for some of these lines (e.g. for SBs and FRIs) at $z > 1.5$ is related to the constant bias imposed by Wilman et al. (2008) at high redshift.

to that found in Starikova et al. (2012); Dolley et al. (2014) and Magliocchetti et al. (2017). Measurements of r_0 for SFGs ($z \leq 1$) presented here are larger than the work of Magliocchetti et al. (2013) (who studied the clustering of far-IR galaxies observed with Herschel), but in general it appears that the measurements from this Chapter are in good agreement with previous work and show similar evolution with redshift.

This increase in r_0 with redshift likely reflects that at higher redshifts, SFGs with higher SFRs are observed. This is because at higher redshifts, more luminous sources are preferentially observed in any flux limited sample (see Figure 2.2). As radio luminosity is known to correlate with SFR (see e.g. Bell, 2003; Garn et al., 2009; Davies et al., 2017), galaxies with higher SFRs are therefore being observed. The clustering of SFGs is known to increase with luminosity (see e.g. Norberg et al., 2002; Zehavi et al., 2005, 2011), stellar mass (Hatfield et al., 2016; Cochrane et al., 2018) and SFR (Cochrane et al., 2018) and can be explained by these sources residing in more massive galaxies which require larger mass haloes to host them.

However, as can be seen in Figure 2.11(b), the values of r_0 measured in this work are much larger than used for normal SFGs in Wilman et al. (2008), where a halo mass of $10^{11}h^{-1}M_{\odot}$ was assumed. This can be seen by the dot-dashed line in Figure 2.11(b). Wilman et al. (2008) separated SFGs into normal and starburst galaxies (which were assumed to have halo masses of $5 \times 10^{13}h^{-1}M_{\odot}$), whereas the distinction is not made here. However if the contribution of starbursts is small, this suggests that the halo masses assumed for SFGs in Wilman et al. (2008), 10^{11} $M_{\odot}$, is low[2].

For AGN, a comparison of the measured r_0 values presented in this work is compared to that of Magliocchetti et al. (2004); Gilli et al. (2005, 2009) and Magliocchetti et al. (2017) in Figure 2.11(c). However, as can be seen from Figure 2.11(c), there is a lot of variation in the values determined by the previous work compared to here. This may reflect the fact that different AGN populations are observed within these

[2]From the SFRs in the VLA 3 GHz catalogue (Smolčić et al., 2017a,b), ∼30% of Clean SFGs have high SFRs (≥ 100 $M_{\odot}\mathrm{yr}^{-1}$).

surveys, which in turn may reside in different environments. The work by Magliocchetti et al. (2017), who compared the clustering of AGN within the same field at 1.4 GHz, are the most comparable for this work. Their value, $r_0 = 7.84^{+1.75}_{-2.31}\,\mathrm{Mpc}\,h^{-1}$ at $z_{med} \sim 1.2$, is consistent (within the uncertainties) with my value for the full redshift sample, $r_0 = 9.6^{+0.7}_{-0.7}\,\mathrm{Mpc}\,h^{-1}$ at $z_{med} = 1.24$. Whilst the r_0 values for SFGs increased with z, for the high redshift sample of AGN investigated in this work, the measured value of $r_0 = 7.3^{+0.9}_{-0.9}\,\mathrm{Mpc}\,h^{-1}$ is smaller than when the full redshift distribution is considered. This decline in r_0 is also observed for the FRI population in Wilman et al. (2008) for $z \geq 1.5$, however this was imposed by Wilman et al. (2008) to prevent the bias growing exponentially. As the lines correspond to assumed, fixed halo masses in Wilman et al. (2008), for a given halo mass r_0 should increase with z. My work therefore suggests that the halo masses in which these AGN reside decline at high redshifts (see Section 2.4.4).

To compare the halo masses that SFGs and AGN reside in, I compare their bias evolution (Figure 2.12). As the populations have very different redshift distributions (and b is known to evolve with redshift as can be seen from the Wilman et al., 2008, models in Figure 2.12), it is not possible to directly compare the bias values for the full SFG and AGN samples. The low redshift samples, however, have much more comparable redshifts for these two populations ($z \sim 0.6-0.7$). At these redshifts, it is clear from Figure 2.12 that AGN are more biased than SFGs. This suggests that AGN reside in larger mass haloes. This reflects the idea that radio AGN inhabit massive ellipticals (McLure et al., 1999; Dunlop & McLure, 2003) which are more massive than SFGs and are also redder in colour (e.g. Roberts & Haynes, 1994; Strateva et al., 2001). Measurements of clustering are both known to increase with stellar mass (Hatfield et al., 2016; Hatfield & Jarvis, 2017; Cochrane et al., 2018) and with redder colours (e.g. Zehavi et al., 2005, 2011). The larger clustering of AGN compared to SFGs therefore reflects that galaxies hosting radio AGN are massive and need large

dark matter haloes to support them.

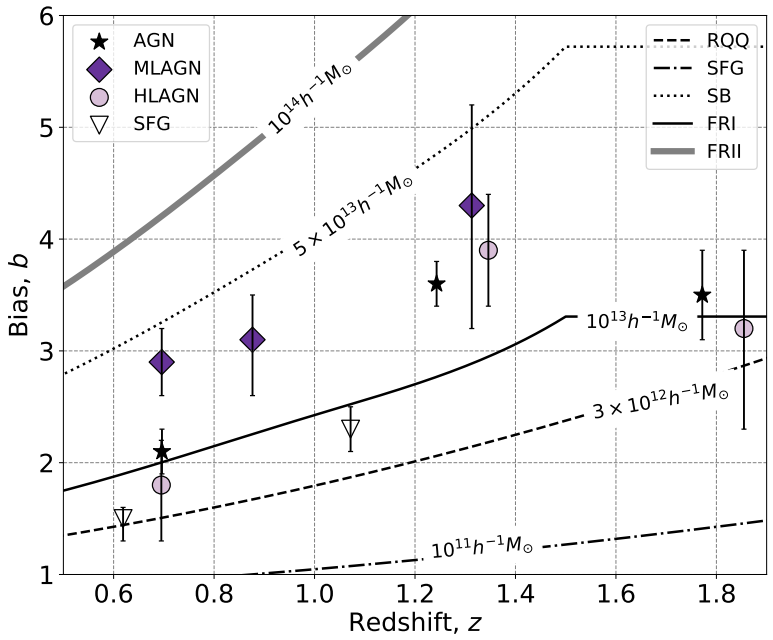


Figure 2.12: The bias as a function of redshift for the different galaxy types investigated here: SFGs (white triangles), AGN (black stars), HLAGN (dark purple diamonds), MLAGN (light purple circles). For all the galaxy sub-populations the results including sources of all redshifts as well as sources with $z \leq 1$ and $z \geq 1$ (except the SFGs and low luminosity AGN) are shown. The lines marked are the expected evolution of bias with redshift for different galaxy types from Wilman et al. (2008). These are for Radio Quiet Quasars (RQQ, dashed); normal Star Forming Galaxies (SFG, dot-dashed); StarBursts (SB, dotted); Fanaroff-Riley Type 1 AGN (FRI, solid black) and Fanaroff-Riley Type 2 AGN (FRII, solid grey).

By using the bias distribution from Wilman et al. (2008) (Figure 2.12) for populations of different halo mass, I make estimates here for the typical halo masses that these radio source populations reside in. By comparing the amplitude of bias to those assumed in Wilman et al. (2008), Figure 2.12 shows a halo mass for SFGs in the range of $M_h \sim 3 - 5 \times 10^{12}\, h^{-1}\, \mathrm{M}_\odot$. This is between the assumed bias models for FRI ($10^{13}h^{-1}M_\odot$) sources and RQQs ($3 \times 10^{12}h^{-1}M_\odot$). For AGN, on the other hand, I find a typical halo mass of $M_h \sim 10^{13} - 3 \times 10^{13}\, h^{-1}\, \mathrm{M}_\odot$, between that of FRI sources

and starbursts. This is much lower than for that of FRII sources, where a halo mass of $10^{14}h^{-1}M_{\odot}$ was assumed by Wilman et al. (2008). At high redshifts, though, the bias for the AGN population appears to flatten, suggesting that AGN are residing in smaller halo masses at higher redshifts at $z \gtrsim 1$.

2.4.3 AGN split by Accretion Efficiency

For AGN split by efficiency, the low redshift samples are most comparable, due to their similar redshift range. For the low redshift samples, Figure 2.12 shows that the inefficient MLAGN are much more clustered than their efficiently accreting counterparts. At these redshifts ($z \sim 0.7$) HLAGN have a bias of $b = 1.8^{+0.4}_{-0.5}$, whilst the MLAGN are more strongly clustered, $b = 2.9^{+0.3}_{-0.3}$. This agrees with the work of Hardcastle (2004); Tasse et al. (2008); Gendre et al. (2013); Ching et al. (2017) and Miraghaei & Best (2017) who suggest that the LERG population are in dense environments. This implies that the efficiency of AGN is related to the dark matter halo environments in which these galaxies reside. This may be correlated to how these AGN are fuelled. In more clustered, larger mass haloes it is likely that the environments are hotter (an increase in temperature with mass has been suggested for galaxy clusters, see e.g. Shimizu et al., 2003; Main et al., 2017) and so it may be more challenging to efficiently fuel the AGN. This is because the material in these hotter environments that will fuel the AGN will have larger thermal energy. Due to this increased thermal energy, accreting such material will be challenging. Ineson et al. (2015) indeed suggest that LERGs (MLAGN) reside in hotter environments than HERGs (HLAGN).

However, it is possible that these conclusions are affected by how the random catalogues are generated. As discussed in Section 2.2.1, whether a source is included within the random catalogue (or not) is most important for those sources that are near the flux limit. In Wilman et al. (2008) sources are categorised as FRI and

FRII sources, not HL/MLAGN. As such, I used the full AGN population to generate random samples for these two sub-populations. This may not be accurate near the flux limit if the fainter sources are dominated by one population. Therefore, when considering the clustering of HLAGN and MLAGN at low redshifts, I also compared the bias determined using a more conservative, 22μJy flux limit. Using this more conservative flux limit, I also found that MLAGN are significantly more clustered than HLAGN, with MLAGN found to have a bias of $b = 3.6^{+0.4}_{-0.4}$ at $z_{med} = 0.7$ whilst the HLAGN had a smaller bias, $b = 1.1^{+0.3}_{-0.4}$ at a similar redshift. This reinforces the conclusions that MLAGN reside in more massive haloes than HLAGN.

At $z \sim$1.3, the high redshift MLAGN population are at a similar median redshift to the HLAGN full population and have similar bias: for MLAGN $b = 4.3^{+0.9}_{-1.1}$, whilst for HLAGN $b = 3.9^{+0.5}_{-0.5}$. This shows that whilst MLAGN are much more clustered at low redshifts compared to HLAGN, the two populations may be similarly clustered at higher redshifts. This suggests that the relationship between these sources and their dark matter haloes evolves differently for the two populations. Indeed the populations themselves are known to evolve differently from studies of their luminosity functions (see e.g. Best & Heckman, 2012; Pracy et al., 2016). However, I have only considered three redshift ranges in this work (due to the small number of sources) and it is important to investigate the evolution in the bias for these different populations using future surveys (see Chapter 6). This will allow a much greater understanding of the evolution of these populations with redshift, not being limited to averaging across large redshift ranges. Finally, for the HLAGN at high redshifts, the bias of these sources also appears to flatten as it does for the full AGN population (see Section 2.4.4).

As presented in Figure 4 of Heckman & Best (2014), most HERGs are FRII sources, whilst LERGs are a mixture of both FRI and FRII type sources. From Figure 2.12 it is clear that the assumed halo masses of FRII sources in Wilman et al.

(2008) are much larger than observed here for HLAGN, or for any of the populations. MLAGN appear to reside in haloes of masses $M_h \sim 3-4\times10^{13}\,h^{-1}$ $\mathrm{M}_\odot$ whilst HLAGN have halo masses of $M_h \sim 1-2\times10^{13}\,h^{-1}$ $\mathrm{M}_\odot$. This therefore suggests that the halo masses assumed in Wilman et al. (2008) may be too high for some populations (or too low for SFGs) and needs to be reconsidered. This is especially important for the work discussed in Ferramacho et al. (2014) and Raccanelli et al. (2015), who use the differences in the bias for different radio source populations to trace non-Gaussianity (i.e. that random fluctuations in the early Universe after the Big Bang were not completely Gaussian in nature).

2.4.4 Halo Mass Evolution for AGN

The clustering results for the HLAGN population and full AGN sample appear to suggest that the bias is flattening for $z > 1.2$, whilst data above this redshift was not available for the MLAGN population. This is also seen in Figure 2.12 for some of the populations in Wilman et al. (2008) but, as mentioned earlier, this was artificially imposed for these populations in order to ensure that the bias did not grow exponentially. As I observe flattening of the bias this suggests the halo masses which are hosting these sources are declining at these high redshifts. By making use of the package `Colossus` (Diemer, 2018) and the bias model from Tinker et al. (2010), it is possible to convert the measured bias to the halo masses that these correspond to, given the median redshifts. These are shown in Figure 2.13 for (a) all AGN, (b) HLAGN and (c) MLAGN. The SFGs are not included in this as the high redshift SFG bias was poorly constrained. Figure 2.13 shows that whilst there is no evolution in the halo masses that MLAGN reside in (panel c) over the redshift ranges considered, both the full AGN and HLAGN are supported by lower mass haloes at the highest redshifts. A small decline in halo mass with redshift has also been observed for the clustering of X-ray AGN (Allevato et al., 2016), however this was at much larger

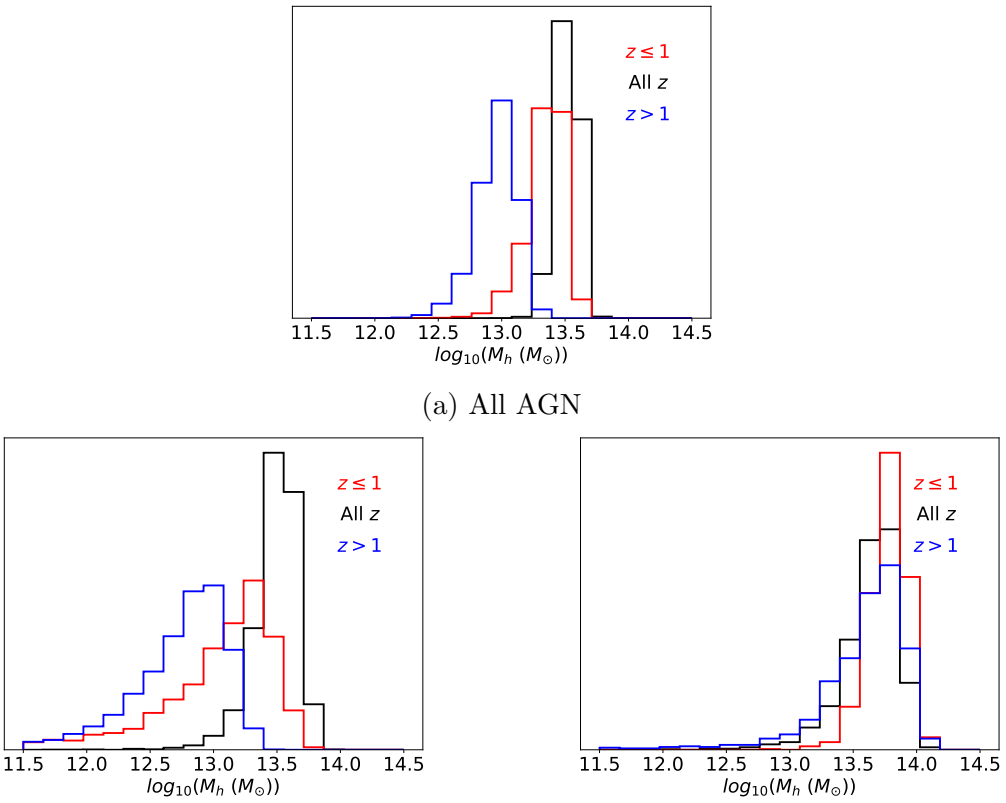
redshifts $z \gtrsim 3$. A decline in the halo mass needed to support powerful AGN may suggest that a rich gas supply is more abundant at these higher redshifts, allowing these bright AGN to be fuelled by smaller haloes.

2.5 Conclusions

In this Chapter, I have presented work from Hale et al. (2018) where I investigated the angular clustering of radio selected AGN and SFGs from the VLA 3 GHz COSMOS Survey (Smolčić et al., 2017a,b). Through determining the angular two-point correlation function, $\omega(\theta)$, for both AGN and SFGs, as well as AGN split by accretion efficiency, I investigated how biased these sources are and determined the masses of the dark matter haloes in which they reside. The angular clustering was measured for the entire samples as well as for a low ($z \leq 1$) and high ($z > 1$) redshift subsample.

In this work, I found that SFGs appear to inhabit lower mass haloes than AGN at $z \leq 1$, with an average halo mass of $M_h \sim 3 - 5 \times 10^{12}\, h^{-1}\, \mathrm{M}_\odot$ for SFGs whilst a larger halo mass of $M_h \sim 1 - 2 \times 10^{13}\, h^{-1}\, \mathrm{M}_\odot$ for AGN. This suggests that more massive haloes are needed on average to support an AGN. This supports the idea that AGN often inhabit massive elliptical galaxies (McLure et al., 1999; Dunlop & McLure, 2003) which will need to be supported by more massive haloes (see e.g. Hatfield et al., 2016; Tinker et al., 2017; Cochrane et al., 2018). Furthermore, by investigating AGN split by accretion efficiency, I found there is a difference in the haloes that these two populations reside in. At $z \sim 0.7$, MLAGN have a bias of $b = 2.9^{+0.3}_{-0.3}$ whilst I measure a much lower bias of $b = 1.8^{+0.4}_{-0.5}$ for HLAGN. This suggests that the environment that these AGN reside in is correlated with their accretion mode, which is also found in work by Gendre et al. (2013); Ineson et al. (2015); Ching et al. (2017) and Miraghaei & Best (2017). This suggests that MLAGN reside in larger, more clustered haloes. As these are likely hotter, the energy of the material which is being

accreted onto the AGN will be greater than in cooler environments. This may make it more challenging to accrete. The work in this Chapter therefore suggests there is a link between environmental properties and the host properties of radio sources. To investigate this further, I next study the relationship between environment and the observed luminosity distribution of radio sources for these different populations, by investigating the conditional luminosity function.



(a) All AGN

(b) HLAGN

(c) MLAGN

Figure 2.13: Evolution of the halo masses, M_h, that the different AGN populations in habit using the full samples (black), low redshift samples (red) and high redshift samples (blue). This is for (a) all AGN, (b) HLAGN and (c) MLAGN.

Chapter 3

The Conditional Luminosity Function

Investigating how different galaxy populations evolve is crucial in understanding what factors may be important in driving their evolution. Frequently this is investigated through studying the Luminosity Function of galaxies across the electromagnetic spectrum (LF, e.g. Bell, 2003; Blanton et al., 2003; Mauch & Sadler, 2007; Loveday et al., 2012; Best & Heckman, 2012; Bowler et al., 2015; Bouwens et al., 2015; Novak et al., 2017) as well as its evolution. This traces the distribution of the number density of a given population of sources for different luminosities. In the radio, this has been used to trace how Star Forming Galaxies (SFGs) evolve compared to Active Galactic Nuclei (AGN) (e.g. Mauch & Sadler, 2007; Prescott et al., 2016), as well as how different AGN populations e.g. High/Low Excitation Radio Galaxies (H/LERGs) (see Best & Heckman, 2012; Pracy et al., 2016; Prescott et al., 2016; Williams et al., 2018) evolve. This not only improves understanding of which populations dominate at different epochs in the Universe's history, but also how quickly these populations may evolve. This allows investigations into whether the evolution is driven by pure luminosity evolution (PLE) or pure density evolution (PDE) (see e.g. Pracy et al.,

2016; Prescott et al., 2016; Smolčić et al., 2017).

However, to truly understand why galaxies are evolving, it is necessary to consider other factors that may be affecting or driving this evolution. One such factor that is often considered is the relationship between galaxies and the dark matter haloes that they inhabit. This gives information on the environment in which these galaxy populations reside and this can be investigated through the clustering of galaxies (see e.g. Lindsay et al., 2014a,b; Hatfield et al., 2016; Magliocchetti et al., 2017; Hale et al., 2018), as investigated in Chapter 2.

In this Chapter, I present work investigating whether it is possible to constrain the Conditional Luminosity Function (CLF) of AGN and SFGs in the VLA 3 GHz COSMOS Survey Smolčić et al. (2017a,b). The CLF (see e.g. Yang et al., 2003; van den Bosch et al., 2004, 2013; Ballantyne, 2017) describes the distribution of sources with luminosity, within a dark matter halo of a given mass. This builds on the work of Chapter 2, where I investigated the clustering of AGN and SFGs in the COSMOS field to now investigate how this clustering relates to the observed luminosity function of these sources. This not only gives information on the number of sources expected at a given luminosity within a halo mass, but also can be used to investigate the typical properties of galaxies that reside within different dark matter haloes.

3.1 The Conditional Luminosity Function

Given by the notation, $\Phi(L|M_h)$, the Conditional Luminosity Function (CLF) is one way in which the underlying distribution of galaxies can be understood by relating the distribution of sources to their luminosities (L) and dark matter halo masses (M_h) (see Yang et al., 2003; van den Bosch et al., 2004; Yang et al., 2008; van den Bosch et al., 2013; Ballantyne, 2017; Trevisan & Mamon, 2017). It describes the luminosity distribution for a population of sources within a given halo mass and can be related

to both the observed luminosity function of sources, as well as their clustering. By using these two observational diagnostics it is therefore possible to disentangle how the dark matter environments and observed luminosity distributions are connected to one another.

To determine the CLF it is crucial to understand the relation between $\Phi(L|M_h)$ and the observed Luminosity Function, $\Phi(L)$, for a population of sources. $\Phi(L)$ describes the number density of sources within a given luminosity bin (typically in $\Delta \log_{10}(L)$) and has been investigated for both AGN and SFGs in many studies (e.g. Mauch & Sadler, 2007; Smolčić et al., 2009; Padovani et al., 2011; McAlpine et al., 2013; Prescott et al., 2016; Smolčić et al., 2017). As this work relies on luminosities at a given frequency, I will specifically use notation in terms of spectral luminosities (L_ν) at all times in order to be clear. Within this notation, the luminosity function can be described by:

$$\begin{aligned} \Phi(L_\nu/\Delta \log_{10}(L_\nu)^{-1}\mathrm{Mpc}^{-3}) &= \frac{A_{sky}}{A_{obs}} \frac{1}{\Delta \log_{10}(L_\nu)} \sum_j \frac{1}{V_{max,j}} \\ \Phi(L_\nu/\mathrm{mag}^{-1}\mathrm{Mpc}^{-3}) &= \frac{2}{5} \frac{A_{sky}}{A_{obs}} \frac{1}{\Delta \log_{10}(L_\nu)} \sum_j \frac{1}{V_{max,j}}, \end{aligned} \tag{3.1}$$

where A_{sky} is the total sky area in deg^2; A_{obs} the sky area of the observations in deg^2; $\Delta log_{10}(L)$ is the width of the luminosity bin. The $\frac{2}{5}$ factor in the second equation originates from the conversion between luminosities and magnitudes. $V_{max,j}$ is the maximum volume that a source could be observed within. For uniform detection across the image, it is calculated as $V(z_{max})$-$V(z_{min})$, where z_{min} is the minimum redshift of the redshift range being considered for $\Phi(L_\nu)$. z_{max}, on the other hand, is the maximum redshift a source could be observed to and remain above the detection limit. If the source could still be observed at redshifts greater than the maximum redshift over which $\Phi(L_\nu)$ is being determined, then z_{max} is given by this maximum redshift (i.e. the upper end of the redshift bin) for the $\Phi(L_\nu)$ being calculated.

$\Phi(L_\nu)$ can be related to the CLF through defining the luminosity function as the

integral of the CLF over all halo masses, taking into account the halo mass function ($n(M_h)$; e.g. Press & Schechter, 1974; Sheth et al., 2001; Jenkins et al., 2001). $n(M_h)$ is the number density of haloes of a given halo mass and this can be expressed as:

$$\Phi(L_\nu) = \int \Phi(L_\nu|M_h)\ n(M_h)\ dM_h. \tag{3.2}$$

Therefore $\Phi(L_\nu|M_h)$ represents the number of sources with a luminosity $L_{1.4\ \mathrm{GHz}}$ in a given $\Delta \log_{10}(L_{1.4\ \mathrm{GHz}})$ bin as a function of halo mass. Following Ballantyne (2017), the average number of sources in a luminosity range is given by[1]:

$$\langle N(M_h)\rangle = \int_{\log_{10}(L_{\nu,1})}^{\log_{10}(L_{\nu,2})} \Phi(L_\nu|M_h) d\log_{10}(L_\nu), \tag{3.3}$$

and, again following the work of Ballantyne (2017), the average luminosity of a population of sources as a function of halo mass can also be found using:

$$\langle L_\nu(M_h)\rangle = \frac{\int L_\nu \Phi(L_\nu|M_h) d\log_{10}(L_\nu)}{\int \Phi(L_\nu|M_h) d\log_{10}(L_\nu)}, \tag{3.4}$$

and the average mass of a population of sources is then:

$$\langle M\rangle = \frac{\int M_h \langle N(M_h)\rangle\ n(M_h)\ dM_h}{\int \langle N(M_h)\rangle\ n(M_h)\ dM_h}. \tag{3.5}$$

However, the luminosity function alone is not sufficient to constrain the CLF as there are many formulations of $\Phi(L_\nu|M_h)$ that could reproduce $\Phi(L_\nu)$. As such, another constraint on the CLF is essential. One such constraint comes from studying the underlying dark matter haloes that a source population reside in, through investigating the clustering of galaxies, as described in Chapter 2. Following Ballantyne (2017), by quantifying how the clustering, and hence the bias (b), varies with luminosity for a galaxy population the CLF can be related to $b(L_\nu)$:

$$b(L_\nu) = \frac{\int \Phi(L_\nu|M_h)\langle N(M_h)\rangle\ b(M_h)\ n(M_h)\ dM_h}{\int \langle N(M_h)\rangle \Phi(L_\nu|M_h)\ n(M_h)\ dM_h}. \tag{3.6}$$

[1]The integral used is here is slightly different to Ballantyne (2017) in the sense that it integrates over $d\log_{10}(L_\nu)$ (not dL_ν) due to the definition of $\Phi(L_\nu)$ which is normalised by $\Delta \log_{10}(L_{1.4\ \mathrm{GHz}})$.

This takes into account the relationship between bias and halo mass, $b(M_h)$, through relations such as those defined by Mo & White (1996); Sheth et al. (2001); Tinker et al. (2010).

The CLF has been described and investigated previously with many galaxy samples, often selected at optical wavelengths (e.g. Yang et al., 2003; van den Bosch et al., 2004, 2007; Yang et al., 2008; van den Bosch et al., 2013; Trevisan & Mamon, 2017). In one of the more recent works, Ballantyne (2017) instead made use of X-ray observations of AGN to parameterise and investigate the evolution of the CLF from $z \sim 0$ to $z \sim 0.9$. From this, they investigated how properties such as AGN lifetime, bias and halo occupation number change between $z \sim 0$ to $z \sim 0.9$. However, the CLF has never been investigated for radio sources. Radio surveys though are now able to produce large area, deep, high resolution surveys, observing thousands of galaxies (see e.g. Smolčić et al., 2017a; Shimwell et al., 2019), opening the possibility to jointly measure $\Phi(L_\nu)$ and large scale clustering ($b(L_\nu)$) to constrain the CLF.

Whilst measurements of $\Phi(L_\nu)$ and $b(L_\nu)$ are able to help disentangle measurements of $\Phi(L_\nu|M_h)$, these are not able to directly probe the CLF and rely on integrals and so accurate knowledge of $n(M_h)$ or $b(M_h)$. Therefore if it were possible to obtain direct measurements for a large sample of AGN or SFGs, then it would be possible to directly measure the CLF. This could be from estimating halo masses using X-ray observations (see e.g. the work of Pratt et al., 2009; Mantz et al., 2010) however may not be possible for all radio sources.

3.1.1 Why Study the CLF?

There are many reasons why studying and quantifying the CLF is important. Firstly, as mentioned previously, the CLF links two important observational properties that are typically investigated when studying galaxy populations: their luminosity distribution and their clustering. This in turns gives us greater understanding of the

environments that galaxy populations reside in (through their clustering) and how this may influence the observed power that these sources emit.

Understanding this relationship will not only provide greater understanding about how different populations evolve in relation to their surroundings, but can have important implications for investigating and understanding the amount of energy injected into the environment by some AGN and how this is dependent on the environment that these sources reside in. There exist relationships between the luminosities of FRI and FRII sources and their jet power in both observations and simulations (see e.g Willott et al., 1999; Bîrzan et al., 2008; Cavagnolo et al., 2010; Hardcastle & Krause, 2014; Croston et al., 2018; Hardcastle, 2018). For FRII sources, Willott et al. (1999) determined one such relation, using arguments about the emission from jets (not observational data). As HERGs are typically FRII sources, it is therefore likely that these radiative mode sources follow the relationship from Willott et al. (1999). For cluster FRI sources, on the other hand, these have been found to follow relations from Bîrzan et al. (2008), who combined radio and X-ray measurements to estimate jet power. Halo mass is a proxy for the environment that sources reside in, with higher mass halos typically corresponding to denser, more clustered environments. Therefore, by studying the relationship between luminosity and halo mass, it is possible to investigate the jet power injected into the environment as a function of environmental properties. This is important for understanding AGN feedback (see Section 1.3) and how the energy that is impacting the region around the sources changes both with redshift and environment. This will be critical in fine tuning feedback models in cosmological simulations (e.g Sijacki et al., 2007).

3.2 Data

3.2.1 Radio Data

In order to study the CLF and its evolution, I made use of the VLA 3 GHz COSMOS Survey (Smolčić et al., 2017a,b). This is the radio survey used and described in Chapter 2, which are some of the deepest, current radio observations over a large field that allows the large scale clustering to be well constrained. Not only this, but the multi-wavelength counterparts determined in Smolčić et al. (2017b) provided classifications on whether a source is determined to be a SFG or an AGN. This therefore allows investigations of how the CLF is dependent on source type. This is also essential due to the differences in bias for the two populations of sources, as described in Chapter 2.

Contrary to Chapter 2, the 5σ (not 5.5σ) is used for this work. This is because it is essential that a large number of sources are available in order to have enough sources in each redshift and luminosity bin to be able to investigate $b(L_\nu)$. Whilst this may introduce some false detections (as described in Chapter 2), these are only a small fraction and it is the 5σ catalogue that has been used for studies of the luminosity function in Novak et al. (2017); Smolčić et al. (2017). However, I continued to use the clean SFG sample, despite the lower number of sources. Not using the clean sample meant that when the luminosity function for SFGs (at $z : 0.3 - 0.8$) was calculated, a high luminosity tail appeared on the luminosity function which appeared anomalous. Although the results from this redshift range will not be presented here, this therefore confirmed that it was best to use the clean SFG sample to avoid any contamination.

3.2.2 Near-IR Data for the Cross Correlation

To investigate clustering as a function of luminosity to determine $b(L_\nu)$, the number of sources in the full AGN or SFG population will be small in each luminosity bin.

This is because of the number density of radio observations (compared to other wavelengths) and the CLF needs to be calculated within small redshift bins. As such, the uncertainties on $\omega(\theta)$ would be larger than in Chapter 2. Therefore I determined $b(L_\nu)$ using the angular cross-correlation function instead of the auto-correlation. This technique has been used previously (Guo et al., 2011; Hartley et al., 2013; Lindsay et al., 2014a) where the numbers of sources are too small to determine the auto correlations. Near-IR data from the COSMOS 2015 Survey (Laigle et al., 2016) is used in order to calculate the cross-correlation function, defined as:

$$\omega_{cross}(\theta) = \frac{\overline{D_1 D_2}(\theta)}{\overline{R_1 R_2}(\theta)} - \frac{\overline{D_1 R_2}(\theta)}{\overline{R_1 R_2}(\theta)} - \frac{\overline{D_2 R_1}(\theta)}{\overline{R_1 R_2}(\theta)} + 1 \tag{3.7}$$

where the subscript 1 and 2 refer to the two catalogues. Here 1 refers to the radio catalogue and 2 the near-IR one.

By calculating the cross-correlation instead of the auto-correlation (as in Chapter 2), the derivation of the bias will now be different as the two datasets have been combined. In Chapter 2 the r_0 was calculated using Equation 1.15. To determine r_0 in terms of what is observed for the cross correlation $N(z)$ is now a combination of the two redshift distributions used. Therefore, in the calculation of $r_{0,CC}$ (where CC stands for cross-correlation), $N(z)$ is taken to be the combined normalised redshift distribution of the two data catalogues (radio and near-IR) following Lindsay et al. (2014a). b^2_{CC} can then be calculated in the same way as in Equation 1.14. By also calculating the auto-correlation for the near-IR COSMOS 2015 data, the bias for the radio data ($b_{AC,1}$) can be calculated using:

$$b^2_{CC}(L_\nu) = b_{AC,1}(L_\nu) \times b_{AC,2}. \tag{3.8}$$

$b_{AC,2}$ can be calculated directly from the auto-correlation of the near-IR data from Laigle et al. (2016), within a given redshift bin. By calculating the bias from the cross correlation of the radio sources in the same redshift bin to the near-IR data ($b^2_{CC}(L_\nu)$)

as a function of luminosity, it is possible to determine the relationship between bias and radio luminosity ($b_{AC,1}(L_\nu)$).

3.3 Calculating the CLF

In order to determine the CLF, $\Phi(L_\nu)$ and $b(L_\nu)$ need to be measured. The steps in order to achieve this are described below.

3.3.1 Mask Sources

Both the radio and near-IR catalogues need to be flagged in order to remove regions that are masked as part of the COSMOS 2015 catalogue (Laigle et al., 2016). The mask used is shown in Figure 3.1 and was applied using the package `pyregion`[2]. The masked data from the COSMOS 2015 catalogue can be seen in Figure 3.1. This is the same as used and seen in Figure 2.1 (in Chapter 2), but the effect of applying it is much more noticeable when considered over the near-IR data.

The radio rms image of the field was also masked in the same way as the catalogues. All rms values across the image that have not been masked are recorded. This is essential to account for the detectability of a source across the image as this will affect the measured luminosity function (see Section 3.3.2).

3.3.2 Calculating the 1.4 GHz Luminosity Function

Next, radio luminosities need to be determined to be able to construct $\Phi(L_\nu)$. For this work, I made calculations of $\Phi(L_\nu)$, at a frequency of 1.4 GHz. Calculating the luminosity function at 1.4 GHz is typically investigated in order to make comparisons with other work. 1.4 GHz was also used to study the evolution of $\Phi(L_\nu)$ for these observations in Novak et al. (2017); Smolčić et al. (2017). Using Equation 1.23, I

[2] `https://pyregion.readthedocs.io/en/latest/index.html`

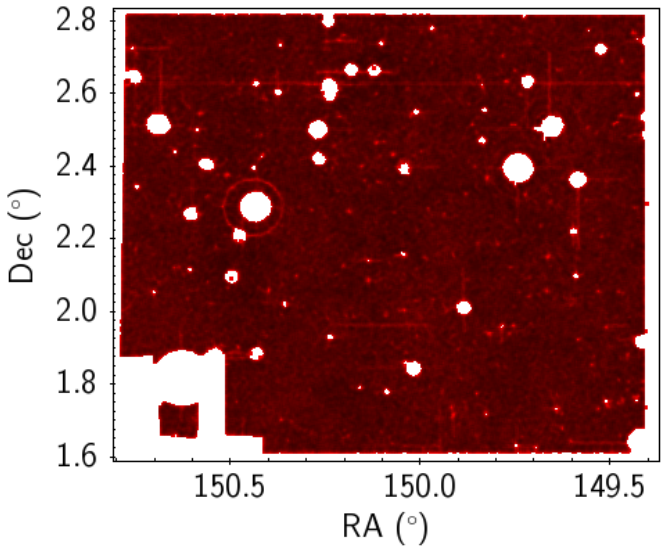
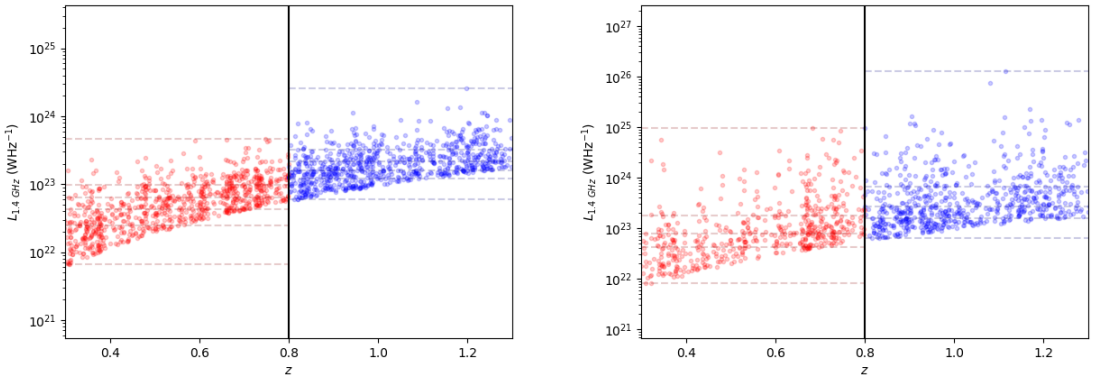


Figure 3.1: The near-IR sources and mask from (Laigle et al., 2016) that are necessary to determine the cross correlation. This illustrates the regions used for the radio and near-IR sources post masking.

calculated the radio luminosity at 1.4 GHz. The luminosities of these sources as a function of redshift are shown in Figure 3.2. Once these luminosities have been calculated, the number of sources from the near-IR and radio catalogues (that are not masked) are further reduced, by only including those sources that are within the specified redshift range being considered.

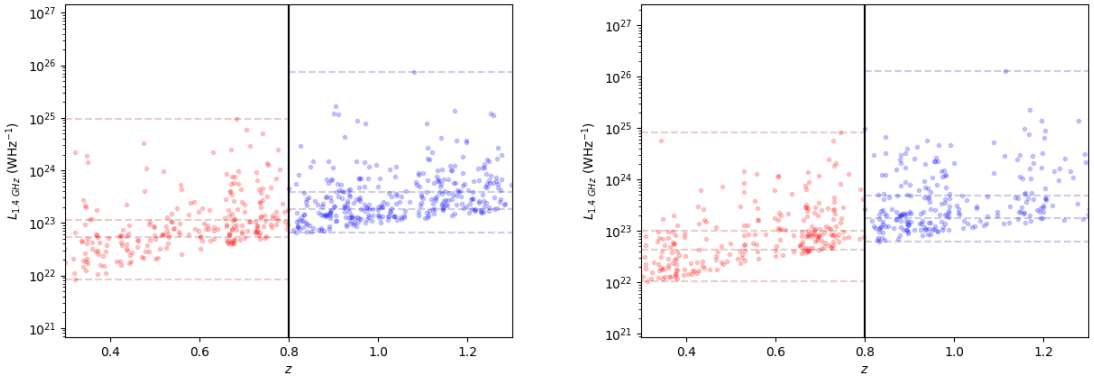
Next the 1.4 GHz luminosity function needs to be determined. For images in which there is a constant uniform noise across the field of view then the radio sources will be uniformly detected across the image. As such, the luminosity function can be calculated using Equation 3.1 (e.g. Page & Carrera, 2000). However, the detection of sources across the radio image used here from Smolčić et al. (2017a) is non-uniform across the image. Therefore the $V_{max,j}$ term in Equation 3.1 must be calculated differently. Instead, the method of Novak et al. (2017) is used to determine the luminosity function, and is calculated using:

$$\Phi(L/\Delta\log_{10}(L_\nu)^{-1}\mathrm{Mpc}^{-3}) = \frac{A_{sky}}{A_{obs}}\frac{1}{\Delta\log_{10}(L_\nu)}\sum_{\Delta\log_{10}(L_\nu)_j}\sum_{\delta z_i}\frac{1}{V_{max,ij}} * A_{Corr,ij}. \quad (3.9)$$



(a) SFGs

(b) AGN



(c) HLAGN

(d) MLAGN

Figure 3.2: Distribution of $L_{1.4\mathrm{GHz}}$ as a function of z for (a) the SFG sample, (b) AGN, (c) HLAGN and (d) MLAGN. The black vertical lines represents the distinction between the redshift bins being considered for this work, whilst the horizontal dashed lines indicate the luminosity bins used to calculate $b(L_\nu)$.

For each source, a correction factor needs to be calculated in order to account for the detectability of sources of a given flux density across the image. Smolčić et al. (2017a) calculated the completeness across the image by running Monte Carlo simulations to determine the recovery rate of simulated sources across the field of view. However, as the image being used here is masked, Table 2 of Smolčić et al. (2017a) is not correct for this work and needs to be adapted. Therefore, A_{Corr} is determined by adapting the completeness correction from Smolčić et al. (2017a) to account for what fraction of the image a source can be detected, given its peak flux in the masked image, compared to the unmasked image.

These corrections can be seen in Figure 3.3, where the dark blue line is the completeness values from Table 2 of Smolčić et al. (2017a). The black line corresponds to the detection fraction at each flux density in the unmasked image. This is determined by calculating the fraction of area (across the total image) where this flux density is greater than 5σ. The grey line is similar to the black line, but is now for the masked VLA image. At low flux densities, the completeness fraction will be higher in the masked image as this is the centre of the image and so will not suffer from the noisier regions that the full unmasked image has. The completeness fraction used here is calculated as the completeness value from Smolčić et al. (2017a) multiplied by the ratio of the grey line to the black line, this is shown as the light blue line in Figure 3.3.

Using this, the correction factor A_{Corr} can be determined for each source, j, at a given redshift value, z_i. As the peak flux is used to calculate the flux limits, this makes the assumption that angular sizes do not change with redshift. This is only true for unresolved sources ($\sim 60\%$ of sources in the 5σ catalogues). However, when I compared $\Phi(L_\nu)$ generated using the peak fluxes to that from total fluxes, the differences in the luminosity functions were negligible and so this assumption should not affect the results significantly.

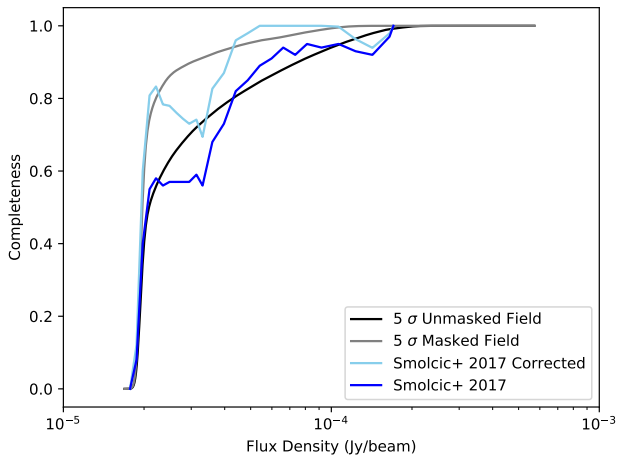


Figure 3.3: Correction factors used here for constructing the luminosity function of sources (light blue line) compared to those found in Smolčić et al. (2017a) for their completeness simulations (dark blue). The black line shows the detection fraction of sources, as a function of flux density, for the unmasked image and the grey line shows the corresponding detection using the masked image.

In Equation 3.9, for each source within a luminosity bin (j), $V_{max,ij} * A_{Corr,ij}$ is found by going through redshift bins (i) of width δz and at each redshift calculating the redshifted peak flux of the source and the detectability correction fraction. By multiplying this correction factor by the difference in volume at the $z_i + \delta z$ to z_i and summing across all redshift bins, I calculate the maximum volume a source could be observed out to, taking into account how detectable sources are across the image.

However, when $\Phi(L_\nu)$ was determined by this method, some of the low luminosity bins appeared to suffer from incompleteness issues. This suggests that A_{corr} at the lowest fluxes does not adequately reflect detection across the image. To be conservative, I therefore set a $\sim$22.5 μJy flux density limit to mitigate this issue.

3.3.3 Redshift Distributions and Luminosity Bins

The redshift distributions for both the near-IR sources and the radio sources need to be determined in order to calculate the clustering length, r_0, and bias for a population

of sources. This is calculated and modelled for both the near-IR data and the radio data over all redshifts and then the redshift range is restricted when calculating $r_0(L_\nu)$. By modelling $N(z)$ over the entire redshift range of all sources, this ensures that the $N(z)$ distribution is not being influenced by the natural sub-structure in redshifts over this range and is instead smoothed out to reflect the distribution as a whole. For the radio data, the luminosity range is split into 3-5 bins (depending on the number of sources available) and the limits of the luminosity bins are calculated so that an even number of sources are contained within each bin. The redshifts for the sources in each luminosity bin are then modelled assuming a smooth functional form, as described in Section 2.2.3.

3.3.4 Random Catalogues

In order to investigate the angular clustering as a function of luminosity, it is necessary to also generate random catalogues to determine $\omega(\theta)$. Therefore for each radio data set that has been split into the redshift range and then by luminosity, a corresponding random catalogue is also generated. This is calculated in the same way as in Chapter 2, however extra conditions are also imposed to ensure that: each random source is not within the masked region; the random sources are within the designated redshift range and luminosity bin and finally that it is of the correct source type for the investigation (i.e. SFG or AGN, as in Chapter 2). Due to the small numbers of sources involved, it is also essential to ensure that $\omega(\theta)$ is not being dominated by the positions of the random catalogue. Therefore the number of randoms generated is either taken as $20\times N_{data}$ or 10,000 sources, whichever is largest.

For the associated near-IR randoms, as there is no significant variation in detection sensitivity across the near-IR image, the associated random catalogue is simply generated by selecting random positions and ensuring the source is not located within a masked region. The near-IR data random catalogue was also generated to have $20\times$

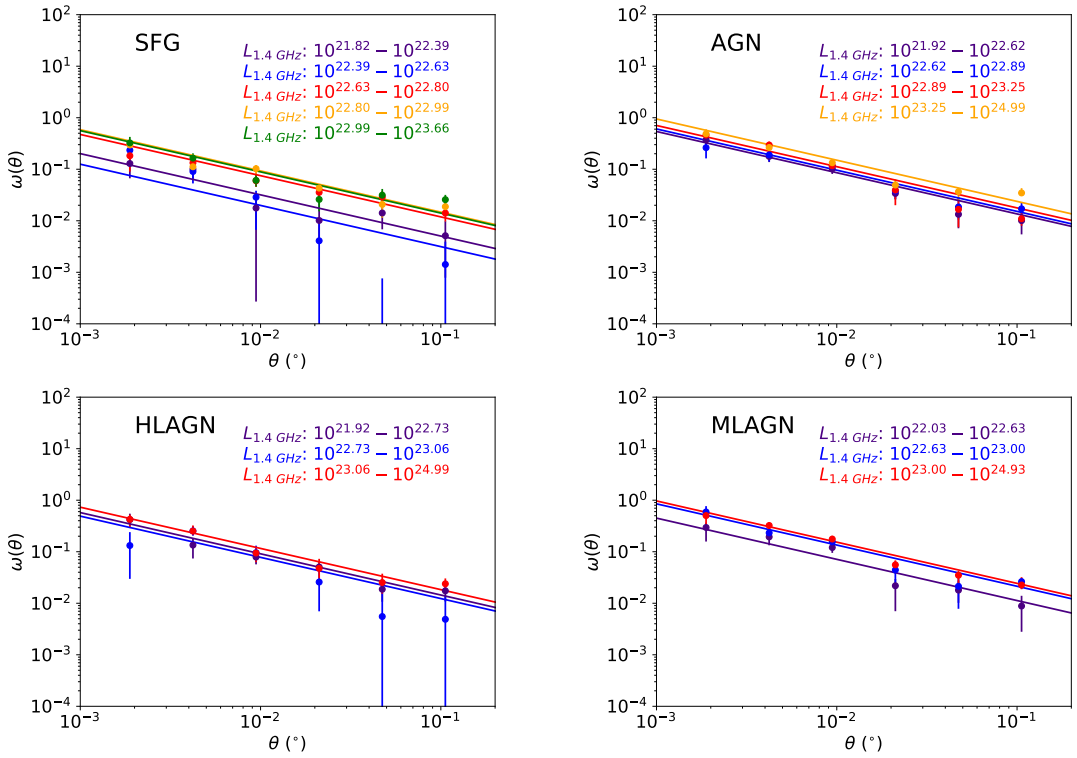
more sources than in the near-IR catalogue (within the z range).

3.3.5 Auto-Correlation and Cross-Correlation Functions

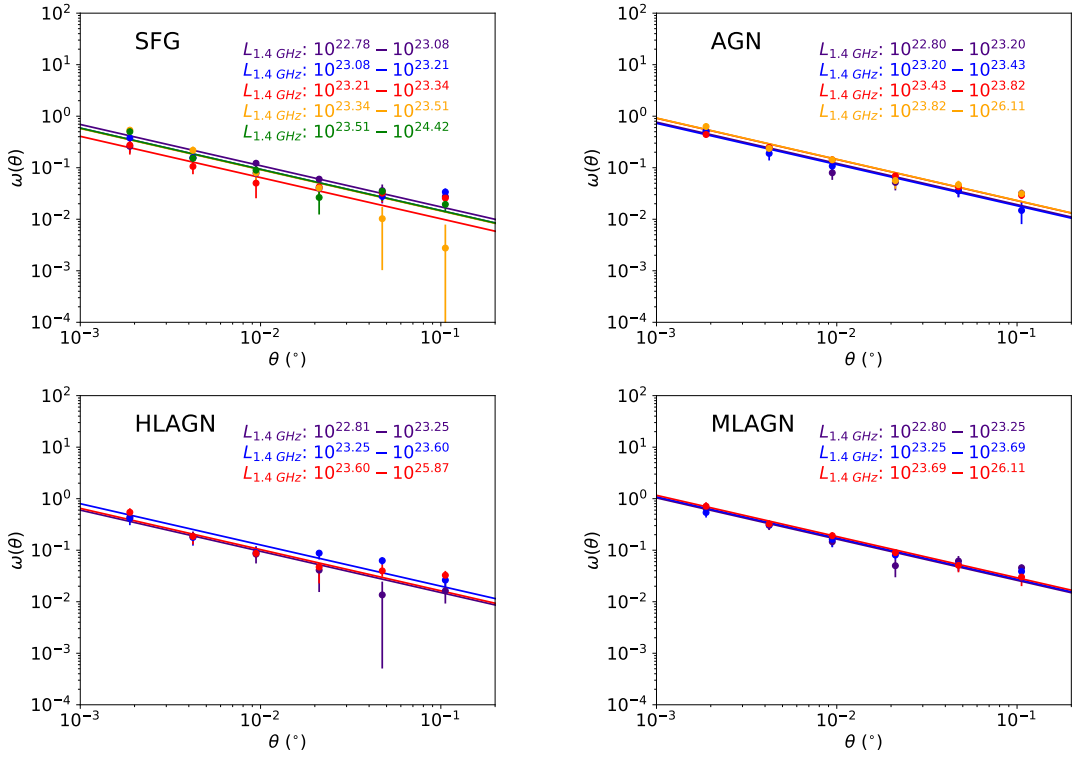
To determine $b(L_\nu)$, the auto-correlation of the near-IR data first needs to be measured. Due to the large number of sources in the near-IR data, the package TreeCorr (Jarvis, 2015) was used to calculate the numbers of pairs of galaxies within different angular separations. Next, the cross-correlation functions were calculated within each luminosity bin using Equation 3.7. This takes into account both the data and random catalogues of the radio and near-IR data. These are fit in the same way as described in Section 2.2.2, where I fix the slope to be 0.8 and account for the integral constraint (see Section 2.2.2; Groth & Peebles, 1977) and are fit over the range $10^{-3} - 0.2°$, in 6 angular bins that were spaced evenly logaritmically. The results of fitting $\omega(\theta)$ for all the luminosity bins used here are shown in Figure 3.4. Whilst some of the observed TPCF seen in Figure 3.4 may appear to have slopes steeper than 0.8, as the numbers of radio source in some luminosity bins are small, I felt it would be appropriate to fix the slope. This should be improved upon in future work when more sources are available. I also note that the measuring of clustering here is limited by the small number of sources which limits these results. This is because cosmic variance cannot be well accounted for in such a small field with a small number of sources.

3.3.6 The bias: $b^2_{CC}(L_\nu)$, $b_{AC,2}$ and $b_{AC,1}(L_\nu)$

Next, the evolution of bias as a function of luminosity, $b(L_\nu)$, needs to be determined in order to be able to constrain the CLF. To do this, first the bias for the cross-correlation, $b^2_{CC}(L)$, is determined for each luminosity bin, as described in Section 3.2.2. Secondly, the bias for the auto-correlation function from the near-IR data also needs to be calculated, $b_{AC,2}$. Using this, and Equation 3.8, it is then possible to calculate the bias as a function of luminosity for the radio data, $b_{AC,1}(L_\nu)$.



(a) $0.3 \leq z < 0.8$



(b) $0.8 \leq z < 1.3$

Figure 3.4: The observed $\omega(\theta)$ for SFGs, AGN, HLAGN and MLAGN for the different luminosity bins used in this work for (a) $0.3 \leq z < 0.8$ and (b) $0.8 \leq z < 1.3$.

3.3.7 Fitting the CLF

Once $\Phi(L_\nu)$ and $b(L_\nu)$ have been measured, it is possible to constrain the CLF using Equations 3.2 and 3.6. To do this, I adopted the functional form from Ballantyne (2017) and determined the variation of $f(L_\nu)$ based on the forms of the observed luminosity functions from Mauch & Sadler (2007). This parameterisation is described in Equations 3.10, 3.11, 3.12 and 3.13.

$$\Phi(L_\nu|M_h) = \left(\frac{M_h}{M_*}\right)^a e^{-M_h/M_{cut}} \; f(L_\nu), \tag{3.10}$$

$$M_{cut} = \left(\frac{L_\nu}{L_{\nu,*}}\right)^c M_N, \tag{3.11}$$

$$f(L) = \frac{1}{\left(\frac{L_\nu}{L_{\nu,*}}\right)^{-b} + \left(\frac{L_\nu}{L_{\nu,*}}\right)^{-d}} \qquad \text{(AGN)}, \tag{3.12}$$

$$f(L) = \left(\frac{L_\nu}{L_{\nu,*}}\right)^{1-b} \exp\left(-\frac{1}{2}\left(\frac{\log_{10}\left(1+\frac{L_\nu}{L_{\nu,*}}\right)^2}{d}\right)\right) \qquad \text{(SFG)}. \tag{3.13}$$

In these Equations, a describes the power law index for the power law distribution of halo masses. The reference halo mass for this power law distribution of halo masses is described by M_*. However, this power law distribution of halo mass needs to be cut off at the high masses as it cannot continue to grow indefinitely (see Figures 3.5 and 3.6). This is quantified by M_{cut} function, which is dependent both on a mass term M_N as well as luminosity. The luminosity dependence is essential to allow $b(L_\nu)$ to evolve with luminosity. The luminosity dependence of the exponential cut-off in haloes is governed by the power law index c. Finally L_* defines a characteristic luminosity scale which for AGN defines the bounding luminosity between the two power law regimes in the luminosity function. For AGN, b and d define the power law slopes at low and high luminosities, whilst for SFGs, b relates to the power law distribution index

of luminosities, with d quantifying a characteristic width of the exponential cut off in luminosity. The effect of varying these parameters can be seen in Figure 3.5 (for SFGs) and Figure 3.6 (for AGN).

To fit the CLF, given this parameterised form, the package `emcee` (Foreman-Mackey et al., 2013) was used. A uniform prior range for each of these parameters needs to be defined as well as a likelihood function. The prior range used for each of the parameters is given in Table 3.1. For AGN, both indices of the luminosity function need to be negative, so b and d were defined to be < 0. I also ensured $|d| > |b|$ to avoid degeneracy. For SFGs on the other hand, d needs to be positive so that the entire term within the exponent is negative. For both AGN and SFGs, c cannot be zero, as $b(L_\nu)$ should evolve and, from Ballantyne (2017), I expect this to be positive. The two halo mass parameters (M_* and M_N) are given in the range $10^{10} - 10^{16} M_\odot$ and are sampled uniformly in logarithmic space. Halo masses above this are not expected to be frequent in the Universe (see the models used in Murray et al., 2013). L_* was able to vary uniformly in logarithmic space between $10^{20} - 10^{25}$ WHz^{-1} for SFGs and $10^{23} - 10^{27}$ WHz^{-1} for AGN. For the likelihood function, a minimisation of χ^2 was used and so the log likelihood function is defined by:

$$\log(\mathcal{L}) = -\frac{1}{2}\left[\sum_i \left(\frac{\Phi_{CLF}(L_i) - \Phi(L_i)}{\sigma_{\Phi(L_i)}}\right)^2 + \sum_j \left(\frac{b_{CLF}(L_j) - b(L_j)}{\sigma_{b(L_j)}}\right)^2\right] \tag{3.14}$$

The 7 parameters of the model for the CLF were fit using 100 walkers and a chain length of 100,000 steps. The $n(M_h)$ and $b(M_h)$ parameterisation that are necessary to constrain the CLF are taken from the package `Colossus` (Diemer, 2018), which was also used in Chapter 2. This assumed a halo mass function model from Tinker et al. (2008) and a bias model from Tinker et al. (2010). The assumed halo masses were taken as '200m', which corresponds to haloes such that they are 200$\times$ the mean matter density of the universe. Both $n(M_h)$ and $b(M_h)$ are evaluated at the median redshift of the radio sample that is being considered. $\sigma_{b(L_j)}$ was taken as the average

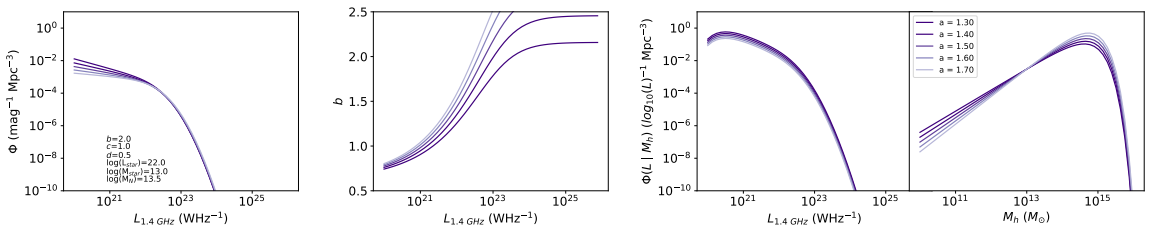
of the errors from the 16^{th} and 84^{th} percentiles.

Due to the small area of this field, and the number of SFGs and AGN needed to constrain the CLF, I chose redshift bins which gave approximately equal numbers of sources in both the low ($0.3 \leq z < 0.8$) and high ($0.8 \leq z < 1.3$) redshift bins for all of the source types. This is a larger bin width than is ideal, as bias is also known to evolve with redshift (see Chapter 2) and, as can be seen in Figure 3.2, this will mean there is evolution of bias within the bin. However for the work presented here, these larger bin widths are necessary due to the small number of sources.

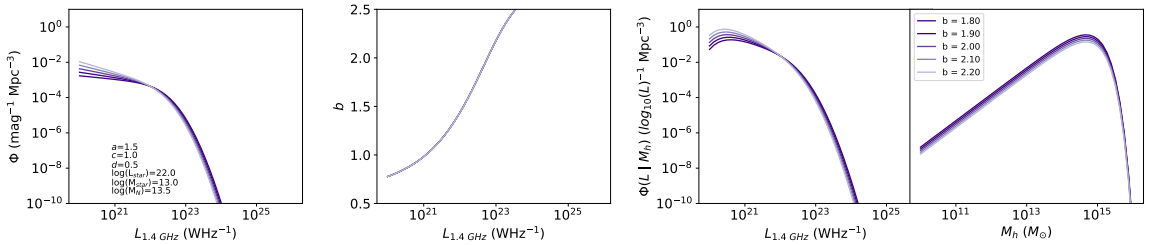
3.3.8 Results from the CLF Fitting

Finally, to investigate how successful the fitting of the CLF was, it is important to ensure that the CLF, using the parameters determined from the `emcee` fitting, can well describe $\Phi(L_\nu)$ and $b(L_\nu)$. To see how well the output chains from `emcee` recover $\Phi(L_\nu)$ and $b(L_\nu)$, the burn-in phase of the chains is first removed. This pre-burn-in phase represents the chains for which the parameters are still converging on best fit parameters from their beginning random distribution along the prior range. In this work, I remove the first 90% of chains. Whilst this is a high burn-in fraction, this aims to ensure that the parameters are near convergence. I also removed any of the chains for walkers that had remained stuck in a high $\log_{10}(\mathcal{L})$ position, since initialisation.

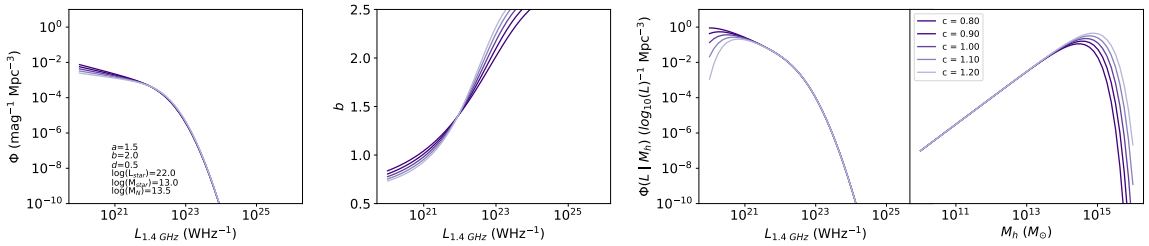
From these post burn-in chains, the recovered CLF fits are constructed by randomly selecting 1000 sub-chains from the post burn-in sample. A sample of sub-chains were used as reconstructing the results and calculating the average properties of the sources would be too time consuming if all chains had been used. Using random sub-samples, a distribution for $\Phi(L_\nu)$ and $b(L_\nu)$ (as well as the calculations of average luminosities etc.) are calculated through propagating the parameters for each sample through the equations defined in Section 3.1. Median values for these functions are then determined, as well as an estimate of the scatter from the 16^{th} and 84^{th}



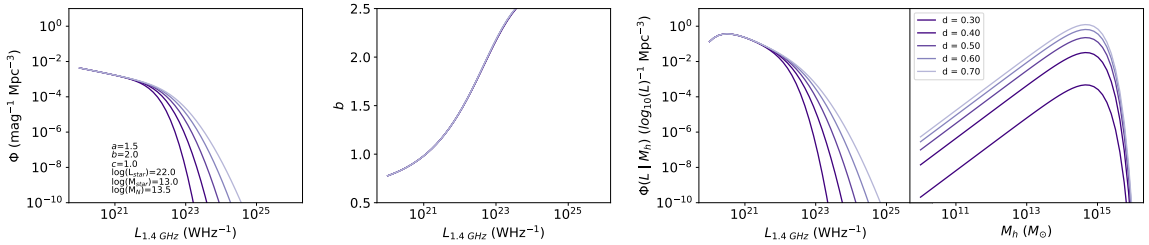
Varying a



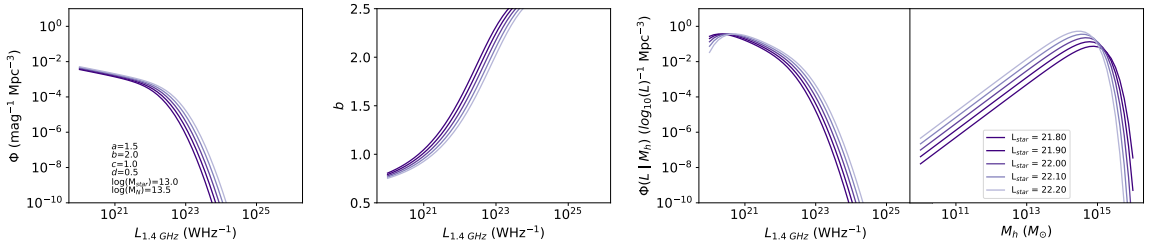
Varying b



Varying c

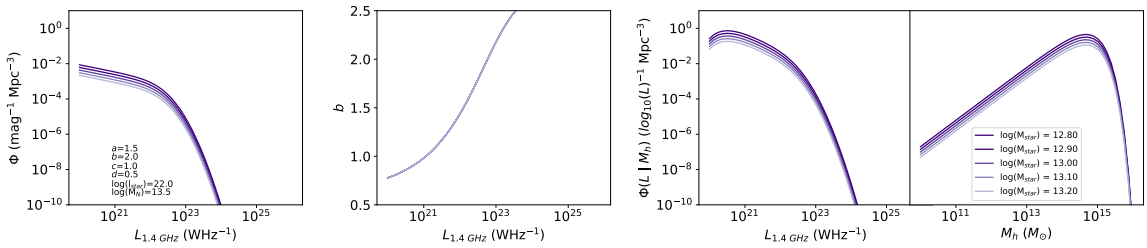


Varying d

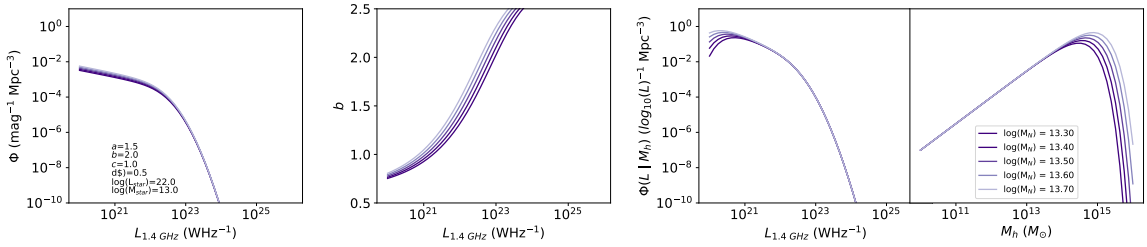


Varying L_*

Figure 3.5: Varying the parameters that model the CLF for SFGs. Each row represents a different parameter and for each row, the effect on $\Phi(L_\nu)$ (left), $b(L_\nu)$ (centre left) and $\Phi(L_\nu|M_h)$ (centre right - as a function of $L_{1.4\ \mathrm{GHz}}$; right - as a function of M_h) is shown. A halo mass of $M_h = 10^{12} M_\odot$ is assumed for panels 1-3 and $L_{1.4\ \mathrm{GHz}} = 10^{23}$ WHz^{-1} is assumed for the fourth panel.



Varying M_*



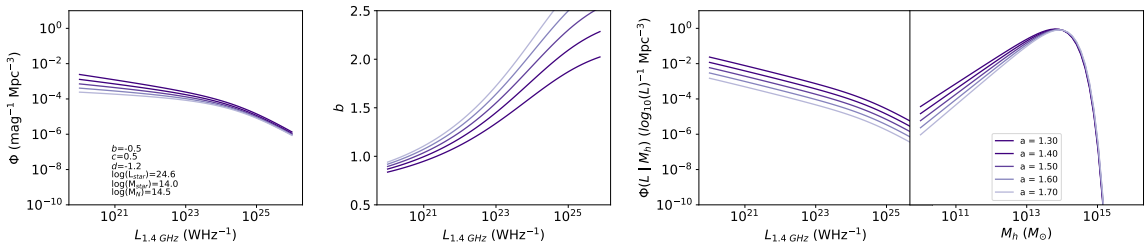
Varying M_N

Figure 3.5: Continued...Varying the parameters that model the CLF for SFGs. Each row represents a different parameter and for each row, the effect on $\Phi(L_\nu)$ (left), $b(L_\nu)$ (centre left) and $\Phi(L_\nu|M_h)$ (centre right - as a function of $L_{1.4\,\mathrm{GHz}}$; right - as a function of M_h) is shown. A halo mass of $M_h = 10^{12} M_\odot$ is assumed for panels 1-3 and $L_{1.4\,\mathrm{GHz}} = 10^{23}$ WHz^{-1} is assumed for the fourth panel.

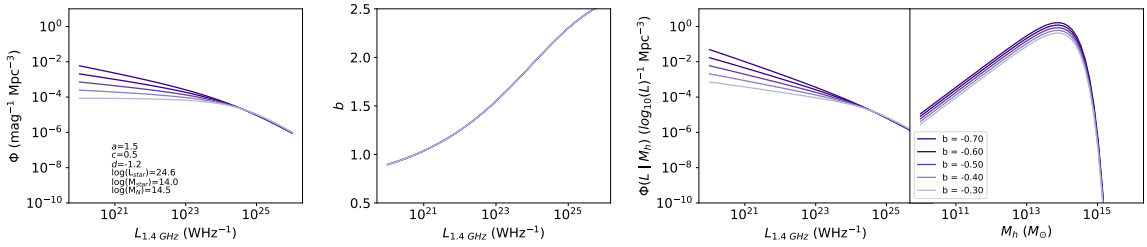
percentiles. Also recorded is the modal bin of the PDF distribution, as at times the median values may not give an accurate representation the modal value.

	a	b	c	d	$\log_{10}(\frac{L_*}{\mathrm{WHz}^{-1}})$	$\log_{10}(\frac{M_N}{M_\odot})$	$\log_{10}(\frac{M_*}{M_\odot})$
SFG	$[0,4]$	$[0,4]$	$[0,2]$	$[0,2]$	[20,25]	$[10,16]$	$[10,16]$
AGN	$[0,4]$	$[-2,0]$	$[0,2]$	$[-2,0]$	[23,27]	$[10,16]$	$[10,16]$
				$\lvert d\rvert > \lvert b\rvert$			

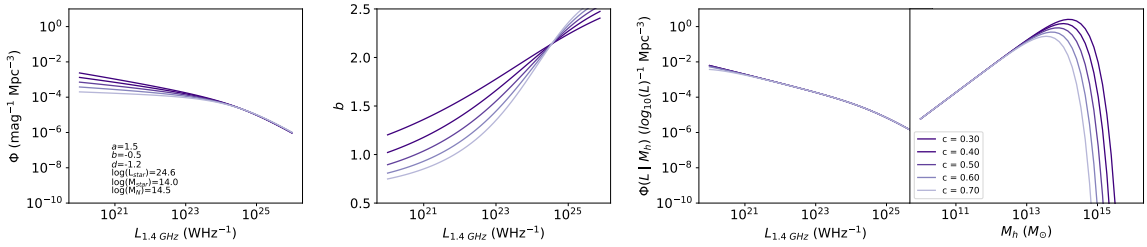
Table 3.1: Prior ranges used in `emcee` (Foreman-Mackey et al., 2013) for the 7-parameter fitting of the CLF model.



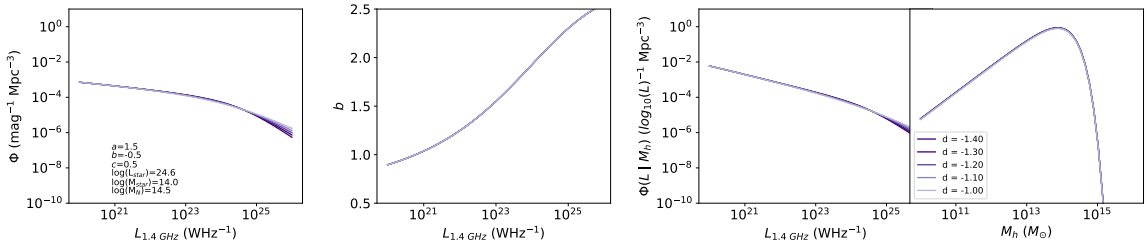
Varying a



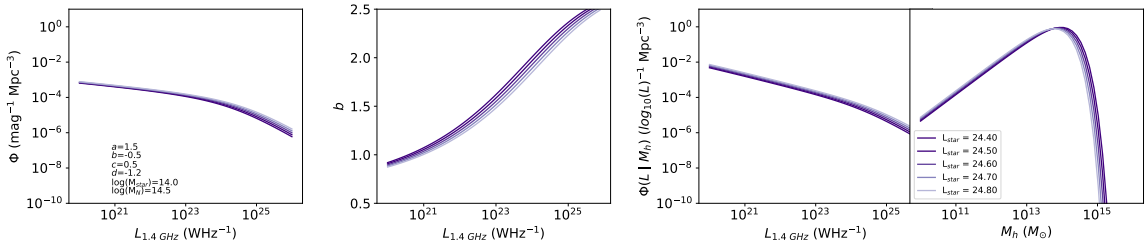
Varying b



Varying c

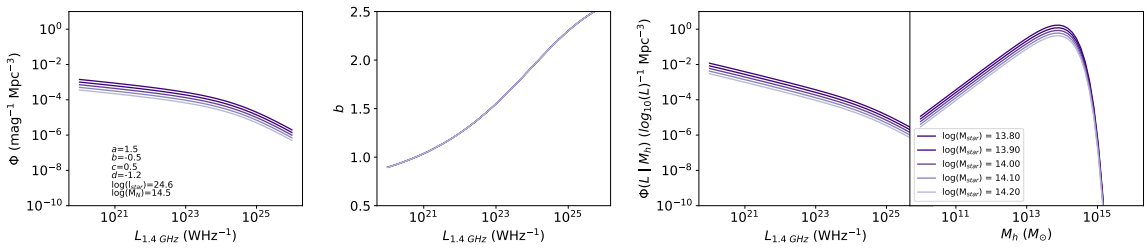


Varying d

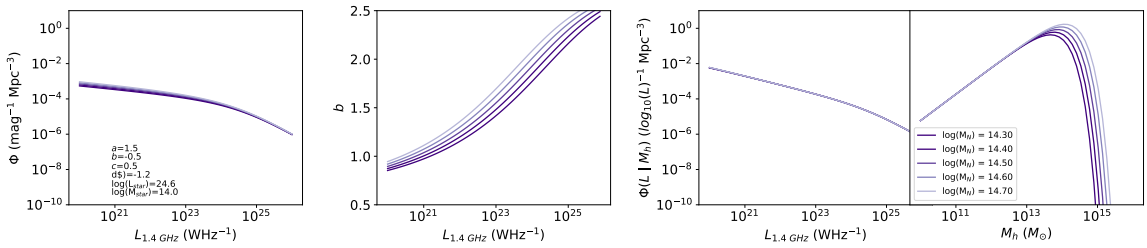


Varying $L*$

Figure 3.6: Varying the parameters that model the CLF for AGN. Each row represents a different parameter and for each row, the effect on $\Phi(L_\nu)$ (left), $b(L_\nu)$ (centre left) and $\Phi(L_\nu|M_h)$ (centre right - as a function of $L_{1.4\ \mathrm{GHz}}$; right - as a function of M_h) is shown. A halo mass of $M_h = 10^{12} M_\odot$ is assumed for panels 1-3 and $L_{1.4\ \mathrm{GHz}} = 10^{23}$ WHz^{-1} is assumed for the fourth panel.



Varying M_*



Varying M_N

Figure 3.6: Continued...Varying the parameters that model the CLF for AGN. Each row represents a different parameter and for each row, the effect on $\Phi(L_\nu)$ (left), $b(L_\nu)$ (centre left) and $\Phi(L_\nu|M_h)$ (centre right - as a function of $L_{1.4\ \mathrm{GHz}}$; right - as a function of M_h) is shown. A halo mass of $M_h = 10^{12} M_\odot$ is assumed for panels 1-3 and $L_{1.4\ \mathrm{GHz}} = 10^{23}$ WHz^{-1} is assumed for the fourth panel.

		Bin 1	**Bin 2**	**Bin 3**	**Bin 4**	**Bin 5**
SFG	$\log_{10}(L)$	21.82-22.39	22.39-22.63	22.63-22.80	22.80-22.99	22.99-23.66
	z_{med}	0.37	0.49	0.60	0.65	0.69
SFG	$\log_{10}(L)$	22.78-23.08	23.08-23.21	23.21-23.34	23.34-23.51	23.51-24.42
	z_{med}	0.37	0.49	0.60	0.65	0.69
AGN	$\log_{10}(L)$	21.92-22.62	22.62-22.89	22.89-23.25	23.25-24.99	
	z_{med}	0.41	0.61	0.67	0.67	
AGN	$\log_{10}(L)$	22.80-23.20	23.20-23.43	23.43-23.82	23.82-26.11	
	z_{med}	0.93	1.01	1.11	1.09	
HLAGN	$\log_{10}(L)$	21.92-22.73	22.73-23.06	23.06-24.99		
	z_{med}	0.44	0.63	0.68		
HLAGN	$\log_{10}(L)$	22.81-23.25	23.25-23.60	23.60-25.87		
	z_{med}	0.94	1.10	1.13		
MLAGN	$\log_{10}(L)$	22.03-22.63	22.63-23.00	23.00-24.93		
	z_{med}	0.40	0.66	0.67		
MLAGN	$\log_{10}(L)$	22.80-23.25	23.25-23.69	23.69-26.11		
	z_{med}	0.92	0.99	0.98		

Table 3.2: The luminosity bins used for each galaxy type in order to investigate $b(L)$ as well as the corresponding median redshifts for the sources in each bin.

Type	z	$\bar{z}$	$L_{1.4\mathrm{GHz}}$ (WHz^{-1})	N	a	b	c	d	$\log_{10}(\frac{L_*}{\mathrm{WHz}^{-1}})$	$\log_{10}(\frac{M_N}{M_\odot})$	$\log_{10}(\frac{M_*}{M_\odot})$
SFG	0.3-0.8	0.69	All	1061	$1.28^{+0.09}_{-0.05}$	$1.85^{+0.22}_{-0.30}$	$1.88^{+0.09}_{-0.18}$	$0.32^{+0.17}_{-0.14}$	$22.89^{+0.45}_{-0.59}$	$13.25^{+0.41}_{-0.46}$	$13.99^{+0.90}_{-1.11}$
					$1.27^{+0.05}_{-0.05}$	$1.80^{+0.11}_{-0.11}$	$1.96^{+0.04}_{-0.04}$	$0.27^{+0.03}_{-0.03}$	$22.96^{+0.11}_{-0.11}$	$13.22^{+0.10}_{-0.10}$	$14.20^{+0.20}_{-0.20}$
SFG	0.8-1.3	1.15	All	1052	$1.76^{+1.38}_{-0.72}$	$1.02^{+0.72}_{-0.65}$	$0.41^{+0.42}_{-0.26}$	$0.43^{+0.05}_{-0.07}$	$22.65^{+0.53}_{-0.44}$	$12.47^{+1.09}_{-0.62}$	$12.11^{+2.17}_{-0.69}$
					$1.10^{+0.10}_{-0.10}$	$0.93^{+0.13}_{-0.13}$	$0.33^{+0.07}_{-0.07}$	$0.38^{+0.06}_{-0.06}$	$22.50^{+0.11}_{-0.11}$	$11.83^{+0.14}_{-0.14}$	$11.81^{+0.20}_{-0.20}$
AGN	0.3-0.8	0.67	All	601	$1.42^{+0.41}_{-0.10}$	$-0.74^{+0.14}_{-0.19}$	$0.79^{+0.33}_{-0.24}$	$-1.18^{+0.23}_{-0.42}$	$23.95^{+1.10}_{-0.69}$	$14.00^{+0.62}_{-0.43}$	$14.83^{+0.78}_{-0.80}$
					$1.30^{+0.09}_{-0.09}$	$-0.76^{+0.06}_{-0.06}$	$0.58^{+0.06}_{-0.06}$	$-1.07^{+0.04}_{-0.04}$	$23.13^{+0.13}_{-0.13}$	$13.76^{+0.10}_{-0.10}$	$14.81^{+0.11}_{-0.11}$
AGN	0.8-1.3	1.09	All	735	$1.16^{+0.63}_{-0.03}$	$-0.44^{+0.25}_{-0.23}$	$0.82^{+0.78}_{-0.52}$	$-1.10^{+0.15}_{-0.37}$	$23.68^{+0.89}_{-0.49}$	$13.75^{+0.55}_{-0.69}$	$14.10^{+1.08}_{-1.29}$
					$1.17^{+0.10}_{-0.10}$	$-0.48^{+0.05}_{-0.05}$	$0.33^{+0.07}_{-0.07}$	$-0.99^{+0.04}_{-0.04}$	$23.13^{+0.13}_{-0.13}$	$13.76^{+0.13}_{-0.13}$	$14.14^{+0.14}_{-0.14}$
HLAGN	0.3-0.8	0.68	All	262	$1.79^{+0.96}_{-0.46}$	$-0.80^{+0.22}_{-0.31}$	$0.44^{+0.19}_{-0.12}$	$-1.43^{+0.35}_{-0.37}$	$24.16^{+1.52}_{-0.91}$	$14.10^{+0.83}_{-0.70}$	$14.03^{+0.99}_{-0.78}$
					$1.41^{+0.10}_{-0.10}$	$-0.72^{+0.07}_{-0.07}$	$0.34^{+0.07}_{-0.07}$	$-1.30^{+0.05}_{-0.05}$	$23.13^{+0.13}_{-0.13}$	$13.98^{+0.11}_{-0.11}$	$13.59^{+0.11}_{-0.11}$
HLAGN	0.8-1.3	1.13	All	366	$1.35^{+1.45}_{-0.28}$	$-0.28^{+0.20}_{-0.27}$	$0.40^{+1.01}_{-0.24}$	$-1.47^{+0.26}_{-0.31}$	$23.51^{+0.40}_{-0.30}$	$13.59^{+0.75}_{-1.18}$	$12.97^{+2.01}_{-1.03}$
					$1.12^{+0.10}_{-0.10}$	$-0.06^{+0.06}_{-0.06}$	$0.20^{+0.07}_{-0.07}$	$-1.36^{+0.04}_{-0.04}$	$23.43^{+0.09}_{-0.09}$	$14.20^{+0.13}_{-0.13}$	$11.96^{+0.16}_{-0.16}$
MLAGN	0.3-0.8	0.67	All	261	$1.34^{+0.05}_{-0.03}$	$-0.79^{+0.13}_{-0.12}$	$1.72^{+0.20}_{-0.38}$	$-1.04^{+0.13}_{-0.23}$	$23.39^{+0.33}_{-0.26}$	$14.11^{+0.21}_{-0.18}$	$15.39^{+0.42}_{-0.47}$
					$1.30^{+0.06}_{-0.06}$	$-0.74^{+0.06}_{-0.06}$	$1.95^{+0.05}_{-0.05}$	$-1.03^{+0.05}_{-0.05}$	$23.39^{+0.13}_{-0.13}$	$14.01^{+0.08}_{-0.08}$	$15.72^{+0.09}_{-0.09}$
MLAGN	0.8-1.3	0.98	All	252	$1.43^{+1.17}_{-0.27}$	$-0.69^{+0.13}_{-0.17}$	$0.21^{+0.33}_{-0.13}$	$-1.06^{+0.26}_{-0.58}$	$25.16^{+1.14}_{-1.58}$	$14.33^{+1.01}_{-0.81}$	$13.76^{+1.51}_{-0.96}$
					$1.18^{+0.10}_{-0.10}$	$-0.66^{+0.06}_{-0.06}$	$0.20^{+0.07}_{-0.07}$	$-0.86^{+0.05}_{-0.05}$	$25.80^{+0.13}_{-0.13}$	$14.22^{+0.12}_{-0.12}$	$12.96^{+0.13}_{-0.13}$

Table 3.3: Table showing the type of source considered; z range and median z; $L_{1.4\mathrm{GHz}}$ for each $b(L_\nu)$ bin; numbers of sources in each bin and finally the parameters (a, b, c, d, $\log_{10}(L_*)$, $\log_{10}(M_N)$ and $\log_{10}(M_*)$) found from fitting the CLF model. These modelled parameters are quoted as the median value for each parameter and the associated uncertainties from the 16^{th} and 84^{th} percentiles as well as the modal bin for each parameter.

3.4 Results and Discussion

As SFGs and AGN are known to have differences in their bias (as discussed in Chapter 2), I will discuss the results for the two populations separately. With the AGN population, my work from Chapter 2 also showed that efficient and inefficient AGN also have different biases, especially at low redshift and so again these two populations are considered separately. However for these HLAGN and (quiescent) MLAGN, the smaller numbers of sources means that I used fewer luminosity bins when investigating the bias.

3.4.1 SFGs

The results of fitting $\Phi(L_\nu|M_h)$ for SFGs is shown in Figure 3.7 for $0.3 \leq z < 0.8$ and in Figure 3.8 for $0.8 \leq z < 1.3$. For the lower redshift SFGs, the parameters are relatively well constrained, however parameter c is hitting the edge of the prior range. This prior range was determined based on the values of Ballantyne (2017), however attempts to extend this parameter range resulted in the same behaviour at the new upper bound limit (taken as 4). As well as this, using the new limits, M_N tended towards larger halo masses that appear unreasonable (i.e. tending to halo masses of $> 10^{16} M_\odot$). At higher redshifts, b is poorly constrained and there is also a double peaked nature to M_*. As can be seen in Figure 3.5, both b and c influence the faint end of the luminosity function, whilst c also influences the bright end of $b(L_\nu)$ and M_* affects the normalisation of $\Phi(L_\nu)$. A lack of constraints on these parameters reflects the limitations of this dataset. Whilst the observations are deep, there will be challenges arising from incompleteness in the catalogues, for example through not having multi-wavelength matches for all sources, as well as small numbers of the brightest objects due to a lack of area to observe these rare objects. Not only this but, as Figure 3.5 indicates, many of these parameters have similar effects on

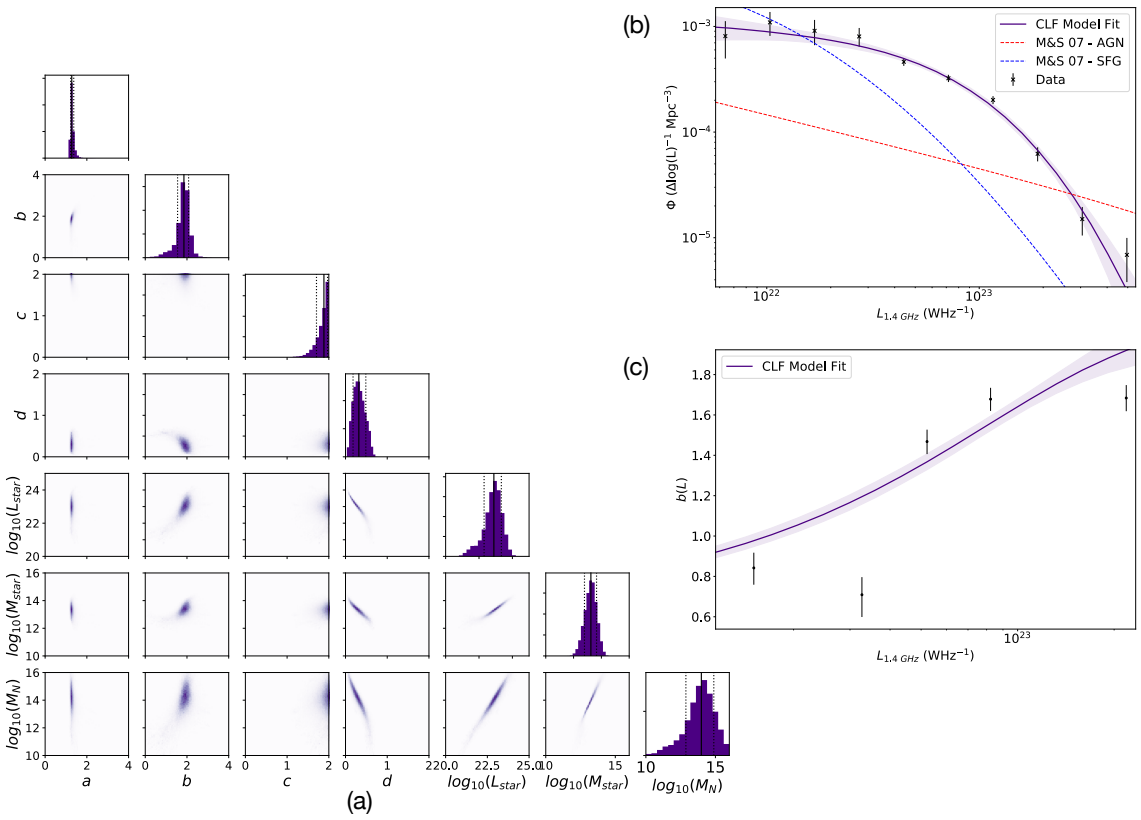


Figure 3.7: $\Phi(L_\nu|M_h)$ and the recovered distributions for $\Phi(L_\nu)$ and $b(L_\nu)$ for SFGs with $0.3 \leq z < 0.8$. Panel (a) shows the corner plot of the `emcee` (post burn-in) chains for the 7 parameters fit in this model. Panel (b) shows the observed luminosity function (black points) and recovered distribution from modelling the CLF (purple). Also shown are the local $\Phi(L_\nu)$ for SFGs (blue) and AGN (red) from Mauch & Sadler (2007) in Panel (b). Panel (c) shows the observed $b(L_\nu)$ distribution (black) and recovered $b(L_\nu)$ from fitting the CLF (purple).

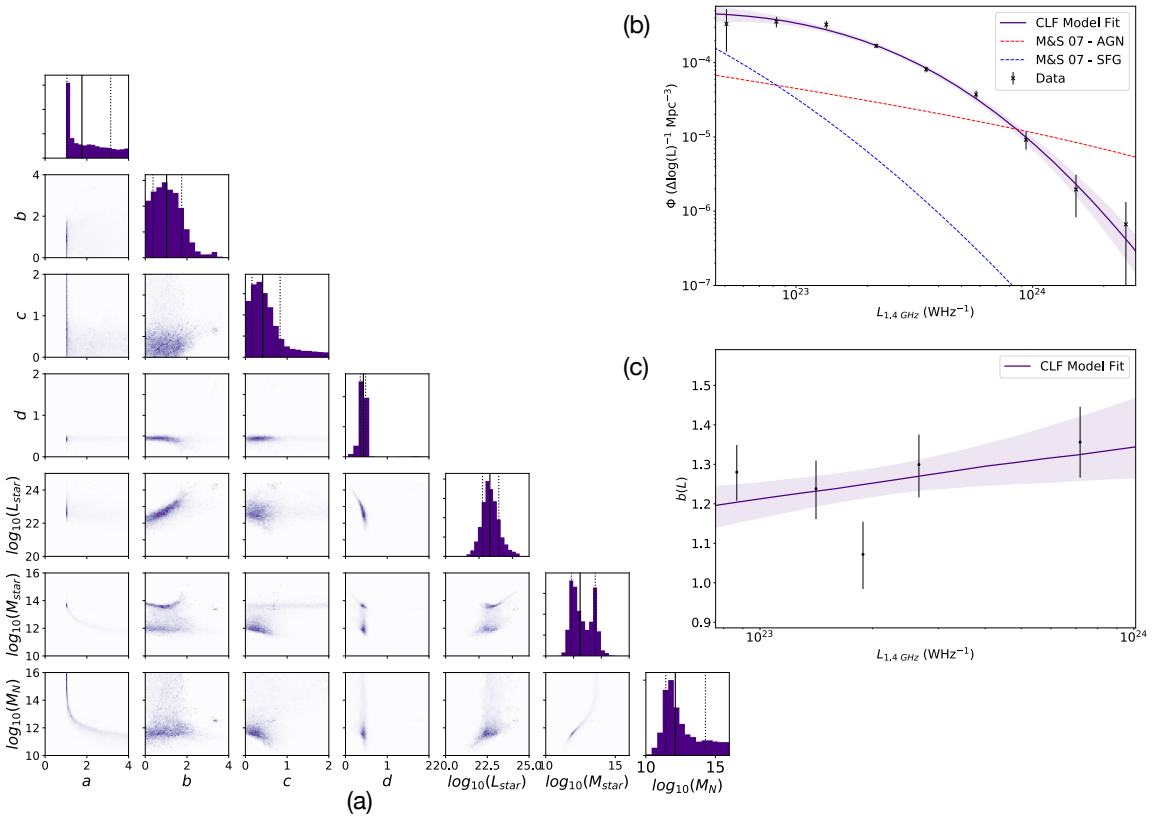


Figure 3.8: $\Phi(L_\nu|M_h)$ and the recovered distributions for $\Phi(L_\nu)$ and $b(L_\nu)$ for SFGs with $0.8 \leq z < 1.3$. Panel (a) shows the corner plot of the `emcee` (post burn-in) chains for the 7 parameters fit in this model. Panel (b) shows the observed luminosity function (black points) and recovered distribution from modelling the CLF (purple). Also shown are the local $\Phi(L_\nu)$ for SFGs (blue) and AGN (red) from Mauch & Sadler (2007) in Panel (b). Panel (c) shows the observed $b(L_\nu)$ distribution (black) and recovered $b(L_\nu)$ from fitting the CLF (purple).

both $\Phi(L_\nu)$ and $b(L_\nu)$, as such this will make fitting the CLF challenging and cause parameters to be degenerate, especially when only a handful of luminosity bins are used to determine $b(L_\nu)$.

Despite a lack of constraints on some parameters, Figures 3.7 and 3.8 show that the model can well reproduce the observed luminosity functions (panel b), as well as the observed $b(L_\nu)$ distribution, shown in panel (c). However, it should also be noted that the values of $\Phi(L_\nu)$ for the SFGs investigated here appear too small at lower luminosities, compared to the results from Mauch & Sadler (2007). This may reflect the fact that only those SFGs determined to be "clean" are included in this work and so some low luminosity SFGs may be missing. Furthermore, SFGs may also be missing from the sample if they were unable to be classified in Smolčić et al. (2017b) (only $\sim$86% of sources had multi wavelength counterparts). Both of these reasons explains why the luminosity function appears low and why the completeness corrections made would be unable to account for this.

The median values for the parameters describing the CLF from the post burn-in chains and the uncertainties from the 16^{th} and 84^{th} percentiles are given in Table 3.3. These values are also shown on the corner plot in Figures 3.7 and 3.8 (panel a) by the solid black line (for the median value) and dashed line for the 16^{th} and 84^{th} percentiles. However, as can be seen, some of the probability distribution functions (PDFs) for these parameters do not appear Gaussian in nature. Therefore, the peak bin in the PDF for each parameter is given in Table 3.3.

Despite limitations in the fitting, the functional form of $\Phi(L_\nu|M_h)$ reproduces the observed luminosity and bias distributions over these redshift ranges. As such, it is likely that the inferred properties of these SFGs, such as average mass (described in Section 3.1), will be accurate over these luminosity ranges. These quantities will be discussed in Section 3.4.3 where I compare the results from the different source populations.

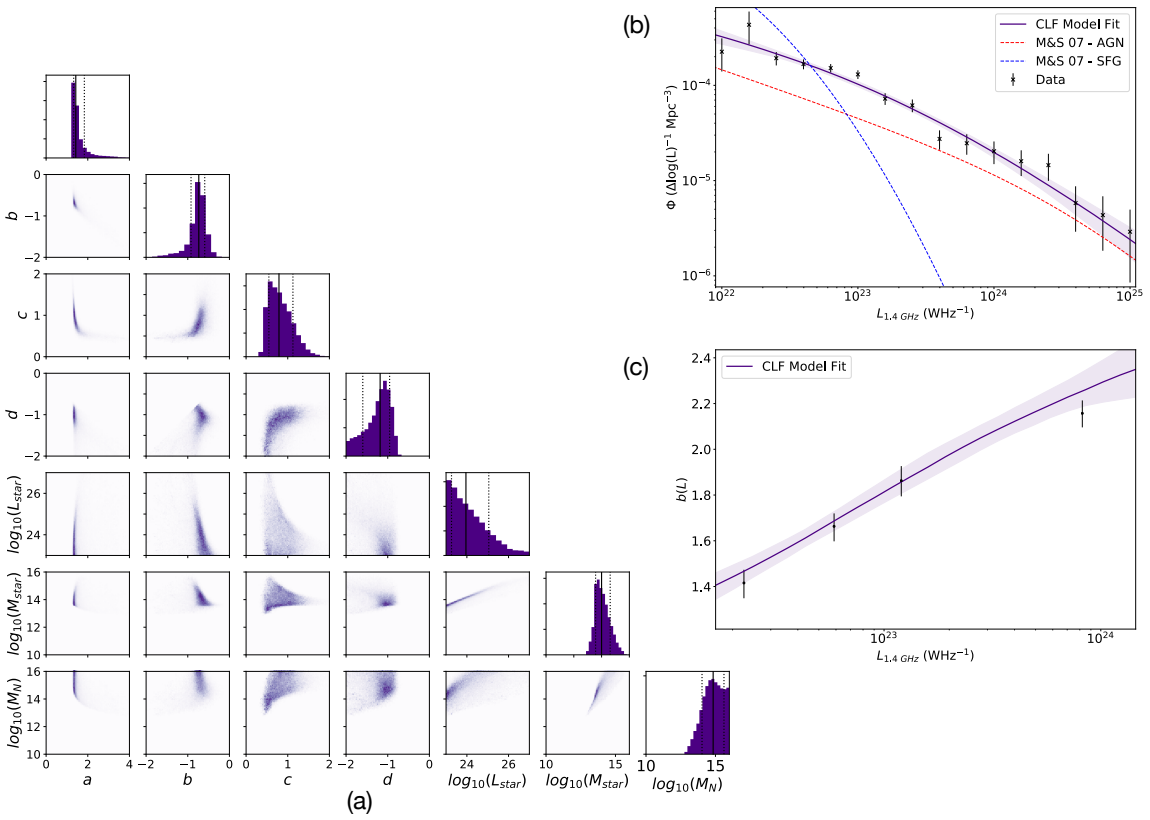


Figure 3.9: $\Phi(L_\nu|M_h)$ and the recovered distributions for $\Phi(L_\nu)$ and $b(L_\nu)$ for AGN with $0.3 \leq z < 0.8$. Panel (a) shows the corner plot of the `emcee` (post burn-in) chains for the 7 parameters fit in this model. Panel (b) shows the observed luminosity function (black points) and recovered distribution from modelling the CLF (purple). Also shown are the local $\Phi(L_\nu)$ for SFGs (blue) and AGN (red) from Mauch & Sadler (2007) in Panel (b). Panel (c) shows the observed $b(L_\nu)$ distribution (black) and recovered $b(L_\nu)$ from fitting the CLF (purple).

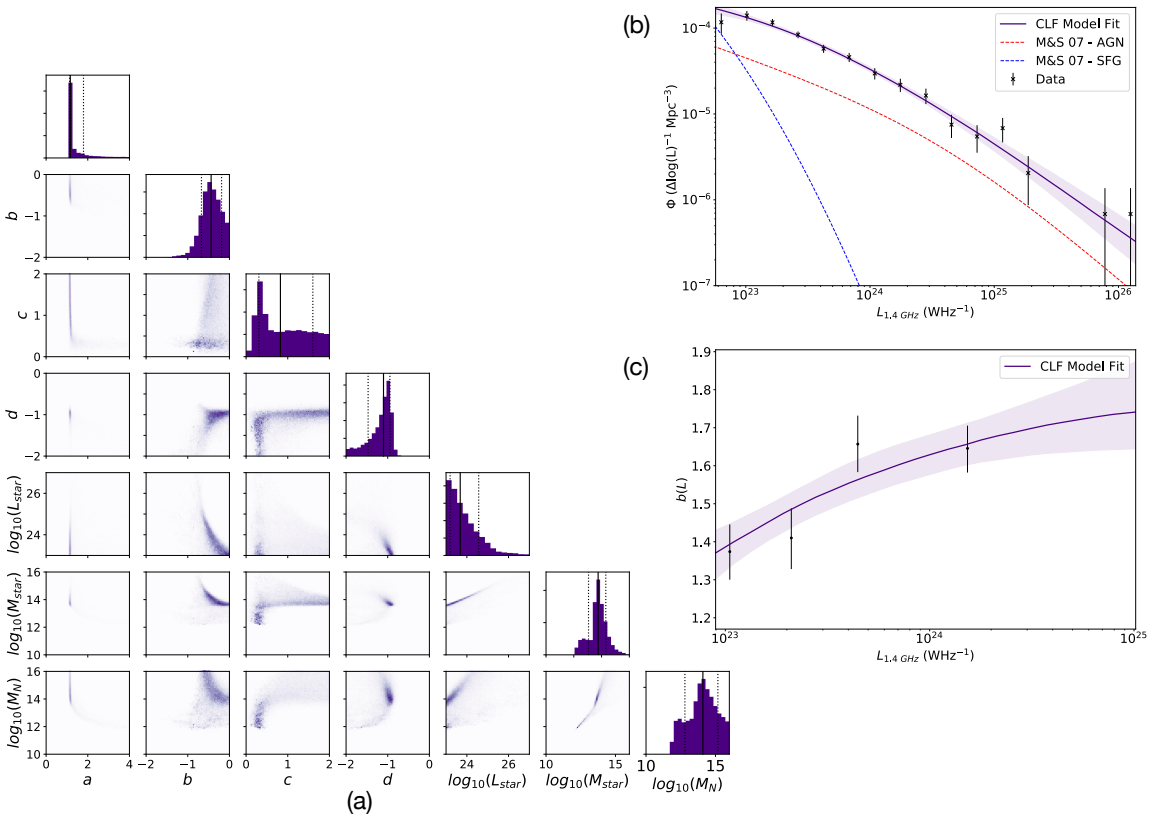


Figure 3.10: $\Phi(L_\nu|M_h)$ and the recovered distributions for $\Phi(L_\nu)$ and $b(L_\nu)$ for AGN with $0.8 \leq z < 1.3$. Panel (a) shows the corner plot of the `emcee` (post burn-in) chains for the 7 parameters fit in this model. Panel (b) shows the observed luminosity function (black points) and recovered distribution from modelling the CLF (purple). Also shown are the local $\Phi(L_\nu)$ for SFGs (blue) and AGN (red) from Mauch & Sadler (2007) in Panel (b). Panel (c) shows the observed $b(L_\nu)$ distribution (black) and recovered $b(L_\nu)$ from fitting the CLF (purple).

3.4.2 AGN

For the full AGN population, the fitting of $\Phi(L_\nu|M_h)$ as well as the reconstructions of $\Phi(L_\nu)$ and $b(L_\nu)$ can be seen in Figure 3.9 for $0.3 \leq z < 0.8$ and Figure 3.10 for $0.8 \leq z < 1.3$. Most parameters are well fit within the prior ranges, with the exception of $\log_{10}(L_*)$. For both the low and high redshift bin, the value of $\log_{10}(L_*)$ is pushing up against the lower value of the prior range (23). For the local luminosity function measured by Mauch & Sadler (2007), the value of L_* is found to be $10^{24.59}$ WHz^{-1} for AGN and this is expected to evolve with redshift (see e.g. Brown et al., 2001; Sadler et al., 2007; Prescott et al., 2016; Smolčić et al., 2017). However, both the low and high redshift luminosity functions only have a few bins that are above the expected L_* value, as such the high luminosity power law will struggle to be constrained. At the low luminosity end, on the other hand, the uncertainties on the luminosity function are smaller and so any apparent flattening in the power law here will force L_* towards lower luminosities. For the $0.8 \leq z < 1.3$ bin, the data point at the lowest luminosity appears too low. This suggests there are deficiencies in the data that are not fully being accounted for in the completeness corrections. This creates an anomalously low value L_*, because it artificially appears that there is a break at low luminosities. Again, due to tails in the PDFs, the median values and uncertainties quoted in Table 3.3 may not be appropriately describing the modal parameters and so the modal bin is also given in Table 3.3. This is especially true for $\log_{10}(L_*)$.

For HLAGN, on the other hand, the results can be seen in Figure 3.11 for $0.3 \leq z < 0.8$ and Figure 3.12 for $0.8 \leq z < 1.3$. The $0.3 \leq z < 0.8$ redshift bin suffers from similar problems to the AGN, as described above. Here, the parameters are well constrained except L_* which is again preferentially pushed to lower luminosities, as discussed above. For the higher redshift bin, these fits suggest a value of b which is flat or positive, reflecting the flattening in $\Phi(L_\nu)$ at low luminosities, which may arise from the fact that the completeness correction described in Section 3.3.2 may not have

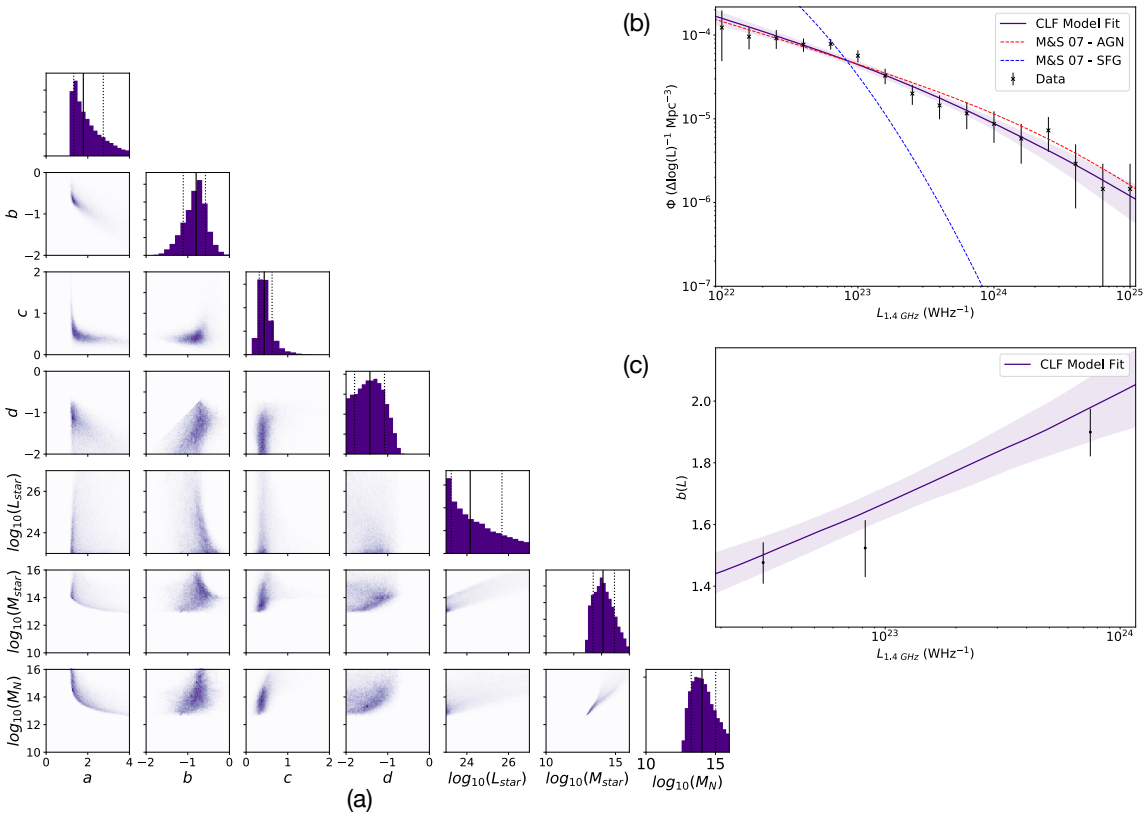


Figure 3.11: $\Phi(L_\nu|M_h)$ and the recovered distributions for $\Phi(L_\nu)$ and $b(L_\nu)$ for HLAGN with $0.3 \leq z < 0.8$. Panel (a) shows the corner plot of the `emcee` (post burn-in) chains for the 7 parameters fit in this model. Panel (b) shows the observed luminosity function (black points) and recovered distribution from modelling the CLF (purple). Also shown are the local $\Phi(L_\nu)$ for SFGs (blue) and AGN (red) from Mauch & Sadler (2007) in Panel (b). Panel (c) shows the observed $b(L_\nu)$ distribution (black) and recovered $b(L_\nu)$ from fitting the CLF (purple).

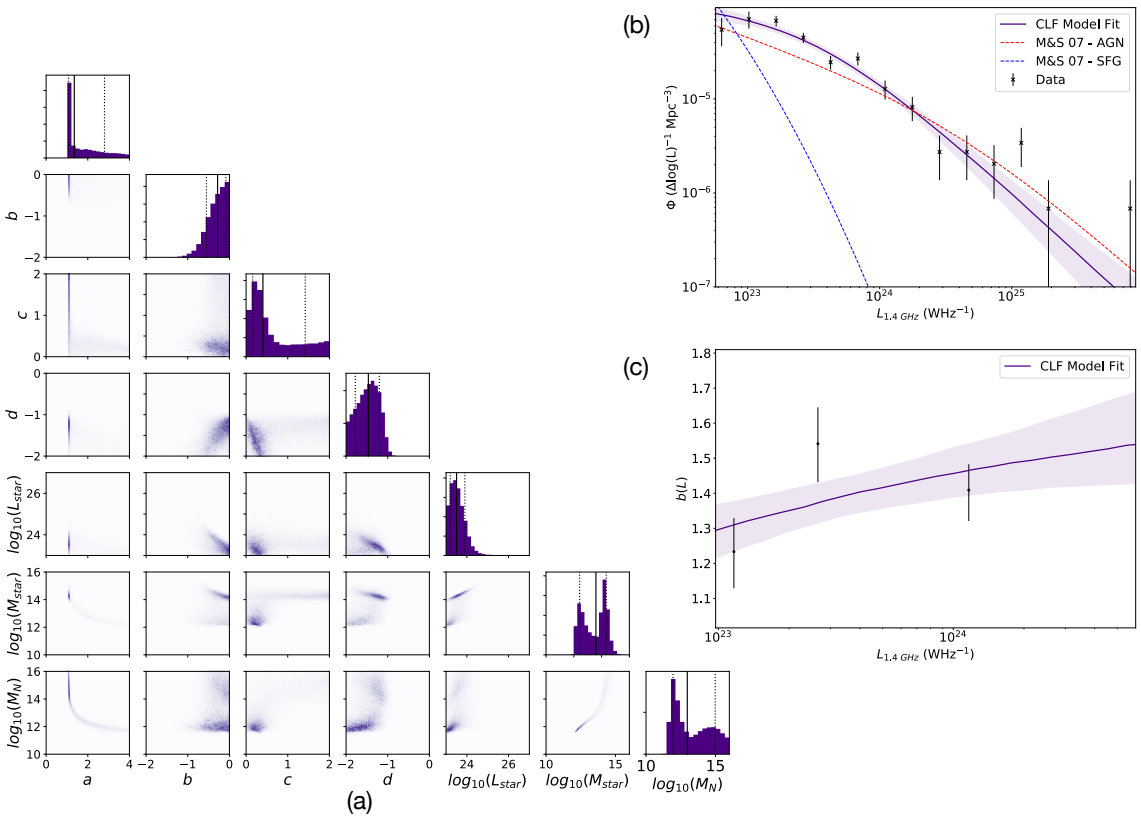


Figure 3.12: $\Phi(L_\nu|M_h)$ and the recovered distributions for $\Phi(L_\nu)$ and $b(L_\nu)$ for HLAGN with $0.8 \leq z < 1.3$. Panel (a) shows the corner plot of the `emcee` (post burn-in) chains for the 7 parameters fit in this model. Panel (b) shows the observed luminosity function (black points) and recovered distribution from modelling the CLF (purple). Also shown are the local $\Phi(L_\nu)$ for SFGs (blue) and AGN (red) from Mauch & Sadler (2007) in Panel (b). Panel (c) shows the observed $b(L_\nu)$ distribution (black) and recovered $b(L_\nu)$ from fitting the CLF (purple).

been large enough. Also it appears that there are degeneracies in determining M_* and M_N which make both appear double peaked. These arise from the value of a, which in many cases appears peaked at a value of ~ 1.2 but also has an excess tail to larger a (which are more consistent with the values found in Ballantyne, 2017). However these different parameters are also able to well recreate $\Phi(L_\nu)$ and $b(L_\nu)$ as can be seen in panels (b) and (c) of Figures 3.11 and 3.12. Finally, the results for MLAGN can be seen in Figures 3.13 for $0.3 \leq z < 0.8$ and Figure 3.14 for $0.8 \leq z < 1.3$. For the lower redshift bin, $0.3 \leq z < 0.8$, again L_* is pushed to the lower end of the prior range. However, c and M_N also favour the edge of the chosen prior ranges, which for M_N are unphysical. At higher redshifts (Figure 3.14) it is L_* again that is poorly constrained, across the entire redshift range. This reflects the fact the $b \sim d$, suggesting that the data appears to be well fit by a single power law. The full parameter fits are given in Table 3.3.

3.4.3 SFGs vs. AGN

Using the results from Sections 3.4.1 and 3.4.2, it is possible to compare the average properties of these different source populations as well as the differences in the observed luminosity functions and bias evolution. As the work in Chapter 2 suggests that the different AGN populations will inhabit different environments, I compare the results for SFGs to those from the MLAGN and HLAGN populations separately.

The comparisons of the recovered luminosity functions of SFGs to the different AGN populations are shown in Figure 3.15. This is shown for SFGs (blue) compared to HLAGN (purple) and MLAGN (red) and the luminosity range for which there is available data observed is indicated by the darker shaded regions. It can be seen that SFGs dominate at the lowest luminosities, with the luminosity value for which the SFGs dominate over the AGN increasing to higher luminosities in the higher redshift bin. In the $0.8 \leq z < 1.3$ bin, the luminosity at which the number density

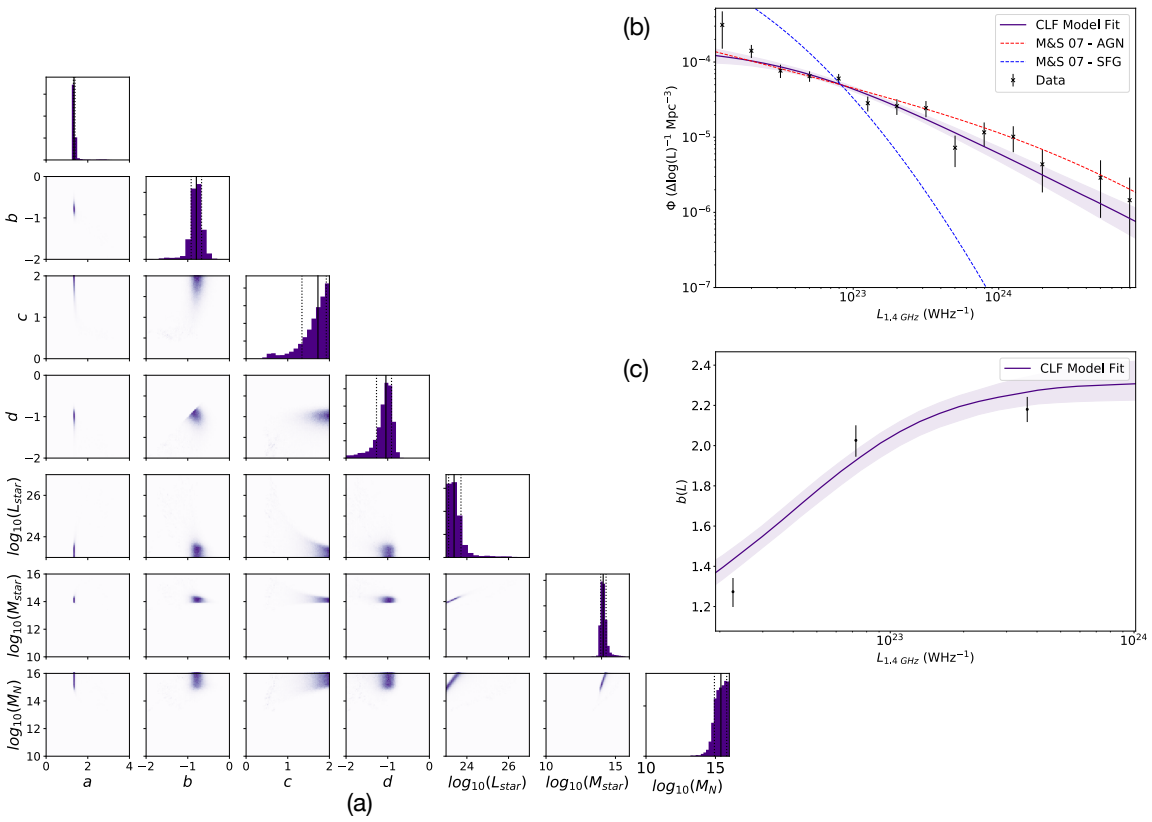


Figure 3.13: $\Phi(L_\nu|M_h)$ and the recovered distributions for $\Phi(L_\nu)$ and $b(L_\nu)$ for MLAGN with $0.3 \leq z < 0.8$. Panel (a) shows the corner plot of the `emcee` (post burn-in) chains for the 7 parameters fit in this model. Panel (b) shows the observed luminosity function (black points) and recovered distribution from modelling the CLF (purple). Also shown are the local $\Phi(L_\nu)$ for SFGs (blue) and AGN (red) from Mauch & Sadler (2007) in Panel (b). Panel (c) shows the observed $b(L_\nu)$ distribution (black) and recovered $b(L_\nu)$ from fitting the CLF (purple).

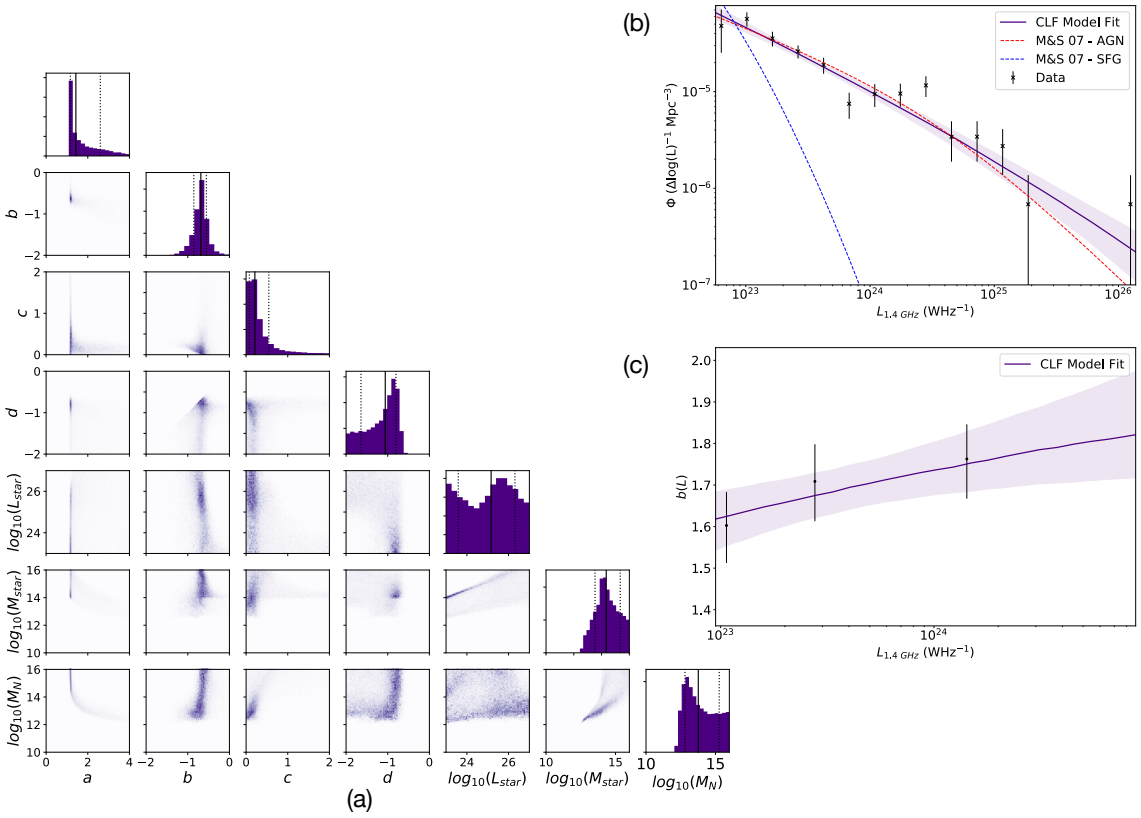
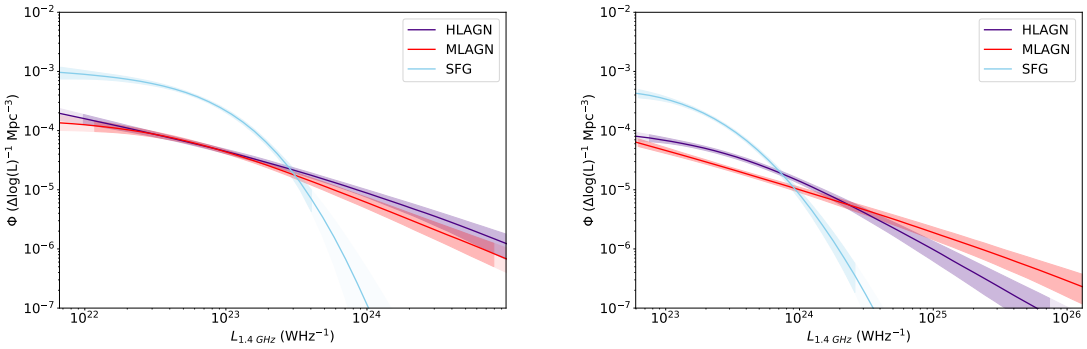
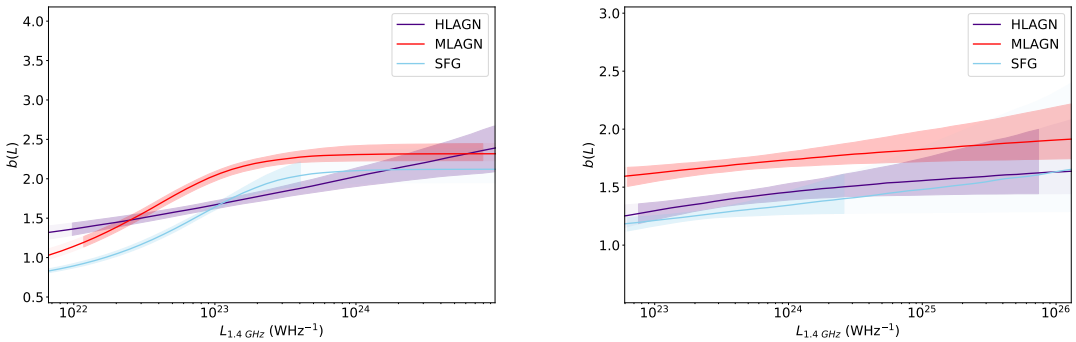


Figure 3.14: $\Phi(L_\nu|M_h)$ and the recovered distributions for $\Phi(L_\nu)$ and $b(L_\nu)$ for MLAGN with $0.8 \leq z < 1.3$. Panel (a) shows the corner plot of the `emcee` (post burn-in) chains for the 7 parameters fit in this model. Panel (b) shows the observed luminosity function (black points) and recovered distribution from modelling the CLF (purple). Also shown are the local $\Phi(L_\nu)$ for SFGs (blue) and AGN (red) from Mauch & Sadler (2007) in Panel (b). Panel (c) shows the observed $b(L_\nu)$ distribution (black) and recovered $b(L_\nu)$ from fitting the CLF (purple).



(a) $0.3 \leq z < 0.8$ (b) $0.8 \leq z < 1.3$

Figure 3.15: The reconstructed $\Phi(L_\nu)$ from $\Phi(L_\nu|M_h)$ for SFGs (blue) compared to HLAGN (purple) and MLAGN (red) for sources with $0.3 \leq z < 0.8$ (left) and $0.8 \leq z < 1.3$ (right). The darker shaded regions indicate the luminosity range for which there is available data over.



(a) $0.3 \leq z < 0.8$ (b) $0.8 \leq z < 1.3$

Figure 3.16: The reconstructed $b(L_\nu)$ from $\Phi(L_\nu|M_h)$ for SFGs (blue) compared to HLAGN (purple) and MLAGN (red) for sources with $0.3 \leq z < 0.8$ (left) and $0.8 \leq z < 1.3$ (right). The darker shaded regions indicate the luminosity range for which there is available data over.

of SFGs and AGN overlap is at $\sim 8 \times 10^{23}$ WHz^{-1}, compared to $\sim 3 \times 10^{23}$ WHz^{-1} for $0.3 \leq z < 0.8$. Figure 3.15 also shows that the luminosity functions for the MLAGN and HLAGN, whilst similar in the lower redshift bin, behave differently in the $0.8 \leq z < 1.3$ bin. For $0.8 \leq z < 1.3$, the MLAGN appear more numerous at higher luminosities, whilst the HLAGN (despite the fact they appear to suffer from incompleteness) are more numerous at lower luminosities. Whilst work in the literature (e.g. Best et al., 2014; Pracy et al., 2016) suggests that HERGs dominate at high luminosities and MLAGN at low luminosities, the lack of MLAGN dominating at these luminosities is likely, in part, due to the fact the clean MLAGN sample was used for this work and, as well as this, the populations representing LERGs may be different between this work and others. At high luminosities, the MLAGN population appear to have higher number densities. Whilst the luminosity function of inefficient AGN has been shown to increase at high luminosities (Best et al., 2014), this lack of HLAGN at high luminosities is unexpected given the dominance of HLAGN at high luminosities and the rapid evolution in their space density with redshift (see e.g. Pracy et al., 2016). An explanation for this is currently difficult to determine. It appears though that the classification of sources into HLAGN and MLAGN given here may not be complete and may also not exactly reflect the populations that are defined and used in previous work.

The differences in the bias can be seen in Figure 3.16 where, over a large range of luminosities, the bias of the MLAGN population is significantly higher than the HLAGN and SFGs, in agreement with the work from Chapter 2. Whilst in Chapter 2 (at low redshifts) the overall MLAGN populations had much greater bias compared to the HLAGN and SFGs, this work suggests that this is true over a large range of luminosities. For the $0.3 \leq z < 0.8$ redshift bin, there is strong evolution in the bias as a function of luminosity for these inefficient accreters (MLAGN), but a flattening at high luminosities. For SFGs, this increase in clustering strength with

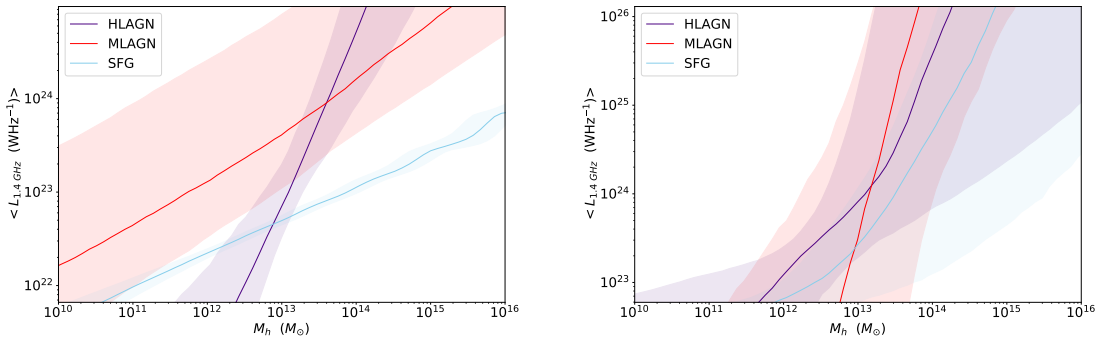
luminosity reflects previous work by Dolley et al. (2014), who also observed these trends. An increase in bias with luminosity has also been observed for the AGN dominated sample of radio sources in Lindsay et al. (2014a). This increase is likely to be related to more luminous sources existing in more massive galaxies, which need to be supported by massive haloes (Hatfield et al., 2016; Tinker et al., 2017). At some luminosities, though, the bias of HLAGN and MLAGN appear much more similar (or indeed smaller for MLAGN). This occurs at low luminosities ($L_{1.4\ \mathrm{GHz}} \lesssim 2 \times 10^{23}$ WHz^{-1}) as well as at high luminosities ($L_{1.4\ \mathrm{GHz}} \gtrsim 3 \times 10^{24}$ WHz^{-1}). In the higher redshift bin, the evolution in bias is much more gradual, but the MLAGN are still on average the most biased population, suggesting they reside in more massive haloes than the HLAGN and SFGs. This gradual increase with luminosity in this redshift bin, compared to the lower redshift one, suggests that there is less of a dependence on halo mass with luminosity at higher redshift. In this redshift range the HLAGN and SFGs are much more similar in terms of their bias, so whilst on average the MLAGN appear to reside in more massive haloes, as seen in Chapter 2, the bias between these two populations is consistent within the uncertainties.

Using the functional forms of the CLF and Equations 3.4 and 3.5, it is also possible to investigate the typical halo masses that these sources reside in as well as the typical luminosities for a population of galaxies for different halo masses. The typical luminosity, for sources residing in haloes of differing masses can be seen in Figure 3.17, over the luminosity ranges for which there is data. This suggests that in the lower mass haloes ($\lesssim 5 \times 10^{13} M_{\odot}$), the MLAGN are more luminous than the HLAGN, whilst at higher halo masses this is reversed. Over all of the halo masses considered in Figure 3.17, the SFG population have the lowest average luminosities. This is expected due to the distribution of their luminosity functions, which is preferential towards smaller luminosities.

In Main et al. (2017), their investigations included the study of the relationship

between radio luminosity at 1.4 GHz to halo mass in order to study radio mode accretion (this corresponds to jet mode, as discussed in Section 1.3, see Figure 4 of Heckman & Best, 2014). They found a power law relation of $L_{1.4\mathrm{GHz}} \propto M_{2500m}^{2.51}$, where M_{2500m} is defined similarly to M_{200m} where a halo is defined as the region where the density is 2500 times the mean matter density of the Universe. The value of M_{2500m} will be smaller than M_{200m} as M_{2500m} will only include the densest part of the cluster, which will be the small central region, whereas using 200m will allow larger radii and hence larger halo masses to be probed. Therefore, I am unable to do a direct comparison to their work in this Thesis. However fitting the results for both the HLAGN and MLAGN over these redshift ranges, I find $L_{1.4\ \mathrm{GHz}} \propto M_h^{\sim 1.8}$ for $0.3 \leq z < 0.8$ and $L_{1.4\ \mathrm{GHz}} \propto M_h^{\sim 1.2}$ for $0.8 \leq z < 1.3$ for the HLAGN population and $L_{1.4\ \mathrm{GHz}} \propto M_h^{\sim 0.5}$ $(0.3 \leq z < 0.8)$ and $L_{1.4\ \mathrm{GHz}} \propto M_h^{\sim 3.2}$ $(0.8 \leq z < 1.3)$ for MLAGN. The latter is a similar scaling relation to that observed in simulated radio relics (Skillman et al., 2011) whilst for HLAGN in the higher redshift bin, the relations are more similar to that found for X-ray AGN at $z \sim 0.9$ in Ballantyne (2017) where a relation of $L \propto M_h^{1.2}$ is found. However, Ballantyne (2017) did observe a much flatter power law slope (~ 0.8) for $z \sim 0$ whilst, for the HLAGN population observed here, the relation is steeper for the lower redshift bin.

Finally, the average halo masses that these sources reside in was also investigated, using Equation 3.5. The results for this are shown in Figure 3.18. This suggests that MLAGN, in agreement with Chapter 2, reside in the most massive haloes for both the redshifts bins considered. However, whilst for all of the source types investigated here (SFGs, HLAGN and MLAGN) the evolution in the halo masses as a function of luminosity evolves much more quickly at lower redshift than at higher redshift, which is also seen in Ballantyne (2017). The average mass of haloes that the MLAGN, HLAGN and SFGs reside in all appear to be smaller for the same given luminosity at higher redshifts. This implies that for the same powerful emission, this can be

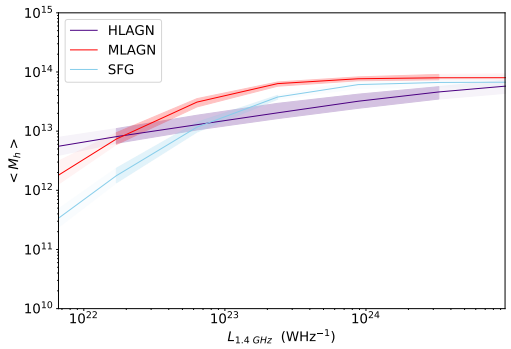


(a) $0.3 \leq z < 0.8$ (b) $0.8 \leq z < 1.3$

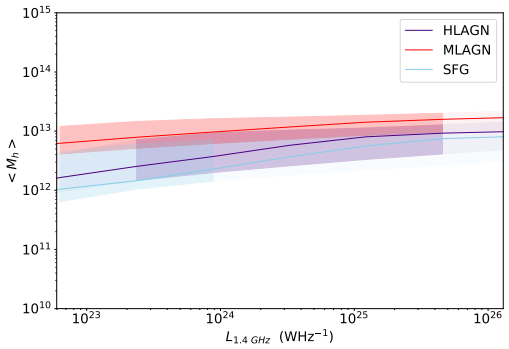
Figure 3.17: Comparison of the typical luminosities for SFGs (blue) compared to HLAGN (purple) and MLAGN (red) recovered from $\Phi(L_\nu|M_h)$.

supported by a smaller mass halo at $0.3 \leq z < 0.8$ compared to $0.8 \leq z < 1.3$. This supports the findings not only of Chapter 2 for HLAGN, but also now extends this finding to the other two populations investigated. Whilst Ballantyne (2017) observed such a behaviour at high luminosities, $\gtrsim 10^{44}$ erg s^{-1}, he did not observe this at the lowest luminosities.

Finally, it is also possible to use $\Phi(L_\nu|M_h)$ to investigate the average ratio of $< L_\nu > /M_h$ as a function of halo mass. Hatfield et al. (2016) measured the ratio of $\frac{M_*}{M_h}$ as a function of halo mass to investigate the halo mass at which it is the maximum efficiency for producing stars. In a similar vein, by considering the ratio of $< L_\nu > /M_h$ it may be possible to investigate how efficient different halo masses are at producing luminous SFGs and AGN. This can be seen in Figure 3.19, which suggests that at lower redshifts, the MLAGN and SFGs are decreasing in their efficiency when they are residing in higher mass haloes. At higher redshifts, though, the results are a lot more uncertain, but suggest that MLAGN are increasing in their efficiency when they reside in higher mass haloes. For HLAGN and SFG the haloes appear to grow in efficiency but only above halo masses of $\sim 10^{13}$ M$_\odot$. Below this they are declining in efficiency.

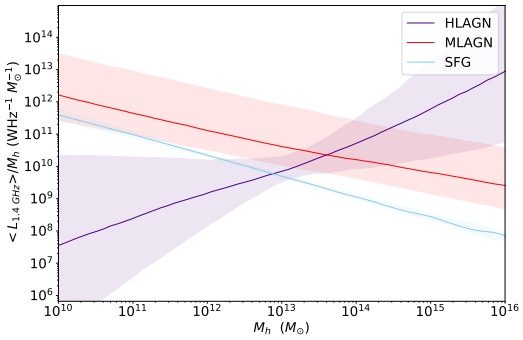


(b) $0.3 \leq z < 0.8$

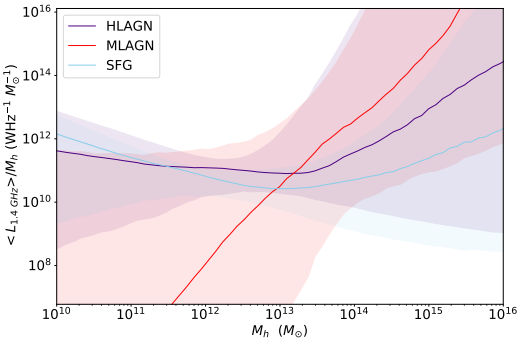


(b) $0.8 \leq z < 1.3$

Figure 3.18: Comparison of the typical halo masses that SFGs (blue) reside in compared to HLAGN (purple) and MLAGN (red) recovered from $\Phi(L_\nu|M_h)$.



(b) $0.3 \leq z < 0.8$



(b) $0.8 \leq z < 1.3$

Figure 3.19: Comparison of the typical luminosity to halo mass ratio for SFGs (blue) compared to HLAGN (purple) and MLAGN (red) recovered from $\Phi(L_\nu|M_h)$.

3.4.4 Jet Power and Environment Relation

For both FRI and FRII type sources, there are expected relations between the radio luminosity and jet power (see Willott et al., 1999; Bîrzan et al., 2008; Cavagnolo et al., 2010; Heckman & Best, 2014; Croston et al., 2018). This work investigating the CLF therefore provides the opportunity to investigate how this jet power may be related to the environment in which AGN reside, using halo mass as a proxy for environment. For FRII sources, one relation that is commonly used in the literature is the relation found by Willott et al. (1999). By making assumptions about the emission in jets, Willott et al. (1999) found a relation between radio luminosities and jet power that has been shown to hold well in comparisons to observational evidence (see e.g. Cavagnolo et al., 2010; Daly et al., 2012) and is used commonly in the literature (Fernandes et al., 2011; Kalfountzou et al., 2017; Ceraj et al., 2018). The relations from Willott et al. (1999) links jet power (Q) and radio luminosity (at 151 MHz; $L_{151\mathrm{MHz}}$). In the relation from Willott et al. (1999), the jet power is given by:

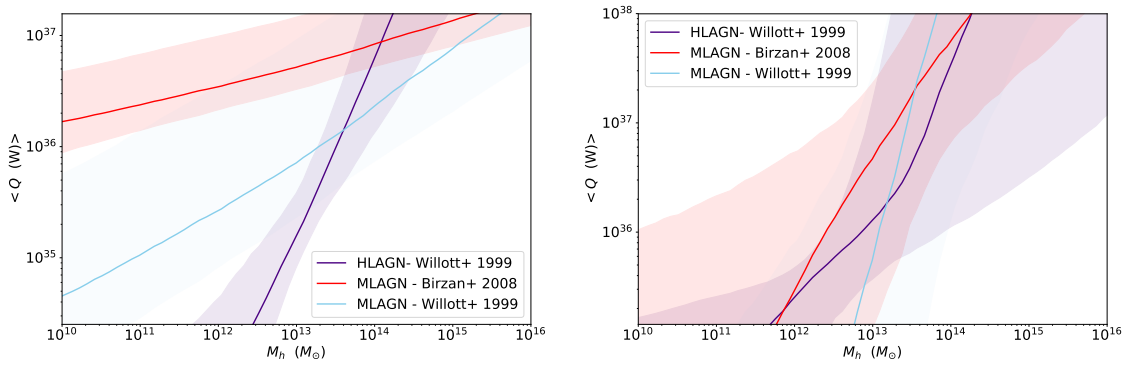
$$Q \sim 3 \times 10^{38} f^{\frac{3}{2}} \left(\frac{L_{151\ \mathrm{MHz}}\ (\mathrm{WHz^{-1}sr^{-1}})}{10^{28}} \right)^{\frac{6}{7}} \mathrm{W}, \tag{3.15}$$

where f is a factor to take into account uncertainties that are involved in calculating the jet power from the luminosity. This is thought to be in the range $\sim 1-20$ and has been taken as $f \sim 10$ in many previous studies (e.g. Fernandes et al., 2011; Gürkan et al., 2015), which I also adopt here.

Bîrzan et al. (2008), on the other hand, studied VLA observations of cluster galaxies including many FRI radio galaxies and used observations of the gas in the cluster to estimate the halo sizes and find the relations between jet power and radio luminosities. This relation is given by:

$$Q \sim 10^{35} \times 10^{1.85} \left(\frac{L_{1.4\ \mathrm{GHz}}\ (\mathrm{WHz^{-1}})}{10^{24}} \right)^{0.35} \mathrm{W}. \tag{3.16}$$

As the HLAGN sources are thought to be mainly of FRII morphology (see Figure 4



(a) $0.3 \leq z < 0.8$ (b) $0.8 \leq z < 1.3$

Figure 3.20: Average jet power for AGN as a function of halo mass using the relations from Willott et al. (1999) for both MLAGN (blue) and HLAGN (purple) as well the relation Bîrzan et al. (2008) for MLAGN (red).

of Heckman & Best, 2014) the relation from Willott et al. (1999) can be used in order to infer how the jet power varies with environment for these sources. For the quiescent MLAGN, on the other hand, these will be a mixture of FRI and FRII morphologies (again see Figure 4 of Heckman & Best, 2014) and so both the relations from Willott et al. (1999) and Bîrzan et al. (2008) will be necessary in order to investigate how the jet power of these sources may be environment dependent. However, I note there may be other relations that may be appropriate (see Cavagnolo et al., 2010; Croston et al., 2018). The result of using Equations 3.15 and 3.16 to investigate this dependence can be seen in Figure 3.20 where the left hand panel shows this relation in the lower redshift bin ($0.3 \leq z < 0.8$) and the right hand panel shows the same figure but for the higher redshift bin ($0.8 \leq z < 1.3$). As the relative fractions of FRI and FRII sources in the MLAGN are unknown, both relations are shown however, in practice, the real jet power will be a compromise between the relative contributions of the two populations.

From Figure 3.20 it can be seen that the HLAGN at the lower redshifts have a much faster increase in the jet powers as halo masses increase, whilst the two MLAGN

components of jet power have a much shallower increase with halo mass. At higher redshifts, on the other hand, the much larger uncertainties in these jet powers suggests all the components have similar evolutions to one another of the jet power as a function of halo mass. Deeper and wider observations with future radio surveys will provide much larger samples of HLAGN and MLAGN. This will be possible with surveys such as with the deep LOFAR fields (van Haarlem et al., 2013) or with MIGHTEE (Jarvis et al., 2017). Using these larger surveys, this will provide much better constraints of the CLF and so develop our understanding of how the feedback of energy from AGN to its surroundings depends on the dark matter environments in which it resides.

3.5 Limitations of this Work

Whilst the study of the CLF using radio observations for the first time is exciting, there are clear limitations to this work. In particular the fitting of the CLF is not as well constrained for SFGs and AGN as I would have hoped. The reasons for this are likely due to a variety of factors, including:

- *Constraining the bright end of the Luminosity Function for AGN*: The parameter d represents the bright end of the luminosity function of AGN, with $log_{10}(L*)$ representing the luminosity at which $\Phi(L_\nu)$ transitions from the low luminosity slope to the high luminosity one. To constrain this bright end of $\Phi(L_\nu)$, it is essential to have bright AGN with luminosities $\gtrsim 10^{25}$ WHz^{-1}. However as these bright sources are rare, large areas are required. Without these large areas, $\Phi(L_\nu)$ is poorly constrained, as can be seen in panel (c) of Figures 3.7 - 3.14. Therefore, surveys with good ancillary data but larger areas are essential. Surveys such as MIGHTEE (Jarvis et al., 2017) as well as the deep tier surveys with LOFAR (van Haarlem et al., 2013), will therefore be essential.

- *The CLF model choice*: This work may also be limited by the choice of CLF

model that is assumed. The model used is the one also used in Ballantyne (2017), however in previous work such as Yang et al. (2003) and Cooray & Sheth (2002) with optical data, a different functional form was assumed. Ballantyne (2017) use the functional form described in Equations 3.10 as halo masses in X-ray AGN had been observed to follow a similar distribution (see Chatterjee et al., 2013). However the dependence of $\Phi(L_\nu|M_h)$ on M_h (when the luminosity dependence is removed) should reflect $< N >$, which is known to behave as the halo mass dependent form described in Equation 3.10 for satellite galaxies but behave differently for central galaxies. The functional form for central galaxies is given in Equation 3.17 (where erf is the error function). If this were to be used, this would also need to be adapted so that there would be some luminosity evolution in the bias. Therefore using a functional form that is known to have worked successfully in the past in (Ballantyne, 2017) and that can recreate the observed $b(L_\nu)$ and $\Phi(L_\nu)$ seemed appropriate for this work. However with larger datasets it will be important to compare how successful these different models are and whether one is preferred. For deeper observations such as with CHILES Con-Pol (Hales & Chiles Con Pol Collaboration, 2014) or MIGHTEE (Jarvis et al., 2017) where it is expected that the '1-halo' clustering term may be observed, a combination of the centrals and satellite galaxy functional forms would be appropriate.

$$\Phi(L_\nu|M_h) = \frac{f}{2}\left[1 + erf\left(\frac{log_{10}(\frac{M}{M_*})}{\sigma_M}\right)\right] \; f(L_\nu) \tag{3.17}$$

- *The Evolution of Bias*: One thing that is crucial to linking the galaxy populations to the halo masses in which they reside is the evolution of bias with luminosity. However this is currently limited to small numbers of sources in each luminosity bin. This not only limits the constraints possible on the bias,

but also the number of luminosity bins (and ranges of luminosities) in which $b(L_\nu)$ can be investigated. As well as this, by splitting in L_ν bins that have equal numbers within each bin, this means that a diverse range of luminosities will exist within a given L_ν bin. This can be seen in Figure 3.2 and is not ideal. As such, if more sources were available, it may be more appropriate to split evenly in luminosity (or log luminosity) bins. This was not possible for this work as splitting luminosities in this way lead to $\sim$100s of sources in one bin, but $\lesssim 20$ sources in the high luminosity bin. This resulted in too few sources in the final luminosity bins to accurately investigate the clustering. In order to obtain better constraints on $b(L_\nu)$ it is therefore essential to have more sources such as from the surveys described above and this will be further discussed in Chapter 6. The relatively small number of sources available is also challenging as the redshift bins need to be small, due to the relationship between b and z (see e.g. Mo & White, 1996; Wilman et al., 2008; Tinker et al., 2010, and Chapter 2). As well as this, the large redshift bins means that the luminosity redshift distribution in Figure 3.2 is affected by the flux limit of the survey, which in turn means that sources of a given luminosity that are detectable at lower redshifts are not observable at higher redshifts within the same redshift bin. Therefore whilst the bin widths used here were necessary in order to be able to describe the evolution of bias with luminosity, they are larger than ideally wanted and so should be reduced for future studies.

- *The Classification of AGN and SFGs*: Finally, as described when discussing the data used in Chapter 2, there are sources within this catalogue that can be considered to both be SFGs and MLAGN. By using a clean sample of the sources, this ensures that I am drawing conclusions about sources known to be part of the given source type. This does however limit the work as, for example, the luminosity functions are lower than expected. Therefore, to improve in

the future, either more multi-wavelength data is needed on observed sources in order to get better classifications as to whether a source is an AGN or SFG (or has both an AGN and high SFRs), or it may be more sensible to consider the CLF for all galaxies as a function of likelihood of being of a certain source type. This may also be especially important in comparisons between HLAGN and MLAGN, because these are defined based on accretion efficiency for which there may not necessarily be a clear dichotomy between the two populations (see e.g. Whittam et al., 2018).

3.6 Conclusions

In this Chapter, I have made the first measurement of the Conditional Luminosity Function (CLF) for SFGs and AGN at radio wavelengths using observations from the VLA 3 GHz COSMOS Survey (Smolčić et al., 2017a) and its counterpart multi-wavelength classification catalogue (Smolčić et al., 2017b). The CLF, $\Phi(L_\nu|M_h)$ describes the number of sources at a given luminosity within dark matter haloes of a given mass and can be constrained by both investigating the clustering for a population of sources as well as their luminosity distribution. Studying the CLF allows investigations into how the environments and luminosities of sources are related, as well as the evolution of this relationship over the age of the universe. Constraining the CLF also provides a model for testing accretion models in simulations more rigorously.

This work is in broad agreement with the conclusions of Chapter 2 that the average halo masses that different populations of galaxies reside in vary dependent on source type. Over the redshift ranges considered, the inefficient MLAGN reside in the most massive haloes with more efficient AGN and SFGs residing in lower mass haloes. However, for all of these populations it appears that these average halo masses

are smaller at larger redshifts, suggesting less massive haloes are needed to support powerful radio sources in the past. I used relations from Willott et al. (1999) and Bîrzan et al. (2008) in order to quantify how the jet power that is injected into the environment is related to the dark matter environments that these different AGN inhabit. Whilst MLAGN will likely have a jet power distribution that is a combination of relations from both Willott et al. (1999) and Bîrzan et al. (2008), and so it is hard to quantify the total jet power, it appears the relation of jet powers compared to environments increase with luminosity much more steeply for more massive environments of HLAGN compared to the MLAGN populations. With future work studying larger populations of HLAGN, this work will provide better understanding of the physically motivated dependences of AGN feedback on environments.

There are, however, limitations to the work presented. These include having too few sources due to a lack of area, which not only means that these results are more likely to be affected by cosmic variance, which is not taken into account here, but also means that constraints on the bias and the bright end of the luminosity function are poorer, which affects how well the CLF can be constrained. Nevertheless, even if the parameters are not as tightly constrained as hoped, the CLF reproduces the observed $\Phi(L_\nu)$ and $b(L_\nu)$ functions successfully. The challenges in constraining the CLF are especially true for the AGN populations, where the high luminosity constraints on $\Phi(L_\nu)$ are poor and so the true CLF may be affected by difficulties in measuring L_*, for example. Despite this, the work presented in this Chapter is exciting, as the CLF has never been investigated in the radio where the comparison between SFGs and AGN can be made due to the large numbers of both types of sources that can be observed. The work presented here will therefore also be revolutionised with future surveys such as MIGHTEE (Jarvis et al., 2017), the LOFAR deep surveys (Röttgering, 2010; van Haarlem et al., 2013) and LoTSS (Shimwell et al., 2019).

Chapter 4

Radio Source Extraction with ProFound

As modern radio continuum surveys (such as those described in Section 1.6) increase in sensitivity and have more combinations of both long and short baselines, continuum observations will not be limited to observing point sources and will observe more extended emission. For these radio surveys, in which thousands of galaxies will be detected, it is not only important to accurately define the flux density arising from these sources, but it is crucial to consider these galaxies within a multi-wavelength context. This is necessary to obtain redshifts and model galaxy properties, which are essential for detailed studies of these sources (see Chapters 2, 3 and 6). Whilst typical source extraction software used on radio images rely on fitting Gaussian morphologies to sources, with the increase in numbers of resolved sources and extended emission it will be important to use source extraction software that is not constrained by morphology and that works within a multi-wavelength context.

In this Chapter, I shall discuss work from Hale et al. (2019a), where I investigated whether ProFound (Robotham et al., 2018) can be used to accurately extract source catalogues from radio continuum images. ProFound was designed as an optical and

near-IR source extractor (e.g. Robotham et al., 2018; Davies et al., 2018; Turner et al., in prep) and was not designed with radio continuum observations in mind. The challenge of extracting sources at radio frequencies compared to other wavelengths arises from the correlated noise. This is due to convolution of measurements with the sampling function (see Section 1.5). This is typically less of an issue in the optical and near-IR, where noise is less correlated.

Within this Chapter, I perform three investigations to test how successfully ProFound can quantify the emission from galaxies within radio images. By using deep radio continuum observations of the XMM-LSS field; images of known extended sources and simulated images, I investigate how ProFound may be used to measure source properties within radio continuum images.

4.1 Commonly Used Source Extraction Software

Before discussing whether ProFound can be used in order to produce accurate catalogues of radio flux densities, it is important consider the other software available that is commonly used within the literature. A variety of different software is available, many of which have been compared in Hopkins et al. (2015). However, in this Chapter, I shall only discuss three commonly used software.

PyBDSF

Firstly, PyBDSF[1] (Mohan & Rafferty, 2015), has been used frequently within the literature, including for LOFAR observations (Williams et al., 2016; Hardcastle et al., 2016; Shimwell et al., 2019, and the work of Chapter 5) and for TGSS-ADR (Intema et al., 2017). It produces a smooth background sky model before identifying bright pixels above a given threshold compared to the sky. This emission (with the sky subtracted) is then modelled as a combination of Gaussians. Two thresholds are

[1] `http://www.astron.nl/citt/pybdsm/index.html`

used to define a source. Firstly, `thresh_isl` determines which pixels are included and modelled within the source, whilst `thresh_pix` determines whether a source is included within the source catalogue. This criteria to include a source within the catalogue is dependent on whether the peak flux of the source is greater than `thresh_pix`$\times rms + mean(map)$, where $mean(map)$ is the mean value of the map within the region bounded by the `thresh_isl` criteria. Sources can be output as both a list of the Gaussian components as well as a source list. In the source list, Gaussian components that are sufficiently close, relative to their flux, are combined together[2]. The commands for running PyBDSF simply using a given `thresh_isl` (isl) and `thresh_pix` (pix) for an image (image.fits) are described below. For observations at a given frequency (freq) with a clean beam defined with a major/minor axis (maj_ax/min_ax) and position angle (pa), these commands are:

```
img = bdsf.process_image(image.fits, thresh_isl=isl, thresh_pix=pix,
frequency=freq, beam=(maj_ax, min_ax, pa))
```

PyBDSF, however, also has many other parameters that can be changed in order to help it better model radio sources, including when there is extended emission (see `http://www.astron.nl/citt/pybdsm/examples.html#image-with-extended-emission`). As I wanted to test the response of the different source extraction software to extended emission, I also used PyBDSF with the settings defined in the tutorial throughout this Chapter. This specifies the extra parameters of: `flagging_opts=True`, `flag_maxsize_bm=100` (this allows Gaussians with sizes up to 100 times the beam size to be fit), `atrous_do=True` (this allows Gaussians of different scales to be fit) and finally `rms_map=False` (this specifies the rms map is unknown) and `mean_map='zero'` (this ensures the background averages to zero so that the extended emission does not contribute to a positive background). The commands to run PyBDSF with these

[2] `http://www.astron.nl/citt/pybdsm/algorithms.html#grouping-of-gaussians-into-sources`

settings (which will be referred to as the atrous version of PyBDSF) are given below.

```
img = bdsf.process_image(image.fits, thresh_isl=isl, thresh_pix=pix,
frequency=freq, beam=(maj_ax, min_ax, pa), flagging_opts=True,
flag_maxsize_bm=100, atrous_do=True, rms_map=False, mean_map='zero')
```

In previous work, PyBDSF is typically run using a value of `thresh_isl=3.0` and `thresh_pix=5.0` (e.g. Williams et al., 2016). These values were also used when it was compared to other software for a source extraction data challenge (Hopkins et al., 2015). The output source catalogue (where Gaussian components have been combined together, if appropriate) is used for this work.

AEGEAN

AEGEAN (Hancock et al., 2012, 2018) is another piece of software that fits Gaussian components to sources. It has been used for the large MWA GLEAM Survey (Hurley-Walker et al., 2017). The premise for AEGEAN is similar to PyBDSF, where pixels above a threshold are combined into a source and modelled as a Gaussian. This uses a criterion like `thresh_isl`, to determine which pixels contribute to the source model (known as `floodclip`). The complementary parameter that determines whether a source is included in the catalogue is the parameter `seedclip`. This is similar to `thresh_pix` and uses a criterion based on the peak flux compared to the rms background. However, in contrast to PyBDSF, no criterion based on the mean value of the map within the source is used. Again, the catalogue can be output as either a component or source list, whilst it can also provide these sources as a ds9 (Joye & Mandel, 2003) region file. To generate the model of the background emission, a complementary package BANE (Background and Noise Estimation; `https://github.com/PaulHancock/Aegean/wiki/BANE`) is used. Finally, the package AeRes (`https://github.com/PaulHancock/Aegean/wiki/AeRes`) is also

available to produce residual maps and source models for the catalogues generated by AEGEAN. The commands to run BANE and AEGEAN to create a catalogue using a `floodclip`/`seedclip` (`flood`/`seed`) value are:

```
BANE image.fits
aegean image.fits -background image_bkg.fits -noise image_rms.fits -
beam maj_ax min_ax pa -seedclip seed -floodclip flood -table aegean-
catalogue.fits
```

Here `image_bkg.fits` and `image_rms.fits` are the background and rms images respectively that are created by BANE. In Hopkins et al. (2015) values of `seedclip=5` and `floodclip=4` were used. In this work I make use of the Gaussian component catalogue. Although this is not what is used for PyBDSF, this choice relates to discrepancies found when using the source catalogue which had fluxes much smaller than expected.

BLOBCAT

Finally, the software used for extracting the catalogue for the data used in Chapters 2 and 3, (Smolčić et al., 2017a), is BLOBCAT. It has also been used many other times in the literature (e.g. Radcliffe et al., 2018; Butler et al., 2018). This follows a similar premise to that of PyBDSF and AEGEAN where bright pixels above a sky background are modelled as Gaussian components. However, in contrast to PyBDSF and AEGEAN, only pixels which are bright enough above a given threshold are included in the source model. This means that sources are not constrained to be Gaussian morphologies, whilst the emission within the source region is. The limitation of this software, in the context of this work, is that it does not support a SIN projection scheme[3]. The data used for these investigations (see Section 4.3) is in

[3]Note: ProFound on the other hand is unable to work on the NCP projection scheme

the SIN projection scheme and as such BLOBCAT was not used in this work.

4.2 ProFound

ProFound[4] (Robotham et al., 2018) is a recently developed source extraction software package, coded in R, that uses the extraction of pixels in order to determine sources in an image and measure their properties, without the assumption of a morphology. As with PyBDSF and AEGEAN, the sky is first modelled before bright pixels that are above a given sky value (`skycut`) are assigned as a source (or 'segment' in ProFound)[5]. Below I give an overview of the method ProFound uses to create a source catalogue and is paraphrased from the full details given in Robotham et al. (2018).

1. *Determine an initial sky model:* Grid up the image and use a median box-car method to determine the average sky properties across the image.

2. *Create an initial source model:* Use a user designated threshold to determine an initial source catalogue by finding those pixels that are above a given threshold compared to the sky background.

3. *Improve upon the sky model:* Use the sources from Step (2) to remove emission believed to be real and then repeat Step (1) to make a better model of the sky.

4. *Source properties:* Measure the source properties for the segments found in Step (2).

5. *Repeat until convergence:* Repeat the processes in Steps (1), (2) and (3) through dilating the segments until the flux have converged.

[4] `https://github.com/asgr/ProFound`

[5] Each segment includes the pixels that have started from the bright pixel which initiated the segment and those pixels that have grown outwards from this and remain above the threshold limit.

6. *Make final source catalogue:* Measure the final properties for the final segments and generate a catalogue of sources from this.

The dilation process which is used by ProFound selects bright pixels first before dilating the segment into the pixels surrounding the segment, provided they contain sufficient flux above the sky background level. This is therefore not restricted to dilating segments in certain directions in order to adhere to certain morphologies, the only restriction being that segments are unable to dilate into other segments. This means that, for example, an FRII type source with a bright core and connected emission along the jets may be split into three neighbouring segments as there may be one segment originating from the core and two segments originating from bright hotspots.

It is possible, however, to ensure that neighbouring segments that touch can be combined into a single source. This is by setting the parameter `groupstats=TRUE`. Continuum images at the current flux limits have a low enough source density and are sufficiently high resolution that the images are not near confusion. This can be seen in Figure 4.1 where the comparison of the same patch of sky at near-IR (from VIDEO; Jarvis et al., 2013) to radio (from Heywood et al., in prep) can be seen. The low source density in the radio is obvious from these images. Therefore with current radio continuum images, `groupstats` can be set to `TRUE`. This will not necessarily be true with future observations from surveys such as CHILES Con-Pol (Hales & Chiles Con Pol Collaboration, 2014), MIGHTEE (Jarvis et al., 2017) and the SKA (see Section 1.6). The commands used to run ProFound are:

```
image_blind=profoundProFound(image, plot=FALSE, skycut=scut, rot-
stats=TRUE, boundstats=TRUE, nearstats=TRUE, groupstats=TRUE,
groupby='segim', verbose=TRUE)
```

This produces a sky (`image_blind$sky`), rms (`image_blind$skyRMS`) and segment

map (`image_blind$group$groupim` if grouped or `image_blind$segim` if not). A corresponding model for grouped sources can be generated through:

```
segim_model=image_blind$group$groupim
segim_model[image_blind$group$groupim!=0]=1
segim_model[image_blind$group$groupim==0]= as.numeric(NaN)
model=(image$imDat-image_blind$sky)*segim_model
```

4.2.1 The Challenges for ProFound with Radio Continuum Images

For radio continuum images (as described in Section 1.5) the visibilities are convolved with the dirty beam and, although the sources will be cleaned in the data reduction process, the background emission will still exhibit structure. As point sources are correlated on the same sizes as unresolved sources, it is likely that, near the flux limit, noise may be confused as real emission. In the optical and near-IR, the noise is typically less correlated, see Figure 4.1, and so the distinction of sources from the background is easier (at a given threshold limit). Therefore with near-IR and optical images, smaller `skycut` values can be used. When using ProFound on radio continuum images, it is therefore crucial to ensure that a sufficiently large `skycut` is used so that the extracted catalogue does not contain large numbers of noise spikes. Too large a `skycut`, though, will mean that real sources will be missed from the catalogue and large fractions of the flux of sources will not be contained within the associated segment.

When an appropriate `skycut` has been determined, it is important to remember that the sum of the pixel fluxes measured within a segment by ProFound in radio maps will not correspond to the total flux density of a source. Instead two corrections, described below, need to first be applied.

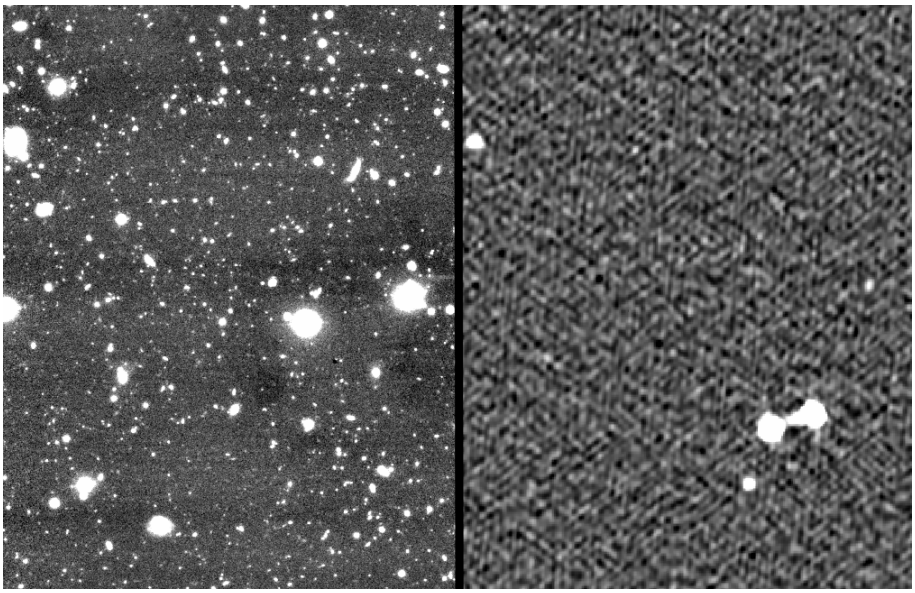


Figure 4.1: Comparison images of the same field of view in the near-IR from VIDEO (left; y-band) (Jarvis et al., 2013), and radio (right; at 1.5 GHz) (Heywood et al., in prep).

4.2.1.1 Map Pixel Values to Total Flux Densities

Radio continuum images have pixel values given in the unit Jy/beam. Therefore, in order to obtain total flux densities in Jy, a conversion factor needs to be applied. With source extraction software such as AEGEAN and PyBDSF, this is applied when the source catalogue is generated. This uses the relationship between the peak values of each Gaussian component as well as its major and minor axes, to generate the total flux:

$$S_{peak} = \frac{S_{int} \times 2 \times \pi \times b_{maj} \times b_{min}}{2 \times \pi \times \sigma_{maj} \times \sigma_{min}}, \tag{4.1}$$

(see Hales et al., 2012). This relates the peak flux (S_{peak}), total flux (S_{int}), major/minor axes of the beam (b_{maj}/b_{min}) and major/minor axes of the source ($\sigma_{maj}/\sigma_{min}$).

However, ProFound does not assume a Gaussian morphology and so Equation 4.1 should not be used. Instead the pixel flux densities per beam in each segment are summed together and a correction needs to be applied to convert from Jy/beam to

Jy. This conversion[6] is given by:

$$S(\mathrm{Jy}) = \frac{\Sigma(S_i)}{N_{pixels\ per\ beam}} = \frac{\Sigma(S_i)}{\frac{2\times\pi\times b_{maj}\times b_{min}}{8\times ln(2)\times\delta_{RA}\times\delta Dec}} = \Sigma(S_i)\frac{8\times ln(2)\times\delta_{RA}\times\delta Dec}{2\times\pi\times b_{maj}\times b_{min}}, \quad (4.2)$$

where $\Sigma(S_i)$ is the sum of all the pixel fluxes in Jy/beam and δ_{RA} and δ_{Dec} are the angular size of the pixels in the RA and Dec directions.

4.2.1.2 Beam Correction

Furthermore, when PyBDSF and AEGEAN calculate the total flux density of a source using Equation 4.1, this inherently accounts for the fact that not all of the flux of the object may be included within the island region. For these, only an accurate estimation of the peak flux and major and minor axes of the source is necessary, which can be generated by modelling a Gaussian component (or multiple components) within the island. ProFound on the other hand cannot do this as no morphology is assumed. Therefore, ProFound may not fully sample the full extent of a source and may underestimate the total flux density. This will most affect faint sources, where much of the emission is below the threshold limit, as well as those sources which are unresolved.

To correct for this, the shape of the Gaussian clean beam can be used to quantify what fraction of the flux is enclosed within the segment. This is illustrated in Figure 4.2. In Figure 4.2, Panel 1 shows an example segment mask for a source (1 within the segment and 0 outside it) and the Gaussian beam is shown in Panel 2. The result of multiplying (1) and (2) together is shown in Panel 3. The sum of the total flux within the masked beam (Panel 3) can then be compared to the total flux within Panel (2) to quantify the fraction of the beam flux that is missing. The correction that needs to be applied to the total flux densities is given by:

[6] as discussed in
`https://newton.cx/~peter/2011/12/reference-the-ultimate-resolved-source-cheatsheet/`

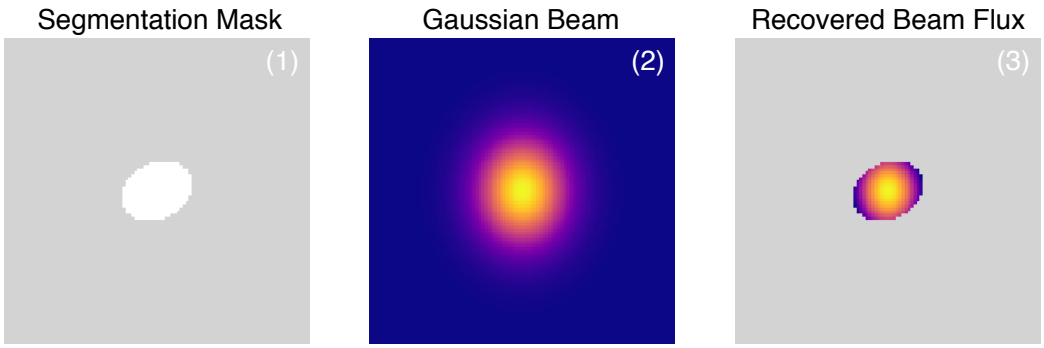


Figure 4.2: Illustration of the method in which beam corrections are applied to a source of a given segmentation map. The mask of the segment is shown in panel (1) and the Gaussian beam is shown in panel (2). By comparing the flux of the Gaussian beam contained within the segmentation mask (panel 3) a correction factor can be determined.

$$S(corrected) = S(\mathrm{Jy}) \times \frac{\Sigma(S_{i,Panel2})}{\Sigma(S_{i,Panel3})} \tag{4.3}$$

These corrections will technically only be valid for unresolved sources and not for extended sources, especially those close to the flux detection limit. However the corrections will be minimal for those extended sources which are bright and contain the majority of their flux above the detection level.

4.3 Data

In order to test how successfully ProFound extracts radio sources it is essential to see how well ProFound can be applied to real observations. However, as the true underlying sky distribution of sources and their fluxes is unknown, this does not allow comparisons of the input and recovered flux densities. Therefore, in this work, I compare how well ProFound runs on both existing continuum observations as well as on simulated data. For comparisons of its success on continuum images, I used images of the sources extracted, their flux density distributions, as well as the residuals as indicators of ProFound's success. I also created simulated mock images

to compare the accuracy of measured fluxes. The data used for this work is described below.

4.3.1 Blind Continuum Data

Firstly, it is important to consider radio continuum images of large fields which will allow a variety of morphologies to be observed. Therefore, I made use of recent VLA observations of the XMM-LSS field, from Heywood et al. (in prep). These observations were made using the VLA, observing a total area of $\sim$7.5 deg^2 at 1.5 GHz to a depth of $\sim$16 μJy/beam rms at 4.5" resolution through combining 32 pointings. This field was chosen to overlap with near-IR observations with the VISTA telescope in the VIDEO Survey (Jarvis et al., 2013), as well as the wealth of ancillary multi-wavelength data that exists over the field (see e.g. Pierre et al., 2004; Tasse et al., 2007; Mauduit et al., 2012; Oliver et al., 2012; Davies et al., 2018; Aihara et al., 2018; Chen et al., 2018). For this work, I used the 1.2 $\times$ 1.2 deg^2 of this field where the central 1 deg^2 of the field overlaps with CFHTLS Deep 1 Field (CFHTLS D1; Cuillandre et al., 2012; Hudelot et al., 2012, `http://www.cfht.hawaii.edu/Science/CFHLS/T0007/`, centred at (36.5°, -4.5°)). This can be seen in Figure 4.3. This field has both a mixture of compact, Gaussian sources as well as extended morphologies including AGN with jets. This diverse range of morphologies, as well as the sensitivity of these observations, made it an ideal image in order to test ProFound's success on blind continuum imaging.

4.3.2 Known Extended Sources

However, as deep continuum field imaging is likely to be dominated by compact objects, it is also important to specifically see how ProFound works on known extended morphologies. As ProFound does not specify source morphologies, tracing complicated non-Gaussian emission is where the benefit of ProFound is likely to

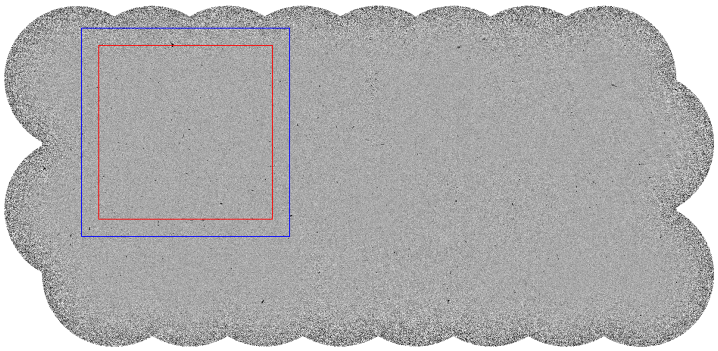


Figure 4.3: The full field (7.5 deg^2) of observations from Heywood et al. (in prep) of the XMM-LSS field using a flux scale between a range of -5×10^{-5}Jy to 10^{-4}Jy. The blue box indicates the 1.2 × 1.2 deg^2 region which are used these investigations. The red region is the central deg^2 for which the sources extracted will be considered over.

Source	Reference	Resolution (")	Frequency (MHz)	Image Size (arcsec^2)
3C16	Leahy & Perley (1991)	1.25	1477	76.8×76.8
3C19	`http://www.jb.man.ac.uk/atlas/object/3C19.html`	0.15	1534	11.5×11.5
3C28	Feretti et al. (1984) `http://www.jb.man.ac.uk/atlas/object/3C28.html`	1.10	1424	76.8×76.8
3C42	Leahy & Perley (1991)	1.20	1477	76.8×76.8
3C47	Leahy (1996)	1.00	1650	76.8×76.8

Table 4.1: Citations and information on the 3C sources used in Section 4.5.2 from `http://www.jb.man.ac.uk/`. Also quoted is the frequency of these observations (all at ∼1.5 GHz) and the respective angular resolutions of each observation.

lie. Therefore the results of using ProFound on continuum imaging of sources from the 3C surveys (see Section 1.6; Edge et al., 1959; Laing et al., 1983) is also investigated. FITS images of five sources were downloaded from `http://www.jb.man.ac.uk/atlas/` which contains information and images on AGN from Laing et al. (1983) (3CRR sample). These five sources were chosen to given an idea of how the different software handled extended morphologies. The sources chosen were 3C16, 3C19, 3C28, 3C42 and 3C47 and information on these sources is given in Table 4.1.

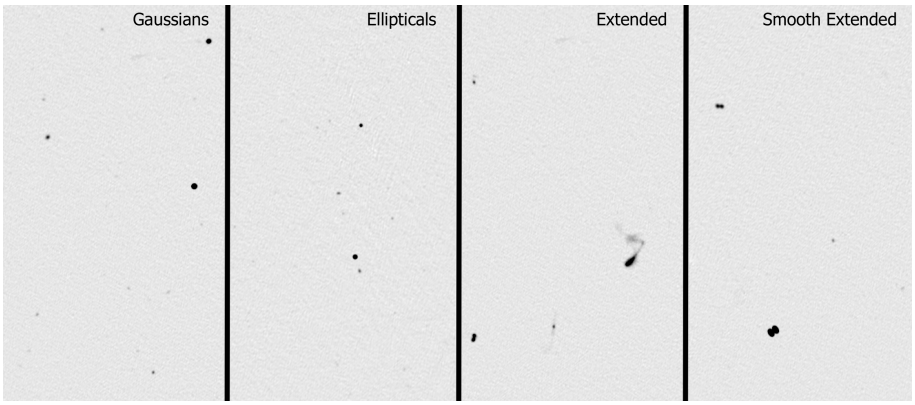


Figure 4.4: Cut-out images ($\sim 0.13^{\circ} \times 0.25^{\circ}$) from the four simulated radio images for Gaussians (left); ellipses (centre left); complex extended morphologies (centre right) and smooth extended lobed objects (right).

4.3.3 Simulated Data

Finally, it is essential to use simulated radio data in order to compare how accurately the flux densities of sources are being recovered. As described above, real continuum images have a multitude of source morphologies, all of which ProFound and any other source detection software need to be able to accurately measure in order to produce true catalogues of the sources within the images. As different morphologies may be favoured by one source extraction software compared to another, I constructed four simulations each with a differing morphology. However as mentioned in Section 4.2.1, the noise in radio images is known to be correlated, therefore it is challenging to create a realistic sky background. As such, I used the residual image generated when PyBDSF is run over the image (using the default parameters discussed in Section 4.1) from Heywood et al. (in prep) in order to have a realistic background. All simulated sources can then be added into this image to create mock images of the radio skies. The morphologies in the four simulations were: Gaussian sources; elliptical disks convolved with the beam; sources with complicated extended morphologies and finally extended sources with smooth morphologies. These are all described in more details below and example cut-outs from each of these are shown in Figure 4.4.

- *Gaussians:* Gaussians of varying sizes, position angles and flux densities were modelled and injected into the image. Flux densities were generated by randomly selecting a 1.4 GHz flux (and scaled to 1.5 GHz) from the S^3 (Wilman et al., 2008) catalogue (described in Sections 1.4.1 and 2.2.1). Position angles were randomly assigned to each Gaussian component (uniformly sampled from 0 to 2π) whilst major/minor axes were generated through sampling the distribution of sizes from the catalogue of sources extracted by PyBDSF when run over the VLA image, using the default parameters described in Section 4.1. These were all ensured to be larger than the beam. Flux densities from S^3 were restricted to only include those with total flux densities $> 3\sigma_{map}$, where σ_{map} is taken here as 16 μJy. This is the typical noise in the image converted to a flux density under the assumption of being a point source. 1000 Gaussians were simulated in the image. The distribution of flux densities and major and minor axes of the sources can be seen in Figure 4.5(a).

- *Elliptical disks:* Similar to the Gaussian simulations, 1000 sources were injected into the image with flux densities assigned from S^3. In this case elliptical components were convolved with the Gaussian beam. The sizes of the elliptical components were the component sizes associated with the source's flux from Wilman et al. (2008).[7] The distribution of flux densities and sizes can be seen in Figure 4.5(b).

- *Complex extended morphologies:* The source models that ProFound generated from the VLA XMM-LSS image (to be described in Section 4.5.1) for large sources were re-injected into the image. 'Large' here is defined such that `R50`$\geq$3.5, with `R50` the approximate semi-major axis containing 50% of the flux. This corresponded to 105 sources within the central square degree. These were

[7]Note: some sources in S^3 have multiple components. Here though, only individual components irrespective of whether the component was part of a multi-component source were used here.

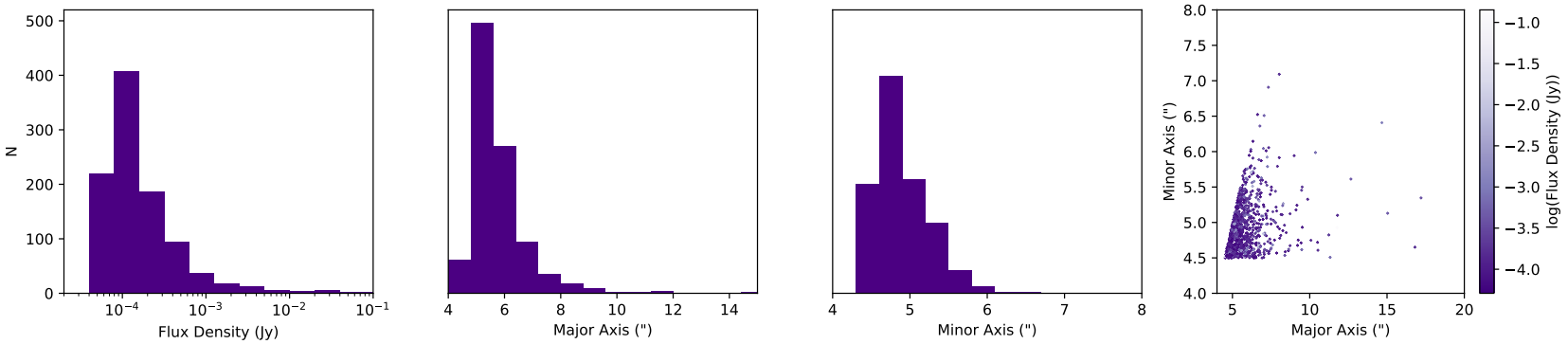
injected into the image a total of 5 times at random positions and the flux multiplied by a factor randomly sampled logarithmically between 10^{-2} and 1 in order to make the sources more comparable to the noise, and make re-extraction of sources more challenging. As the recovery of sources was small in these simulations, each simulation was repeated 5 times and the results combined together in order to improve understanding of the recovery fraction, given that the morphologies can be complex.

- *Extended Lobed Objects:* Similar to the disk simulations, extended sources generated from elliptical components from Wilman et al. (2008) were injected into the image. However, in this case, only multi-component sources (i.e. FRI and FRII type sources) were injected. 500 lobed sources were injected into the image.

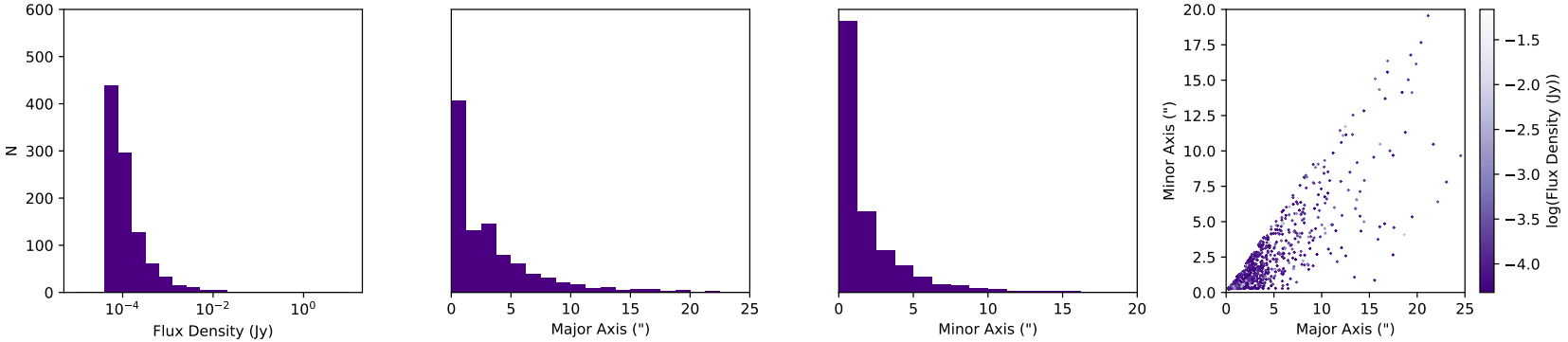
4.4 Source Detection Parameters

Whilst ProFound has been used for near-IR and optical images (see Robotham et al., 2018; Davies et al., 2018; Turner et al., in prep), the source threshold parameters will need to be larger when detecting sources in radio continuum imaging for the reasons discussed in Section 4.2.1. In order to identify which parameters are most appropriate for ProFound, I considered the false detection of sources across the VLA image described in Section 4.3.1. The noise in the image is thought to be symmetric, that is for each positive noise spike there is likely to be a corresponding negative spike. Therefore by comparing the number of sources detected in the negative image (i.e. the -1 $\times$image; hereafter referred to as inverted) to those in the real image (see Figure 4.6), the fraction of false detections can be quantified and the percentage of real detections found from:

$$\%\ Real\ Detections = 100 \times \frac{N_{image} - N_{inv.\ image}}{N_{image}}. \tag{4.4}$$



(a) Gaussian simulations



(b) Elliptical simulations

Figure 4.5: The input flux density distribution (left); major axis distribution (centre left); minor axis distribution (centre right) and comparison of the source sizes and flux densities (right) for the simulated images with (a) Gaussian morphologies and (b) elliptical morphologies convolved with the beam.

The parameter `skycut` is used by ProFound in order to determine which pixels can form segments based upon their flux above the sky background. The corresponding percentage of real detections can be seen in the left hand panel of Figure 4.7 as a function of this `skycut` parameter. This suggested a `skycut` value of 3.5 should be used, as this gave a false detection percentage of 2% and so was a good balance between extracting a large number of sources, whilst also ensuring it is a relatively clean sample of real sources. When the same analysis is run on PyBDSF, now as a function of both `thresh_isl` and `thresh_pix`, the percentages of real detections can be seen in Figure 4.7 (centre). To have a similar number of false detections as with ProFound the `thresh_isl`/`thresh_pix` values should be taken as either 2.5/5.0 or 3.0/4.5. The 3.0/4.5 combination is used here due to its lower `thresh_pix` threshold. This will allow a larger number of faint sources to be detected. For AEGEAN instead, despite having the two fitting parameters, `seedclip` and `floodclip`, `floodclip` does not influence whether a source is included within the catalogue. As such the percentage of real detections of AEGEAN is only calculated as a function of `seedclip`. From Figure 4.7 (right) it can be seen that a value between 4.5 and 5.0 is appropriate. I therefore chose to use a value of `seedclip`=4.5 and `floodclip`=4.

For the rest of the work described in this Chapter I will consider how successful ProFound, AEGEAN and PyBDSF are using the values described above as well as using the default parameters described in Section 4.1 for PyBDSF and AEGEAN. Where extended emission is likely to be present, I also will compare to PyBDSF using the extended emission settings (described in Section 4.1), this will be referred to as the atrous setting.

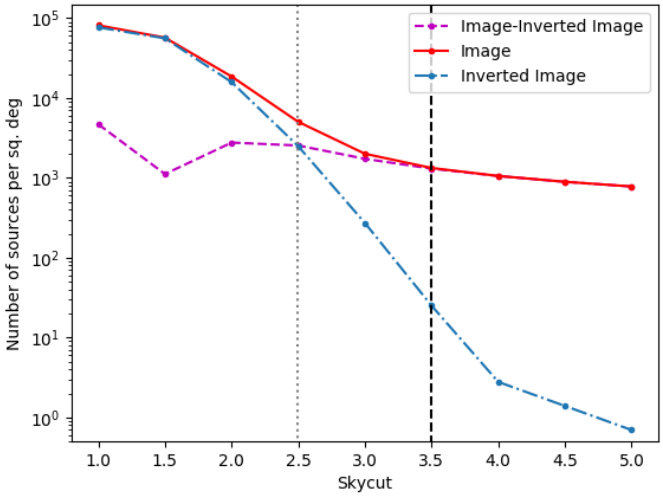


Figure 4.6: The number of source detections with ProFound as a function of `skycut`. This is shown for the real (red) and inverted (blue) image. The difference between detections in the real and inverted images is also shown (magenta). The grey dotted line indicates where only half of the sources detected are real. Finally, the black dashed line indicates the `skycut` value that will be used for this work (a value of 3.5), which had ∼98% real detections.

4.5 Results

Next, I discuss the response of ProFound on the three different datasets discussed in Section 4.3.

4.5.1 Blind Continuum Imaging

The histogram of flux densities for sources within the central square degree of the VLA image from Heywood et al. (in prep) (see Section 4.3.1) can be seen in Figure 4.8 for ProFound, PyBDSF and AEGEAN. The right hand panel of Figure 4.8 corresponds to when the default parameters discussed in Section 4.1 are used for PyBDSF and AEGEAN. The left hand panel, on the other hand, uses the parameters determined in Section 4.4. For ProFound, both sources with and without a beam correction (Section 4.2.1.2) are shown in Figure 4.8 and PyBDSF is shown

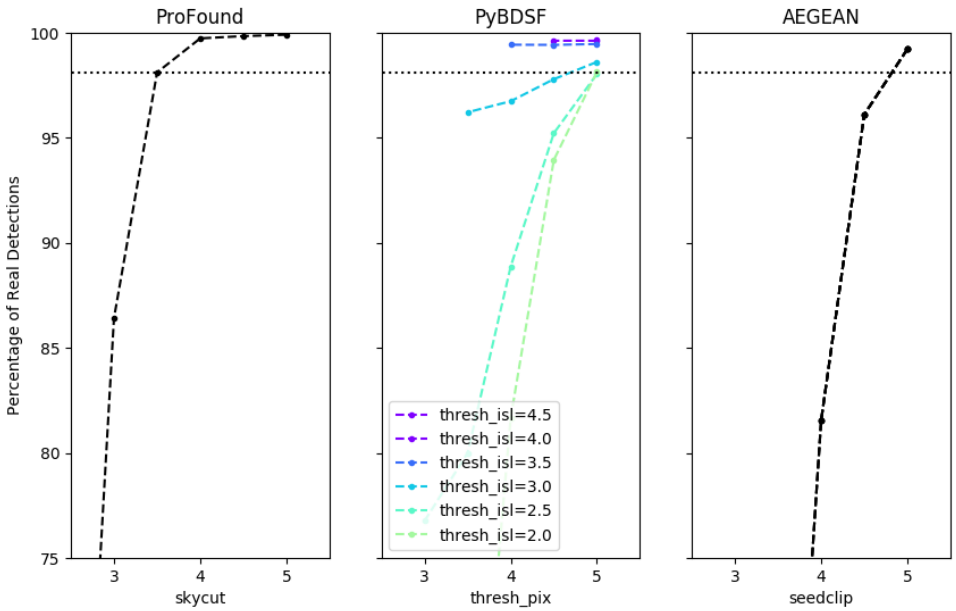


Figure 4.7: The percentage of real detections for the different source extraction software used when their respective detection parameters are varied. This is shown for ProFound (left), PyBDSF (centre) and AEGEAN (right).

both with and without the atrous setting.

ProFound extracted a total of 1,360 sources over the central square degree. On the other hand, PyBDSF found a total of 1,122 sources when the default parameters were used and 1,280 sources when the parameters determined in Section 4.4 were used. AEGEAN on the other hand found a total of 1,192 sources using the default parameters and 1,484 using the parameters from Section 4.4. When the atrous setting was used, PyBDSF instead found 1,332 sources using the default parameters and 1,692 sources using `thresh_isl/thresh_pix` = 3.0/4.5. These differences in the number of sources detected will not only be affected by the different detection thresholds used by the different software, but also on whether a source has been split into multiple components or whether it has been combined into a single source. ProFound is more likely to merge single source components together if segments touch, however this will not necessarily be the case with PyBDSF and, as the components catalogue

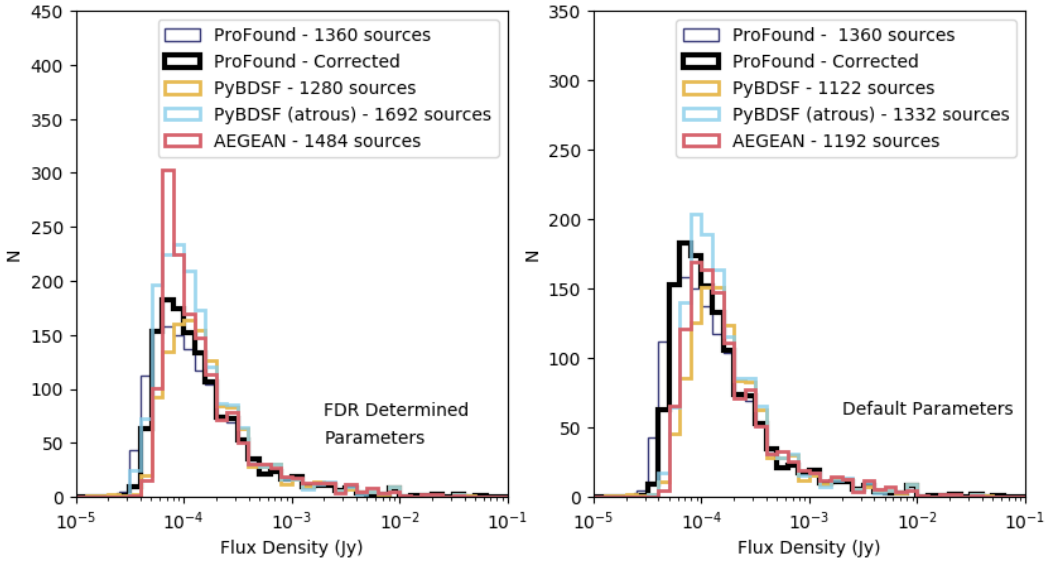


Figure 4.8: Histograms of the measured flux density distributions of sources extracted from the central square degree of the radio continuum image from Heywood et al. (in prep), as discussed in Section 4.3.1, with ProFound (thin dark blue, uncorrected and thick black line, corrected), PyBDSF (yellow and light blue when the atrous setting is used) and AEGEAN (red). The left hand panel shows the measured flux densities when the parameters found in Section 4.4 are used. The right hand panel instead shows the results using the default parameters (see Section 4.1).

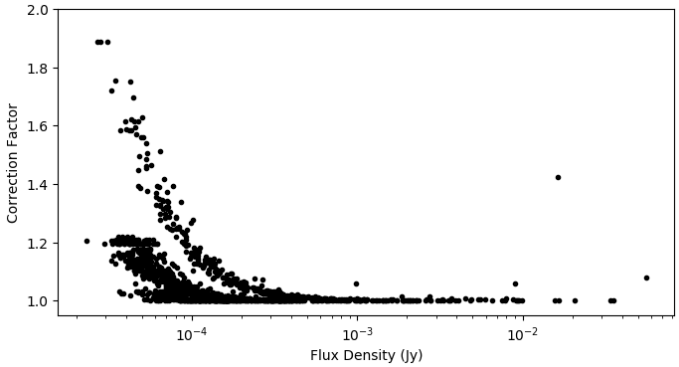


Figure 4.9: Corrections that are applied to the sources extracted with ProFound over the image discussed in Section 4.3.1 to correct for not fully dilating to the full extent of the beam as a function of flux density.

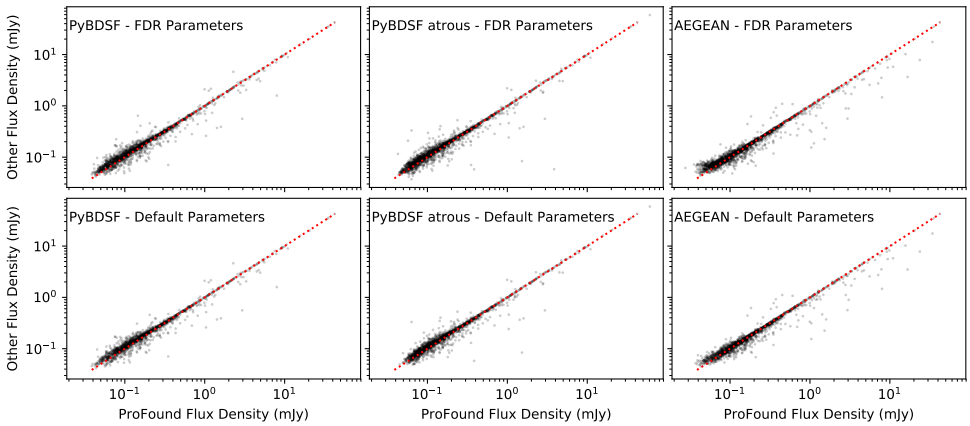


Figure 4.10: Comparisons of the fluxes measured by ProFound to sources in PyBDSF (left and centre for atrous) and AEGEAN (right) that are matched within 5". The upper panel corresponds to when the parameters from Section 4.4 are used and the lower panel to when the default parameters are used.

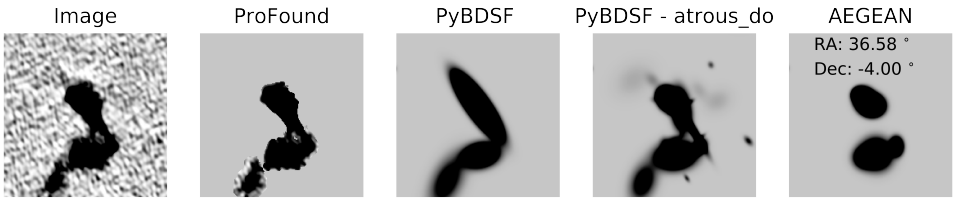
from AEGEAN is used, will not be the case with AEGEAN.

Figure 4.8 shows that sources extracted using ProFound are skewed towards lower flux densities if a beam correction is not applied (dark blue). Applying the corrections (shown in black), which are shown as a function of flux density in Figure 4.9, increases the flux density measurement of many of the faint sources. Figure 4.9 confirms that the beam corrections needed for bright objects are typically small, but these can increase up to factors of ~ 2 for the faintest galaxies. The striping in the distribution of these correction factors, that can be seen in Figure 4.9, arise due to the discretisation of pixels that are included within the segment, i.e. pixels can only be included in whole numbers. The corresponding comparison of flux densities measured for sources matched within 5" between the catalogues observed with ProFound, PyBDSF and AEGEAN can be seen in Figure 4.10. This angular separation was chosen to compromise that sources will be a mixture of unresolved (which should use matching separations $\lesssim$ the beam size) and resolved sources (which may need larger matching separations). The sources where ProFound finds a much higher flux may consist of objects that have been merged into a single source by ProFound but are

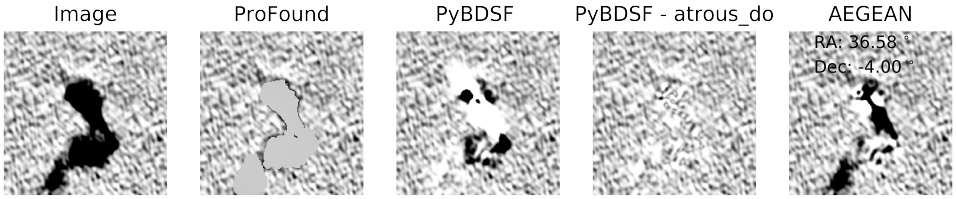
represented by multiple components in PyBDSF and AEGEAN and so their fluxes will be underestimated.

Comparisons of the models determined by the different software for three extended sources can be seen in Figure 4.11. In Figure 4.11, the residuals (image - model) for each source is also shown in the lower panel. What is important to note is that ProFound will include noise within the source model. This is because a smooth sky distribution will be determined using the median box car method. As such, when this background is subtracted from the pixels within the segment that define a source, noise peaks and troughs will be included within the source model. On the other hand, PyBDSF and AEGEAN assume a smooth Gaussian model and so the noise structure will be included in the residuals. Provided that the noise is symmetric and there are not regions with large artefacts, this should not affect the flux density measurements made with ProFound. This is because the noise would sum to zero and not affect the total flux density of the source. Whether this is the case can only be determined if the true flux densities of these sources were known, therefore if the noise does not sum to zero, this will become clear when the simulated images are used (Section 4.5.3). However, the advantage of ProFound can be seen in Figure 4.11 for its ability to fully trace the emission of these sources. It also cannot over-fit components by forcing a source model (which can be seen in Figure 4.11 by the white regions). Finally, ProFound does not produce the extended haloes which can be seen around source 1 of Figure 4.11.

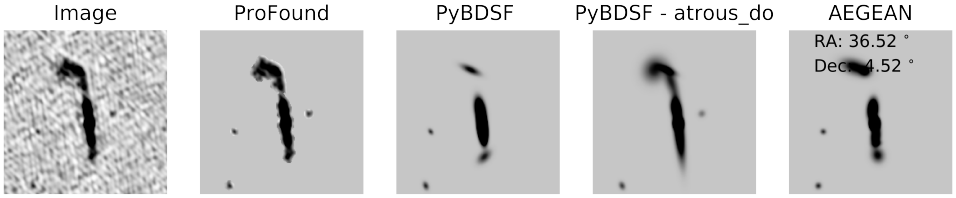
A final comparison was made through investigating the residuals that are left behind once the source model is removed from the image. These residuals should describe the noise properties of the image and should be Gaussian in distribution (e.g. Zwart et al., 2015b). However, the region used for these investigations is noisier in the upper half of the field as the image is part of a mosaic. The upper part of the image is at the edge of this mosaic and so is inherently noisier. Therefore I do not



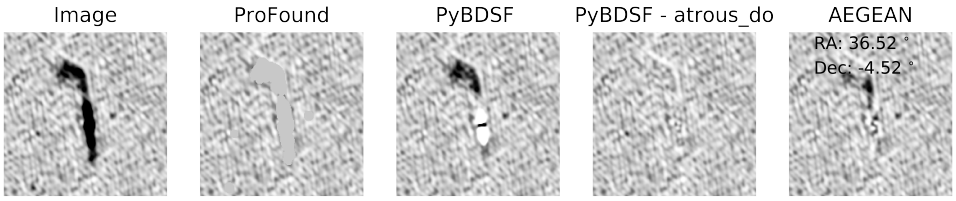
Source 1: Image and models.



Source 1: Image and residuals.



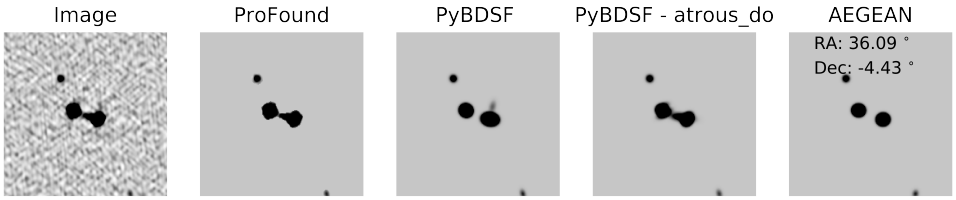
Source 2: Image and models.



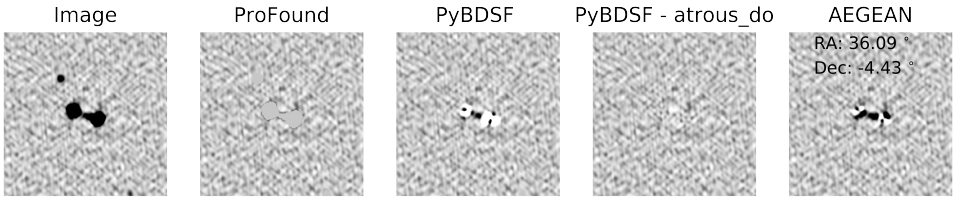
Source 2: Image and residuals.

Figure 4.11: Examples of source models for three extended sources extracted from the observations in Heywood et al. (in prep) (panel 1) and their residuals when sources are extracted using ProFound (panel 2) compared to PyBDSF (panel 3 and 4) and AEGEAN (panel 5). For PyBDSF and AEGEAN the parameters of Section 4.1 are used. The flux density per beam scale for each of the images is the same and ranges between -0.05 mJy/beam and 0.1 mJy/beam.

expect the residuals to be modelled by a single Gaussian component. The histogram of the pixel residuals are shown in Figure 4.12. For the residuals from ProFound this just quantifies the sky distribution within the segments, plus the background distribution where there are no sources. In Figure 4.12, the residuals are shown for the three source extraction software package and for PyBDSF is shown both with and without the atrous setting. A multi component Gaussian model is also shown in grey, which is found by fitting the negative residuals from ProFound with multiple Gaussian components of variable amplitude and σ. The component of this with the smallest noise has an rms value of ~ 0.016 mJy, which is the average rms value of the image (as described in Section 4.3.1).



Source 3: Image and models.



Source 3: Image and residuals.

Figure 4.11: Continued...Examples of source models for three extended sources extracted from the observations in Heywood et al. (in prep) (panel 1) and their residuals when sources are extracted using ProFound (panel 2) compared to PyBDSF (panel 3 and 4) and AEGEAN (panel 5). For PyBDSF and AEGEAN the parameters of Section 4.1 are used. The flux density per beam scale for each of the images is the same and ranges between -0.05 mJy/beam and 0.1 mJy/beam.

From Figure 4.12, it can be seen that ProFound is able to successfully construct residuals that are symmetric in nature, which suggest that ProFound is successfully

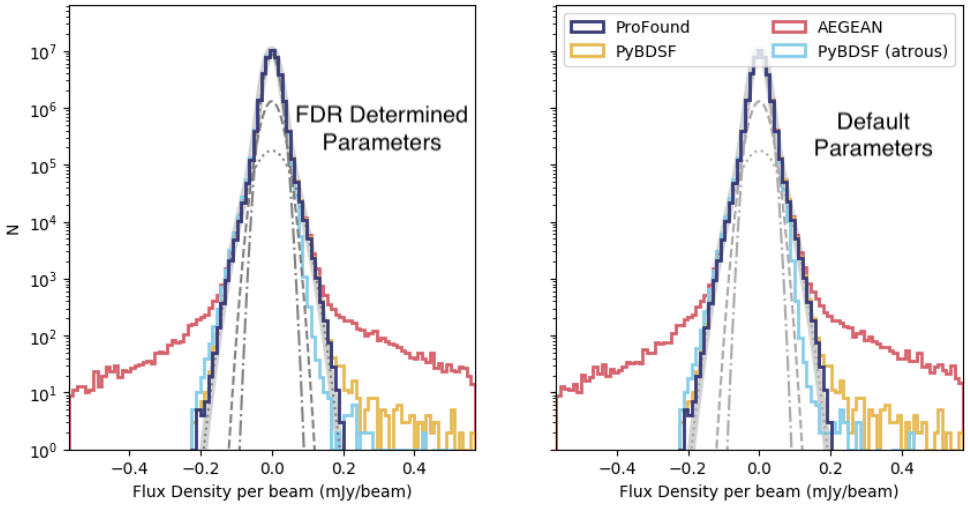


Figure 4.12: Histogram of the pixel residuals that remain in the continuum image from Heywood et al. (in prep) when the sources extracted with ProFound (dark blue) , PyBDSF (yellow, and light blue with the atrous setting) and AEGEAN (red) are removed. The left hand panel corresponds to when the parameters determined in Section 4.4 are used and the right uses the default parameters described in Section 4.1. The grey lines indicate the Gaussian components used to model the noise, with the sum of these Gaussian components shown in the grey solid line.

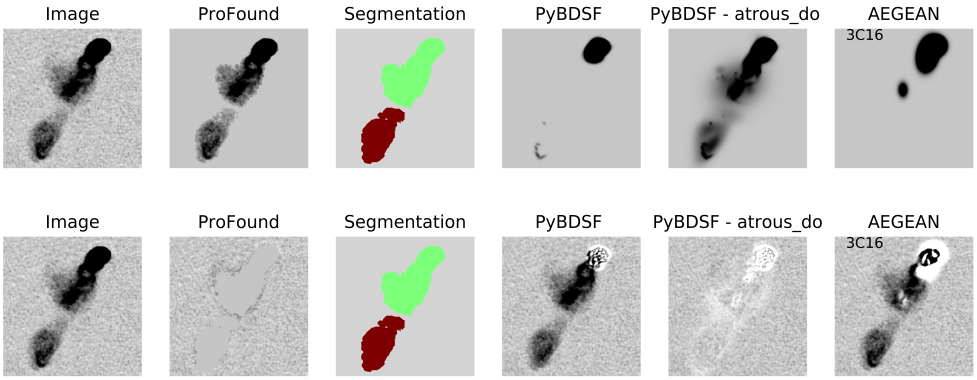
identifying the pixels that are associated with real emission and so these have flux residuals ∼0 Jy/beam in Figure 4.12. For AEGEAN there is both an excess of positive and negative residuals. The negative residuals suggests that there has been an over estimation of emission, whilst a positive excess suggests under fitting. For PyBDSF, on the other hand, there is only a small positive tail of residuals which is improved upon when the atrous setting is used as it it able to better fit extended emission. However with this setting, the residuals are asymmetric, which may suggest the flux from sources has been overfit.

4.5.2 3C Sources

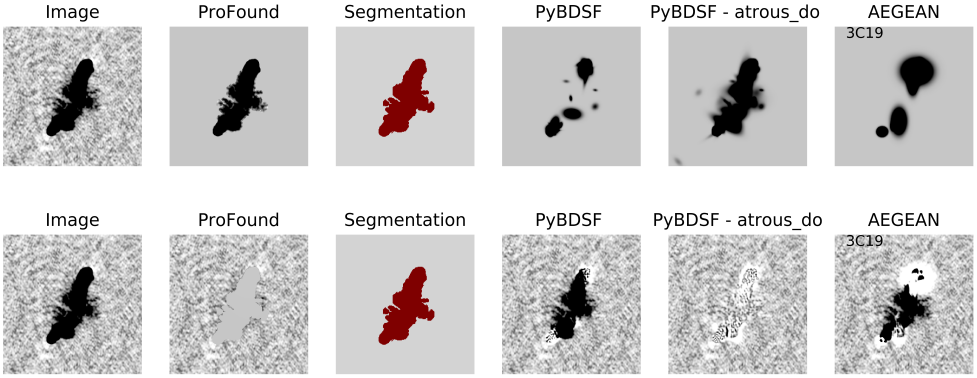
Next, I consider how ProFound handles extended emission using observations of 3C (e.g. Laing et al., 1983) sources. These observations were described in Section 4.3.2 and the details of the observations used given in Table 4.1. The fitting of these sources

using ProFound, PyBDSF and AEGEAN can be seen in Figure 4.13. For each source investigated, this corresponds to two rows in Figure 4.13 with the models given in the top row and the residuals in the bottom row. The PyBDSF and AEGEAN models are shown using the default parameters described in Section 4.1 and used in Hopkins et al. (2015). These were used as the parameters determined in Section 4.4 are not necessarily optimised for this image. The (grouped) segmentation map produced by ProFound is also shown in Figure 4.13, with each different segment (see 3C16 and 3C47) represented by a different colour.

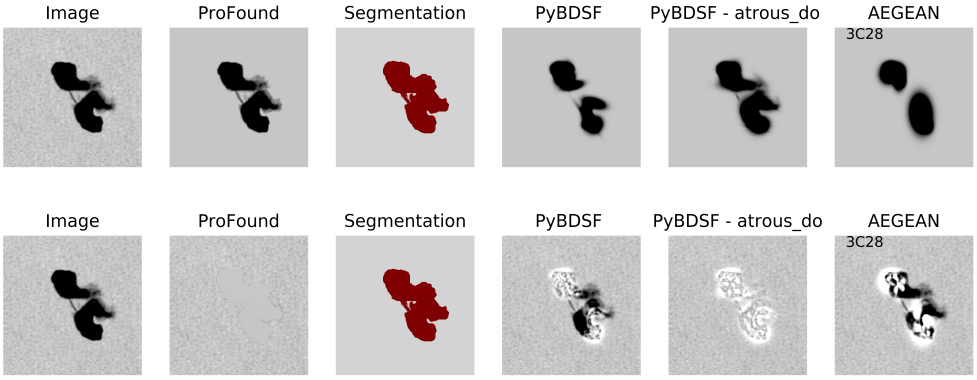
Figure 4.13 shows that ProFound is able to successfully trace the morphology of the different sources which, for AEGEAN and PyBDSF (without the atrous setting), is challenging. This is especially seen for 3C16 and 3C47 in Figure 4.13. It is also important to note that even though PyBDSF with the atrous setting is able to better trace the emission regions of these sources, it does appear to over estimate the flux in components by assuming the sources can be constructed with Gaussians. This is shown by the white (negative) regions in the residual images which are shown in the lower panel for each source. However, as discussed in Section 4.5.1, the model that is found by ProFound will contain noise and the residuals across the segment will be the smooth sky background. This smooth background is visible in Figure 4.13 however the variation across the sources is not visible due to the dynamic range used. This smooth background can be seen as an excess around zero in Figure 4.14, which shows the histogram of the residuals. These residuals, however, demonstrate that ProFound is successful in recovering a Gaussian distribution of residuals suggesting it is accurately identifying the real emission. For both PyBDSF and AEGEAN, on the other hand, there is evidence for both over and under-fitting of components within the residuals histograms. This suggests that both are not as successful at removing the extended emission compared to ProFound.



(a) 3C16



(b) 3C19

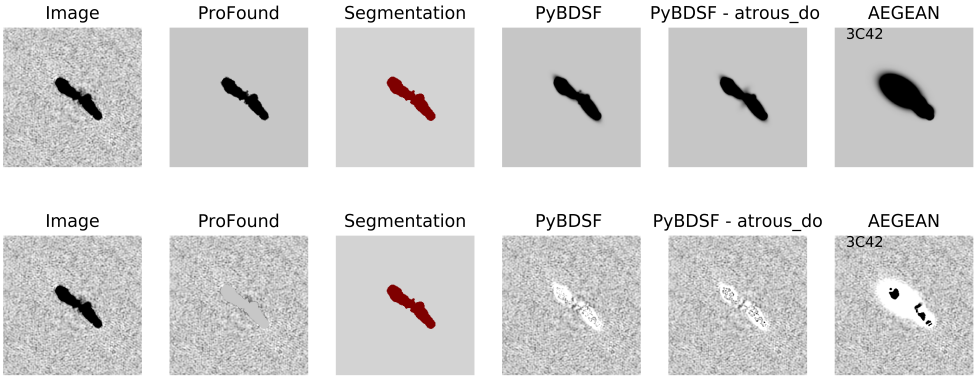


(c) 3C28

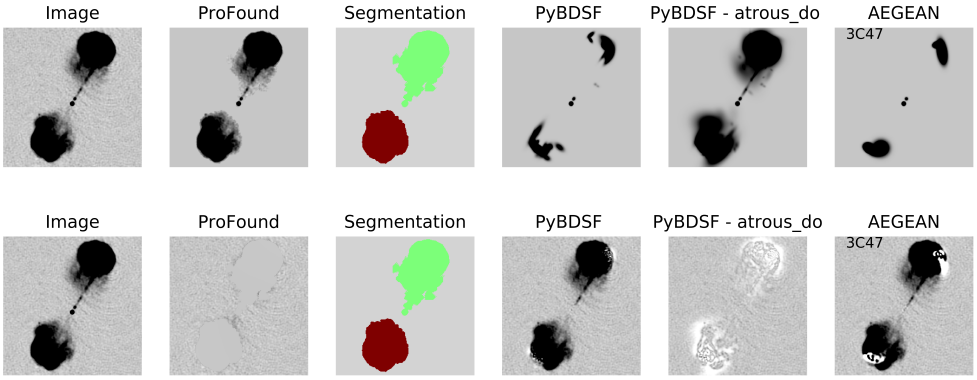
Figure 4.13: Comparison of the model (top row) and residual (bottom row) images from ProFound (panel 2), PyBDSF (panel 4 and panel 5 with the atrous setting) and AEGEAN (panel 6) of five 3C sources (whose images are shown in panel 1). These are shown for the sources: 3C16 (a), 3C19 (b), 3C28 (c), 3C42 (d) and 3C47 (e). The segmentation images from ProFound are shown in panel 3. Here PyBDSF and AEGEAN use the default parameters as described in Section 4.1, and used in Hopkins et al. (2015). Shown are the entirety of the images that are fit.

4.5.3 Simulations

Finally, although the investigations in Sections 4.5.1 and 4.5.2 are good for visually comparing whether ProFound is able to successfully model emission in radio continuum observations, it is important to quantify its accuracy in determining the flux density of a source. As discussed in Section 4.3.3, I created simulated images with four different morphologies: (a) Gaussian sources (b) elliptical disks convolved with



(d) 3C42



(e) 3C47

Figure 4.13: Comparison of the model (top row) and residual (bottom row) images from ProFound (panel 2), PyBDSF (panel 4 and panel 5 with the atrous setting) and AEGEAN (panel 6) of five 3C sources (whose images are shown in panel 1). These are shown for the sources: 3C16 (a), 3C19 (b), 3C28 (c), 3C42 (d) and 3C47 (e). The segmentation images from ProFound are shown in panel 3. Here PyBDSF and AEGEAN use the default parameters as described in Section 4.1, and used in Hopkins et al. (2015). Shown are the entirety of the images that are fit.

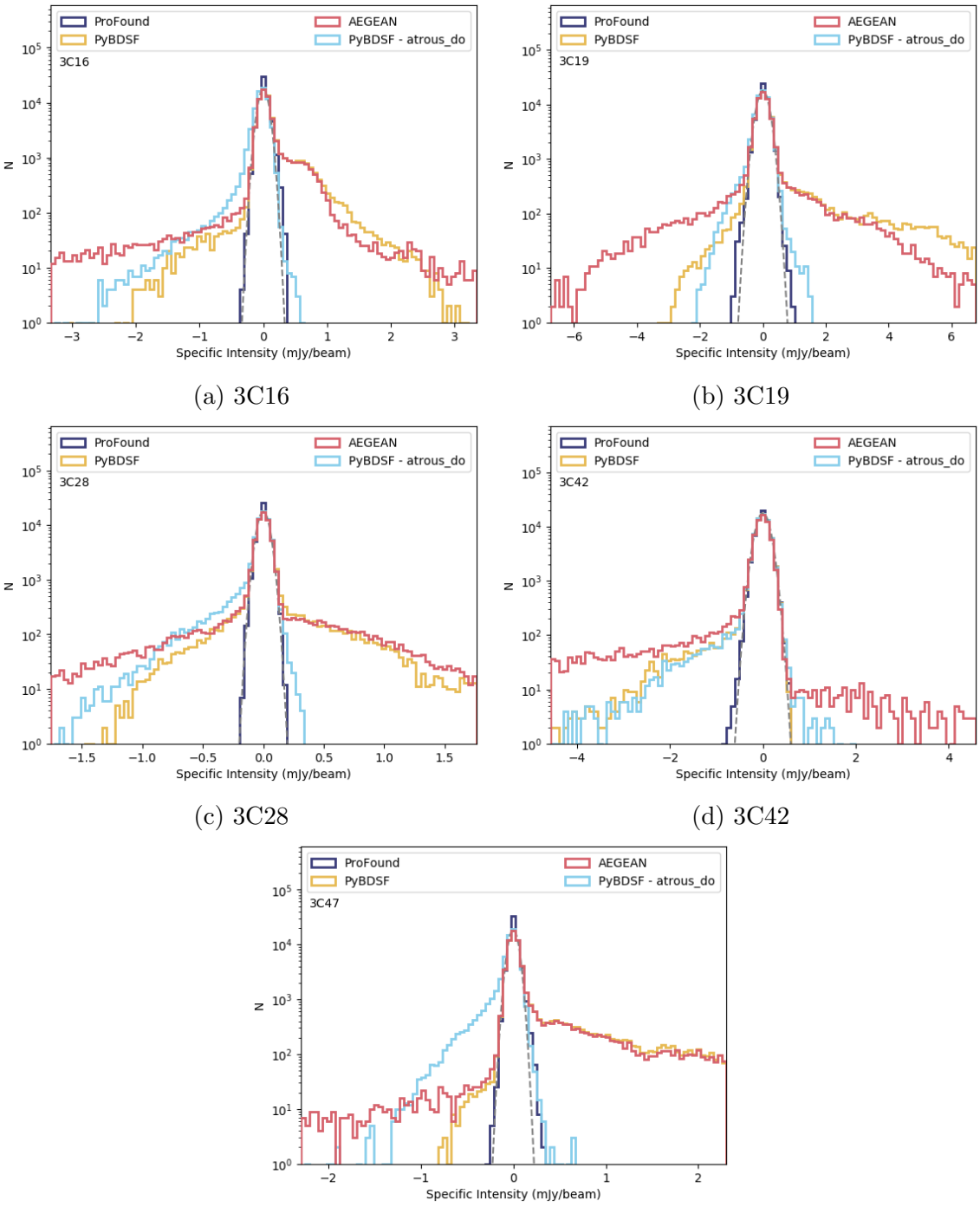
the beam (c) extended sources with complex morphologies and (d) extended sources with smooth morphologies.

The input catalogues (Cat_{inp}) were matched to the measured catalogue for each simulated image (Cat_{im}) in a number of steps. Firstly any sources that were already in the PyBDSF residual image (before any simulated sources were injected) needed to be removed. Therefore each of the software were run over the residual image and the output catalogues from this (Cat_{resid}) cross matched to the output catalogue of the simulation (Cat_{im}) within a 1" positional radius. Any sources that were matched in this step were removed from the output catalogue of the simulation. This produced a final catalogue (Cat_{sim}) that only had those sources that were recovered from the simulation. Cat_{inp} was then matched to Cat_{sim} within a 3" positional radius. This matched catalogue allowed the input and recovered flux densities to be compared. Especially for extended sources, it is possible that AEGEAN and PyBDSF may decompose these sources into multiple components. Therefore all sources that were not matched to the input catalogue (within the 3" radius) were matched to the closest input source provided they were within 20" of an input source.

It is important to note that for Gaussian simulations, there existed a difference between the flux density from S^3 compared to the sum of the pixels that were injected into the image. This was small ($\lesssim$1%), but arose depending on whether a source was injected in the centre of the pixel or not. Despite the differences being small, in order to accurately compare fluxes, the flux densities from PyBDSF and AEGEAN were compared to the flux densities input from S^3, whilst for ProFound, the sum of flux densities in the pixels that were injected into the image were compared to.

To compare the success of the different source extractors in being able to accurately measure these sources, several quantities are compared:

1. *Flux densities:* By comparing the injected flux density to the flux density recovered by each software, this allowed me to quantify the accuracy of source



(a) 3C16

(b) 3C19

(c) 3C28

(d) 3C42

(e) 3C47

Figure 4.14: Histogram of the residuals leftover in the images of 3C sources, described in Section 4.3.2, when the source models from ProFound (dark blue), PyBDSF (yellow, or light blue when the atrous setting is used) and AEGEAN are used. The grey dashed line indicates a Gaussian model used to fit the residuals. This is shown for (a) 3C16, (b) 3C19, (c) 3C28, (d) 3C42 and (e) 3C47.

recovery. A histogram of the ratio of the recovered to the input flux densities are given as well as this ratio as a function of input flux density.

2. *Completeness:* The completeness (see Hopkins et al., 2015) quantifies what fraction of the input sources are recovered as a function of flux density. This is calculated here through matching the input sources to recovered sources (within the centre of the image) using a radius of 3" for single component sources (Gaussian and elliptical simulations) and 15" for the multi component sources. The completeness at a given flux density (S_i) is given by Equation 4.5. This therefore quantifies how successful the software is at finding injected sources.

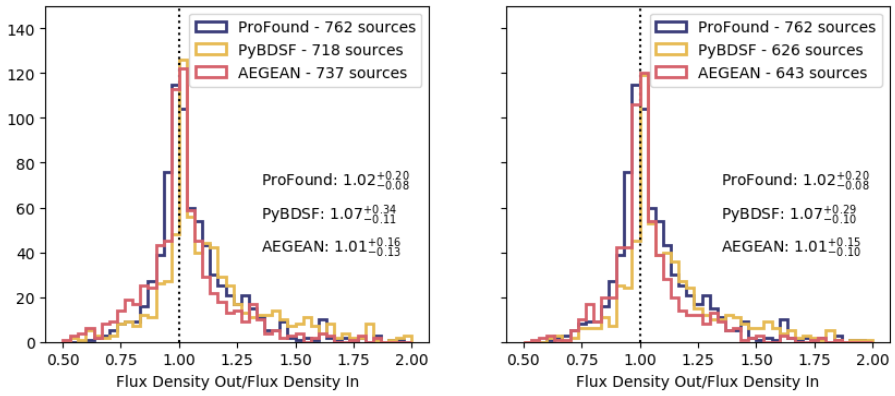
3. *Reliability:* The reliability (also see Hopkins et al., 2015) instead quantifies what fraction of the recovered sources are real (as a function of flux density). This is calculated here by matching the output sources (within the centre of the image) to input sources, using the same matching radii as above. It is defined by Equation 4.6.

4. *Completeness* $\times$ *Reliability:* The product of completeness and reliability is used to describe a compromise between recovering as many sources as possible whilst minimising false detections.

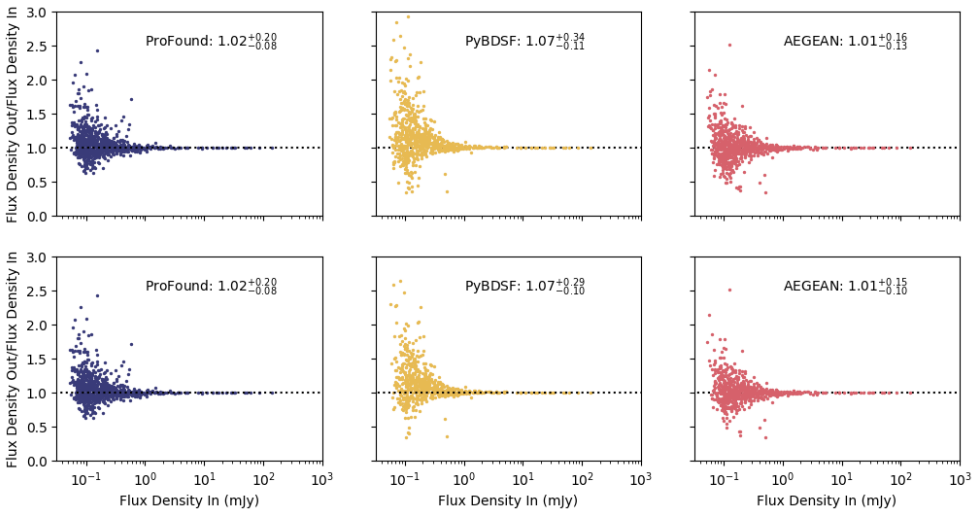
$$Completeness(S_i) = \frac{N_{Matched,\ input}(S_i, S_i + \delta S)}{N_{injected}(S_i, S_i + \delta S)} \tag{4.5}$$

$$Reliability(S_i) = \frac{N_{Matched,\ recovered}(S_i, S_i + \delta S)}{N_{injected}(S_i, S_i + \delta S)} \tag{4.6}$$

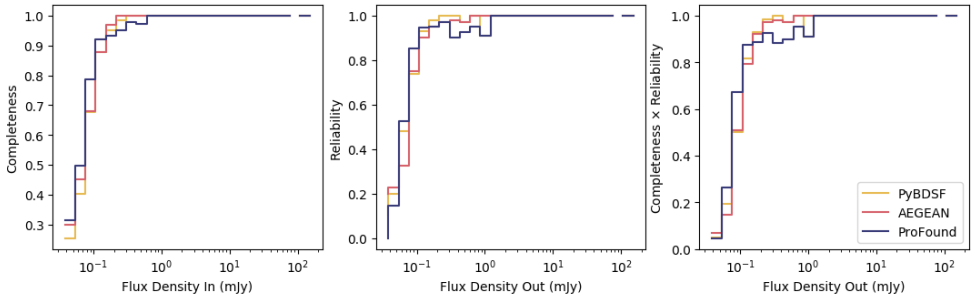
As the thresholds used by the different software will affect measurements of completeness and reliability, these are shown in Figures 4.15 - 4.18 using the detection parameters found in Section 4.4. The values of completeness and reliability will be affected by the matching radii between sources as well as whether sources have been merged together into a single object.



(a) Histogram comparison of flux densities recovered to those injected

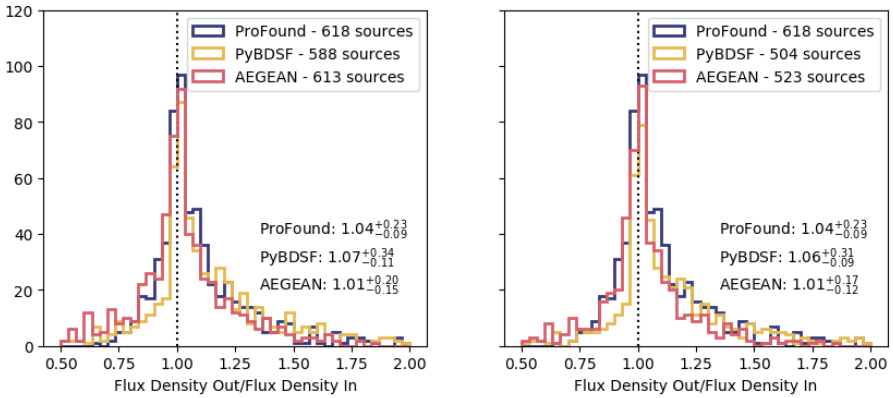


(b) The ratio of recovered to injected flux densities as a function of input flux density

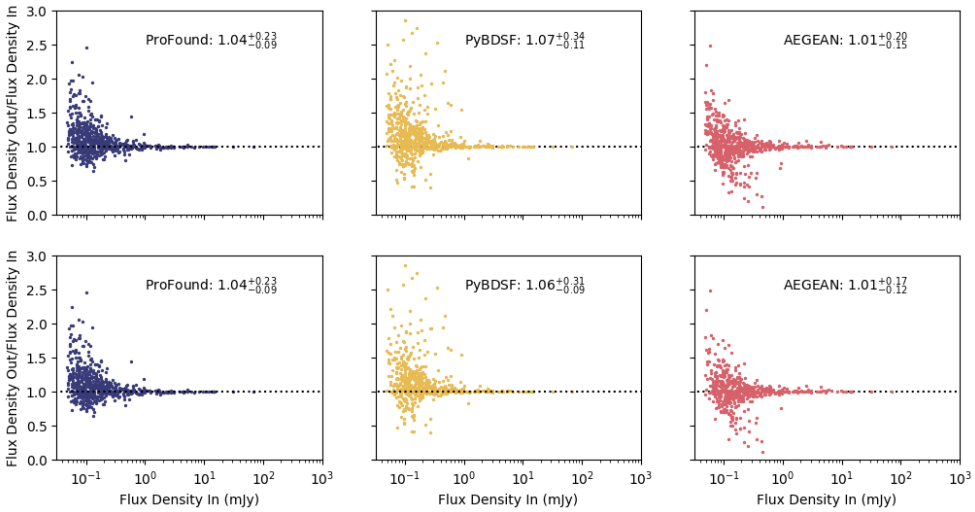


(c) The completeness (left), reliability (centre) and completeness × reliability (right) as a function of flux density.

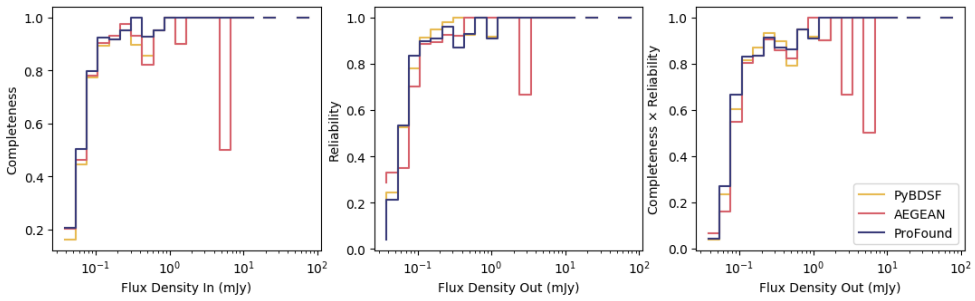
Figure 4.15: The results from running ProFound (dark blue), PyBDSF (yellow) and AEGEAN (red) over the simulated image in which Gaussian sources of different sizes and fluxes are injected into the image. The histogram of recovered to injected flux densities are shown in (a) for the parameters from Section 4.4 (left) and default parameters (right). In (b) are the ratios of recovered to injected flux densities as a function of input flux density for the parameters from Section 4.4 (upper) and default parameters (lower) and finally in (c) is the completeness, reliability and completeness × reliability for the parameters determined in Section 4.4.



(a) Histogram comparison of flux densities recovered to those injected

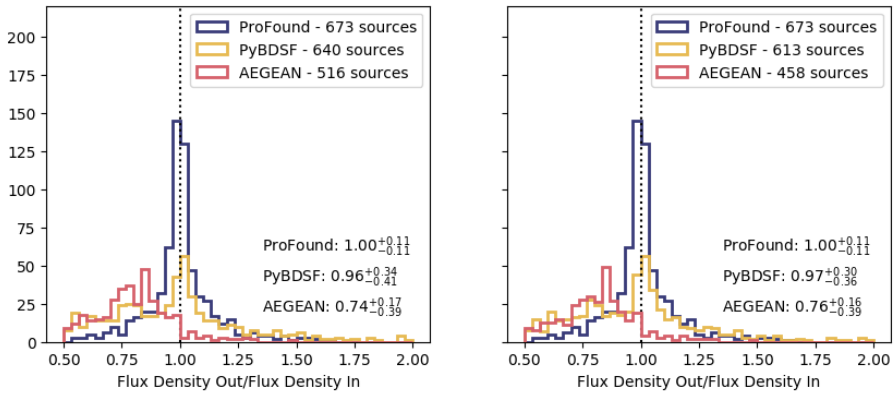


(b) The ratio of recovered to injected flux densities as a function of input flux density

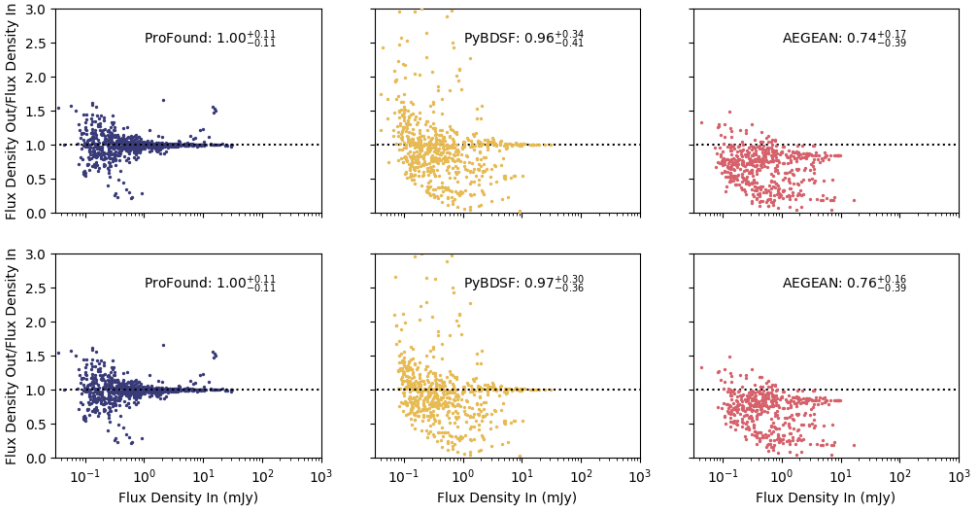


(c) The completeness (left), reliability (centre) and completeness × reliability (right) as a function of flux.

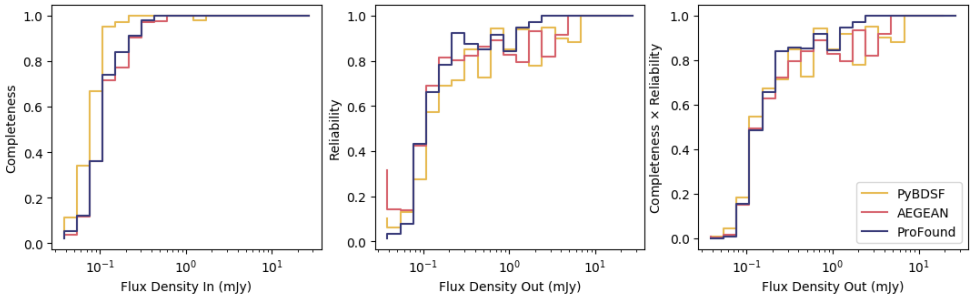
Figure 4.16: The results from running ProFound (dark blue), PyBDSF (yellow) and AEGEAN (red) over the simulated image in which elliptical sources of different sizes and fluxes are injected into the image. The histogram of recovered to injected flux densities are shown in (a) for the parameters from Section 4.4 (left) and default parameters (right). In (b) are the ratios of recovered to injected flux densities as a function of input flux density for the parameters from Section 4.4 (upper) and default parameters (lower) and finally in (c) is the completeness, reliability and completeness × reliability for the parameters determined in Section 4.4.



(a) Histogram comparison of flux densities recovered to those injected

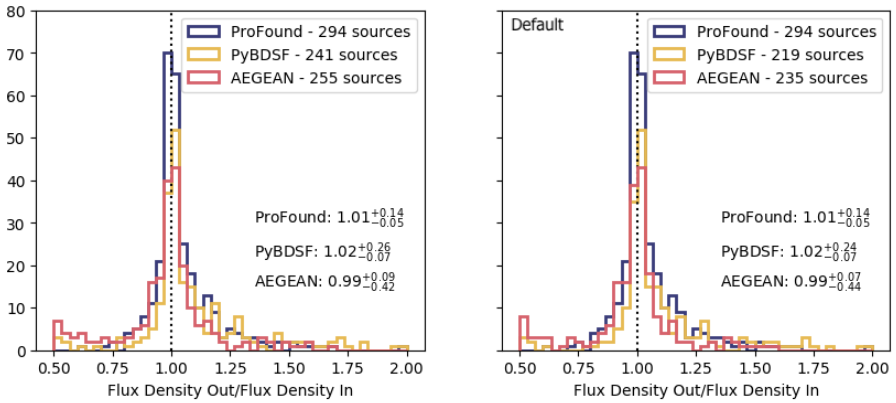


(b) The ratio of recovered to injected flux densities as a function of input flux density

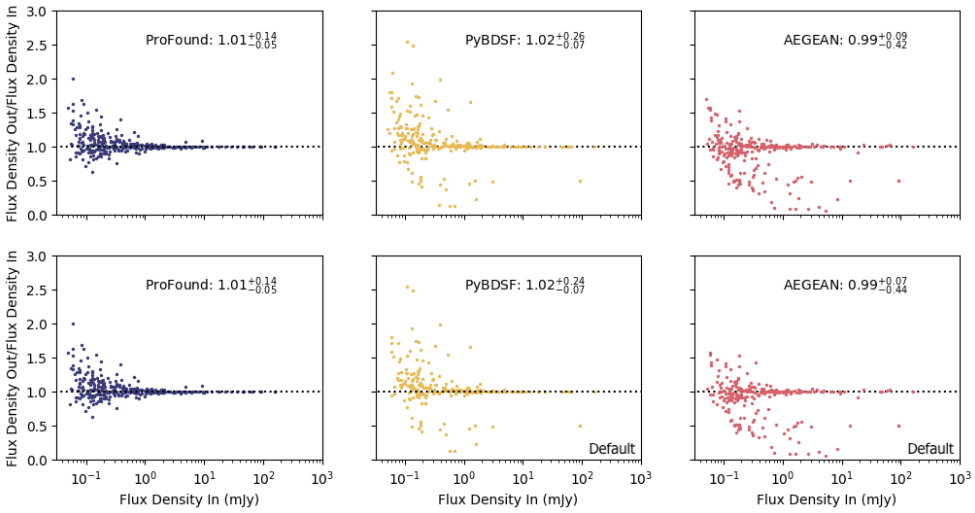


(c) The completeness (left), reliability (centre) and completeness × reliability (right) as a function of flux.

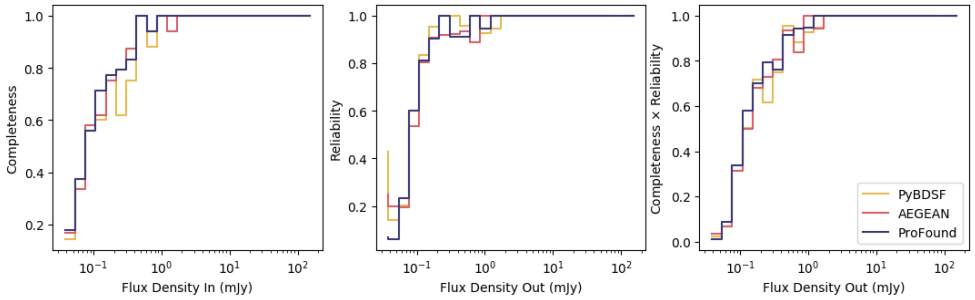
Figure 4.17: The results from running ProFound (dark blue), PyBDSF (yellow) and AEGEAN (red) over the simulated image in which sources with complex extended morphologies are injected into the image. The histogram of recovered to injected flux densities are shown in (a) for the parameters from Section 4.4 (left) and default parameters (right). In (b) are the ratios of recovered to injected flux densities as a function of input flux density for the parameters from Section 4.4 (upper) and default parameters (lower) and finally in (c) is the completeness, reliability and completeness × reliability for the parameters determined in Section 4.4.



(a) Histogram comparison of flux densities recovered to those injected



(b) The ratio of recovered to injected flux densities as a function of input flux density



(c) The completeness (left), reliability (centre) and completeness × reliability (right) as a function of flux.

Figure 4.18: The results from running ProFound (dark blue), PyBDSF (yellow) and AEGEAN (red) over the simulated image in which sources with smooth extended morphologies are injected into the image. The histogram of recovered to injected flux densities are shown in (a) for the parameters from Section 4.4 (left) and default parameters (right). In (b) are the ratios of recovered to injected flux densities as a function of input flux density for the parameters from Section 4.4 (upper) and default parameters (lower) and finally in (c) is the completeness, reliability and completeness × reliability for the parameters determined in Section 4.4.

The results for the different simulations can be seen in Figures 4.15 - 4.18 where the median flux ratios as well as uncertainties in these values by using the 16^{th} and 84^{th} percentiles are quoted. This can be seen for ProFound, PyBDSF and AEGEAN. For PyBDSF, the emission is smooth in the Gaussian, elliptical and smooth extended source simulations. Therefore, PyBDSF without the atrous setting is used. For the extended sources with complex morphologies, on the other hand, PyBDSF with the atrous setting is used. In each of Figures 4.15 - 4.18, panel (a) shows the histogram of flux ratios, (b) shows the flux ratio as a function of input flux density and finally panel (c) shows the results for completeness and reliability.

For the simulations in which Gaussian sources are injected, the results can be seen in Figure 4.15. For these PyBDSF, AEGEAN and ProFound all peak around a recovered to input flux ratio of 1. The flux ratio for ProFound is $1.02^{+0.20}_{-0.08}$, suggesting that ProFound is, on average, able to accurately determine the flux densities for Gaussian sources. However, as the errors are asymmetric, ProFound is more likely to over estimate the flux densities of sources instead of underestimate them. For PyBDSF on the other hand, the recovered flux ratio when the parameters from Section 4.4 are used is $1.07^{+0.34}_{-0.11}$ ($1.07^{+0.29}_{-0.10}$ for the default values) whilst with AEGEAN I find a ratio of $1.01^{+0.16}_{-0.12}$ ($1.01^{+0.15}_{-0.10}$ for the default values). This suggests that ProFound is able to accurately measure the fluxes of Gaussian sources and produce comparable results to PyBDSF and AEGEAN, which were designed to accurately model Gaussian sources. The accurate detection of Gaussian sources is crucial to any source detection software in radio continuum images as all unresolved sources will be Gaussian in morphology. The completeness and reliability of the different software can be seen in Figure 4.15(c), which shows that ProFound has comparable reliability and completeness to both AEGEAN and PyBDSF for sources of a Gaussian morphology. There is a small decline in reliability at some high flux densities, however this appears to originate from closely located, large Gaussians

being merged together. These are therefore considered as one source in ProFound but may be considered as separate sources by PyBDSF and AEGEAN.

For elliptical sources, the results of the simulation are shown in Figure 4.16. From these, I measure a flux ratio for ProFound of $1.04^{+0.23}_{-0.09}$ whilst I find a flux ratio of $1.07^{+0.34}_{-0.11}$ ($1.06^{+0.31}_{-0.09}$ using the default parameters) using PyBDSF and $1.01^{+0.20}_{-0.15}$ ($1.01^{+0.17}_{-0.12}$ using the default parameters) for AEGEAN. As can be seen in Figure 4.16(a), these values all peak around a flux ratio of 1. However the flux ratios for PyBDSF and ProFound are on average over-estimating the flux density of simulated sources compared to the input value. In general, though, by just considering the flux ratio for this simulation, I again find that ProFound is also able to extract accurate flux densities for these morphologies with comparable results to PyBDSF and AEGEAN. The Completeness×Reliability (see Figure 4.16c, right) also suggest that ProFound is able to also extract elliptical sources accurately compared to both AEGEAN and PyBDSF. Therefore ProFound also is able to successfully model sources with a more disk like morphology.

For the complex extended morphologies, I measure a flux ratio of recovered to input fluxes with ProFound of 1.00 ± 0.11. This is shown in Figure 4.17. For PyBDSF (with the atrous setting), the flux ratio also peaks around a value of ~ 1 for which I find a ratio of $0.96^{+0.34}_{-0.41}$ ($0.97^{+0.30}_{-0.36}$ using the default parameters). This suggests that with this atrous setting utilised, the flux densities of these complex morphology sources are well modelled. However, the scatter in the flux ratio found with PyBDSF is $\sim$3 times larger than with ProFound. AEGEAN on the other hand, has a flux ratio that peaks around $\sim$0.75. This is likely to be a combination of not being able to model extended emission and to not having matched all large components of a source together. Whilst the completeness of PyBDSF is typically larger than found by ProFound and AEGEAN (see Figure 4.17c) its reliability is lower which suggests that PyBDSF is detecting more spurious sources. The combi-

nation of Completeness×Reliability, which determines a compromise in accuracy of these software, are comparable between all three at small flux densities, however at the higher flux densities ProFound appears more accurate.

Finally, all three software packages have similar accuracy in detecting smooth extended sources, this can be seen in Figure 4.18. I measured a flux ratio of $1.01^{+0.14}_{-0.05}$ for ProFound, $1.02^{+0.26}_{-0.07}$ ($1.02^{+0.24}_{-0.07}$ using the default values) for PyBDSF and $0.99^{+0.09}_{-0.42}$ ($0.99^{+0.07}_{-0.44}$ using the default values) for AEGEAN. All three of these software packages that are compared show comparable completeness and reliability, suggesting ProFound is able to measure the flux density of these sources with similar accuracy to that of PyBDSF and AEGAN.

Combined, these simulations all indicate that ProFound is able to accurately act as a source detection software for radio continuum images, for a variety of morphologies, with similar success to both PyBDSF and AEGEAN for smooth sources. My simulations suggest that ProFound is, however, more accurate in the recovery of complex extended morphologies. All these suggest ProFound can be successfully used for determining source catalogues with radio continuum observations.

4.6 Conclusions

In this Chapter, I have investigated whether the near-IR and optically motivated source extraction software ProFound (Robotham et al., 2018) can be used in detecting sources from radio continuum observations. To assess the accuracy of ProFound, I investigated the detection of sources over already observed radio continuum observations as well as for simulated radio images. These simulated images took into account the variety of source morphologies that are typically observed in radio continuum imaging. These investigations have all shown that ProFound is able to perform just as successfully as the commonly used packages of AEGEAN and PyBDSF, but is

more accurate at tracing and extracting emission from sources that have complicated morphologies, such as AGN with large jets. This work suggests that ProFound can be used successfully to extract radio catalogues in future continuum radio surveys.

For different radio continuum imaging, different software will be preferred in different situations. For surveys in which the observations are dominated by unresolved sources, those detection software which fit Gaussian components are likely to be useful because these unresolved sources will be Gaussian in nature. This is especially important for telescopes such as the MWA (Tingay et al., 2013). However, for telescopes that have combinations of long and short baselines and have the resolution to detect extended emission (e.g. for LOFAR (van Haarlem et al., 2013) or MeerKAT (Jonas, 2009; Jonas & MeerKAT Team, 2016)), ProFound may be more suitable.

Whilst some of the advantages and limitations of using ProFound in a radio continuum context have been discussed through this work, there are also many other advantages. For example, as discussed in Robotham et al. (2018); Davies et al. (2017), segments that are generated using one image (typically with high resolution) are also able to be used to obtain fluxes at other wavelengths. This is through converting these segments to the same pixel resolution and then either forcing a measurement, or dilating these segments outwards to enclose the entire flux of a source. This can be used to extract fluxes for galaxies over a wealth of multi frequency imaging and may be useful in constructing spectral energy distributions (SEDs) of galaxies, using a consistent method. This is useful for detected sources, but by forcing flux measurements at the positions of known galaxies, has the potential to be used to obtain source counts below 5σ. Both of these uses will be discussed in Chapter 6.

For radio continuum observations, this work has shown that ProFound is competitive with currently available source extraction software, but it also has advantages, especially when sources are not restricted to unresolved morphologies. ProFound therefore has significant potential for future sensitive radio observations.

Chapter 5

LOFAR Observations of the XMM-LSS Field

In order to gain as much information as possible about radio galaxies, it is important to observe sources at many frequencies. The VLA COSMOS observations (Smolčić et al., 2017a) that were used in Chapters 2 and 3 are, in the context of this Thesis, high frequency radio observations. At low frequencies (<1 GHz), older electron populations are observed (see e.g. Brienza et al., 2017), AGN emission is dominated by the lobes (see Peterson, 1997) and steep spectrum sources are preferentially observed (due to Equation 1.19). The LOw Frequency ARray (LOFAR; van Haarlem et al., 2013) is one pathfinder for SKA-LOW that is observing the low-frequency radio sky (see Section 1.6). This includes recent observations over $\sim$420 deg^2 of the northern sky (Shimwell et al., 2019).

Multi-frequency radio continuum observations are essential for many reasons. Not only can they help in determining the source type (see e.g. Callingham et al., 2017) but they can also be important in disentangling the synchrotron components of radio emission from free-free emission (see Section 1.4 and e.g. Galvin et al., 2018). As such, it is useful to observe radio galaxies where there is wealth of ancillary radio and

multi-wavelength data. Using LOFAR to observe sources over such fields of the sky will be important in constraining the low-frequency emission of these extragalactic sources. Whilst LOFAR will observe many of these fields, including Lockman Hole and Elais-N1 to great depths (Röttgering, 2010; van Haarlem et al., 2013), these are northern fields. These are optimal for observations with LOFAR, as the antennas are located in the Netherlands and northern Europe and so they are in regions in which the telescope is highly sensitive. The XMM-LSS field (described in Chapter 4) however is another field that has a wealth of information across large wavelength ranges (Tasse et al., 2007; Heywood et al., in prep; Chen et al., 2018; Jarvis et al., 2013; Oliver et al., 2012; Aihara et al., 2018). Unfortunately though, its location, at a declination of $\sim$-4.5°, is not ideal for LOFAR, whose optimal declination is $\sim +52°$.

In this Chapter, I present observations from Hale et al. (2019b) of the XMM-LSS field at 144 MHz, observed with LOFAR for a total of 12 hours. This was reduced using recently developed pipelines (see Tasse et al., 2018; Shimwell et al., 2019) which have been essential in producing the recent LOFAR Two-metre Sky Survey (LoTSS Shimwell et al., 2019). Using the catalogues produced by these observations allowed me to determine the success of LOFAR observing at low elevations, which is important for the continuation of LoTSS. This catalogue also allows comparisons with past radio observations over the region of sky in order to discuss the properties of these sources.

5.1 Challenges of Observing at Low Frequencies

As touched upon in Section 1.5, there are challenges that exist with low frequency radio observations which are less prevalent at higher frequencies. At low frequencies changes in the electron content in the ionosphere (e.g. Wilson et al., 2009; Williams et al., 2016; Heald et al., 2018) cause phase changes in the radio waves. This leads to challenges in accurate astrometry for these sources, with the apparent positions able

to vary by large amounts (O(∼10"); Heald et al., 2018). This is more problematic for observations with a large field of view because variations in the ionosphere can differ significantly across the image. With LOFAR, the field of view for single pointings is ∼20 deg^2 (Williams et al., 2016; Hardcastle et al., 2016; Shimwell et al., 2019). As such, a single correction to account for the ionosphere cannot be applied across the image as this will leave behind residual ionospheric errors across the field of view.

In order to help deal with the varying ionospheric effects across the field of view, a technique of facet calibration has been developed and implemented in pipelines such as FACTOR (based on van Weeren et al., 2016, and see `http://www.astron.nl/citt/facet-doc/`) and in the DDF-pipeline (see e.g. `https://github.com/saopicc/DDFacet` and Tasse et al., 2018; Shimwell et al., 2019). The basic premise for both of these are to:

1. Find bright sources within the field of view that have sufficient flux density so that self-calibration can be performed.

2. Tile up the sky around these bright sources into "facets" so that each facet only contains the sky that is closest to the central bright source (or sources).

3. Self calibrate for phase and amplitude errors on these bright sources in order to account for the ionosphere.

4. The assumption is then made that the calibration solutions that need to be applied to the bright central source can also be applied to the rest of the facet.

5. By solving for the solutions within each facet, it is possible to improve the fidelity of the radio images by reducing ionospheric effects across the field of view.

This method is known as direction dependent calibration, where calibration is performed for different regions within the field of view. Whilst direction dependent

calibration is used here for low frequency observations due to the ionospheric effects, with more sensitive observations at higher frequencies (e.g. with MIGHTEE; Jarvis et al., 2017) direction dependent calibration will also be important due to effects such as the changing beam pattern (see Tasse et al., 2018).

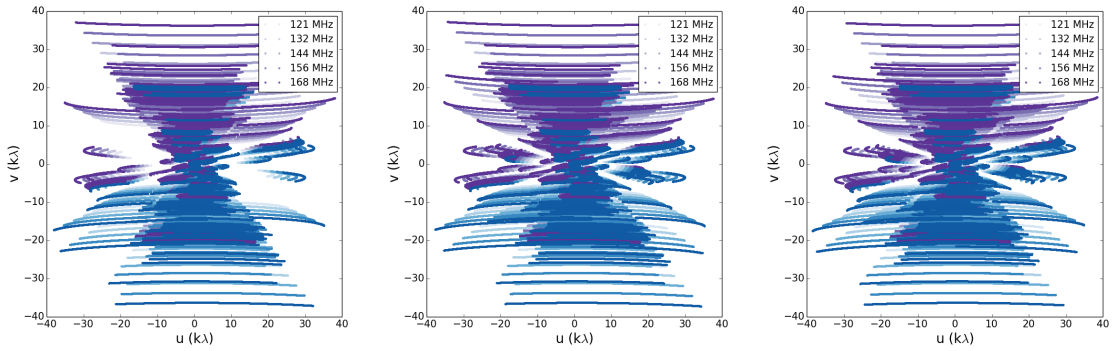
5.2 Observations

In this Chapter I present observations of the XMM-LSS field observed with LOFAR. This made use of three observations of the XMM-LSS field, each of four hours in duration using the High Band Antenna (HBA) which operates in the region of ∼110-190 MHz. As mentioned in Section 1.6, LOFAR has both a combination of a dense core and stations spread over the Netherlands combined with stations across other countries within Europe. Whilst the European baselines will dramatically improve the resolution of observations, obtaining sub-arcsecond resolution, they are challenging to calibrate. Whilst there are observations of this field using many of the European stations, the long baseline pipeline is currently in development and so only the visibilities using the Netherlands stations are combined for these observations. For the three observations used in this Chapter, the details of these and the primary calibrators used (see Section 1.5) are given in Table 5.1.

As described in Section 1.5, not all spatial scales are observed using interferometers. The scales that these observations are sensitive to are determined by the distribution of baselines between antennas used in the array. The corresponding *uv*-coverage for these observations are shown in Figure 5.1 where the *uv*-coverage for five frequency bands is shown. This is for bands between ∼121 MHz (light purple) to ∼168 MHz (dark purple) with the conjugate points also shown in blue. These conjugate points represent the fact that a given visibility at (u,v) will also correspond to a visibility at (-u,-v) depending on the convention of the direction taken. Figure

ID	Target	Date	Duration	Pointing Position
Obs 1	3C48	22/12/2015	10 mins	01:37:41.30 +33:09:35.1
Obs 1	XMM-LSS	22/12/2015	4 hours	02:20:00.00 -04:30:00.0
Obs 2	3C147	27/02/2017	10 mins	05:42:36.155 +49 51 07.28
Obs 2	XMM-LSS	27/02/2017	4 hours	02:20:00.00 -04:30:00.0
Obs 3	3C147	01/03/2017	10 mins	05:42:36.155 +49 51 07.28
Obs 3	XMM-LSS	01/03/2017	4 hours	02:20:00.00 -04:30:00.0

Table 5.1: Observation details of the three four hour observations of the XMM-LSS field used to generate the final images and catalogues discussed in this Chapter.



(a) Observation 1 (b) Observation 2 (c) Observation 3

Figure 5.1: *uv*-coverage of the three four hour observations of the XMM-LSS field used for this work. Five different frequency bands are shown in purple ranging from 121 MHz (lightest) to 168 MHz (darkest). Their conjuguate points are also shown in blue.

5.1 shows that there is incomplete sampling within the *uv*-plane. This will result in artefacts within the dirty beam that will likely affect the quality of the image that is possible with these observations.

5.3 Data reduction

5.3.1 Calibration and Direction Independent Cleaning

As described in Section 1.5, the process of going from the observed visibilities to the final, cleaned image takes many steps. This includes firstly calibrating the data using

observations of well-known radio sources, to determine the response of the different antennas as a function of frequency, as well as calibrating the observed measured values by the detector to the physical flux densities that these correspond to. For each of these observations this initial processing was done using the PreFactor pipeline[1] (see de Gasperin et al., 2019; Shimwell et al., 2019), which is the automated calibration pipeline used by the LOFAR surveys team. For these observations, the PreFactor pipeline was run by Tim Shimwell (University of Leiden) using SURFsara[2], which is a high performance computing facility based in the Netherlands. Through this calibration process, averaging of the data into 8s intervals with two channels per sub-band (see Mechev et al., 2017) was used to reduce the size of the observations.

Solely performing direction independent cleaning limits observations to this resolution (see e.g. Mahony et al., 2016; Shimwell et al., 2017). Higher resolution imaging is possible though by accounting for the direction dependent effects introduced by the ionosphere and the beam. Two pipelines have been developed to perform direction dependent cleaning on LOFAR observations. Firstly, Factor (van Weeren et al., 2016, , `http://www.astron.nl/citt/facet-doc/`) was developed and used for the first LOFAR observations (Williams et al., 2016; Hardcastle et al., 2016). More recently, is the DDF-pipeline (DDF; see Shimwell et al., 2019)[3]. This has been used for the most recent, large sky observations of the LoTSS Survey Shimwell et al. (2019). These are described below.

5.3.2 Direction Dependent Cleaning

Factor

When first reducing LOFAR observations of the XMM-LSS field, I used Factor (`http://www.astron.nl/citt/facet-doc/`; van Weeren et al., 2016) for direction

[1] `https://github.com/lofar-astron/prefactor`
[2] `https://www.surf.nl/en/about-surf/subsidiaries/surfsara/`
[3] `https://github.com/mhardcastle/ddf-pipeline`

dependent cleaning. FACTOR tiles the sky up into facets around the bright sources, which can be seen in Figure 5.2. This also included regions around bright sources outside the field of view, that may affect observations within the field of view. Each facet was then calibrated individually with two rounds of phase only calibration to correct for the ionosphere and then two rounds of amplitude and phase calibration to reduce artefacts relating to the beam. These stages of cleaning can be seen for four of the facets in Figure 5.3. This calibration can then be applied to the facet as a whole before the individual facets are tiled together.

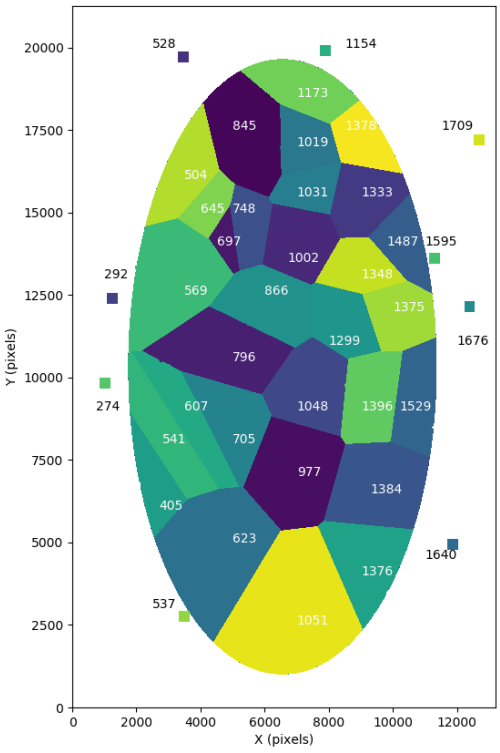


Figure 5.2: The facets across the field of view determined by FACTOR to perform direction dependent cleaning.

Whilst the improvement to the fidelity of these images using FACTOR is obvious, FACTOR does have some limitations. Firstly, each facet is calibrated independently and performed sequentially until the whole field of view is complete. This means that there can be discontinuities in the images across facet boundaries. This can be seen,

for example in Figure 9 of Williams et al. (2016). Most challenging though, is that the process is slow. For an initial reduction where only 40 sub-bands were used (there are ~ 230 sub-bands across the full bandwidth), this took approximately a month to reduce a large portion of the field. This is therefore not practical to reduce the data for the large sky survey, LoTSS (Shimwell et al., 2019).

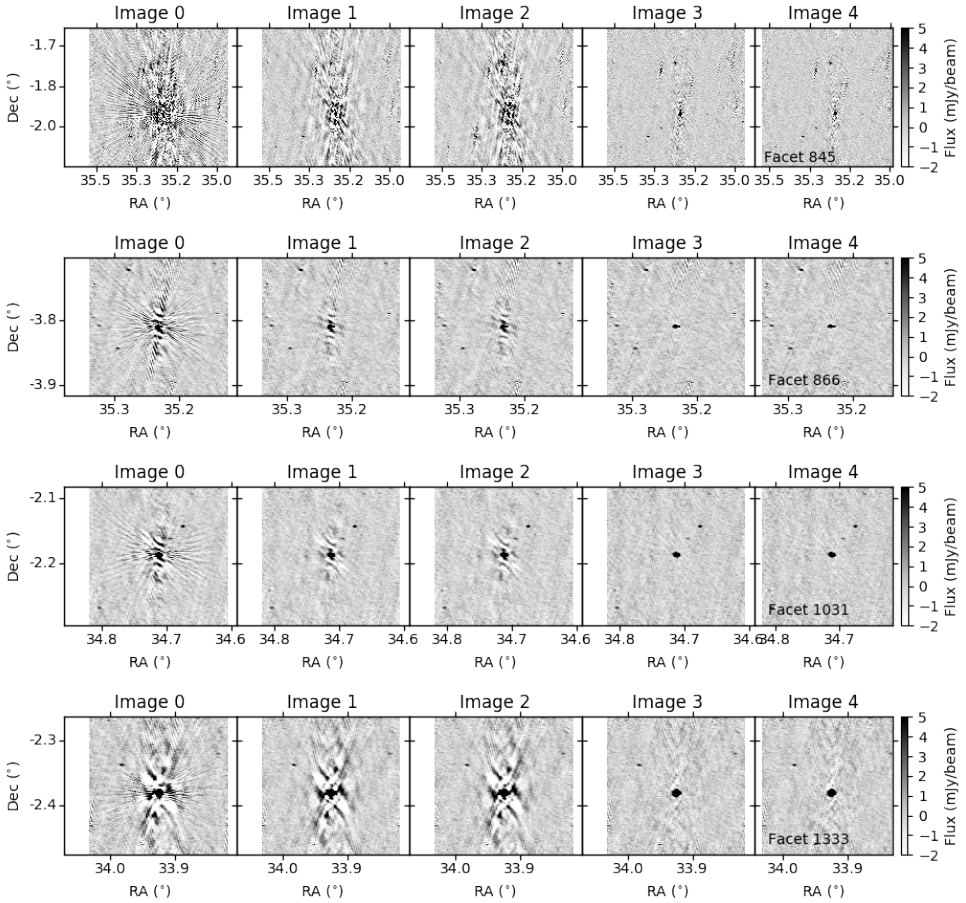


Figure 5.3: Direction dependent facet calibration of the bright sources from four facets using FACTOR (`http://www.astron.nl/citt/facet-doc/`; van Weeren et al., 2016) using two rounds of phase only calibration and two rounds of phase and amplitude calibration.

DDF

Instead, the new LOFAR pipeline that was developed for LoTSS and is used here combined two packages. The first is KILLMS (hereafter KMS; Tasse, 2014b,a; Smirnov &

Tasse, 2015) which solves for the gains to satisfy the RIME (Section 1.5.4). kMS uses Jones matrices to solve these equations for the different directions. This is combined with an imager, DDFacet[4] (Tasse et al., 2018) which applies the gains solutions found in kMS and cleans the image. For use with the LOFAR surveys, these packages were combined together to form the DDF pipeline[5] (Shimwell et al., 2019). Using this combined pipeline, the field of view is again split up into facets around the bright sources and the gains solved for. In contrast to Factor, however, the directions are solved for simultaneously. This means that the effects of bright sources on neighbouring facets can be accounted for and also, as facets taper towards the edge, there are no discontinuities in the image between neighbouring facets. DDF also uses rounds of phase only, as well as phase and amplitude calibrations and associated cleaning in order to correct for the direction dependent effects. DDF was used to reduce all three observations simultaneously with the help of Wendy Williams (University of Hertfordshire). DDF was much more user friendly compared to Factor as no human intervention was necessary and the pipeline was much quicker. To reduce all sub-bands of the three observations took one week.

A "parset" file was used to specify the parameters associated with the cleaning, such as the weighting scheme to use. This weighting scheme determines how the *uv*-plane is sampled and whether to give more weight to longer or shorter baselines or whether to weight these equally. For these observations, a Briggs weighting scheme was used (see `https://casa.nrao.edu/docs/casaref/imager.weight.html`, `http://www.aoc.nrao.edu/dissertations/dbriggs/`; Briggs et al., 1999; Sault & Oosterloo, 2007) with a final robustness of -0.5. In the Briggs weighting scheme, a compromise is made between equally weighting all *uv*-points (natural; which has a Briggs robustness of -2) and equally weighting all pixels within the *uv*-plane (uniform; which has a Briggs robustness of 2). This will provide a compromise between observing

[4] `https://github.com/saopicc/DDFacet`
[5] `https://github.com/mhardcastle/ddf-pipeline`

extended emission whilst also producing high resolution images. The parset file also contained the location of the data as well as locations of previous radio observations to be used for flux and astrometric checks within the cleaning process. Observations at 74 MHz (from Cohen et al., 2003) and at 325 MHz (from Tasse et al., 2006) were used to check the flux scale, whilst the TGSS-ADR (Intema et al., 2017) catalogue (Section 1.6) was used as an initial sky model.

The output from DDF produced both a beam corrected and uncorrected image. Because the sensitivity will decline towards the edge of the field of view (not including secondary sidelobes in the primary beam), sources observed within these outer regions will therefore appear dimmer. The beam corrected image corrects for this drop off in sensitivity. For the final image described in this Chapter, this included only regions where the primary beam response was $\geq 50\%$.

5.4 Image Quality and Data Catalogues

The final image of the XMM-LSS field that was produced using DDF is shown in Figure 5.4 for the full 27 deg^2. This is elliptical in shape, contrary to the circular pointings observed in Williams et al. (2016); Hardcastle et al. (2016); Shimwell et al. (2017, 2019). This relates to the low elevation of the observed field to the antennae in LOFAR. This shape will be important to consider for the tiling strategy of LoTSS (Shimwell et al., 2019) at lower elevations. The corresponding central 4 deg^2 is shown in Figure 5.5. These images have a central frequency of 144 MHz and a resolution of 7.5×8.5".

The associated source catalogue was generated using PyBDSF (Mohan & Rafferty, 2015). This is one of the packages described in Chapter 4. PyBDSF has been used for all previous LOFAR observations (Mahony et al., 2016; Williams et al., 2016; Hardcastle et al., 2016; Shimwell et al., 2019). The PyBDSF parameters used for

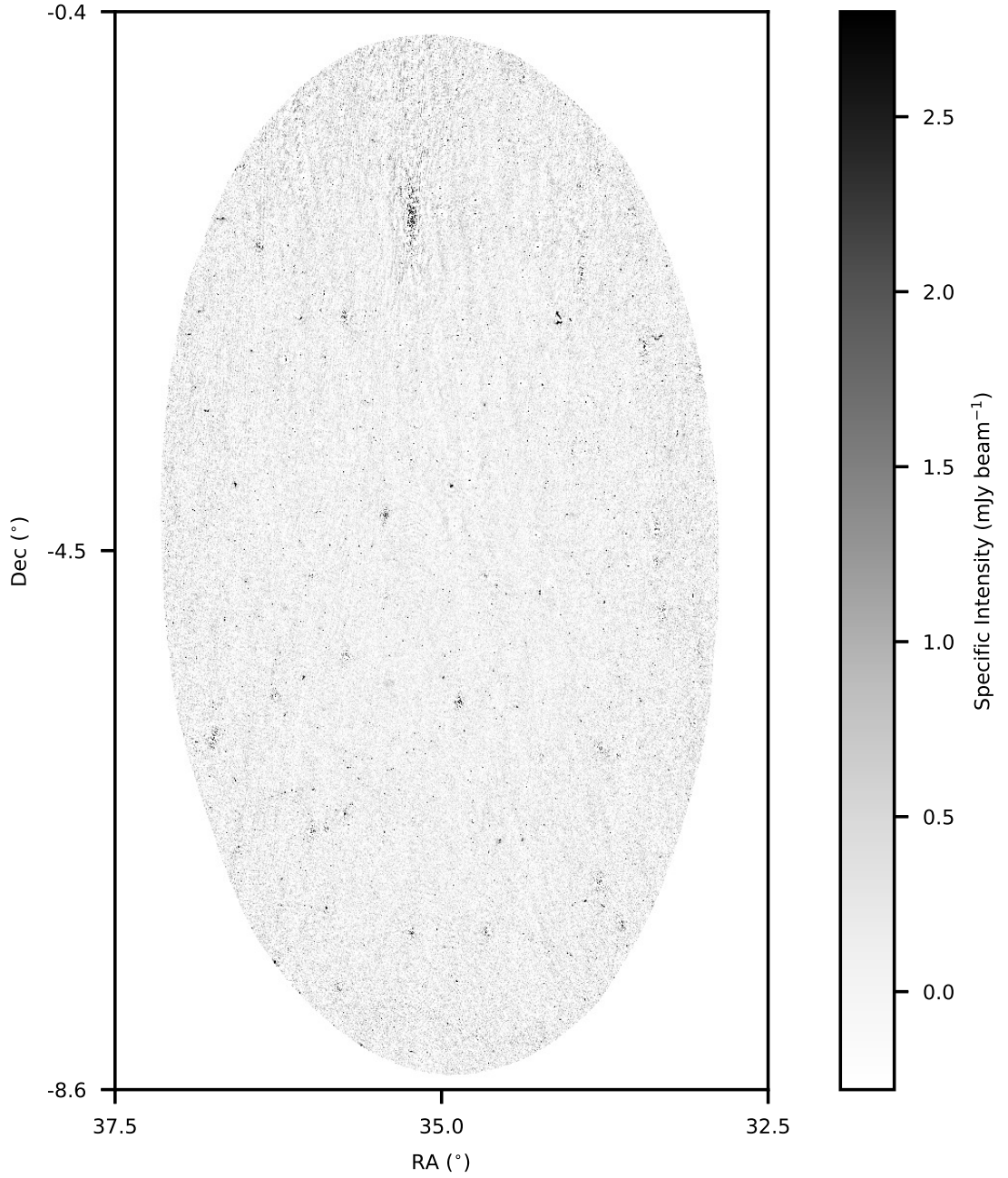


Figure 5.4: Image of the full ∼27 deg^2 image of the XMM-LSS field observations imaged with LOFAR at 144 MHz. This is shown on a flux scale between -$\sigma_{\rm cen}$ and 10$\sigma_{\rm cen}$, where $\sigma_{\rm cen}$ is the central rms of the image, 0.28 mJy beam^{-1}.

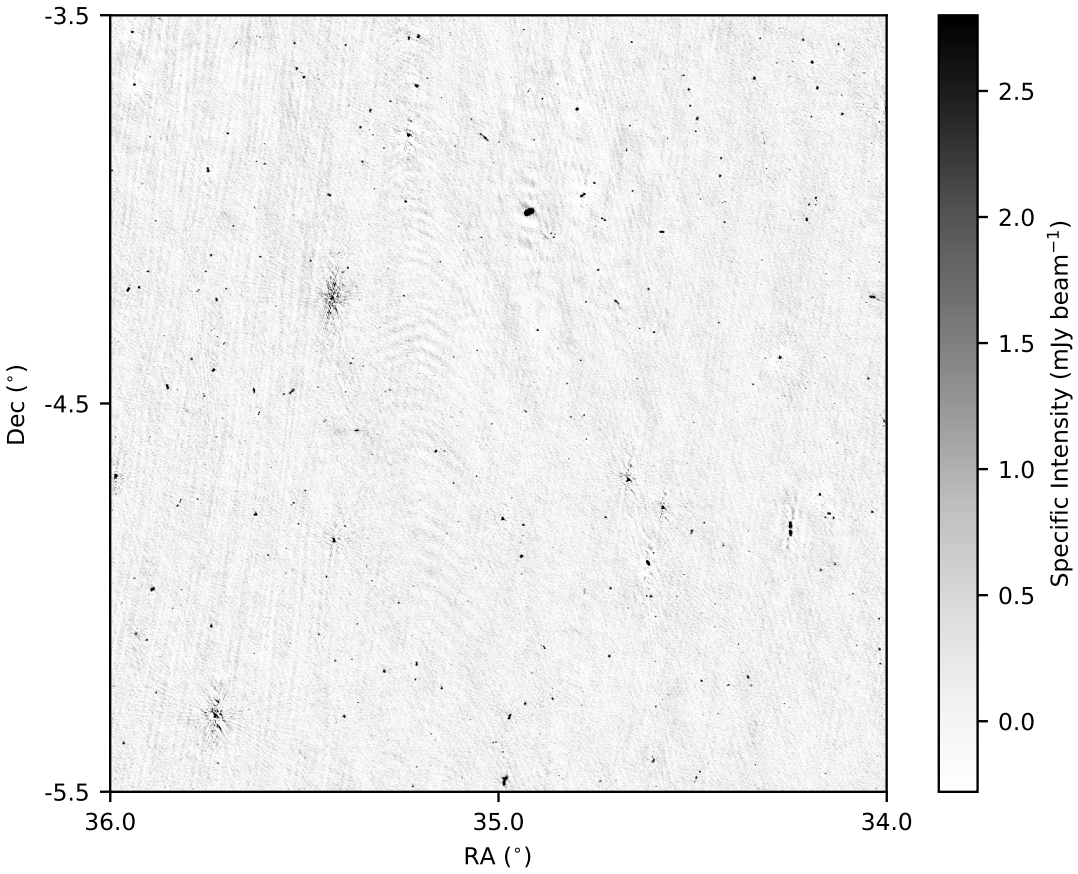


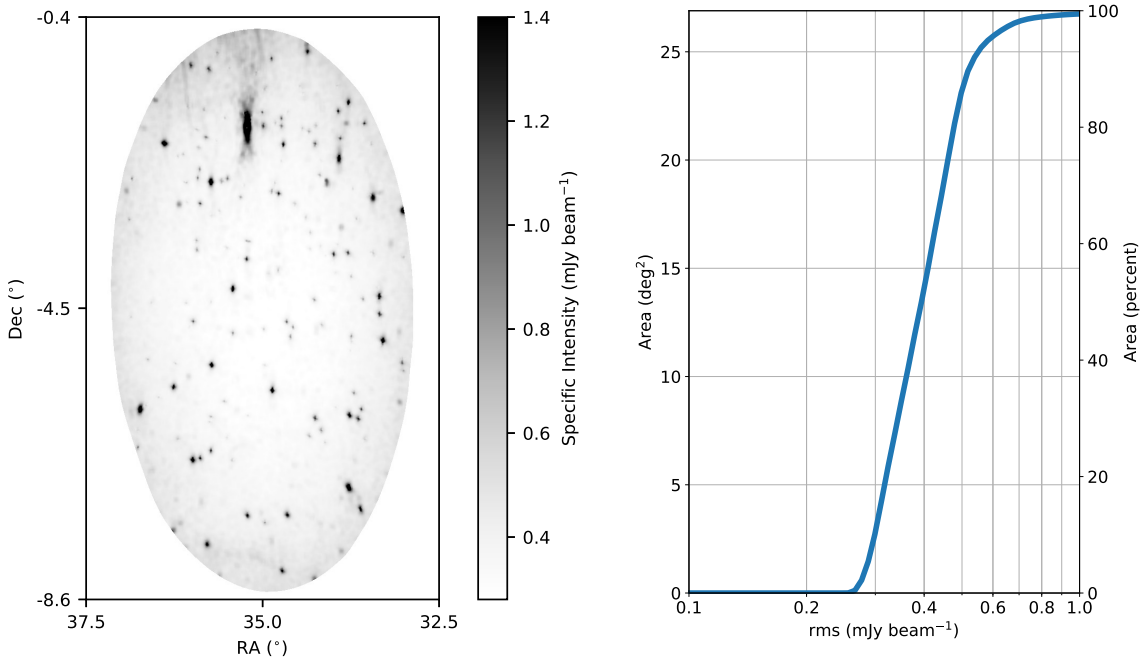
Figure 5.5: Image of the central $\sim$4 deg^2 image of the XMM-LSS field observations imaged with LOFAR at 144 MHz. This is shown on a flux scale between -$\sigma_{\rm cen}$ and $10\sigma_{\rm cen}$, where $\sigma_{\rm cen}$ is the central rms of the image, 0.28 mJy beam^{-1}.

these previous studies have been developed to best reflect the data. This includes using settings such as the `atrous` setting described in Chapter 4 in order to account for the large amounts of extended emission that will exist within the image. Following the work of Williams et al. (2016), the PyBDSF parameters that were chosen are given in Table 5.2. The source catalogue is used for all further work in this Chapter. This is where the individual Gaussian components found by PyBDSF have been merged, if this is appropriate (see Chapter 4). This catalogue contains 3,169 sources.

PyBDSF also produces, alongside the catalogue, an associated rms map for the image. This is shown in Figure 5.6(a). This rms map has a central value of $\sim$280 μJy/beam, with a corresponding median rms across the entire image of $\sim$400 μJy/beam. Figure 5.6(b) shows the corresponding area (or the percentage of total area) that has an rms value below a given flux density per beam. Figure 5.6(a) illustrates how the rms can often appear larger around bright sources.

To ensure that all multi-component sources within the 3,169 sources were combined together into a single source, I performed a visual inspection of every source. This was used to both identify whether a source needed combining with other components (for example if it was a lobed FRII galaxy) as well as identifying artefacts that may have accidentally been misidentified as a source. For the sources that were considered to be matches or possible matches, these were each visually inspected using `ds9` (Joye & Mandel, 2003), making use of comparisons to the observations from Heywood et al. (in prep) (see Section 4.3.1) where possible.

For some sources, $\sim$50, two sources appeared to be co-located. This consisted of a smaller, central component surrounded by a much larger Gaussian component. For these sources, the combined and uncombined fluxes of the two components were compared to those from TGSS-ADR. This comparison implied these sources should be merged. After merging together components and removing artefacts, 3,044 sources remained within the final catalogue. For the final catalogue it was essential to update



(a) (b)

Figure 5.6: The rms coverage of the LOFAR observations of the XMM-LSS field. Panel (a) shows the rms coverage across the image with a flux scale varying between $\sigma_{\rm cen}$ and $5\sigma_{\rm cen}$. Panel (b) shows the area of the image (and percentage of the total area) that has an rms value less than a given flux density.

thresh_isl=3.0	thresh_pix=5.0	rms_box=(160,50)
rms_map=True	mean_map='zero'	ini_method='intensity'
adaptive_rms_box=True	adaptive_thresh=150	rms_box_bright=(60,15)
group_by_isl=False	group_tol=10.0	atrous_do=True
flagging_opts=True	flag_maxsize_fwhm=0.5	advanced_opts=True
blank_limit=None	atrous_jmax=3	

Table 5.2: The parameters used in PyBDSF to determine the output final catalogue as used in Williams et al. (2016).

the information from the original PyBDSF catalogue to account for these merged components.

To assign properties to the merged components, I followed the method of Williams et al. (2016) and used a flux weighted position to describe the source location whilst the total flux density is the sum of the individual flux densities of all merged components. The peak flux density adopted corresponds to the maximum peak flux of the components that were merged together. Determining size properties of composite sources, on the other hand, is more challenging and so the approach of Hardcastle et al. (2016) was adopted. This led to a new column in the final catalogue referred to as the 'Composite Size'. This corresponded to the maximum distance between combined components. This size is likely to be underestimated though for the $\sim$50 sources described above that were comprised of a small and larger co-spatial component. I also included in the catalogue identifiers to indicate whether a source was bright and surrounded by many artefacts as well as another column to indicate that a source was close to the edge of the field of view. These sources were identified during the visual inspection of sources. For those sources close to the edge of the field, this flag was to identify those sources where flux may have been missing as it was cut off by the edge of the image.

The first few lines from the catalogue are presented in Table 5.3. A later row is also included to highlight the different information available for a composite source. This catalogue also includes the value of the rms at the central location of the source as well as a correction to the position of the source using the catalogues from FIRST (Becker et al., 1995; Helfand et al., 2015). The method to determine this positional offset is discussed in Section 5.5.2.

5.4.1 Source Images and comparison with ProFound

Images of six of these large, composite sources are shown in Figure 5.7 (left panel for each source). This includes two of the largest, complex morphology sources that are observed with LOFAR within this field, which highlight LOFAR's sensitivity to extended emission. The corresponding model that PyBDSF generated for these sources is also shown in Figure 5.7 (right panel for each source). As I mentioned in Chapter 4, extended sources may be more challenging to extract with PyBDSF than ProFound. Therefore, I also make comparisons in Figure 5.7 to the source model produced with ProFound (central panel for each source), using the same `skycut` as determined in Chapter 4 to highlight how ProFound may be used in these investigations. Whilst some extended haloes may be observed with PyBDSF for some of these sources, in general they successfully model the source. As the parameters used in PyBDSF have been tailored to work successfully with these low frequency observations, the PyBDSF catalogue is likely to therefore accurately measure source flux densities. This is especially true here as the parameters used for ProFound have not been tailored to the image. The image of the field centre (Figure 5.5) also suggests that the image may have some noisier regions, more so than within the VLA image from Heywood et al. (in prep). This is also evident in the top right extended source in Figure 5.7, which appears to have negative emission surrounding the centre of the source. This may be more challenging for ProFound to deal with and would need further investigation to find the appropriate parameters in ProFound to use. Therefore the catalogue from PyBDSF is used for this work.

IAU_source_ID	Source_ID	Pre_matched _source_ID	RA (°)	E_RA (°)	DEC (°)	E_DEC (°)	Total_flux (mJy)	E_Total_flux (mJy)	... ...
J022831.74-044607.2	0	0	37.13225	0.00003	-4.76867	0.00004	48.38	1.05	...
J022831.48-041625.0	1	1	37.13118	0.00008	-4.27361	0.00006	7.42	0.89	...
J022827.74-042647.2	2	2	37.11558	0.00024	-4.44643	0.00021	5.34	0.80	...
J022826.77-041752.8	3	3	37.11155	0.00005	-4.29800	0.00004	13.20	0.89	...
J022814.35-050242.1	17	17	37.05885	1.10995	-5.04834	-0.15114	85.58	1.33	...

IAU_source_ID	...	Peak_flux (mJy/beam)	E_peak_flux (mJy/beam)	Maj (")	E_Maj (")	Min (")	E_Min (")	PA (°)	E_PA (°)	Composite _Size (")	...
J022831.74-044607.2	...	22.87	0.49	13.12	0.37	8.39	0.17	123.56	2.77	nan	...
J022831.48-041625.0	...	7.09	0.52	9.01	0.73	7.42	0.50	77.26	16.65	nan	...
J022827.74-042647.2	...	3.08	0.55	10.95	2.01	10.09	1.74	90.29	93.35	nan	...
J022826.77-041752.8	...	11.92	0.53	9.00	0.43	7.85	0.33	114.95	14.14	nan	...
J022814.35-050242.1	...	21.29	0.49	nan	nan	nan	nan	nan	nan	36.59	...

IAU_source_ID	...	N _sources	Matched _ID_1	...	Bright	Edge	rms (mJy/beam)	RA _FIRST (°)	DEC _FIRST (°)
J022831.74-044607.2	...	1	0	...	1	0	0.49	37.13234	-4.76854
J022831.48-041625.0	...	1	0	...	0	0	0.52	37.13127	-4.27348
J022827.74-042647.2	...	1	0	...	0	0	0.51	37.11568	-4.44630
J022826.77-041752.8	...	1	0	...	0	0	0.52	37.11164	-4.29787
J022814.35-050242.1	...	2	20	0	0	0	0.53	37.05894	-5.04821

Table 5.3: Extract of the source catalogue for the 3,044 sources observed in the XMM-LSS field. Included in this are the source ID, positional information, flux measurements (and errors), size information and flags to indicated whether multiples components have been combined together into a single source or whether the source is bright or near the edge of the field of view. Also included is information on the rms at the source location as well as a corrected position to FIRST.

5.4.2 Noise Compared to Previous LOFAR Observations

The value of the noise found for this image will be different to those found in Williams et al. (2016); Hardcastle et al. (2016) due to differences in the exposure time of these observations as well as the differences in the declination of the observations. As the LOFAR surveys plan to observe the entire northern sky (van Haarlem et al., 2013; Shimwell et al., 2019), it is therefore useful to compare the sensitivity of these observations to those at higher declinations. The sensitivity of observations has a dependence on exposure time (Δt) and declination given by:

$$\text{Sensitivity} \propto \frac{1}{\sqrt{\Delta t}} \frac{1}{\cos^2(\text{Dec} - \text{Dec}_{\text{best}})} \tag{5.1}$$

where whilst the time dependence is a known dependence (see the description in Taylor et al., 1999), the $\cos^2$ dependence on declination was determined through observations of the MSSS Survey (Heald et al., 2015). I was informed of this sensitivity dependence by T. Shimwell (in private communication). The optimal sensitivity of LOFAR is at $\sim 52^\circ$ (see Shimwell et al., 2019). Previous observations with LOFAR by Williams et al. (2016) of the Boötes field observed for 8 hours at 34.5° Declination. These observations had a central rms of $\sim$ 0.11 mJy/beam. For Shimwell et al. (2017) instead, each pointing was 8 hours in duration and had sensitivities of $\sim$0.1 mJy/beam (without direction dependent calibration) at a declination of $\sim 51^\circ$. Accounting for differences in observation duration and declination, these both suggest that a central rms of $\sim$0.3 mJy/beam is expected for the observations of the XMM-LSS field that are presented in this Chapter. This is similar to what I have observed for these observations. For the single pointing observation of the Boötes Field (Williams et al., 2016), 50% of the area of the image had an rms less than $\sim$ 0.17 mJy/beam which would suggest 50% of this image should have rms values of $\sim$ 0.4 mJy/beam, which is in agreement with that seen in Figure 5.6. This suggests the data reduction process has successfully cleaned the image to an expected noise level.

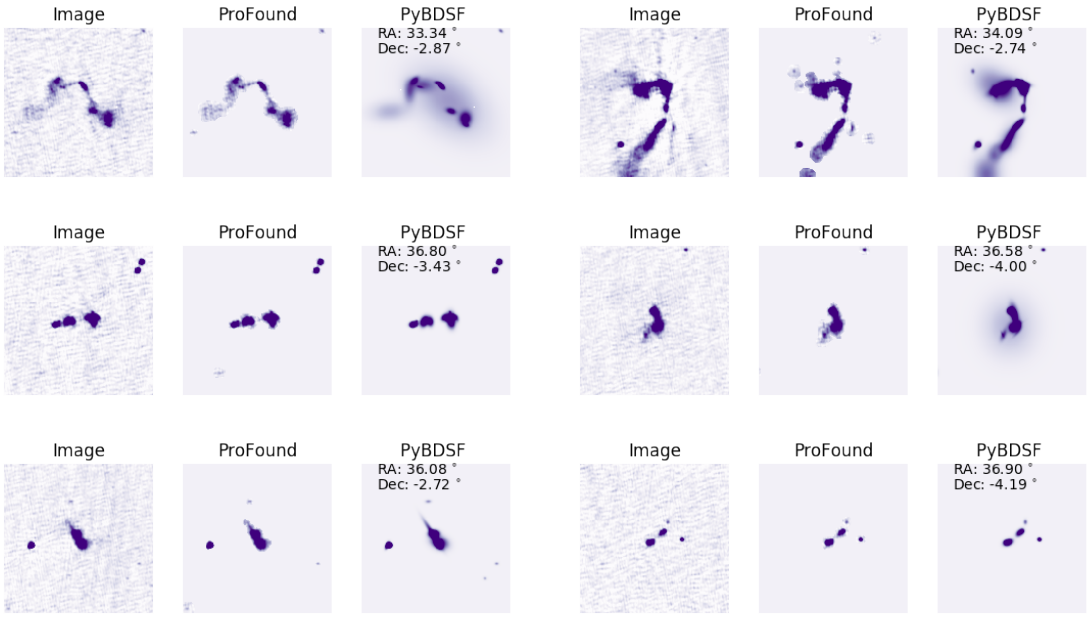


Figure 5.7: Example images (left) of six large, extended sources observed in the LOFAR XMM-LSS field observations as modelled with both ProFound (centre) and PyBDSF (right). Each image is 7.5' across.

5.5 Comparison to Other Radio Observations

It is important to assess the accuracy of these observations. Accurate fluxes are essential for calculations of the luminosities of these sources (as in Chapter 3) as well as in determining the source count distribution (see Figure 1.6). Futhermore, accurate positional information is key for follow up observations of radio sources using higher resolution instruments, such as for WEAVE-LOFAR (Smith et al., 2016). It is therefore important to make comparisons with previous observations, in this case at radio frequencies. High radio frequencies are especially important for comparisons of astrometry as they are less affected by the ionosphere, which can move the positions of sources (see Section 5.1). The radio observations that I make comparisons to are described below. Whilst some of these observations were introduced in Section 1.6, the information is reiterated here as their resolution and sensitivity will be important for these investigations. The large area radio catalogues used, in order of frequency,

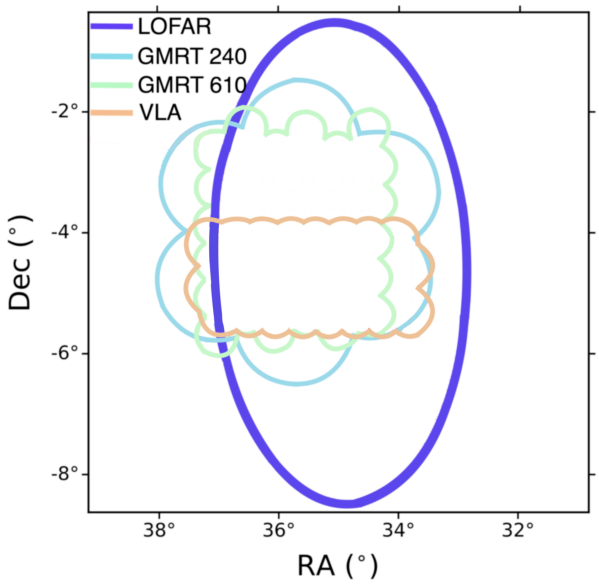


Figure 5.8: Locations of overlapping target radio observations of the XMM-LSS field. The outline of the region observed with LOFAR is shown in purple along with the outlines of observations from Heywood et al. (in prep) (orange) and Tasse et al. (2007) at 240 MHz (blue) and 610 MHz (green).

are:

- *TGSS-ADR (Intema et al., 2017)*: As introduced in Section 1.6, this survey is based on observations with the GMRT at 150 MHz. The observations cover the northern sky at 25" resolution and were used, as discussed above, to provide a sky model for the DDF-pipeline. As these observations are at a similar frequency to those of the XMM-LSS field with LOFAR, presented here, they are useful for comparing source flux densities for some of the brighter sources within the image. The typical rms noise for these observations is ∼3.5 mJy/beam.

- *GLEAM (Hurley-Walker et al., 2017)*: The MWA GLEAM Survey again operates at similar frequencies to LOFAR (70-230 MHz) making it useful for flux comparisons. However its resolution (2') is much poorer than LOFAR or the GMRT and it is also less sensitive (rms ∼6-10 mJy/beam).

- *FIRST (Becker et al., 1995; Helfand et al., 2015)*: At 1.4 GHz, FIRST has 5"

resolution and observes the northern sky to a typical rms of ∼0.15 mJy/beam. It will be important in assessing the astrometric accuracy of the LOFAR XMM-LSS field observations.

As well as these, I made use of smaller area, targeted observations of the XMM-LSS field. The overlap between these observations are presented in Figure 5.8. These observations are:

- *240 MHz Observations (Tasse et al., 2007)*: These observations were targeted pointings mosaiced together to observe the XMM-LSS field using the GMRT at 240 MHz. They have ∼14 deg^2 overlap with the LOFAR field of view (Figure 5.8). These observations are at ∼15" resolution to a depth of ∼2.5 mJy/beam depth.

- *610 MHz Observations (Tasse et al., 2007)*: These observations with the GMRT are instead at 610 MHz and have ∼11 deg^2 of overlap with the XMM-LSS field observations presented in this Chapter (Figure 5.8). The observations have 6.5" resolution and an rms sensitivity of ∼0.3 mJy/beam.

- *VIDEO VLA (Heywood et al., in prep)*: Finally the 1.5 GHz targeted VLA observations used in Chapter 4 have ∼7 deg^2 of overlap with these LOFAR observations (Figure 5.8). The images in Heywood et al. (in prep) have 4.5" resolution and a typical rms depth of ∼16 μJy/beam.

To make comparisons between the observations presented in this work to previous observations, I used a positional cross match within 5". However as each of these observations have different sensitivities and resolution, further cuts need to be made to ensure that these do not affect flux and astrometry comparisons. I followed the work of Williams et al. (2016) to ensure that:

- There are no neighbouring sources (in the higher resolution observations) within N". Here N is the beam size of the poorer resolution observations. This ensures

that a source within the poorer resolution observations is not in fact multiple sources when observed at higher resolution.

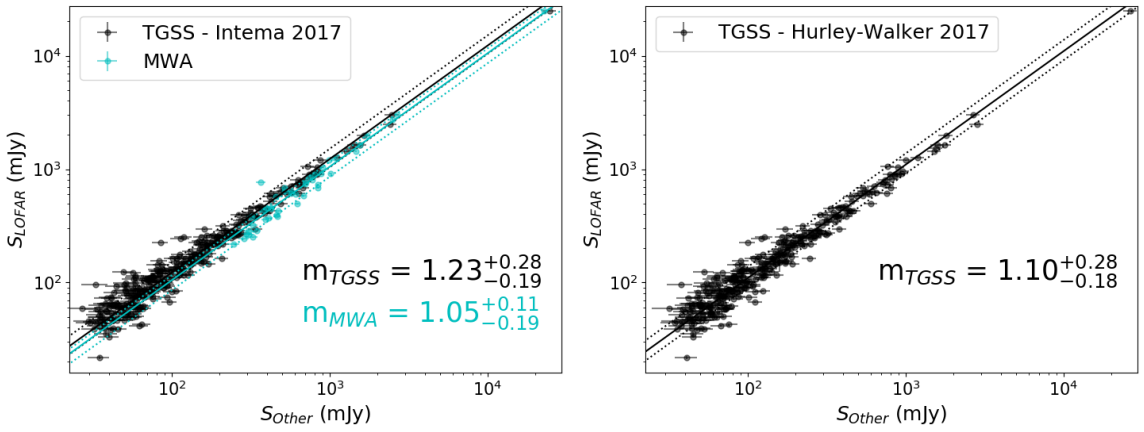
- The size of the source in the higher angular resolution catalogue is small, less than double the size of the angular resolution from the poorer resolution observations. This is to prevent differences in fluxes arising from extended emission in one catalogue that is not detected within another survey due to differing baselines. It also helps reduce issues that may arise from multi component sources.

- Only those sources of high signal to noise are included. This reduces any spurious noise detections that may be observed as a source. The criteria used is: $(S_{peak}/\sigma_{rms} \geq 10)$.

With these source cuts imposed it is possible to make comparisons between these previous observations to the sources observed within the LOFAR image.

5.5.1 Flux comparison

Using the criteria described above left 319 matches between the LOFAR and TGSS-ADR catalogues and 45 sources between LOFAR and MWA. The number of matches between LOFAR and MWA may be underestimated if there are large positional offsets in the MWA catalogue, which would reduce the number of sources that can be compared with. In Hurley-Walker et al. (2017) the mean absolute RA offset is ≤ 4" whilst for declination it is ≤ 2" and so any sources with large offsets would be missed in this analysis. Using the matches with TGSS-ADR and MWA, these resulted in a median flux ratio $(S_{\rm LOFAR}/S_{\rm TGSS})=1.23^{+0.28}_{-0.19}$ between LOFAR and TGSS-ADR and a ratio of $1.05^{+0.11}_{-0.19}$ between LOFAR and GLEAM. The uncertainties quoted on these values arise from the $16^{\rm th}$ and $84^{\rm th}$ percentiles. The comparisons of these fluxes and



(a) (b)

Figure 5.9: Panel (a) shows comparisons of the integrated flux recorded by LOFAR to that recorded by TGSS-ADR (black Intema et al., 2017) and MWA (cyan Hurley-Walker et al., 2017). The best fit lines from the median ratio (solid) and the uncertainties (dashed) are also shown with the corresponding median values and uncertainties given in the bottom left corner. Panel (b) again shows the comparison of LOFAR fluxes to TGSS-ADR but using the corrected data catalogue of Hurley-Walker (2017).

their median offsets are shown in Figure 5.9(a). The median flux ratio from comparisons with MWA observations are closer to 1 and similar to the flux ratios observed in Hardcastle et al. (2016) and Shimwell et al. (2017) for comparisons with TGSS (∼1.1). However only a small number of MWA sources are matched with the LOFAR catalogue. In comparison, when TGSS-ADR was compared to, ∼4 times more source matches were found but the flux ratio is much larger (median ratio=1.23). This suggests there is an issue with the flux scaling of one (or both) of these observations.

However, Hurley-Walker (2017) identified that there are systematic offsets as a function of location between the MWA flux measurements compared to TGSS-ADR. As such, Hurley-Walker (2017) created a corrected catalogue to TGSS-ADR which aimed to account for these differences. These corrections are in the range of ∼1.05-1.25 for the region overlapping with these LOFAR observations. Using this catalogue instead (see Figure 5.9b) reduces the flux ratio observed between LOFAR and TGSS-

ADR, with a median flux ratio of $1.10^{+0.28}_{-0.18}$. This is much more comparable to the work of Hardcastle et al. (2016); Shimwell et al. (2017). The correction factors to the TGSS sources in Hurley-Walker (2017) across the field of view are shown in Figure 5.10. This highlights a correlation between the correction factor needed and the declination of the source. This suggests that the TGSS-ADR observations are systematically offset over this particular field of view. This declination dependence is not observed across all overlap regions though (see Figure 3 of Hurley-Walker, 2017).

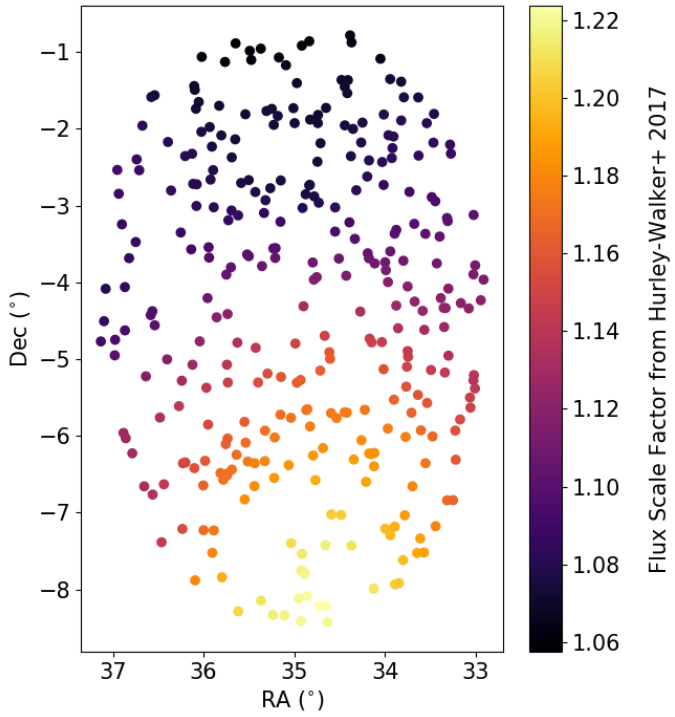


Figure 5.10: The correction factors determined in Hurley-Walker (2017) to correct fluxes from TGSS-ADR as a function of position.

Whilst Williams et al. (2016); Hardcastle et al. (2016) and Shimwell et al. (2019) apply flux corrections to the source catalogue using the flux comparisons with TGSS-ADR, the conflicts between the different flux scalings means that I do not apply such corrections here. However, for the remainder of this Chapter I make comment on the

effect of not having corrected these fluxes to TGSS-ADR.

5.5.2 Positional offsets

Another metric to assess these observations is through considering the astrometric accuracy. Therefore comparisons of the positions for these sources from both radio catalogues are considered. The sources observed in GLEAM, however, are not compared to, due to its poor resolution and low source density of matches. Offsets in RA are defined as ΔRA= $\mathrm{RA_{LOFAR}}$ - $\mathrm{RA_2}$, with ΔDec defined in a similar way. In both of these, the $_2$ relates to the secondary catalogue that the LOFAR observations are matched to. These are calculated for those sources which satisfy the criterion described in Section 5.5. The positional offsets between these LOFAR observations and the five remaining catalogues are shown in Figure 5.11. These are quantified in Table 5.4 by quoting the median offset and uncertainties from the 16^{th} and 84^{th} percentiles. Also included in Table 5.4 is an estimate for the mean and width of a Gaussian that fits the histogram distribution (see Figure 5.11) of these offsets.

From Figure 5.11 and Table 5.4, comparisons with the GMRT from TGSS-ADR and the GMRT 240 MHz observations from Tasse et al. (2007) have the smallest positional offsets in both RA and Dec. These offsets are found to be, on average, $\lesssim 0.15$". Contrary to this, at higher frequencies much larger RA and Dec offsets are observed, in the range of $\sim 0.35 - 0.5$". As higher frequency radio observations are less susceptible to the effects of the ionosphere, the positional information from these are likely to be more accurate. This suggests there are positional offsets of $\sim$0.4" that need to be accounted for. However, for the GMRT 610 MHz observations, whilst the Declination offsets are in the same direction as those from FIRST and the VIDEO VLA observations, the right ascension offset acts in the opposite direction. The results from comparisons with FIRST and VIDEO VLA however are much more similar to one another and, as seen in Table 5.3.

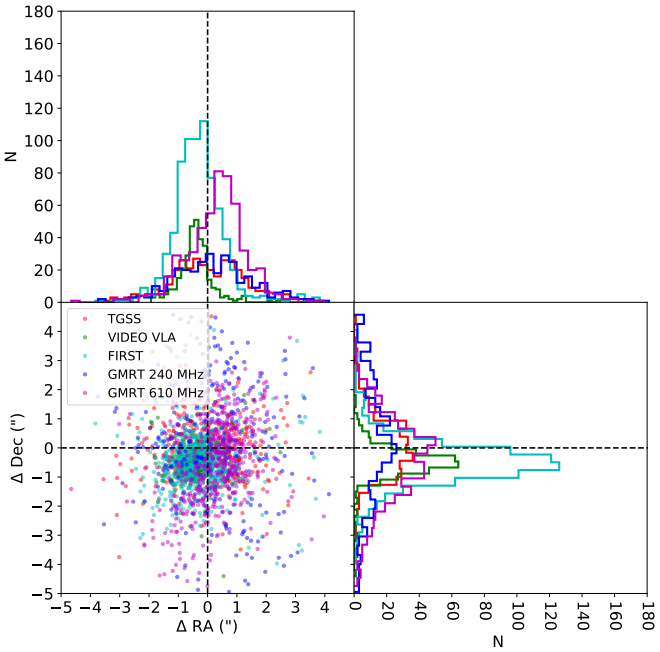


Figure 5.11: Positional RA and Dec offsets of the LOFAR sources compared to TGSS-ADR (red; Intema et al., 2017), GMRT 240 MHz observations (dark blue; Tasse et al., 2007), GMRT 610 MHz observations (magenta; Tasse et al., 2007), FIRST (light blue; Becker et al., 1995; Helfand et al., 2015) and VIDEO VLA (green; Heywood et al., in prep).

For Declination especially, the GMRT 240 MHz and 610 MHz observations have much larger uncertainties in these offsets ($\sigma \sim 1-2$) compared to those in RA ($\sigma \sim 0.8-1.3$), whilst the VLA (FIRST and VIDEO VLA) and TGSS-ADR observations have more comparable RA and Dec uncertainties. This suggests there is a discrepancy between observations that is more prominent in Declination that is not seen with more recent observations with the GMRT in the TGSS-ADR survey.

A positional correction to the sources observed here with LOFAR is applied using the observed comparisons with FIRST (see Table 5.3). This is a correction of ΔRA= -0.34" and ΔDec= -0.51".

Catalogue	Matches within 5"	RA offset (") *Histogram*	RA offset (") *Gaussian*	Dec offset (") *Histogram*	Dec offset (") *Gaussian*
TGSS-ADR	319	$-0.06^{+1.16}_{-1.10}$	$\mu = -0.00$ $\sigma = 1.19$	$-0.03^{+1.06}_{-0.93}$	$\mu = -0.04$ $\sigma = 0.97$
GMRT 240	310	$+0.10^{+1.38}_{-1.20}$	$\mu = +0.11$ $\sigma = 1.26$	$+0.14^{+2.09}_{-2.02}$	$\mu = +0.22$ $\sigma = 1.99$
GMRT 610	532	$+0.36^{+0.76}_{-1.06}$	$\mu = +0.41$ $\sigma = 0.81$	$-0.41^{+1.29}_{-1.43}$	$\mu = -0.37$ $\sigma = 1.36$
FIRST	738	$-0.34^{+0.72}_{-0.64}$	$\mu = -0.35$ $\sigma = 0.66$	$-0.51^{+0.65}_{-0.60}$	$\mu = -0.51$ $\sigma = 0.58$
VLA	317	$-0.40^{+0.37}_{-0.47}$	$\mu = -0.41$ $\sigma = 0.35$	$-0.47^{+0.43}_{-0.42}$	$\mu = -0.50$ $\sigma = 0.39$

Table 5.4: Astrometric accuracy of the XMM-LSS field LOFAR observations compared to radio observations from TGSS-ADR (Intema et al., 2017), GMRT observations (Tasse et al., 2007), FIRST (Becker et al., 1995; Helfand et al., 2015) and VIDEO VLA (Heywood et al., in prep). Quoted are the median offsets as well as uncertainties from the 16^{th} and 84^{th} percentiles and a best fit model of a Gaussian distribution.

5.6 Spectral Indices

Throughout this Thesis a spectral index, α (Equation 1.19), has been assumed, with a fixed value of 0.7. Whilst this is similar to what is observed in many surveys (e.g. Hardcastle et al., 2016; Intema et al., 2017; Smolčić et al., 2017a), it is important to consider whether this is consistent with these observations. Using the catalogues described above and imposing the requirement of sources in the catalogues being small, isolated and high signal to noise, I was able to compute the spectral indices of these LOFAR sources.

However, the spectral index would automatically be biased if no further cuts on the data were imposed, this is because of differences in the flux limits for the different surveys. If only the spectral index for those sources which were detected in both catalogues were calculated, this will bias the measured α values. This is because, excluding the VLA-VIDEO observations, the LOFAR sensitivity is deeper (assuming

a typical 0.7 spectral index) and so not imposing a flux cut will bias the measured spectral indices towards flatter slopes. I therefore imposed another flux cut such that only those LOFAR sources which could be detected in the secondary catalogue (at 5σ) with a spectral index of α=2 or less are included. As a spectral index of 2 is larger than expected for most sources, this ensures that α will not be biased to smaller values. The calculated spectral indices are shown as a histogram in Figure 5.12 and the distributions of these are given in Table 5.5, using the same format as in Table 5.4.

Table 5.5 shows that the higher frequency measurements with the VLA (i.e. FIRST and VIDEO VLA) and the GMRT (610 MHz) have spectral index values peaking around the assumed value of $\sim 0.7-0.8$ that is used throughout this Thesis. The median values for α and their uncertainties are: $0.70^{+0.27}_{-0.24}$ (GMRT 610 MHz), $0.79^{+0.26}_{-0.22}$ (FIRST), and $0.76^{+0.18}_{-0.24}$ (VLA). However, this is very different to the values found for the 240 MHz observations which suggests a spectral index of $1.26^{+0.52}_{-0.36}$. This is much steeper than the expected value of 0.7. However the frequencies considered here (between the two catalogues) are much more similar. Any issues in the flux scalings between the two catalogues will be amplified here, due to the small frequency range being considered.

As the value of α is determined through measuring fluxes, any offsets in the flux scale (see Section 5.5.1) will affect the measured values. If the flux correction from the TGSS-ADR catalogue in Intema et al. (2017) was applied ($S_{\rm LOFAR}/S_{\rm TGSS}$=1.23) this would decrease measurements of α, resulting in $\alpha^{144}_{240} \sim 0.85$. However the values for the spectral indices using the higher frequency measurements would also decline, in the range ∼0.55-0.7 which is towards the lower side of what is expected. Given that the true flux densities may be a compromise between the uncorrected fluxes and those corrected with TGSS-ADR from Intema et al. (2017), this suggests the assumption of $\alpha \sim 0.7$ used throughout this Thesis is appropriate and can continue to be used.

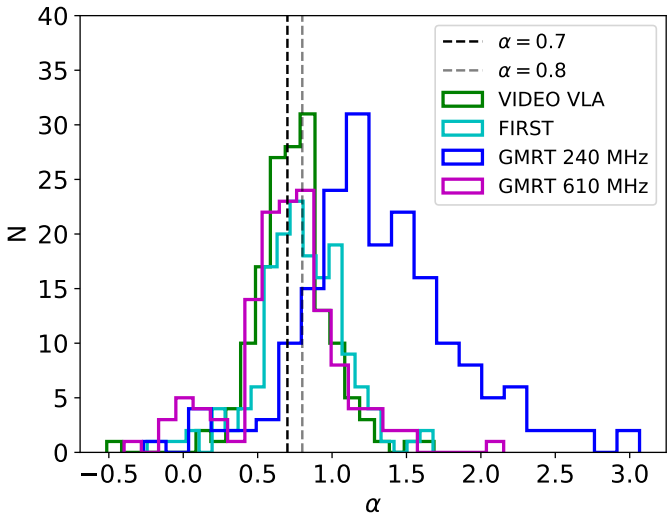


Figure 5.12: Histograms showing the measured spectral index, α, of LOFAR observations compared to radio observations from GMRT 240 MHz observations (dark blue; Tasse et al., 2007), GMRT 610 MHz observations (magenta; Tasse et al., 2007), FIRST (light blue; Becker et al., 1995; Helfand et al., 2015) and VIDEO VLA (green; Heywood et al., in prep).]

Catalogue	Frequency (MHz)	Matches within 5"	α *Histogram*	α *Gaussian*
GMRT	240	192	$1.26^{+0.52}_{-0.36}$	$\mu = 1.25$, $\sigma = 0.41$
GMRT	610	141	$0.70^{+0.27}_{-0.24}$	$\mu = 0.72$, $\sigma = 0.22$
FIRST	1400	157	$0.79^{+0.26}_{-0.22}$	$\mu = 0.81$, $\sigma = 0.24$
VLA	1500	155	$0.76^{+0.18}_{-0.24}$	$\mu = 0.73$, $\sigma = 0.20$

Table 5.5: Measured spectral indices of sources in the XMM-LSS field LOFAR observations compared to radio observations from the GMRT (Tasse et al., 2007), FIRST (Becker et al., 1995; Helfand et al., 2015) and VIDEO VLA (Heywood et al., in prep). Quoted are the median offsets as well as uncertainties from the 16^{th} and 84^{th} percentiles and a best fit model of a Gaussian distribution.

5.7 Source counts

Finally I present the source counts observed for these observations. The source counts show the distribution of extragalactic radio sources as a function of flux density (see Figure 1.6). However, the catalogue extracted by PyBDSF will not be a true representation of the underlying source distribution. This is due to limitations in the detection of sources across the image. Difficulties in detecting sources may arise from:

- As discussed in Section 4.4, close to the detection limit, noise may be confused as real emission and included within the measured catalogue. Therefore a correction will need to decrease the measured source counts to reflect the percentage of false detections.
- Non-uniform noise across the image (as seen in Figure 5.6) results in faint sources unable to be detected equally across the image. This is especially an issue at the edge of the field of view and around bright sources.
- As the inclusion of sources within a catalogue is often dependent on the peak flux of a source (see Chapter 4), the peak flux will be smaller for a resolved source than for an unresolved one. As such, larger sources with the same total flux will be preferentially be harder to detect (resolution bias).
- Finally, as identified in Chapter 4, PyBDSF may struggle in detecting source of given morphologies or fluxes. The observed source counts will therefore need to be corrected to account for any limitations in the detection of sources using PyBDSF.

All of these effects need to be corrected for in order to determine the intrinsic population of sources as a function of flux density.

5.7.1 False detection

As described in Section 4.4, it is possible to determine the number of false detections using the negative image. Whilst in the work of Section 4.4 I only calculate the total number of false detections, the source counts need to be corrected for this effect as a function of flux density. This is because it is the fainter flux density bins that preferentially contain noise that has been mistaken for a real source. However, when the false detection fraction was first investigated, there were significant corrections necessary at much higher fluxes. This is because the `adaptive_rms_box` setting used in PyBDSF (see Table 5.2) ensures that around bright sources a smaller region is used to estimate the rms than at other locations in the image. This reduces the regions of high rms in regions further from a bright source, but will increase the rms close to it. With higher rms values near bright sources, fewer artefacts will be included within the source catalogue. When sources are measured in the negative image, artefacts are more likely to be included in the catalogue. This is because there are no bright sources and so the rms box size is not smaller near the source. To correct for this, I masked out the regions (using the package `PyRegion`; `https://pyregion.readthedocs.io/en/latest/index.html`) surrounding bright sources (>0.1 Jy). This masked out all regions within an angular radius of 200".

5.7.2 Completeness simulations

To account for detection sensitivity, resolution bias and completeness I followed the method in Section 4.3.3 and used simulations to determine the recovery rate of sources as a function of flux density. I used elliptical sources with sizes from Wilman et al. (2008) convolved with the LOFAR clean beam. Whilst the work in Section 4.3.3 used the multi-component sources from Wilman et al. (2008) separately to the single components, here a combination of both source types are injected into the image. Flux densities are associated to the sources using the scaled 1.4 GHz flux densities to

144 MHz from Wilman et al. (2008). As the sizes in Wilman et al. (2008) are realistic size distributions of sources, this should therefore account for resolution bias. Whilst work by Prandoni et al. (2001); Mahony et al. (2016) and Williams et al. (2016) adopt the Windhorst correction (Windhorst et al., 1990) in order to correct for resolution bias, here I combine resolution bias, completeness and non-uniform detection across the image into a single simulation.

To determine the necessary correction factors, 100 simulations were constructed in which 1,500 simulated sources were input into the final image. Only sources that had total flux densities >1 mJy at 144 MHz were included. Comparisons between the recovered and input sources at different flux densities (accounting for sources already within the image) therefore allowed me to calculate the total completeness correction necessary to apply to each flux bin. For a flux density bin (i) this completeness correction is given by:

$$\text{Completeness correction}_{\text{i}} = \frac{N_{\text{injected,i}}}{N_{\text{recovered\_from\_injected,i}}}, \tag{5.2}$$

where $N_{\text{injected,i}}$ is the number of sources within the flux bin that are injected into the image and $N_{\text{recovered\_from\_injected,i}}$ is the number recovered. There are two methods by which $N_{\text{recovered\_from\_injected}}$ can be determined. In Method 1, I calculated $N_{\text{recovered\_from\_injected}}$ by subtracting the source count distribution of the catalogue described in Section 5.4 from the output catalogue of each simulation. In Method 2, $N_{\text{recovered\_from\_injected}}$ is calculated through positional matching of the input source catalogue to the output sources, within a given angular radius. Similar to Section 4.3.3, a 5" radius is used to match to the PyBDSF catalogue of sources in the original image to those recovered from the simulated image. Any matches are removed from the output catalogue of the simulations. This corrected output catalogue is then positionally matched to the input simulation catalogue within a 15" radius. This resulted in ∼98% of sources from the corrected output catalogue being matched to a simulated

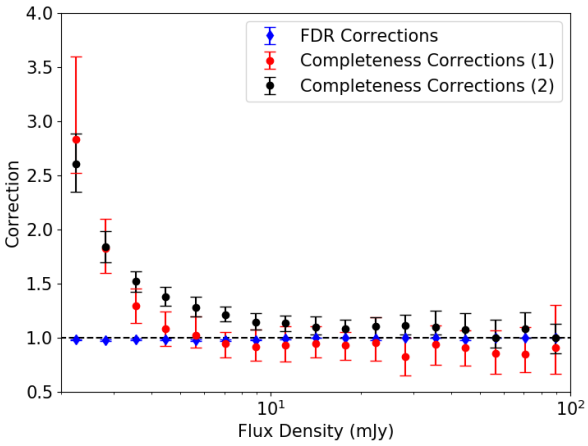


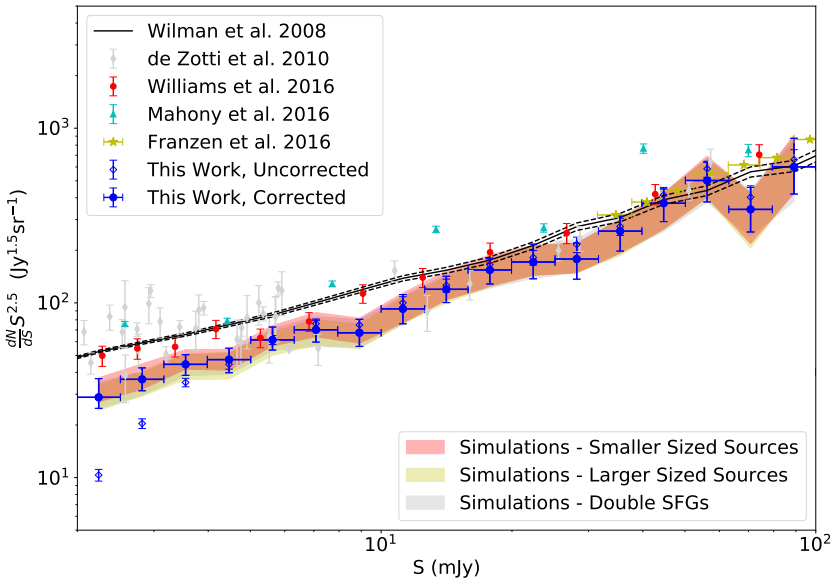
Figure 5.13: Correction factors calculated and then applied to the obtained source counts from the LOFAR XMM-LSS catalogue. These corrections are for false detections (blue) and completeness (using methods 1 (red) and 2 (black)).

input source.

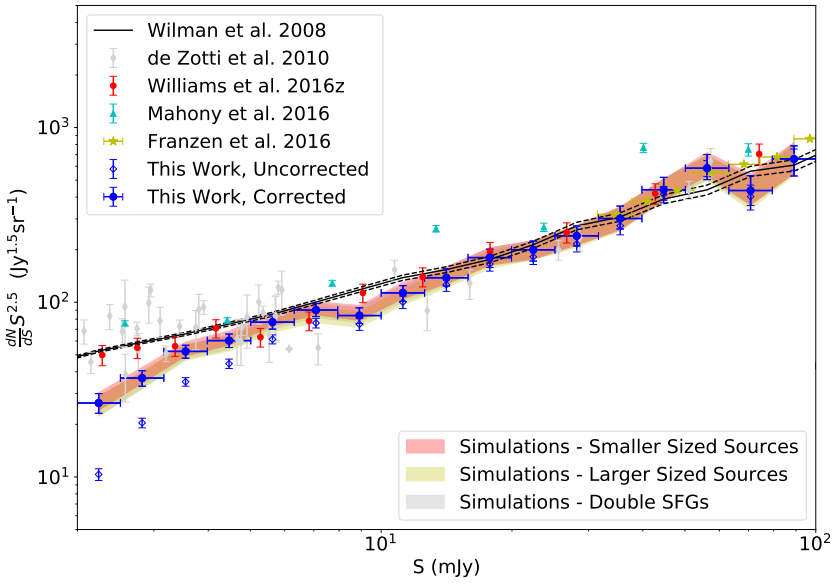
Both Methods 1 and 2 take into account Eddington bias, where faint sources are pushed above the detection threshold if they land on a noise spike. However, Method 1 will be affected when sources merge together within the image and are measured as a single source. Finally, Method 1 can account for differences in the extracted flux densities to what is simulated, see Section 4.3.3. However, due to the latter two reasons, correction factors <1 can be obtained. This will occur as sources move in and out of flux density bins. As 100 simulations were run, the correction factor is calculated as the median correction factor for a flux density bin. The associated uncertainties are calculated again from the 16^{th} and 84^{th} percentiles.

5.7.3 Corrected Source Counts

The correction factors determined using the methods described in Sections 5.7.1 and 5.7.2 are shown in Figure 5.13. These are applied by multiplying the corrections in each bin with the naive source counts. The associated errors are combined in quadrature.



(a) Method 1.



(b) Method 2.

Figure 5.14: Measured (blue open diamonds) and corrected (blue filled circles) source counts from the LOFAR XMM-LSS catalogue presented in this Chapter compared to previous studies from Franzen et al. (2016) (yellow stars); Williams et al. (2016) (red circles); de Zotti et al. (2010); Mahony et al. (2016) (cyan diamonds) and Wilman et al. (2008). The shaded regions indicate the results when using smaller/larger source sizes, see Section 5.7.4. This is shown using completeness Method 1 (a) and Method 2 (b).

The final source counts are presented in Figure 5.14 and the associated values are given in Tables 5.6 and 5.7 for Methods 1 and 2. The uncorrected source counts are indicated by blue open markers, whilst the corrected values use filled blue markers. Also shown in Figure 5.14 are the measured source counts from Mahony et al. (2016); Williams et al. (2016) and Franzen et al. (2016) (at ∼150 MHz), as well as from the compilation of source counts in de Zotti et al. (2010). The scaled source counts from Wilman et al. (2008) are also presented for comparison.

As Figure 5.13 shows, the completeness corrections dominate compared to those calculated to account for false detections. The false detection rate corrections are ∼1 across the range of flux densities considered. The completeness corrections, on the other hand, deviate from ∼1 at flux densities $\lesssim$4 mJy (Method 1) and $\lesssim$10 mJy (Method 2). Below these flux densities, the correction factors increase steadily to a maximum value of ∼3. As mentioned above, some of the completeness correction factors found using Method 1 are smaller than 1. This suggests that the recovered flux densities for these sources are different to those input in the simulation. This is expected as using PyBDSF for sources with an elliptical morphology in the simulations of Chapter 4 determined preferentially larger flux densities. Therefore the recovered source count distribution will differ from the input distribution (even at high fluxes) as sources move between the flux density bins. This will also be affected by the merging together of sources.

The comparison of corrected source counts determined in this Chapter, compared to previous work (Figure 5.14), shows that the source counts determined here are similar to previous work. Using Method 1 (panel a), the source counts appear systematically lower than previous work, especially at flux densities $\lesssim$10 mJy. Whilst at high flux densities, the source counts are also underestimated due to the reasons discussed above. With Method 2 (Figure 5.14b), on the other hand, the observed source counts are much more similar to previous observation for flux densities $\gtrsim$3

mJy. However the source counts are still low at the faintest flux densities. As these source count measurements rely on accurate flux density measurements, it is also important to discuss the results if the flux corrections determined in Section 5.5.1 were applied to the data. When the 1.23 flux density correction factor from comparisons with TGSS-ADR were applied, this moved the source counts obtained in this Chapter further away from the previous observations shown in Figure 5.14. This suggests such a large flux density correction is inappropriate for this data.

As the source counts using Method 2 are much higher than Method 1 at high fluxes, this agrees with the idea that the accuracy of measuring flux densities for this source morphology is important, as well as the effect of merging sources. However, the fact though that using both methods giver a source count distribution that is lower than expected (at the faintest fluxes) suggests that there is another unknown issue that is affecting the detection of sources, which has not been accounted for here. This will be discussed in Section 5.7.4.

5.7.4 Further Source Count Simulations

As the corrected source counts at low flux densities were biased low for Methods 1 and 2, I further investigated whether this could be a result of the assumed source model from Wilman et al. (2008). I therefore conducted three further simulations (each with 100 runs) to determine whether the distribution of the input simulated sources may be affecting the results presented above. These further simulations involved:

1. *Increasing the number of SFGs within the Wilman et al. (2008) input catalogue:* Deep source counts by Smolčić et al. (2017a) have suggested that the source count distribution at the faintest fluxes are in fact larger than used in Wilman et al. (2008). Therefore I doubled the number of SFGs in the input S^3 model to increase the number of faint SFGs being injected into the simulated images.

2. *Increasing or decreasing the sizes of the components of the Wilman et al. (2008) sources by 20%:* This would help investigate whether resolution bias has been sufficiently accounted for.

These further tests are shown in Figure 5.14 as shaded regions. These suggest that changes to the assumed source populations will have a minimal impact on the distribution of the corrected source counts. This therefore leaves the conclusion that either there is an unknown correction that is not properly accounted for, or that the assumed source morphology for these simulations is important. This morphology was chosen though to reflect what are thought to be accurate source populations.

5.8 Conclusions

In this Chapter I have presented observations of the XMM-LSS field at 144 MHz using three four hour observations with LOFAR. Whilst this is a challenging field to observe with LOFAR, due to its low elevation, the work in this Chapter has shown it is possible to produce a high fidelity image at $\sim$8" resolution of this field. After inspection of the sources, the removal of artefacts and the merging of multi-component sources, this resulted in a catalogue of 3,044 sources. This final image was observed to have a central rms of $\sim$280 μJy/beam and a median rms of $\sim$400 μJy/beam across the entire $\sim$27 deg^2 of the image. Whilst the rms values are larger than observed with other LOFAR observations (Williams et al., 2016; Hardcastle et al., 2016; Shimwell et al., 2019), these rms measurements are in line with the expectations given the known sensitivity dependence of the telescope with declination and the sensitivity response with exposure time.

Through comparing this work to previous observations, the ratio of fluxes compared to TGSS-ADR is larger than with previous LOFAR observations. The comparison between the fluxes measured in these two catalogues resulted in a flux ratio of

$S_{\rm LOFAR}/S_{\rm TGSS}$=1.23. Previous work with LOFAR has instead found flux ratios $\sim$1.1 (e.g. Hardcastle et al., 2016; Shimwell et al., 2017). This ratio is reduced when the fluxes are instead compared with the MWA ($\sim$1.05) or using a rescaled version of the TGSS-ADR catalogue from Hurley-Walker (2017) (ratio of $\sim$ 1.1). Previous radio observations across the field were also used to compare the astrometric accuracy of these observations, which are typically constrained to $\lesssim$ 0.4". However, there is variation between the astrometric uncertainties found between the different surveys for which these observations are compared to. Good astrometry is essential for follow up observations with WEAVE-LOFAR (Smith et al., 2016). Previous radio observations have also allowed the spectral index distribution of sources to be measured. This suggests a value of $\alpha\sim$0.7 for comparisons with many of the observations compared to at other frequencies, in agreement with the value of assumed throughout this Thesis. There is however much steeper spectral indices measured when compared to observations at 240 MHz, although this difference in frequency are small between these measurements which means any flux offsets will make a large difference to the measured spectral index.

Finally, in this Chapter, I have also presented the measured and corrected source counts within this field compared to previous observations. This suggests that whilst the source counts determined are consistent with previous measurements, at the lowest flux densities the number of sources are underestimated. This suggests that there is an unknown factor that is reducing the number of sources observed and so needs to be accounted for. Despite this, the work of this Chapter has shown that it is possible to continue the work of LoTSS (Shimwell et al., 2019) to lower elevations and create high fidelity images across the entire northern sky using LOFAR.

S (mJy)	S_c (mJy)	N	$\frac{dN}{dS}S^{2.5}$ (Jy$^{1.5}$)	FDR	Completeness Correction	Corrected $\frac{dN}{dS}S^{2.5}$ (Jy$^{1.5}$)	Comp. SFG	Comp. Small	Comp. Large
2.00 - 2.51	2.24	169^{+13}_{-13}	$10.34^{+0.80}_{-0.80}$	0.98 ± 0.01	$2.84^{+0.76}_{-0.32}$	$28.81^{+8.03}_{-3.91}$	$2.90^{+0.54}_{-0.40}$	$3.11^{+0.57}_{-0.42}$	$2.84^{+0.47}_{-0.40}$
2.51 - 3.16	2.82	236^{+15}_{-15}	$20.40^{+1.30}_{-1.30}$	0.98 ± 0.01	$1.83^{+0.27}_{-0.23}$	$36.51^{+5.94}_{-5.11}$	$1.73^{+0.25}_{-0.22}$	$1.91^{+0.33}_{-0.29}$	$1.69^{+0.31}_{-0.18}$
3.16 - 3.98	3.55	287^{+16}_{-16}	$35.04^{+1.95}_{-1.95}$	0.98 ± 0.01	$1.29^{+0.16}_{-0.16}$	$44.48^{+6.03}_{-6.01}$	$1.31^{+0.14}_{-0.19}$	$1.39^{+0.17}_{-0.18}$	$1.25^{+0.22}_{-0.17}$
3.98 - 5.01	4.47	258^{+16}_{-16}	$44.50^{+2.76}_{-2.76}$	0.98 ± 0.01	$1.08^{+0.16}_{-0.16}$	$47.35^{+7.69}_{-7.53}$	$1.02^{+0.14}_{-0.11}$	$1.08^{+0.14}_{-0.12}$	$1.02^{+0.15}_{-0.16}$
5.01 - 6.31	5.62	252^{+15}_{-15}	$61.39^{+3.65}_{-3.65}$	0.98 ± 0.01	$1.02^{+0.18}_{-0.11}$	$61.41^{+11.35}_{-7.72}$	$1.02^{+0.15}_{-0.11}$	$1.07^{+0.16}_{-0.11}$	$0.97^{+0.15}_{-0.12}$
6.31 - 7.94	7.08	221^{+14}_{-14}	$76.05^{+4.82}_{-4.82}$	0.98 ± 0.01	$0.94^{+0.11}_{-0.13}$	$70.15^{+9.48}_{-10.51}$	$0.93^{+0.18}_{-0.17}$	$1.02^{+0.16}_{-0.16}$	$0.89^{+0.17}_{-0.10}$
7.94 - 10.00	8.91	154^{+12}_{-12}	$74.85^{+5.83}_{-5.83}$	0.98 ± 0.01	$0.92^{+0.16}_{-0.13}$	$67.43^{+12.95}_{-11.07}$	$0.92^{+0.16}_{-0.14}$	$0.93^{+0.17}_{-0.11}$	$0.92^{+0.08}_{-0.16}$
10.00 - 12.59	11.22	146^{+12}_{-12}	$100.24^{+8.24}_{-8.24}$	0.99 ± 0.01	$0.93^{+0.18}_{-0.15}$	$92.43^{+19.22}_{-16.55}$	$0.92^{+0.16}_{-0.15}$	$0.93^{+0.20}_{-0.14}$	$0.91^{+0.15}_{-0.12}$
12.59 - 15.85	14.13	130^{+11}_{-11}	$126.08^{+10.67}_{-10.67}$	1.00	$0.95^{+0.16}_{-0.13}$	$119.61^{+22.45}_{-19.23}$	$0.97^{+0.16}_{-0.11}$	$1.00^{+0.20}_{-0.17}$	$0.97^{+0.13}_{-0.16}$
15.85 - 19.95	17.78	121^{+11}_{-11}	$165.76^{+15.07}_{-15.07}$	1.00	$0.93^{+0.12}_{-0.13}$	$154.32^{+24.35}_{-26.00}$	$0.88^{+0.17}_{-0.12}$	$0.93^{+0.19}_{-0.16}$	$0.91^{+0.15}_{-0.14}$
19.95 - 25.12	22.39	94^{+9}_{-9}	$181.90^{+17.42}_{-17.42}$	0.99 ± 0.01	$0.95^{+0.24}_{-0.17}$	$171.41^{+45.51}_{-34.03}$	$0.96^{+0.18}_{-0.17}$	$1.00^{+0.17}_{-0.19}$	$0.98^{+0.18}_{-0.17}$
25.12 - 31.62	28.18	79^{+8}_{-8}	$215.94^{+21.87}_{-21.87}$	1.00	$0.83^{+0.17}_{-0.17}$	$178.31^{+41.73}_{-41.62}$	$0.88^{+0.18}_{-0.16}$	$0.86^{+0.14}_{-0.16}$	$0.84^{+0.16}_{-0.12}$
31.62 - 39.81	35.48	71^{+8}_{-8}	$274.13^{+30.89}_{-30.89}$	1.00	$0.94^{+0.18}_{-0.19}$	$257.50^{+56.14}_{-59.46}$	$0.89^{+0.22}_{-0.16}$	$0.90^{+0.15}_{-0.17}$	$0.89^{+0.17}_{-0.17}$
39.81 - 50.12	44.67	76^{+8}_{-8}	$414.49^{+43.63}_{-43.63}$	0.99 ± 0.01	$0.91^{+0.16}_{-0.17}$	$371.98^{+77.26}_{-80.09}$	$0.84^{+0.17}_{-0.19}$	$0.85^{+0.15}_{-0.16}$	$0.81^{+0.19}_{-0.12}$
50.12 - 63.10	56.23	76^{+8}_{-8}	$585.48^{+61.63}_{-61.63}$	1.00	$0.86^{+0.21}_{-0.19}$	$501.84^{+136.13}_{-123.40}$	$0.85^{+0.28}_{-0.18}$	$0.89^{+0.29}_{-0.19}$	$0.88^{+0.25}_{-0.18}$
63.10 - 79.43	70.79	37^{+6}_{-6}	$402.63^{+65.29}_{-65.29}$	1.00	$0.85^{+0.25}_{-0.17}$	$342.90^{+115.04}_{-88.26}$	$0.81^{+0.26}_{-0.17}$	$0.79^{+0.21}_{-0.19}$	$0.76^{+0.24}_{-0.19}$
79.43 - 100.00	89.13	43^{+6}_{-6}	$660.95^{+92.23}_{-92.23}$	1.00	$0.91^{+0.40}_{-0.24}$	$600.86^{+274.99}_{-180.84}$	$0.88^{+0.46}_{-0.28}$	$0.93^{+0.45}_{-0.26}$	$0.91^{+0.35}_{-0.22}$

Table 5.6: Calculated source counts and associated errors for LOFAR observations of the XMM-LSS field. This includes information on the flux densities of the bin, raw counts (and associated error) in the given flux density bin, source counts, and correction factors (Method 1) as described in Sections 5.7.1 and 5.7.2. Finally the corrected source counts using these corrections are presented, as well as estimates for the completeness corrections (denoted comp. in table) using the models in Section 5.7.3.

S (mJy)	S_c (mJy)	N	$\frac{dN}{dS}S^{2.5}$ (Jy$^{1.5}$)	FDR	Completeness Correction	Corrected $\frac{dN}{dS}S^{2.5}$ (Jy$^{1.5}$)	Comp. SFG	Comp. Small	Comp. Large
2.00 - 2.51	2.24	169^{+13}_{-13}	$10.34^{+0.80}_{-0.80}$	0.98 ± 0.01	$2.61^{+0.28}_{-0.26}$	$26.49^{+3.50}_{-3.35}$	$2.49^{+0.31}_{-0.23}$	$2.67^{+0.25}_{-0.23}$	$2.51^{+0.25}_{-0.24}$
2.51 - 3.16	2.82	236^{+15}_{-15}	$20.40^{+1.30}_{-1.30}$	0.98 ± 0.01	$1.84^{+0.14}_{-0.15}$	$36.81^{+3.71}_{-3.74}$	$1.77^{+0.16}_{-0.12}$	$1.92^{+0.17}_{-0.16}$	$1.78^{+0.19}_{-0.13}$
3.16 - 3.98	3.55	287^{+16}_{-16}	$35.04^{+1.95}_{-1.95}$	0.98 ± 0.01	$1.52^{+0.10}_{-0.10}$	$52.32^{+4.43}_{-4.49}$	$1.51^{+0.09}_{-0.10}$	$1.56^{+0.14}_{-0.11}$	$1.49^{+0.15}_{-0.11}$
3.98 - 5.01	4.47	258^{+16}_{-16}	$44.50^{+2.76}_{-2.76}$	0.98 ± 0.01	$1.38^{+0.09}_{-0.08}$	$60.33^{+5.45}_{-5.27}$	$1.36^{+0.09}_{-0.11}$	$1.42^{+0.11}_{-0.10}$	$1.33^{+0.07}_{-0.09}$
5.01 - 6.31	5.62	252^{+15}_{-15}	$61.39^{+3.65}_{-3.65}$	0.98 ± 0.01	$1.28^{+0.10}_{-0.08}$	$76.92^{+7.60}_{-6.75}$	$1.27^{+0.10}_{-0.07}$	$1.32^{+0.11}_{-0.07}$	$1.23^{+0.07}_{-0.07}$
6.31 - 7.94	7.08	221^{+14}_{-14}	$76.05^{+4.82}_{-4.82}$	0.98 ± 0.01	$1.21^{+0.08}_{-0.06}$	$90.24^{+8.09}_{-7.41}$	$1.20^{+0.09}_{-0.06}$	$1.24^{+0.09}_{-0.06}$	$1.15^{+0.09}_{-0.08}$
7.94 - 10.00	8.91	154^{+12}_{-12}	$74.85^{+5.83}_{-5.83}$	0.98 ± 0.01	$1.14^{+0.08}_{-0.07}$	$83.92^{+8.94}_{-8.34}$	$1.15^{+0.09}_{-0.07}$	$1.18^{+0.09}_{-0.07}$	$1.09^{+0.06}_{-0.07}$
10.00 - 12.59	11.22	146^{+12}_{-12}	$100.24^{+8.24}_{-8.24}$	0.99 ± 0.01	$1.14^{+0.06}_{-0.08}$	$113.14^{+11.35}_{-12.11}$	$1.12^{+0.09}_{-0.07}$	$1.18^{+0.08}_{-0.08}$	$1.08^{+0.06}_{-0.04}$
12.59 - 15.85	14.13	130^{+11}_{-11}	$126.08^{+10.67}_{-10.67}$	1.00	$1.10^{+0.10}_{-0.06}$	$138.09^{+17.05}_{-14.15}$	$1.11^{+0.09}_{-0.08}$	$1.14^{+0.09}_{-0.09}$	$1.08^{+0.11}_{-0.08}$
15.85 - 19.95	17.78	121^{+11}_{-11}	$165.76^{+15.07}_{-15.07}$	1.00	$1.09^{+0.08}_{-0.09}$	$180.18^{+21.04}_{-21.82}$	$1.09^{+0.11}_{-0.06}$	$1.11^{+0.09}_{-0.07}$	$1.05^{+0.08}_{-0.05}$
19.95 - 25.12	22.39	94^{+9}_{-9}	$181.90^{+17.42}_{-17.42}$	0.99 ± 0.01	$1.11^{+0.08}_{-0.11}$	$199.62^{+24.37}_{-27.49}$	$1.09^{+0.12}_{-0.06}$	$1.11^{+0.09}_{-0.07}$	$1.10^{+0.07}_{-0.06}$
25.12 - 31.62	28.18	79^{+8}_{-8}	$215.94^{+21.87}_{-21.87}$	1.00	$1.11^{+0.10}_{-0.08}$	$239.93^{+33.15}_{-29.95}$	$1.11^{+0.11}_{-0.07}$	$1.12^{+0.10}_{-0.08}$	$1.11^{+0.10}_{-0.11}$
31.62 - 39.81	35.48	71^{+8}_{-8}	$274.13^{+30.89}_{-30.89}$	1.00	$1.10^{+0.15}_{-0.07}$	$301.54^{+53.91}_{-38.87}$	$1.10^{+0.10}_{-0.06}$	$1.11^{+0.12}_{-0.11}$	$1.08^{+0.15}_{-0.08}$
39.81 - 50.12	44.67	76^{+8}_{-8}	$414.49^{+43.63}_{-43.63}$	0.99 ± 0.01	$1.08^{+0.15}_{-0.08}$	$440.65^{+78.40}_{-56.35}$	$1.08^{+0.13}_{-0.08}$	$1.08^{+0.11}_{-0.08}$	$1.06^{+0.13}_{-0.06}$
50.12 - 63.10	56.23	76^{+8}_{-8}	$585.48^{+61.63}_{-61.63}$	1.00	$1.00^{+0.17}_{-0.09}$	$585.48^{+115.41}_{-81.99}$	$1.00^{+0.14}_{-0.09}$	$1.06^{+0.14}_{-0.13}$	$1.00^{+0.13}_{-0.11}$
63.10 - 79.43	70.79	37^{+6}_{-6}	$402.63^{+65.29}_{-65.29}$	1.00	$1.08^{+0.15}_{-0.08}$	$436.18^{+94.14}_{-78.29}$	$1.08^{+0.12}_{-0.08}$	$1.10^{+0.17}_{-0.10}$	$1.07^{+0.13}_{-0.14}$
79.43 - 100.00	89.13	43^{+6}_{-6}	$660.95^{+92.23}_{-92.23}$	1.00	$1.00^{+0.13}_{-0.14}$	$660.95^{+124.41}_{-132.82}$	$1.00^{+0.18}_{-0.11}$	$1.00^{+0.20}_{-0.10}$	$1.00^{+0.13}_{-0.13}$

Table 5.7: Calculated source counts and associated errors for LOFAR observations of the XMM-LSS field. This includes information on the flux densities of the bin, raw counts (and associated error) in the given flux density bin, source counts, and correction factors (Method 2) as described in Sections 5.7.1 and 5.7.2. Finally the corrected source counts using these corrections are presented, as well as estimates for the completeness corrections (denoted comp. in table) using the models in Section 5.7.3.

Chapter 6

Conclusions and Future Work

The work presented in this Thesis has provided details of multiple studies where modern radio surveys are used and can be used in the future to advance the study of galaxy evolution. This has included investigating the evolution in the relationship between galaxy populations and the dark matter environments they reside in, as well as making preparations for future studies of galaxy evolution with radio surveys.

In Chapter 2, I investigated the dark matter environments that different radio populations inhabit. By measuring their angular clustering and determining the bias of these galaxies, my work suggests that SFGs reside in less massive haloes than AGN, reflecting ideas that radio AGN reside in the most massive elliptical galaxies (McLure et al., 1999; Dunlop & McLure, 2003) which are likely to need massive dark matter haloes to support them. By further studying the clustering of radiatively efficient and inefficient AGN (analogues of H/LERGs), I found that the inefficient AGN reside, at low redshifts, in much more massive haloes than their efficient counterparts. This reflects previous investigations into the environment of H/LERGs (Hardcastle, 2004; Gendre et al., 2013; Heckman & Best, 2014) and supports the theory that these larger mass haloes contain hotter gas, that is more challenging to accrete efficiently onto a black hole. Indeed, Ineson et al. (2015) found that inefficient AGN reside in

hotter environments and an increase in temperature with halo mass is also observed for clusters (Main et al., 2017). This hotter gas will have more kinetic energy and therefore will be more challenging to accrete onto the black hole, making the accretion inefficient. The inefficient AGN, however, reside in more similar haloes to that of the efficient AGN at higher redshifts ($z \sim 1.3$, see Chapter 2) suggesting that the relationship between these AGN and their environments evolve differently for the two populations.

Building on the work of studying the clustering of radio populations, my research into the Conditional Luminosity Function (CLF) in Chapter 3 has shown that it is possible to relate the observed galaxy clustering to the luminosity distribution of sources by describing how the number of radio sources is dependent on both luminosity and halo mass. This work has shown that despite limitations in the data, the CLF is able to successfully reproduce both $\Phi(L_\nu)$ and $b(L_\nu)$. By considering the modelled form of the CLF for SFGs, HLAGN and MLAGN, I confirm the results of Chapter 2. This suggests that the MLAGN population, on average, reside in the most massive haloes. My work also suggests that whilst the typical masses of haloes that all three populations reside in increase with luminosity, lower halo masses are expected for the same luminosity at high redshift compared to low redshift. That is to say that lower mass haloes were needed in the past to host an AGN or SFG of a given luminosity. Finally, by making use of the relations of Willott et al. (1999) and Bîrzan et al. (2008), I have used the CLF to quantify how the jet power of these AGN may be related to their dark matter environment. This work however is limited by the current available data, which needs to be of larger area to obtain the necessary statistics of luminous AGN and SFGs. More sensitive observations over those fields with a wealth of ancillary data is also essential. This is in order to better understand how the relationship between galaxies, their environments and their observed properties evolve over time.

In Chapter 4, I have shown how ProFound (Robotham et al., 2018), an optical and near-IR motivated source extraction software, can successfully recover radio emission in continuum imaging for a variety of source morphologies. This provides a tool for extracting flux densities for sources from radio images regardless of shape and not limited to only fitting Gaussian components. Furthermore, as ProFound has a multi-wavelength aspect to it, its success at radio frequencies provides the framework to be able to extract flux information across the electromagnetic spectrum using a consistent method. I showed that ProFound is competitive with PyBDSF and AEGEAN, which are two commonly used source extraction software packages used on radio continuum images, but that fit smooth Gaussian emission to a source. ProFound however is more successful than PyBDSF and AEGEAN when complex, extended morphologies are considered. Finally, in Chapter 5, I showed that it is possible to make deep, high-resolution low frequency observations of the XMM-LSS field, despite its low elevation to the LOFAR array. This field is well studied across the electromagnetic spectrum (Tasse et al., 2007; Oliver et al., 2012; Jarvis et al., 2013) and allows for deep, multi-frequency radio observations which can be combined with other deep radio observations at other frequencies (Heywood et al., in prep; Jarvis et al., 2017) in order to study how these populations evolve.

Combined, this work has allowed me to develop a much greater understanding of the use of extragalactic radio surveys in studying how different galaxy populations evolve in relation to their environments. However, as discussed in Chapter 3, there are limitations to the work described here as, despite using state of the art radio observations, there is a need for larger areas and deeper observations within fields with excellent ancillary data to better constrain the evolution of different galaxy populations. Fortunately, with the advent of future deep extragalactic radio surveys (e.g. Norris et al., 2011; Jarvis et al., 2017; Hales & Chiles Con Pol Collaboration, 2014; Shimwell et al., 2019) data will soon be available to develop upon the work

presented here as well as provide opportunities for further work, as described below.

6.1 Future Work

6.1.1 Galaxy Clustering

The work presented in Chapters 2 and 3 relied on using the angular clustering of galaxies and Limber inversion in order to infer the bias of a population of galaxies. However, if spectroscopic redshifts were available for all sources then direct calculation of the redshift space correlation function (e.g. Hawkins et al., 2003; Ross et al., 2007) is possible. This would allow a more direct calculation of the bias. With the upcoming WEAVE-LOFAR survey (Smith et al., 2016), redshifts will be obtained for those extragalactic galaxies observed with LOFAR over three tiers. Tier 1 will obtain redshifts for those sources within the LoTSS survey (Shimwell et al., 2019) that have flux densities of $S_{150\ \mathrm{MHz}} > 10$ mJy. Tiers 2 and 3, on the other hand, will provide highly complete redshift information for fainter sources in those fields that have a wide range of ancillary data, such as the Boötes, COSMOS and Elais-N1 fields. In the deepest fields, this may allow the detection of both the '1-' and '2-halo' clustering terms (see e.g. Cooray & Sheth, 2002; Yang et al., 2003) for the first time with radio observations. These clustering terms are also likely to be detected using surveys such as MIGHTEE (Jarvis et al., 2017) and CHILES Con-Pol (Hales & Chiles Con Pol Collaboration, 2014), due to their sensitive, high resolution observations.

With observations of both the small and large scale clustering, full HOD (Halo Occupation Distribution) modelling (e.g. Cooray & Sheth, 2002; Zheng et al., 2005, 2007; Hatfield et al., 2016) of sources will become more appropriate in order to understand the haloes that galaxy populations are residing in. HOD modelling is a way to better understand the dark matter haloes that a population of galaxies inhabits, by quantifying the properties of these haloes, such as the masses needed to both host

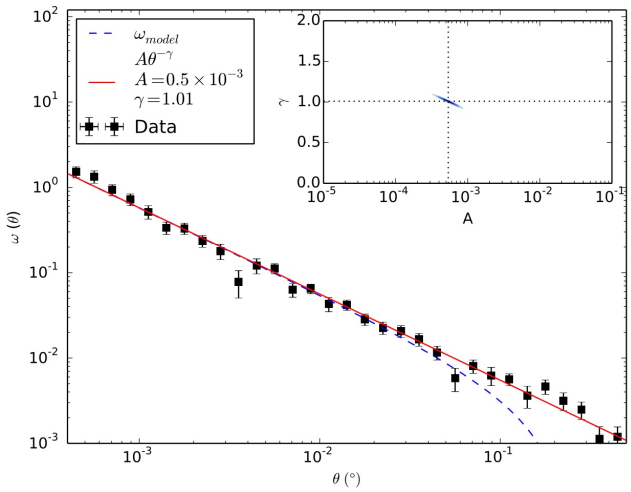


Figure 6.1: Predicted clustering constraints with MIGHTEE (Jarvis et al., 2017) created by me for Jarvis et al. (2017) using the clustering of all sources in Chapter 2 but accounting for differences in sensitivity and area.

central and satellite galaxies as well as the power law distribution of halo masses. This therefore does not limit understanding to just the typical masses of the haloes that a galaxy population resides in.

By investigating the large and small scale clustering over those fields which have ancillary data also means that these properties can be investigated in relation to Star Formation Rates (SFR), stellar masses (M_*) (e.g. Hatfield et al., 2016; Tinker et al., 2017) and galaxy type (as in Chapters 2 and 3). It also allows the possibility to investigate the clustering of satellite SFGs around AGN that exist within the same dark matter halo. This would allow a direct investigation into how SFGs may be influenced by the AGN that co-exist within the same dark matter halo and whether the type of AGN (e.g. HERG/LERG) may affect how SFGs cluster around them. This would make use of cross correlations between the two populations, using the cross-correlation function described in Chapter 3. This will also allow the models of Hatfield & Jarvis (2017) to be investigated and this research would be novel at radio frequencies.

6.1.2 The Relationship between Galaxies and Environments

Also, as described in Chapter 3, initial determination of the CLF was limited by the numbers of sources available. As such, surveys over larger areas and with similar amounts of ancillary data are essential. Surveys such as MIGHTEE (Jarvis et al., 2017) and the deeper LOFAR tier surveys (Röttgering, 2010; van Haarlem et al., 2013) will be essential in providing the necessary datasets to constrain the CLF to high redshifts for SFGs and AGN, whilst the wide area of EMU (Norris et al., 2011) and LoTSS (Shimwell et al., 2019) will be essential to constraining the CLF for bright AGN. The predicted constraints on clustering for MIGHTEE (Jarvis et al., 2017) can be seen in Figure 6.1. This was calculated by me for Jarvis et al. (2017) using the clustering constraints on all sources in the VLA COSMOS Survey (Smolčić et al., 2017a,b) (fitting both for the amplitude and slope of $\omega(\theta)$) and extrapolating to MIGHTEE accounting for differences in survey area and depths. This made use of the simulated skies from Wilman et al. (2008) to determine how the differences in sensitivity will affect the number of detected sources. Therefore using these studies should revolutionise measurements of $b(L_\nu)$ and hence improve the constraints of the CLF. This will allow a greater understanding of how the environment of galaxies are related in the evolution of galaxies, and the variation with time. Studying the CLF for different populations of radio sources will also help improve predictions of realistic extragalactic radio skies that will be observed with the SKA.

6.1.3 Studying Radio Sources in a Multi-wavelength Context

Finally, as the work of Chapter 4 has shown, it is possible to accurately measure the flux densities of radio sources from continuum images using ProFound (Robotham et al., 2018). Whilst spectral energy distributions (SEDs) of sources can be measured by combining source catalogues at different wavelengths, the natural follow on from the work in Chapter 4 is that SEDs can be measured using a consistent source ex-

traction method. For large populations of galaxies, radio emission is not included in SED modelling (but see da Cunha et al., 2015; Miettinen et al., 2017). SEDs of radio continuum emission are instead typically fit only over radio frequencies (e.g. Clemens et al., 2008; Galvin et al., 2018). Whilst radio continuum measurements cannot be used to measure redshifts and stellar masses, for SFGs the unbiased SFR can be estimated with radio continuum observations (see e.g. Bell, 2003; Davies et al., 2017; Gürkan et al., 2018). This therefore may be useful in disentangling the star formation that may be obscured at other wavelengths. For MIGHTEE (Jarvis et al., 2017), the work of Chapter 5 will be important in contributing low frequency radio continuum to the SEDs of sources observed over the XMM-LSS field.

Furthermore, the multi-wavelength aspect of ProFound allows the use of segments at one wavelength to force photometry at radio wavelengths. This may be important for sub-5σ studies such of source counts (e.g. Zwart et al., 2015b) and luminosity functions (e.g. Malefahlo et al., in prep). In these studies, pixel fluxes at the position of multi wavelength sources can be used to investigate sources below the nominal detection threshold. ProFound would reduce errors in the fluxes of these sub-5σ sources, which may arise from offset radio emission (if they are extended). However, with both of these multi-wavelength studies, whilst using ProFound can extract fluxes in a consistent method, it may be limited by the fact that radio morphologies may be different to those observed at other wavelengths. ProFound, however, does allow an aperture found at one wavelength to be expanded in the other image by either retaining the segment shape or growing the segment in any direction. The latter may especially be important for AGN which can have very different morphologies at radio frequencies compared to other wavelengths.

All these studies follow on from the work presented in this Thesis and will be important for studying galaxy evolution of SFGs and AGN from radio surveys, within a multi-wavelength context.

References

G. E. Addison, G. Hinshaw, & M. Halpern. *Monthly Notices of the Royal Astronomical Society*, 436:1674–1683, Dec 2013.

H. Aihara, R. Armstrong, S. Bickerton, et al. *Publications of the Astronomical Society of Japan*, 70:S8, January 2018.

V. Allevato, F. Civano, A. Finoguenov, et al. *Astrophysical Journal*, 832:70, November 2016.

AMI Consortium et al. *Monthly Notices of the Royal Astronomical Society*, 415: 2708–2722, Aug 2011.

R. Antonucci. *Annual Review of Astronomy and Astrophysics*, 31:473–521, Jan 1993.

P. N. Appleton, D. T. Fadda, F. R. Marleau, et al. *The Astrophysical Journal Supplement Series*, 154:147–150, Sep 2004.

Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, et al. *Astronomy and Astrophysics*, 558:A33, October 2013.

J. E. Baldwin, R. C. Boysen, S. E. G. Hales, et al. *Monthly Notices of the Royal Astronomical Society*, 217:717–730, December 1985.

D. R. Ballantyne. *Monthly Notices of the Royal Astronomical Society*, 464:613–625, January 2017.

J. K. Banfield, O. I. Wong, K. W. Willett, et al. *Monthly Notices of the Royal Astronomical Society*, 453:2326–2340, November 2015.

R. H. Becker, R. L. White, & D. J. Helfand. *Astrophysical Journal*, 450:559, September 1995.

R. H. Becker, M. D. Gregg, I. M. Hook, et al. *Astrophysical Journal*, 479:L93–L96, April 1997.

E. F. Bell. *Astrophysical Journal*, 586:794–813, April 2003.

J. Benjamin, C. Heymans, E. Semboloni, et al. *Monthly Notices of the Royal Astronomical Society*, 381:702–712, October 2007.

A. S. Bennett. *Memoirs of the Royal Astronomical Society*, 68:163, January 1962.

A. J. Benson, R. G. Bower, C. S. Frenk, et al. *Astrophysical Journal*, 599:38–49, December 2003.

P. N. Best & T. M. Heckman. *Monthly Notices of the Royal Astronomical Society*, 421:1569–1582, April 2012.

P. N. Best, L. M. Ker, C. Simpson, et al. *Monthly Notices of the Royal Astronomical Society*, 445:955–969, Nov 2014.

L. Bîrzan, B. R. McNamara, P. E. J. Nulsen, et al. *Astrophysical Journal*, 686: 859–880, Oct 2008.

C. Blake & J. Wall. *Monthly Notices of the Royal Astronomical Society*, 337:993–1003, Dec 2002.

C. Blake, T. Mauch, & E. M. Sadler. *Monthly Notices of the Royal Astronomical Society*, 347:787–794, Jan 2004.

C. Blake, A. Collister, & O. Lahav. *Monthly Notices of the Royal Astronomical Society*, 385:1257–1269, Apr 2008.

M. R. Blanton, D. W. Hogg, N. A. Bahcall, et al. *Astrophysical Journal*, 592:819–838, Aug 2003.

D. C. J. Bock, M. I. Large, & E. M. Sadler. *Astronomical Journal*, 117:1578–1593, March 1999.

A. Bonaldi, M. Bonato, V. Galluzzi, et al. *Monthly Notices of the Royal Astronomical Society*, 482:2–19, Jan 2019.

R. J. Bouwens, G. D. Illingworth, P. A. Oesch, et al. *Astrophysical Journal*, 803:34, Apr 2015.

R. G. Bower, A. J. Benson, & Robert A. Crain. *Monthly Notices of the Royal Astronomical Society*, 422:2816–2840, Jun 2012.

R. A. A. Bowler, J. S. Dunlop, R. J. McLure, et al. *Monthly Notices of the Royal Astronomical Society*, 452:1817–1840, September 2015.

W. N. Brandt & D. M. Alexander. *Astronomy and Astrophysics Reviews*, 23:1, January 2015.

M. Brienza, L. Godfrey, R. Morganti, et al. *Astronomy and Astrophysics*, 606:A98, Oct 2017.

D. S. Briggs, F. R. Schwab, & R. A. Sramek. In G. B. Taylor, C. L. Carilli, & R. A. Perley, editors, *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*, page 127, Jan 1999.

M. J. I. Brown, R. L. Webster, & B. J. Boyle. *Astronomical Journal*, 121:2381–2391, May 2001.

B. F. Burke & F. Graham-Smith. *An Introduction to Radio Astronomy*. Cambridge University Press, 2014.

A. Butler, M. Huynh, J. Delhaize, et al. *Astronomy and Astrophysics*, 620:A3, Nov 2018.

J. R. Callingham, R. D. Ekers, B. M. Gaensler, et al. *Astrophysical Journal*, 836:174, Feb 2017.

B. W. Carroll & D. A. Ostlie. *An introduction to modern astrophysics and cosmology*. Pearson, Addison-Wesley, 2006.

S. M. Carroll, W. H. Press, & E. L. Turner. *Annual Review of Astronomy and Astrophysics*, 30:499–542, 1992.

K. W. Cavagnolo, B. R. McNamara, P. E. J. Nulsen, et al. *Astrophysical Journal*, 720:1066–1072, Sep 2010.

L. Ceraj, V. Smolčić, I. Delvecchio, et al. *Astronomy and Astrophysics*, 620:A192, Dec 2018.

S. Chatterjee, My L. Nguyen, A. D. Myers, et al. *Astrophysical Journal*, 779:147, Dec 2013.

C.-T. Chen, W. Brandt, B. Luo, et al. volume 231 of *American Astronomical Society Meeting Abstracts*, page 231.03, January 2018.

S. Chen & D. J. Schwarz. *Astronomy and Astrophysics*, 591:A135, June 2016.

J. H. Y. Ching, S. M. Croom, E. M. Sadler, et al. *Monthly Notices of the Royal Astronomical Society*, 469:4584–4599, Aug 2017.

M. S. Clemens, O. Vega, A. Bressan, et al. *Astronomy and Astrophysics*, 477:95–104, Jan 2008.

R. K. Cochrane, P. N. Best, D. Sobral, et al. *Monthly Notices of the Royal Astronomical Society*, 475:3730–3745, Apr 2018.

A. S. Cohen, H. J. A. Röttgering, N. E. Kassim, et al. *Astrophysical Journal*, 591: 640–661, July 2003.

A. S. Cohen, W. M. Lane, W. D. Cotton, et al. *Astronomical Journal*, 134:1245–1262, September 2007.

A. L. Coil. *The Large-Scale Structure of the Universe*, page 387. 2013. doi: 10.1007/978-94-007-5609-0_8.

M. Colless, G. Dalton, S. Maddox, et al. *Monthly Notices of the Royal Astronomical Society*, 328:1039–1063, Dec 2001.

J. J. Condon. *Annual Review of Astronomy and Astrophysics*, 30:575–611, 1992.

J. J. Condon, W. D. Cotton, E. W. Greisen, et al. *Astronomical Journal*, 115:1693–1716, May 1998.

J. J. Condon, W. D. Cotton, E. B. Fomalont, et al. *Astrophysical Journal*, 758:23, October 2012.

A. Cooray & R. Sheth. *Physics Reports*, 372:1–129, December 2002.

C. M. Cress, D. J. Helfand, R. H. Becker, et al. *Astrophysical Journal*, 473:7, December 1996.

J. H. Croston, J. Ineson, & M. J. Hardcastle. *Monthly Notices of the Royal Astronomical Society*, 476:1614–1623, May 2018.

J-C. J. Cuillandre, K. Withington, P. Hudelot, & others. In *Observatory Operations: Strategies, Processes, and Systems IV*, volume 8448 of *Society of Photo-Optical*

Instrumentation Engineers (SPIE) Conference Series, page 84480M, Sep 2012. doi: 10.1117/12.925584.

E. da Cunha, F. Walter, I. R. Smail, et al. *Astrophysical Journal*, 806:110, Jun 2015.

R. A. Daly, T. B. Sprinkle, C. P. O'Dea, et al. *Monthly Notices of the Royal Astronomical Society*, 423:2498–2502, Jul 2012.

L. J. M. Davies, M. T. Huynh, A. M. Hopkins, et al. *Monthly Notices of the Royal Astronomical Society*, 466:2312–2324, April 2017.

L. J. M. Davies, A. S. G. Robotham, S. P. Driver, et al. *Monthly Notices of the Royal Astronomical Society*, page 1478, June 2018.

M. Davis & P. J. E. Peebles. *Astrophysical Journal*, 267:465–482, April 1983.

M. Davis, A. Meiksin, M. A. Strauss, et al. *Astrophysical Journal, Letters*, 333: L9–L12, October 1988.

F. de Gasperin, T. J. Dijkema, A. Drabent, et al. *Astronomy and Astrophysics*, 622: A5, Feb 2019.

G. de Zotti, M. Massardi, M. Negrello, et al. *Astronomy and Astrophysics Review*, 18:1–65, Feb 2010.

I. Delvecchio, V. Smolčić, G. Zamorani, et al. *Astronomy and Astrophysics*, 602:A3, June 2017.

DES Collaboration et al. *Physical Review D*, 98:043526, Aug 2018a.

DES Collaboration et al. *arXiv e-prints*, art. arXiv:1811.02374, Nov 2018b.

V. Desjacques, D. Jeong, & F. Schmidt. *Physics Reports*, 733:1–193, Feb 2018.

R. H. Dicke, P. J. E. Peebles, P. G. Roll, et al. *Astrophysical Journal*, 142:414–419, July 1965.

B. Diemer. *The Astrophysical Journal Supplement Series*, 239:35, December 2018.

Scott Dodelson. *Modern cosmology*. 2003.

T. Dolley, M. J. I. Brown, B. J. Weiner, et al. *Astrophysical Journal*, 797:125, December 2014.

J. L. Donley, A. M. Koekemoer, M. Brusa, et al. *Astrophysical Journal*, 748:142, April 2012.

Y. Dubois, C. Pichon, C. Welker, et al. *Monthly Notices of the Royal Astronomical Society*, 444:1453–1468, Oct 2014.

J. S. Dunlop & R. J. McLure. In R. Bender & A. Renzini, editors, *The Mass of Galaxies at Low and High Redshift*, page 268, Jan 2003. doi: 10.1007/10899892_67.

D. O. Edge, J. R. Shakeshaft, W. B. McAdam, et al. *Memoirs of the RAS*, 68:37–60, 1959.

A. C. Fabian. *Annual Review of Astronomy and Astrophysics*, 50:455–489, Sep 2012.

B. L. Fanaroff & J. M. Riley. *Monthly Notices of the Royal Astronomical Society*, 167:31P–36P, May 1974.

L. Feretti, I. M. Gioia, G. Giovannini, et al. *Astronomy and Astrophysics*, 139:50–54, October 1984.

C. A. C. Fernandes, M. J. Jarvis, S. Rawlings, et al. *Monthly Notices of the Royal Astronomical Society*, 411:1909–1916, Mar 2011.

L. D. Ferramacho, M. G. Santos, M. J. Jarvis, et al. *Monthly Notices of the Royal Astronomical Society*, 442:2511–2518, August 2014.

L. Ferrarese & D. Merritt. *Astrophysical Journal*, 539:L9–L12, Aug 2000.

D. Foreman-Mackey, D. W. Hogg, D. Lang, et al. *Publications of the Astronomical Society of the Pacific*, 125:306, March 2013.

T. M. O. Franzen, C. A. Jackson, A. R. Offringa, et al. *Monthly Notices of the Royal Astronomical Society*, 459:3314–3325, July 2016.

W. L. Freedman, B. F. Madore, B. K. Gibson, et al. *Astrophysical Journal*, 553: 47–72, May 2001.

T. J. Galvin, N. Seymour, J. Marvil, et al. *Monthly Notices of the Royal Astronomical Society*, 474:779–799, February 2018.

T. Garn, D. A. Green, J. M. Riley, et al. *Monthly Notices of the Royal Astronomical Society*, 397:1101–1112, August 2009.

S. T. Garrington, B. Anderson, C. Baines, et al. In Jr. Oschmann, J. M., editor, *Ground-based Telescopes*, volume 5489 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 332–343, October 2004. doi: 10.1117/12.553235.

M. A. Gendre, P. N. Best, J. V. Wall, et al. *Monthly Notices of the Royal Astronomical Society*, 430:3086–3101, April 2013.

R. Gilli, E. Daddi, G. Zamorani, et al. *Astronomy and Astrophysics*, 430:811–825, February 2005.

R. Gilli, E. Daddi, R. Chary, et al. *Astronomy and Astrophysics*, 475:83–99, November 2007.

R. Gilli, G. Zamorani, T. Miyaji, et al. *Astronomy and Astrophysics*, 494:33–48, January 2009.

J. F. R. Gower, P. F. Scott, & D. Wills. *Memoirs of the RAS*, 71:49, 1967.

E. J. Groth & P. J. E. Peebles. *Astrophysical Journal*, 217:385–405, Oct 1977.

Q. Guo, S. Cole, C. G. Lacey, et al. *Monthly Notices of the Royal Astronomical Society*, 412:2277–2285, April 2011.

G. Gürkan, M. J. Hardcastle, M. J. Jarvis, D. J. B. Smith, N. Bourne, L. Dunne, S. Maddox, R. J. Ivison, & J. Fritz. *Monthly Notices of the Royal Astronomical Society*, 452(4):3776–3794, Oct 2015.

G. Gürkan, M. J. Hardcastle, D. J. B. Smith, et al. *Monthly Notices of the Royal Astronomical Society*, 475:3010–3028, Apr 2018.

C. L. Hale, M. J. Jarvis, I. Delvecchio, et al. *Monthly Notices of the Royal Astronomical Society*, 474:4133–4150, March 2018.

C. L. Hale, A. S. G. Robotham, L. J. M. Davies, et al. *accepted to Monthly Notices of the Royal Astronomical Society*, art. arXiv:1902.01440, Feb 2019a. URL `https://academic.oup.com/mnras/advance-article/doi/10.1093/mnras/stz1462/5511783`.

C. L. Hale, W. Williams, M. J. Jarvis, et al. *Astronomy and Astrophysics*, 622:A4, Feb 2019b.

C. A. Hales & Chiles Con Pol Collaboration. In *Exascale Radio Astronomy*, volume 2, April 2014.

C. A. Hales, T. Murphy, J. R. Curran, et al. *Monthly Notices of the Royal Astronomical Society*, 425:979–996, Sep 2012.

S. E. G. Hales, J. E. Baldwin, & P. J. Warner. *Monthly Notices of the Royal Astronomical Society*, 263:25–30, July 1993.

J. P. Hamaker, J. D. Bregman, & R. J. Sault. *Astronomy and Astrophysics Supplement Series*, 117:137–147, May 1996.

A. J. S. Hamilton. *Astrophysical Journal*, 417:19, November 1993.

A. J. S. Hamilton. *Monthly Notices of the Royal Astronomical Society*, 322:419–425, April 2001.

P. J. Hancock, T. Murphy, B. M. Gaensler, et al. *Monthly Notices of the Royal Astronomical Society*, 422:1812–1824, May 2012.

P. J. Hancock, C. M. Trott, & N. Hurley-Walker. *Publications of the Astronomical Society of Australia*, 35:e011, March 2018.

M. J. Hardcastle. *Astronomy and Astrophysics*, 414:927–929, February 2004.

M. J. Hardcastle. *Monthly Notices of the Royal Astronomical Society*, 475:2768–2786, Apr 2018.

M. J. Hardcastle & M. G. H. Krause. *Monthly Notices of the Royal Astronomical Society*, 443:1482–1499, Sep 2014.

M. J. Hardcastle, D. A. Evans, & J. H. Croston. *Monthly Notices of the Royal Astronomical Society*, 376:1849–1856, April 2007.

M. J. Hardcastle, G. Gürkan, R. J. van Weeren, et al. *Monthly Notices of the Royal Astronomical Society*, 462:1910–1936, October 2016.

W. G. Hartley, O. Almaini, A. Mortlock, et al. *Monthly Notices of the Royal Astronomical Society*, 431:3045–3059, June 2013.

G. Hasinger, N. Cappelluti, H. Brunner, et al. *The Astrophysical Journal Supplement Series*, 172:29–37, Sep 2007.

P. W. Hatfield & M. J. Jarvis. *ArXiv e-prints*, June 2017. URL `https://arxiv.org/abs/1606.08989`.

P. W. Hatfield, S. N. Lindsay, M. J. Jarvis, et al. *Monthly Notices of the Royal Astronomical Society*, 459:2618–2631, July 2016.

E. Hawkins, S. Maddox, S. Cole, et al. *Monthly Notices of the Royal Astronomical Society*, 346:78–96, Nov 2003.

G. Heald, J. McKean, & R. Pizzo. *Low Frequency Radio Astronomy and the LOFAR Observatory*. Springer International Publishing, 2018. doi: 10.1007/978-3-319-23434-2.

G. H. Heald, R. F. Pizzo, E. Orrú, et al. *Astronomy and Astrophysics*, 582:A123, Oct 2015.

T. M. Heckman & P. N. Best. *Annual Review of Astronomy and Astrophysics*, 52:589–660, August 2014.

D. J. Helfand, R. L. White, & R. H. Becker. *Astrophysical Journal*, 801:26, March 2015.

I. Heywood, M. J. Jarvis, & J. J. Condon. *Monthly Notices of the Royal Astronomical Society*, 432:2625–2631, July 2013.

I. Heywood, M. J. Jarvis, A. J. Baker, et al. *Monthly Notices of the Royal Astronomical Society*, 460:4433–4452, August 2016.

I. Heywood et al. in prep.

J. A. Hodge, R. H. Becker, R. L. White, et al. *Astronomical Journal*, 142:3, Jul 2011.

J. A. Högbom. *Astronomy and Astrophysics, Supplement*, 15:417, June 1974.

D. W. Hogg. *ArXiv Astrophysics e-prints*, May 1999. URL `https://arxiv.org/abs/astro-ph/9905116`.

A. M. Hopkins, M. T. Whiting, N. Seymour, et al. *Publications of the Astronomical Society of Australia*, 32:e037, October 2015.

E. Hubble. *Proceedings of the National Academy of Science*, 15:168–173, March 1929.

P. Hudelot, J. Ch. Cuillandre, K. Withington, et al. *VizieR Online Data Catalog*, art. II/317, Sep 2012.

J. D. Hunter. *Computing in Science and Engineering*, 9:90–95, May 2007.

N. Hurley-Walker. *ArXiv e-prints*, March 2017.

N. Hurley-Walker, J. R. Callingham, P. J. Hancock, et al. *Monthly Notices of the Royal Astronomical Society*, 464:1146–1167, January 2017.

O. Ilbert, M. Salvato, E. Le Floc'h, et al. *Astrophysical Journal*, 709:644–663, February 2010.

J. Ineson, J. H. Croston, M. J. Hardcastle, et al. *Monthly Notices of the Royal Astronomical Society*, 453:2682–2706, November 2015.

H. T. Intema, P. Jagannathan, K. P. Mooley, et al. *Astronomy and Astrophysics*, 598:A78, February 2017.

M. Jarvis. TreeCorr: Two-point correlation functions. Astrophysics Source Code Library, August 2015.

M. J. Jarvis, D. J. B. Smith, D. G. Bonfield, et al. *Monthly Notices of the Royal Astronomical Society*, 409:92–101, November 2010.

M. J. Jarvis, D. G. Bonfield, V. A. Bruce, et al. *Monthly Notices of the Royal Astronomical Society*, 428:1281–1295, January 2013.

M. J. Jarvis, A. R. Taylor, I. Agudo, et al. *ArXiv e-prints*, September 2017.

A. Jenkins, C. S. Frenk, S. D. M. White, et al. *Monthly Notices of the Royal Astronomical Society*, 321:372–384, Feb 2001.

J. Jonas & MeerKAT Team. In *Proceedings of MeerKAT Science: On the Pathway to the SKA. 25-27 May, 2016 Stellenbosch, South Africa (MeerKAT2016). Online at* `https://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=277</A>,id.1`, page 1, 2016.

J. L. Jonas. *IEEE Proceedings*, 97:1522–1530, August 2009.

B. J. Jones, V. J. Martínez, E. Saar, et al. *Reviews of Modern Physics*, 76:1211–1266, October 2004.

E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001. URL `http://www.scipy.org/`. [Online; accessed <today>].

W. A. Joye & E. Mandel. In H. E. Payne, R. I. Jedrzejewski, & R. N. Hook, editors, *Astronomical Data Analysis Software and Systems XII*, volume 295 of *Astronomical Society of the Pacific Conference Series*, page 489, Jan 2003.

N. Kaiser. *Astrophysical Journal, Letters*, 284:L9–L12, September 1984.

E. Kalfountzou, J. A. Stevens, M. J. Jarvis, et al. *Monthly Notices of the Royal Astronomical Society*, 471:28–58, October 2017.

K. I. Kellermann, R. Sramek, M. Schmidt, et al. *Astronomical Journal*, 98:1195–1207, October 1989.

S. M. Kent. *Astronomical Journal*, 91:1301–1327, June 1986.

S. M. Kent. *Astronomical Journal*, 93:816–832, April 1987.

S. M. Kent. *Astronomical Journal*, 96:514–527, August 1988.

L. Koutoulidis, M. Plionis, I. Georgantopoulos, et al. *Monthly Notices of the Royal Astronomical Society*, 428:1382–1394, January 2013.

J. Kristian, A. Sandage, & J. A. Westphal. *Astrophysical Journal*, 221:383–394, Apr 1978.

C. Laigle, H. J. McCracken, O. Ilbert, et al. *Astrophysical Journal, Supplement*, 224: 24, June 2016.

R. A. Laing & A. H. Bridle. *Monthly Notices of the Royal Astronomical Society*, 336: 328–352, October 2002.

R. A. Laing, J. M. Riley, & M. S. Longair. *Monthly Notices of the Royal Astronomical Society*, 204:151–187, July 1983.

S. D. Landy & A. S. Szalay. *Astrophysical Journal*, 412:64–71, July 1993.

J. P. Leahy. *Vistas in Astronomy*, 40:173–177, January 1996.

J. P. Leahy & R. A. Perley. *Astronomical Journal*, 102:537, August 1991.

D. Nelson Limber. *Astrophysical Journal*, 117:134, Jan 1953.

S. N. Lindsay, M. J. Jarvis, & K. McAlpine. *Monthly Notices of the Royal Astronomical Society*, 440:2322–2332, May 2014a.

S. N. Lindsay, M. J. Jarvis, M. G. Santos, et al. *Monthly Notices of the Royal Astronomical Society*, 440:1527–1541, May 2014b.

E. N. Ling, J. D. Barrow, & C. S. Frenk. *Monthly Notices of the Royal Astronomical Society*, 223:21P–27P, December 1986.

M. S. Longair. *High Energy Astrophysics*. Cambridge University Press, February 2011.

J. Loveday, P. Norberg, I. K. Baldry, et al. *Monthly Notices of the Royal Astronomical Society*, 420:1239–1262, February 2012.

Z. Ma, J. Zhu, W. Li, et al. *arXiv e-prints*, art. arXiv:1806.00398, Jun 2018.

M. Magliocchetti, S. J. Maddox, E. Hawkins, et al. *Monthly Notices of the Royal Astronomical Society*, 350:1485–1494, June 2004.

M. Magliocchetti, P. Popesso, D. Rosario, et al. *Monthly Notices of the Royal Astronomical Society*, 433:127–137, July 2013.

M. Magliocchetti, P. Popesso, M. Brusa, et al. *Monthly Notices of the Royal Astronomical Society*, 464:3271–3280, January 2017.

J. Magorrian, S. Tremaine, D. Richstone, et al. *Astronomical Journal*, 115:2285–2305, June 1998.

E. K. Mahony, R. Morganti, I. Prandoni, et al. *Monthly Notices of the Royal Astronomical Society*, 463:2997–3020, December 2016.

R. A. Main, B. R. McNamara, P. E. J. Nulsen, et al. *Monthly Notices of the Royal Astronomical Society*, 464:4360–4382, Feb 2017.

E. Malefahlo et al. in prep.

R. Mandelbaum, C. Li, G. Kauffmann, et al. *Monthly Notices of the Royal Astronomical Society*, 393:377–392, February 2009.

A. Mantz, S. W. Allen, H. Ebeling, et al. *Monthly Notices of the Royal Astronomical Society*, 406(3):1773–1795, Aug 2010.

T. Mauch & E. M. Sadler. *Monthly Notices of the Royal Astronomical Society*, 375: 931–950, March 2007.

J.-C. Mauduit, M. Lacy, D. Farrah, et al. *Publications of the Astronomical Society of the Pacific*, 124:714, July 2012.

K. McAlpine, D. J. B. Smith, M. J. Jarvis, et al. *Monthly Notices of the Royal Astronomical Society*, 423:132–140, June 2012.

K. McAlpine, M. J. Jarvis, & D. G. Bonfield. *Monthly Notices of the Royal Astronomical Society*, 436:1084–1095, Dec 2013.

H. J. McCracken, B. Milvang-Jensen, J. Dunlop, et al. *Astronomy and Astrophysics*, 544:A156, Aug 2012.

M. M. McGilchrist, J. E. Baldwin, J. M. Riley, et al. *Monthly Notices of the Royal Astronomical Society*, 246:110–122, September 1990.

R. J. McLure & M. J. Jarvis. *Monthly Notices of the Royal Astronomical Society*, 337:109–116, November 2002.

R. J. McLure, M. J. Kukula, J. S. Dunlop, et al. *Monthly Notices of the Royal Astronomical Society*, 308:377–404, September 1999.

A. Mechev, J. B. R. Oonk, A. Danezi, et al. In *Proceedings of the International Symposium on Grids and Clouds (ISGC) 2017*, page 2, March 2017.

O. Miettinen, I. Delvecchio, V. Smolčić, et al. *Astronomy and Astrophysics*, 606:A17, Sep 2017.

H. Miraghaei & P. N. Best. *Monthly Notices of the Royal Astronomical Society*, 466: 4346–4363, Apr 2017.

H. Mo, F. C. van den Bosch, & S. White. *Galaxy Formation and Evolution*. Cambridge University Press, May 2010.

H. J. Mo & S. D. M. White. *Monthly Notices of the Royal Astronomical Society*, 282: 347–361, September 1996.

N. Mohan & D. Rafferty. PyBDSF: Python Blob Detection and Source Finder. Astrophysics Source Code Library, February 2015.

E. Momjian, C. L. Carilli, F. Walter, et al. *Astronomical Journal*, 147:6, Jan 2014.

R. Morganti. *Frontiers in Astronomy and Space Sciences*, 4:42, Nov 2017.

S. G. Murray, C. Power, & A. S. G. Robotham. *Astronomy and Computing*, 3:23–34, Nov 2013.

S. T. Myers & S. S. G. (. VLASS Survey Team. In *American Astronomical Society Meeting Abstracts #231*, volume 231 of *American Astronomical Society Meeting Abstracts*, page 231.08, January 2018.

T. Naab & J. P. Ostriker. *Annual Review of Astronomy and Astrophysics*, 55:59–109, Aug 2017.

R. Nan, D. Li, C. Jin, et al. *International Journal of Modern Physics D*, 20:989–1024, Jan 2011.

R. Narayan, R. Mahadevan, & E. Quataert. In M. A. Abramowicz, G. Björnsson, & J. E. Pringle, editors, *Theory of Black Hole Accretion Disks*, pages 148–182, January 1998.

J. F. Navarro, C. S. Frenk, & S. D. M. White. *Astrophysical Journal*, 462:563, May 1996.

J. F. Navarro, C. S. Frenk, & S. D. M. White. *Astrophysical Journal*, 490:493–508, Dec 1997.

H. Netzer, P. Lira, B. Trakhtenbrot, et al. *Astrophysical Journal*, 671:1256–1263, December 2007.

P. Norberg, C. M. Baugh, E. Hawkins, et al. *Monthly Notices of the Royal Astronomical Society*, 332:827–838, June 2002.

R. P. Norris, A. M. Hopkins, J. Afonso, et al. *Publications of the Astronomical Society of Australia*, 28:215–248, August 2011.

Ray P. Norris. *Nature Astronomy*, 1:671–678, September 2017.

M. Novak, V. Smolčić, J. Delhaize, et al. *Astronomy and Astrophysics*, 602:A5, June 2017.

S. J. Oliver, J. Bock, B. Altieri, et al. *Monthly Notices of the Royal Astronomical Society*, 424:1614–1635, August 2012.

R. A. Overzier, H. J. A. Röttgering, R. B. Rengelink, et al. *Astronomy and Astrophysics*, 405:53–72, July 2003.

P. Padovani, N. Miller, K. I. Kellermann, et al. *Astrophysical Journal*, 740:20, Oct 2011.

M. J. Page & F. J. Carrera. *Monthly Notices of the Royal Astronomical Society*, 311: 433–440, Jan 2000.

D. C. Pan, M. S. Vogeley, Hoyle, et al. *Monthly Notices of the Royal Astronomical Society*, 421:926–934, April 2012.

J. A. Peacock. *arXiv Astrophysics e-prints*, September 2003.

J. A. Peacock & R. E. Smith. *Monthly Notices of the Royal Astronomical Society*, 318:1144–1156, November 2000.

John A. Peacock. *Cosmological Physics*. Cambridge University Press, 1999.

T. J. Pearson. *Monthly Notices of the Royal Astronomical Society*, 171:475–505, June 1975.

T. J. Pearson & A. C. S. Readhead. *Annual Review of Astronomy and Astrophysics*, 22:97–130, Jan 1984.

P. J. E. Peebles. *Astrophysical Journal, Letters*, 189:L51, April 1974.

P. J. E. Peebles. *The large-scale structure of the universe*. Princeton University Press, 1980.

A. A. Penzias & R. W. Wilson. *Astrophysical Journal*, 142:419–421, July 1965.

B. M. Peterson. *An Introduction to Active Galactic Nuclei*. Cambridge University Press, February 1997.

M. Pierre, I. Valtchanov, B. Altieri, et al. *Journal of Cosmology and Astroparticle Physics*, 9:011, September 2004.

J. D. H. Pilkington & J. F. Scott. *Memoirs of the RAS*, 69:183, 1965.

Planck Collaboration et al. *Astronomy and Astrophysics*, 571:A16, Nov 2014.

Planck Collaboration et al. *Astronomy and Astrophysics*, 594:A13, Sep 2016.

Planck Collaboration et al. *arXiv e-prints*, art. arXiv:1807.06209, July 2018.

M. B. Pracy, J. H. Y. Ching, E. M. Sadler, et al. *Monthly Notices of the Royal Astronomical Society*, 460:2–17, July 2016.

I. Prandoni & N. Seymour. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 67, April 2015.

I. Prandoni, L. Gregorini, P. Parma, et al. *Astronomy and Astrophysics*, 365:392–399, January 2001.

G. W. Pratt, J. H. Croston, M. Arnaud, et al. *Astronomy and Astrophysics*, 498: 361–378, May 2009.

M. Prescott, T. Mauch, M. J. Jarvis, et al. *Monthly Notices of the Royal Astronomical Society*, 457:730–744, March 2016.

W. H. Press & P. Schechter. *Astrophysical Journal*, 187:425–438, Feb 1974.

J. R. Primack. *New Astronomy Reviews*, 49:25–34, May 2005.

A. Raccanelli, O. Doré, D. J. Bacon, et al. *Journal of Cosmology and Astroparticle Physics*, 1:042, January 2015.

J. F. Radcliffe, M. A. Garrett, T. W. B. Muxlow, et al. *Astronomy and Astrophysics*, 619:A48, Nov 2018.

S. C. Read, D. J. B. Smith, G. Gürkan, et al. *Monthly Notices of the Royal Astronomical Society*, 480:5625–5644, Nov 2018.

M. J. Rees. In R. M. Wald, editor, *Black Holes and Relativistic Stars*, page 79, 1998.

N. Rees. *Monthly Notices of the Royal Astronomical Society*, 244:233–246, May 1990.

John A. Regan, Turlough P. Downes, Marta Volonteri, et al. *arXiv e-prints*, art. arXiv:1811.04953, Nov 2018.

R. Rengelink. In H. J. A. Röttgering, P. N. Best, & M. D. Lehnert, editors, *The Most Distant Radio Galaxies*, page 399, Jan 1999.

A. Rest, D. Scolnic, R. J. Foley, et al. *Astrophysical Journal*, 795:44, Nov 2014.

E. Retana-Montenegro & H. J. A. Röttgering. *Astronomy and Astrophysics*, 600:A97, Apr 2017.

A. G. Riess, A. V. Filippenko, P. Challis, et al. *Astronomical Journal*, 116:1009–1038, September 1998.

A. G. Riess, L. Macri, S. Casertano, et al. *Astrophysical Journal*, 699:539–563, Jul 2009.

A. G. Riess, L. M. Macri, S. L. Hoffmann, et al. *Astrophysical Journal*, 826:56, Jul 2016.

J. M. W. Riley, E. M. Waldram, & J. M. Riley. *Monthly Notices of the Royal Astronomical Society*, 306:31–34, June 1999.

M. S. Roberts & M. P. Haynes. *Annual Review of Astronomy and Astrophysics*, 32: 115–152, 1994.

A. S. G. Robotham, L. J. M. Davies, S. P. Driver, et al. *Monthly Notices of the Royal Astronomical Society*, 476:3137–3159, May 2018.

N. Roche & S. A. Eales. *Monthly Notices of the Royal Astronomical Society*, 307: 703–721, Aug 1999.

N. P. Ross, J. da Ângela, T. Shanks, et al. *Monthly Notices of the Royal Astronomical Society*, 381:573–588, Oct 2007.

H. J. A. Röttgering. In *ISKAF2010 Science Meeting*, page 50, Jan 2010.

V. C. Rubin, Jr. Ford, W. K., & N. Thonnard. *Astrophysical Journal*, 238:471–487, Jun 1980.

V. C. Rubin, D. Burstein, Jr. Ford, W. K., et al. *Astrophysical Journal*, 289:81–104, Feb 1985.

E. M. Sadler, R. D. Cannon, T. Mauch, et al. *Monthly Notices of the Royal Astronomical Society*, 381:211–227, October 2007.

R. J. Sault & T. A. Oosterloo. *arXiv e-prints*, art. astro-ph/0701171, Jan 2007.

W. Saunders, M. Rowan-Robinson, & A. Lawrence. *Monthly Notices of the Royal Astronomical Society*, 258:134–146, September 1992.

J. Schaye, R. A. Crain, R. G. Bower, et al. *Monthly Notices of the Royal Astronomical Society*, 446:521–554, Jan 2015.

E. Schinnerer, M. T. Sargent, M. Bondi, et al. *The Astrophysical Journal Supplement Series*, 188:384–404, Jun 2010.

B. P. Schmidt, N. B. Suntzeff, M. M. Phillips, et al. *Astrophysical Journal*, 507:46–63, November 1998.

J. R. Shakeshaft. In H. C. van de Hulst, editor, *Radio astronomy*, volume 4 of *IAU Symposium*, page 218, 1957.

F. Shankar, D. H. Weinberg, & J. Miralda-Escudé. *Astrophysical Journal*, 690:20–41, January 2009.

Y. Shen, M. A. Strauss, N. P. Ross, et al. *Astrophysical Journal*, 697:1656–1673, Jun 2009.

R. K. Sheth & G. Tormen. *Monthly Notices of the Royal Astronomical Society*, 308: 119–126, Sep 1999.

R. K. Sheth, H. J. Mo, & G. Tormen. *Monthly Notices of the Royal Astronomical Society*, 323:1–12, May 2001.

M. Shimizu, T. Kitayama, S. Sasaki, et al. *Astrophysical Journal*, 590:197–206, Jun 2003.

T. W. Shimwell, H. J. A. Röttgering, P. N. Best, et al. *Astronomy and Astrophysics*, 598:A104, February 2017.

T. W. Shimwell, C. Tasse, M. J. Hardcastle, et al. *Astronomy and Astrophysics*, 622: A1, Feb 2019.

D. Sijacki, V. Springel, T. Di Matteo, et al. *Monthly Notices of the Royal Astronomical Society*, 380:877–900, Sep 2007.

Joseph Silk & Gary A. Mamon. *Research in Astronomy and Astrophysics*, 12:917–946, Aug 2012.

R. A. Skibba, M. S. M. Smith, A. L. Coil, et al. *Astrophysical Journal*, 784:128, Apr 2014.

S. W. Skillman, E. J. Hallman, B. W. O'Shea, et al. *Astrophysical Journal*, 735:96, Jul 2011.

O. M. Smirnov. *Astronomy and Astrophysics*, 527:A106, March 2011.

O. M. Smirnov & C. Tasse. *Monthly Notices of the Royal Astronomical Society*, 449: 2668–2684, May 2015.

D. J. B. Smith, P. N. Best, K. J. Duncan, et al. In *SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*, pages 271–280, December 2016.

V. Smolčić, M. Novak, M. Bondi, et al. *Astronomy and Astrophysics*, 602:A1, June 2017a.

V. Smolčić, I. Delvecchio, G. Zamorani, et al. *Astronomy and Astrophysics*, 602:A2, June 2017b.

V. Smolčić, E. Schinnerer, G. Zamorani, et al. *Astrophysical Journal*, 690:610–618, Jan 2009.

V. Smolčić, M. Novak, I. Delvecchio, et al. *Astronomy and Astrophysics*, 602:A6, June 2017.

D. N. Spergel, L. Verde, H. V. Peiris, E. Komatsu, et al. *The Astrophysical Journal Supplement Series*, 148:175–194, Sep 2003.

S. Starikova, S. Berta, A. Franceschini, et al. *Astrophysical Journal*, 751:126, June 2012.

I. Strateva, Ž. Ivezić, G. R. Knapp, et al. *Astronomical Journal*, 122:1861–1874, Oct 2001.

G. Swarup, S. Ananthakrishnan, V. K. Kapahi, et al. *Current Science*, 60:95, Jan 1991.

G. P. Szokoly, J. Bergeron, G. Hasinger, et al. *Astrophysical Journal, Supplement*, 155:271–349, December 2004.

C. Tasse. *Astronomy and Astrophysics*, 566:A127, June 2014a.

C. Tasse. *ArXiv e-prints*, October 2014b.

C. Tasse, A. S. Cohen, H. J. A. Röttgering, et al. *Astronomy and Astrophysics*, 456: 791–800, September 2006.

C. Tasse, H. J. A. Röttgering, P. N. Best, et al. *Astronomy and Astrophysics*, 471: 1105–1116, September 2007.

C. Tasse, P. N. Best, H. Röttgering, et al. *Astronomy and Astrophysics*, 490:893–904, November 2008.

C. Tasse, B. Hugo, M. Mirmont, et al. *Astronomy and Astrophysics*, 611:A87, April 2018.

G. B. Taylor, C. L. Carilli, & R. A. Perley, editors. *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*, 1999.

M. Taylor. TOPCAT: Tool for OPerations on Catalogues And Tables, Jan 2011.

P. Taylor & C. Kobayashi. *Monthly Notices of the Royal Astronomical Society*, 442: 2751–2767, August 2014.

M. Tegmark, M. R. Blanton, M. A. Strauss, et al. *Astrophysical Journal*, 606:702–740, May 2004.

P. Thomasson. *Quarterly Journal of the Royal Astronomical Society*, 27:413–431, September 1986.

A. R. Thompson. In G. B. Taylor, C. L. Carilli, & R. A. Perley, editors, *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*, page 11, 1999.

S. J. Tingay, R. Goeke, J. D. Bowman, et al. *Publications of the Astronomical Society of Australia*, 30:e007, January 2013.

J. Tinker, A. V. Kravtsov, A. Klypin, et al. *Astrophysical Journal*, 688:709–728, December 2008.

J. L. Tinker, B. E. Robertson, A. V. Kravtsov, et al. *Astrophysical Journal*, 724: 878–886, December 2010.

J. L. Tinker, J. R. Brownstein, H. Guo, et al. *Astrophysical Journal*, 839:121, Apr 2017.

M. Trevisan & G. A. Mamon. *Monthly Notices of the Royal Astronomical Society*, 471:2022–2038, October 2017.

R. Turner et al. in prep.

C. M. Urry & P. Padovani. *Publications of the Astronomical Society of the Pacific*, 107:803, September 1995.

F. van den Bosch, X. Yang, & H. J. Mo. In R. Dettmar, U. Klein, & P. Salucci, editors, *Baryons in Dark Matter Halos*, page 41, December 2004.

F. C. van den Bosch, X. Yang, H. J. Mo, et al. *Monthly Notices of the Royal Astronomical Society*, 376:841–860, April 2007.

F. C. van den Bosch, S. More, M. Cacciato, et al. *Monthly Notices of the Royal Astronomical Society*, 430:725–746, April 2013.

S. van der Walt, S. C. Colbert, & G. Varoquaux. *Computing in Science and Engineering*, 13:22–30, Mar 2011.

M. P. van Haarlem, M. W. Wise, A. W. Gunst, et al. *Astronomy and Astrophysics*, 556:A2, August 2013.

R. J. van Weeren, W. L. Williams, M. J. Hardcastle, et al. *Astrophysical Journal, Supplement*, 223:2, March 2016.

D. E. Vanden Berk, G. T. Richards, A. Bauer, et al. *Astronomical Journal*, 122: 549–564, August 2001.

M. Vogelsberger, S. Genel, V. Springel, et al. *Nature*, 509:177–182, May 2014.

A. M. von Benda-Beckmann & V. Müller. *Monthly Notices of the Royal Astronomical Society*, 384:1189–1199, March 2008.

E. M. Waldram, G. G. Pooley, K. J. B. Grainge, et al. *Monthly Notices of the Royal Astronomical Society*, 342:915–925, July 2003.

Z. L. Wen, J. L. Han, & F. S. Liu. *The Astrophysical Journal Supplement Series*, 199: 34, Apr 2012.

I. H. Whittam, J. M. Riley, D. A. Green, et al. *Monthly Notices of the Royal Astronomical Society*, 462:2122–2137, October 2016.

I. H. Whittam, M. Prescott, K. McAlpine, M. J. Jarvis, & I. Heywood. *Monthly Notices of the Royal Astronomical Society*, 480:358–370, October 2018.

W. L. Williams, R. J. van Weeren, H. J. A. Röttgering, et al. *Monthly Notices of the Royal Astronomical Society*, 460:2385–2412, August 2016.

W. L. Williams, G. Calistro Rivera, P. N. Best, et al. *Monthly Notices of the Royal Astronomical Society*, 475:3429–3452, April 2018.

W. L. Williams, M. J. Hardcastle, P. N. Best, et al. *Astronomy and Astrophysics*, 622:A2, Feb 2019.

C. J. Willott, S. Rawlings, K. M. Blundell, et al. *Monthly Notices of the Royal Astronomical Society*, 309:1017–1033, Nov 1999.

R. J. Wilman, H. J. A. Röttgering, R. A. Overzier, et al. *Monthly Notices of the Royal Astronomical Society*, 339:695–700, March 2003.

R. J. Wilman, L. Miller, M. J. Jarvis, et al. *Monthly Notices of the Royal Astronomical Society*, 388:1335–1348, August 2008.

R. J. Wilman, M. J. Jarvis, T. Mauch, et al. *Monthly Notices of the Royal Astronomical Society*, 405:447–461, June 2010.

T. L. Wilson, K. Rohlfs, & S. Hüttemeister. *Tools of Radio Astronomy*. Springer-Verlag, 2009. doi: 10.1007/978-3-540-85122-6.

R. Windhorst, D. Mathis, & L. Neuschaefer. In *Evolution of the Universe of Galaxies*, volume 10 of *Astronomical Society of the Pacific Conference Series*, pages 389–403, 1990.

C. Wu, O. I. Wong, L. Rudnick, et al. *Monthly Notices of the Royal Astronomical Society*, 482:1211–1230, January 2019.

G. Yang, W. N. Brandt, F. Vito, et al. *Monthly Notices of the Royal Astronomical Society*, 475:1887–1911, Apr 2018.

X. Yang, H. J. Mo, & F. C. van den Bosch. *Monthly Notices of the Royal Astronomical Society*, 339:1057–1080, Mar 2003.

X. Yang, H. J. Mo, & Frank C. van den Bosch. *Astrophysical Journal*, 676:248–261, March 2008.

I. Zehavi, M. R. Blanton, J. A. Frieman, et al. *Astrophysical Journal*, 571:172–190, May 2002.

I. Zehavi, D. H. Weinberg, Z. Zheng, et al. *Astrophysical Journal*, 608:16–24, June 2004.

I. Zehavi, Z. Zheng, D. H. Weinberg, et al. *Astrophysical Journal*, 630:1–27, September 2005.

I. Zehavi, Z. Zheng, D. H. Weinberg, et al. *Astrophysical Journal*, 736:59, July 2011.

Y. Zhang, X. Yang, H. Wang, et al. *Astrophysical Journal*, 779:160, Dec 2013.

Z Zheng, A. A. Berlind, D. H. Weinberg, et al. *Astrophysical Journal*, 633:791–809, Nov 2005.

Z. Zheng, A. L. Coil, & I. Zehavi. *Astrophysical Journal*, 667:760–779, Oct 2007.

J. Zwart, J. Wall, A. Karim, et al. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 172, April 2015a.

J. T. L. Zwart, M. Santos, & M. J. Jarvis. *Monthly Notices of the Royal Astronomical Society*, 453:1740–1753, Oct 2015b.

F. Zwicky. *Helvetica Physica Acta*, 6:110–127, Jan 1933.