Astronomical Catalogue Cross Identification for Data Mining and Statistical Analysis of the Infrared and Faint Radio Sky

Stuart Duncan WESTON

A thesis submitted to

Auckland University of Technology

in fulfilment of the requirements for the degree of

Doctor of Philosophy (PhD)

2020

Institute for Radio Astronomy & Space Research

School of Engineering, Computer & Mathematical Sciences

Declaration of Authorship

I, Stuart Duncan WESTON, declare that this thesis titled, "Astronomical Catalogue Cross Identification for Data Mining and Statistical Analysis of the Infrared and Faint Radio Sky" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: 29th July 2019

The Roads go ever ever on

Roads go ever ever on, Over rock and under tree, By caves where never sun has shone, By streams that never find the sea; Over snow by winter sown, And through the merry flowers of June, Over grass and over stone, And under mountains in the moon.

Roads go ever ever on, Under cloud and under star. Yet feet that wandering have gone Turn at last to home afar. Eyes that fire and sword have seen, And horror in the halls of stone Look at last on meadows green, And trees and hills they long have known.

The Road goes ever on and on Down from the door where it began. Now far ahead the Road has gone, And I must follow, if I can, Pursuing it with eager feet, Until it joins some larger way, Where many paths and errands meet.

The Road goes ever on and on Down from the door where it began. Now far ahead the Road has gone, And I must follow, if I can, Pursuing it with weary feet, Until it joins some larger way, Where many paths and errands meet. And whither then? I cannot say.

The Road goes ever on and on Out from the door where it began. Now far ahead the Road has gone. Let others follow, if they can! Let them a journey new begin. But I at last with weary feet will turn towards the lighted inn, My evening rest and sleep to meet.

J.R.R. Tolkien, The Lord of the Rings

Abstract

Cross-identifying complex radio sources with optical or infrared counterparts in surveys such as the Australia Telescope Large Area Survey (ATLAS) has traditionally been performed by a visual inspection of individual sources. However, with new surveys from the Australian Square Kilometre Array Pathfinder detecting many tens of million of radio sources such an approach is no longer feasible. The Likelihood Ratio (LR) allows the use of additional data about survey objects to be cross matched rather than just position and proximity, such as flux, probability distribution and catalogue surface density. This thesis presents a new software algorithm (LRPY - Likelihood Ratio in PYthon) to automate the process of cross-identifying radio sources with catalogues at other wavelengths using the LR. I demonstrate LRPY by applying it to the ATLAS Data Release 3 and a *Spitzer*-based multi-wavelength catalogue, identifying 3,848 cross matched sources using my LR-based selection criteria. I found that LRPY could be extended to identify radio sources with multiple infrared counterparts many of which I identify as interacting galaxy pairs. In addition I also investigated if LRPY could be used to select radio sources with more complex morphology, such as double lobe radio sources. This extension to the algorithm is shown to work well in identifying the host of many double lobe radio sources. A subset of 1987 cross matched sources in this thesis have flux density values for all four bands in the *Spitzer*/IRAC instrument, which allowed me to use various criteria to distinguish between active galactic nuclei (AGN) and star-forming galaxies (SFG). I found that 936 radio sources ($\approx 47\,\%$) meet both of the Lacy and Stern AGN selection criteria. Of the matched sources, 295 have spectroscopic redshifts and we examine the radio to infrared flux ratio vs redshift, proposing

an AGN selection criterion below the Elvis radio-loud (RL) AGN limit for this dataset. Taking the union of all three AGN selection criteria I have identified 956 cross matched sources as AGN ($\approx 48\,\%$). From this dataset, we find a decreasing fraction of AGN with lower radio flux densities consistent with other results in the literature. I found there is a strong power law correlation demonstrated using the complete Fusion dataset seen in the AGN arm of the MIR $[3.6]-[5.8]$ vs $[4.5]-[8.0]$ plot for the LACY AGN selection wedge, but there was no such strong correlation for the Stern AGN wedge of the MIR $[5.8]-[8.0]$ vs $[3.6]-[4.5]$ colour-colour plot. Due to this I have analysed the cross matched sources in this thesis for a power law relationship, it was apparent that for the MIR $[3.6]-[5.8]$ vs $[4.5]-[8.0]$ colour-colour plots used for Lacy AGN selection the correlation between the AGN candidate infrared flux ratios to the power-law locus is very strong in the AGN arm. The Stern MIR colour-colour criteria $[5.8]-[8.0]$ vs $[3.6]-[4.5]$ shows no strong correlation for the AGN candidate infrared flux ratios to the power-law locus for the cross matched sources. I have also looked at the relationship between the MIR 8.0μm to radio $1.4\,$GHz relationship, there is a clear distinction between AGN and SFG's. I have also looked at the relationship between the Radio $L_{1.4\mathrm{GHz}}$ and MIR luminosities $L_{3.6\,\mu\mathrm{m}}$ and $L_{4.5\,\mu\mathrm{m}}$, as prior work had provided existence of a correlation. What was found is that the higher redshift sources ($z > 0.3$) lie within the top right quadrant of the plots, when the Stern and Lacy AGN are identified we see the majority of these AGN also lie within this same quadrant. Looking at the AGN selected sources we see a slope close to unity at both MIR wavelengths, for $3.6\,\mu$m the slope is $m = 1.004 \pm 0.155$ and at $4.5\,\mu$m it was found to be $m = 0.980 \pm 0.153$. I conclude the thesis by summarizing the work and discussing future work to be undertaken as a result of the research.

Acknowledgements

First I would like thank my primary supervisor Prof. Sergei Gulyaev for hiring me to work at the Institute, and providing the challenge of contributing in establishing ourselves as a functioning geodetic and astronomy VLBI station with the 12m antenna. Also the opportunity to contribute and work on the conversion of a Satellite Earth station into a fully functional radio telescope able to provide valuable data for Science and Geodesy. When that 200 tons of steel first moved under the control of the Field System I had installed and we collected data, I was over the moon. In addition I also thank my other AUT colleagues Tim Natusch, Lewis Woodburn for making a great team in achieving our goal of establishing New Zealand once again in the field of Radio Astronomy; Patricia Sallis for all the administrative help and assistance plus some proof reading over the years.

Next I wish to thank and acknowledge my supervisors Dr. Nicholas Seymour [1] & Dr. Ray Norris [2]. Nick has been a very steady partner in this journey and has shown immense patience, thanks mate. Ray for being a hard critique when the sloppy student needed it most to pull his socks up, we got it published and I hope the wine went down well.

I would like to thank Dr W. Miller Goss of the National Radio Astronomy Observatory (NRAO) for his early encouragement one evening when running him to his accommodation after an Auckland Astronomical Society meeting he had presented at. Also in addition for the material about early New Zealand Radio Astronomy with respect to

[1] International Centre for Radio Astronomy Research, Curtin University, Perth, Australia

[2] CSIRO, P.O. Box 76, Epping, NSW 1710, Australia & Western Sydney University, Penrith, NSW, Australia

Elizabeth Alexander and John Bolton & Gordon Stanley that he provided latterly, it was good to add a Kiwi perspective to the story.

I also thank Dr. L. Dunne and Dr. S. Maddox for hosting myself for initial advice at the University of Canterbury and subsequent helpful discussions about catalogues.

I thank the referee, Jim Condon, for critically reading the manuscript and suggesting substantial improvements which significantly improved the paper Weston et al. (2018).

I am a recipient of a CSIRO Astronomy & Space Science, Australia Telescope National Facility Graduate Research Scholarship. Thus I would like to thank CSIRO Astronomy & Space Science for awarding me a Australia Telescope National Facility Graduate Research Scholarship. The time spent as Duty Astronomer at the ATCA, Narrabri and latterly visiting Marsfield was invaluable. So a big thank you to Robin Wark for her very patient supervision during the learning curve of being a ATCA Duty Astronomer, oh the fun of programming for 64Meg Zoom mode at 2am. Also the many other staff at ATNF who made my stays enjoyable and others for dealing with a stupid Pom operating the New Zealand end of the LBA (Chris Phillips, Jamie Stevens and Cormac Reynolds).

From this I must also thank our colleagues from the University of Tasmania Simon Ellingsen, Jamie McCallum & Gabor Orosz for being so cooperative and helpful with other projects such as AuScope, the first fringe test between the Warkworth 30m and the Hobart 26m and latterly for the VLBI at 7.6 & 7.8 GHz on a Methanol Maser flare of G358 (exciting stuff).

To an old friend who understands my struggle with the written word Mark Stevenson and for being a non-technical proof reader for my dyslexia, a big thank you. He also

kept nagging me about how it was progressing, invaluable in hindsight.

This research has made use of the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration. The tools and resources now available at the stroke of a fey keys via a computer these days with an internet connection is truly amazing compared to my very early undergraduate days in the library researching articles on microfiche (who remembers those days?).

Finally all the many great people I have met within the VLBI community over the last decade, it really is an international family where without very close international cooperation the data would not be collected and the science produced.

Last but not least my family for putting up with the trials and tribulations of reaching this destination.

PS: Is there any one else I missed? If I have then also a big thank you.

Contents

[3]Parts of this section were first published in Weston et al. (2018)

List of Figures

List of Tables

List of Acronyms

LAH	**L**ist **A**bbreviations **H**ere
AGN	**A**ctive **G**alactic **N**uclei
ASCII	**A**merican **S**tandard **C**ode for **I**nformation **I**nterchange
ASKAP	**A**ustralian **S**quare **K**ilometre **A**rray **P**athfinder
ATCA	**A**ustralia **T**elescope **C**ompact **A**rray
ATLAS	**A**ustralia **T**elescope **L**arge **A**rea **S**urvey
CANDLES	**C**osmic **A**ssembly **N**ear-IR **D**eep**L**egacy **S**urvey
CDFS	**C**handra **D**eep **F**ield **S**outh
COSMOS	The Cosmic Evolution Survey
CSIRO	**C**ommonwealth **S**cientific and **I**ndustrial **R**esearch **O**rganisation
DEC	**Dec**lination
ELAIS	**E**uropean **L**arge **A**rea **I**SO **S**urvey
EMU	**E**volutionary **M**ap of the **U**niverse
ESA	**E**uropean **S**pace **A**gency
FIRST	**F**aint **I**mages of the **R**adio **S**ky at **T**wenty-Centimeters
FK5	Fifth Fundamental Catalogue
FWHM	**F**ull **W**idth **H**alf **M**aximum
GOODS	**G**reat **O**bservatories **O**rigins **D**eep **S**urvey
GPL	**G**eneral **P**ublic **L**icense
HELP	**H**erschel **E**xtragalactic **L**egacy **P**roject
HST	**H**ubble **S**pace **T**elescope
ICRS	**I**nternational **C**elestial **R**eference **S**ystem
IRAC	**I**nfra **R**ed **A**rray **C**amera

IRD	**I**nfra **R**ed **D**ouble
IVS	**I**nternational **V**LBI **S**ervice for Geodesy & Astrometry
JVLA	Karl G. **J**ansky **V**ery **L**arge **A**rray
LF	**L**unimosity **F**unction
LRPY	**L**ikelihood **R**atio for **Py**thon
MeerKAT	Originally the **K**aroo **A**rray **T**elescope MeerKAT is a precursor for the SKA-mid array in South Africa
MIPS	**M**ultiband **I**maging **P**hotometer on the **S**pitzer
MIR	**M**id **I**nfra**R**ed
NIR	**N**ear **I**nfra**R**ed
NN	**N**earest **N**eighbour
NRAO	**N**ational **R**adio **A**stronomy **O**bservatory
NVSS	**N**RAO **V**LA **S**ky **S**urvey
OzDES	The **Oz** (Australian) **D**ark **E**nergy **S**urvey
PAH	**P**olycyclic **A**romatic **H**ydrocardons
PP	**P**oisson **P**robability
RA	**R**ight **A**scension
RDBMS	**R**elational **D**ata**b**ase **M**anagement **S**ystem
RLAGN	**R**adio **L**oud **A**ctive **G**alactic **N**uclei
PDF	**P**robability **D**istribution **F**unction
RLAGN	**R**adio **L**oud **A**ctive **G**alactic **N**uclei
SED	**S**pectral **E**nergy **D**istribution
SDSS	**S**loan **D**igital **S**ky **S**urvey
SFG	**S**tar **F**orming **G**alaxies
SFR	**S**tar **F**orming **R**ate
SKA	**S**quare **K**ilometer **A**rray
SMBH	**S**uper **M**assive **B**lack **H**ole
SNR	**S**ignal to **N**oise **R**atio
SPIRE	**S**pectral and **P**hotometric **I**maging **R**eceiver
SUMSS	The **S**ydney **U**niversity **M**olonglo **S**ky **S**urvey
SWIRE	**S**pitzer **W**ide-area **I**nfra**R**ed **E**xtragalactic survey

TENIS	**T**aiwan **E**CDFS **N**ear-**I**nfrared **S**urvey
TOPCAT	**T**ool for **OP**erations on **C**atalogues **A**nd **T**ables
VIKING	**VI**STA **K**ilo-Degree **In**frared **G**alaxy Survey
VIMOS	**Vi**sible **M**ulti-**O**bject **S**pectrograph
VISTA	**V**isible and **I**nfrared **S**urvey **T**elescope for **A**stronomy
VLA	**V**ery **L**arge **A**rray
VLBI	**V**ery **L**ong **B**aseline **I**nterferometry
VLT	**V**ery **L**arge **T**elescope
WISE	**W**ide-field **I**nfrared **S**urvey **E**xplorer
XID	Cross (**X**) **ID**entification

Physical Constants

Speed of Light	c	=	$2.997\,924\,58 \times 10^8$ ms^{-1} (exact)
Hubble constant	H_0		70 km s^{-1} Mpc^{-1}
Omega matter	Ω_M		0.27
Omega lambda	Ω_Λ		0.73
Solar Luminosity	$L_\odot$		3.828×10^{26} W

Symbols

a	distance	m
Jy	Jansky	10^{-26} W m^{-2} Hz^{-1}
P	power	W (Js^{-1})
ω	angular frequency	$rads^{-1}$
z	redshift	

Dedicated to Those who came before and those who will come after.

Moon rise over the Australian Telescope Compact Array (ATCA), which was used for the ATLAS survey resulting in the catalogue used in this thesis. I was fortunate to visit this site on several occasions as Duty Astronomer as part of my CSIRO Astronomy & Space Science, Australia Telescope National Facility Graduate Research Scholarship. Credit: Stuart Weston

Chapter 1

Introduction

In this opening chapter I provide a brief overview of the first multi-wavelength astronomical observations and the problem of cross-identifying objects between the different wavelengths, one of these highly significant observations took place here in New Zealand. I then introduce more recent large scale surveys and the multi-wavelength cross-identifying issues.

1.1 The early days of Radio Astronomy

In 1933 Karl Jansky (Jansky, 1933) working for Bell Labs was tasked with investigating static that interfered with the development of long range radio communication. He identified one source of static that increased and decreased in amplitude, this cyclic nature coinciding with the period of 23 hours and 56 minutes, characteristic of the motion of fixed stars and the Milky Way (a sidereal day). He thus concluded that the source was beyond the Earth. The maximum amplitude of the static coincided with the constellation Sagittarius (what we now know to be the centre of the Milky Way galaxy) being at the zenith. This "hiss" Jansky concluded was radio electromagnetic radiation coming from the Milky Way. Although he played no further significant role, his discoveries led to a realization that other regions of the electro-magnetic spectrum

were available to astronomers. It is not unreasonable to claim that he was responsible for the birth of the field of study we now know as radio astronomy.

Another significant piece of work was done by Grote Reber (Reber, 1944) using a home built 9.57m (31.4 ft) dish producing a radio intensity map at 160 MHz. He comments in his paper that the radio contour map produced showed that the brightest areas corresponded with the Milky Way. Due to the angular resolution no individual sources could be resolved, but this was a significant observation and development after Karl Jansky's observations.

Also during this period the field of radio science had advanced at a phenomenal pace especially during the Second World War for Radar and Radio Communications. Several research and operational groups had noticed a correlation between radio interference and solar activity especially at sunset and sunrise. For example the "Norfolk Island Affect" was reported by Elizabeth Alexander (Alexander, 1946) here in New Zealand (An interesting summary of Elizabeth Alexander is provided in (Harris, 2017)), and by a separate group in the United Kingdom (Lovell and Banwell, 1946). There was also a similar report by the early Radio Physics group in Australia (Pawsey et al., 1946). After the War some of these people moved back to academic research which lead to the development of radio as a new field to enhance and extend the study of astronomy, and for astronomy to become multi-wavelength.

1.2 Early Cross Identification of Radio Sources

The next significant step was to overcome the positional errors with the early radio instruments (minutes of arc) for the identified discrete radio sources. Due to the positional errors they could not be cross matched with known optical astronomical objects with any certainty. The angular resolution θ of an instrument which can be approximated by:

$$\theta \approx \frac{\lambda}{D} \tag{1.1}$$

where λ is the wavelength of the observed electromagnetic radiation, and D is the diameter of the telescope's objective. The resulting angular resolution θ is in radians. To increase the angular resolution by building an antenna with a very large objective entailed structural and engineering problems with an associated significant financial cost to solve; an alternative was to build large arrays of dipoles or by the use of interferometry.

Interferometry is a method where by one wave front from a single source is received at two spatially separated locations. The distance between these two spatially separated locations is D in Equation 1.1. Michelson and Pense (Michelson and Pease, 1921) used such a system on the Mt Wilson 100-inch optical telescope as seen in Figure 1.1 to measure stellar diameters. As these two components are combined; where they are in phase constructive interference is obtained and where they are out of phase destructive interference results producing what is commonly referred to as "fringes". This technique was very difficult on the optical instruments of the time as the fringes were not stable and vibrated randomly due to atmospheric fluctuations (for a informative review see (Brown, 1968b) and (Brown, 1968a)). But it was realised very early on that with radio astronomy as the received electromagnetic radiation is converted to an electrical signal ((Ryle and Vonberg, 1946) and (Ryle, 1952)) and the wavelengths observed (100's MHz) were 10's of metres with a much lower atmospheric impact made implementation of the interferometry technique achievable (see Figure 1.2 for diagram of an early interferometer).

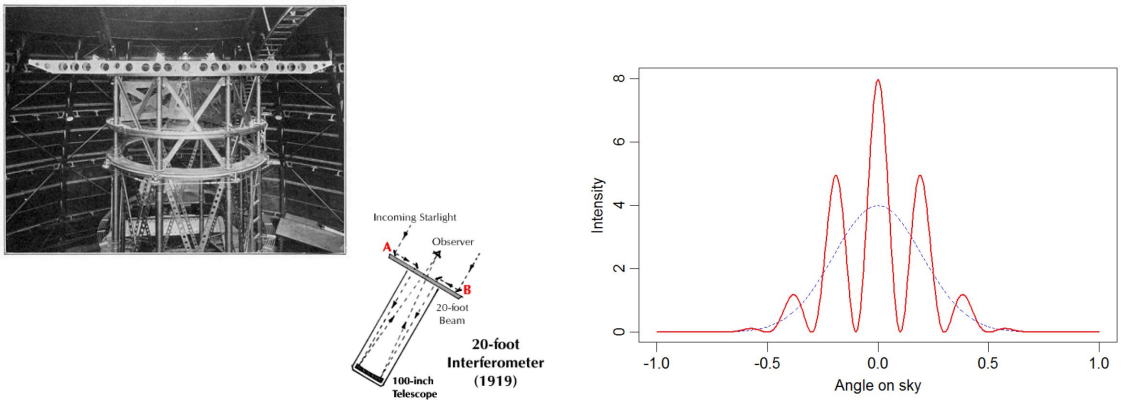


FIGURE 1.1: The image on the left hand side shows the optical interferometer instrument mounted on the Mt Wilson 100-inch optical telescope, and the light path. On the right hand side is a representation of the fringes (red solid line) against the blue dashed line which would be the envelope of the interferometric pattern for a uniform aperture.

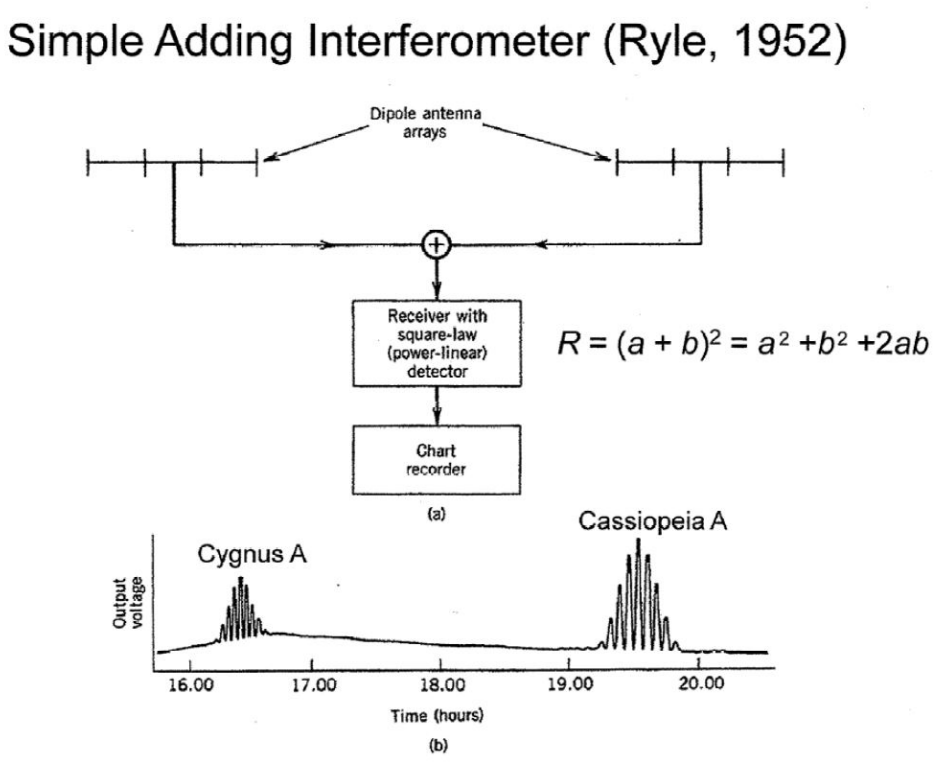


FIGURE 1.2: The top part shows a simple block diagram of the Cambridge Interferometer where the signals are combined additively (a). The bottom part shows the recording obtained for two radio sources, Cygnus A and Cassiopeia A (b). This figure is taken from (Ryle, 1952)

In an attempt to overcome the angular resolution issue, Bolton and Stanley (Bolton and Stanley, 1948) undertook an experiment using a cliff top interferometer (see Figure 1.3),

initially at Dover Heights on the east coast of Australia near to Sydney. They provided the first positive evidence of a discrete non-solar radio source with sufficient angular resolution (10 secs in RA and 7' in Dec) to associate this radio source with Cygnus A. In addition they were able identify two components to the source.

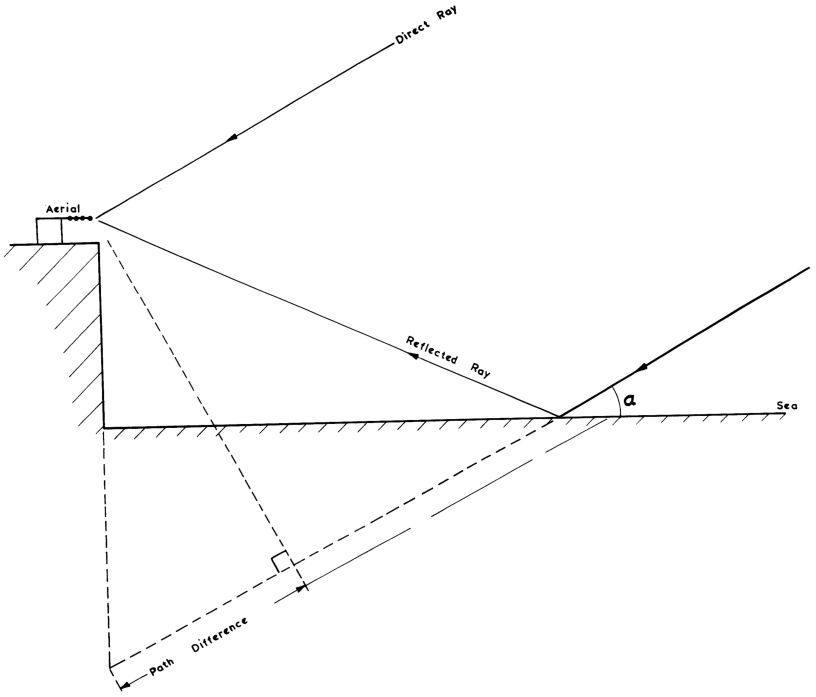


FIGURE 1.3: The cliff top interferometer (Lloyd's mirror interferometer) setup as used by Bolton and Stanley in 1948

Later Bolton and Stanley (Bolton et al., 1949) used the same equipment in the North Island of New Zealand at Leigh on the east coast and Piha[1] on the west coast. These two locations are north of Auckland and separated by about 100 km across the isthmus, making the logistics of moving the instruments manageable. Most importantly it allowed the radio sources to be observed rising and setting something that was not possible from Dover Heights, Australia. From these observations they were able to obtain more precise positions where D in Equation 1.1 at 100 MHz was several hundred metres due to the natural heights of the cliffs at the locations selected. They achieved sub-arcminute resolution in right ascension from their work and felt confident to match

[1] An account of this event is given by Coney (2013)

their radio sources to known optical objects Taurus A, Virgo A and Centaurus A with errors of 30” to 60” in Right Ascension and 7’ to 10’ in Declination (reproduced in Figure 1.4). At the time Virgo A (NGC4486) and Centaurus A (NGC 5128) were generally classed as extra-galactic nebulae, both now known to be galaxies.

Source	Position (Epoch 1948)		Possible associated visible object		
	Right ascension	Declination	Object	Spectrum	Remarks
Taurus A	$5^h\ 31^m\ 00^s \pm 30^s$	$+22^\circ\ 01' \pm 7'$	N.G.C. 1952 (Messier 1)	Continuous. Weak emission lines of H, He, forbidden lines of N, O and Si	The Crab nebula, expanding shell of an old super-nova
Virgo A	$12^h\ 28^m\ 06^s \pm 37^s$	$+12^\circ\ 41' \pm 10'$	N.G.C. 4486 (Messier 87)	Continuous	Spherical nebula—unresolved
Centaurus A	$13^h\ 22^m\ 20^s \pm 60^s$	$-42^\circ\ 37' \pm 8'$	N.G.C. 5128	Continuous. Weak emission lines, $H\beta$, $H\gamma$, $H\delta$, and λ 4686	Unresolved nebula crossed by a marked obscuring band

FIGURE 1.4: The table reproduced from (Bolton et al., 1949) showing the possible optical counterparts with position errors from their radio observations.

In Figure 1.5 is a facsimile of the original chart recorder output of the fringe pattern produced while Bolton and Stanley observed Taurus A rising at Pakiri Hill, New Zealand with the cliff top interferometer.

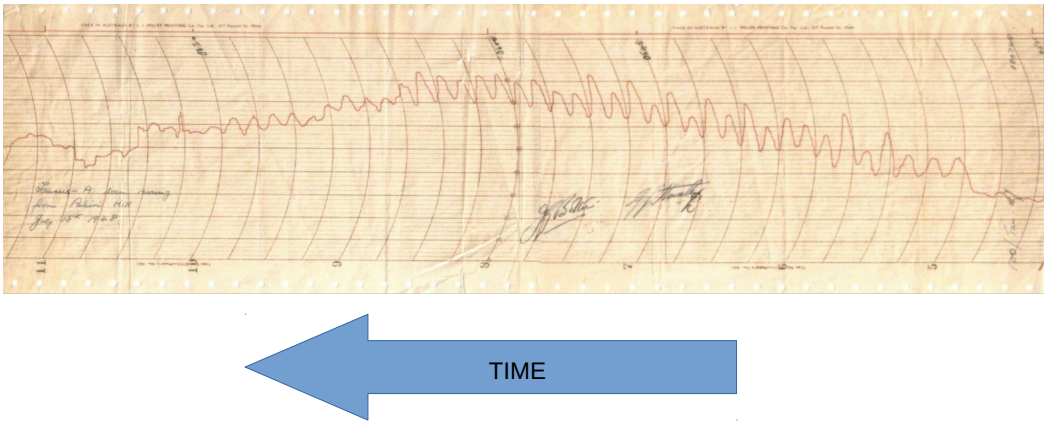


FIGURE 1.5: The cliff top interferometer chart recorder output for the fringes of Taurus A rising from Pakiri Hill, July 18th 1948. Credit : Greenwood Family Archive and Miller Goss

In optical astronomy, surveys and cataloging was already very well established, so the next logical progression for radio was to also start searching the sky for additional discrete radio sources and recording their positions, a survey at radio frequencies to produce catalogues. The first of such work published was undertaken by Ryle of Cambridge University using the Cambridge Interferometer producing a series of catalogues, the first being (Ryle et al., 1950). This first catalogue listed 50 sources and the authors proposed, as the distribution of sources showed no concentration within the galactic plane, that they could be outside the Milky Way galaxy. This work continued to produce the third Cambridge (3C) catalogue in 1959 (Edge et al., 1959) culminating in the revised 3C catalogue (Bennett, 1962) with 328 discrete radio sources.

From these sky surveys two populations of sources were identified, the Class I objects were generally concentrated about the galactic plane but the Class II were more uniformly distributed. Some of these Class I objects had been previously identified (Bolton et al., 1949). At the time the majority of the Class II objects were not associated with optically visible objects, but indications were that they could be extra-galactic and at great distance. During 1962 a series of lunar occultations of one of these strong radio sources 3C273, were observed by Hazard, Mackey and Shimmins (Hazard et al., 1963), with the newly completed Parkes radio telescope. The analysis of these observations gave the position of the two components with sub-arcsecond precision. Most importantly the position was given in the optical reference frame, which allowed a unique identification of the optical counterpart. Such arc-second precision had not been achieved previously. It was soon found (Schmidt, 1963) that component B was coincident with a bright 13th magnitude stellar object, while component A coincided with a faint wisp or jet which ends 20" from B. Schmidt's spectrum of the bright 'star' obtained with the 200-inch Mt Palomar telescope. The spectrum obtained by Schmidt could only be explained by a redshift of 0.158 which allowed identification of the Balmer lines of Hydrogen, in addition to emission from other lines. Now with two similar stellar-like objects, 3C48 and 3C273, the possibility of a chance-identification was vanishingly small. With a redshift of 0.158 these objects were quite clearly very distant in effect at cosmological distances. These new objects were called "quasars" (quasi-stellar), and the remainder

of 1963 saw a dramatic change in our understanding of the Universe (a review of this work is provided by (Hazard et al., 2015)).

These initial and subsequent surveys are provided invaluable data and tools for cosmology and astrophysics. Catalogues now exist of objects at different wavelengths, depths and resolution which allow for different properties to be examined and compared, leading to a more complete understanding of these objects.

1.3 Rationale and Significance of the Study

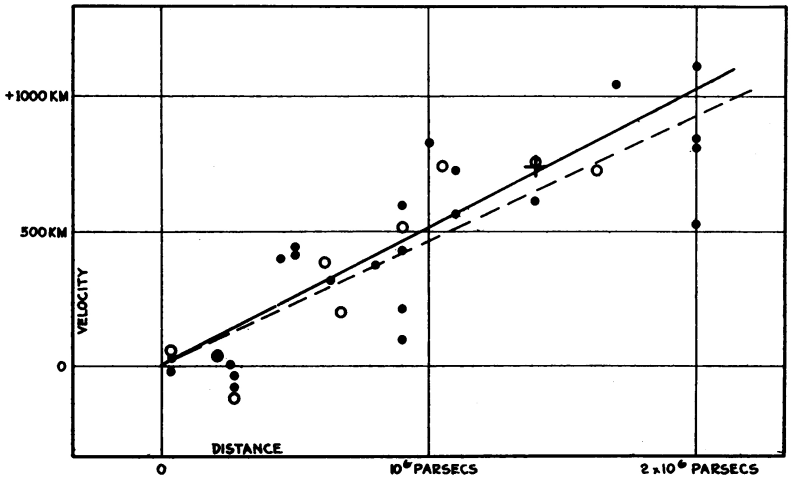


FIGURE 1

Velocity-Distance Relation among Extra-Galactic Nebulae.

Radial velocities, corrected for solar motion, are plotted against distances estimated from involved stars and mean luminosities of nebulae in a cluster. The black discs and full line represent the solution for solar motion using the nebulae individually; the circles and broken line represent the solution combining the nebulae into groups; the cross represents the mean velocity corresponding to the mean distance of 22 nebulae whose distances could not be estimated individually.

FIGURE 1.6: The figure reproduced from Hubble (1929)

In the early part of the last century observational measurements of the doppler-shifted spectral lines of what were then called Nebulae (later to be recognised as galaxies independent from our own) by Vesto Slipher (1922) in the local region showed a consistent redshift indicating they were moving rapidly away from the Earth, some at approximately $\sim$ 1300 km s^{-1} (Slipher, 1921). Slipher's careful work and measurements determined that of his sample of 41 galaxies, 36 had redshifts and the remaining 5 had blue shifts. If we had a steady state universe it would be expected to find an approximately even distribution of doppler-shifts, as it would be expected that there would be an equal number of galaxies approaching (blue-shift) as moving away (redshift); but not a bias to either red or blue. Edwin Hubble (see Figure 1.6) perfected a method to measure the distances of galaxies using variable stars called Cepheids (Hubble, 1925). This method provided the observational data later used to propose H_0 (the Hubble constant, for an expanding universe) (Hubble, 1929). More recently Hubble's model needing revising due

to observational evidence of an expanding and accelerating universe (Schmidt et al., 1998) requiring a non-zero value for the cosmological constant (Λ[2]) . In 1928 cosmologist Howard Robertson used Slipher's redshifts and Hubble's published distances (Robertson, 1929) and provided for the first time a more rigorous relationship between a galaxy's velocity and its distance:

$$velocity\ of\ galaxy = constant \times distance$$

This relationship lead to the constant we now call the Hubble Constant which defines a linear correlation between redshift and distance; its value has been refined over the years to $\sim$ 70 km s^{-1} Mpc^{-1}. It was also found that the redshifted galaxies were also moving away from each other, not just from us the observers in the Milky Way, leading to astronomers proposing the hypothesis of an expansion of the universe.

So the universe is and has been expanding and the density of objects would have been greater in the past due to the smaller volume. To reduce the difference in the density of objects between some time in the past z_0 and a different time say $z_0 + dz$ astronomers have defined the "comoving space density". The comoving space density is the number of objects per Mpc^3 at a redshift z divided by $(1+z)^3$ (Hogg, 1999), scaling the density down to a value it would have today (at redshift $z = 0$). The comoving space density of a constant number of non evolving objects will not change with z, and thus the expansion of the universe, but a change in this density implies that one of the following must be true: the number of objects is varying, the objects are evolving or both.

Large surveys have proved to be powerful tools (Norris et al., 2006) providing large datasets allowing statistical analyses on a larger and larger number of objects. To study cosmology and extragalactic astronomy it is required to go deeper (higher redshift) and wider (survey area), requiring major time allocations on instruments. To investigate this change in object comoving density with morphology, astrophysics uses the Luminosity Function (LF), defined to be the number of objects in a sample that have an absolute magnitude between M and $M + dM$. The LF for any class of object per

[2]The cosmological constant Ω_Λ term was in the past taken to be zero, but more recently it has been realised to be non-zero (Schmidt et al., 1998) and has an important effect on modelling the geometry of the universe

Mpc^3 for a range of redshifts will indicate if there are evolutionary effects for the different classes of objects such as quasars, active galactic nuclei (AGN) and star-forming galaxies (SFG) in relation to the age of universe

In the 1990's a series of radio surveys commenced, producing a significant increase in the total number of known radio sources. For example the NRAO VLA Sky Survey (NVSS) surveyed the whole northern sky north of $\delta = -40^\circ$ at 1.4 GHz between 1993 and 1996, producing a catalogue of about 1.8 million sources above $S \approx 2.5$ mJy at a resolution of $\theta = 45$". NVSS is still the largest ever radio survey, and its survey paper (Condon et al., 1998) is the second most highly-cited paper in radio astronomy.

In addition there was the complementary Faint Images of the Radio Sky at Twenty-Centimeters (FIRST) again using the NRAO VLA (Becker et al., 1995) which surveyed a smaller area than NVSS but still $10,000$ square degrees between 1993 and 2004 but with a higher resolution (1.8" pixel size) and greater sensitivity (1 mJy) to yield a catalogue of over 800,000 sources. The area observed was designed to coincide with the Palomar Observatory Sky Survey around the North and South Galactic Caps. It found many sources that are not present in the NVSS catalogue, but is insensitive to some extended NVSS sources.

The above two surveys covered the northern hemisphere due to the location of the instrument; a corresponding southern hemisphere survey was provided by the Sydney University Molonglo Sky Survey (SUMMS) ((Bock et al., 1999) and (Mauch et al., 2003)). This was a radio imaging survey of the sky south of declination $\delta = -30^\circ$ with a total area of 8100 square degrees to a sensitivity between 6 and 10 mJy (depending on declination) containing $211,064$ sources.

In Figure 1.7 is shown the historical increase in the number of radio sources detected as surveys went deeper and wider. With current instruments we are approaching the limit due to the available finite resource of time as demonstrated on Figure 1.8 where the dashed line corresponds roughly to a few months of observing time.

In 2007 the Cosmic Evolution Survey (COSMOS) (Scoville et al., 2007) was designed to probe the evolution of galaxies over the redshift range $0.5 < z < 6$ and was based

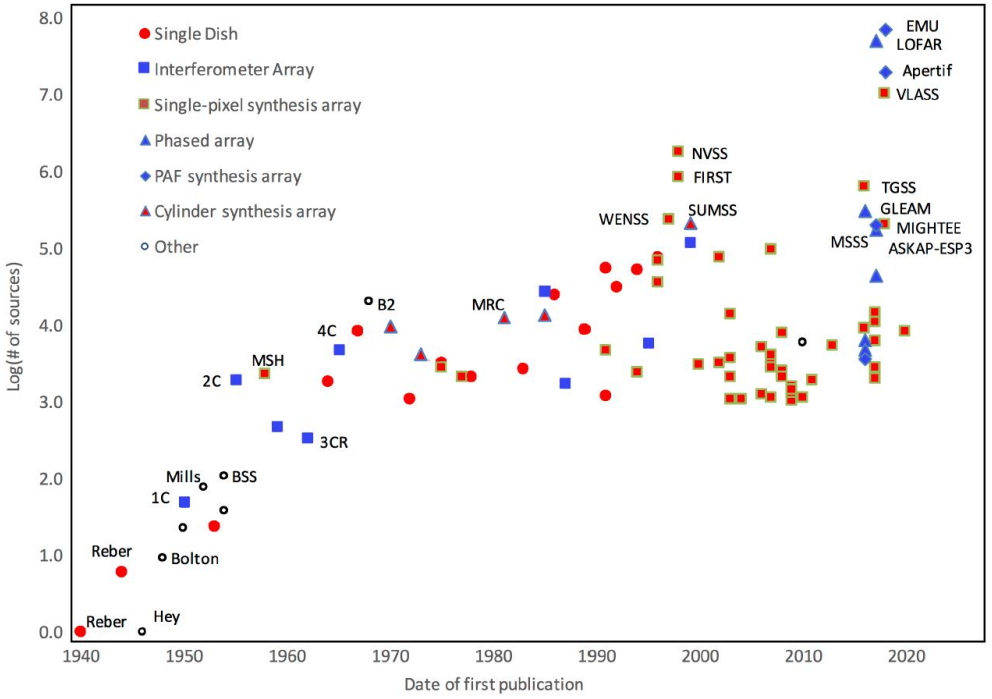


FIGURE 1.7: A plot of the number of known extragalactic radio sources discovered by surveys as a function of time. Credit: (Norris, 2017)

on the Hubble Space Telescope imaging of a 2 square degree area centered at RA 10:00:28.6, Dec +02:12:21.00 (J2000), this area was also observed at other wavelengths from X-ray to Radio. Just in the radio domain various instruments have observed for $\approx$ 1900 hours alone across multiple bands from 244 MHz to 35 GHz. Over 2 million galaxies have been detected providing a huge multi-wavelength data set allowing detailed investigation and analysis of galaxy formation and evolution, producing over 200 papers in scientific journals (Caltech, 2015).

More recently radio surveys have gone deeper but due to the time constraint they have not covered the same large areas. For example the Australia Telescope Large Area Survey (ATLAS) was a project to image a smaller area 7 square degrees at 1.4 GHz but to a sensitivity of 15 μJy for CDFS (Norris et al., 2007) using 173 hours of integration (21 pointing centers of 8.2 hours each) and ELAIS (Middelberg et al., 2007) using 231 hours of integration with a resolution of $\approx$ 11"$\times$5" these resulted in identifying 2150 sources for data release 1. This increase in sensitivity has resulted in surveys going from finding mainly AGN (which are very radio bright even at high redshift) as in the earlier surveys to now finding star forming galaxies (SFG) out to $z \approx 1$, thus opening up a new

area of parameter space. There is some overlap between FIRST and ATLAS in area and the *Spitzer* Wide-area InfraRed Extragalactic survey (SWIRE Xu et al. 2002a), thus providing the multi-wavelength, resolution and depth data required for more complex analysis as (e.g. Mao et al. 2012).

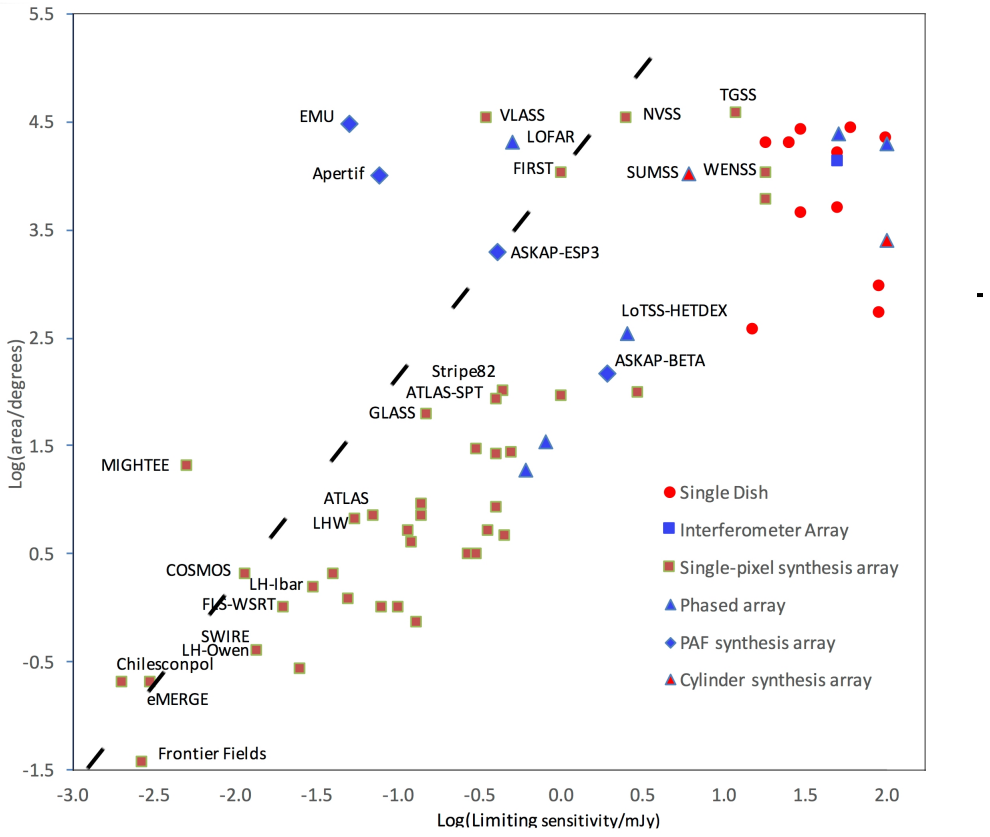


FIGURE 1.8: The sky area vs sensitivity of modern radio surveys. The dashed line marks the boundary of existing surveys. Credit: (Norris, 2017)

With the new instruments being commissioned such as the Australian Square Kilometre Array pathfinder (ASKAP) (Johnston et al., 2007) and MeerKAT (Jonas, 2009) moving into this new space of very large area and high sensitivity shown in Figure 1.8 (reproduced from (Norris, 2017) which provides a summary review of Radio Surveys to-date). For these new even larger surveys the number of objects identified will increase to the millions. One such future survey is the Evolutionary Map of the Universe[3] (EMU) (Norris et al., 2011), which will be a radio sky survey at $1.2 - 1.4$ GHz and a resolution of ≈ 10" using the new ASKAP antenna array to make a deep radio survey of the Southern

[3]There are similarities between ATLAS and EMU in terms of frequency, sensitivity and resolution, and the ATLAS data is being used to test many of the technical and scientific processes for EMU.

Sky. Due to size of the survey area (entire sky south of $+30^\circ$ declination) and depth a very large number of new sources will be identified, estimated at about 70 million (Norris et al., 2011). Deep in this context means that the survey will try to achieve a sensitivity $\sim$10 μJy rms or better (slightly better than ATLAS) and as a result probe sources such as SFGs to a redshift of $z \approx 1$, powerful star bursts to even higher redshifts and AGNs to the edge of the visible universe.

To identify the object types and provide their redshift, data from other wavelengths in different surveys, catalogues or follow up observations such as the ATLAS spare fibre program for OzDES (D'Andrea and OzDES, 2014) are required. Having this data we are then faced with the dilemma of how to cross identify the $\sim$ 70 million sources with these surveys or follow up observations in other wavelengths in an automated pipeline. In addition will be the problem of how to manage the resultant large catalogue for EMU data users to conduct further science mining the data.

So there now exist low resolution radio images with corresponding high resolution optical images and it is not immediately obvious which the true host is; this is demonstrated in the Figure 1.9 [4].

[4] I have made extensive use of SAOImageDS9 (Joye and Mandel, 2003) for generating these postage stamp images and some of the larger survey images here and in latter chapters

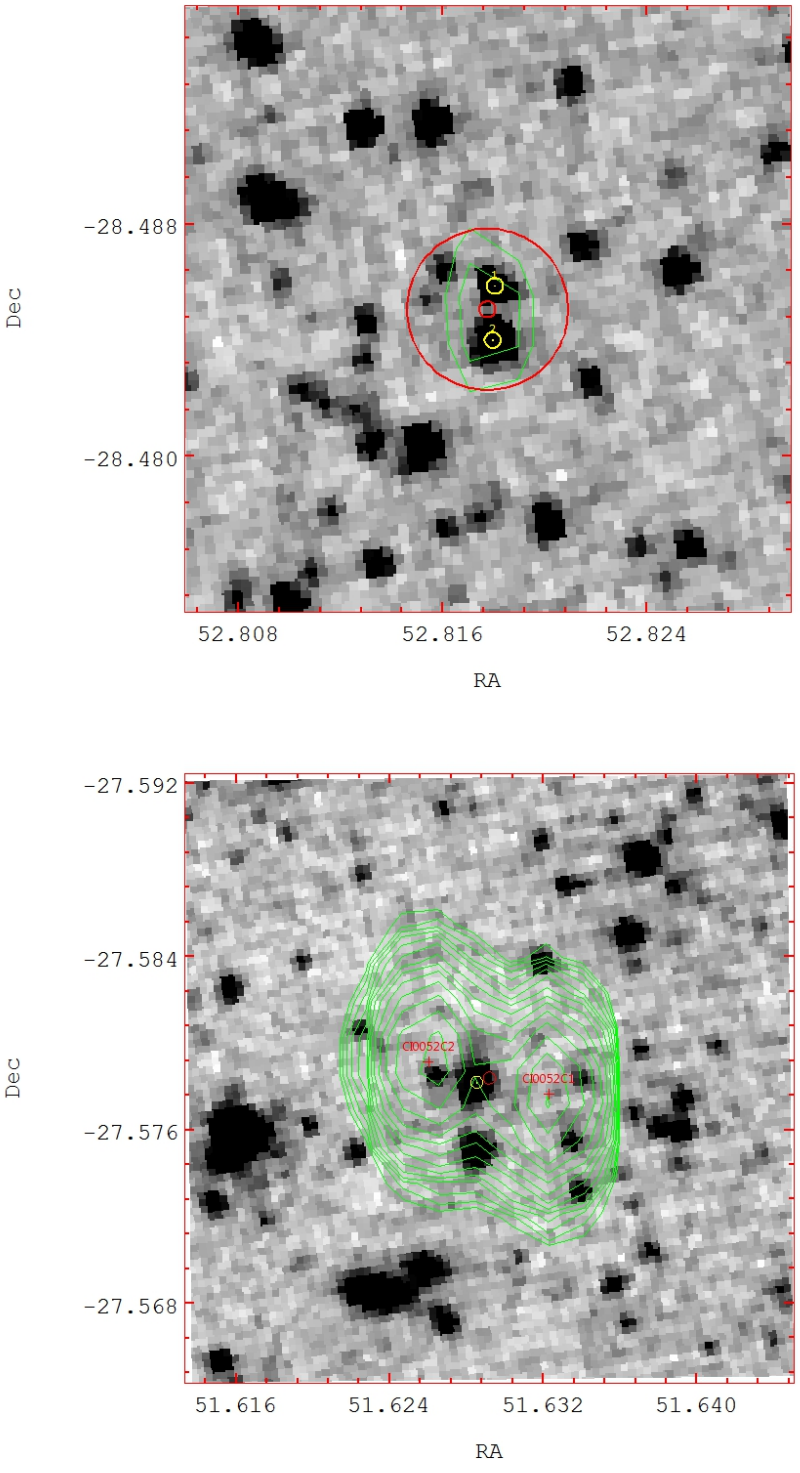


FIGURE 1.9: Example of low resolution radio contours from ATLAS overlaid onto the higher resolution background infrared image. In the top image there is a compact radio source CI0765 (green contours) with two infrared candidates within the 10" search radius marked by open yellow circles. The bottom image is a more complex radio double source CI0052 with the two component centers marked by red crosses with the radio flux weighted mean between these marked by a red circle and a possible IR candidate close to this marked by a yellow circle.

As we move into this era of very large multi-wavelength surveys, the task of cross-matching millions of sources between the catalogues is well beyond a human manual visual method. Computer algorithms are required to automate this process. In addition some method of database storage such as a Relational Database Management System (RDBMS) with data mining algorithms is also necessary. Such environments used for data analysis and reporting are known as a data warehouse. These are commonly used in the commercial world and we ask the question, "can commercial disciplines usefully be brought to bear on this matter?"

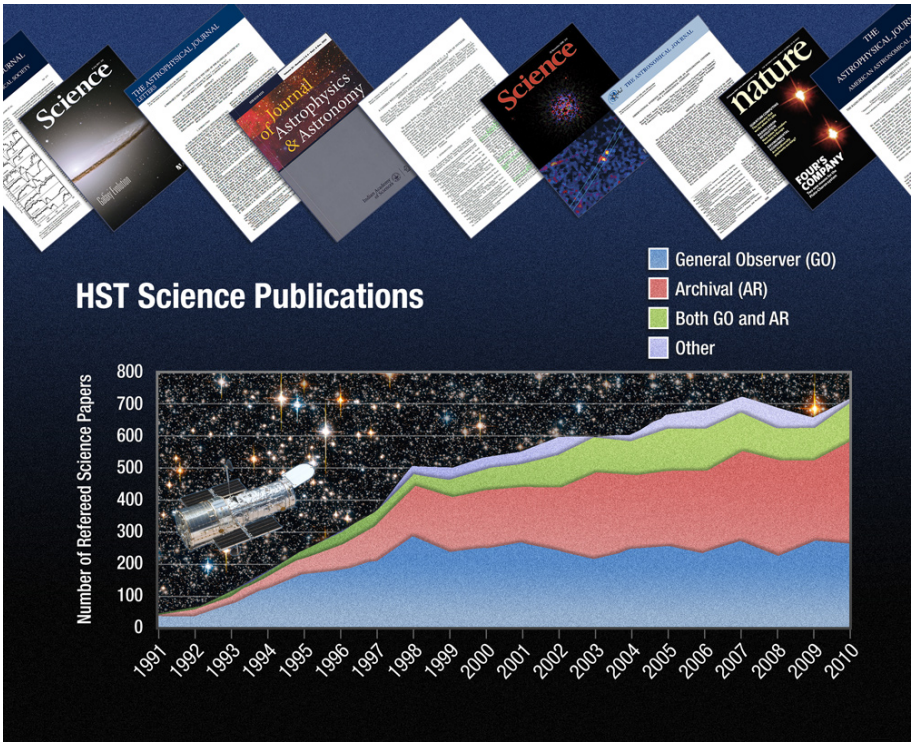


FIGURE 1.10: HST Science Publications. Credit: NASA, ESA, and A. Feild (STScI) (Villard, 2011)

As the next generation of radio surveys probe unexplored areas of observational parameter space and generate very large archives of data. We are moving into a realm of data driven study and research requiring new tools and algorithms. I will end this section by looking at publications from the Hubble Space Telescope (HST) as shown in Figure 1.10, where we see the growth in publications from re-processing of archived data. The number of science papers written based on Hubble archived data has increased to the point where it has eclipsed the number of papers resulting from new observations, as the time available on the HST is a finite resource. What will this mean for ASKAP,

MeerKAT and other instruments and the forthcoming SKA. We are entering an era of very large and complex data mining for years to come by astrophysicists and cosmologists. What discoveries will be made with this avalanche of data ?

1.4 Thesis Outline

This thesis focuses on the development of a new algorithm based on the Likelihood Ratio with some extensions to assist in identifying more complex radio and infrared sources. From the cross identifications obtained, I then in some detail investigate the radio and infrared properties of these galaxies. The thesis is organised as follows:

Chapter 2 describes the several sets of survey data used in this thesis with an overview of some of the more commonly used cross-identification methods. I conclude with a short section of some of the new methods and algorithms being developed and applied.

Chapter 3 I describe in some detail the algorithm developed to implement the Likelihood Ratio. I also describe the computational method employed to process the algorithm using Python and MySQL.

Chapter 4 In this chapter I undertake a comparison of the methods employed in this work. I compare a nearest neighbour approach to the Likelihood Ratio, in addition comparing Poisson Probability with the Likelihood Ratio.

Chapter 5 I look at refinements to the Likelihood Ratio in identifying radio doubles (multiple radio sources) being associated with one infrared source. We also extend previous work in extracting what we have called "Infrared Doubles" (ID's), these being two or more infrared galaxies in a local cluster (with very similar z) resulting in one source of radio emission.

Chapter 6 Taking the results from the previous chapters I investigate the infra-red and radio properties of the cross-identified (matched) galaxies.

Chapter 7 describes the conclusions of this work and presented in this thesis with a look at future extensions and applications.

The Chandra Deep Field South, observed in the U-, B-, and R-bands with ESO"s VIMOS and WFI instruments. The U-band VIMOS observations were made over a period of 40 hours and constitute the deepest image ever taken from the ground in the U-band. The image covers a region of 14.1 x 21.6 arcminutes on the sky and shows galaxies that are 1 billion times fainter than can be seen by the unaided eye. The VIMOS R-band image was assembled by the ESO/GOODS team from archival data, while the WFI B-band image was produced by the GABODS team. Credit:ESO/ Mario Nonino, Piero Rosati and the ESO GOODS Team

Chapter 2

Background

This chapter introduces and explains the different wavelength surveys used later in this thesis for multi-wavelength catalogue cross identification. There are two radio and one infrared catalogues, with a follow up optical survey to the radio survey providing optical spectroscopic data for obtaining photometric redshifts for the radio galaxies. In addition various methods used for matching objects between different wavelength surveys are introduced.

2.1 The Survey Data

In this section the different surveys and catalogues used for this thesis are introduced starting with the ATLAS radio survey and the Data Fusion *Spitzer* catalogue. The cross-identification method investigated and the later analysis is focused on the ATLAS DR3 catalogue and the Fusion *Spitzer* catalogue. This is followed by an overview of the other surveys also used in this thesis to provide additional data for investigation and confirming hypothesis to be applied for the cross-identified sources, OzDES provides redshifts and FIRST a comparison and calibration data set being shallower but much larger than ATLAS.

In Figure 2.1 I show how these different wavelength surveys cover the sky. The areas marked in grey show the coverage of FIRST, the blue area show the OzDES coverage

and the areas marked in red are from Fusion *Spitzer* and show where the CDFS (RA $\approx 53^\circ$ and Dec $\approx -28^\circ$) and ELAIS (RA $\approx 9^\circ$ and Dec $\approx -43^\circ$) fields lie. At this scale it is not possible to show how ATLAS, Fusion *Spitzer* and OzDES overlap within the CDFS and ELAIS fields. In later sections of this chapter more detailed maps of these areas are provided to show the survey overlap. In conjunction with Figure 2.1 is Table 2.1 giving a comparison of the different Radio Surveys used, and with the future EMU radio survey which has similarities with ATLAS in terms of frequency, sensitivity and resolution, as such the ATLAS data is being used to test many of the technical and scientific processes for EMU.

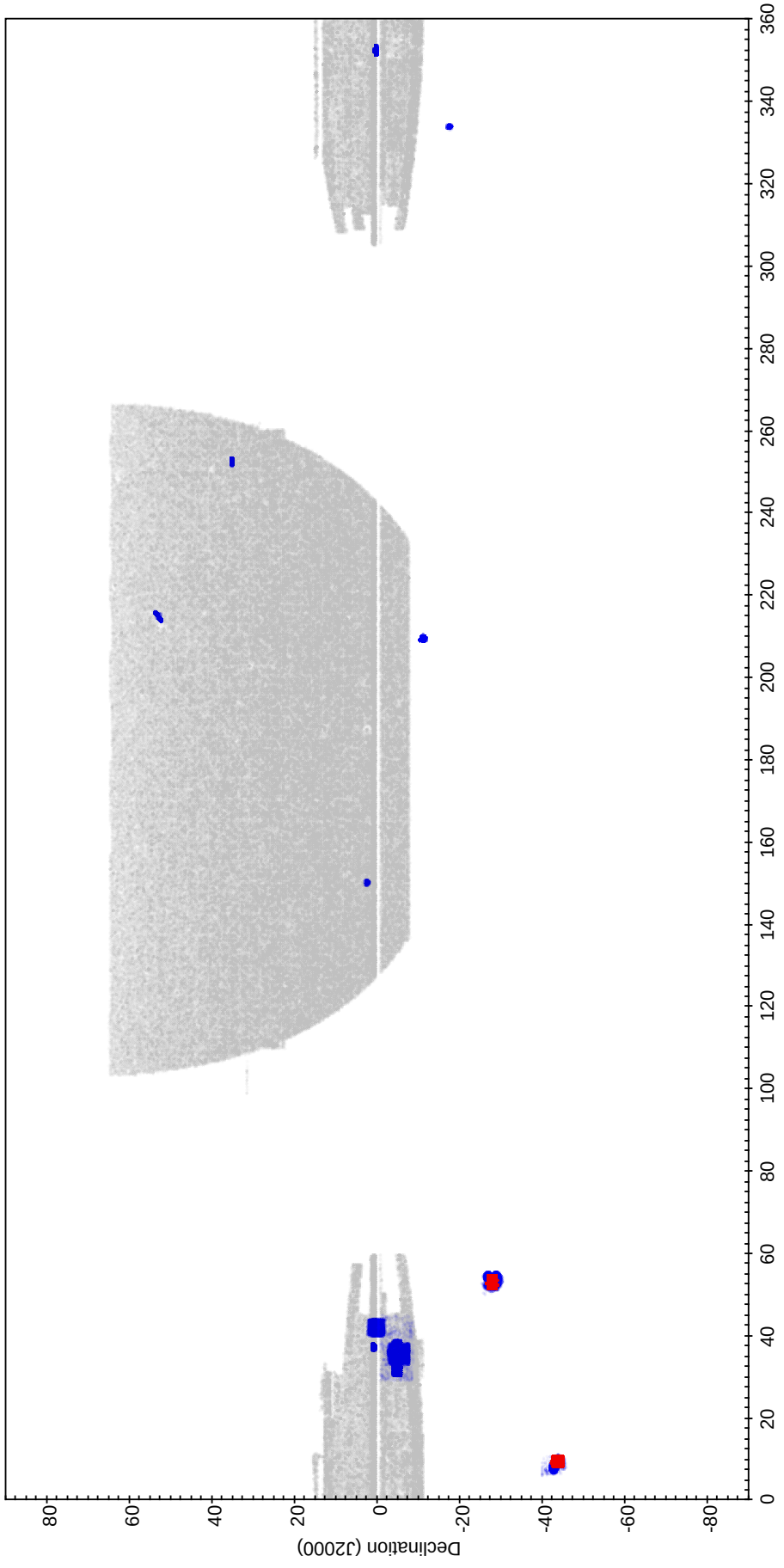


FIGURE 2.1: Whole Sky map showing the coverage of the surveys used in this thesis. The grey area marks FIRST, the blue area marks the OzDES fields and the red is for the Fusion *Spitzer* fields.

TABLE 2.1: Radio Survey Comparison

Survey	Area	Sensitivity	Resolution	Frequency
FIRST	10^4 deg^2	0.13 mJy	5”	1.4 GHz
ATLAS-CDFS	≈ 3.5 deg^2	65 μJy	17×7”	1.4 GHz
ATLAS-ELAIS	≈ 3.5 deg^2	75 μJy	12×8”	1.4 GHz
EMU	entire sky south of $+30^\circ$ Dec	10 μJy	10”	1.4 GHz

2.1.1 The ATLAS DR3 Radio Survey and Catalogue

The ATLAS survey was completed with the Australia Telescope Compact Array (ATCA) between June 2009 and June 2010 June and covers $1.3 - 1.8$ GHz, over an area coinciding with the *Chandra* Deep Field South (CDFS) survey obtained from the X-ray *Chandra* satellite between 1999 and 2000 (Giacconi et al., 2001) and the European Large Area ISO Survey - South 1 (ELAIS-S1 or ES1) (Rowan Robinson et al., 1999).

The ATLAS survey observed an area of the sky coinciding with the Chandra Deep Field South (CDFS) survey obtained from the X-ray Chandra satellite between 1999 and 2000 (Giacconi et al., 2001) and the European Large Area ISO Survey - South 1 (ELAIS-S1 or ES1) (Rowan-Robinson et al., 2010). These areas have also been observed with the *Spitzer* Wide-area InfraRed Extragalactic survey (SWIRE) (Xu et al., 2002b), thus providing the multi-wavelength data required for the further work by ATLAS team members producing such analysis as Mao et al. (2012).

The ATLAS DR3 component source catalogue presented in Franzen et al. (2015) has found a total of 5146 radio source components down to 5σ over both fields. There are 3079 source components down to 65 μJy in CDFS and 2067 source components down to 75 μJy in ELAIS S1. The restoring beam for the two ATLAS fields is given in Table 2.2.

TABLE 2.2: Radio Restoring Beam for each ATLAS field

Field	Major Axis (arcsec)	Minor Axis (arcsec)	Position Angle (degrees)
CDFS	16.8	6.9	1.0
ELAIS	12.2	7.6	-11.0

It should be noted that Middelberg et al. (2007) identified a positional offset between the ATLAS ELAIS field and catalogued SWIRE counterparts, mean offset (0.08 ± 0.03)" in right ascension and (0.06 ± 0.03)" in declination for ATLAS DR1; these have been corrected for in the following work. Also the initial work was based on a visual cross-identification between the surveys by very experienced members of the ATLAS team.

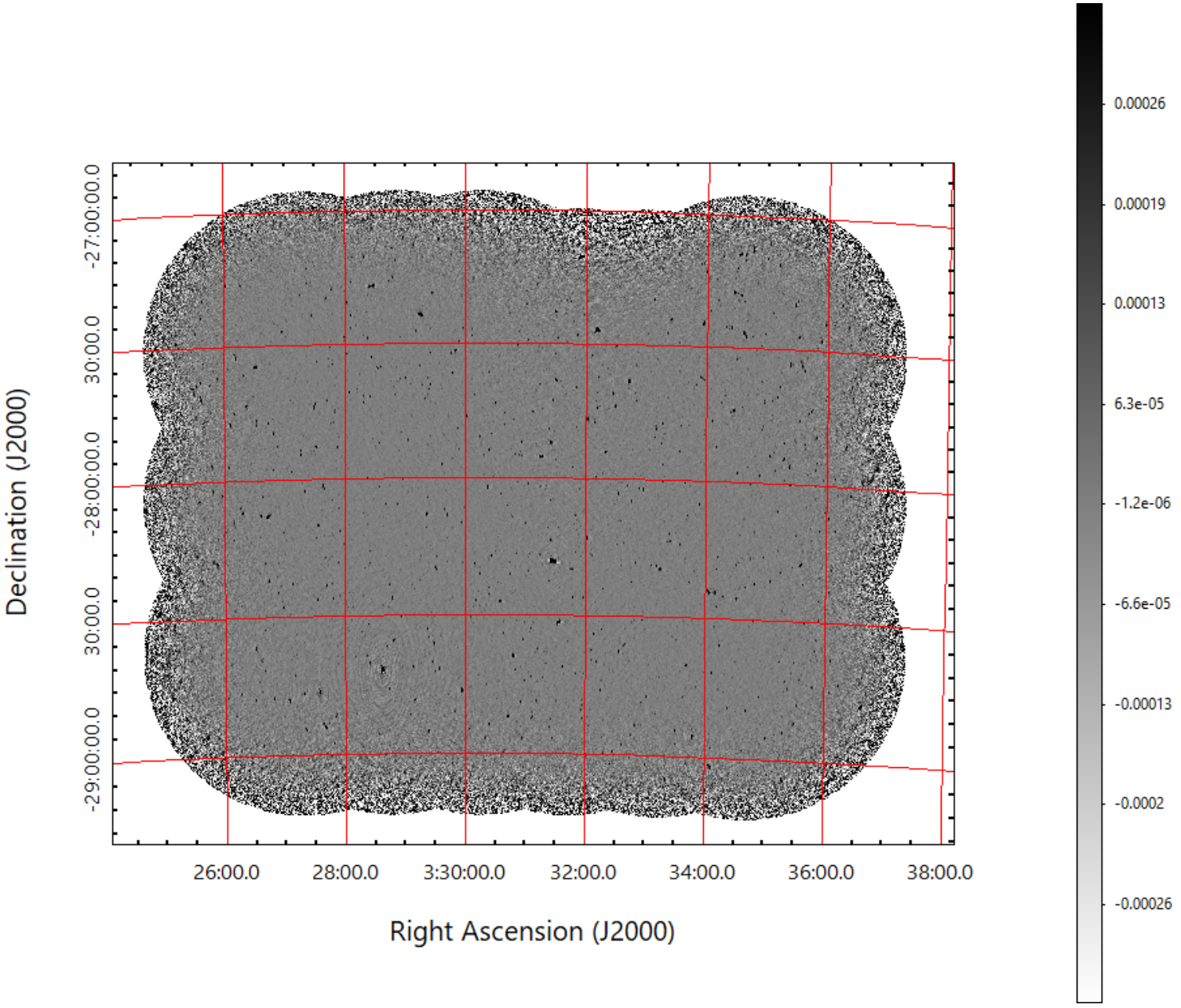


FIGURE 2.2: An inverse greyscale of ATLAS DR3 radio image for the CDFS field. On the right hand side is the greyscale colour bar for the image pixel values in mJy/Beam.

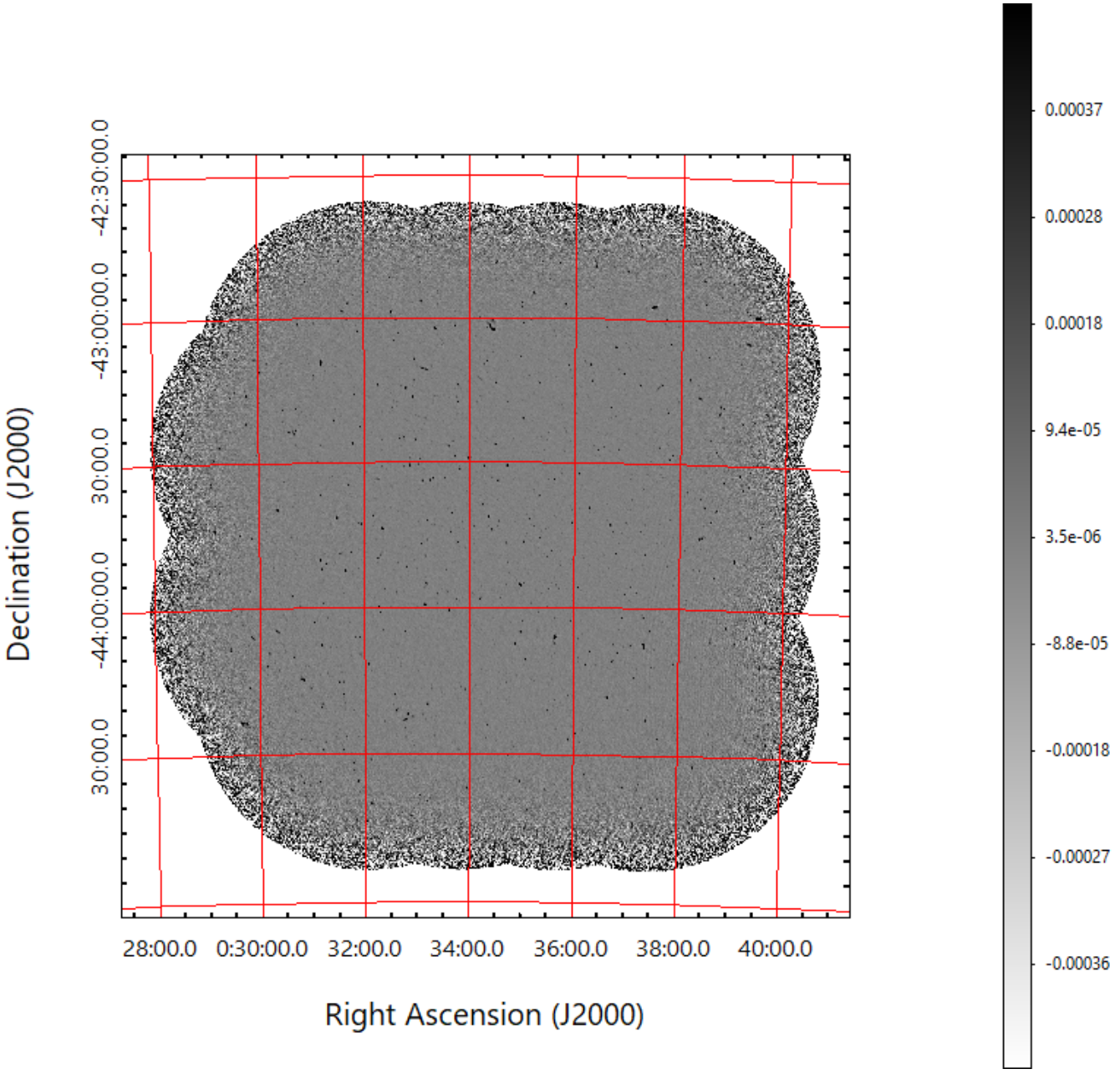


FIGURE 2.3: An inverse greyscale of ATLAS DR3 radio image for the ELAIS field. On the right hand side is the greyscale colour bar for the image pixel values in mJy/Beam.

2.1.2 The Fusion *Spitzer* UV to Mid-IR Catalogue

The *Spitzer* Space Telescope had several large programs during its first year of operation, one of which was the *Spitzer* Wide-area InfraRed Extragalactic survey (SWIRE) (Xu et al., 2002b). The SWIRE survey imaged ≈ 50 square degrees divided among six different directions on the sky, detecting over two million galaxies. A key science goal was to study galaxy evolution for the space $0.5 < z < 3$ in the following IRAC (Infrared Array Camera covering the mid-infrared) Infrared wavelengths $3.6, 4.5, 5.8$ and $8.0\,\mu$m, also the MIPS (far-infrared camera) $24, 70$ and $160\,\mu$m. Early galaxies are shrouded in dust which blocks optical and ultraviolet light; infrared and radio are not obscured by dust so these infrared and Radio wavelengths are important tools to see through this dust.

The Fusion *Spitzer* catalogue is a multi-wavelength far-UV to mid/far-IR catalogue of *Spitzer* selected sources, hereafter referred to as the "Fusion" catalogue (Vaccari et al., 2010a; Vaccari, 2015), which has coverage of the CDFS and ES1 fields thus overlapping the ATLAS survey. This catalogue is based on detections at $3.6\,\mu$m with the IRAC instrument (Fazio et al., 2004) on board the *Spitzer* Space Telescope (Werner et al., 2004), down to a flux density of $4.6\,\mu$Jy in the CDFS field and $4.8\,\mu$Jy in the ELAIS-S1 field. There are 391,518 *Spitzer* IRAC sources in ELAIS-S1 (see Figure 2.4) and 462,638 in CDFS (see Figure 2.5). We note that this catalogue contains very few photometric and spectroscopic redshifts pertaining to radio sources, but the Herschel Extragalactic Legacy Project (HELP) Vaccari (2016) is in the process of putting together multi-wavelength data, compute photometric redshifts and physical parameters for sources in ATLAS (and ASKAP/EMU Early Science) fields.

The cross identification undertaken using this Fusion catalogue makes use of the IRAC $3.6\,\mu$m flux density, as this is the most complete set of infrared flux values (93% of the catalogue entries had $3.6\,\mu$m flux density values, where as 63% of the entries have $4.5\,\mu$m flux values and the longer wavelengths are even more incomplete) in the catalogue.

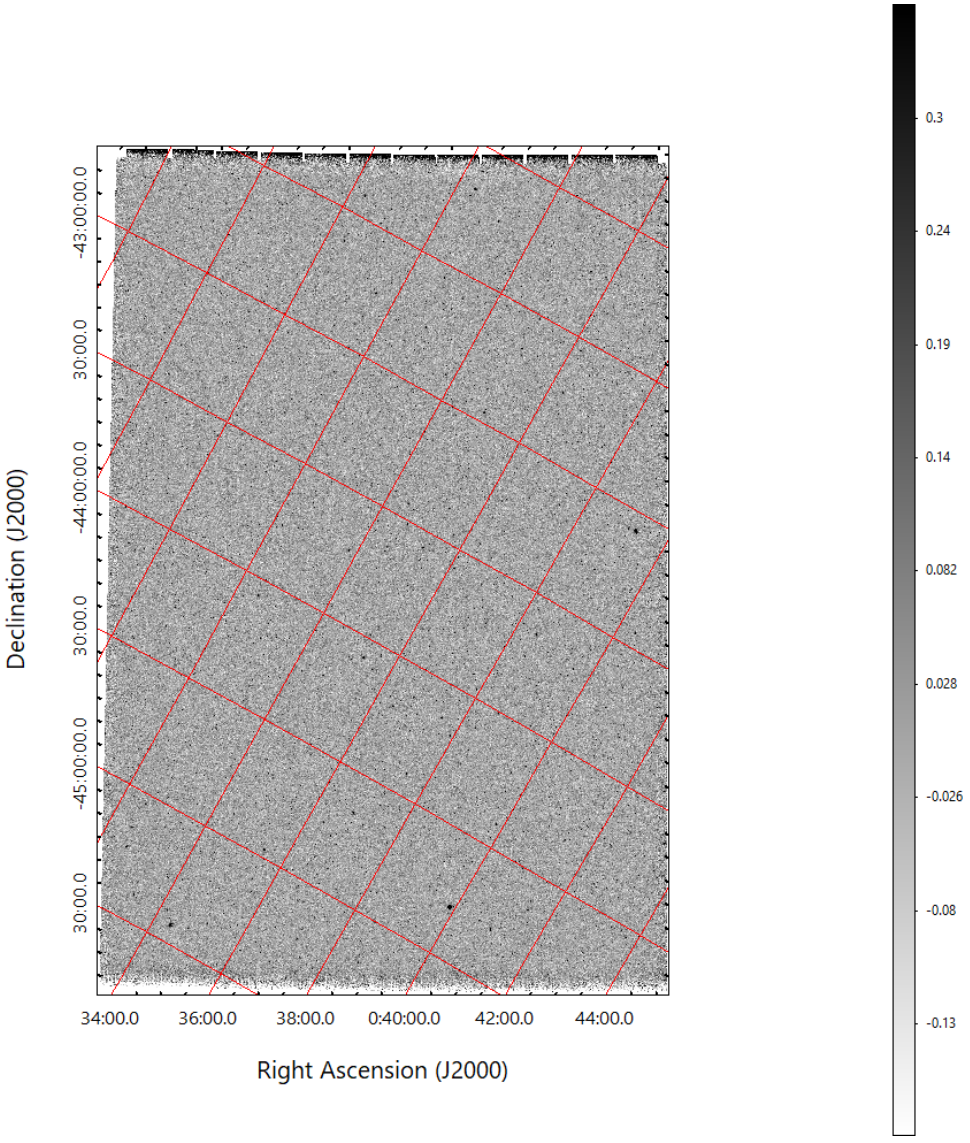


FIGURE 2.4: An inverse greyscale of the SWIRE IRAC Band 1 Infrared image for the ELAIS field. On the right hand side is the greyscale colour bar for the image pixel values in Jy.

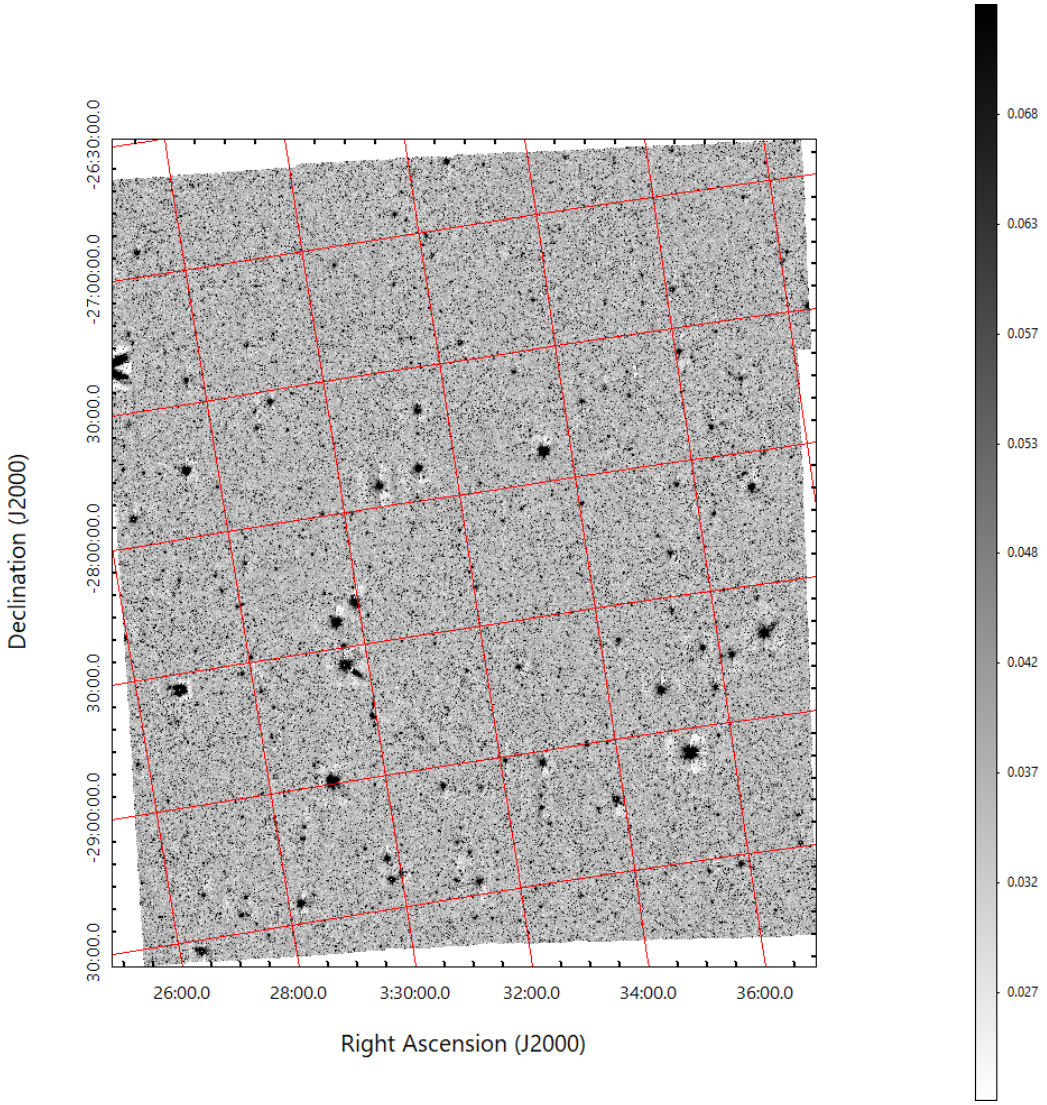


FIGURE 2.5: An inverse greyscale of the SWIRE IRAC Band 1 Infrared image for the CDFS field. On the right hand side is the greyscale colour bar for the image pixel values in Jy.

2.1.3 The OzDES Survey

The Australian Dark Energy Survey (OzDES, 2016) using the 4m Anglo-Australian Telescope started in 2013 with the goal of measuring the redshifts of tens of thousands of galaxies. Using a multi-fibre 2D positioner, 400 individual fibres can be aligned with known positions for galaxies within the field of view for a single pointing. These fibres are then fed back to a spectrograph to obtain very detailed spectra of the galaxies that the fibres align with, thus allowing precise redshift values (z) to be obtained for specific galaxies. Some of the fields to be observed with OzDES also correspond with the CDFS and ES1 of ATLAS. The spectroscopic redshifts obtained by OzDES is also being used to provide an important source of calibration data for photometric redshifts.

Spectroscopic redshifts are obtained by observing a frequency (or wavelength) within which known spectral lines exist, and then measuring the shift in frequency (or wavelength) of these lines to their laboratory positions. A photometric redshift (Baum, 1962) is an estimate for the recession velocity of an astronomical object without measuring its spectrum, and is less reliable than the spectroscopic method in part due to the assumptions made about the nature of the spectrum for the astronomical source. With the large sky surveys conducted from the late 1990s photometric redshift was adopted as it was not possible to get sufficient telescope time to observe spectrosopic redshifts for all the objects. But with the development of multi-fibre devices allowed multiple fibres to be positioned on many objects with in the field of view. An example of this is the AAOmega spectrograph with the Two Degree Field system (2dF) (Smith et al., 2004) which allows the aquisition of upto 392 simultaneous spectra of objects within a two degree field. As a result of this technical development the issue of telescope time to obtain individual spectrographic redshifts of many objects has been minimised, and the obtaining of accurate redshifts.

A proposal (Norris et al., 2013) was submitted as part of the OzDES survey to use some of the AAOmega spare fibres to conduct a spectroscopic follow up survey of the radio-detected galaxies in the ATLAS survey. The AAOmega spectrograph has spare fibres for redundancy, if these are not required for the primary project then others can use them.

The ATLAS project requested to use about 30 fibres to get redshifts and spectroscopic typing of the ATLAS radio sources.

This thesis uses the 2016-02-25 version of the Global Redshift Catalogue which includes OzDES and literature spectra in the DES deep fields. Use has been made of some initial data released by OzDES (Yuan et al., 2015) with some preliminary redshifts available to try and match the XID's in this thesis between the radio and infrared sources and increase the number of redshifts available from Fusion.

2.1.4 The FIRST Survey

The FIRST survey (Becker et al., 1994) was a systematic survey of the northern sky at 20 cm wavelength using the NRAO Very Large Array (VLA), it covered over $10,000$ square degrees, achieving images with 1.8" pixel size, a typical RMS of 0.15 mJy and a resolution of 5". The final catalogue has $946,432$ sources providing a statistically large number of sources in the catalogue (White, 2019). The survey area also coincides with the Sloan Digital Sky Survey (SDSS).

The work of Magliocchetti et al. (1998) used the February 27 1997 data release of the FIRST survey, which contained $236,000$ entries for identifying radio doubles in this survey. I applied the same techniques as Magliocchetti et al. (1998) to reproduce the work and results (See Section 5.2) using the final and much larger FIRST catalogue to verify this was still valid. The results of this were then used to verify several of the hypotheses in this thesis, which are then applied to the resultant catalogue from my work in Weston et al. (2018) from the ATLAS-DR3 cross-identification with the Fusion catalogue for further analysis.

In addition the deep JVLA 1.4 GHz survey (Miller et al., 2013) contains a sub-region of the ATLAS CDFS Field which is used in Section 5.1 for confirmation of the Infrared Doubles.

2.1.5 Catalogue Comparison

In this sub-section I compare the different surveys used in this thesis and their sky coverage.

It is important to note that the sources in the radio catalogues do not fully overlap with Fusion for the CDFS field. All the sources in the ATLAS and Fusion catalogues are overlaid as shown in the following Figures 2.6 and 2.7. While the Fusion catalogue completely covers the ATLAS observations in ELAIS-S1, part of the ATLAS CDFS data is not covered by the Fusion catalogue and so approximately 20 Radio sources are unavailable for cross matching. Hence the analysis is restricted to the following sub-region for CDFS $51.7^\circ \leq \text{RA} \leq 54.2^\circ$ and $-29.0^\circ \leq \text{Dec} \leq -27.2^\circ$. This region has been placed so that it is inside Fusion and 100" from the edge. Also the analysis for ELAIS-S1 is restricted to $7.3^\circ \leq \text{RA} \leq 9.7^\circ$ and $-44.6^\circ \leq \text{Dec} \leq -42.9^\circ$.

The last two figures of this section show the ATLAS DR3 sources overlaid onto the OzDES sources for comparison of the field coverage. In Figure 2.8 the ATLAS CDFS sources are overlaid on the OzDES coverage, then in Figure 2.9 the ATLAS ELAIS sources are overlaid onto OzDES.

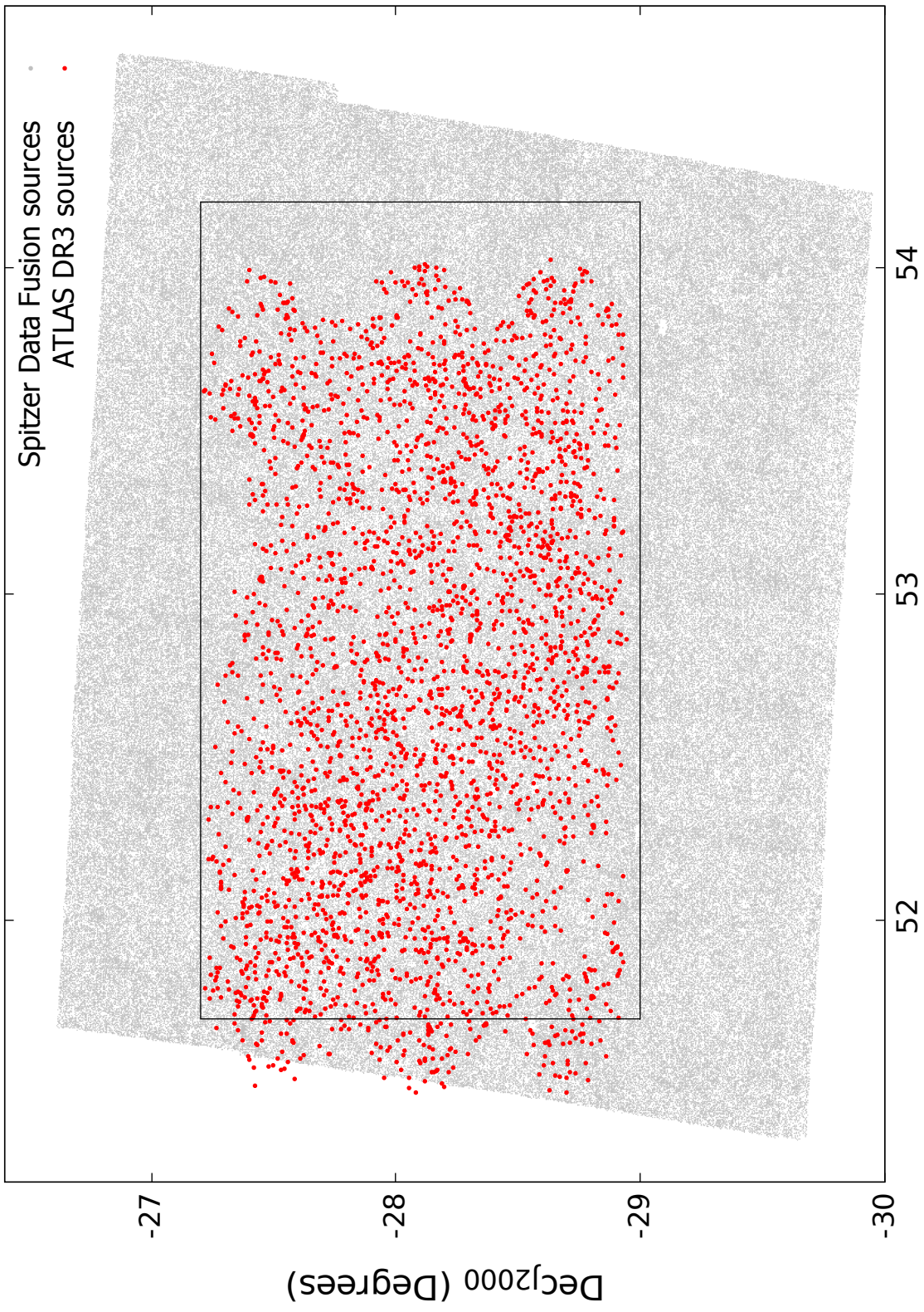


FIGURE 2.6: Sky distribution of ATLAS and Fusion sources in the CDFS field. The grey background dots are the Fusion sources and the larger red foreground dots are the ATLAS sources.

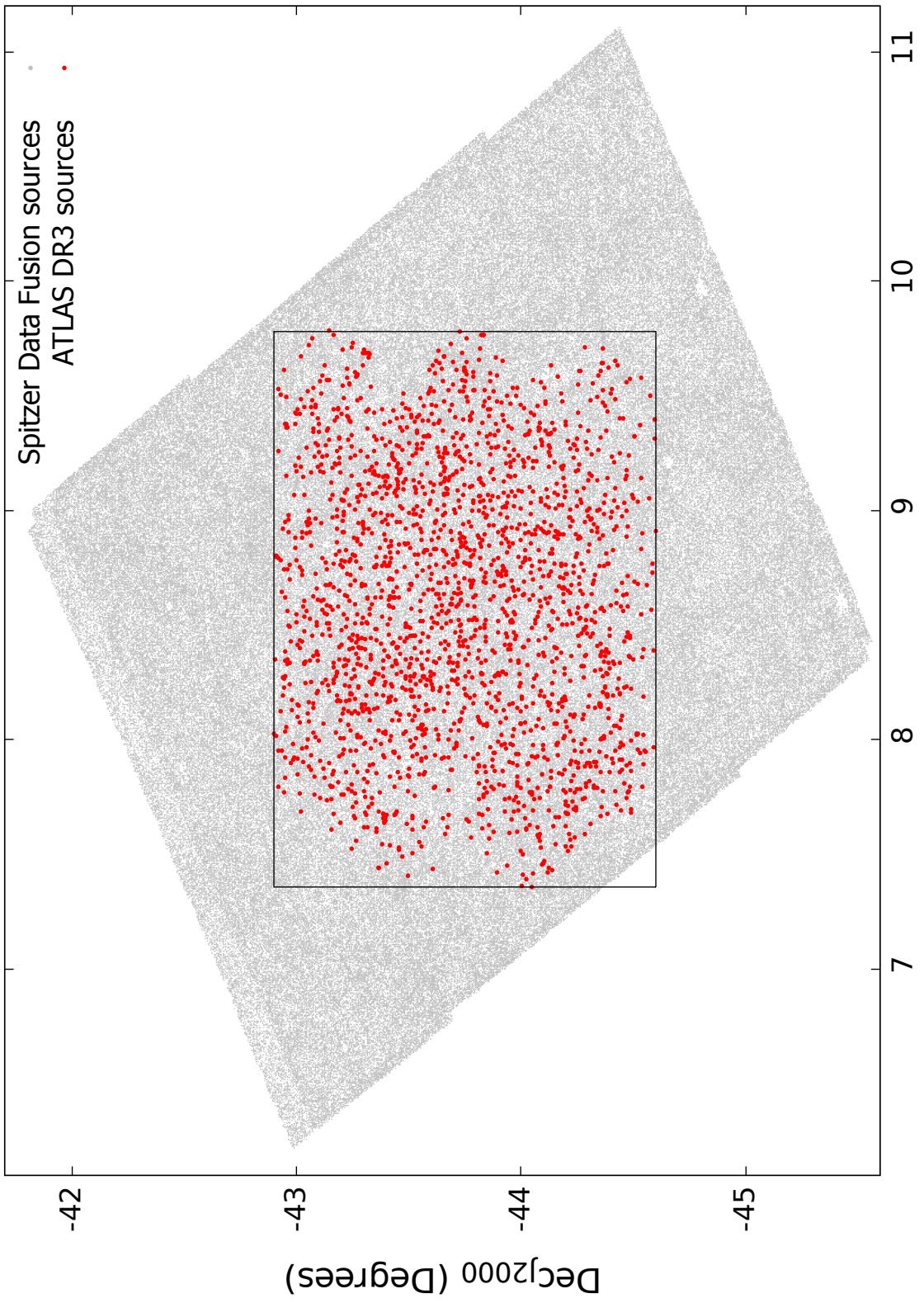


FIGURE 2.7: Sky distribution of ATLAS and Fusion sources in the ELAIS field. The grey background dots are the Fusion sources and the larger red foreground dots are the ATLAS sources.

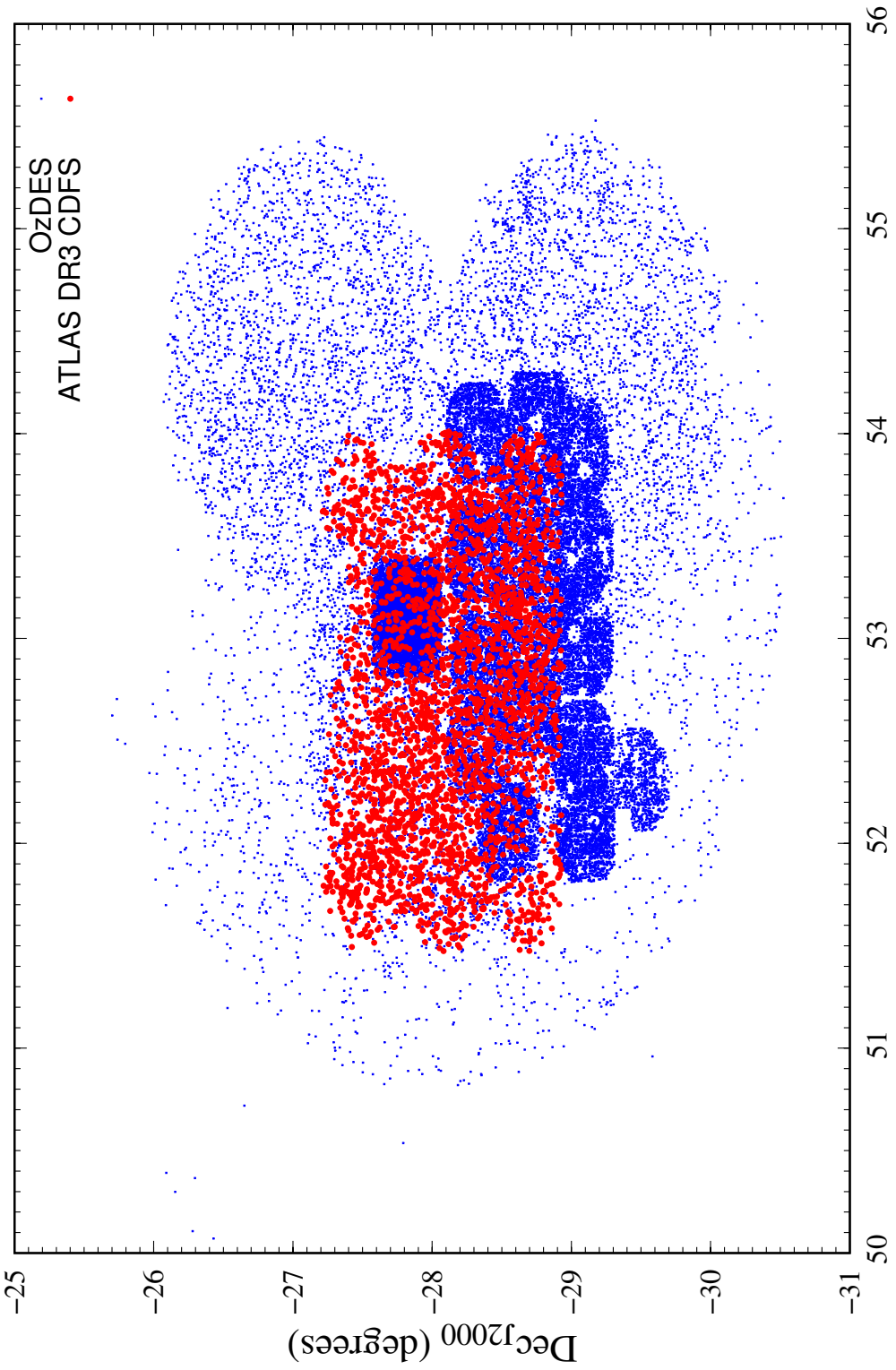


FIGURE 2.8: Sky distribution of ATLAS and OzDES sources in the CDFS field. The blue background dots are the OzDES sources and the larger red foreground dots are the ATLAS DR3 sources.

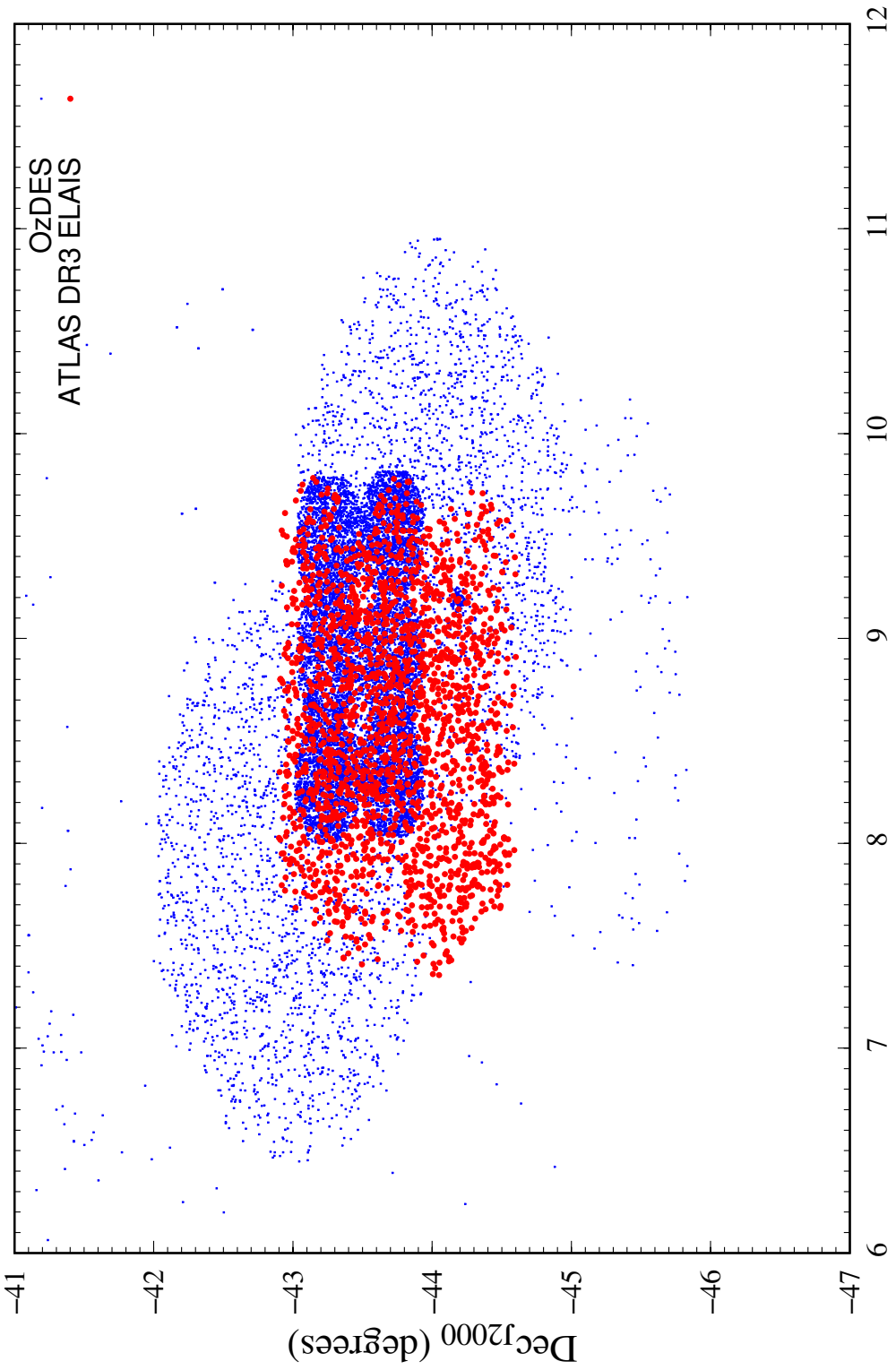


FIGURE 2.9: Sky distribution of ATLAS and OzDES sources in the ELAIS field. The blue background dots are the OzDES sources and the larger red foreground dots are the ATLAS DR3 sources.

2.2 Cross-Identification Methods

Before proceeding to undertake any analysis and science, first the sources from different surveys and catalogues which have used different wavelengths must be matched with a radio equivalent. This matching is a key problem when merging surveys from different wavelengths to determine which sources are associated with one another and which are unrelated (Hardcastle et al., 2010).

The primary focus of this thesis is to investigate a better and more refined way to cross-identify objects between catalogues. In the following three sub-sections are reviewed three different methods used to date. The first method Nearest Neighbour is in common usage and it requires only the coordinates of the sources. The other two use additional information from the catalogues such as flux and source densities to assist in selecting a match.

2.2.1 Nearest Neighbour

The Nearest neighbour search is simply a proximity search. In this case it would be looking for the nearest infrared source to an ATLAS radio source.

To demonstrate this in Figure 2.10 there is a radio source at the center of a 10" search radius. Within the area defined by this radius are three infrared candidates whose centers lie within the search radius, with one infrared source that overlaps the search area but whose center lies outside. The method is to take the three sources whose centers lie within the search area as candidates and calculate their angular separation from the radio source, and then to select the closest which in this case would be the infrared candidate with an angular separation of 5" from the Radio Source. No information other than angular separation is used to select the closest infrared candidate to the radio source.

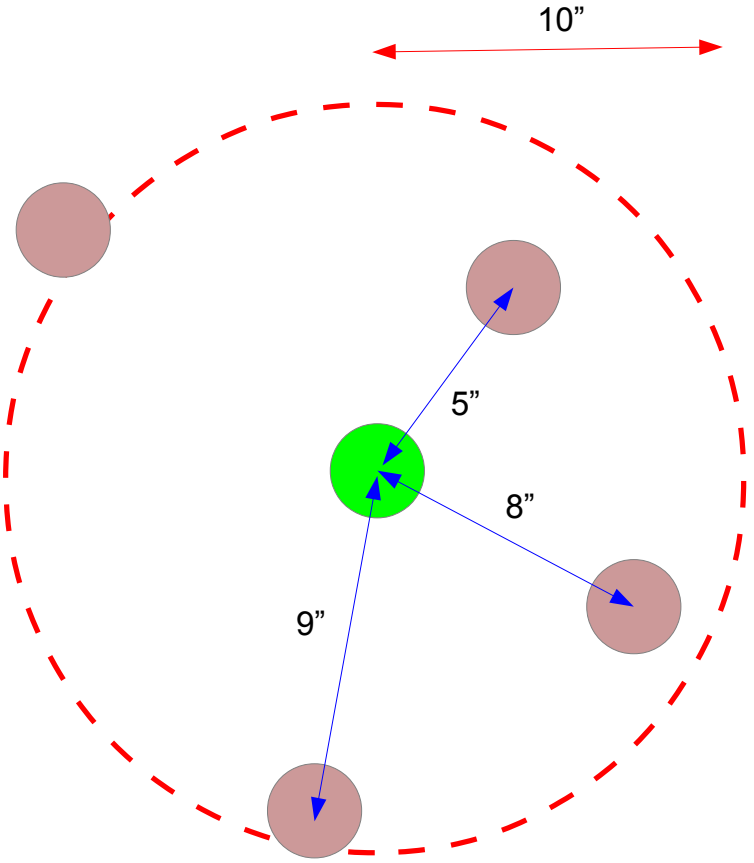


FIGURE 2.10: Nearest Neighbour : The green filled circle is the radio source, the light red circles are close-by infrared sources. The red dotted line denotes the outer edge of the 10” radius search area centered on the radio source.

As the angular distances are very small in this case a maximum of 10" which is $\approx$ 0.0028 of a degree, it is possible to approximate the sky to a two dimensional plane allowing the use of Pythagoras Theorem. Therefore for angular separation $\Delta\theta$ we get:

$$\begin{aligned} \Delta\text{RA} &= \text{RA}_{\text{radio}} - \text{RA}_{\text{offset}} - \text{RA}_{\text{IR}} \\ \Delta\text{Dec} &= \text{Dec}_{\text{radio}} - \text{Dec}_{\text{offset}} - \text{Dec}_{\text{IR}} \\ \Delta\theta &= \sqrt{(\Delta\text{RA} \times \cos(\text{Dec}))^2 + (\Delta\text{Dec})^2} \end{aligned} \tag{2.1}$$

There are two additional terms due to the Radio and Infrared catalogues having coordinate offsets $\text{RA}_{\text{offset}}$ and $\text{Dec}_{\text{offset}}$.

2.2.2 Poisson Probability

A more sophisticated method than a nearest neighbour approach is the Poisson probability (PP) as described and used by Downes et al. (1986), who applied it to a small dataset of 188 radio sources which were cross identified against optical images. They applied a search radius about the radio source position and then determined the probability that an optical candidate within the search area is the correct identification.

The Poisson probability method differs to a nearest neighbour approach by using additional information available from the catalogues such as stellar magnitude or radio flux, redshift and surface density of objects. This is done by estimating the probability that a given source is unassociated: Given a candidate counterpart with magnitude m at a distance r from the radio position and the surface density ρ_A of objects brighter than m, find the expected number of events μ within the limits (distance $\leqslant r$, magnitude $\leqslant m$):

$$\mu = \rho_A \pi r^2 \tag{2.2}$$

where ρ_A is a function of m, the Poisson probability of there being at least one unassociated source within the search radius r_s is:

$$P = 1 - e^{-\mu}. \tag{2.3}$$

If $P \ll 1$ then the object is unlikely to be a chance association, but it is not possible to say that the probability of it being the correct counterpart is $(1-P)$ as shown by Downes et al. (1986). They consider a probability P^*; the radius r within which $P \leqslant P^*$ is a function of magnitude, therefore of surface density ρ_A. They show that the expected number of events with $P \leqslant P^*$ is thus :

$$E = \int \pi r^2(\rho_A) d\rho_A. \tag{2.4}$$

For $P \ll 1$, $\pi r^2(\rho_A) = P^*/\rho_A$ and E diverges logarithmically; but still E is finite because there is an upper limit to the search radius. Thus there is a critical Poisson probability level P_c defined by:

$$P_c = \pi r_s^2 \rho_{AT} \tag{2.5}$$

where r_s is the search radius and ρ_{AT} is the density at the detection limit for the catalogue being matched. Then P_c defines E as follows:

$$E = P_c \qquad P^* \geq P_c$$

$$E = P^*(1 - ln(P_c/P^*)) \quad P^* \leq P_c \ .$$

Now for each candidate there is the probability that it is a chance coincidence P_{not}:

$$P_{not} = 1 - e^{-E}. \tag{2.6}$$

If there are several candidates within the search area, they calculated the probability for each of them and took the one least likely to be there by chance as the best candidate. I note that there were only two occurrences ($\approx 1\%$) of this in their work. As will be shown latter, I had far more occurrences of this situation ($\approx 33\%$) between the ATLAS Radio and Fusion catalogues (see Chapter 4).

Later Dunlop et al. (1989) used this method for identification between radio galaxies and quasars observed with the Parkes Selected Regions study and the follow up optical and infrared observations of the same objects. They used a limit for the search radius such that the problem of multiple candidates did not arise. An et al. (2018) also used the Poisson probability method but as a method to identify ”possible” candidates between a single-dish survey of submillimeter galaxies and a ALMA follow up survey. The identified candidates were then used as a training set for the developed supervised

machine-learning algorithm to identify multi-wavelength counterparts to sources in sub-millimeter surveys, no mention is made of how a situation of $n > 1$ possible candidates is handled.

I found that due to the density of sources within the Fusion survey catalogue there can be $1 < n < 5$ possible infrared candidates within the search area for a radio source. The Poisson probability method has no way to assign some sort of weight or reliability to the n possible candidates within a search area, so how do we differentiate these n possible candidates? Is taking the one with the lowest P_{not} valid? In a latter section of this work I present a comparison of the Poisson probability method with the likelihood ratio method by showing a plot of likelihood ratio values with a colour code for their corresponding P_{not} values.

2.2.3 Likelihood Ratio

In the catalogues there is additional information rather than just the positions of the objects, for example flux, and it is possible to determine other information such as source density and flux distribution. The question arises: can this additional information be used to assist in selecting in our case the infrared source from a list of potential candidates to a radio source?

Consider the Figure 2.11 for a simplistic representation, there are three candidates within the $10''$ search radius (they have been shaded to represent different infrared flux density value, where a lighter shade indicates a lower flux density). In this case should the closer but weaker infrared candidate at 7" or the higher flux candidate at 8" be the most probable candidate?

Richter (1975) presented an early statistical treatment to cross match optical sources to the low resolution 5C3 radio survey using a method applying the statistical separation of real and chance identifications. Richter's work lead to de Ruiter et al. (1977) who matched optical sources to radio sources detected with the Westerbork Synthesis Radio Telescope using a probability ratio, referred to as the Likelihood Ratio. This technique uses the ratio of the following (a) the *a priori* probability $dp(r|id)$, is based on their measured separation between r and $r+dr$ because of radio and optical position uncertainties, assuming that radio source and optical counterpart are intrinsically located at the same position; (b) the *a priori* that the optical object is an unrelated background or foreground source. This method was refined by Sutherland and Saunders (1992) where the Likelihood Ratio (LR) was defined as the ratio between the probability that a candidate source is the correct identification and the corresponding probability for an unrelated background source, and is given as:

$$LR = \frac{q(m)f(r)}{n(m)} \tag{2.7}$$

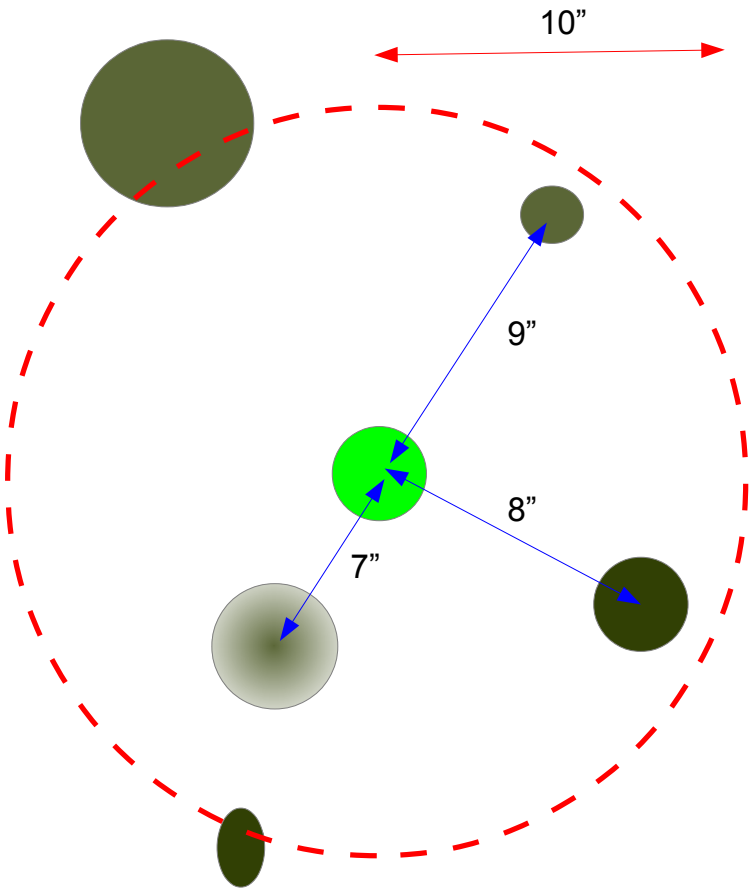


FIGURE 2.11: The green filled circle is the radio source, the light gray circles are close by infrared sources shaded to indicate different infrared Flux values. The red dotted line denotes the outer edge of the 10” radius search area centered on the radio source.

where $q(m)$ is the probability distribution of the true counterparts as a function of magnitude m; $f(r)$ represents the radial probability distribution of offsets between the counterpart centroids in the two catalogues, and the term $n(m)$ is the surface density of background and foreground objects.

Recently there has been a resurgence in the use of the LR technique to cross identify low resolution long wavelength surveys to optical data with higher resolution. For example, Ciliegi et al. (2003) used this method to find optical counterparts for the VLA 6 cm Lockman Hole survey. Ciliegi et al. (2005) used the same technique to identify optical and NIR counterparts for the VLA 1.4 GHz survey in the VIMOS VLT deep survey. More recently Smith et al. (2011) used the technique with some further refinements of Sutherland and Saunders (1992) to identify optical counterparts to 250 μm sources from the *Herschel*-ATLAS survey. The refinements from the Smith et al. (2011) technique has been followed by Fleuren et al. (2012) with some modifications of their own when matching sources between a near-infrared (NIR) VISTA VIKING and *Herschel*-ATLAS SPIRE catalogues.

In a latter section 3.1 the terms in Equation 2.7 are broken down and discussed with reference to the Fusion catalogue and the ATLAS DR3 catalogue.

2.2.4 More Recent Techniques

There has been revival in exploring existing and new methods for cross-identifying astronomical sources in multiple observations which are briefly discussed in the following paragraphs.

One such new method using a Bayesian approach was presented by Budavári and Szalay (2008). This has been taken by Salvato et al. (2018) to produce a new algorithm "Nway", which extends the previous priors of distance and sky density by adding one or more additional priors such as colour and magnitude.

An automated procedure is presented by Fan et al. (2015) using the ATLAS radio and CDFS infrared catalogues. This uses Bayesian hypothesis testing to achieve reliable associations of more complex radio morphology sources such as doubles and triples to a single infrared source.

Another avenue explored is by Citizen Science and Galaxy Zoo (Banfield et al., 2015) , working with the FIRST and ATLAS radio catalogues against the WISE infrared catalogue. The project had ≈ 6900 registered volunteers who managed to classify 1,155,000 sources with a consensus amongst the volunteers > 0.5 for more than 75 per cent of the classifications, showing the ability to cross match and classify a large numbers of sources by crowdsourcing. This method still will not scale for the number of sources expected to be found by EMU and other surveys, but could provide a robust large training set for a Machine Learning method such as discussed in Alger et al. (2018).

2.3 Summary

In this Chapter I have started by looking at the surveys used for this thesis in Section 2.1, starting with the Radio Survey and catalogue ATLAS followed by the Fusion catalogue which are used for cross identification. The OzDES survey is briefly discussed as this is used for providing redshift data, and the FIRST survey and catalogue which is used latter to test some of the refinements to the LR Algorithm.

In the final Section 2.2 in this chapter I have compared some of the methods employed to date for cross-identification between catalogues such as a simple Nearest Neighbour followed by the Poisson Probability and finally the Likelihood Ratio. The section concludes with a brief overview for some of the more recent techniques being explored by others.

This picture depicts one-tenth of the SWIRE survey field called ELAIS-N1. In this image, the bright blue sources are hot stars in our own Milky Way, which range anywhere from 3 to 60 times the mass of our Sun. The fainter green spots are cooler stars and galaxies beyond the Milky Way whose light is dominated by older stellar populations. The red dots are dusty galaxies that are undergoing intense star formation. The faintest specks of red-orange are galaxies billions of light-years away in the distant universe.Credit: NASA/JPL-Caltech/C. Lonsdale (IPAC/Caltech) and the SWIRE Team

Chapter 3

Method

In this chapter the Likelihood Ratio method is investigated in detail for multi-wavelength survey catalogue cross identification. A refinement of the method and the subsequently developed algorithm was presented and summarised in Weston et al. (2018). In this work we apply the algorithm to cross match the ATLAS DR3 radio (which is the low resolution) catalogue with the Fusion (the Infrared higher resolution) catalogue.

3.1 Modified Likelihood Ratio Technique

In Section 2.2.3 I introduced the Likelihood Ratio:

$$LR = \frac{q(m)f(r)}{n(m)} \tag{3.1}$$

I will now expand the terms as applied to the ATLAS and Fusion catalogues in this work.

In the following sub-sections the terms in Equation 3.1 are discussed with reference to the Fusion catalogue and the ATLAS catalogue. As the Fusion catalogue provides 3.6 μm flux densities, I use the flux densities, S_ν, rather than magnitudes from this point.

3.1.1 Survey Position Offsets

As highlighted in Section 2.1.1 different survey coordinates can have offsets as demonstrated in Figure 3.1 from Ciliegi et al. (2005). It was noted by Middelberg et al. (2007) that a positional offset between the ATLAS ELAIS field and catalogued SWIRE counterparts existed:

> We have tested for systematic radio-infrared position offsets by calculating the average offsets of 533 sources which consist of a single radio component and a catalogued SWIRE counterpart, and have $SNR > 10$. The offsets have a mean of $(0.08 \pm 0.03$") in right ascension and $(0.06 \pm 0.03$") in declination. Although the offset is formally significantly different from zero, we note that it is less than a tenth of a pixel in the radio image.

The offset arises because the optical reference frame is based on the FK5 catalog of 1535 stars (Fricke et al., 1988), and the reference frame used for radio sources was based on VLBI observations of extra-galactic sources (Argue et al., 1984). The VLBI reference frame is continuously monitored and improved by the continued work of the International VLBI Service for Geodesy & Astrometry (IVS) (Nothnagel et al., 2017). The current International Celestial Reference Frame (ICRS) at optical wavelengths is now defined by the Hipparcos Catalogue of 118,218 stars (ESA, 1997) (This will be superseded by the European Space Agency Gaia mission (Mignard et al., 2018)). As a result when comparing catalogues at different wavelengths there can be positional differences due to the different reference frames used for the astrometry. These differences need to be quantified if they exist, and coordinates corrected or offset allowed for as in our case, before any meaningful cross identification can be undertaken.

Middelberg et al. (2007) do not say which they think had an error, radio or infrared, but with it being as stated as "tenth of a pixel in the radio image" and the Radio having the poorer resolution, I thus allow for an error in position between the two. Those offsets look consistent with what is also seen taking the high reliability XID's and plotting Δx,Δy between the radio and infrared (IR) positions.

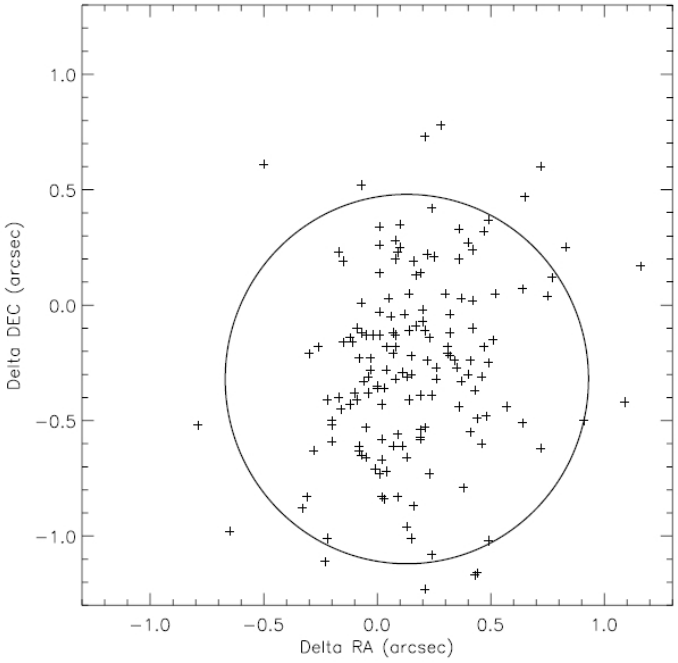


FIGURE 3.1: Position offsets for strong ($> 10\sigma$) radio point-like sources identified with a point-like optical counterpart. The circle of radius 0.8" (with the mean offset removed) encloses 90% of the sources (Ciliegi et al., 2005)

To demonstrate the systematic offset problem described above, I show Figures 3.2[1] and 3.3[1] in which the Right Ascension and Declination position offsets between the ATLAS DR3 sources and the Fusion candidates are plotted for the two fields. Here I used the selection criteria from Section 3.1.6. The top plot in each figure is a scatter plot for the position offsets and the bottom plot is a density plot of the same data using a bin size of 0.15". No significant systematic offset is visually noticeable from these plots for the two fields.

In particular for the ELAIS field we find the average offsets $\Delta RA = 0.133$" and $\Delta Dec = -0.032$"; for the CDFS field $\Delta RA = 0.029$" and $\Delta Dec = 0.094$". These figures are consistent with the positional offsets obtained by Franzen et al. (2015) for ATLAS DR3 which are $\Delta RA_{min} = 0.016$" and $\Delta Dec_{min} = 0.033$".

[1] These figures and some in later chapters were produced with TOPCAT (Taylor, 2005) as it had an easy to use ability to transpose a scatter plot to density. TOPCAT was also very useful when first reading catalogues and for draft plotting.

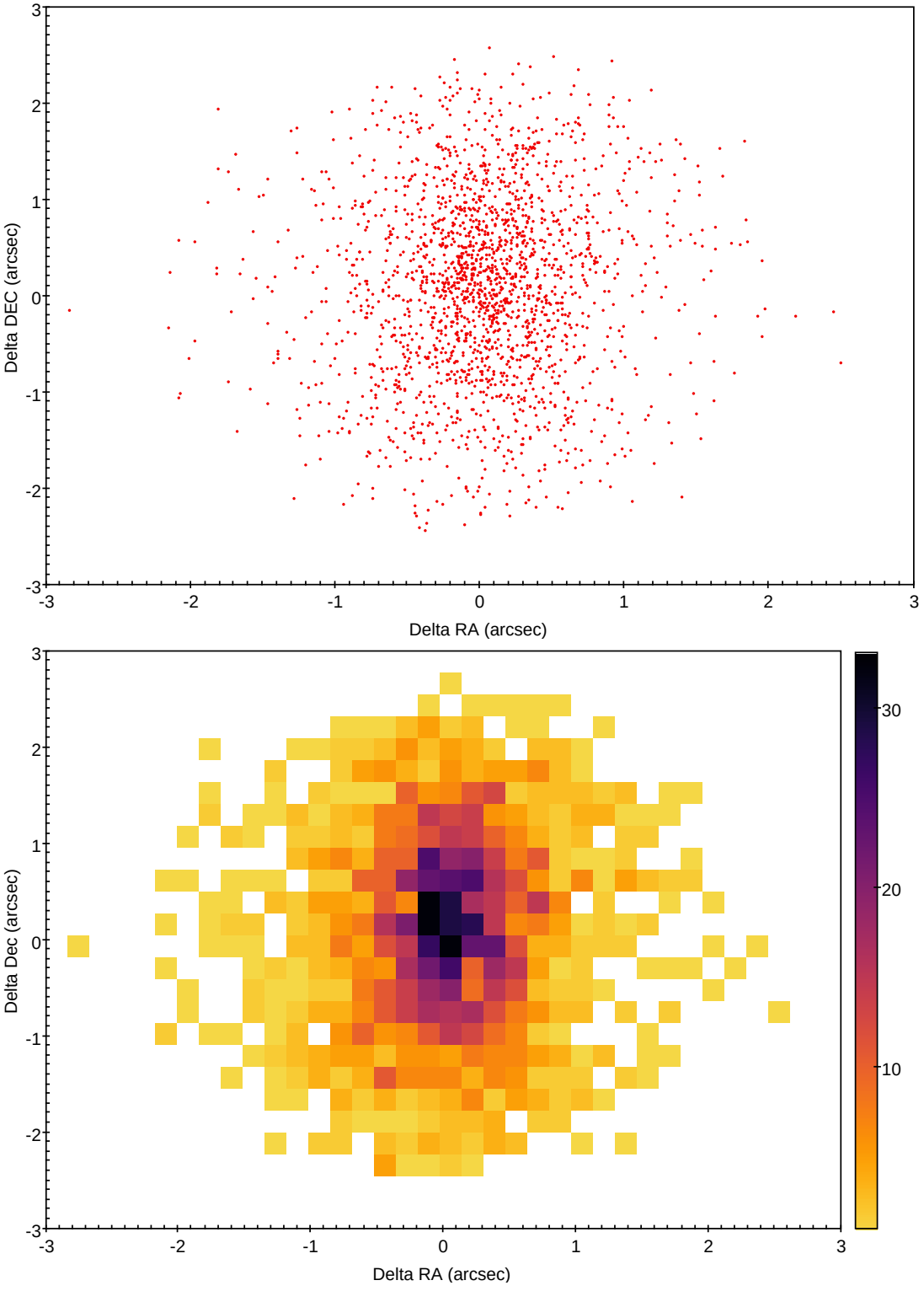


FIGURE 3.2: The top plot shows the catalog position offsets for the LR matched sources in the CDFS field for the ATLAS DR3 Radio and the Fusion catalogues. The bottom plot shows a density plot of the top plot using a bin size of $0.15 \times 0.15''$.

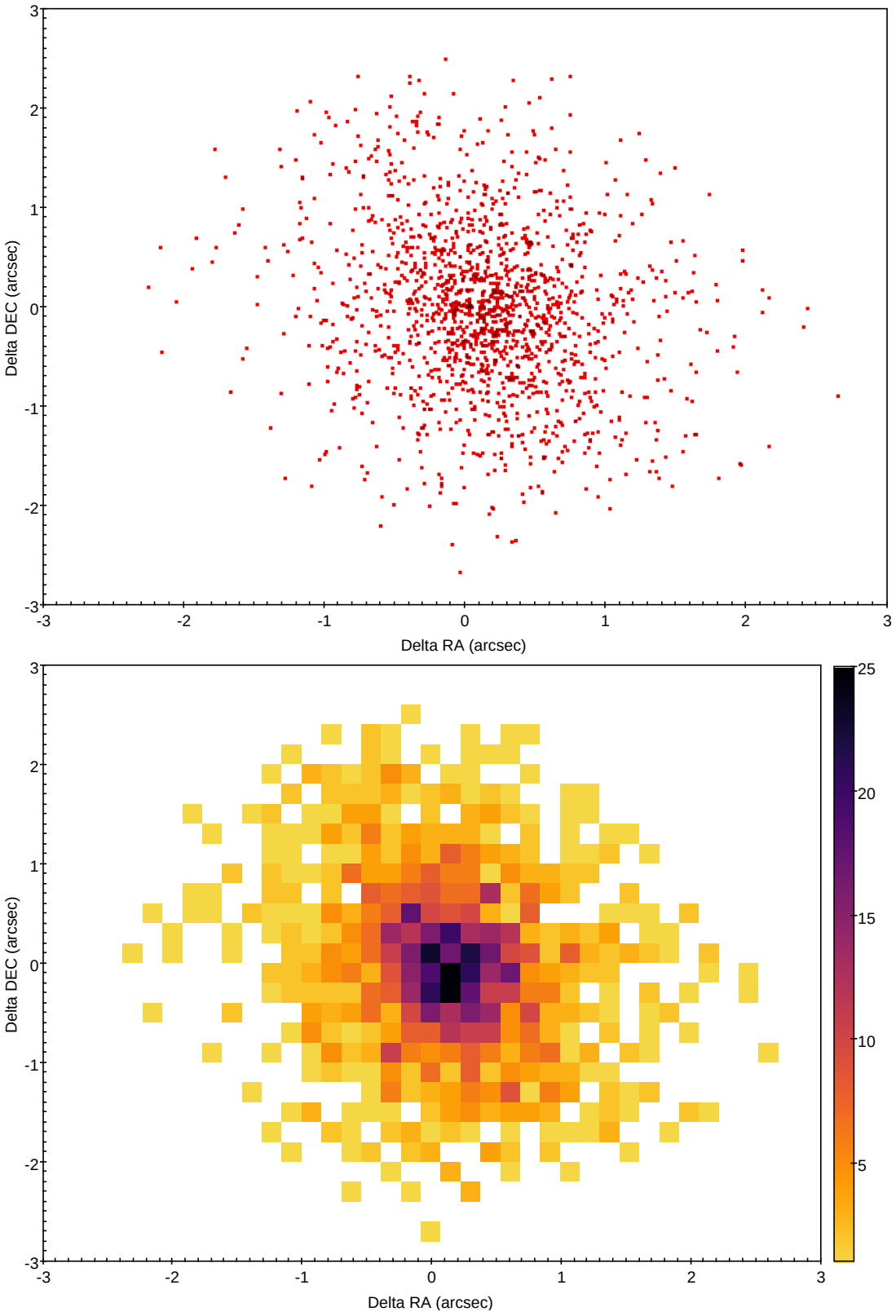


FIGURE 3.3: The top plot shows the catalog position offsets for the LR matched sources in the ELAIS field for the ATLAS DR3 Radio and the Fusion catalogues. The bottom plot shows a density plot of the top plot using a bin size of $0.15 \times 0.15''$.

3.1.2 The Radial Probability Distribution Function

Here we follow the standard approach to the definition of the LR (Sutherland and Saunders, 1992). Therefore we use $f(r)$ in Equation 3.1 as the probability distribution function (PDF) of the positional errors (see also the definition of $f(r)$ given by Fleuren et al. (2012)). We note a confusion in definition of $f(r)$ in Smith et al. (2011), where they first define $f(r)$ as "the radial probability distribution of *offsets* between the 250 μm positions and the SDSS r-band centroid", that is as the PDF of the offsets between objects of two catalogues, then (in the next paragraph) as the "probability distribution function of the positional error". The difference between the two definitions is significant, because the "probability distribution function of the positional errors" is determined by the Gaussian function, whence the PDF of the offsets between objects of two catalogues is described by the Rayleigh distribution function [2]. In our case $f(r)$ is a two-dimensional Gaussian distribution of the form:

$$f(r) = \frac{1}{2\pi\sigma^2}\exp\left(\frac{-r^2}{2\sigma^2}\right). \tag{3.2}$$

Here, r is the angular distance (in arcseconds) from the radio source position, and σ is the combined positional error given by:

$$\sigma = \sqrt{{\sigma_{\rm Posn}}^2 + {\sigma_{\rm Atlas}}^2 + {\sigma_{\rm Fusion}}^2}. \tag{3.3}$$

The Fusion absolute position uncertainty $\sigma_{\rm Fusion}$ is taken as 0.1" (Vaccari et al., 2010b) and the ATLAS absolute position uncertainty $\sigma_{\rm Atlas}$ is taken from Huynh et al. (2005), who argued that the positional accuracy of 1.4 GHz ATCA observations for 10σ detections is 0.6".

The positional uncertainty term $\sigma_{\rm Posn}$ of the individual lower resolution ATLAS sources depends on the signal to noise ratio (SNR) and the full-width at half maximum (FWHM) of the radio restoring beam (the point spread function). We use the value for $\sigma_{\rm Posn}$ as provided in Ivison et al. (2007) and used in Huynh et al. (2005):

[2] Jim Condon, who was the reviewer of our paper Weston et al. (2018), brought our attention to this confusion in Smith et al. (2011)

$$\sigma_{\mathrm{Posn}} \simeq 0.6 \left(\frac{FWHM}{SNR} \right) \tag{3.4}$$

As the position angle of the restoring beam is negligible for both fields (see Table 2.2) we can assume it is zero, and using the equation of the ellipse for the beam shape, we get :

$$\sigma_{\mathrm{Posn}} = \frac{0.6}{SNR} \times \left(\left(\frac{\sin\theta}{\Phi_{\mathrm{Min}}} \right)^2 + \left(\frac{\cos\theta}{\Phi_{\mathrm{Maj}}} \right)^2 \right)^{-1/2} \tag{3.5}$$

where θ is the Position Angle of the candidate Fusion counterpart relative to the radio source defined clockwise from North. The SNR values are taken from the ATLAS radio catalogue for each source. The terms Φ_{Min} and Φ_{Maj} are the values of minor and major axes of the beam given in Table 2.2.

Figure 3.4[3] shows the distribution of the values of $f(r)$ with radius from Equation 3.2 for the individual candidate Fusion counterparts found within an initial search radius of 10". It can be seen that $f(r)$ is very small for $r > 6$". Figure 3.5[3] we show the $f(r)$ values on a log scale for r up to $r < 10$". It can be seen at $r < 6$" that $f(r)$ has become very small with a range of values $10^{-12} < f(r) < 10^{-3}$. As $LR \propto f(r)$ then the LR will also become very small and Nearest Neighbours outside a search radius of 6" can be ignored; in the final algorithm I constrained the work to those matches within a 6" search radius. I also further discuss the rationale for choosing a final search radius of 6" in Section 3.1.5.

[3]These figures and many others in this work were produced with GNUPLOT (Williams et al., 2019)

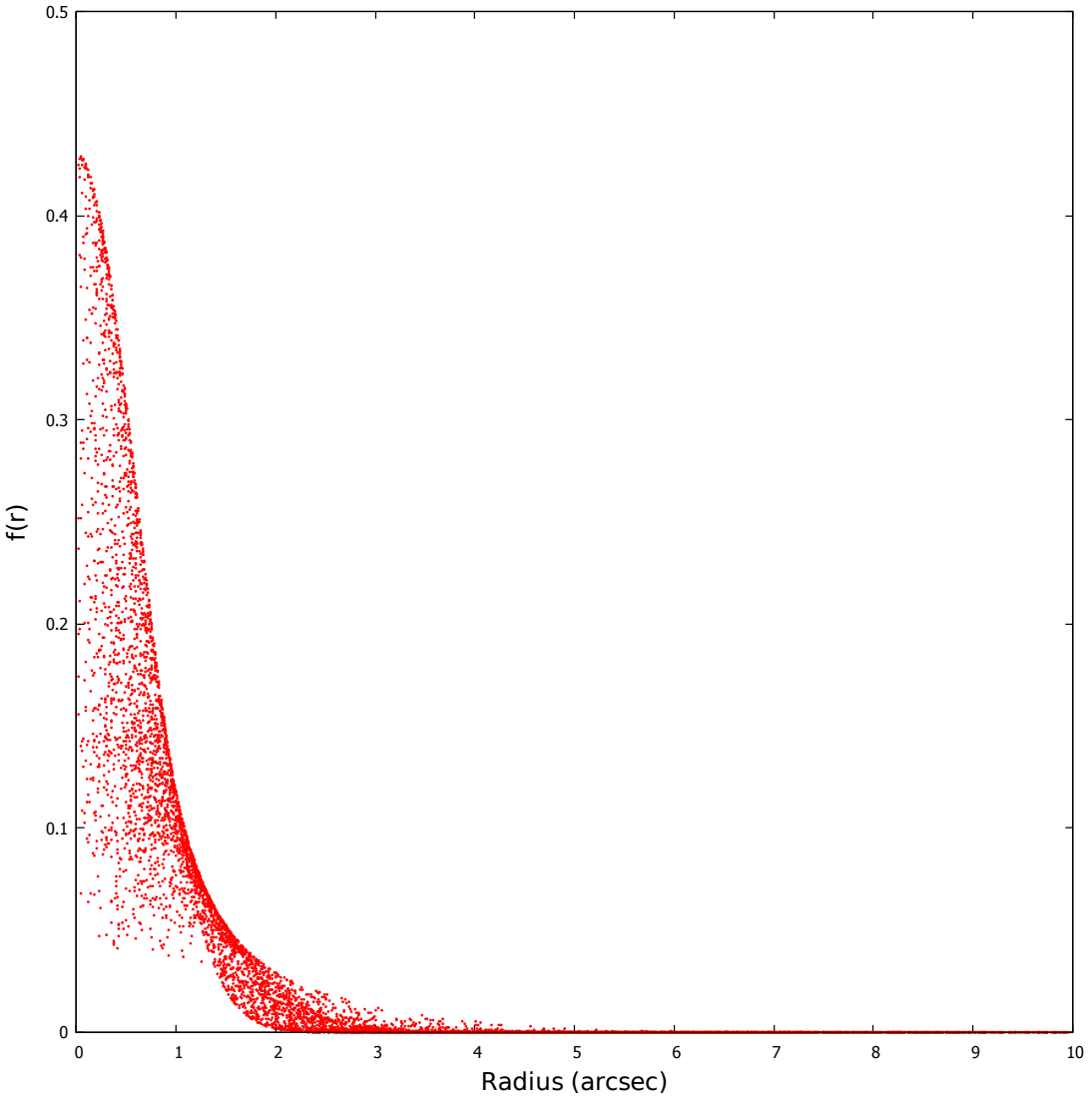


FIGURE 3.4: A plot of $f(r)$ vs r for all Fusion candidates to the ATLAS sources within a search radius of 10".

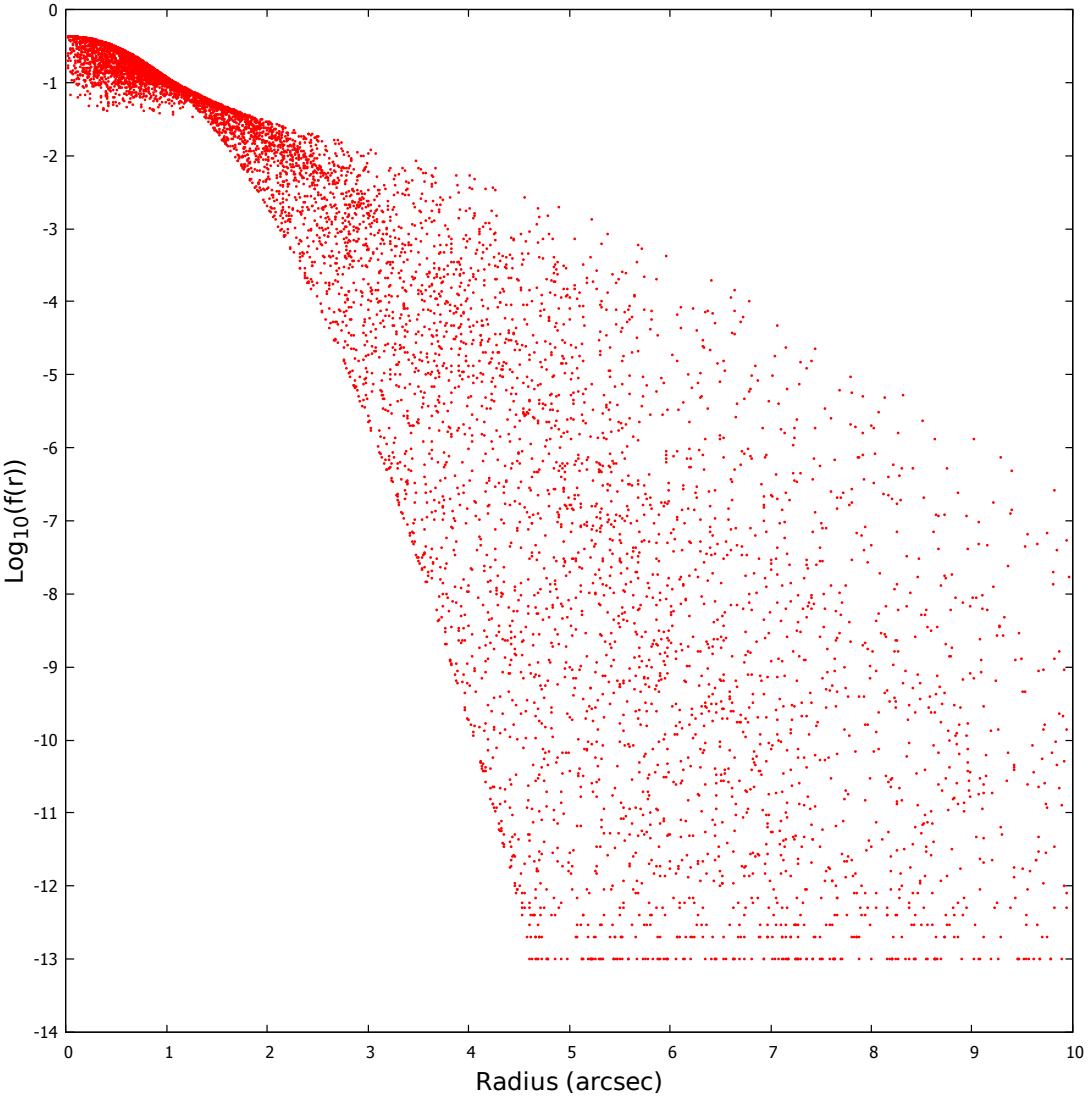


FIGURE 3.5: Plots of $Log_{10} f(r)$ vs r for all Fusion candidates to the ATLAS Radio sources within a search radius of 10".

3.1.3 The Background Flux Density Probability Function

The quantity $n(S_\nu)$ is the surface density of background and foreground Fusion sources with flux density S_ν. The surface density of Fusion sources not related to ATLAS radio sources can be obtained from the Fusion catalogue by one of two methods, both of which have been implemented within a new algorithm Likelihood Ratio in PYthon (LRPY) :

1. Use all Fusion sources with an annulus of 6" $< r <$ 100" around each radio candidate — this is referred to as the *local* method.

2. Use all Fusion sources from the area of overlap between the two catalogues (defined in section 2.1.5) — this is referred to as the *global* method and is the default in LRPY.

With the *local* method care must be taken that the annuli are not too close to the edge of the field, as this can result in a lower count for the background sources as they encompass regions beyond the survey with no sources. To mitigate this edge effect, only annuli 100" from the inside edge of the area are used. The flux densities are binned and the resultant $n(S_\nu)$ values are then divided by the total area covered to produce a density function, for the CDFS and ELAIS-S1 fields (see Figure 3.6). The values of S_ν are stored in a database lookup table for use later in the final LR calculations.

There are advantages and disadvantages to both methods: the *local* method can account for variations in depth and density of a catalogue, and for very large surveys the entire catalogue is not required. But the *global* method can provide better statistics, if area is limited and depth is uniform, which can be important for both bright and faint flux densities where numbers are small. We use the global method as default as it best suits our situation, with the Fusion catalogue being uniform in depth.

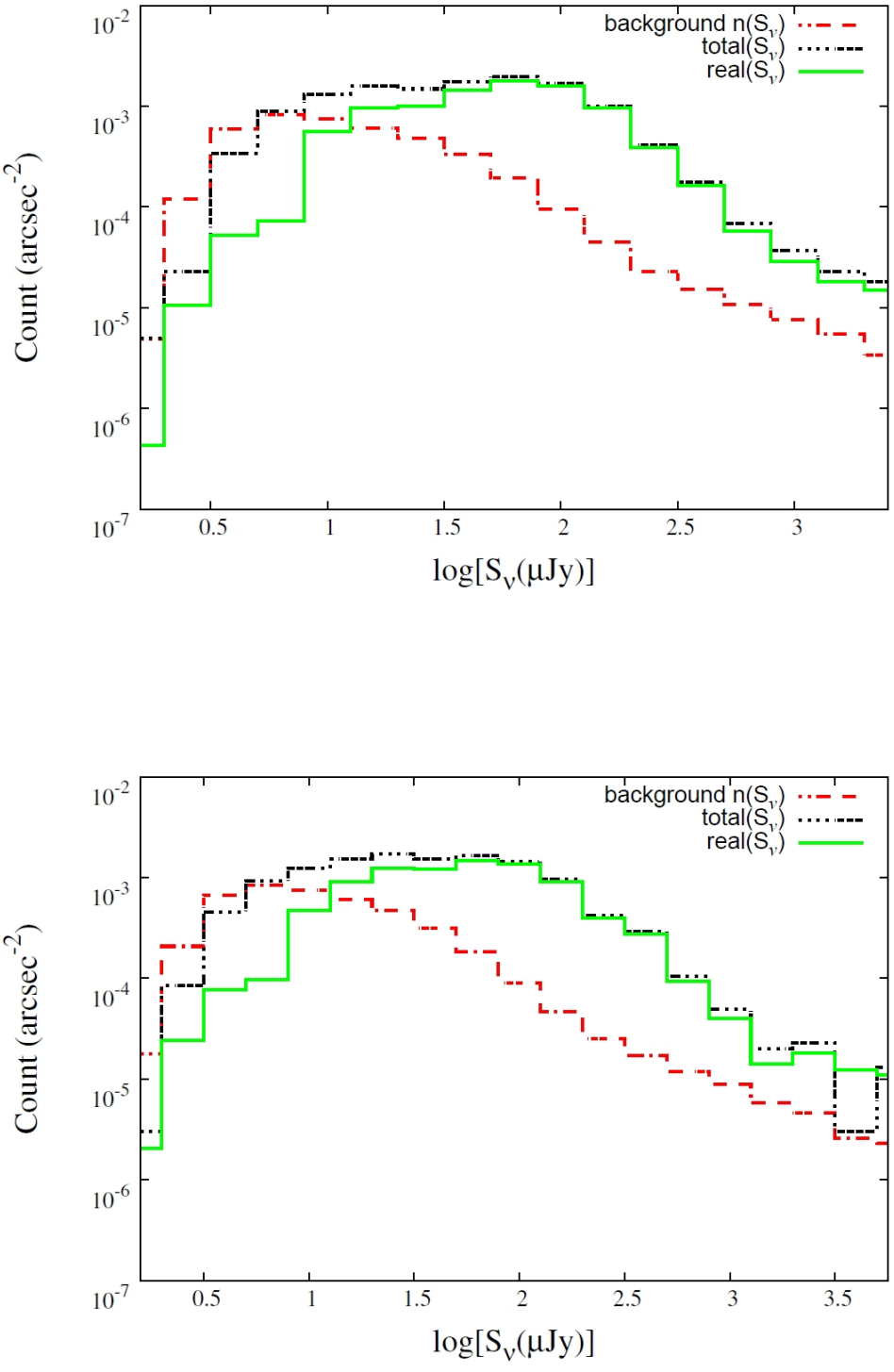


FIGURE 3.6: Histograms of the Fusion values for $n(S_\nu)$ background (red dashed line), $total(S_\nu)$ (black dotted line) and $real(S_\nu)$ (green solid line) for CDFS (top) and ELAIS-S1 (bottom). Note that S_ν is the 3.6 μm flux density.

3.1.4 The True Counterpart Probability Distribution

The true counterpart probability distribution, $q(S_\nu)$, is the probability that a true Fusion counterpart to a radio source has a flux density of S_ν at $3.6\,\mu$m :

$$q(S_\nu) = \frac{real(S_\nu)}{\sum_{S_\nu} real(S_\nu)} \times Q_0 \tag{3.6}$$

Here $real(S_\nu)$ is the background subtracted distribution of flux densities of Fusion sources around an ATLAS source. The coefficient Q_0 represents the probability that a real counterpart is above the detection limit in the matching catalogue; it does not depend on the search radius. To determine $real(S_\nu)$ we take:

$$real(S_\nu) = total(S_\nu) - n(S_\nu) \tag{3.7}$$

where $n(S_\nu)$ is the surface density of unrelated background/foreground sources introduced in the previous subsection and $total(S_\nu)$ is the surface density of all Fusion sources to be matched within the search radius, r. These include the true counterpart (if above the detection limit) plus unrelated background and foreground sources. These values are kept in the same LRPY database table as $n(S_\nu)$ for use later by the algorithm.

The distributions of $real(S_\nu)$ and $total(S_\nu)$, as well as distribution of $n(S_\nu)$, are shown in Figure 3.6. It should be noted that for the invalid physical condition where $n(S_\nu) > total(S_\nu)$ (i.e. when the background exceeds the measured distribution), a method is adopted to set $real(S_\nu)$ to be positive. This occurs at faint and bright flux densities when there is a small number of Fusion sources in a given flux density bin. To keep our estimate of $real(S_\nu)$ positive and physical, we replace negative values of $real(S_\nu)/total(S_\nu)$ with a value determined from the last positive value at faint and bright flux densities. This adaptation ensures we account for potential counterparts at the extreme flux density values.

A reasonably accurate determination of Q_0 in Equation 3.6 is required. To estimate Q_0 by summing $real(S_\nu)$ and dividing by the total number of ATLAS sources, then Q_0

would likely be over-estimated due to source clustering and genuine multiple matches (which we deal with in 5.1). While this simple method finds values of $Q_0 = 0.845$ for CDFS and 0.822 for ELAIS-S1, I undertake the following process to estimate its value more accurately. Following Fleuren et al. (2012) who, to avoid these issues, estimate the value $1 - Q_0$, which in this case will be the fraction of ATLAS sources *without* a Fusion counterpart, which are referred to here as 'blanks'. These 'blanks' will principally be ATLAS sources with true counterparts below the Fusion detection limit, or ATLAS sources with true Fusion counterparts outside the search radius. The latter case is possible when ATLAS sources are complex and the Fusion counterpart may well not correspond to any radio component, but lie between components (lobes) which can be separated by tens of arcseconds.

The true fraction of blanks, $1 - Q_0 = \overline{S_t}$, will be greater than the observed fraction of blanks, $\overline{S_o}$, because a fraction of true blanks will have random (i.e. physically unrelated) Fusion sources within the search radius. Hence, we do not wish to falsely count such sources as matches. Therefore, $\overline{S_t}$ equals $\overline{S_o}$ plus some fraction of true blanks 'contaminated' by random Fusion sources. Hence,

$$\overline{S_t} = \overline{S_o} + \overline{S_t} \times \frac{N_s}{N} \tag{3.8}$$

where N_s is the number of sources, out of N with randomly generated positions, containing one or more Fusion sources within the search radius, and N is the total number of radio sources. If we define N_r as the number of N randomly generated sources which do not have a Fusion counterpart within the search radius, such that $N = N_s + N_r$, it is straightforward to show:

$$\overline{S_t} = \overline{S_o} + \overline{S_t} \times \left(\frac{N - N_r}{N} \right) \tag{3.9}$$

$$\frac{\overline{S_t}}{N} = \frac{\overline{S_o}}{N_r} \tag{3.10}$$

Hence, one can determine the fraction of true blanks, $\overline{S_t}/N$, as a function of search radius, r, by determining the ratio of the number of observed blanks, $\overline{S_o}$, to the number of blanks from a randomly generated catalogue, N_r, as a function of r. We calculate this result for our case by counting the number of observed blanks with increasing search radii across 0" $< r <$ 20" and repeat for a catalogue of N randomly generated positions of Fusion sources. We present these results in Figure 3.7 showing, as a function of radius, the fraction of observed blanks, $\overline{S_o}/N$, and the fraction of random blanks, N_r/N and their ratio which equals $\overline{S_t}/N$. As the radius, r, increases to encompass all true counterparts, this result tends toward $1-Q_0$. We can fit the distribution in Figure 3.7 with the following expression:

$$\frac{\overline{S_t}(r)}{N} = 1 - Q_0 \times (1 - e^{-r^2/2\sigma^2}) \tag{3.11}$$

from Fleuren et al. (2012) where σ is positional uncertainty. This function returns unity at $r = 0$ and $1-Q_0$ for large r. By fitting for Q_0 and taking σ as the maximum value for the field ($\sigma_{CDFS} = 1.08$" , $\sigma_{ELAIS-S1} = 0.868$") to the function, using a non-linear least squares fit, we obtain for both these fields the values and uncertainties for Q_0 presented in Table 3.1.

TABLE 3.1: Estimated fraction, of the non-blanks, Q_0 (ATLAS sources with a true counterpart), and its error δQ_0.

Field	Q_0	δQ_0
CDFS	0.831	0.018
ELAIS-S1	0.825	0.017

These values are fairly similar to the ones from our earlier crude estimate, but with the CDFS being a little higher and ELAIS-S1 being slightly lower. We note that this function must pass through (0,1) by definition, but may deviate within the best match search radius due to physical clustering of sources or from the existence of multiple true components.

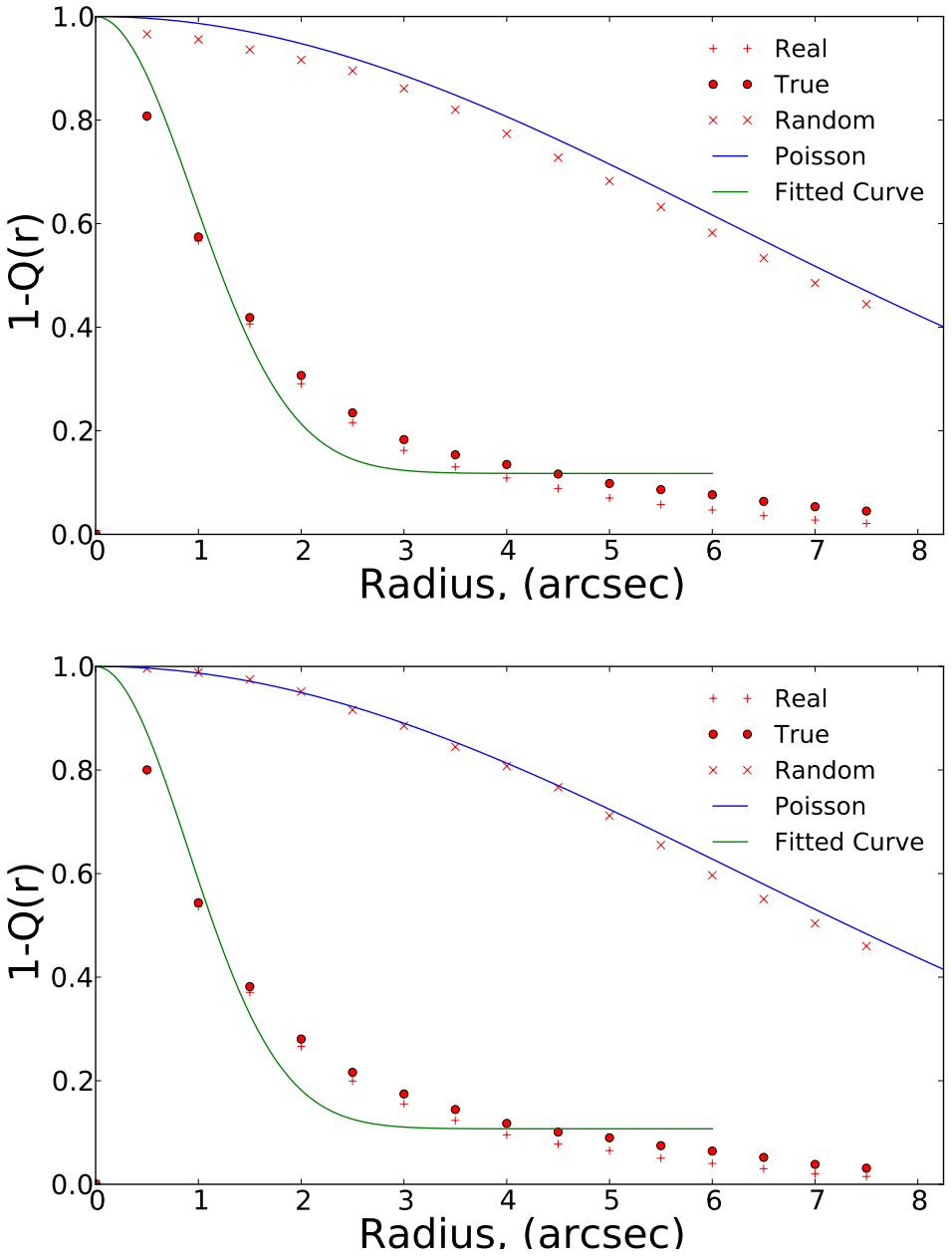


FIGURE 3.7: Estimation of Q_0 for CDFS (top) and ELAIS-S1 (bottom) determined from fitting the ratio, $\overline{S_t}$ (red filled circles), of the fraction of observed blanks, $\overline{S_o}$ (plusses), and the fraction of random blanks, N_r (crosses). The green line represents the functional fit to the ratio (Equation 3.11), and the blue line is an estimate of the fraction of random blanks from Poisson statistics using Equation 3.14. Taking (Fleuren et al., 2012), the dependence of Q on the search radius can be presented in the form $Q(r) = Q_0 \exp(-r^2/2\sigma^2)$.

It is possible to model the distribution of random blanks, N_r, in Figure 3.7. The probability that an observed area of sky, $a = \pi r^2$, has one or more random Fusion sources is given by the Poisson distribution $P(a) = 1 - e^{-a\lambda}$, where λ is the density of Fusion sources. Hence, from Equation 3.8, we can write:

$$\overline{S_t} = \overline{S_o} + \overline{S_t}(1 - e^{-a\lambda}) \tag{3.12}$$

which we can rearrange to :

$$\overline{S_t} = \overline{S_o} e^{a\lambda} \tag{3.13}$$

and therefore from Equation 3.10 we get:

$$\frac{N_r}{N} = e^{-a\lambda} \tag{3.14}$$

In Figure 3.7 we overlay this function on the random blanks distribution with radius using a density of Fusion sources of $\lambda = 0.004$ arcsec^{-2}, for both fields. We note this theoretical determination matches our empirical determination well.

3.1.5 The Search Radius

Fleuren et al. (2012) deal with 1,376,606 near-IR sources in the area of 56 $\deg^2$, which results in density of near-IR sources of $\lambda = 6.8$ arcmin^{-2} and mean intersource distance of $r_0 = (\pi\lambda)^{-1/2} \sim 13$". They chose the search radius $r = 10$", which is 77% of the mean intersource distance. In our case, the Fusion source density is higher (~ 15.1 arcmin^{-2}) and therefore the mean intersource distance is smaller: $r_0 = 8.7$". To be consistent with Fleuren et al. (2012), I chose the search radius at 77% of our mean intersource distance: $r = 8.7" \times 0.77 \sim 6"$. Also as shown in section 3.1.2 the function $f(r)$ exponentially decreases making the LR vanishingly small, $< 10^{-3}$, outside $r = 6"$.

3.1.6 Proposed Selection Criteria

In this section we analyse different aspects of the resultant cross-matches and present how we determine criteria for selecting true matches from the LR and reliability values.

Due to the high density of background sources in the Fusion catalogue there can be $0 \leqslant n \leqslant 5$ possible candidate Fusion counterparts for a given radio source within the search radius of 6" (see Table 3.2). Included in Table 3.2 is the expected number of radio sources in each field with n Fusion potential counterparts from a random distribution, i.e. via Poisson statistics. The numbers we find are higher than those from Poisson statistics suggesting (a) potentially more than one galaxy is contributing to the radio emission and (b) there may be clustering around the host galaxies of radio sources. The former option is discussed in Section 5.1 and we noted the latter point in Section 3.1.4.

TABLE 3.2: Statistics of Fusion counterparts inside the $6''$ search radius around ATLAS sources. The first column is the number of Fusion matches; the second column is the number of ATLAS sources with the corresponding number of Fusion matches (n) for CDFS; the third column is the percentage of the total. Columns four and five are the same, but for the ELAIS-S1 field.

	CDFS			ELAIS S1		
n	*Poisson*	*Count*	*%*	*Poisson*	*Count*	*%*
(*Matches*)						
0	1905	378	12.2	1307	177	8.3
1	914	1657	53.8	628	1157	54.7
2	219	832	27.0	151	615	29.1
3	35	185	6.0	24	138	0.6
4	4	23	1.0	3	24	1.1
5	0.4	3	0.1	0.3	2	0.1
Totals		3078			2113	

To select from these n possible candidates a reliability value for each can be determined thus:

$$R_j = \frac{LR_j}{\sum_{i=1}^{n} LR_i + (1 - Q_0)} \tag{3.15}$$

where R_j is the reliability that the candidate Fusion counterpart j of n possible counterparts is associated to the radio source. The sum is taken over all n potential candidates

within the 6" search radius and Q_0 is the probability that the true Fusion counterpart is above the detection limit (determined in 3.1.3 and presented in Table 3.1). Plots of Reliability versus Likelihood Ratio for each candidate counterpart for both fields are presented in Figure 3.8.

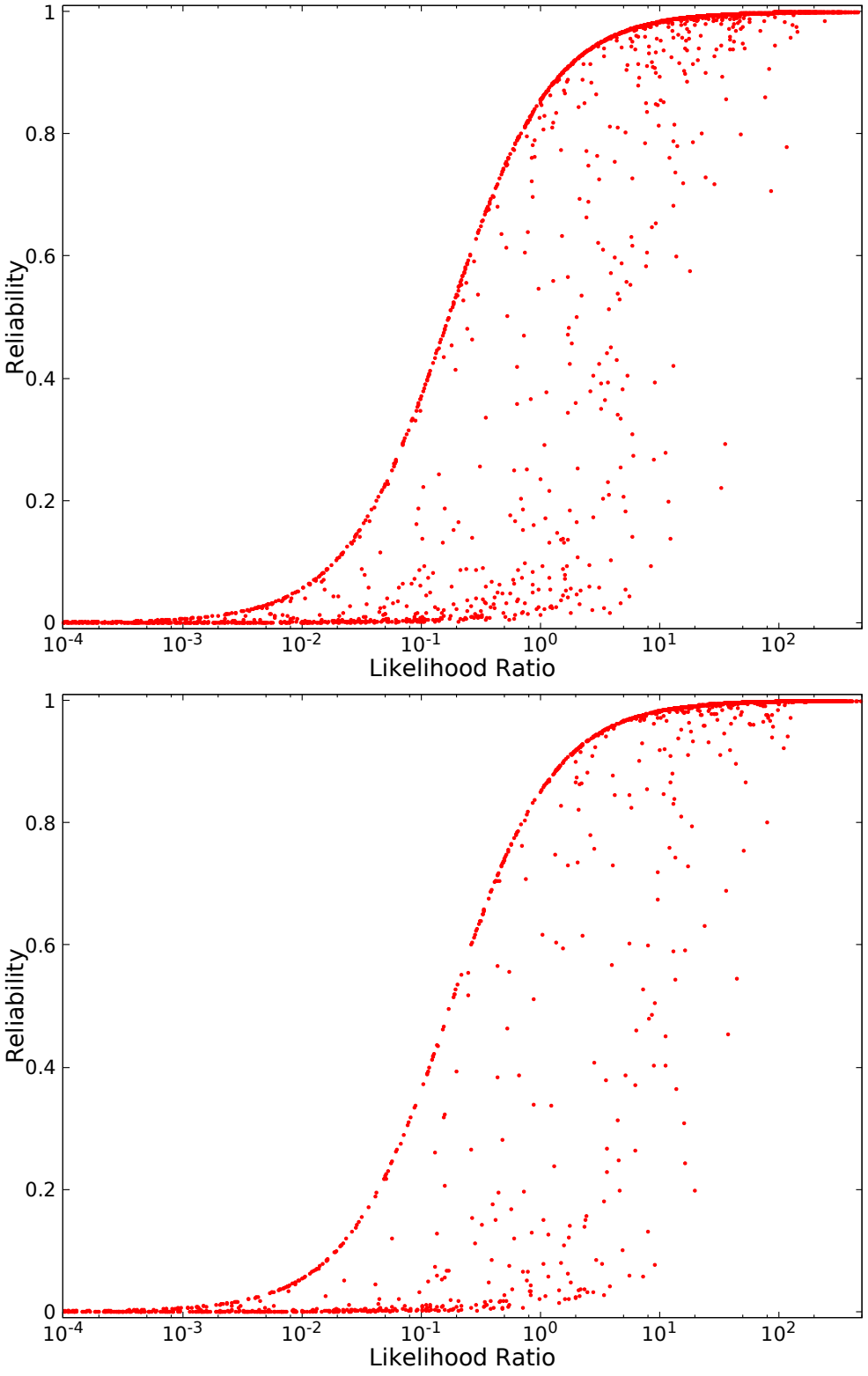


FIGURE 3.8: Plots showing the variation of the reliability, R, as a function of the likelihood ratio or CDFS (top) and ELAIS S1 (bottom). For both plots there is some symmetry of data points around $R \approx 0.5$ (which will be discussed in section 5.1) could be used to identify potential Fusion pairs being related to one radio source.

There is always a trade-off between maximising the number of radio sources with 'reliable' counterparts and minimising the contamination of false associations. Equation 3.15 permits us to compare the relative likelihood of an association between an ATLAS and a Fusion source in the situation where we have two or more potential counterparts. Determining the appropriate cut-off values in LR and Reliability is therefore crucial for any scientific analysis.

Reliability can also be calculated for the case of a single Fusion source, n= 1 (one Fusion source in the search radius):

$$R_j = \frac{LR}{LR + (1 - Q_0)}. \tag{3.16}$$

Hence, once a LR cut-off, LR_c, is determined, the corresponding cut-off value of reliability, R_c, can be calculated for single sources, as we know Q_0 (here we take $Q_0 = 0.85$).

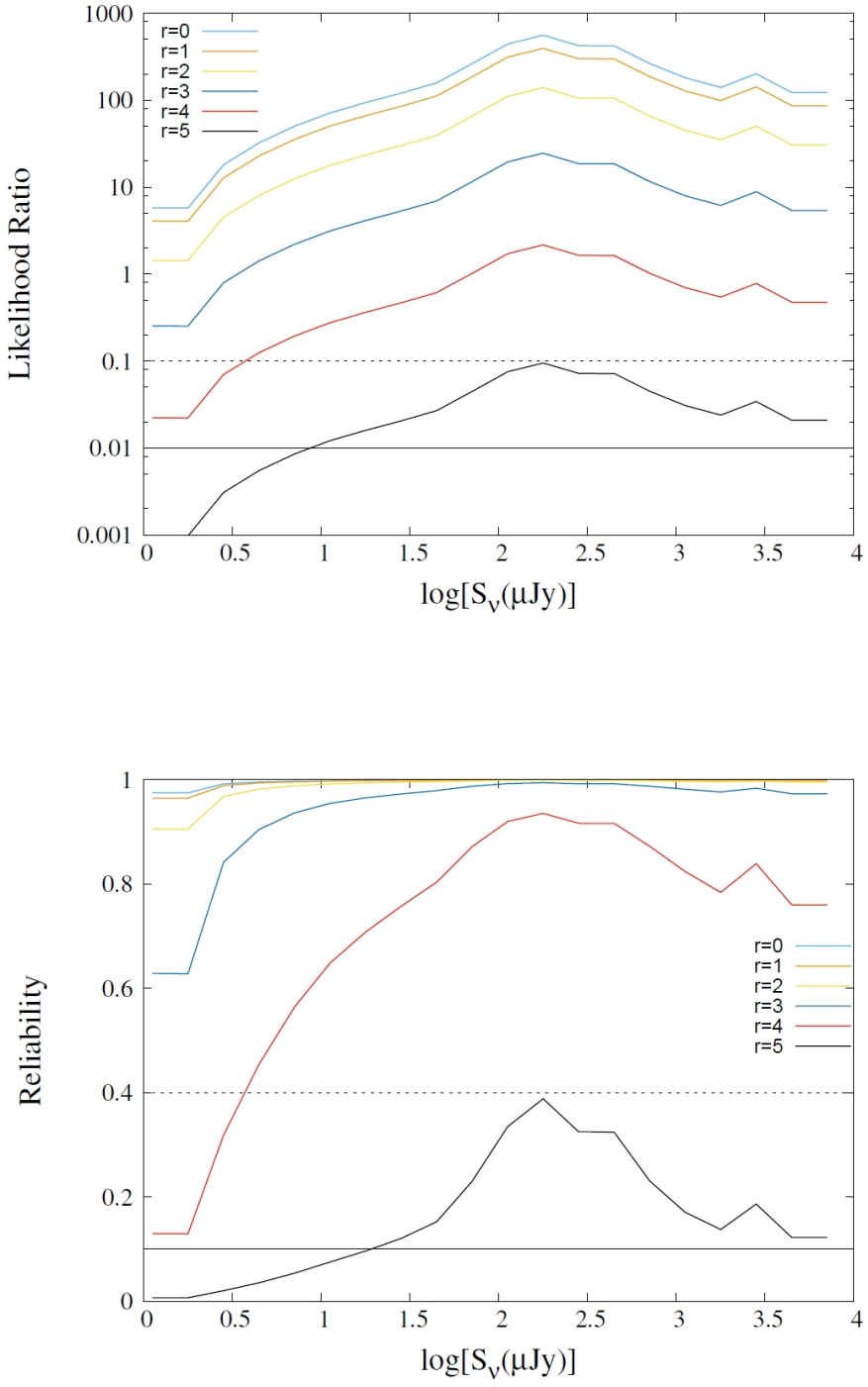


FIGURE 3.9: Likelihood Ratio against $\log(S_\nu)$ (Top) and Reliability against $\log(S_\nu)$ (Bottom) for extreme values of the positional uncertainty $\sigma = 1.2$" in CDFS. Families of curves are computed for distances r between a candidate Fusion counterpart and the ATLAS source in the range $r = 0$" to 5". Distributions of $real(S_\nu)$ and $n(S_\nu)$ used to determine the likelihood ratio are taken from the CDFS field statistics. Horizontal solid lines corresponding to suggested cut-off values presented in 3.1.6, for $LR = 0.01$ and $R = 0.1$ are drawn on the figures. The horizontal dashed lines show a much stronger selection criteria $LR = 0.1$ and $R = 0.4$. Note that S_ν is the 3.6μm flux density.

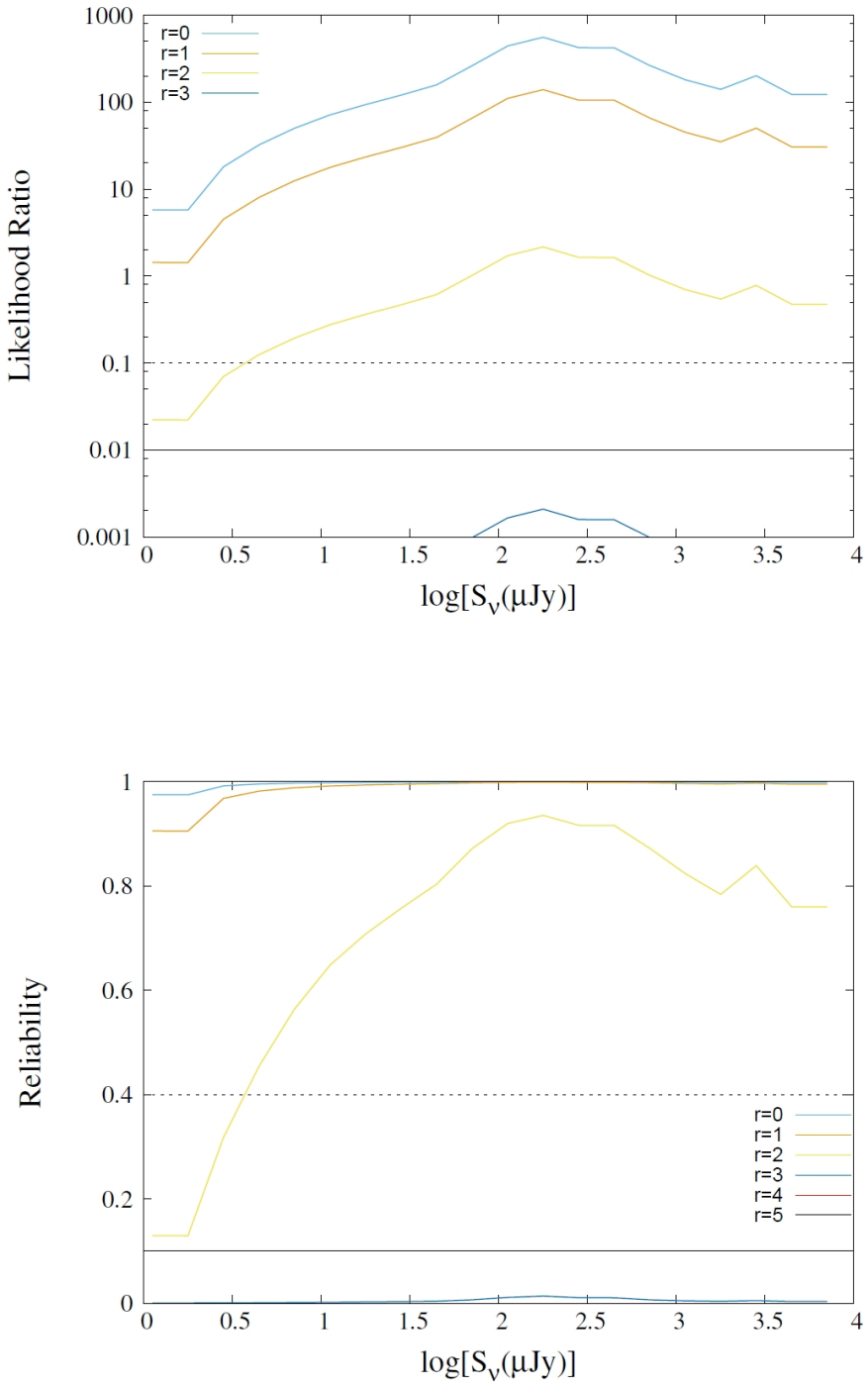


FIGURE 3.10: Likelihood Ratio against $\log(S_\nu)$ (Top) and Reliability against $\log(S_\nu)$ (Bottom) for extreme values of the positional uncertainty $\sigma = 0.6$" in CDFS. Families of curves are computed for distances r between a candidate Fusion counterpart and the ATLAS source in the range $r = 0$" to 5". Distributions of $real(S_\nu)$ and $n(S_\nu)$ used to determine the likelihood ratio are taken from the CDFS field statistics. Horizontal solid lines corresponding to suggested cut-off values presented in 3.1.6, for $LR = 0.01$ and $R = 0.1$ are drawn on the figures. The horizontal dashed lines show a much stronger selection criteria $LR = 0.1$ and $R = 0.4$. Note that S_ν is the 3.6μm flux density.

Figures 3.9 and 3.10 show the families of theoretical curves LR vs. S_ν and R vs. S_ν for the range of r (distance between the radio source and Fusion candidate) from 0" to 5" (all inside the search radius of 6") and $Q_0 = 0.85$. They are calculated for the set of $real(S_\nu)$ and $n(S_\nu)$ we observe in the CDFS field for radio sources detected. The upper plots are computed for $\sigma = 1.2$", which is close to the maximum value of σ we deal with in the CDFS field (Section 3.1.1); the bottom plots correspond to $\sigma = 0.6$" (close to the minimum value of σ in CDFS field).

We can choose the LR_c for single Fusion sources in such a way that for $\sigma = 1.2$" almost all single Fusion sources within $r = 5$" are considered as true counterparts. This condition is fulfilled when $LR_c = 0.01$ (horizontal solid line in Figures 3.9 and 3.10) and corresponds to a reliability cutoff of $R_c = 0.055$ for CDFS and 0.054 for ELAIS-S1. These cut-off values are shown in graphs with horizontal solid lines. The horizontal dashed lines show a much stronger criterion for cut-off values of $LR_c = 0.1$ and the corresponding $R_c = 0.37$ for CDFS and 0.36 for ELAIS-S1. In this case all Fusion sources with $r > 4$" are excluded from consideration as possible counterparts.

Another approach to determining a value for the reliability cut-off, where those candidates with a reliability greater than R_c can be treated as true counterparts, was used by Smith et al. (2011), who estimated the number of false cross-matches using :

$$N_{\rm false}(R_c) = \sum_{R_i^{Max} \geqslant R_c} (1 - R_i) \tag{3.17}$$

Figure 3.11 shows $N_{\rm false}$ as a function of R_c for our two fields. Smith et al. (2011) used a reliability limit of 0.8 which gave them a contamination rate of 4.2 %. Bonzini et al. (2012) selected only those candidates with a reliability greater than 0.6 as the threshold, to ensure the expected number of spurious associations was below 5 % of the auxiliary catalogue, while at the same time maximising the number of identified sources. Using a similar acceptable contamination threshold at 5 % for our datasets, results in $R_c = 0.1$ for both CDFS and ELAIS-S1 fields (Figure 3.11). As we discussed above, this value of R_c corresponds to $LR_c = 0.01$ for single Fusion sources.

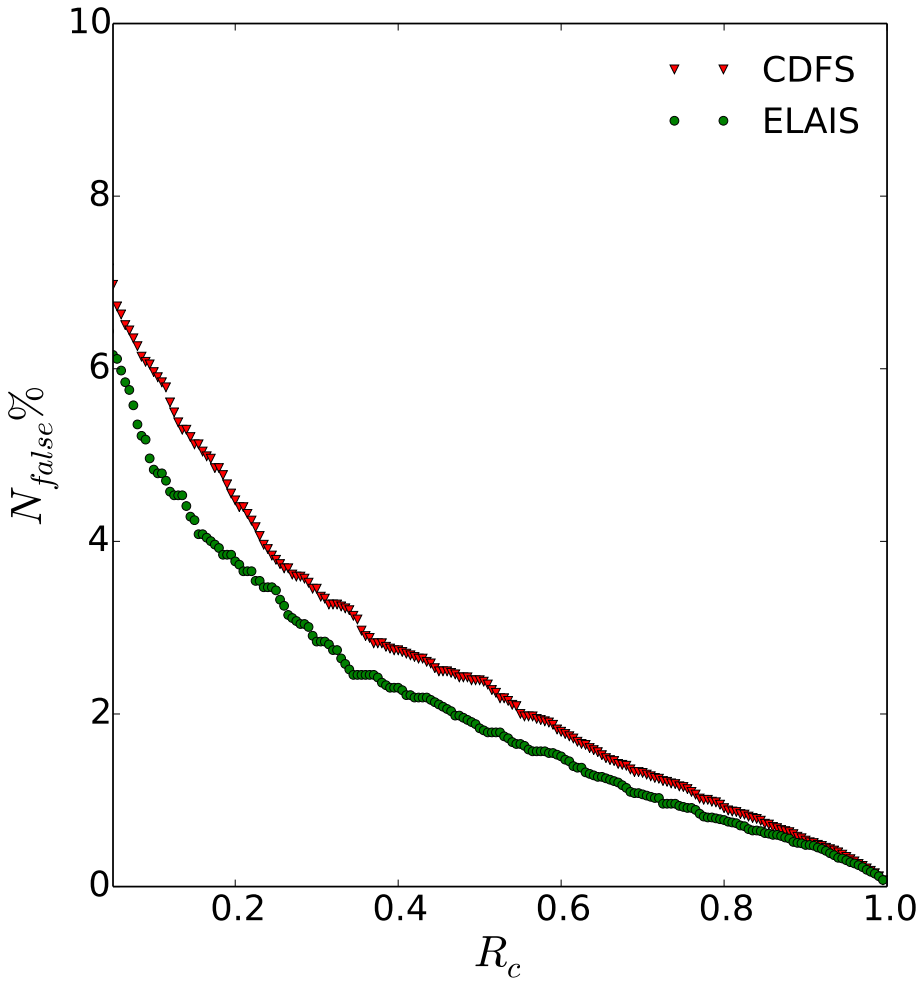


FIGURE 3.11: Estimated percentage of the false cross-matches, N_{false}, as a function of the Reliability cut-off, R_c, for CDFS (red) and ELAIS-S1 (green) determined with Equation 3.17.

Using the 5% contamination threshold, I accept only Fusion counterparts with $LR \geq LR_c$. Here I use $LR_c = 0.01$ and we reject all Fusion counterparts below this value. We apply this to all Fusion sources, whether they are single or multiple.

For the ATLAS fields using a LR cutoff of 0.01 (reliability cutoff of 0.1) and using Equation 3.17, we have for CDFS $N_{false} = 159$ which is 5.2%; and for ELAIS-S1 we have $N_{false} = 99$ which is 4.8%. Using this cutoff there are 2135 ATLAS sources with at least one match in the CDFS field and 1580 in the ELAIS-S1 field.

3.2 Computational Method

In this section the computational method employed for the LRYP algorithm is presented.

3.2.1 The Relational Database Design

The ATLAS DR3 catalogue comes as a flat ascii text file of 20 columns. This has been normalised into five relational database tables, where each uses the component id as the primary key. For example each ATLAS component has an entry in the *field*_coords table as defined in 3.3. Where *field* is the name of the field (CDFS/ELAIS) for which this table stores information for. The naming convention allows the code and table definitions to be the same and just having to provide the catalogue *field* that is required to be processed.

TABLE 3.3: *field*_coords relational database table definition. Note: "DEC" is a reserved word in MySQL, so DECL was used instead to denote declination.

Column	Datatype	Indexed	Description
ID	VARCHAR(8)	PK	Component identification number
RA	DOUBLE	Y	Right Ascension (degrees, J2000)
RA_ERR	DOUBLE	-	Error in right ascension (arcsec)
DECL*	DOUBLE	Y	Declination (degrees, J2000)
DECL_ERR*	DOUBLE	-	Error in declination (arcsec)

TABLE 3.4: *field*_radio_properties relational database table definition

Column	Datatype	Indexed	Decription
ID	VARCHAR(8)	PK	Component identification number
SNR	DOUBLE	-	signal-to-noise ratio of raw detection
RMS	FLOAT	-	local rms noise level (mJy/beam)
BWS	FLOAT	-	local bandwidth smearing value
SP	FLOAT	-	Fitted source peak (mJy/beam)
SP_ERR	FLOAT	-	Error in fitted source peak (mJy/beam)
SINT	FLOAT	-	Integrated flux density (mJy)
SINT_ERR	FLOAT	-	Error in integrated flux density (mJy)
OBS_FREQ	FLOAT	-	Frequency at which the peak and integrated flux was measured (MHz)

* "DEC" is a reserved word in MySQL, so DECL was used instead to denote declination

TABLE 3.5: *field*_deconv relational database table definition

Column	Datatype	Indexed	Description
ID	VARCHAR(8)	PK	Component identification number
DECONV	FLOAT	-	Deconvolved angular size (arcsec)
DECONV_ERR	FLOAT	-	Error in deconvolved angular size (arcsec)
V	FLOAT	-	Visibility area

TABLE 3.6: *field*_name relational database table definition

Column	Datatype	Indexed	Description
ID	VARCHAR(8)	PK	Component identification number
SURVEY	VARCHAR(6)	-	Survey - ATLAS
NAME	VARCHAR(30)	-	Full catalogue name

TABLE 3.7: *field*_sindex relational database table definition

Column	Datatype	Indexed	Description
ID	VARCHAR(8)	PK	Component identification number
SINDEX	DOUBLE	-	Spectral index of source between 1400 and 1710 MHz
SINDEX_ERR	DOUBLE	-	Error on spectral index

TABLE 3.8: *field*_matches relational database table definition

Column	Datatype	Indexed	Description
CID	CHAR(8)	PK	Component identification number
SWIRE_INDEX_SPITZER	INT(11)	Y	SWIRE identification
F_R	DECIMAL(15,13)	-	$f(r)$ value for the ATLAS and SWIRE
		-	pair of source's
DX	FLOAT	-	difference between Radio and IR
		-	RA position in arcsec
DY	FLOAT	-	difference between Radio and IR
		-	Dec position in arcsec
R_DECDEG	FLOAT	-	The radial distance between the
		-	sources in decimal degrees
R_ARCSEC	FLOAT	-	The radial distance between the
		-	sources in arcsec
SNR	FLOAT	-	signal-to-noise ratio of raw detection
FLUX	FLOAT	-	Magnitude/Flux of none radio source
LR	DECIMAL(28,9)	-	The Likelihood Ratio value for this
		-	pair or sources
RELIABILITY	DECIMAL(28,9)	-	The Reliability value for this pair
		-	of sources
P_NOT	DECIMAL(15,13)	-	The Poisson Probability value for this pair
		-	of sources **not used**

Working tables are also used in the algorithm for determining Q_0 and the matches table for the cross identifications.

The Fusion catalogue is a very large flat ascii table with many columns. For this work it has not been normalised but should be for future work where the data volumes get significantly larger. Also for much larger datasets thought should be given to the partitioning of the tables.

All objects are provided in the electronic appendix.

3.3 Python Code

The method described in Chapter 3 forms the LRPY algorithm which along with the selection rules presented in Section 3.1.6 have been coded in Python and the full set of files is available from *github* (Weston, 2018) under a GNU General Public License (GPL) V3.0. Anyone is permitted to use this code for research purposes, but we ask that they cite the paper (Weston et al., 2018). What follows is an overview of the Python programs and functions which implement the LRPY algorithm for this work:

TABLE 3.9: Likelihood Ratio Python Program List in order of execution

Name	Description
likelihoodratio_main.py	The main program which calls the following
area_none_radio_survey.py	To define and determine the area for different scenarios such as *local* or *global* defined in Section 3.1.3
radio_pairs.py	Radio pair search
populate_matches.py	Nearest Neighbour Search
f_r.py	Calculate $f(r)$ for NN's (see 3.1.2
n_m.py	Calculate $n(m)$ (see 3.1.3)
q_0.py	To determine $Q0$ (see 3.1.4)
total_m.py	Determine the total surface density of all background sources (see 3.1.4)
real_m.py	Determine the real surface density of background sources (see 3.1.4)
q_m.py	Determine the parameter $q(m)$ for a for a possible candidate
plot_m.py	Plot the figures as seen in Figures 3.6
likelihoodratio.py	Calculate the LR for NN's
reliability.py	Calculate the Reliability for NN's
plot_lr_vs_rel.py	Plot Results

TABLE 3.10: Poisson Probability Python Program List in order of execution

Name	Description
pp_main	The main program which calls the following
area_none_radio_survey.py	To define and determine the area
populate_matches.py	Nearest Neighbour Search
n_m.py	Calculate $n(m)$
mu_c.py	Calculate μ_c
p_c.py	Calculate P_c
mu_star.py	Calculate μ_*
plot_p_not.py	verification plot P_{not}
plot_rel_vs_pnot.py	verification plot Rel vs P_{not}
plot_lr_vs_pnot.py	verification plot LR vs P_{not}

It should be possible to adapt the SQL within the Python code to equate with different table definitions for other survey data. Also in the electronic appendix is further code used for generating the postage stamps of the radio contours over the Spitzer IR images.

3.4 Results and Catalogue

In this section is described the catalogue containing the results of the ATLAS cross-identification with Fusion using the LRPY algorithm discussed earlier in this chapter.

3.4.1 Results of Cross-identification

The results of our cross-identification of the ATLAS catalogue with the Fusion catalogue are presented in the Table 3.11. As we described in the previous Section 2.1.5, and illustrated in Figures 2.6 and 2.7, approximately 96% of the total number of ATLAS sources are covered by the Fusion catalogue, which makes 2922 radio sources in the CDFS field and 2113 in the ELAIS field. So there are in total 5035 radio sources for XID with the Fusion catalogue. Not all radio sources we deal with have Fusion candidates inside the search radius used in this work (6"). This number of blanks is small consisting of 222 for CDFS and 177 for ELAIS. So the number of "candidates" (radio sources with one or more Fusion sources in the search area) drops to 2700 for the CDFS field and 1936 for the ELAIS field.

TABLE 3.11: Results of cross-identification of ATLAS sources with FUSION sources using the LRPY code. We present the total number of radio sources, the number having Fusion coverage, the number with any Fusion counterpart within 6" and the number with high reliability Fusion counter parts per Section 3.1.6 and 5.1.

Field	CDFS	ELAIS-S1	both
Total	3078	2113	5191
with Fusion coverage	2922	2113	5035
with any Fusion candidates	2700	1936	4636
with high reliability XID	2222	1626	3848

We found that a large percentage of these candidate radio sources have just one Fusion source in the search radius $\approx 40\%$ (see Table 3.2). About 40% of (nonblank) radio sources have two or more (up to 5) Fusion sources within the search radius.

Applying the LR criteria for ”single” sources and both LR and Reliability criteria for the situation when two or more Fusion sources are in the search radius, we find that about $\approx 84\%$ of candidates correspond to the criteria that are used in this work for cross-identification in Sections 4.1 and 4.2. The ATLAS sources without secure Fusion counterparts likely have counterparts below the Fusion detection limit.

3.4.2 Catalogue

The information is divided into two tables, one for each field CDFS and ELAIS-S1. Example subsets are given in Table 3.12 for CDFS and Table 3.13 for ELAIS-S1, the entire catalogues for the two fields are provided in ASCII format with the electronic supplement.

TABLE 3.12: ATLAS/FUSION SWIRE Cross-Identification Catalogue for the CDFS field. A description of the table is given in Section 3.4. (This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content, a full copy of the catalogue is available online.)

(1) ATLAS ID	(2) *RA* (J2000)	(3) *Dec* (J2000)	(4) $S_{\rm Int}$ mJy	(5) SWIRE ID	(6) $RA_{\rm IR}$ deg	(7) $Dec_{\rm IR}$ deg	(8) $S_{3.6\mu{\rm m}}$ μJy	(9) $\sigma_{3.6\mu{\rm m}}$ μJy	(10) $\log_{10}({\rm LR})$	(11) Reliability
CI0001C1	52.1516	-28.6982	132.5	432065	52.1520	-28.6988	2565	7.07	-1.154	0.290
CI0002	51.8917	-28.7726	158.151	428929	51.8917	-28.7725	101.84	0.8	2.556	0.999
CI0003	53.5387	-28.4055	74.8	158805	53.5388	-28.4053	67.29	1.09	2.011	0.998
CI0005C1	51.9088	-28.0239	19.8	456752	51.9079	-28.0232	5.16	0.43	-6.966	0.000
CI0005C2	51.9127	-28.0357	0.692	456300	51.9127	-28.0358	24.75	0.81	1.906	0.997
CI0005C3	51.9067	-28.0251	69.6	456683	51.9071	-28.0250	200.42	1.63	1.424	0.993
CI0007	53.9722	-27.4613	118.207	63449	53.9722	-27.4612	398.33	2.01	2.458	0.999
CI0008	52.1943	-28.4379	56.6902	303864	52.1940	-28.4383	68.82	0.88	0.693	0.966
CI0009	53.8646	-27.3308	95.6314	209688	53.8646	-27.3307	157.62	1.77	2.610	0.999
CI0010	53.6121	-27.7338	55.7524	190007	53.6121	-27.7338	201.51	1.55	2.546	0.999

TABLE 3.13: ATLAS/FUSION SWIRE Cross-Identification Catalogue for the ELAIS-S1 field. A description of the table is given in Section 3.4. (This table is available in its entirety in a machine-readable form in the online journal. A portion is shown here for guidance regarding its form and content, a full copy of the catalogue is available online.)

(1) ATLAS ID	(2) *RA* (J2000)	(3) *Dec* (J2000)	(4) $S_{\rm Int}$ mJy	(5) SWIRE ID	(6) $RA_{\rm IR}$ deg	(7) $Dec_{\rm IR}$ deg	(8) $S_{3.6\mu{\rm m}}$ μJy	(9) $\sigma_{3.6\mu{\rm m}}$ μJy	(10) $\log_{10}({\rm LR})$	(11) Reliability
EI0001	9.06966	-43.15964	160.032	87322	9.06957	-43.15968	103.24	1.53	2.565	0.999
EI0002C2	7.87001	-43.68909	193.85	373161	7.87013	-43.68917	46.44	0.96	2.038	0.998
EI0003	8.28863	-43.99076	69.6057	220919	8.28844	-43.99079	141.9	1.8	2.554	0.999
EI0004C1	8.67872	-43.50959	49.97	237221	8.67851	-43.50948	89.18	1.45	2.352	0.999
EI0004C3	8.67261	-43.51230	0.538	237403	8.67237	-43.51141	228.49	2.27	-3.161	0.003
EI0005	8.01844	-44.19189	35.5887	350583	8.01828	-44.19192	48.61	0.84	2.042	0.998
EI0006	8.20550	-44.36404	39.7013	341374	8.20543	-44.36403	373.25	2.28	2.420	0.999
EI0007	9.34721	-44.37919	44.6797	159474	9.34722	-44.37909	51.05	0.81	2.308	0.999
EI0008	9.19112	-43.09654	30.4775	87851	9.19100	-43.09654	10.76	0.68	1.705	0.996
EI0009C1	9.32550	-44.50327	49.35	162876	9.32449	-44.50391	133.17	1.3	-7.356	0.000

The cross identification catalogue columns in the preceding tables are organized as follows:

Column (1) - ATLAS DR3 Identification number of the radio source "cid"

Column (2) - Right Ascension (J2000) of the radio source, decimal degrees

Column (3) - Declination (J2000) of the radio source, decimal degrees

Column (4) - Integrated Radio Flux Densities (mJy) at 1.4 GHz

Column (5) - Fusion Identification number "swire_index_*Spitzer*"

Column (6) - Right Ascension (J2000) of the IR candidate, decimal degrees

Column (7) - Declination (J2000) of the IR candidate, decimal degrees

Column (8) - IR Flux Density at 3.6 μm (μJy)

Column (9) - IR Flux Density Uncertainty at 3.6 μm (μJy)

Column (10) - Log_{10} of Likelihood Ratio of the IR candidate

Column (11) - Reliability of the IR candidate

These two tables are available in their entirety including all Fusion sources within 6" of an ATLAS source in a machine-readable format in the supplementary material. No filtering on Reliability or Likelihood Ratio has been undertaken. Column 10 has been presented with its Log_{10} value to make the column width manageable, also Column 11 is presented to three decimal places. Both columns in the supplementary material will be presented to a higher precision. From our findings in Section 3.1.6 we recommend the following selection criteria for accepting identifications: $LR_c \geq 0.01$ and the Reliability is greater than R limit given by using Equation 5.2 with $\beta = 0.4$. The data is also available as a series of normalised relational database tables where, by using the index columns "cid" and "swire_index_spitzer" as the relationship to join the tables, it is possible for the reader to construct their own version of the catalogue or work with the data in other ways.

3.5 Summary

This chapter started by discussing in detail the implementation of the LR method for the work in Section 3.1. A different approach to previous work by others for the selection of matches has been presented in Section 3.1.6, which uses a combination of the Reliability and LR values.

It was noticed that there are some single Radio sources that have multiple IR components with similar Reliability values (centered about $0.3 < R < 0.7$). It had been noticed and reported in prior work that maybe there was perhaps some local clustering of IR sources merging to form one apparent source of Radio emission. This scenario is explored in a later chapter which confirms that this is in fact the case for some of these identified cases.

The LR method works on a principle of matching in this case one Radio source to one IR candidate, what it can not handle is the possibility that a Radio source can have more complex morphology such as bright side lobes and multiple components for one IR source (Galaxy). Later in this work we will present a possible adaptation of this LR algorithm to accommodate at least Radio sources with two components being matched to one IR candidate.

Also presented in Sections 3.2 and 3.3 was how the algorithm was implemented using a Relational Database and Python code.

Finally in Section 3.4 from this work is presented the resultant catalog of matched sources between ATLAS and Fusion with a detailed description. No selection has been made, all matches are provided even if they don't match the selection criteria chosen in this work. This allows others to use the catalogue and define there own selection criteria or other extensions and modifications in the post processing.

I have spent a large part of my time at AUT operating the Warkworth 12m dish for IVS, LBA and SpaceX. Credit: Stuart Weston

Chapter 4

Comparison of XID methods

In this chapter I review the two different methods of cross identification and conduct some comparison between them. In the first section 4.1 I look at the Likelihood Ratio with respect to a simple nearest neighbour method. In the second section 4.2 I compare the Posisson Probability method again with a simple nearest neighbour method and in addition compare with the Likelihood Ratio.

4.1 Comparing Nearest Neighbour to Likelihood Ratio

In the following sub-sections we will compare the two methods Nearest Neighbour (NN) and Likelihood Ratio (LR) first statistically and then visually. For the statistical comparison I look at the global statistics from the various methods. For the visual comparison I look at the IR images with the ATLAS radio contours overlaid for the candidates, of which a subset are presented in this work to show the different scenarios.

4.1.1 Statistical Comparison

The initial step in my LR algorithm is to find all Fusion sources within a specified search radius 6" of the ATLAS Radio Sources. Due to the high density of sources within the

Fusion catalogue there can be $1 \leqslant n \leqslant 5$ possible candidates. For this work of matching the two catalogues ATLAS to Fusion I obtained the following.

For Nearest Neighbour where $n = 1$ (only one possible match) as cross identifications if a purely positional closeness was a valid selection criteria, then we would expect a very high number of these to have high LR values certainly above the selected cut-off as just using a NN approach all of these would be selected as cross matches. Smith et al. (2011) showed that for sources where $n = 1$ when looking at their LR values only $\approx 48\%$ of these had a Reliability ($R \geqslant 0.8$) above their selected cut-off ($R_c = 0.8$), demonstrating an advantage of the LR technique over a simple NN algorithm. It should be noted here that though Reliability (R) was initially introduced for the case of several candidates in the search radius (Equation 3.15), it can be used formally for the case of one IR source ($n = 1$) in the search radius. Even in this case R can be $\ll 1$ if the LR of the candidate source is $\ll 1 - Q_0$ as was highlighted earlier with Equation 3.16.

Using the ATLAS to Fusion 6" search radius candidates and our proposed selection criteria in Section 3.1.6, are presented the following breakdown in Tables 4.1 (for the CDFS field) and 4.2 (for the ELAIS field) of the distribution of the number of Fusion candidates to the ATLAS DR3 sources, and the fraction of reliable counterparts. The first column is the number of matches; the second column is the number of ATLAS sources with the corresponding number of $n(Matches)$ with Fusion for CDFS; the third column is the number of sources that meet the selection criteria (defined in Section 3.1.6); The fourth column is the corresponding percentage of sources (column three to column two).

Taking the table data and plotting $n(Matches)$ against the percentage figures in the last column we obtain the plot in Figures 4.1 and 4.2. In these plots we can see a downward trend in the percentage of matches meeting the cutoff selection criteria (defined in Section 3.1.6), at $n(Matches) = 4$ for both fields we see an anomaly to this trend. But at $n(Matches) \geqslant 4$ we have very low numbers and therefor the percentages are unreliable.

TABLE 4.1: The 6″ distribution of the number of Fusion candidates to the ATLAS DR3 sources in the CDFS field, and the fraction of reliable counterparts. The first column is the number of matches; the second column is the number of ATLAS sources with the corresponding number of n(*Matches*) with Fusion; the third column is the number from column 2 that meet the selection criteria defined in Section 3.1.6. Column 4 is the percentage of column 3 to column 4.

$n(Matches)$	$n(ATLAS_{CDFS})$	n($R \geqslant R_{cutoff}$)	%
1	1657	1285	77.5
2	832	592	71.1
3	185	129	69.7
4	23	10	43.5
5	3	2	66.7
Totals	2700	2018	74.7

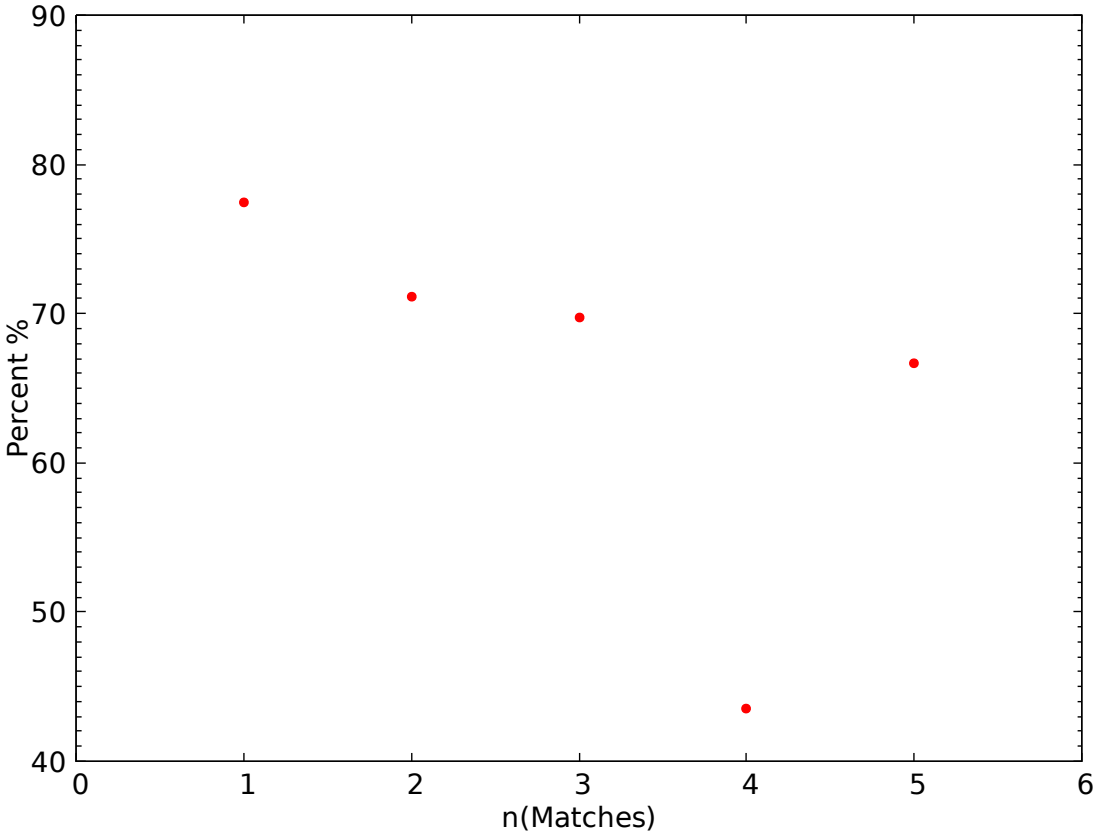


FIGURE 4.1: This figure shows the values for column 4 against column 1 from Table 4.1 for the CDFS field.

TABLE 4.2: The 6″ distribution of the number of Fusion candidates to the ATLAS DR3 sources in the ELAIS field, and the fraction of reliable counterparts. The first column is the number of matches; the second column is the number of ATLAS sources with the corresponding number of n(*Matches*) with Fusion for ELAIS; the third column is the number from column 2 that meet the selection criteria defined in Section 3.1.6. Column 4 is the percentage of column 3 to column 4.

$n(Matches)$	$n(ATLAS_{ELAIS})$	n($R \geqslant R_{cutoff}$)	%
1	1157	928	80.2
2	615	472	76.7
3	138	89	64.5
4	24	16	66.6
5	2	1	50.0
Totals	1936	1506	77.8

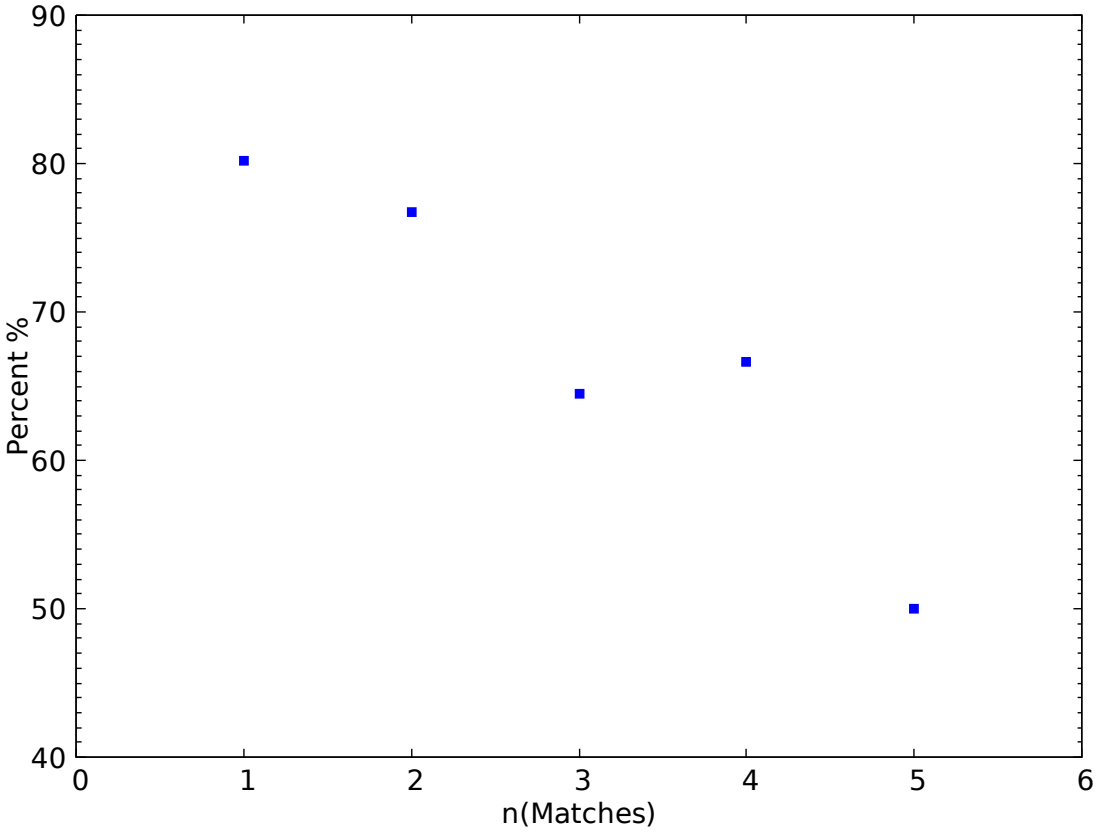


FIGURE 4.2: This figure shows the values for column 4 against column 1 from Table 4.2 for the ELAIS field.

Combining the two fields to get slightly better statistics I present in Table 4.3. There is a higher figure of ≈ 77 to 80% between the two fields for where n = 1 compared to Smith et al. (2011), if this is applied to much larger catalogues as expected with EMU we would have a very large number of sources incorrectly cross-matched. The difference between our result and Smith et al. (2011) is that they took a hard cut-off for matches where R$>$ 0.8, where as my cut-off defined in section 3.1.6 is more relaxed and dependant on R and LR together.

TABLE 4.3: The 6" distribution of the number for all Fusion candidates to the ATLAS DR3 sources in both fields field, and the fraction of reliable counterparts. The first column is the number of matches; the second column is the number of ATLAS sources with the corresponding number of n(*Matches*) with Fusion; the third column is the number from column 2 that meet the selection criteria (defined in Section 3.1.6). Column 4 is the percentage of column 3 to column 2.

n(*Matches*)	n(ATLAS)	n($R \geqslant R_{cut-off}$)	%
1	2814	2213	78.6
2	1447	1064	73.5
3	323	218	67.5
4	47	26	55.3
5	5	3	60.0
Totals	4636	3524	76.0

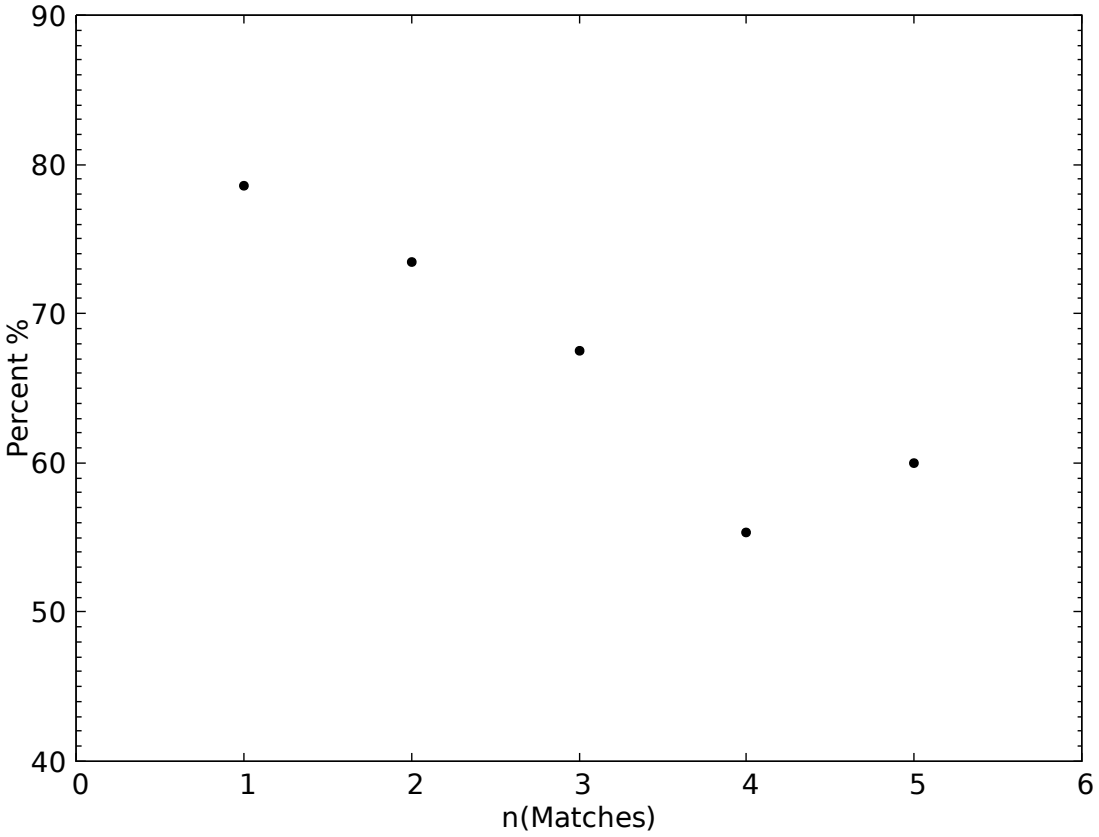


FIGURE 4.3: This figure shows the values for column 4 against column 1 from Table 4.3 presented above for the combined fields.

4.1.2 Visual Comparison

To demonstrate the LR results following is Table 4.4 for a very small sub-set of the sources to demonstrate different scenario's with some of their catalogue and LR properties. Following the table are corresponding postage stamp images (Figures 4.4 and 4.5) of these same sources to visually show the various scenarios. In the postage stamp images are shown the radio contours (green) over the infrared image (grey scale) with the NN candidate marked in yellow and the other candidate's with in the 10" search radius marked in magenta. The small open red circle marks the catalogue position of the ATLAS radio source, and the larger open red circle is the original 10" search radius.

TABLE 4.4: Table of the properties that go with the postage stamp images presented in Figures 4.4 and 4.5.

ATLAS ID	Flux mJy	SWIRE ID	Ang Sep arcsec	Reliability
EI0011	260.46	378652	0.32	0.99
EI0030C1	27.41	251483	4.17	0.000000029
EI0028	15.04	367874	3.15	0.0000012
EI0028	23.31	367909	0.28	0.99
EI0020C1	6.5	245088	5.30	0.0
EI0020C1	28.5	245157	1.70	0.90
EI0020C1	8.81	245158	4.14	0.000000002
CI0014C2	48.31	194762	4.60	0.0
CI0026C1	25.84	346970	4.13	0.000000058
CI0026C1	17.07	347008	2.66	0.024
CI0032C1	210.78	148504	5.54	0.0
CI0033	10.56	174120	0.25	0.99

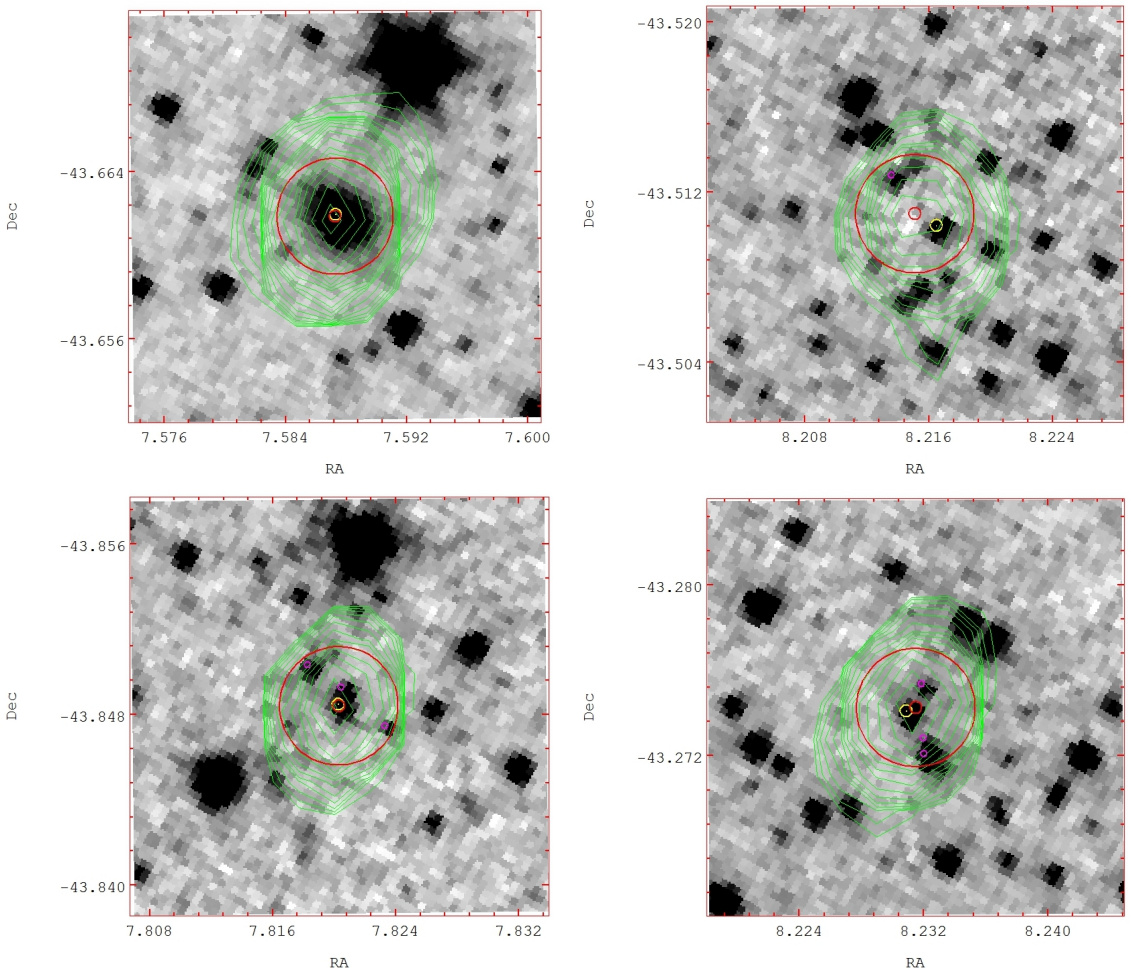


FIGURE 4.4: ELAIS_S1 - Postage stamps images of radio contours over IR image with LR candidate marked in yellow and the NN candidate's marked in magenta. The small open red circle marks the catalogue position of the ATLAS radio source, and the larger open red circle is the original 10" search radius. Using the ATLAS source catalogue identification labels top left is EI0011, top right is EI0030C1, bottom left is EI0028 and bottom right is EI0020C1.

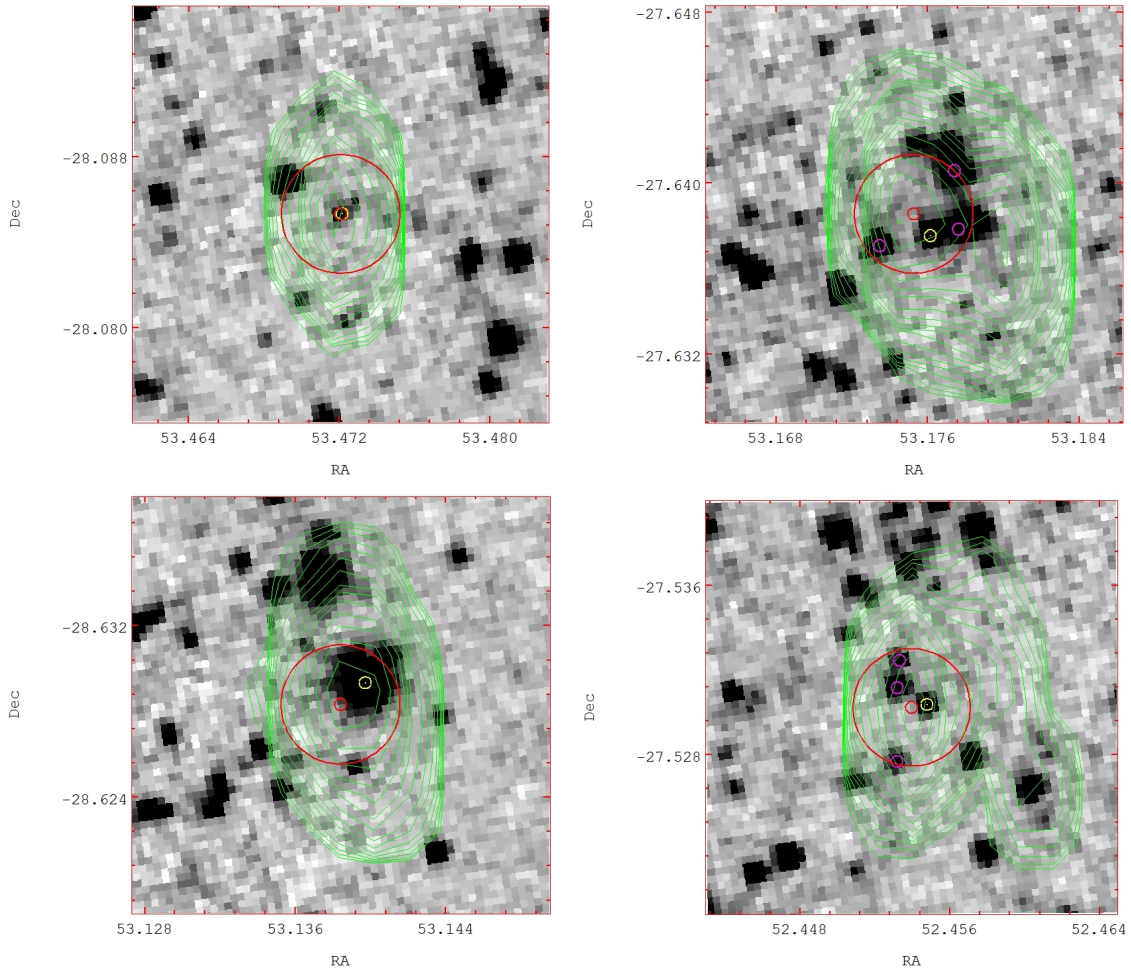


FIGURE 4.5: CDFS - Postage stamps images of radio contours over IR image with LR candidate marked in yellow and the NN candidate's marked in magenta. The small open red circle marks the catalogue position of the ATLAS radio source, and the larger open red circle is the original 10" search radius. Using the ATLAS source catalogue identification labels top left is CI0033, top right is CI0014C2, bottom left is CI0032C1 and bottom right is CI0026C1.

Taking the first set of postage stamp images for the ELAIS field shown in Figure 4.4, I will now discuss them in detail order based on their complexity using information also from Table 4.4.

The first image of ATLAS source **EI0011** shows a very clean scenario, there is only one IR candidate with in 0.32" of the ATLAS catalogue position. Both methods NN and LR would select this as a match, in this case the LR assigns a reliability of 0.9996 to this candidate and would pass the LR selection criteria (see Section 3.1.6) used in this work.

For the ATLAS source **EI0030C1** we have a different scenario, although there are two IR candidates within the initial 10" search radius only one lies within the final 6" search radius. This IR candidate lies 4.17" from the ATLAS catalogue position. Depending on the NN limit that might be selected this is the nearest neighbour, but the LR assigns this IR source a very low reliability of 2.9×10^{-8} and as such this would fail the LR selection criteria used. This clearly demonstrates a difference between NN and LR methods for identifying matches.

This next image of **EI0028** is a little more complex, there are four IR candidates within the initial 10" search radius but only two are within the final 6" search radius chosen. Here the NN candidate at a distance of 0.27" also passes the LR selection criteria with a reliability of 0.9976.

In the final image of **EI0020C1** there are five IR candidates within the initial 10" search radius but only three are within the final 6" search radius. Here the nearest neighbour at a distance of 1.7" also passes the LR selection criteria with a reliability of 0.9, the other two candidates have a very low reliability (one of which doesn't meet the precision of the system).

The next set of postage stamp images for the CDFS field shown in Figure 4.5 I will now discuss what they show similar to the above section.

In the first image for **CI0033** there is a clean field, there is only one IR candidate which lies 0.25" from the ATLAS source catalogue position. Again

NN would select this as match, in this case LR assigns a reliability of 0.993 to this candidate which would pass my LR selection criteria. So here both methods would agree.

In the next image for **CI0014C2** there is a complex field, within the initial 10" search radius there are four possible IR candidates, of which only one lies within the final 6" angular separation of cut-off used in this work at 4.6". Using NN this IR candidate would be selected as a match but with the LR it's reliability is very low and effectively zero for the precision in the table. Thus this IR candidate is not identified as a match to the ATLAS source.

In the image for **CI0032C1** there is only one IR candidate that lies 5.53" from the ATLAS source catalogue position. Using NN this IR candidate would be selected as a match but with the LR it's reliability is very low and effectively zero for the precision in the table. Thus this IR candidate is not identified as a match to the ATLAS source using the LR.

Finally for the image of **CI0026C1** can be see another complex scenario of four IR candidates within the initial 10" search radius, but only two of these are selected using our 6" angular separation of cut-off. Of these two IR candidates one lies at 2.66" from the ATLAS catalogue position with a LR reliability of 0.024 and the other lies at 4.13" from the ATLAS catalogue position with a LR reliability of 5.8×10^{-8}. The IR candidate at 2.66" is the NN but due to its low LR reliability value is not selected as a match in this work.

4.2 Comparing Poisson Probability to Likelihood Ratio

In this section are presented Tables 4.5 and 4.6 where a similar comparison as was made for the LR vs NN in the previous sections of this chapter is here also made between NN and PP. The columns for these tables are: first the number of matches within the 10" search radius n($matches$); next is $N(IRAC_{3.6})$ which is the number of Spitzer IRAC candidates within 10" of the ATLAS DR3 sources for the corresponding n($matches$); and the fraction of reliable counterparts where $P_{not}(\leqslant 0.1)$; and the last column, the ratio (%) of $P_{not}(\leqslant 0.1)/N(IRAC_{3.6})$ where the ATLAS DR3 24NOV2014 data release is used. Also presented with the Tables are the Plots 4.6 and 4.7 for % against n($matches$).

In the Tables following it will be noticed that for $P_{not}(\leq 0.1)$ there is a far higher number of possible matches compared to the LR method, for example where n($matches$) $= 1$ in CDFS there are 672 sources within 10" and of these with the LR method we have 397 matches with a $R \geq 0.8$ (61 %) but for PP we have 598 (89 %). For the ELAIS field where n($matches$) $= 1$ there are 477 sources within 10" and of these with the LR method we have 382 matches with a $R \geqslant 0.8$ (80 %), but for PP we have 409 (86 %) not as significant a difference compared to the CDFS results. The same conditions affecting the total under column 2 from the previous section also applies here.

Looking at Figure 4.6 which corresponds with Table 4.5 the plot shows a steady increase in the % from $1 \leqslant n \leqslant 3$ up to a value of 95%. It stays constant untill $n = 5$ after which we see a slight drop and for $n = 8$ the low numbers make the figure unreliable. For Figure 4.7 which corresponds with Table 4.6 the plot again shows a steady increase in the % fron $1 \leqslant n \leqslant 5$ up to 99% after this point we see a drop down to $\approx 70\%$ and again at $n = 8$ the low numbers make the result unreliable. In Figure 4.8 which corresponds with Table 4.7 I have combined both fields for reference.

TABLE 4.5: The distribution of the number of Spitzer IRAC candidates within $10''$ of the ATLAS DR3 sources in the CDFS field, and the fraction of reliable counterparts where $P_{not} \leq 0.1$, using DR3 24NOV2014 data.

$n(Matches)$	$N(IRAC_{3.6})$	$P_{not}(\leq 0.1)$	%
0	263		
1	672	598	89
2	937	861	92
3	647	613	95
4	357	339	95
5	152	145	95
6	37	34	92
7	11	10	91
8	2	2	100
Totals	3078	2602	84

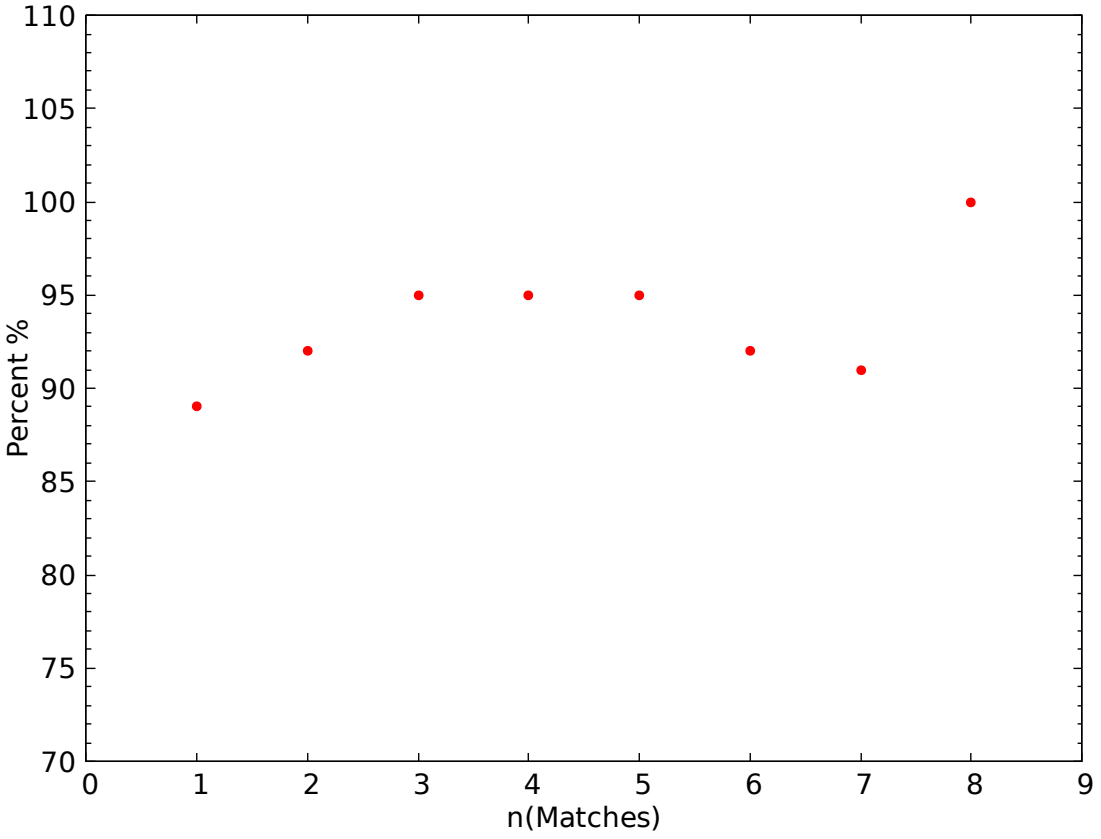


FIGURE 4.6: In this figure are plotted the values for column 4 against column 1 from the above Table 4.5 for the CDFS field.

TABLE 4.6: The distribution of the number of Spitzer IRAC candidates within $10''$ of the ATLAS DR3 sources in the ELAIS S1 field, and the fraction of reliable counterparts where $P_{not} \leq 0.1$, using DR3 24NOV2014 data.

$n(Matches)$	$N(IRAC_{3.6})$	$P_{not}(\leq 0.1)$	%
0	107		
1	477	404	84
2	705	659	87
3	466	434	93
4	241	228	94
5	98	97	99
6	29	22	76
7	8	6	75
8	2	2	100
Totals	2133	1852	87

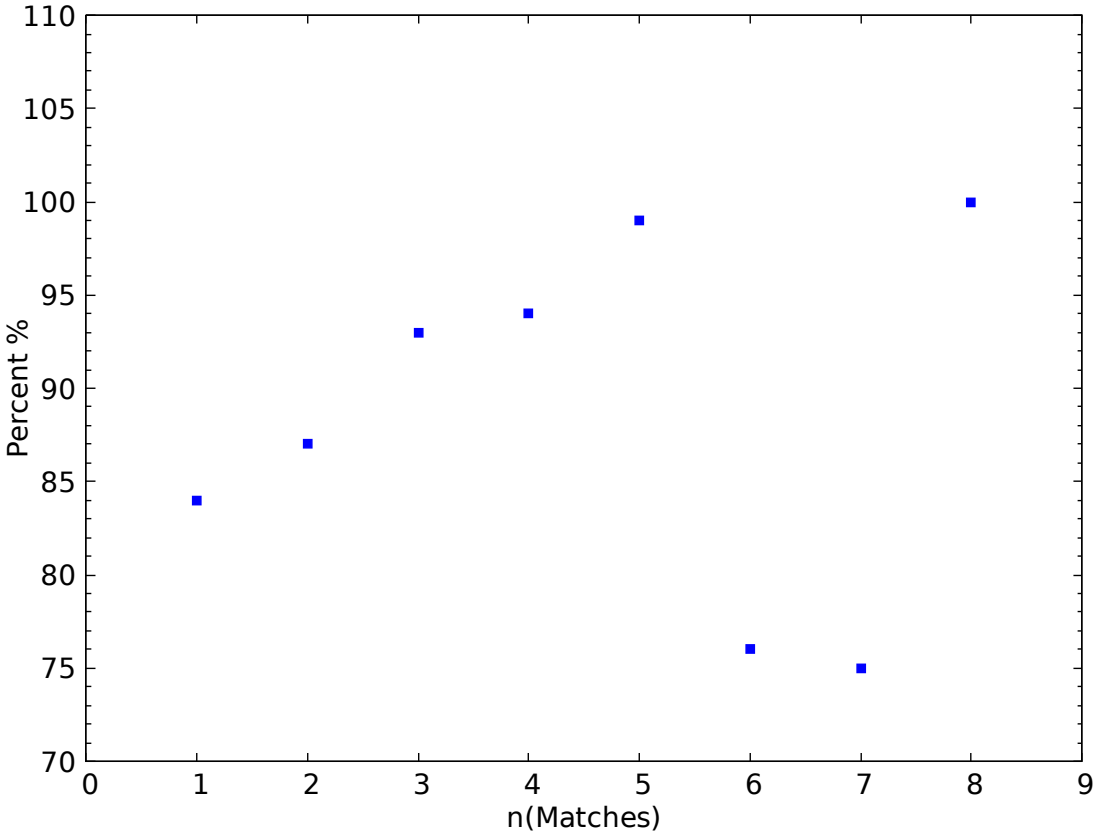


FIGURE 4.7: In this figure are plotted the values for column 4 against column 1 from the above Table 4.6 for the ELAIS field.

TABLE 4.7: The distribution of the number of Spitzer IRAC candidates within $10''$ of the ATLAS DR3 sources in both fields, and the fraction of reliable counterparts where $P_{not} \leq 0.1$, using DR3 24NOV2014 data.

$n(Matches)$	$N(IRAC_{3.6})$	$P_{not}(\leq 0.1)$	%
0	370		
1	1149	1002	87
2	1642	1520	93
3	1113	1047	94
4	598	567	95
5	250	242	97
6	66	56	85
7	19	16	84
8	4	4	100
Totals	5211	4454	85

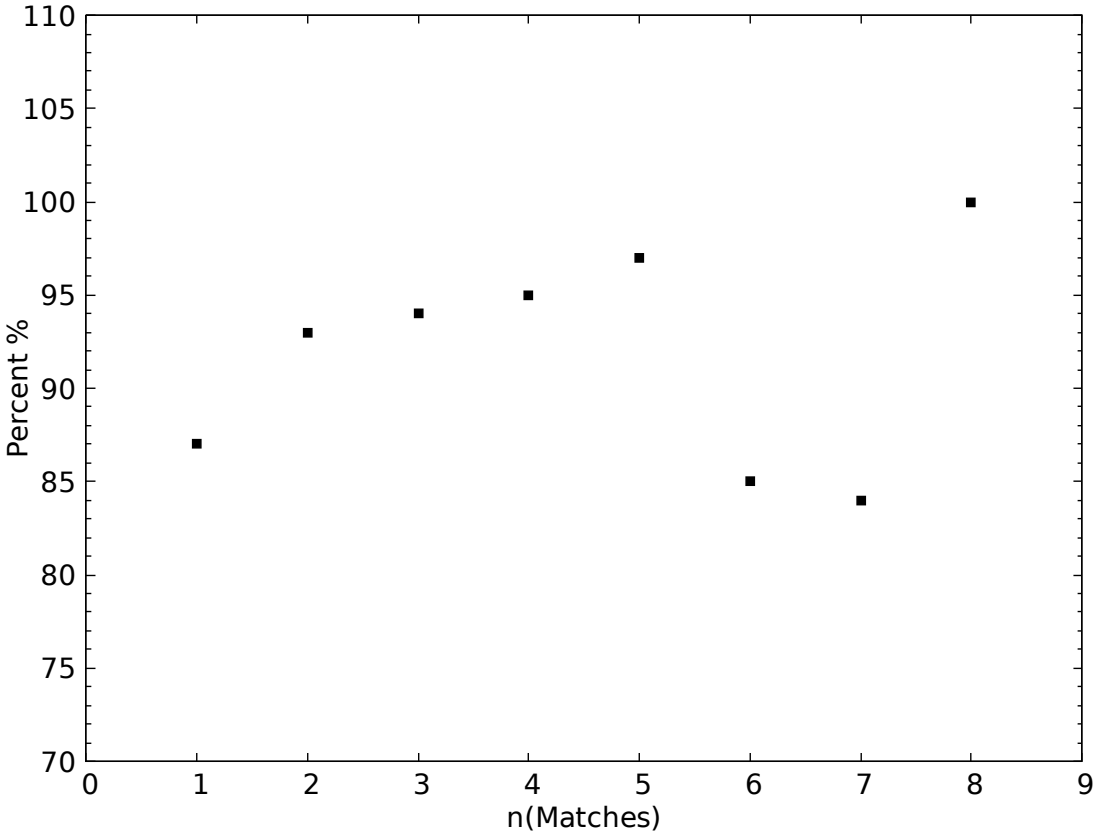


FIGURE 4.8: In this figure are plotted the values for column 4 against column 1 from the above Table 4.7 for both field.

In the following Figure 4.9 I present plots of the P_{not} vs LR values for the two fields. On these figures I have marked the Likelihood Ratio limit (LR $= 0.01$) discussed in Section 3.1.6 with a red dot-dashed line, and the the PP reliable counterparts limit ($P_{not} = 0.1$) from earlier in this section with a red dashed line. It is not possible to define a clear cut off for candidates with a P_{not} below a value to identify them as highly probable candidates by comparing with the Reliability method, without some quite significant contamination. For completeness I include Figure 4.10 where both fields have been combined into one plot.

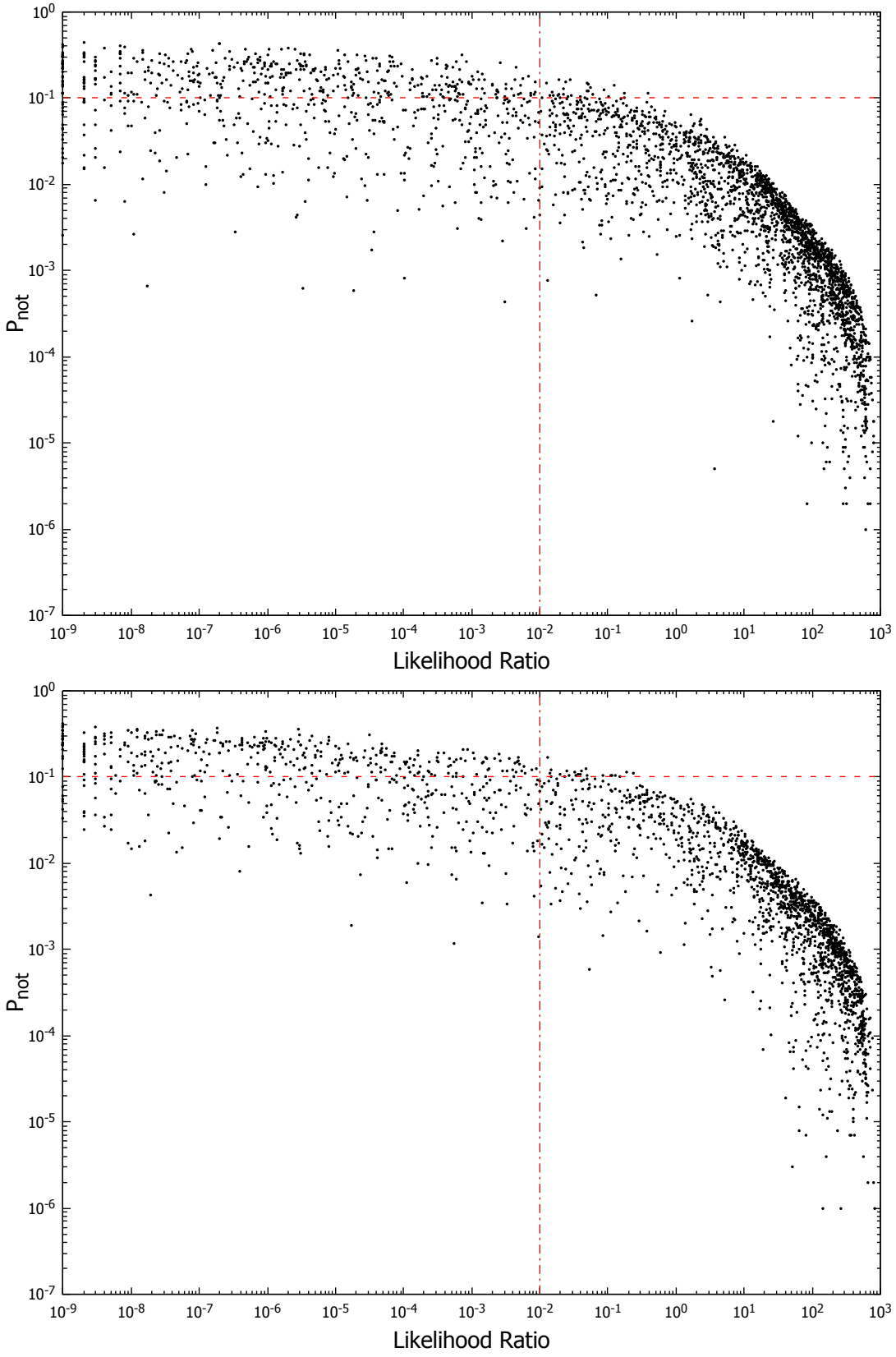


FIGURE 4.9: The top plot is of the P_{not} vs LR for all of the potential IR candidates to the radio sources from the ATLAS DR3 CDFS field using a 6" search radius. The bottom plot is the same but for the ATLAS DR3 ELAIS field using a 6" search radius. On both plots I have marked LR = 0.01 (the proposed LR selection criteria from Section 3.1.6 with a red dot-dashed line and $P_{not} = 0.1$ (the PP reliable counterparts limit from earlier in this section) with a red dashed line.

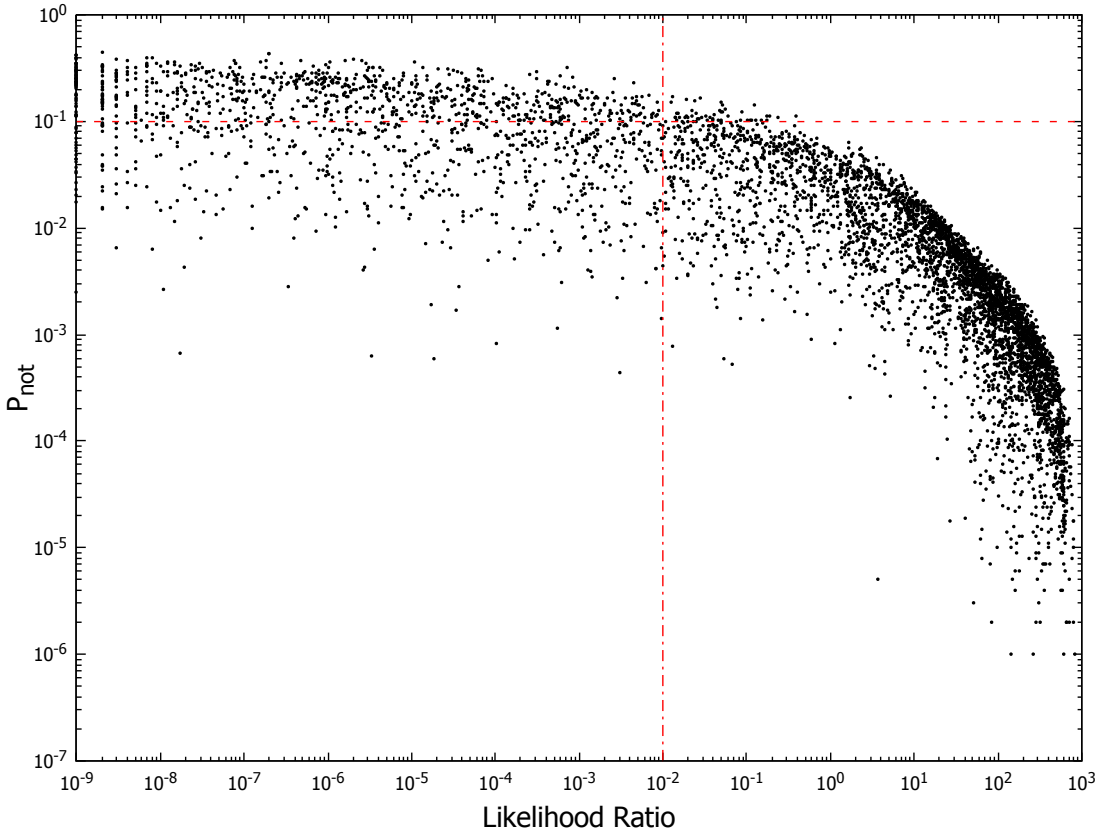


FIGURE 4.10: In this figure are plotted the values for all ATLAS DR3 IR candidates using a 6" search radius. On the plot I have marked the $P_{not} = 0.1$ with a red dashed line and LR $= 0.01$ with a red dot-dashed line.

In another attempt to compare the two methods, in Figure 4.11 I present the R vs LR plot as presented earlier, but I have binned the candidates by the P_{not} values using a different colour and symbol for each bin as shown in the legend of the figure. It can be seen that there are in affect three groups, first are those for $P_{not} > 0.1$. We have a middle group for $0.1 > P_{not} > 0.01$, and a final group for $P_{not} < 0.01$, these limits have been placed on the plots in Figure 4.9 for reference.

The PP method in its current form fails to distinguish between multiple possible candidates within the search radius, a highly probable scenario (as demonstrated in Tables 4.5 and 4.6) given the depth of the two catalogues being cross matched.

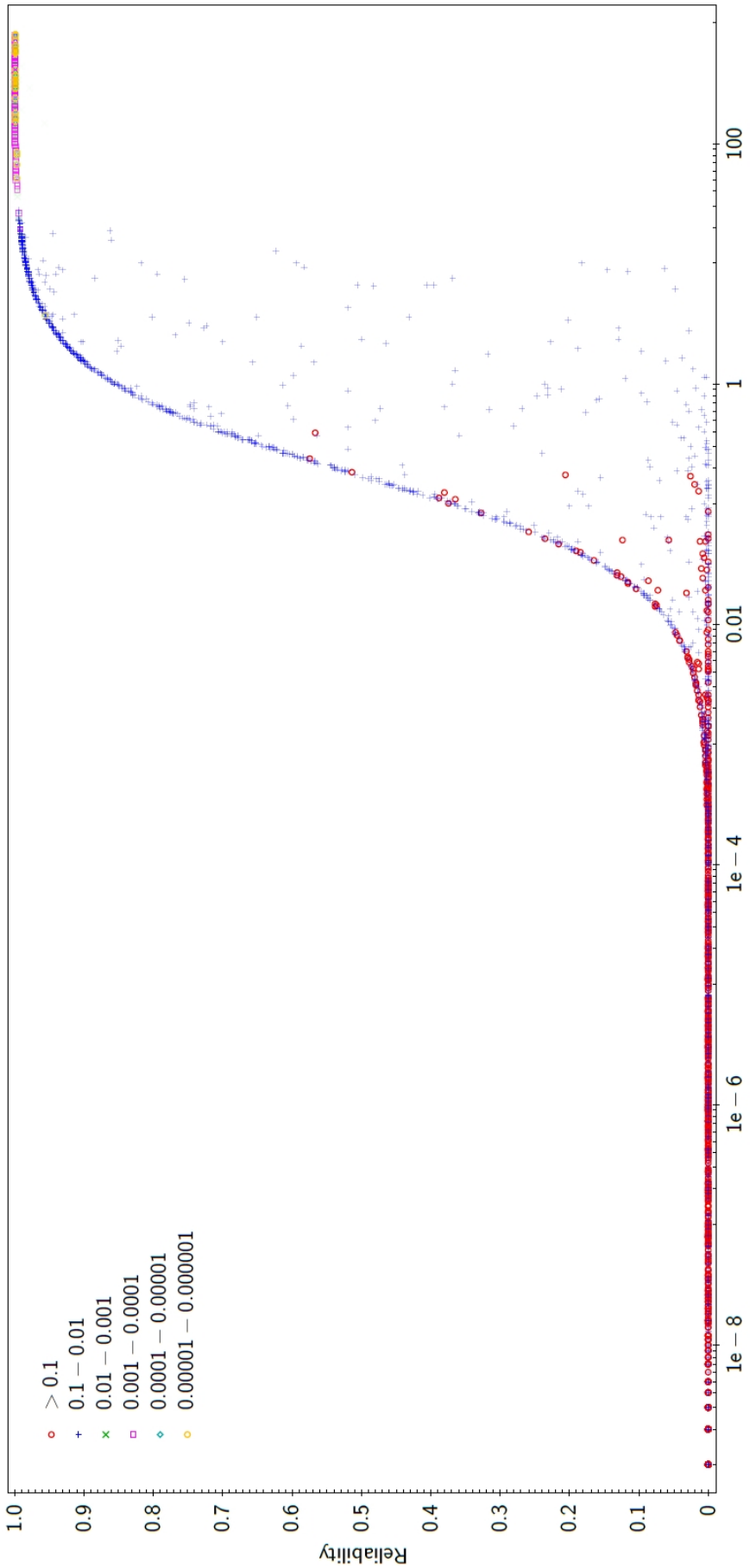


FIGURE 4.11: A plot of R vs LR of ATLAS DR3 with the corresponding P_{not} value bins colour coded by value ranges as shown in the legend.

4.3 Summary

In this chapter I reviewed the two different methods of cross identification and conducted some comparison between them. In the first section I looked at the Likelihood Ratio with respect to a simple nearest neighbour method, I also provided a visual comparison for a sub-set of the sources to demonstrate the different scenarios. Then in the second section I compared the Poisson Probability method again with a simple nearest neighbour method, then compared with the Likelihood Ratio results.

Taking a nearest neighbour approach for those with ATLAS sources with n = 1 Fusion candidates in Table 4.3, we see only 78% pass the LR selection criteria in this work. Several visual examples have also been presented in Section 4.1.2 to highlight this situation and indicate a shortcoming of using a NN approach. Another failing of the NN method is that it can not distinguish between multiple possible candidates within the search radius, it just selects the candidate closest which I have attempted to show might not be the ”true” candidate.

Also I have attempted to shown how the PP method in its current form fails to distinguish between multiple possible candidates within the search radius, a highly probable scenario given the depth and density of the two catalogues being cross matched. In comparison by using the Reliability in the LR method I am able to assign a value to each of the possible candidates and make a quantitative selection based on these values. Also the P_{not} values do not provide any extra information which the LR does, which is used in the following Chapter 5 for the identification of more complex radio source morphology or a radio source resulting from a merging of the radio emission due to insufficient angular resolution.

I worked with the team from the Institute of Radio Astronomy and Space Research, AUT University between 2010 and 2014 converting this former satellite communications 30m dish of Telecom NZ to a Radio Telescope Woodburn et al. (2015). Credit: Stuart Weston

Chapter 5

Refining the Likelihood Ratio Technique

In this Chapter I present two extensions to the Likelihood Ratio method. As was shown earlier in Section 3.1.6 Figure 3.8 there is a symmetry of data points around $R \approx 0.5$. The first section of this chapter deals with the investigation of this symmetry and the resultant identifying of potential IR Doubles (IRD) or clusters of sources in the Fusion (shorter wavelength and higher angular resolution) catalogue being cross matched with the ATLAS Radio catalogue (longer wavelength and lower angular resolution catalogue). Some of this work in the first section of this chapter was published in Weston et al. (2018) and is indicated where appropriate. There are a small number of ATLAS radio sources with complex morphology which have been visually selected from the radio images by experienced investigators. For a small catalogue and image area this is practical, but with the future surveys planned this will not be possible due to the number of objects (10's of millions) in the catalogues and the survey areas (10's of thousand square degrees). I explore if it is possible to use an algorithm with the LR data as a side product to select some or all of these objects from the ATLAS radio survey, as these future surveys will require cross matching. The second section describes a method to identify possible Radio Doubles and Multiple Radio candidates as an extension to the LR and is currently a paper in preparation.

5.1 Multiple IR Candidates

When reviewing the R vs LR plots in Section 3.1.6 Figure 3.8 it was noticed that there is symmetry of some data points in pairs about $0.2 \leqslant R \leqslant 0.8$, these are a small subset of 26 sources (0.8 %) in CDFS and 36 sources (1.7 %) in ELAIS. To better demonstrate this in Figure 5.1 the R vs LR plot is reproduced just for the region of $0.2 \leqslant R \leqslant 0.8$ the area of interest in this discussion. These sources also demonstrate a possible limitation of the LR method, since the method implicitly assumes that there is only one true counterpart to a given source for cross matching. These objects represent a small numeric and thus percentage of the total catalogue, but if this was applied to much larger catalogues or surveys this would become a significant number of missed sources.

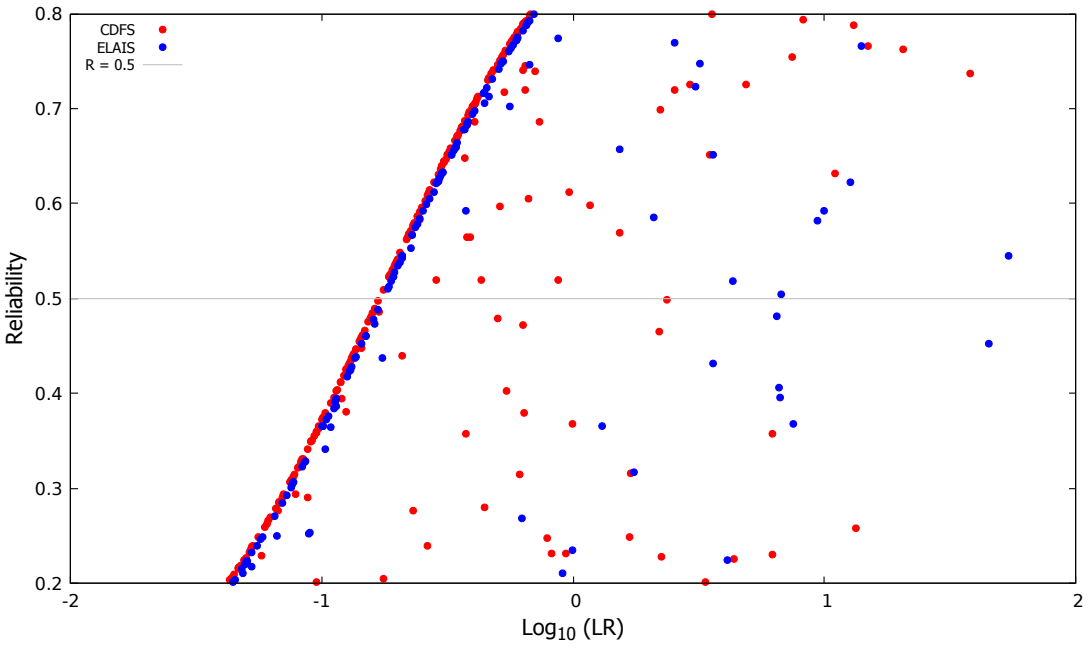


FIGURE 5.1: This is a plot of R vs LR for all possible matches within the 6" search radius for both fields. We have zoomed in to a region centered on $0.2 \leqslant R \leqslant 0.8$ and $-2.0 \leqslant Log_{10}(LR) \leqslant 2.0$ to better demonstrate the symmetry observed.

In Smith et al. (2011) they had noted a similar area in their catalogue and suggested that these could be due to multiple interacting counterparts, as four of their sources had multiple counterparts with a spectroscopic redshifts $\Delta z \lesssim 0.001$. Also Fleuren et al. (2012) highlighted these possible multiple counterparts, and propose that these could be either merging galaxies or members of the same cluster. Fleuren et al. (2012) found

matches to 37 sources (out of 1444) with a mean redshift difference of 0.0011 with a maximum difference of $\Delta z = 0.0187$. Upon investigation of the candidates within the XID catalogue produced by this work in Section 3.4, some of the pairs about $R = 0.5 \pm 0.3$ were found to be due to a radio source with two IR candidates of similar flux density and similar angular separation from the radio source. Thus the possibility was considered that these are interacting pairs of IR sources which could be contributing approximately equally to the radio emission and be grouped together forming what I call in this work a InfraRed Double (IRD).

To explain this consider the situation where there are multiple ($\geq$ 2) IR candidates with similar LR and they are ”close” to each other; I consider the possibility that these are local clusters of IR sources or more specifically an IRD (galaxy merger) both contributing to the Radio emission. To demonstrate this refer to Figure 5.2 where on the left there are two IR sources at a similar distance from the radio source but diagonally opposed about the possible radio source. On the right are two IR sources with similar reliability and distance from radio sources and with a high possibility that they are merging or part of a galaxy cluster both contributing to the radio emission. The latter of the two could be due to a projection affect of a near and far source being in chance alignment. These two possibilities can be destinguished if the redshifts of the two IR sources were available.

The criteria I have chosen to select for possible IRD’s are: the IR sources have a close position to each other (ie not opposed to each other as in the left hand side of Figure 5.2), similar distance from the radio source; similar IR flux which could indicate they also have a similar redshift z.

To demonstrate this I present two postage stamp images in Figure 5.3. On the left hand side two IR sources are very close to each other marked 1 and 2. In the right hand side image we have two IR sources (marked 1 & 3) at similar distances from the radio center but opposed to each other with the radio source as the center.

One ATLAS radio source due to its unresolved peak in a low-resolution radio image could potentially be produced by two or more radio sources blended into one apparent “source” by the large radio beam. I will now proceed to explore the possibility that these are local clusters of IR sources which could be contributing to the one radio emission. In

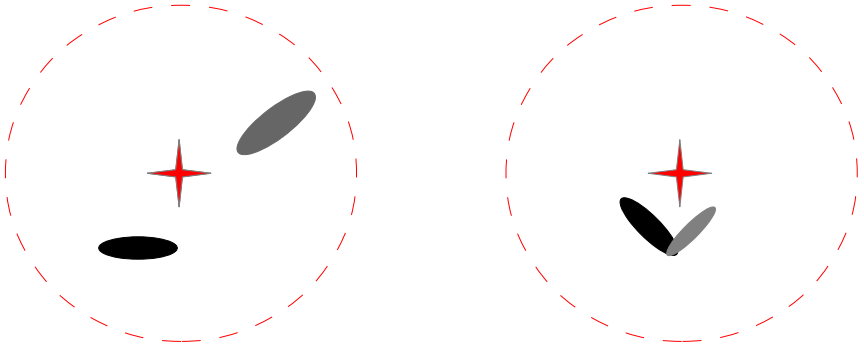


FIGURE 5.2: In the above figures we have a radio source represented by a red four pointed star, two IR sources represented by the ellipses and the defined search radius with a dashed red lines.

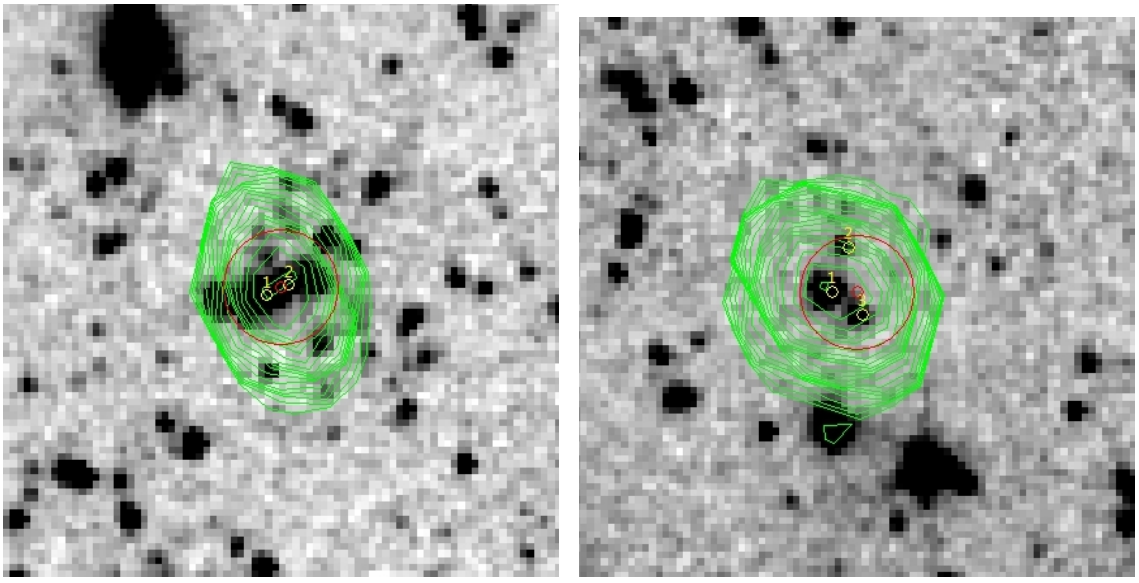


FIGURE 5.3: This shows two real scenarios as described and presented in Figure 5.2. In the two IR images are IR sources (marked by small open yellow circles), the small open red circle is the Radio source position and the larger open red circle is the search radius; overlayed are the green radio flux contours. The left hand image shows a possible IR Radio Double and the right hand image shows three IR sources within the search radius around the Radio source that are unrelated.

this section the LRPY algorithm has been mofified to identify possible double blended radio sources using the background sources from Fusion.

When two Fusion sources with similar LRs are found in the search field around a radio source, the R of both sources is determined by :

$$R = \frac{LR}{(1-Q_0)+LR/0.5}\ , \tag{5.1}$$

which follows from Equation 3.16 when $LR = LR_1 = LR_2$. Equation 5.1 results in $R = 0.5$ when $LR \gg 1-Q_0$.

In the following Figure 5.4 the axis of symmetry for pairs is shown with the red solid curve which follows Equation 5.1. The red dashed curves above and below the axis of symmetry are given by:

$$R = \frac{LR}{(1-Q_0)+LR/(0.5\pm\beta)}\ . \tag{5.2}$$

where in this case $\beta = 0.4$, so that when $LR \gg 1-Q_0$ these tend toward $R = 0.1$ and 0.9 for the dashed lines. Thus the hypothesis is proposed that if both counterparts have $0.1 \leq R \leq 0.9$ then they might both be counterparts Fusion sources and both be contributing to the radio emission. As I stated earlier about my criteria for selecting IRD candidates, the R is proportional to $f(r)$ (see Equation 3.1) and this is a function of angular distance r between the candidate and source, so candidates at similar distances from the source will have a similar $f(r)$. In addition the IR flux of the IR candidate comes into the relationship to determine the R as part of the LR. However, if one counterpart has a reliability $R < 0.1$ then I consider the other counterpart to be the sole true match.

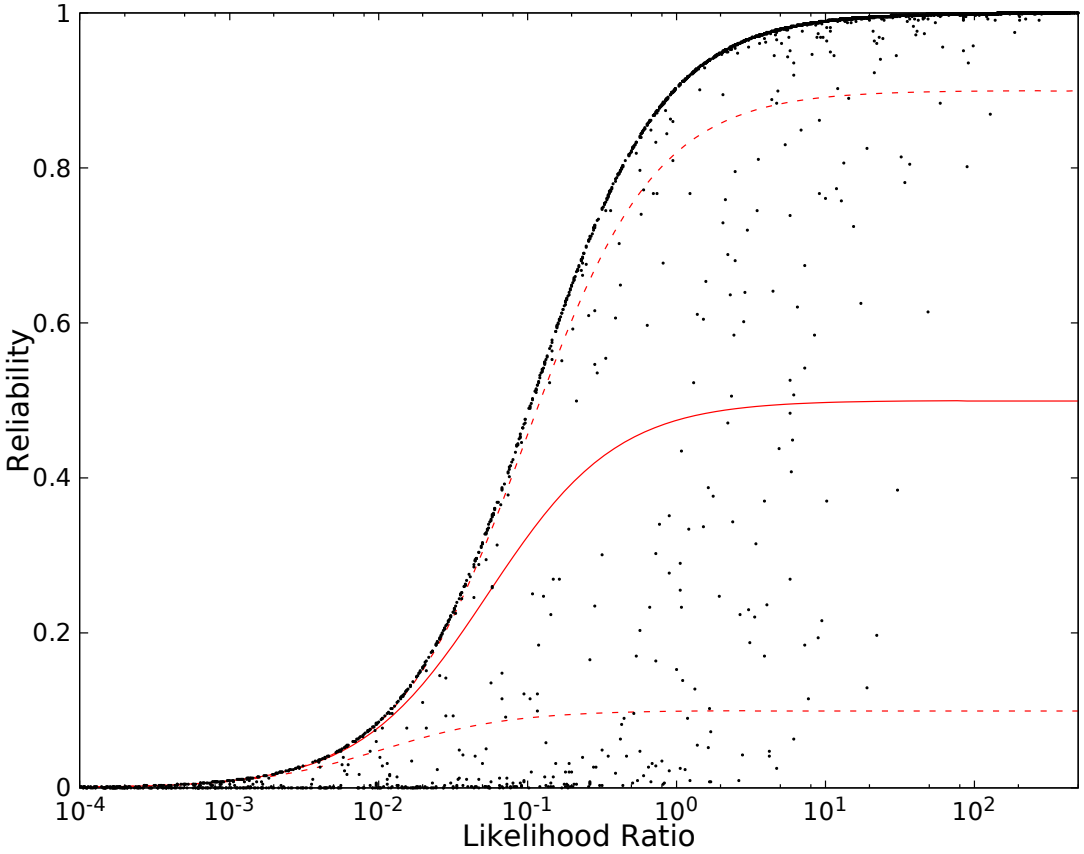


FIGURE 5.4: This is a plot of R vs LR for all possible matches within the 6" search radius for both fields. In addition are show the selection boundarys, the upper and lower selection boundrys are marked with a red dotted line and the axis of symmetry of points is marked by a solid red line.

For example, if there are two sources inside the search radius, one with $R_1 = 0.05$ and the other with $R_2 = 0.95$ (so $R_1/R_2 < 1/19$), the first source is rejected, even if $LR_1 > LR_c$, and the second Fusion source is considered as a single source and sole counterpart. Hence all components of pairs below the lower dashed line are rejected, and all Fusion sources above the upper dashed line are now considered as singles. For this work a value of $\beta = 0.4$ was used based on the *LR* and *R* cut-off values in Figures 3.9 and 3.10. This acceptance zone can be narrowed or widened by decreasing or increasing β in the algorithm, respectively.

There is a relatively small subset of multiple Fusion counterparts between the dashed lines in Figure 5.4 and with $LR > 0.01$. In our case, there are 38 pairs of Fusion counterparts in the CDFS field and 26 in ELAIS-S1 which makes $\approx 2\%$ of all radio sources with cross-identifications in Table 3.11. Hence if this selection is applied to much larger catalogues with a significant number of sources for example the 7×10^7 radio sources expected in EMU, then using this detection rate there could be more than a million with multiple matches which could be missed.

Using the selection rules as outlined earlier in this section, 64 pairs for the two fields were found. To explore the possibility that some of these 64 pairs of galaxies could be members of the same group of galaxies or even physically interacting, a nearest-neighbour match of the Fusion sources with objects from the OzDES survey (presented in Section 2.1.3) was performed. A NN approach was applied because both the Fusion and the OzDES optical surveys use the same reference frame and have high resolution compared to the radio.

If the Fusion source is within $1''$ of an OzDES object, it is considered to be the same object. From this a total of 22 out of 64 doubles were found to have spectroscopic redshifts of *both* galaxies which are presented in Tables 5.1 and 5.2. The columns in these two tables are organised as follows:

Column (1) ATLAS DR3 Identification number of the radio source "cid"

Column (2) 2 lines for the Fusion Identification number "swire_index_*Spitzer*"

Column (3) 2 lines for the angular separation (arcsec) or each Fusion source from the common radio source

Column (4) 2 lines for the NN OzDES sources to the Fusion sources in column 2

Column (5) 2 lines for the redshift "z" from OzDES

Column (6) The difference in redshift from column 5

For each ATLAS source there are two lines in the table for the Fusion sources that match the above selection criteria.

TABLE 5.1: For the ATLAS CDFS field are presented the redshifts for possible IR doubles taken from OzDES, by a nearest neighbour match between Fusion and OzDES (Yuan et al., 2015) within 1″. The lines with the ATLAS ID in bold indicate HST images exist of these galaxies which are presented in more detail latter.

ATLAS ID	Fusion ID	ang sep (arcsec)	OzDES ID	OzDES z	OzDES Δz
CI0069	309081	0.528	281939.8	0.6789	0.0019
	309075	0.045	00036776	0.68084	
CI0099C2	295215	0.106	0076-01223	0.3339	0.0004
	295098	0.285	2940685175	0.3344	
CI0175	333146	0.448	91-274837.9	0.1816	0.0001
	333165	0.223	32564	0.1817	
CI0191	467746	0.321	S117	0.0909	0.0001
	467716	0.121	NAO_0552_119829	0.09078	
CI0548	322386	0.188	57-280213.3	0.5368	0.0005
	322361	0.236	2940877666	0.53631	
CI0561	151844	0.013	20-283323.1	0.3402	0.006
	151810	0.106	0082-01440	0.3462	
CI0632	328609	0.231	NOAO_0334_R126091	0.32776	0.0007
	328657	0.076	NOAO_0552_126052	0.32708	
CI0633	197618	0.277	S477	0.2511	0.0005
	197687	0.258	63053	0.2516	
CI0757	171555	0.1063	0084-00302	0.6633	0.0068
	171511	0.1062	0085-00883	0.6701	
CI1000	178274	0.149	2939983811	0.3380	0.0002
	178269	1.070	49-275932.3	0.3378	
CI1036	183594	0.072	92922	1.0967	0.0007
	183614	0.119	49-275217.7	1.096	
CI1042	162527	0.119	2940728894	0.406	0.006
	162601	0.106	0084-01738	0.400	
CI0418	184058	0.115	0139-01350	0.2864	0.0105
	184031	0.642	26362	0.2759	
CI0961	147928	0.094	2940682513	0.5963	0.0004
	147978	0.106	0082-00421	0.5959	
CI1905	315366	0.106	0079-00072	0.5815	0.2892
	315305	0.106	0078-01319	0.2923	
CI1906	187271	0.189	34546	0.4357	0.2216
	187229	0.055	J033244.87	0.2141	

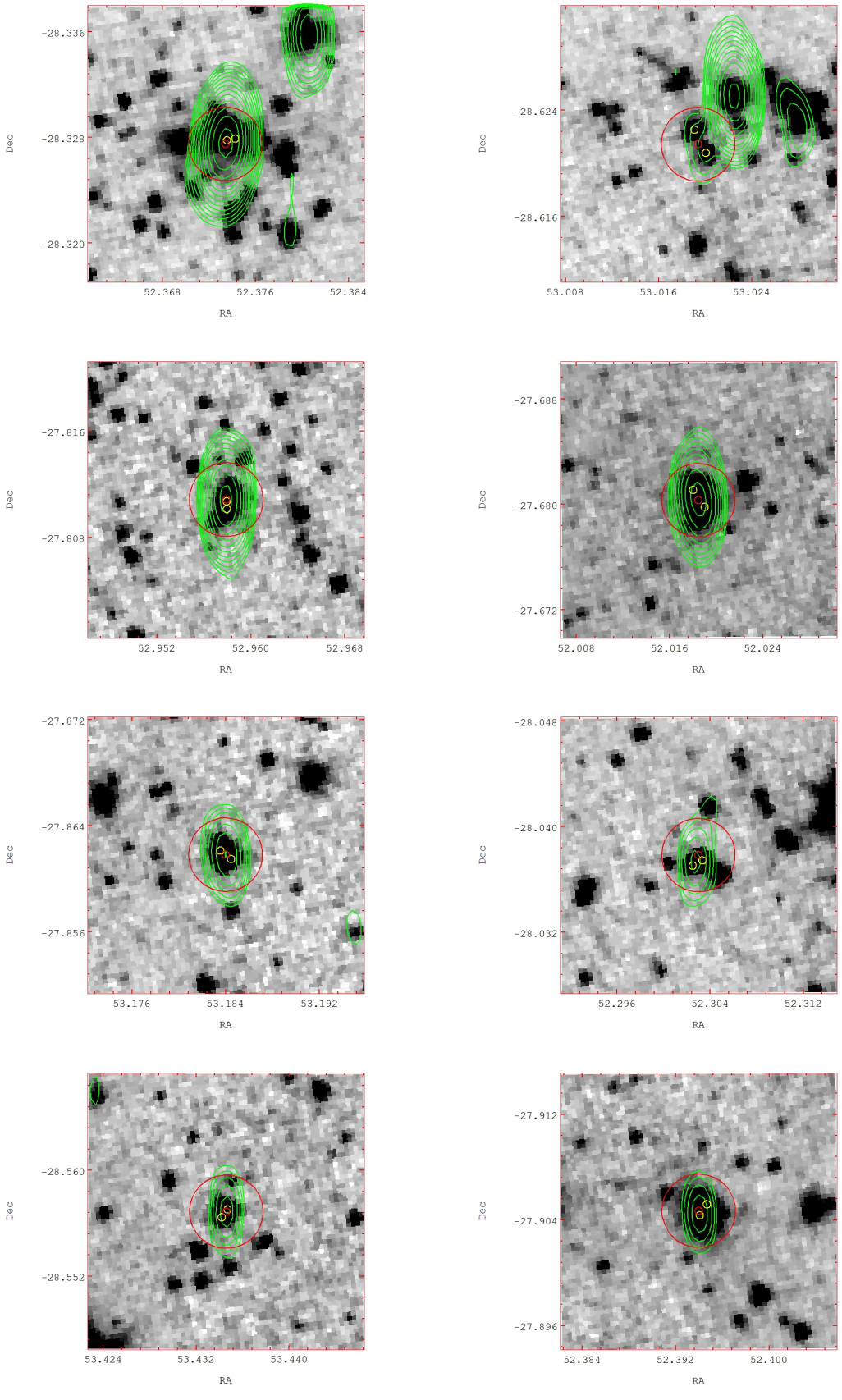
TABLE 5.2: For the ATLAS ELAIS field are presented the redshifts for possible IR doubles taken from OzDES, by a nearest neighbour match between Fusion and OzDES (Yuan et al., 2015) within $1''$.

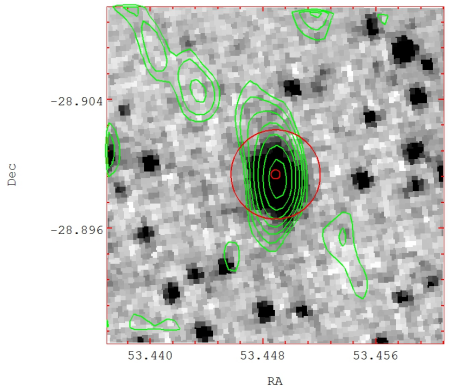
ATLAS ID	Fusion ID	ang sep (arcsec)	OzDES ID	OzDES z	OzDES Δz
EI0151	215007	0.224	2970674536	0.12434	0.0004
	215061	0.191	2970674654	0.12478	
EI0455	221400	0.254	2971105989	0.198	0.0036
	221459	0.136	0092-01998	0.195	
EI0487	101702	0.144	J003459.03	0.330	0.0011
	101761	0.111	J003458.95	0.329	
EI0863	247565	0.100	0094-01686	0.217	0.0064
	247574	0.333	2971175179	0.224	
EI1034	196663	0.20	0096-00993	0.34711	0.0022
	196655	0.11	2971105849	0.3493	
EI1219	73440	0.21	2970777434	0.4001	0.0008
	73483	0.14	2970777513	0.3993	

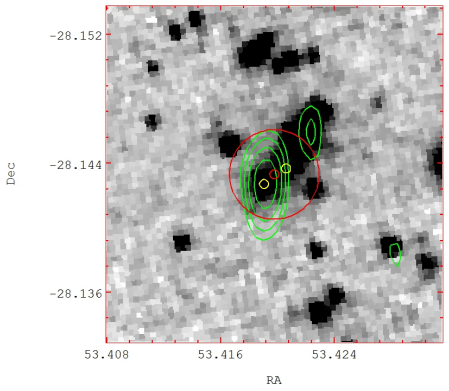
In 20 (out of 22) cases both objects have very close redshifts, $\Delta z/z < 0.01$, and in two cases the Fusion two sources making the pair have significantly different redshifts indicating a chance alignment.

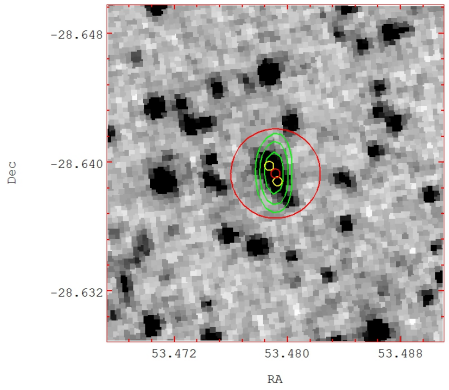
From the preceeding Tables 5.1 and 5.2 are next presented the postage stamp images of these IRDs in the following Figures 5.5 and 5.6. The ATLAS radio contours (starting at $3/\sigma$ and then spaced by a factor of 2) in green are overlaid on the greyscale IR images to demonstrate these objects. The Fusion images are centered on the ATLAS radio candidate coordinates with the radio contours overlayed to visually demonstrate the LR XID's in the CDFS field. The small red circle denotes the ATLAS radio candidate position and the larger red circle is the $10''$ search radius. The small yellow circle denotes the SWIRE IR candidate position with a $R > 0.8$; and the small megenta circles are other IR candidates within the search radius but with a $R < 0.2$. Each image is 75x75 arcseconds.

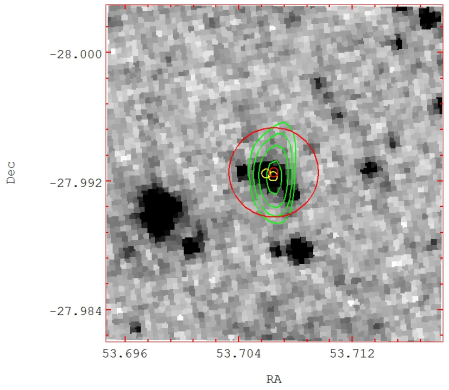
FIGURE 5.5: In this figure are shown the Infrared Doubles in the CDFS Field, the radio contours are in green overlayed onto the IR image with the Infrared sources marked with small yellow open circles. The radio source position is marked with a small open red circle, and the 10" NN search radius is marked with the larger open red cicrle. Using the ATLAS ID they are listed left to right and top to bottom: CI0069, CI0099C2, CI0175, CI0191, CI0418, CI0548, CI0561, CI0632, CI0633, CI0757, CI0961, CI1000, CI1036, CI1042, CI1633, CI1905 and CI1906

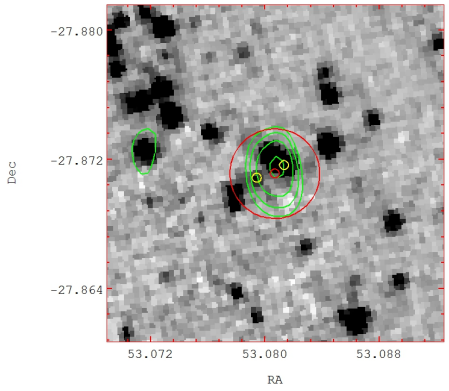


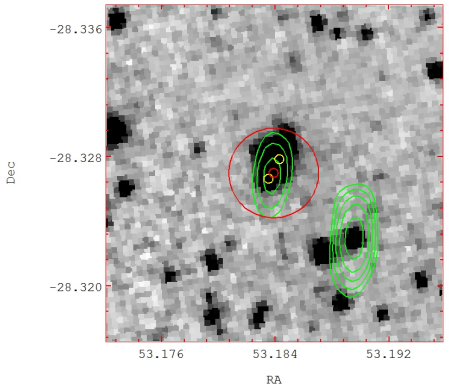

-28.904
-28.896
Dec
53.440
53.448
53.456
RA

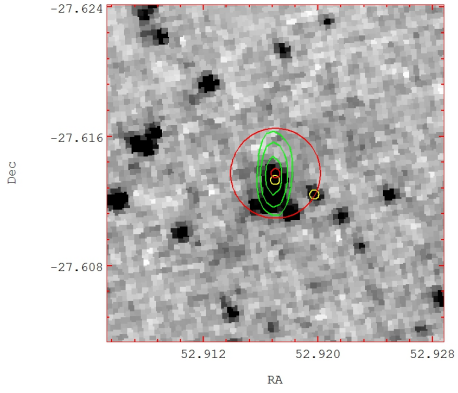

-28.152
-28.144
-28.136
Dec
53.408
53.416
53.424
RA

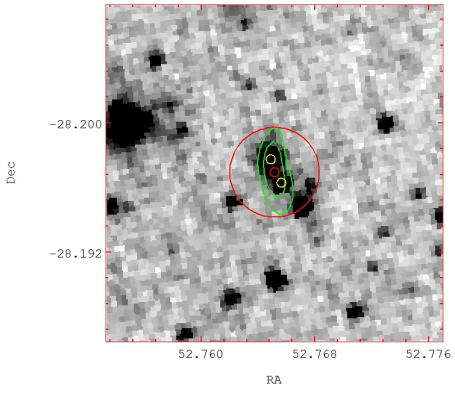

-28.648
-28.640
-28.632
Dec
53.472
53.480
53.488
RA


-28.000
-27.992
-27.984
Dec
53.696
53.704
53.712
RA


-27.880
-27.872
-27.864
Dec
53.072
53.080
53.088
RA


-28.336
-28.328
-28.320
Dec
53.176
53.184
53.192
RA


-27.624
-27.616
-27.608
Dec
52.912
52.920
52.928
RA


-28.200
-28.192
Dec
52.760
52.768
52.776
RA

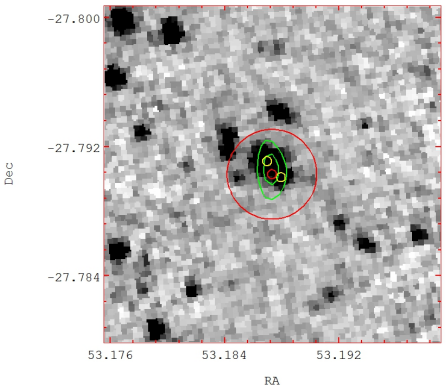
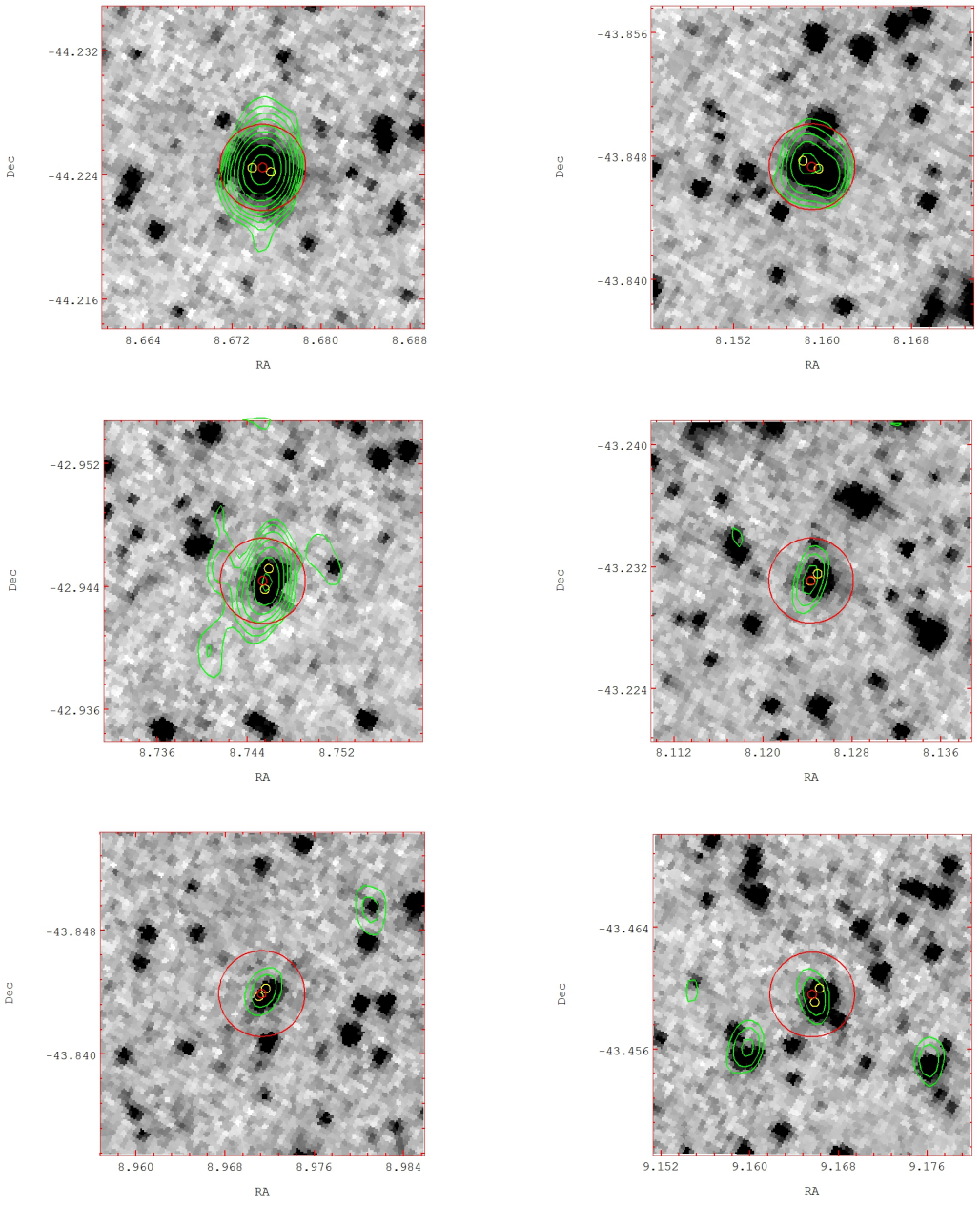

-27.800
-27.792
-27.784
Dec
53.176
53.184
53.192
RA

FIGURE 5.6: In this figure are shown the Infrared Doubles in the ELAIS Field, the radio contours are in green overlayed onto the IR image with the Infrared sources marked with small yellow open circles. The radio source position is marked with a small open red circle, and the 10" NN search radius is marked with the larger open red cicrle. Using the ATLAS ID they are listed left to right and top to bottom: EI0151, EI0455, EI0487, EI0863, EI1034 and EI1219.



For further corroboration I found in the Great Observatories Origins Deep Survey (Dickinson et al., 2003; Renzini et al., 2003, GOODS), two of the CDFS IRDs were found to have Hubble Space Telescope (HST) archive images [1], CI0418 and CI1036. In the following Figures 5.8 and 5.8 are presented these images with the IR source positions marked and the radio contours overlaid. As well as the ATLAS radio contours also shown are the contours from the deep JVLA 1.4 GHz survey of this sub-region of the CDFS (Miller et al., 2013). The HST images clearly indicated that these two pairs of galaxies are interacting via their disturbed morphologies and tidal tails.

[1] Based on observations made with the NASA/ESA Hubble Space Telescope, obtained from the data archive at the Space Telescope Science Institute. STScI is operated by the Association of Universities for Research in Astronomy, Inc. under NASA contract NAS 5-26555.

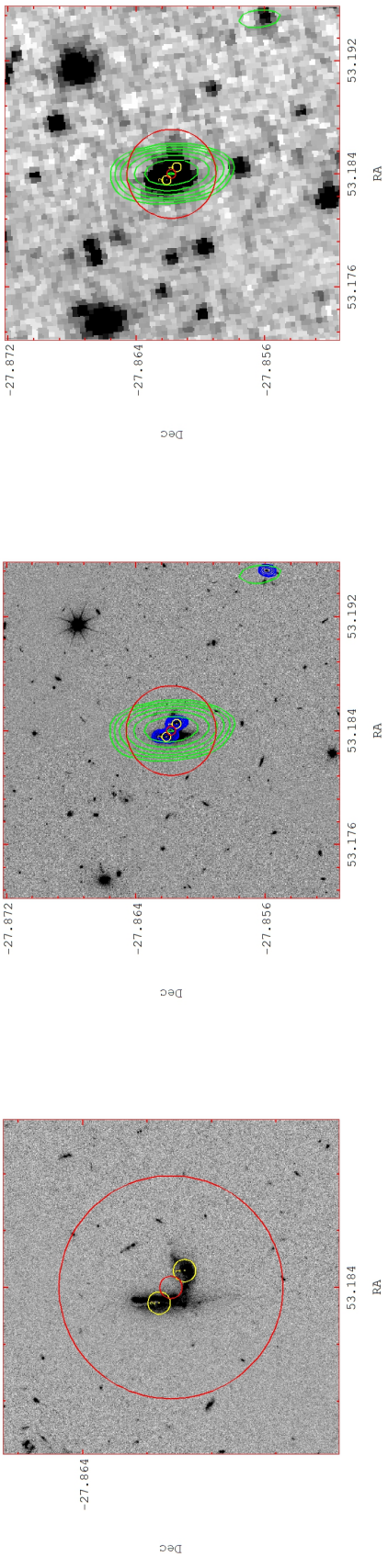


FIGURE 5.7: The image on the left shows the HST image of the area. The middle image shows the IR double candidate for ATLAS source CI0418 comprised of the two Fusion candidates positions overlayed on a HST image. The image on the right covers the same area but has the Spitzer IR image as the background for comparison. In all images the position for the ATLAS radio source is shown by a small red circle, the possible IR candidate positions are shown by small yellow circles. The larger red circle shows the 6″ search radius centered on the radio source used in this work. The ATLAS radio contours are in green and the VLA radio contours are in blue.

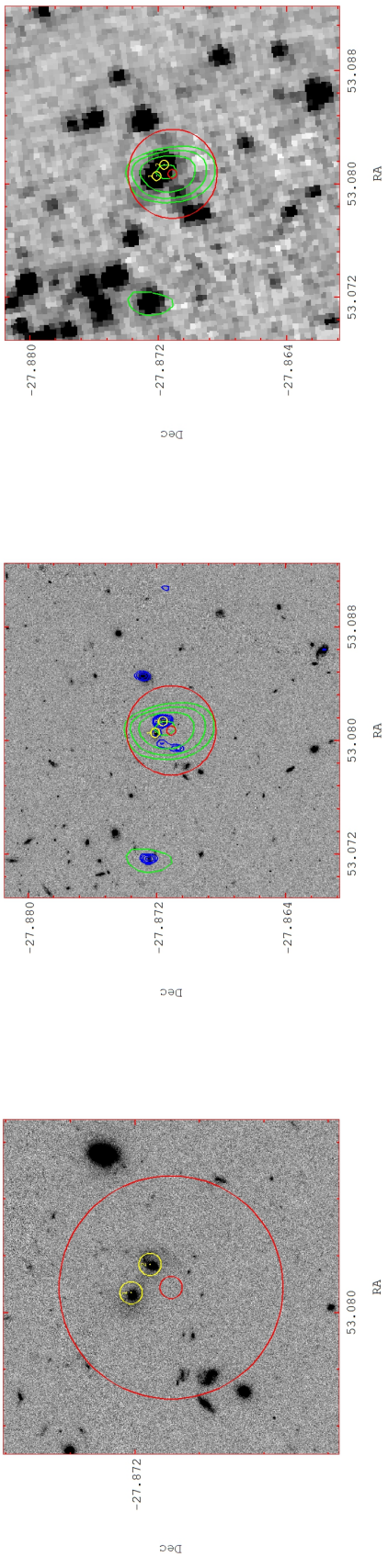


FIGURE 5.8: The image on the left shows the HST image of the area. The middle image shows the IR double candidate for ATLAS source CI1036 comprised of the two Fusion candidates positions overlayed on a HST image. The image on the right covers the same area but has the Spitzer IR image as the background for comparison. In all images the position for the ATLAS radio source is shown by a small red circle, the possible IR candidate positions are shown by small yellow circles. The larger red circle shows the 6″ search radius centered on the radio source used in this work. The ATLAS radio contours are in green and the VLA radio contours are in blue.

5.2 Identifiying double radio components

In this section I look to extend the LRPY algorithm to possibly identify radio pairs and multiple radio component candidates with an automated method. I start by using the FIRST catalogue which covers a very large area to a higher resolution than ATLAS, and which contains a large set of 1378 radio pairs , and can therefore be used to refine a method for identifying Radio Doubles. The results from this are then applied to the LRPY algorithm and applied to the ATLAS DR3 catalogue for identifying possible Radio Double candidates and this work is currently a paper in preparation (Weston et al, in prep).

5.2.1 Radio Doubles

The aim is to extend the LRYP algorithm to flag possible radio pairs and multiple radio components. Taking the work of Magliocchetti et al. (1998) with the FIRST survey, they defined two sets of criteria for combining sources into pairs, the first is where $flux1/flux2 = 1$, this is relaxed for the combined sources if their fluxes differed by a factor less than 4:

$$\frac{1}{4} < \frac{flux1}{flux2} < 4, \tag{5.3}$$

the second was taking those pairs where the following term :

$$\theta_{link} = 100 \left(\frac{flux1 + flux2}{100} \right)^{0.5} arcseconds \tag{5.4}$$

defines the upper limit for possible pairs. These are demonstrated in the following Figure 5.9 reproducing the work of Magliocchetti et al. (1998). This method has also been applied by Sadler et al. (2002) and Ching et al. (2017).

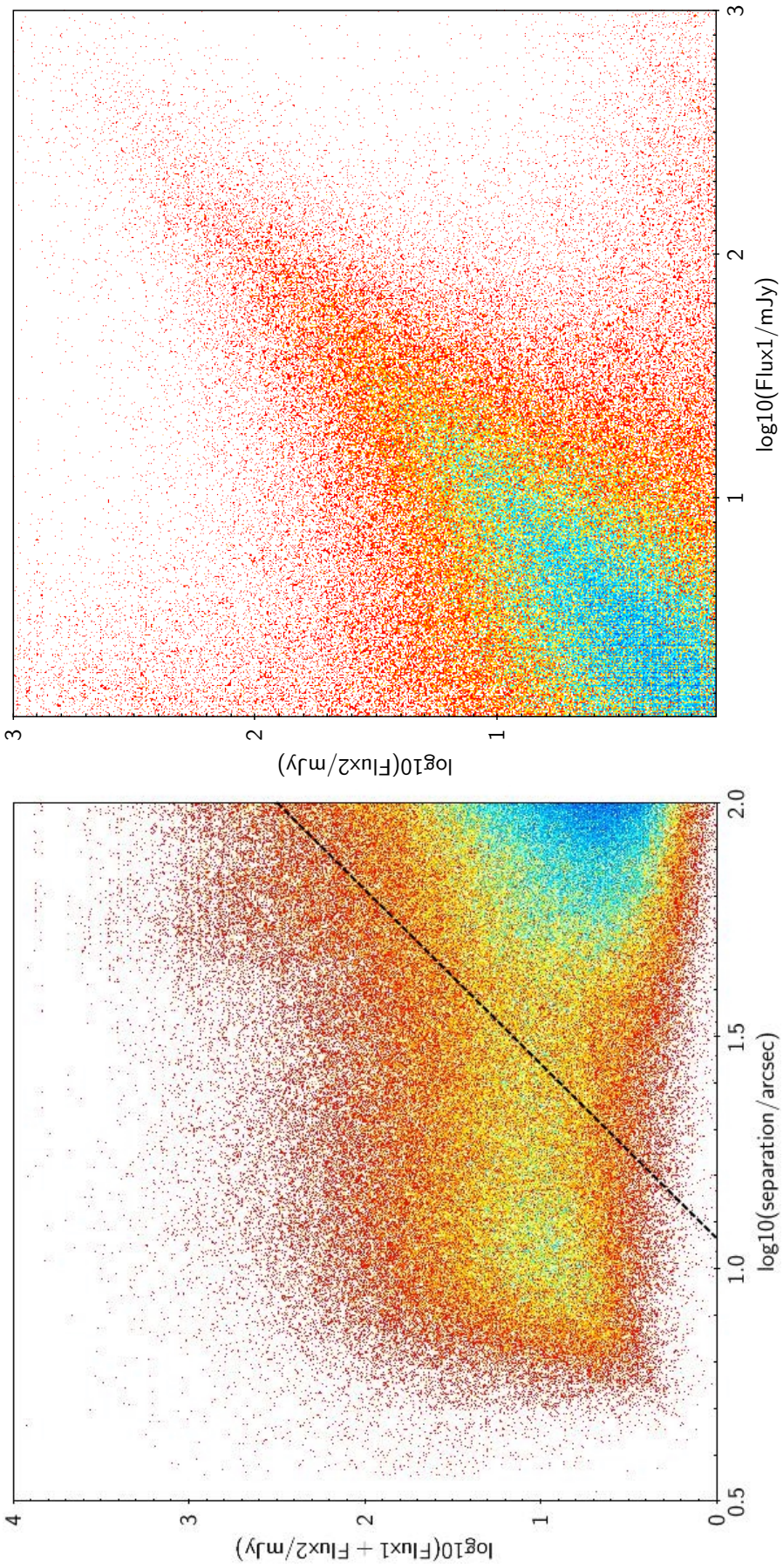


FIGURE 5.9: Top: the FIRST double sources the sum of the fluxes of the components as a function of their separation, the line shows the Magliocchetti et al. (1998) criterion. Bottom: is the flux ratio of the components from the same double sources plotted at the top. The colour is a log density scale with red being lowest and blue being highest.

I have looked for a method to combine the information presented in Figure 5.9 which could be programmed into the LRPY algorithm, which I now explain.

For a source at a distance R and measured flux density f, the luminosity is:

$$L_\nu = 4\pi R^2 f \tag{5.5}$$

If the flux F of a source is measured at a distance R , then the flux f measured from the same source at a distance r is:

$$f = \left(\frac{R}{r}\right)^2 F \tag{5.6}$$

If a standard candle is used, its brightness would change with the distance as $1/r^2$, so "r" is proportional to $1/\sqrt{F}$. Separation can be a measure of the distance to the source, so in the standard candle model θ is proportional to $1/\sqrt{F_1+F_2}$. By dividing by $\sqrt{F_1+F_2}$ then all the sources are effectively brought to the same distance from us , so it is possible to compare their physical properties. One of these properties can be possible correlation between "true separation" ($\theta/\sqrt{F_1+F_2}$) and the flux ratio F_1/F_2.

Looking at the ratio of angular separation to flux difference between nearest neighbour radio sources with an angular separation of less than $100''$ the following term can be derived θ_F :

$$\theta_F = \frac{\theta}{\sqrt{F_1+F_2}} \tag{5.7}$$

Using the FIRST catalogue, and doing a nearest neighbour search on its self with a search radius of 100" and plotting these sources with our new parameter space of F_1/F_2 versus $\theta/\sqrt{F_1+F_2}$, the top plot in Figure 5.10 is obtained.

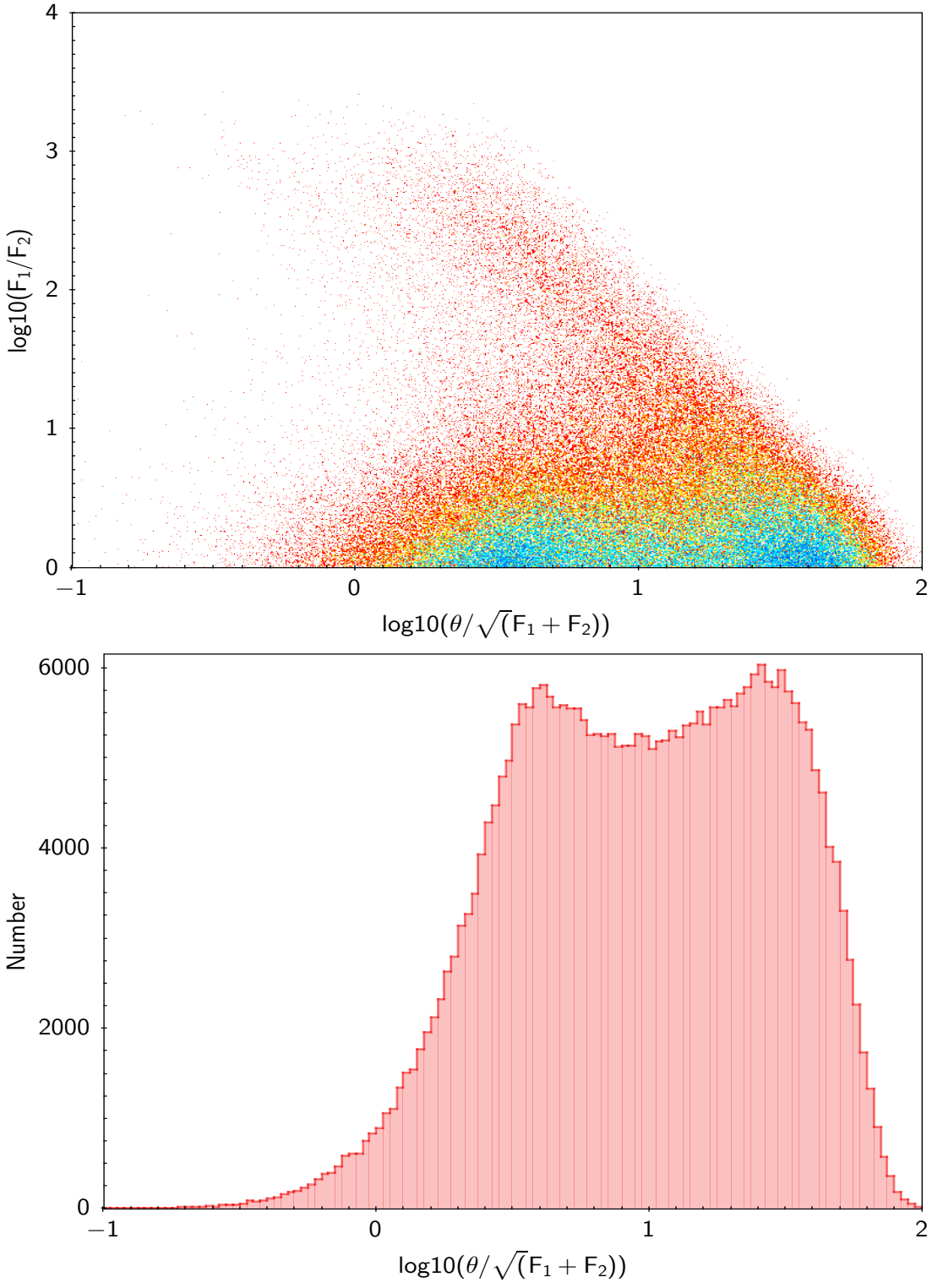


FIGURE 5.10: Top: is the plot of the flux ratio of the components from the same double sources plotted in Figure 5.9. The colour is a log density scale with red being lowest and blue being highest. Bottom: histogram of the values from Equation 5.7 for the nearest neighbour pairs from the FIRST catalogue.

In Figure 5.10 both plots quite clearly shown that there are two clusters of pairs, one centered about $\theta/\sqrt{(F_1+F_2)} \approx 4$ and another at $\theta/\sqrt{(F_1+F_2)} \approx 30$. To verify if this is a real affect I created a random catalogue covering the same area using the same FIRST sources but giving them random RA and Dec positions but keeping their other properties. This was achieved by taking the RA and Dec positions of the sources in the catalogue and moving each by a random $\pm 400''$. The physical properties such as peak flux and integrated flux are thus retained. This will remove any natural clustering and physically associated sources. The same plots as shown in Figures 5.10 are reproduced in Figure 5.11 for the random catalogue.

Figure 5.11 quite clearly shows that if sources were uniformly distributed with no natural clustering and no physically associated sources, then as the search radius increases the probability of finding a random neighbour increases producing ”random” pairs. Thus there is only one concentration of pairs at $\theta/\sqrt{(F_1+F_2)} \approx 30$ there is no concentration of pairs at $\theta/\sqrt{(F_1+F_2)} \approx 4$.

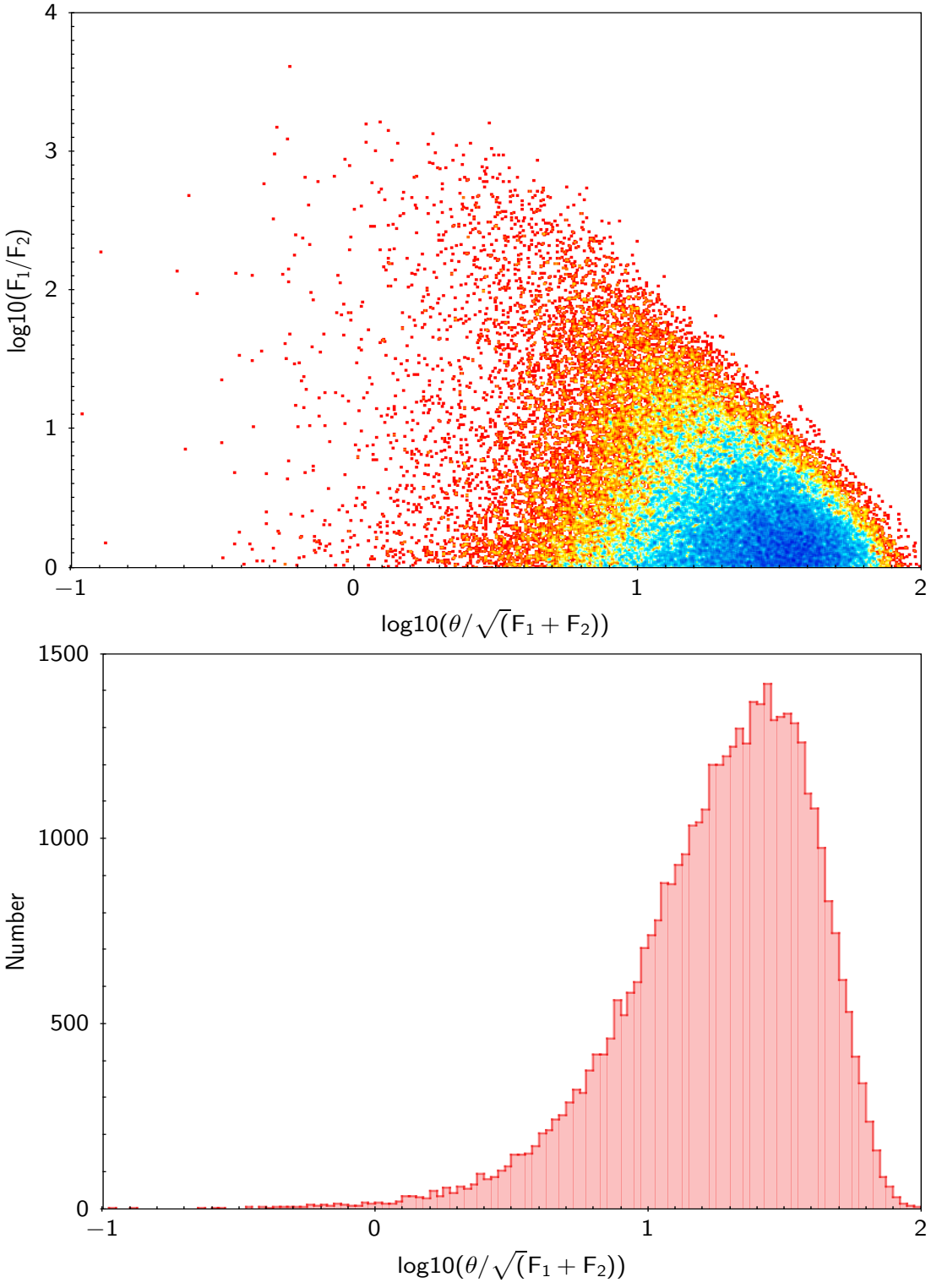


FIGURE 5.11: Top is the plot of the flux ratio of the components from the double sources using a random catalogue covering the same area as FIRST. The colour is a log density scale with red being lowest and blue being highest. Bottom is a histogram of the values from Equation 5.7 using the same random catalogue covering the same area as the FIRST catalogue.

Next the data as shown in Figure 5.10 and 5.11 is gridded with a 50×50 grid placed over the area, then a count of the number of points within each grid square is taken. The top plot in Figure 5.12 shows the results of this gridding for the pairs selected from the FIRST catalogue. The second plot from the top in Figure 5.12 shows the result of the same gridding for the generated random catalogue. Taking an area of the middle plot ([30:40] , [20:30]) where doubles are not expected the average count over these grid squares is determined and used as a weighting factor (4) to be applied to the random grid values.

Next all the random grid square counts are multiplied by this weight factor. Then the new weighted RANDOM grid square counts are subtracted from the REAL grid square counts. This produced the bottom plot in Figure 5.12 showing the residuals, clearly identifying a group of objects which are taken as the real physical pairs, due to natural clustering and physically associated sources.

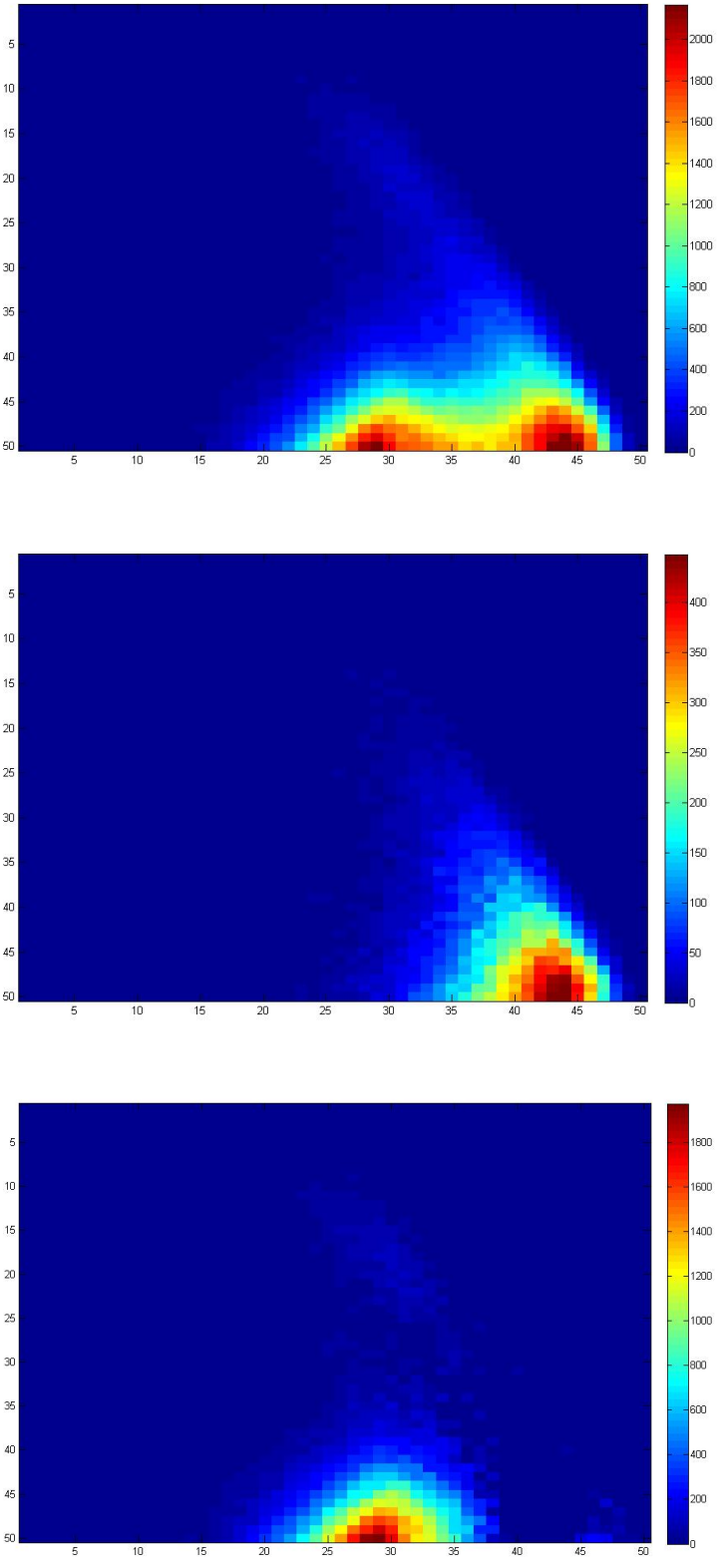


FIGURE 5.12: In the above plots a 50x50 grid was placed onto the data shown in the Figures 5.10 & 5.11 and the points within each grid where summed. The top plot shows the results of this gridding for Figure 5.10. The second plot from the top shows the result of this gridding for Figure 5.11. The bottom plot shows the result of subtracting the Weighted RANDOM grid values from the REAL grid values to show the residuals.

If it is assumed the bottom plot of Figure 5.12 shows the true Radio Doubles. Taking a grid square $[29:50]$ and taking the value of this square (Which is the number of true Radio Doubles for that grid square) and dividing it by the same grid square from the top plot in Figure 5.12 (Which will be the number of all nearest neighbours consisting of true and random) the result will be to obtain the ratio of real to all. Then increase the X-Axis grid squares to use by $29 \pm i$ and Y-Axis by $50 + i$ and loop through from $i = 1;21$, the area for this is given by:

$$A = (2i+1)(i+1) \text{ units square.} \tag{5.8}$$

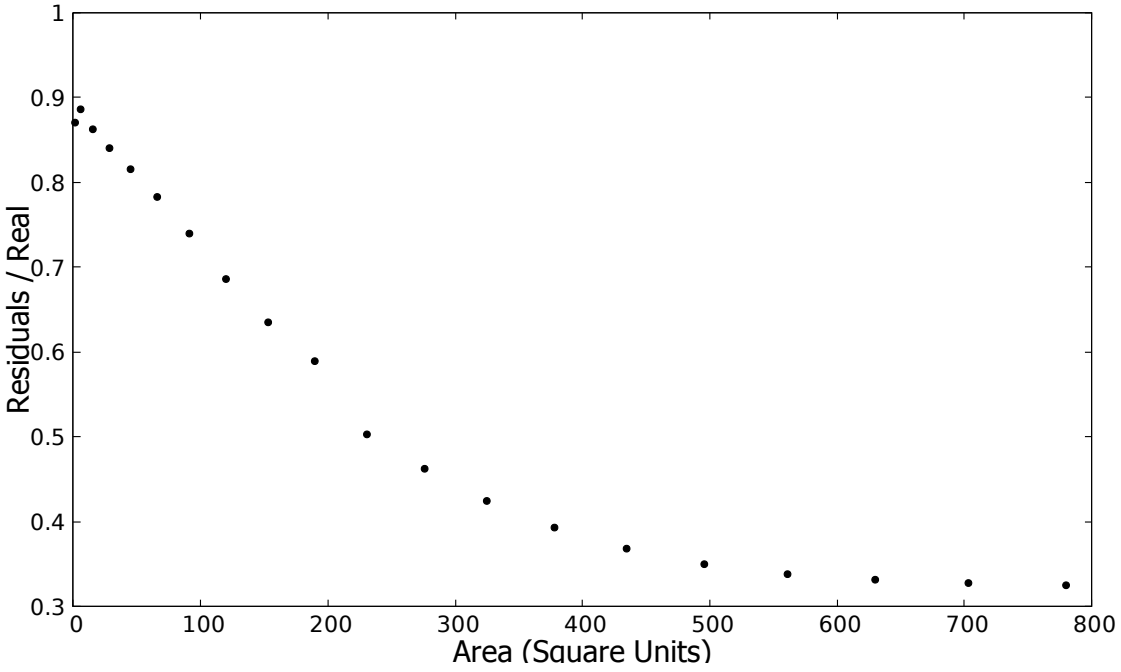


FIGURE 5.13: A plot of the Real to Random Radio Pairs by Square Unit area, using the grid square values from the top and bottom plots in Figure 5.12.

The result of this process is shown in Figure 5.13. As the area given by Equation 5.8 is expanded from the point occupied mostly by double sources, the ratio decreases from 0.88 down to 0.32. The upper value of 0.88 demonstrates that even the area dominated by doubles sources is contaminated with single sources, otherwise the ratio would be 1.0. The lower values of 0.32 gives us the percentage of double sources (32%) of all sources within the catalogue.

5.2.2 FIRST Doubles

A visual check for some of the FIRST doubles produced by the method described in Section 5.2.1 was undertaken to ensure that this method was indeed selecting ”true” radio doubles. So a set of postage stamp images[2] of a random subset of sources (500) in the Magliocchetti et al. (1998) space for doubles were produced of the FIRST radio contours overlayed onto the WISE IR images. Not all 500 sources had full WISE coverage so a subset of 239 was obtained. Undertaking a visual inspection of these, the classification summary is presented in Table 5.3.

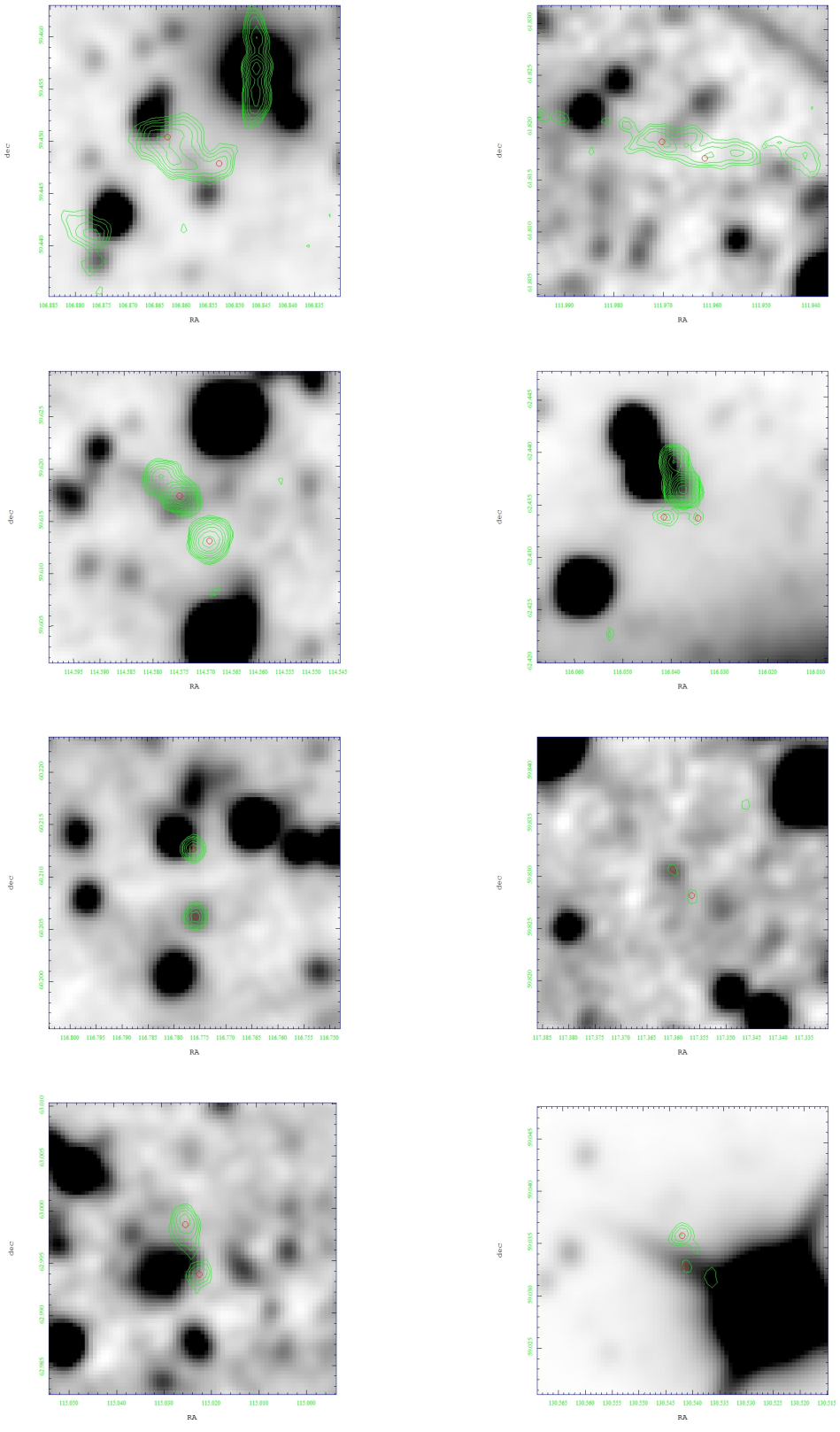
TABLE 5.3: Table showing the totals for different classes from the visual identification.

Class	Number
Double	73
Associated	43
Random	120
Unknown	3
Total	239

In Figure 5.14 I attempt to show visually what the different classes of indentifications looked like that are presented in Table 5.3.

[2]The full set of postage stamp images are provided in the electronic supplement

FIGURE 5.14: FIRST Radio Doubles, here we have created postage stamp images of the FIRST radio contours overlayed on the same sky image from WISE. The first line were identified as True doubles, the second line are Associated, the third line are Random and the last line are Unknown.



Taking the visual identification of these objects and plotting their physical properties using the same parameter space as in the top plot of Figure 5.10, the plot in Figure 5.15 was obtained:

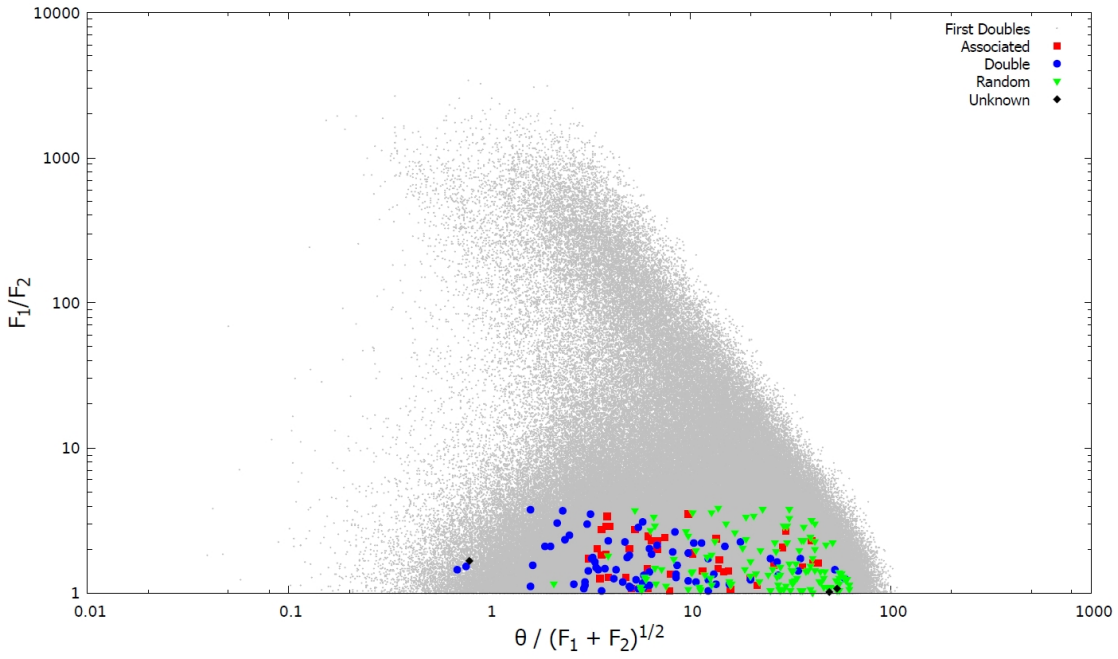


FIGURE 5.15: In the above is plotted the flux ratio of the FIRST doubles from Figure 5.10 in the background as grey. Overlayed are the different doubles identified from the visual identification in Table 5.3. The blue diamonds are the associated sources, the green filled circles are the doubles, the pink squares are the random doubles and the black filled triangles are the unknown objects.

Based on this and the following section it is proposed that the Magliocchetti et al. (1998) selection criteria can be tightened to define an area bounded by $1.0 \leqslant F_1/F_2 \leqslant 2.1$ and $1.9 \leqslant \theta_F \leqslant 11.0$. Taking this area from Figure 5.13 a ratio of the Residuals to Total of 0.837 is obtained.

5.2.3 ATLAS DR3 Doubles

Taking the ATLAS DR1 entries from Norris et al. (2006) CDFS Table 6 where the column f_class contains one or more lower-case letters, for this work those entries with a letter "a" indicating morphology (i.e. double, triple, or core-jet radio source) were used to obtain a list of multi-component radio sources, this was refined to select those visually identified as radio double and triple sources. Then also including Norris et al. (2006) ELAIS Table 6, where a different identification method has been used to flag visually identified radio doubles or triples in the "comm" column searching for the strings "double" and "triple". Both of these sets of multiple radio components have been marked on Figure 5.16 as blue rectangles labelled "ATLAS DR1 doubles". It was noticed that the majority of these objects lie within an area $1.0 < F_1/F_2 < 2.1$ and $1.9 < \theta_F < 11.0$ as shown in Figure 5.16, this also coincides with a concentration of WISE doubles and inside the Magliocchetti et al. (1998) pair selection criteria. But some of these ATLAS doubles do lie outside the Magliocchetti et al. (1998) selection area.

Using the above selection criteria I have selected the potential doubles from the work in this thesis with ATLAS DR3 and overlayed them on the same plot as black diamond outlines. In the algorithm where the above selection criteria are met the two components are marked as a possible radio double (RD), their entries are flagged in the ATLAS catalogue and a new radio source is created with a flux weighted mean position between the two components and the radio properties required by the LR for this new source are derived thus:

- $FLUX_{RD} = FLUX_1 + FLUX_2$
- $SNR = max(SNR_1, SNR_2)$
- $SINT_{ERR} = \sqrt{({SINT_{ERR_1}}^2 + {SINT_{ERR_2}}^2)}$

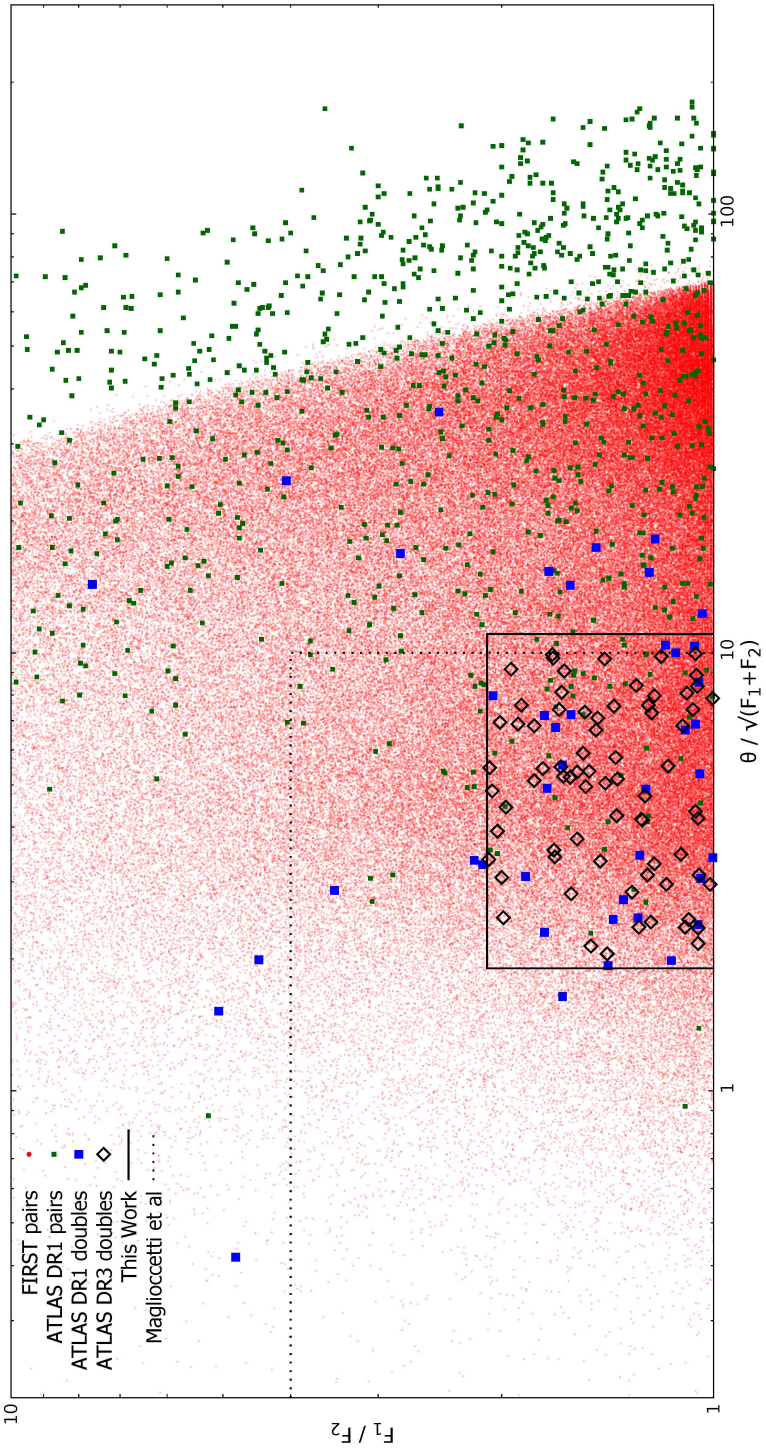


FIGURE 5.16: In this figure are presented two datasets to demonstrate the selection criteria for possible radio double candidates. Taking the FIRST survey a nearest neighbour search on itself using a search radius of 100" has been undertaken and plotted the flux ratio F_1/F_2 of the pairs vs their θ_F from Equation: 5.7 these are represented by small red circles. Next using the ATLAS DR1 dataset are have plotted the nearest neighbour radio pairs with their flux ration F_1/F_2 and θ_F with small green rectangles. The blue filled rectangles represent those sources that were visually identified as radio pairs by (Norris et al., 2006). The black open diamonds are the radio doubles selected from ATLAS DR3 by our algorithm. Two large rectangles are marked , the first using a solid black line is the area defined by our selection criteria $1.0 < F_1/F_2 < 2.1$ and $1.9 < \theta_F < 11.0$, the rectangle with a dashed line uses the selection criteria from (Magliocchetti et al., 1998).

The extension to the algorithm then takes this new source using the flux weighted mean position between the two sources and the radio properties derived in the previous paragraph. It then looks for a IR candidates around this new flux weighted mean position, for any candidates found within the 6" search radius about this new position it calculates their LR and *R* values. These new candidate R values are then compared with the values for the candidates about the components, if the R of the new candidate about the flux weighted mean position meets our selection criteria from Section 3.1.6 and is superior to the component candidates then this new candidate is taken as the source of the radio emission. The database is modified accordingly with this new source and the others are flagged but not deleted.

Some examples of potential radio doubles in postage stamp images are presented on the next page. The left hand image in Figure 5.17 presents a situation where there are no likely candidates near each radio source (marked with a +) , but a very obvious visual candidate (marked with a ⊙) near the flux weighted mean position (marked with a ◯). In the right hand image there are candidates near each radio source, but a larger candidate near the flux weighted position.

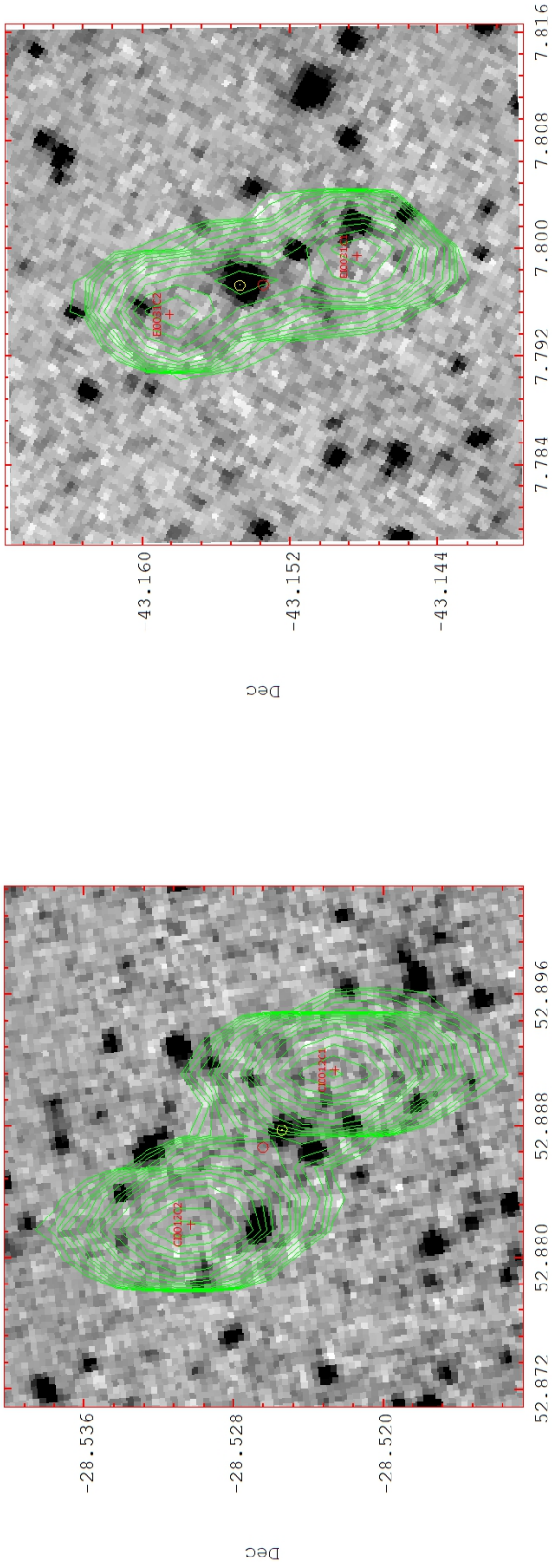


FIGURE 5.17: In the left is a example of a radio double candidate in the CDFS field (CI0012C1), and on the right another example in the ELAIS field (EI0031C1).

Following are presented for the Radio Doubles selected from ATLAS DR3, Table 5.4 (CDFS) and Table 5.5 (ELAIS). The columns in these two tables are organised as follows:

Column (1) - Internal ID for the new created source

Column (2) - ATLAS ID of Component 1

Column (3) - ATLAS ID of Component 2

Column (4) - LR of the new source

Column (5) - *R* of the new source

TABLE 5.4: CDFS Radio Doubles

Radio Double ID	ATLAS ID 1	ATLAS ID 2	LR	R
56	CI0011C2	CI0011C3	0.000000405	0.000002366
60	CI0011C2	CI0011C4	6.037470230	0.972429364
99	CI0021C1	CI0021C3	0.000542224	0.002485028
99	CI0021C1	CI0021C3	0.046477792	0.213009026
135	CI0030C2	CI0030C1	34.42959631	0.995052818
195	CI0035C1	CI0035C2	4.647264872	0.964474749
248	CI0045C4	CI0045C3	0.000000005	0.000000029
286	CI0052C1	CI0052C2	0.665529324	0.795416307
295	CI0054C2	CI0054C1	48.89734749	0.996511484
304	CI0058C1	CI0058C2	5.744737418	0.971065107
315	CI0059C1	CI0059C3	0.000000004	0.000000023
321	CI0059C1	CI0059C4	15.8939	0.9893
442	CI0076C1	CI0076C2	62.08357092	0.997250389
604	CI0106C3	CI0106C2	0.000557063	0.003243460
601	CI0106C3	CI0106C1	9.375089901	0.982068766
630	CI0109C2	CI0109C1	0.000014433	0.000084309
683	CI0112C1	CI0112C2	6.817366401	0.975506148
792	CI0152C1	CI0152C2	42.20646605	0.995960692
914	CI0178C1	CI0178C2	23.01742789	0.992618084
1015	CI0196C1	CI0196C2	49.19364020	0.996532422
1296	CI0285C1	CI0285C2	0.000054659	0.000319212
1314	CI0291C1	CI0291C3	0.000075814	0.000442704
1315	CI0291C2	CI0291C3	0.000000001	0.000000006

TABLE 5.5: ELAIS Radio Doubles

Radio Double ID	ATLAS ID 1	ATLAS ID 2	LR	R
68	EI0024C2	EI0024C1	0.008790136	0.047500606
131	EI0032C1	EI0032C2	197.0044416	0.999106084
148	EI0040C2	EI0040C1	5.654732101	0.969771369
157	EI0040C1	EI0040C3	7.828411913	0.977979993
158	EI0040C2	EI0040C3	0.000000015	0.000000085
182	EI0046C2	EI0046C1	352.5925752	0.999500344
190	EI0047C2	EI0047C1	0.000014506	0.000082291
233	EI0055C1	EI0055C2	0.000000025	0.000000142
277	EI0066C1	EI0066C2	0.472260037	0.728208576
305	EI0071C1	EI0071C2	6.794966377	0.974668231
315	EI0074C1	EI0074C2	0.007721923	0.041970412
334	EI0077C1	EI0077C2	0.000000002	0.000000011
344	EI0080C1	EI0080C2	14.56004437	0.988038862
353	EI0082C1	EI0082C2	48.27119253	0.996361770
431	EI0104C2	EI0104C1	3.455591841	0.951467489
521	EI0124C2	EI0124C1	0.000000625	0.000003546
671	EI0150C2	EI0150C1	44.73755332	0.996075528
761	EI0169C1	EI0169C2	0.001499545	0.008435663

This extension of the LRPY algorithm method fails for more complex radio sources with three or more components. In Figure 5.18 I show ATLAS radio source EI0032 which has three components (C1, C2 & C3), overlayed onto the greyscale IR image. In the three images it has three possible combinations making three possible doubles via this algorithm. Possibly this method can be extended to see if there are several doubles which have common components and try to distinguish the more complex morphology and classify it accordingly. This can also be seen from the Tables 5.4 and 5.5 where there are several doubles made from three or more component sources as indicated by the ATLAS ID naming (CI0106C1 to CI0106C3).

Looking at some of these multiple component sources I wonder if a minimum separation between components should also be defined, as looking at the postage stamp images it is not apparent that they are double or more complex sources. The source finder used for ATLAS appears to have identified multiple components which could be small scale structure within the source and not true doubles or triples with clearly well separated components with no IR source.

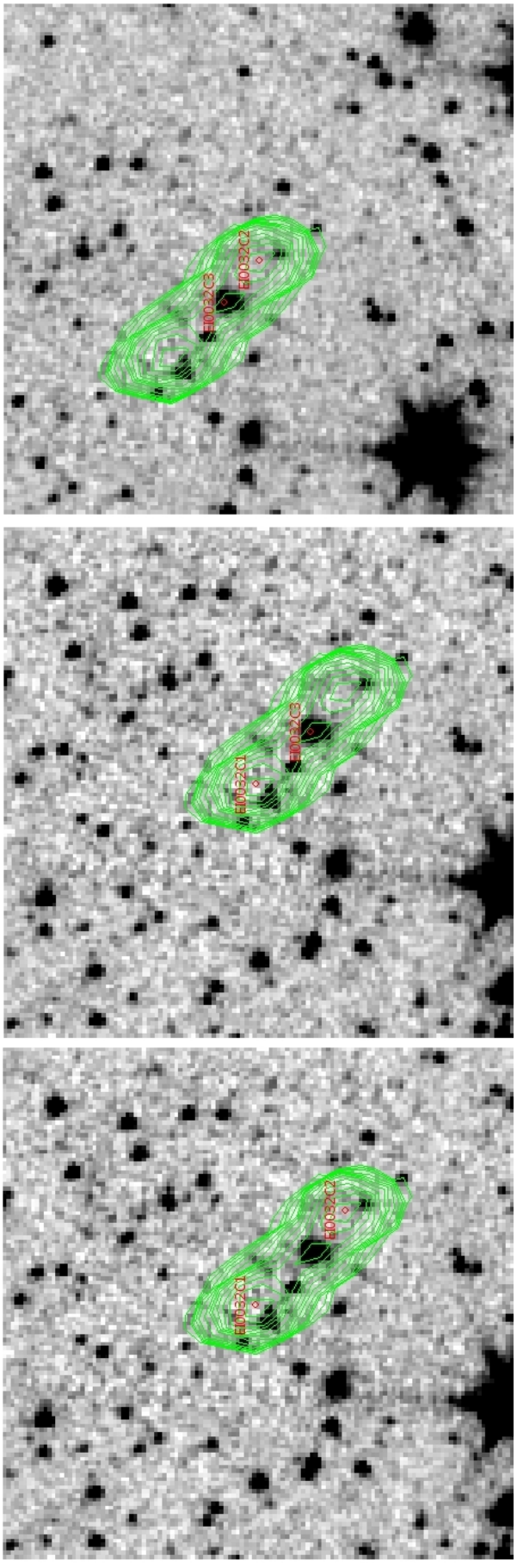
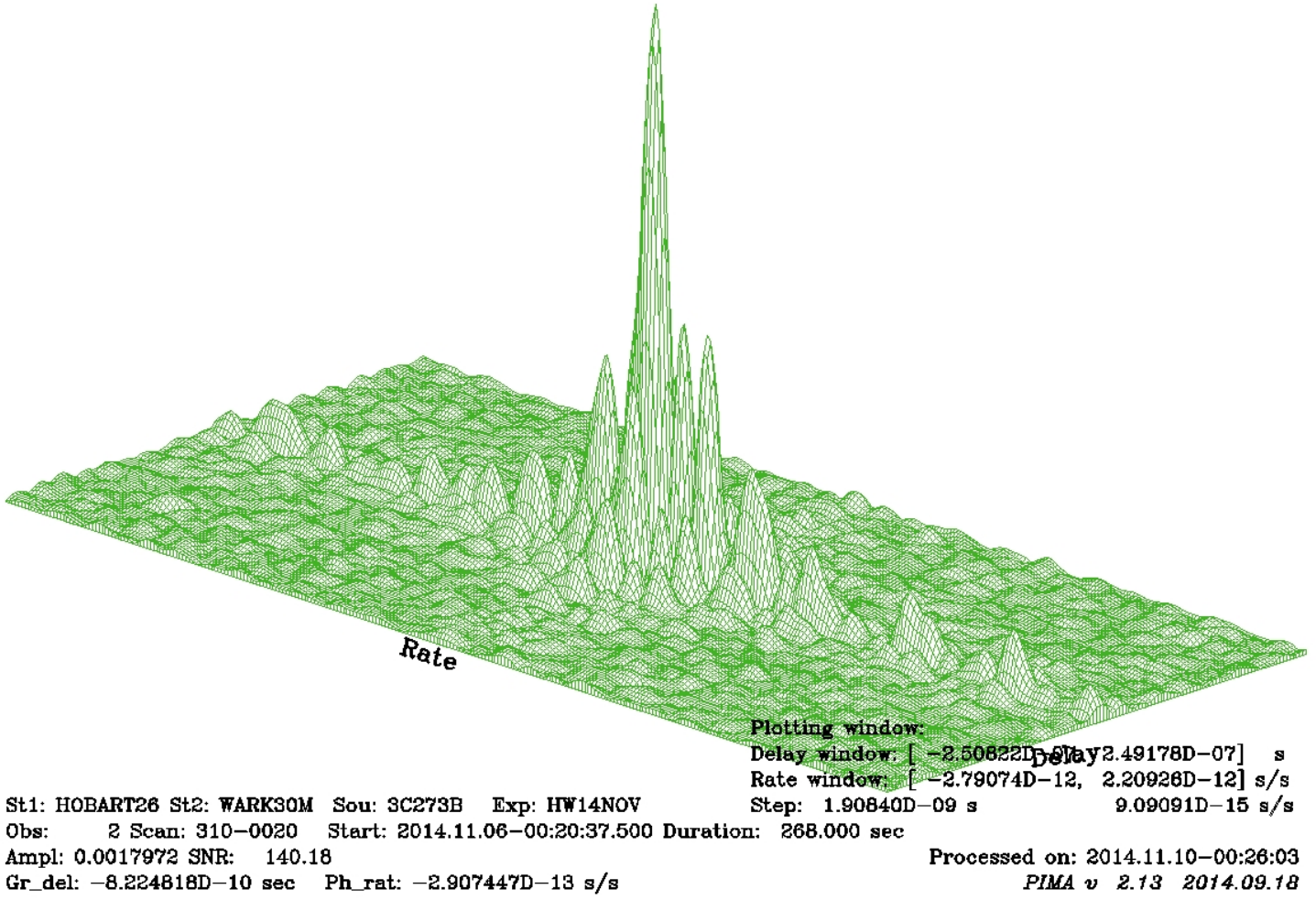


FIGURE 5.18: In this figure I show a more complex radio source EI0032 containing three components. The small open red circle marks the catalogue position for the radio source/component, by its side is the catalogue identification.

5.3 Summary

It was shown in Chapter 3 that there is a symmetry of data points around $R \approx 0.5$, in this Chapter Section 5.1 I highlighted the apparent pairs of XID's about where $0.1 \leqslant R \leqslant 0.9$ and the following investigation of these pairs. From matching with OzDES to obtain their redshift z, we found the majority (32 from 34) of these infrared doubles have a very similar redshift $\Delta z < 0.01$. There are two exceptions where the IRD candidates each have a $\Delta z > 0.22$ and quite clearly are random alignments. For two of the IRD candidates with low values of Δz I obtained HST images and a corresponding FIRST radio images (FIRST has higher resolution than ATLAS), from these it can quite clearly be seen that they are interacting pairs of galaxies with disturbed morphology and tidal tails. This highlights a small sub-group (34) in this data that could have been missed and in the much larger surveys to come would result in a significant number of potentially interesting objects being missed.

In Section 5.2 I looked to see if the algorithm could be extended to identify radio doubles. It is able to select a subset of which over half are probably radio doubles, so this could be a useful method to filter from a larger catalogue a sub-set to review in more detail perhaps visually by the investigator. Currently the algorithm is unable to identify more complex radio morphology such as triples or even more complex sources by identifying 2 or more possible doubles. By further refinement, such as also searching for a common source to two or more doubles the algorithm it is felt could be tuned to also select possible triples; but this is beyond the scope of this work, but in further work it is hoped to further tune and refine this extension to the algorithm. Some such algorithm will be important for the larger surveys to come as the number of objects (10's of millions) in the catalogues and the survey areas (10's thousand square degrees) covered will make it impractical for investigators to search for these objects manually.



A major milestone which I was involved in was the fist VLBI using the AUT Warkworth 30m with the UTas Hobart 26m and Ceduna in obtaining the first fringe on C-Band in 2014. Credit: Leonid Petrov

Chapter 6

Host Galaxy Properties

Using the new catalogue produced by the work in this thesis of cross matched sources between the ATLAS DR3 radio and Fusion IR catalogue, this provides additional information for the ATLAS radio galaxies rather than just radio flux such as redshifts and multi-wavelength IR flux data. This additional data can now be used to look at the source properties in more detail, allowing for more complex analysis and comparisons. The primary question to be addressed is whether our radio sources are AGNs or SFGs. As already stated in Section 1.3, as radio surveys have gone deeper they have moved from being AGN dominated to being more balanced in that SFGs are becoming visible. Some of the work in this Chapter was presented in Weston et al. (2018) and is indicated where appropriate.

6.1 Mid-Infrared and Radio Properties of the Matched Galaxies[1]

In this section the Mid-InfraRed (MIR) properties of the cross-identified sources are used in order to examine the nature of the faint ATLAS radio sources. It has been shown by Stern et al. (2005) and Lacy et al. (2004) that IRAC MIR colours can reliably

[1] Parts of this section were first published in Weston et al. (2018)

isolate AGNs. These methods work by being sensitive to the hot ($\sim 1000\,\mathrm{K}$) dust around the AGN nucleus causing excess emission in the MIR compared to regular SFGs.

6.1.1 Infrared Colour-Colour Plots

First I present MIR colour-colour diagrams which are used to compare the magnitude or flux of astronomical objects at different wavelengths. This in turns allows the examining of the type and evolution of the sources. In this subsection I first compare the mid-infrared colour-colour plots for the cross-identified sources, which provides the values for fluxes of the sources at 3.6μm, 4.6μm, 5.8μm and 8.0μm from the Infrared Array Camera (IRAC) (Eisenhardt et al., 2004).

Also included are the colour evolution of two different parts of the MIR spectra for the sources, $[3.6]-[4.6]$ vs redshift and $[5.8]-[8.0]$ vs redshift. Polycyclic aromatic hydrocarbons (PAH's - (Helou et al., 2000); (Peeters et al., 2004)) are observed in emission from diffuse dust clouds and are tracers of molecular gas in SFGs. So the significance of these MIR bands and PAH's is that the interstellar medium obscures most direct tracers of star formation which lie within the visible and ultraviolet part of the electromagnetic spectrum, but PAH molecules are vibrationally excited by visible and ultraviolet photons and the resultant MIR emissions lines are observable through the interstellar medium. These changes in colour for SFGs are due to the MIR spectral features shown in Figure 6.2 by the solid-black line (The other lines are for a suite of template SED's for different dust parameters). These features shift wavelength as redshift increases. For example the very pronounced feature in the $\approx 5.8\mu$m will move towards the 8.0μm IRAC MIR band and the feature at $\approx 8.0\mu$m will move out of the IRAC MIR bands.

In comparison, AGN do not show the same pronounced PAH's spectral features for SFGs as shown in Figure 6.3 for a generic AGN SED. The unified model (see Figure 6.1) for an AGN consists of a central super massive black hole (SMBH) which range in mass from $10^5 M_\odot$ to $10^9 M_\odot$, around this is a very hot accretion disk. The observers viewing angle is important to how such a AGN would be seen and classified. The accretion disk produces mainly X-Ray, UV and visible light. Perpendicular to the plain

of rotation centred about the axis of rotation relativistic jets of charged particles are emitted, as these jets encounter resistance when penetrating the interstellar material of the host galaxy and the intergalactic medium beyond, very bright radio emission lobes are produced. The X-Ray and UV emission from the accretion disk is obscured by dust, this obscured radiation causes the dust to get very hot which results in the observed ”IR bump” marked in Figure 6.3. This very host dust around the AGN is too hot to stimulate the PAH MIR emission. Some, although not all, AGN produce jets which we observe in the radio spectrum. Many of the radio sources will be powered by such jets, but many will also be powered by star-formation. We can observe this difference in the MIR SED of these sources. Due to this difference in SED between AGNs and SFGs, the MIR results in a smoother evolutionary track of AGN, as the ”IR bump” moves through the MIR band with redshift.

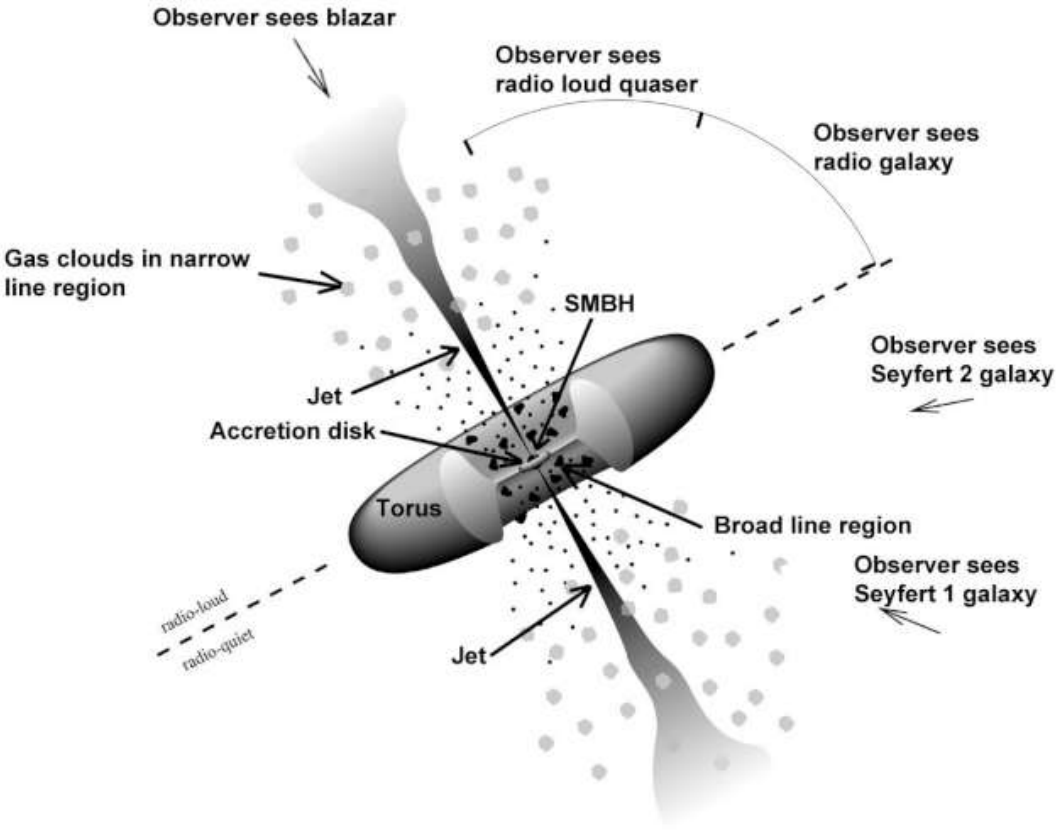


FIGURE 6.1: The Unified Model of an AGN. Credit https://fermi.gsfc.nasa.gov/science/eteu/agn/

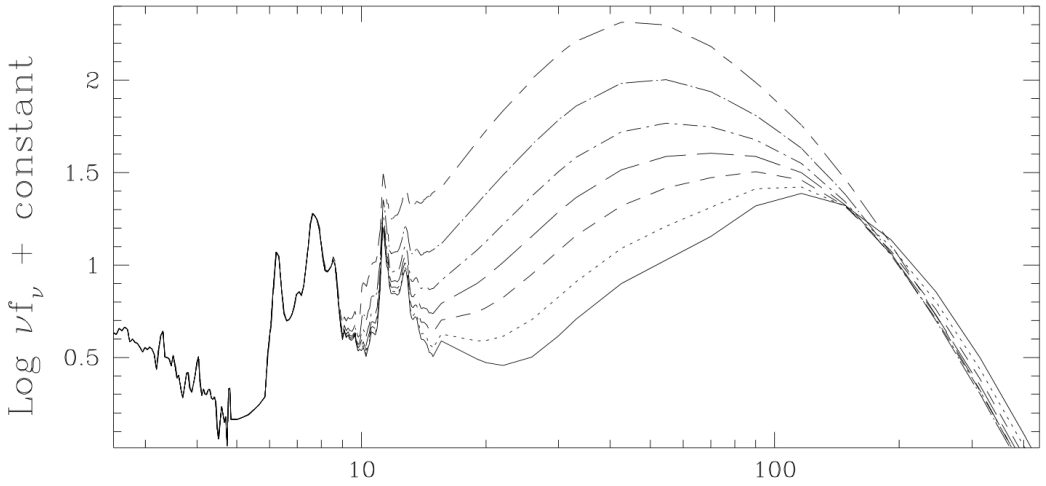


FIGURE 6.2: The spectral energy distribution (SED) of a normal star-forming galaxy at IR wavelengths, taken from Dale and Helou (2002), the x-axis is Wavelength in μm. The solid-black line SED contains spectral features due to PAH's such as the feature centred on 8.0μm that I would like to highlight (the other lines are for a suite of template SED's for different dust parameters).

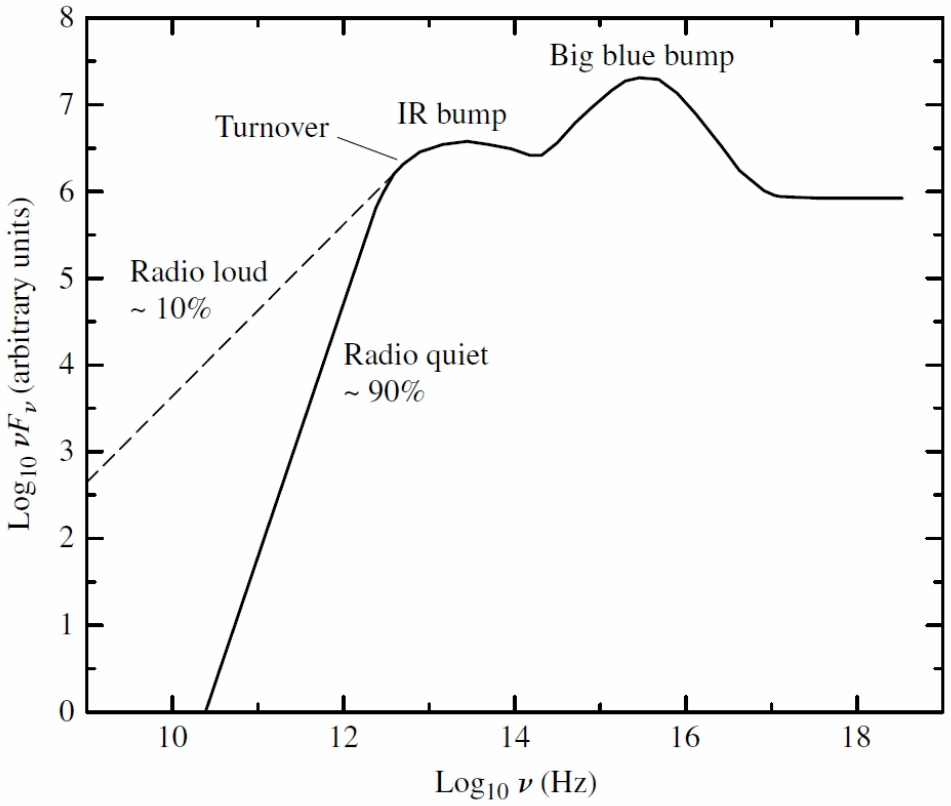


FIGURE 6.3: The spectral energy distribution observed for many types of AGN, taken from Carroll and Ostlie (2006). The AGN heats up the surrounding torus of dust and gas causing it to radiate into the near-IR and it has been found that the MIR spectrum can be used to identify AGNs.

Earlier work by Eisenhardt et al. (2004) presented a vertical spur in the $[3.6]-[4.6]$ versus $[5.8]-[8.0]$ colour-colour diagram which may be associated with AGN (where the magnitude difference $[i]-[j]=-2.5\log\left(S_i/S_j\right)$, where i and j are the wavelengths of the *Spitzer* IRAC bands in μm). This is also supported by Stern et al. (2005) who proposed a region in this parameter space which separates AGN from Galactic stars and SFG. Lacy et al. (2004) presented a $[8.0]-[4.5]$ versus $[5.8]-[3.6]$ colour-colour diagram and also identified an area to select AGN.

Taking the cross-matches (based on the selection criteria in Section 3.1.6) of the ATLAS catalogue against the Fusion catalogue, they will now be used to determine if a source is an AGN or SFG. This is done by plotting these sources in the same colour-colour diagrams as Stern et al. (2005) and Lacy et al. (2004). Not all of the cross-identifications have IR detections in the four *Spitzer* bands and, as such, comparisons can only be made where data is available at all wavelengths. The numbers of radio sources with cross-matches and their break down is presented in Table 6.1.

In the following colour-colour and colour-redshift plots I have included the evolutionary tracks for M82 and NGC4429 taken from Seymour et al. (2007). M82 is a galaxy with high star formation and the SED will be dominated by the PAH's spectral features in the MIR. NGC4429 contains a large old stellar population with little gas and dust, as a result there is little star formation and as a result optically looks redder. Thus their resultant evolutionary tracks provide a reference to mark the boundary between AGN and non-AGN galaxies.

In the following Figures 6.5 and 6.4 are presented two colour-colour plots for the Fusion counterparts to the ATLAS sources. Following Stern et al. (2005), Figure 6.5 presents the $[5.8]-[8.0]$ versus $[3.6]-[4.5]$ colour-colour plot. Figure 6.4 is $[3.6]-[5.8]$ versus $[4.5]-[8.0]$ as per Lacy et al. (2004). Evolutionary tracks from redshift 0 to 2 for a late type starburst galaxy (M82) and an early type galaxy (NGC 4429) are included in both figures. Markers have been placed to indicate $z=0$, $z=1$ and $z=2$. The spectral energy distribution (SED) tracks have been taken from Seymour et al. (2007) based on the work from Devriendt et al. (1999). It can be seen how these evolutionary tracks generally remain outside the Stern et al. (2005) and Lacy et al. (2004) AGN selection

'wedge' (grey shaded areas) and neither would be selected as an AGN candidate if located below $z = 2$ which is inside the designed ATLAS survey limit, confirming the Stern and Lacy AGN selection areas as highlighted in grey.

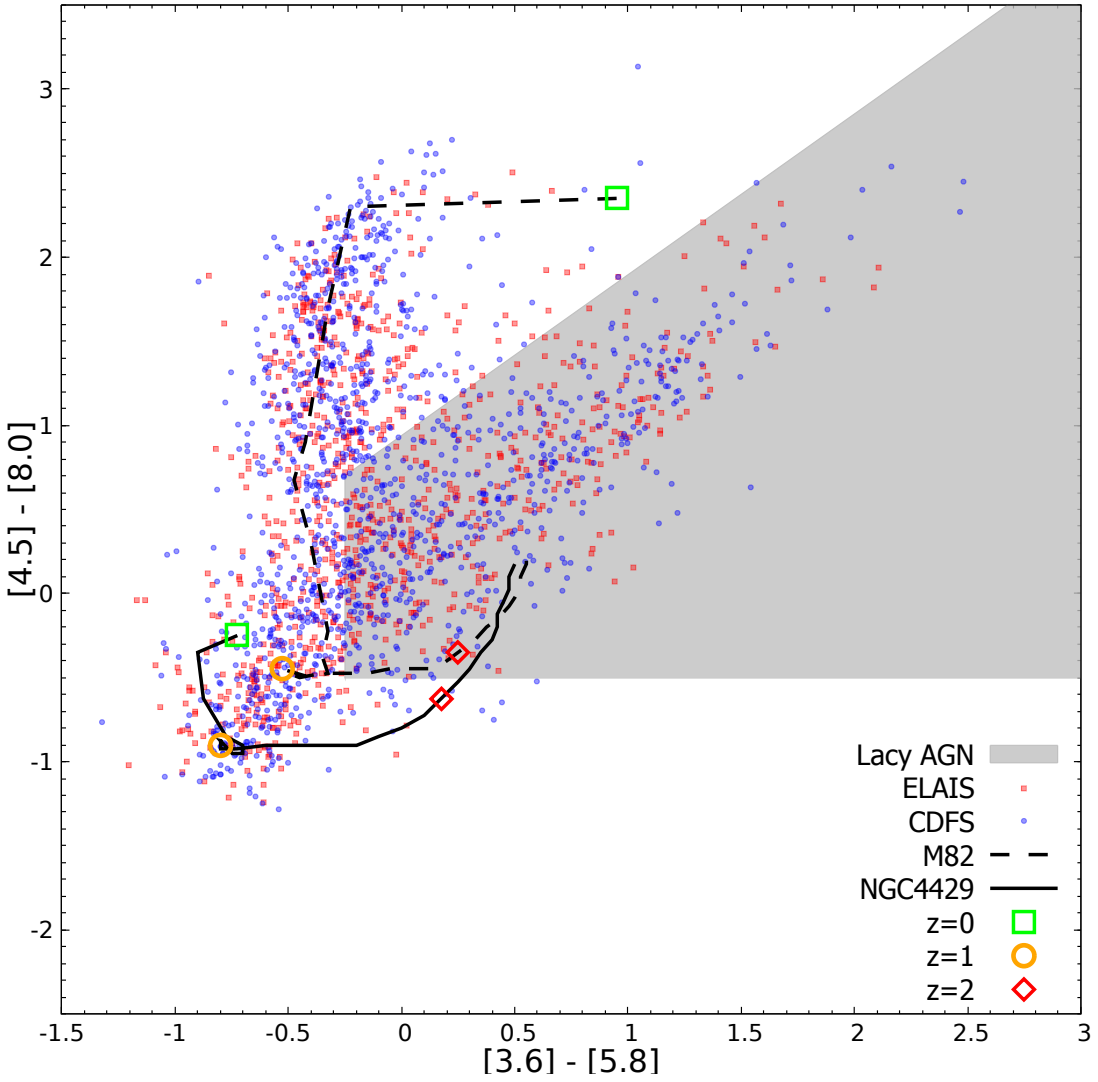


FIGURE 6.4: The $[3.6]-[5.8]$ vs. $[4.5]-[8.0]$ colour-colour diagrams of the Fusion counterparts to the ATLAS sources (as determined in Section 3.1.6) with detections in all four IRAC bands. The grey shaded area showing the location of the Lacy et al. (2004) selection for AGN. The evolutionary tracks for M82 and NGC4429 from $z = 0$ to $z = 2$ taken from Seymour et al. (2007) are included.

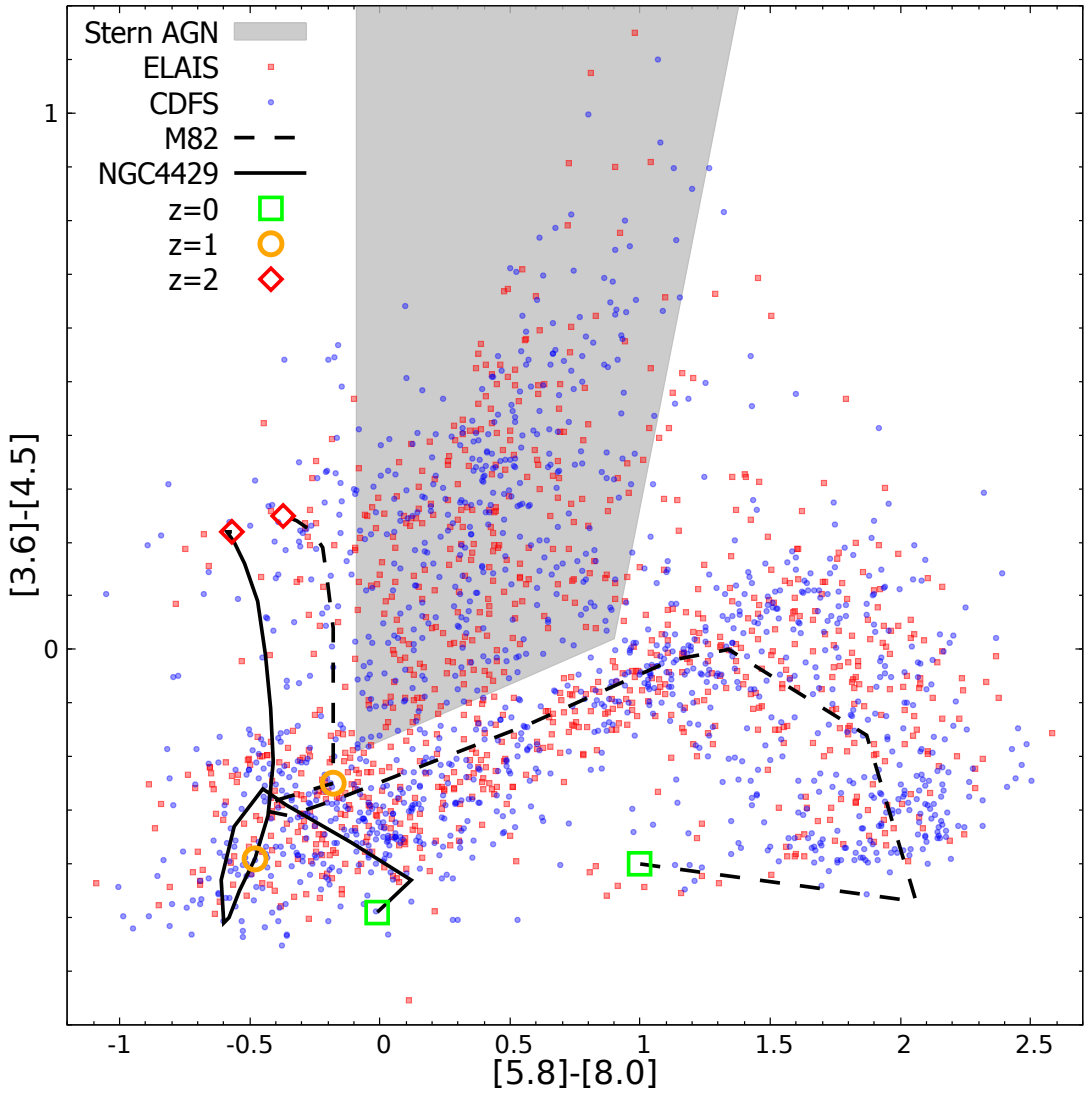


FIGURE 6.5: The $[5.8]-[8.0]$ vs. $[3.6]-[4.6]$ colour-colour diagrams of the Fusion counterparts to the ATLAS sources (as determined in Section 3.1.6) with detections in all four IRAC bands. The grey shaded area shows the location of the Stern et al. (2005) selection for AGN. Also included are the evolutionary tracks for M82 and NGC4429 from $z = 0$ to $z = 2$ taken from Seymour et al. (2007).

Many sources in Figure 6.5 are spread along the evolutionary track of M82 to a redshift of $z = 1$ as there are very few sources in the $z = 1$ to $z = 2$ region of this track. Of note is the vertical spur in the Stern AGN zone grey shaded is consistent with Eisenhardt et al. (2004) and Stern et al. (2005). In Figure 6.4 there is a clear fork with the right hand arm entering the Lacy AGN zone grey shaded. This is also consistent with Lacy et al. (2004). It should be noted that Mao et al. (2012), using the ATLAS DR1 data release and associated spectroscopic classifications, showed that many spectroscopic AGN lay outside the Stern and Lacy wedges.

The next set of Figures 6.6 and 6.7 show the MIR colour evolution of the galaxies, by comparing the different IRAC bands using the flux ratio between $[5.8]-[8.0]$ and $[3.6]-[4.5]$ and how these individual ratios evolve with redshift. First Figure 6.6 shows the evolution of the PAH's spectral features shown in Figure 6.2 at $\approx 8.0\mu$m as the redshift increases, these spectral features move from the 8.0μm IRAC band towards and into the 5.8μm IRAC band. In the figure for $0 < z < 0.3$ we can see a strong correlation for the majority of the galaxies with M82 a star burst galaxy, with a range of 1.0 to ≈ 2.0 for $[5.8]-[8.0]$. When we are at $z > 1$ the vast majority of the galaxies lie well above M82 along a line with some scatter indicating that star formation is not a significant factor in there MIR SED at these redshifts. Across all redshifts we see a range of -1.0 to ≈ 2.5 for $[5.8]-[8.0]$, showing the large contributory factor of the PAH's spectral features in these bands between SFGs and non-SFGs.

In Figure 6.7 the PAH's spectral features are not as strong as in the previous plot, so the change as these features move between the bands is less pronounced with a range of -0.5 to ≈ 0.5 for $[3.6]-[4.5]$. For $0 < z < 0.3$ we can see a strong correlation for the majority of the galaxies with M82 a star burst galaxy. For $0.3 < z < 1.0$ it can be seen that some sources are close to the modelled evolution for NGC4429 but an equal number are well above the M82 track. For $z > 1$ the majority of the galaxies lie well above M82, indicating that star formation is not a factor in these galaxies.

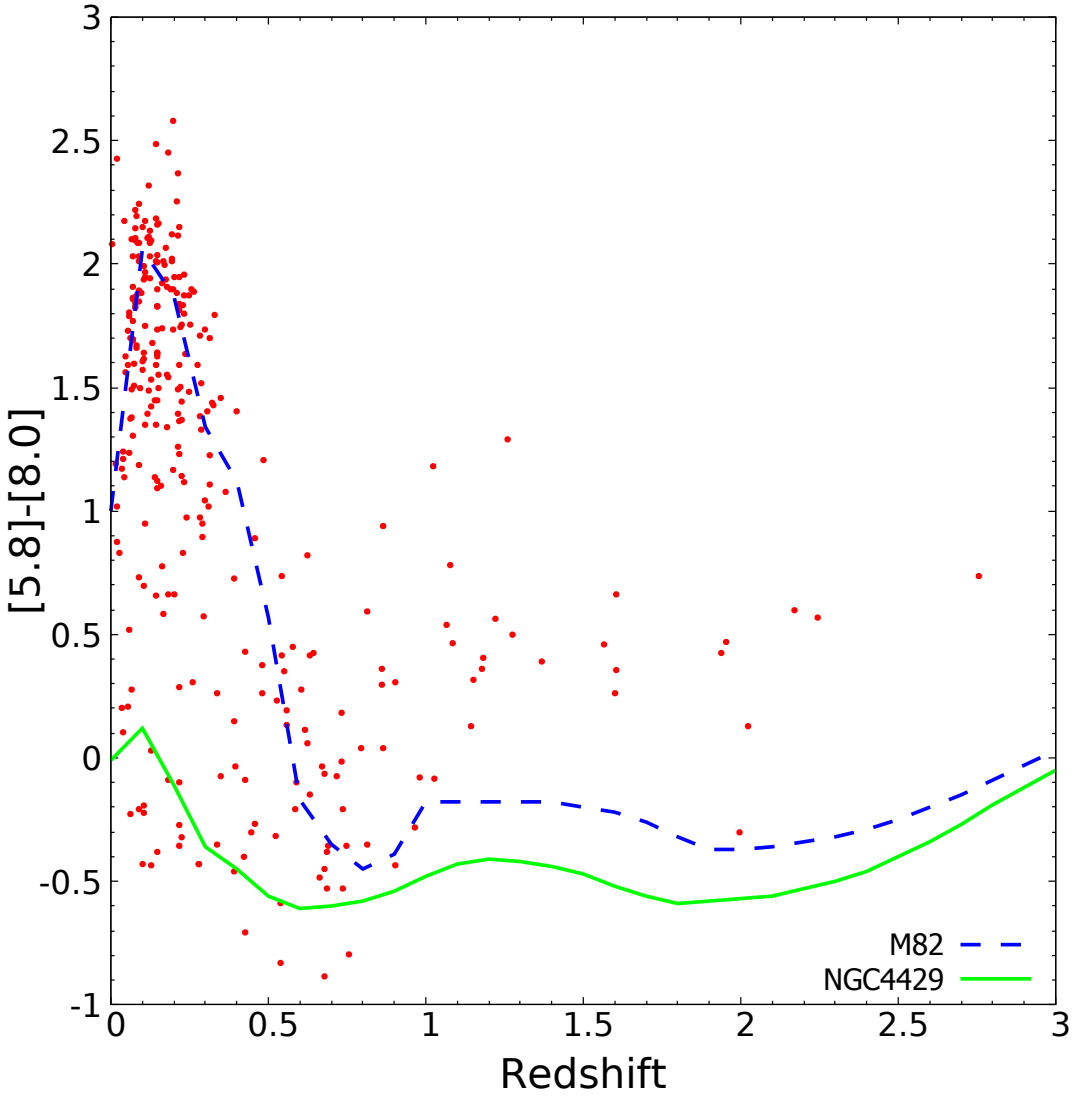


FIGURE 6.6: The $[5.8]-[8.0]$ colour evolution of the Fusion counterparts to the ATLAS sources against redshift z. Also included are the evolutionary tracks for M82 and NGC4429 from $z = 0$ to $z = 3$ taken from Seymour et al. (2007).

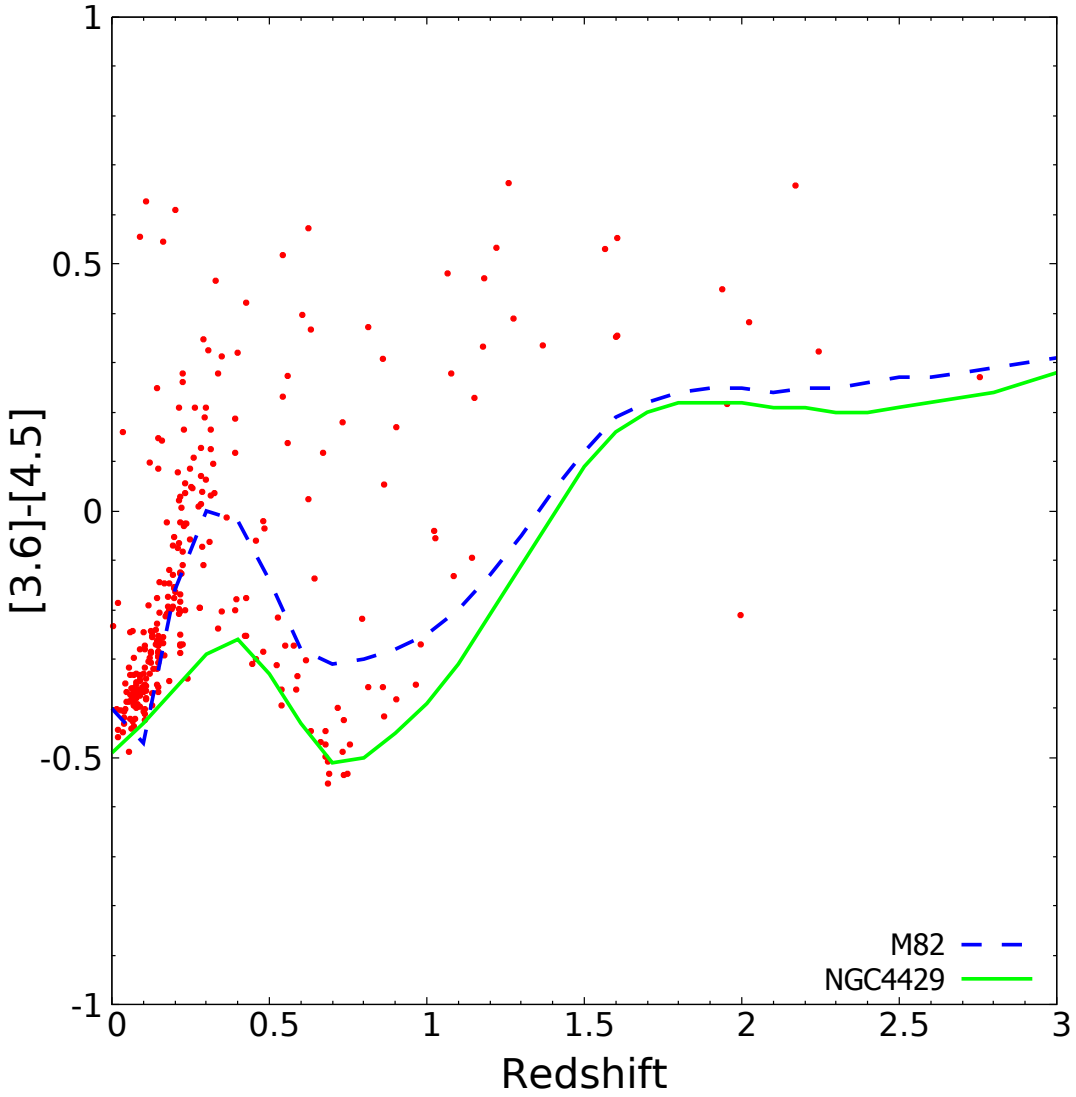


FIGURE 6.7: The $[3.6]-[4.6]$ colour evolution of the Fusion counterparts to the ATLAS sources against redshift z. Also included are the evolutionary tracks for M82 and NGC4429 from $z = 0$ to $z = 3$ taken from Seymour et al. (2007).

In Figure 6.8 is shown the Type 1 AGN selection suggested by Richards et al. (2006) where the criteria is $[3.6] - [4.5] > -0.1$. Applying this to the data set, there are 580 sources from the ELAIS field and 454 from the CDFS fields; a total 1034 out of 1984 (52%) with complete IRAC data and as determined by our XID selection criteria in Section 3.1.6 would be classified as Type 1 AGN. It has been noted by Mao et al. (2012) that this provides a selection with considerable contamination by SFG. Looking at the evolutionary track of NGC4429 compared to Figure 6.5 using the Stern method, it can be seen that it crosses the threshold into the AGN space between $0 < z < 1$, M82 stays to the left of the threshold until $1 < z < 2$ which is consistent with Stern and Lacy.

By using the selection criteria in Section 3.1.6 for Fusion cross-identifications, in the Lacy AGN selection zone there are a total of 848 XIDs and for the Stern AGN selection zone there are a total of 533 XIDs. A total of 956 XIDs satisfy the union of the Stern and Lacy selection criteria for AGN.

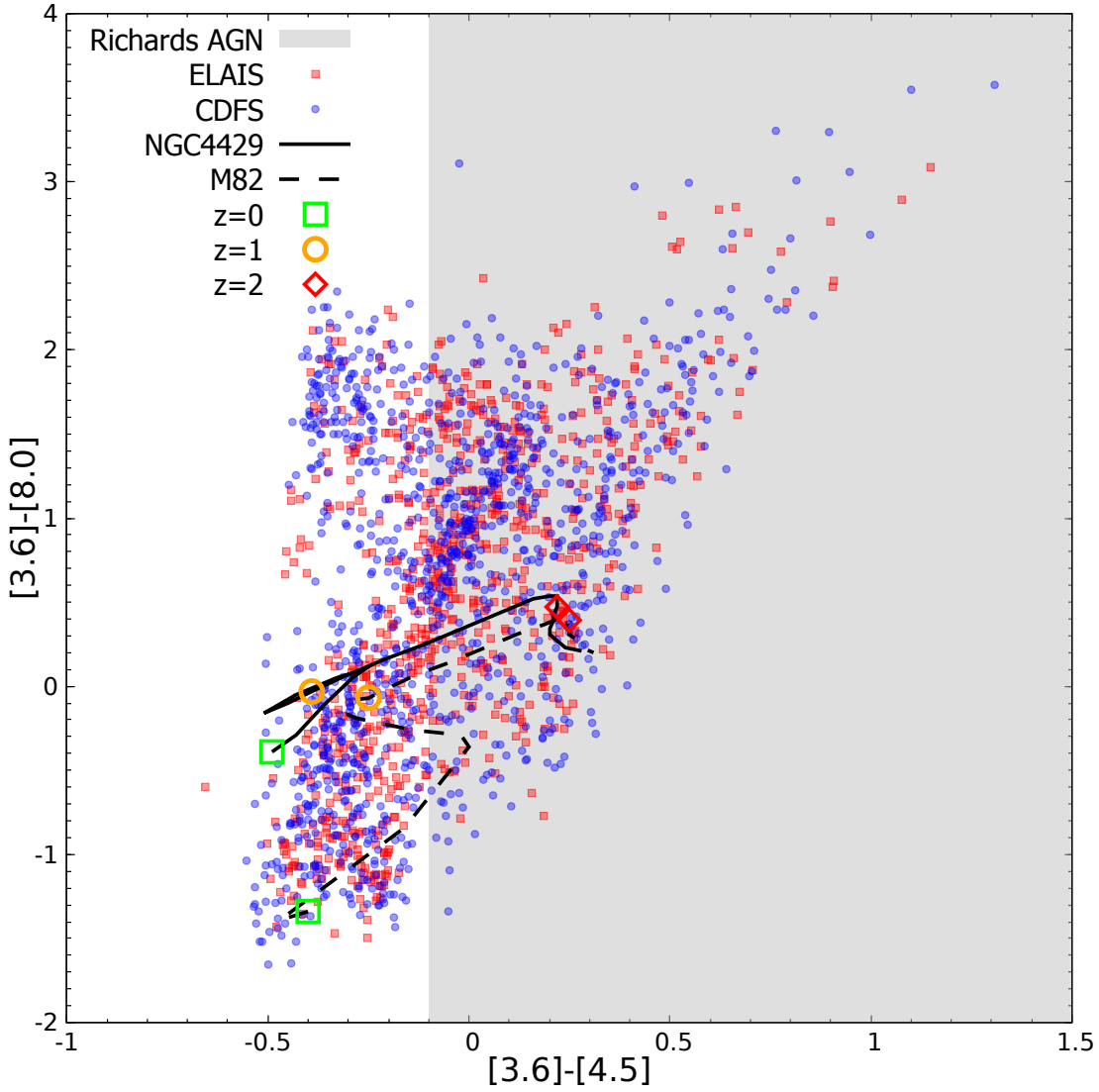


FIGURE 6.8: The $[3.6]-[4.5]$ vs. $[3.6]-[8.0]$ colour-colour diagrams of the Fusion counterparts to the ATLAS sources (as determined in Section 3.1.6) with detections in all four IRAC bands. The grey shaded area showing the location of the Richards et al. (2006) selection for AGN. The evolutionary tracks for M82 and NGC4429 from $z = 0$ to $z = 2$ taken from Seymour et al. (2007) are included.

6.1.2 Flux Density Ratio of $S_{1.4\,\mathrm{GHz}}$ to $S_{3.6\,\mu\mathrm{m}}$ versus z

In order to account for the relative radio emission from radio loud AGN (RLAGN), the radio to 3.6 μm flux density ratio is examined for all of the cross-matched sources with known redshifts for the sample in Figure 6.9. Now to compare these to tracks of known sources shifted to higher redshifts (i.e. compare the ratio of the observed frame 1.4 GHz and 3.6 μm flux densities shifted with redshift). For comparison are included the tracks for the radio-loud and radio-quiet AGN from Seymour et al. (2008) based on Elvis et al. (1994) templates and the two galaxies used in the previous section (the starburst M82 and the quiescent galaxy NGC4429). The galaxy template tracks are relatively flat although some variation is seen with redshift. To be noted is that the AGN templates are for unobscured AGN which do not include any potential obscuration of the AGN by dust from a torus or the host galaxy. Any obscuration would increase these flux ratios by suppressing the observed 3.6 μm flux density as it gets shifted to the Near-InfraRed (NIR) and optical rest-frame at higher redshift.

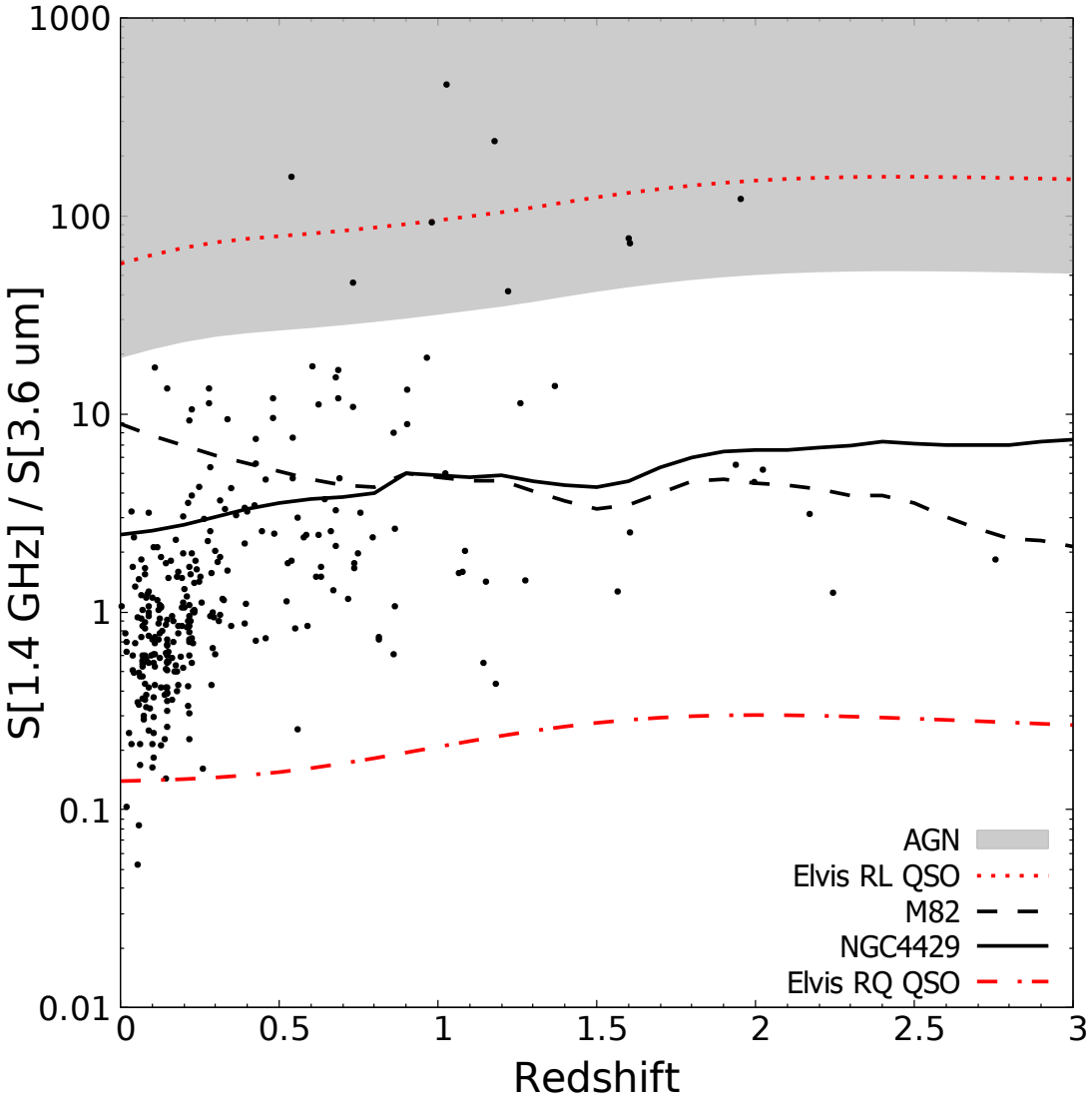


FIGURE 6.9: The ratio between the radio 1.4 GHz and Fusion 3.6 μm flux density plotted as a function of redshift for all XIDs (determined in Section 3.1.6). The red dotted line near the top of the figure indicates the loci of a classical radio-loud QSO from Seymour et al. (2008) based on Elvis et al. (1994), and the red dot-dashed line in the lower part of the figure indicates the loci for radio-quiet QSO from Seymour et al. (2008) based on Elvis et al. (1994). The grey shaded area denotes the population that identifies the radio loud AGN. Also included are the evolutionary tracks for M82 and NGC4429 from $z = 0$ to $z = 3$ taken from Seymour et al. (2007).

The redshifts for 295 sources come from the OzDES global redshift catalogue (Yuan et al., 2015, Lidman et al. in prep). It can be seen that most of the redshifts are at $z < 0.3$, which is due to the targeting of the brightest optical counterparts by OzDES and earlier surveys (Mao et al., 2012), although there is a tail to $z \sim 2.8$. This $z < 0.3$ grouping typically have a flux ratio from ~ 0.2 to 2, below that for the starburst and quiescent galaxy tracks. Why does our group of SFGs have lower flux ratios than these two sources? To first approximation it is possible to say that, if these are star forming galaxies, the radio emission traces the star-formation rate (SFR) and the 3.6 μm emission traces the stellar mass. Hence the galaxies in this group are likely similar to M82, but with a lower specific SFR (SFR per unit stellar mass). As the ATLAS sources are selected on a SFR proxy, radio flux, and lie at higher redshift, it is likely the higher stellar masses are pulling the observed ratio down despite the higher SFRs compared to M82.

In terms of identifying which sources in this plot have radio emission powered by AGN, it is possible to use the RL AGN track as a guide (Seymour et al., 2008). Allowing for uncertainty in the models and the fluxes it can be suggested that any radio source with a ratio greater than one third of the track from the RL AGN, marked by the grey shaded area in Figure 6.9, is likely powered by an AGN. In this work are found only nine sources above this line which also have a redshift value. there are also 57 matches with no redshift and the 3.6 μm flux is below the detection limit. Taking their radio flux and dividing by the 3.6 μm flux limit would place these matches in the grey AGN region.

6.1.3 Flux Density Ratio of $S_{1.4\,\mathbf{GHz}}$ to $S_{8.0\,\mu\mathbf{m}}$ versus z

As was shown in Figure 6.2 there is a very strong SED feature at $\approx 8.0\mu$m due to a PAHs which is used as a marker for SFGs. At low redshifts the 8.0μm PAH feature will dominate the MIR in the IRAC 8.0μm band. As we move to higher redshift this feature will be less dominate and radio will be become the more dominate.

Taking the XID's from this thesis and finding those with redshift and 8.0μm flux values, we obtain a total set of 66 galaxies. As can be seen in Figure 6.10 there is a cluster of sources bound by $z < 0.5$ and $S[1.4\,\mathrm{GHz}]/S[8.0\,\mu\mathrm{m}] < 1.0$. These lie below the evolutionary track of M82 (a star burst galaxy) and thus are star forming galaxies.

We have a total of 66 galaxies with values for z and $S[1.4\,\mathrm{GHz}]/S[8.0\,\mu\mathrm{m}]$, of these 21 (32%) have a $S[1.4\,\mathrm{GHz}]/S[8.0\,\mu\mathrm{m}] > 1.0$ and there are 45 with a flux ratio < 1.0 (68%). Within the rectangle bounded by $z < 0.5$ and $S[1.4\,\mathrm{GHz}]/S[8.0\,\mu\mathrm{m}] < 1.0$ there are 44 (66%).

There will be some selection affect at the higher redshifts and not all ATLAS sources have $\approx 8.0\mu$m and redshift values.

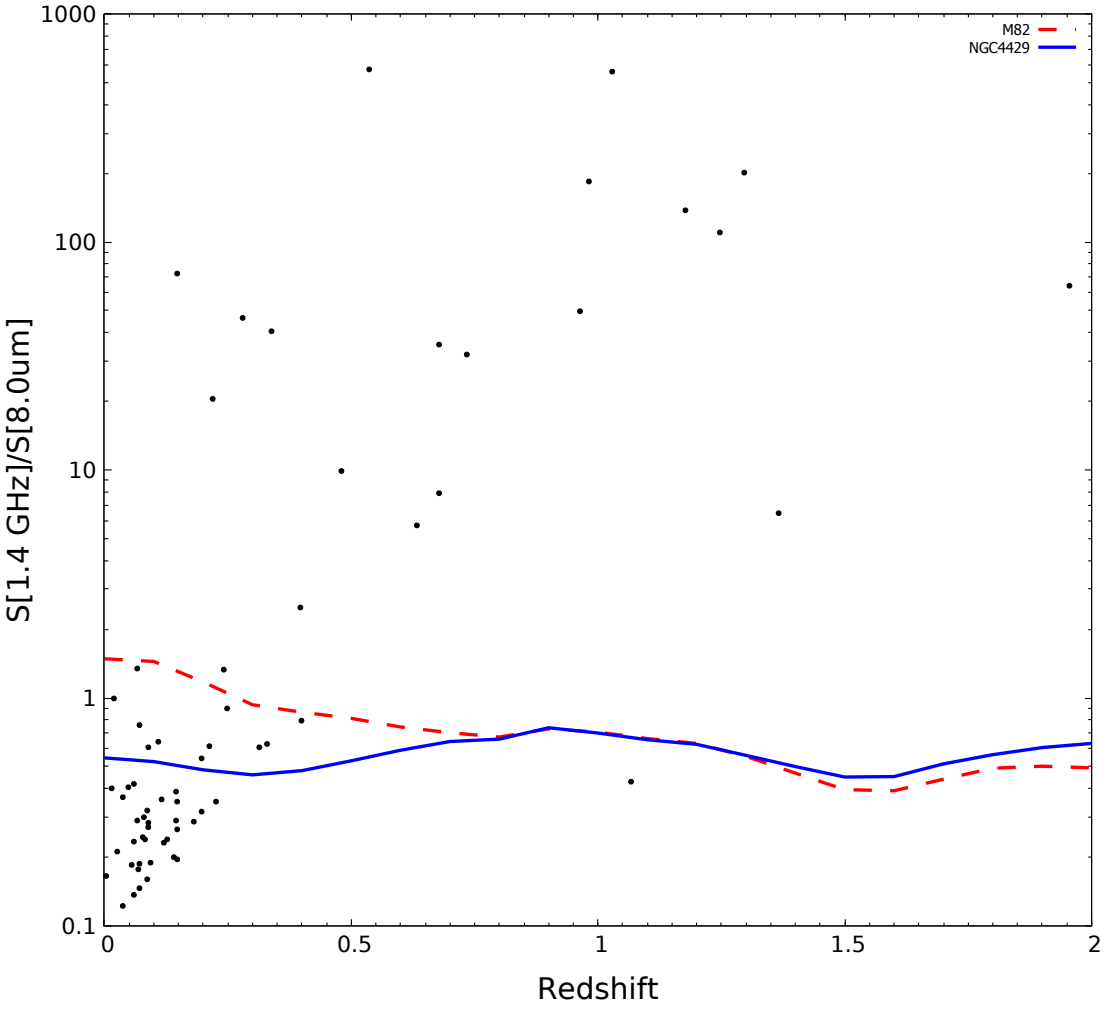


FIGURE 6.10: The ratio between the ATLAS radio 1.4 GHz and Fusion 8.0 μm flux density plotted as a function of redshift for all XIDs (determined in Section 3.1.6). Also included are the evolutionary tracks for M82 (red dashed line) and NGC4429 (blue solid line) from $z = 0$ to $z = 2$ taken from Seymour et al. (2007).

6.1.4 The IRAC $S_{8.0\mu m}$ to Radio $S_{1.4\,GHz}$ Relationship

It was shown by Mingo et al. (2016) plotting $S_{1.4\,GHz}$ against WISE 12.0μm (W3 band) magnitude that two separate populations existed. They highlighted the SFGs and AGNs and demonstrated that a population with $10 < W3 < 13$ magnitude was clearly dominated by AGNs with some contamination by spirals and elliptical galaxies. The population for $2 < W3 < 10$ magnitude was dominated by starburst galaxies with some spirals and a very small number of elliptical galaxies.

In the Fusion data there exists $S_{24.0\mu m}$ MIPS data, but no similar separation of two groups existed when taking the $S_{24.0\mu m}$ against $S_{1.4\,GHz}$ values. Looking at Figure 6.2 there is a strong PAH spectral feature at $\approx 12.0\,\mu$m and another at $\approx 8.0\,\mu$m.Taking the $S_{8.0\mu m}$ IRAC data for the XID's in this work, I noticed a similar split in the sources producing two distinct groups (Figure 6.11). There is clearly one group with low IR flux $S_{8.0\mu m} < 200\,\mu$Jy and a wide range of Radio Flux between $0.2 < S_{1.4\,GHz} < 110\,m$Jy. The second group extends over a large range of IR flux values $200 < S_{8.0\mu m} < 11,000\,\mu$Jy and has a smaller range of radio flux values $0.2 < S_{1.4\,GHz} < 5\,m$Jy.

In Figure 6.12 I have marked those galaxies (red dots ●) that via the Stern or Lacy selection criteria would be identified as AGN. In the top plot of the figure for the Lacy AGN selection criteria, it can be seen that the majority of these sources lie within the first group described in the previous paragraph. In the bottom plot of the figure for the Stern AGN selection criteria, a slightly different distribution is apparent but the majority again lie in this first group with some appearing in the second grouping.

There are a total of 212 sources with values for all the fields required. I draw a blue vertical line at $S_{8.0\mu m} = 200.0\,\mu$Jy which visually appears to separate the majority of the Sten/Lacy AGN in the figures. There are a total of 76 sources (out of 212) that match the Lacy AGN selection criteria, and most of them (72 or 94%) have $S_{8.0\mu m} < 200.0\,\mu$Jy. There is a smaller number (47) meeting the Stern AGN selection criteria and for which 28 (59%) have $S_{8.0\mu m} < 200.0\,\mu$Jy. For the whole set there are a total of 109 sources where $S_{8.0\mu m} < 200.0\,\mu$Jy, giving 66% which meet the Lacy AGN selection criteria and

25% meet the Stern AGN selection criteria. It is important to note that this data is not complete for all fluxes, it is a small subset (212) of the total number of XID's (1987).

So it becomes apparent that AGNs have low IR flux ($S_{8.0\mu\mathrm{m}}$). When plotted against their Radio Flux they can be separated from SFGs. In summary, the plot of Radio $S_{1,4\,\mathrm{GHz}}$ versus IR $S_{8.0\mu\mathrm{m}}$ can be considered as an additional useful tool for distinguishing between AGNs and SFGs.

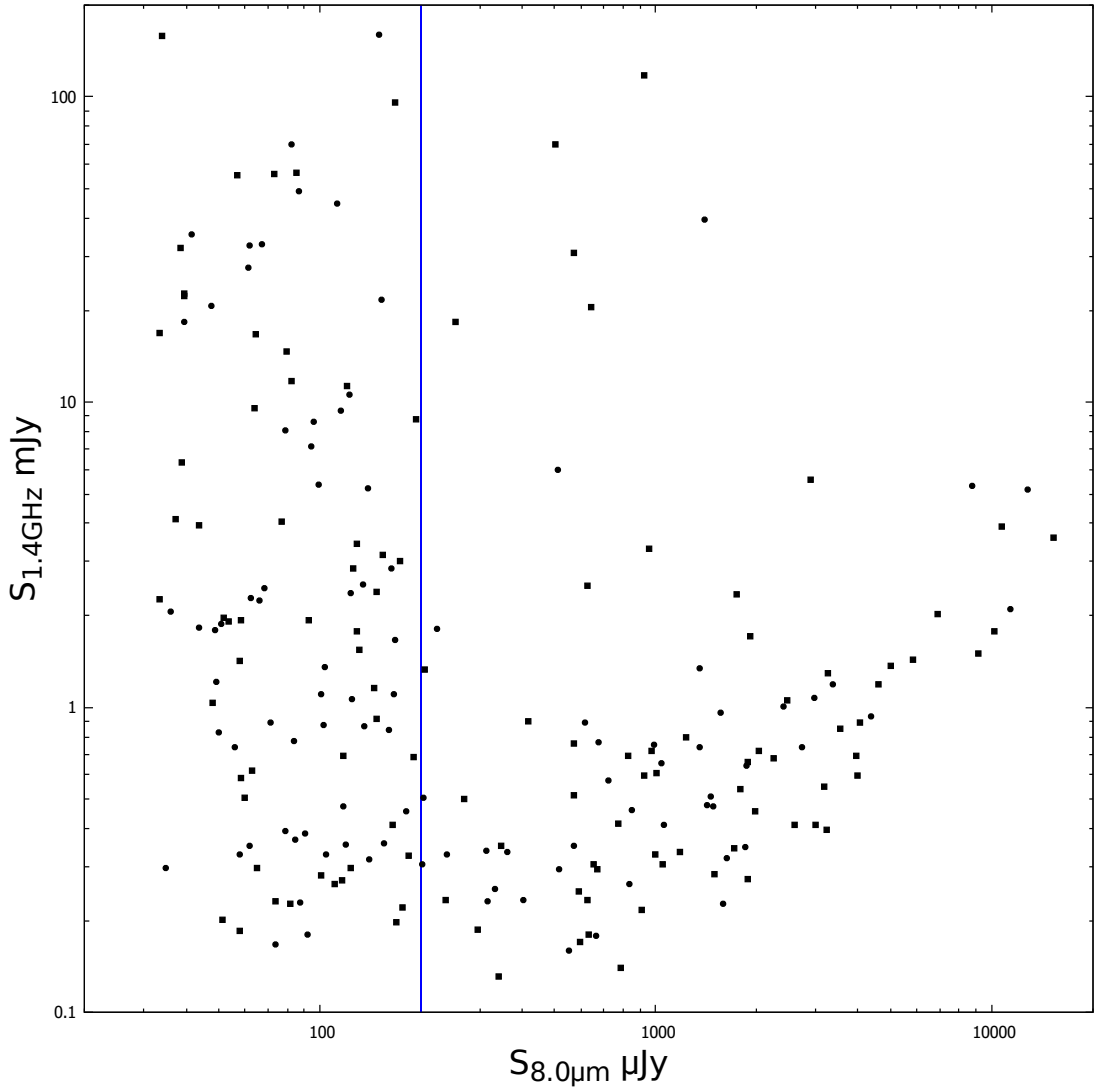


FIGURE 6.11: $S_{8.0\mu m}$ versus $S_{1.4GHz}$ For all ATLAS DR3 cross identifications that have a radio flux $S_{1.4GHz}$ and IRAC $S_{8.0\mu m}$ values. The blue vertical line indicates the limit $S_{8.0\mu m} = 200.0\mu$Jy below which the AGN reside.

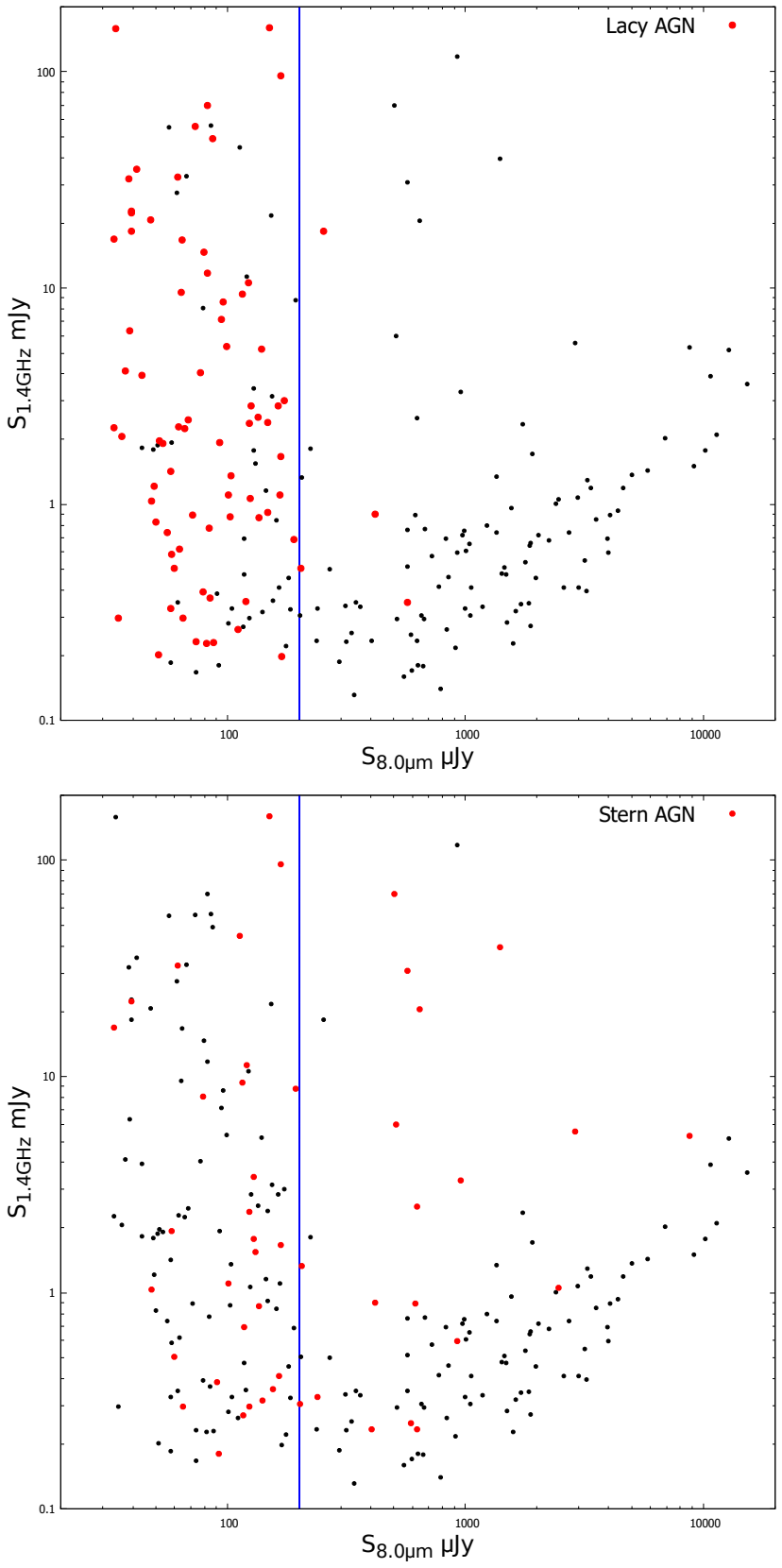


FIGURE 6.12: The top plot shows $S_{8.0\mu\mathrm{m}}$ versus $S_{1.4GHz}$ for all ATLAS DR3 cross identifications that have a radio flux $S_{1.4\mathrm{GHz}}$ and IRAC $S_{8.0\mu\mathrm{m}}$ values with the Lacy wedge AGNs highlighted in red (●). The bottom plot shows $S_{8.0\mu\mathrm{m}}$ versus $S_{1.4\mathrm{GHz}}$ for all ATLAS DR3 cross identifications that have a radio flux $S_{1.4\mathrm{GHz}}$ and IRAC $S_{8.0\mu\mathrm{m}}$ values with the Stern wedge AGNs highlighted in red.The blue vertical line indicates the limit $S_{8.0\mu\mathrm{m}} = 200.0\mu$Jy below which the AGN reside.

6.1.5 Results of AGN Identification

Taking the identification of Fusion counterparts to determine if our sample of ATLAS sources are AGN or SFG by the various methods in the previous sections is summarised in Table 6.1. There are found 1987 cross matched radio sources with flux values for all four IRAC bands. Of these ($\approx 27\,\%$) meet the Stern AGN selection criteria, and ($\approx 43\,\%$) Lacy, in addition 48 ($\approx 2\,\%$) of these sources lie above a line one third of the RL AGN track. Taking the union of all of these across the three selection criteria there are 956 ($\approx 48\,\%$) which are possible AGN.

TABLE 6.1: Results of the classification of cross-identification of ATLAS sources. The first row presents the total number of XIDs with a complete set of IRAC bands. The following rows show the AGN identified by the three methods (Stern, Lacy and Flux Density ratio) followed by the total number of AGN identified, i.e. the union of the preceding three sets, and the percentage.

Field	CDFS	ELAIS-S1	both
# with Fusion XID and all IRAC bands	1153	834	1987
# AGN			
Stern	298	235	533
Lacy	490	358	848
Ratio	29	19	48
Total $\cup$	550 (48%)	406 (49%)	956 (48%)

6.2 AGN Power Law Discussion

In Condon et al. (1991) it is mentioned that the FIR-radio correlation can be fitted with a power-law given by $L_{FIR} \propto L^{\gamma}_{151MHz}$ with $\gamma = 0.86 \pm 0.03$. Interestingly Lacy et al. (2004) makes no mention of the power-law for MIR AGN, and I will show that there is in fact a very strong correlation of AGN sources in this parameter space to a power-law. In comparison Stern et al. (2005) does mention the power-law, and notes that quasars have a power-law SED, $f_\nu \propto \nu^{-\alpha}$ and in the figures for z vs $[3.6]-[4.5]$ and z vs $[5.8]-[8.0]$ marks the power-law for $0.5 < \alpha < 2.0$ but interestingly does not include this in the colour-colour figure with the plot of $[5.8]-[8.0]$ vs $[3.6]-[4.5]$ showing the AGN selection wedge.

Vanden Berk et al. (2001) who used the SDSS observed wavelength range of 3800-9100 Å to look at the spectra of quasars, they highlight that the continuum spectra was well fitted by two power laws with a break at $\approx 5000\mathring{A}$. They note that the frequency power-law index, α_ν, is -0.44 from ≈ 1300 to $5000\mathring{A}$ and -2.45 from ≈ 5000 to $9100\mathring{A}$.

In Alonso-Herrero et al. (2007) was presented a clear demonstration of applying the power-law to radio galaxy IR colours for the selection of AGN. Also Donley et al. (2012) noted that the IRAC power-law selection only identifies the most robust colour selected AGN, and demonstrated the power-law locus (the line on which a source with a perfect IRAC power-law SED would fall) on the Lacy and Stern AGN selection wedges

For the work that follows I use the following for the power-law, flux density depends on frequency as $F(\nu) = A\nu^{-\alpha}$, where α is the spectral index. For the ratio of fluxes we get :

$$\frac{F(\nu_1)}{F(\nu_2)} = \left(\frac{\nu_1}{\nu_2}\right)^{-\alpha} = \left(\frac{\lambda_1}{\lambda_2}\right)^{\alpha} \tag{6.1}$$

and

$$\frac{F(\nu_3)}{F(\nu_4)} = \left(\frac{\lambda_3}{\lambda_4}\right)^{\alpha} \tag{6.2}$$

Applying logarithm to both sides of 6.1 and 6.2 and denoting :

$$y \equiv \log \frac{F(\nu_1)}{F(\nu_2)} \tag{6.3}$$

and

$$x \equiv \log \frac{F(\nu_3)}{F(\nu_4)} \tag{6.4}$$

we get

$$y = \alpha \log \frac{\lambda_1}{\lambda_2} \tag{6.5}$$

and

$$x = \alpha \log \frac{\lambda_3}{\lambda_4} \tag{6.6}$$

And finally,

$$y = \beta x \tag{6.7}$$

where

$$\beta = \frac{\log(\lambda_1/\lambda_2)}{\log(\lambda_3/\lambda_4)} \tag{6.8}$$

For the Lacy plot the IRAC wavelengths are $\lambda_1 = 4.5\mu m$, $\lambda_2 = 8.0\mu m$, $\lambda_3 = 3.6\mu m$ and $\lambda_4 = 5.8\mu m$. They are for Stern 3.6, 4.5, 5.8 and $8.0\mu m$. Therefore, for the à la Stern option $\beta = 0.693$, and for the à la Lacy option $\beta = 1.2036$. These theoretical slopes are shown as green solid lines in the following plots for this section, and the range of α has been selected for each such that the line fits within the different wedges and available data points.

Taking all the sources from Fusion *Spitzer* that have complete flux values for all IRAC bands, this gives a set of 21248 sources providing a statistically large sample for checking the power-law correlation. The resultant Stern and Lacy AGN selection wedges are presented in Figure 6.13, where the top plot is for the Lacy criteria and the bottom plot is for Stern. What stands out very clearly is the strong correlation in the $[3.6]-[5.8]$ vs $[4.5]-[8.0]$ colour-colour plot for the right hand arm of the Y that quite clearly follows the power-law locus, which lies within the Lacy AGN wedge. The correlation in the $[5.8]-[8.0]$ vs $[3.6]-[4.5]$ colour-colour plot for those objects within the Stern AGN

selection wedge for the power-law locus is not as ”strong” with significant scatter about the locus.

I highlight that there is a very high density of sources in the area that the evolutionary tracks for M82 and NGC4429 reside at for $z \approx 1$. In the top plot for Lacy this high density area lies within $-1.1 < [3.6] - [5.8] < -0.7$ and $-1.4 < [4.5] - [8.0] < -0.7$, the bottom plot for Stern this area lies within $-1.0 < [5.8] - [8.0] < -0.3$ and $-0.6 < [3.6] - [4.5] < -0.3$.

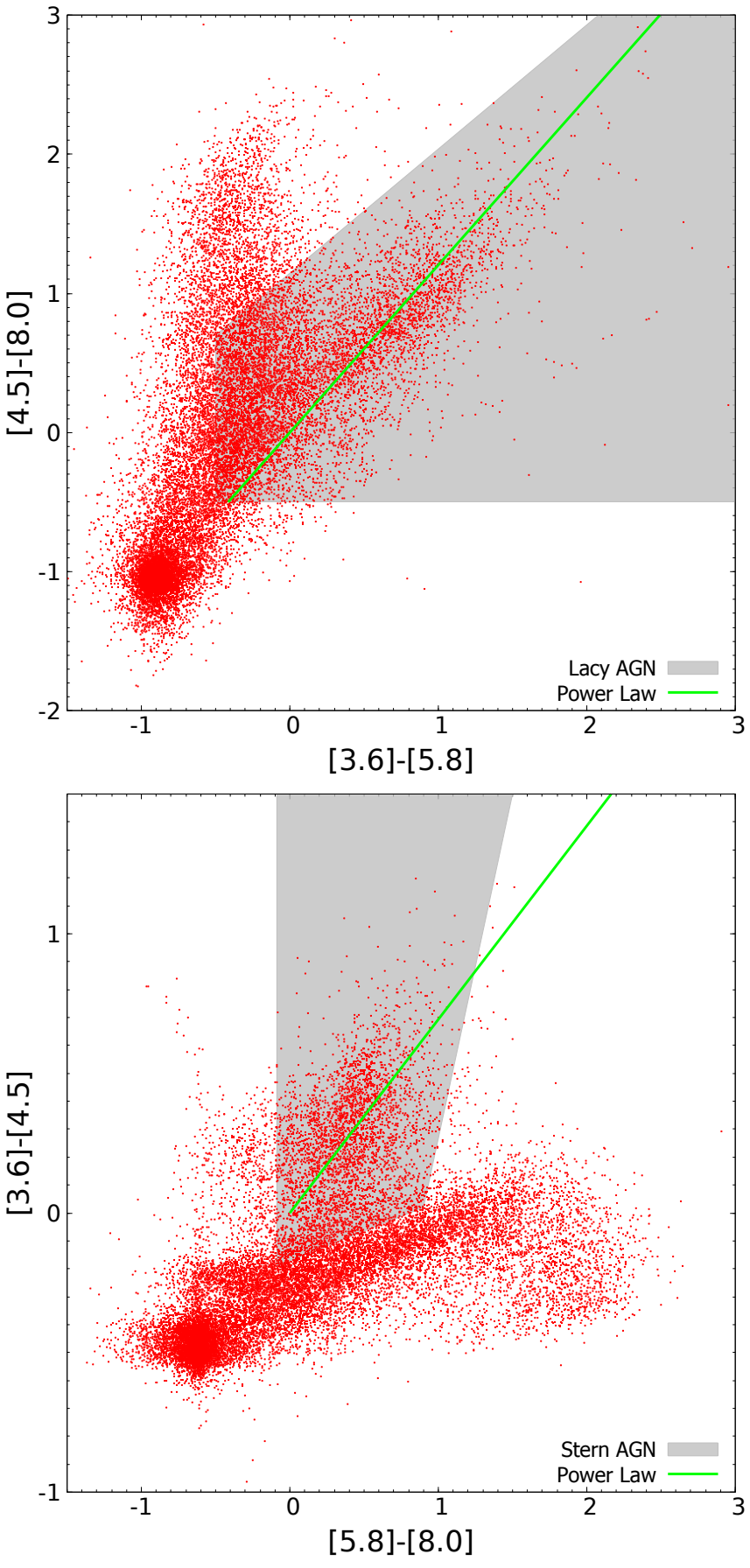


FIGURE 6.13: The same colour-colour diagrams of all Fusion with detections in all four IRAC bands. The grey shaded area showing the location of the appropriate AGN selection region. The green line is the theoretical power-law for the ratio of the IR frequencies.

In Figure 6.14 I have repeated the same plots as shown in Figure 6.13, but only for the XID's from this work. In the top plot of Figure 6.14 is the Lacy AGN selection criteria, again the dispersion of AGNs around the theoretical power-law is very small and the sources are practically on the power-law locus straight line.

The bottom plot in Figure 6.14 is for the Stern AGN selection criteria, here it can be seen that there is a lot of dispersion of AGN candidates about the power-law locus straight line.

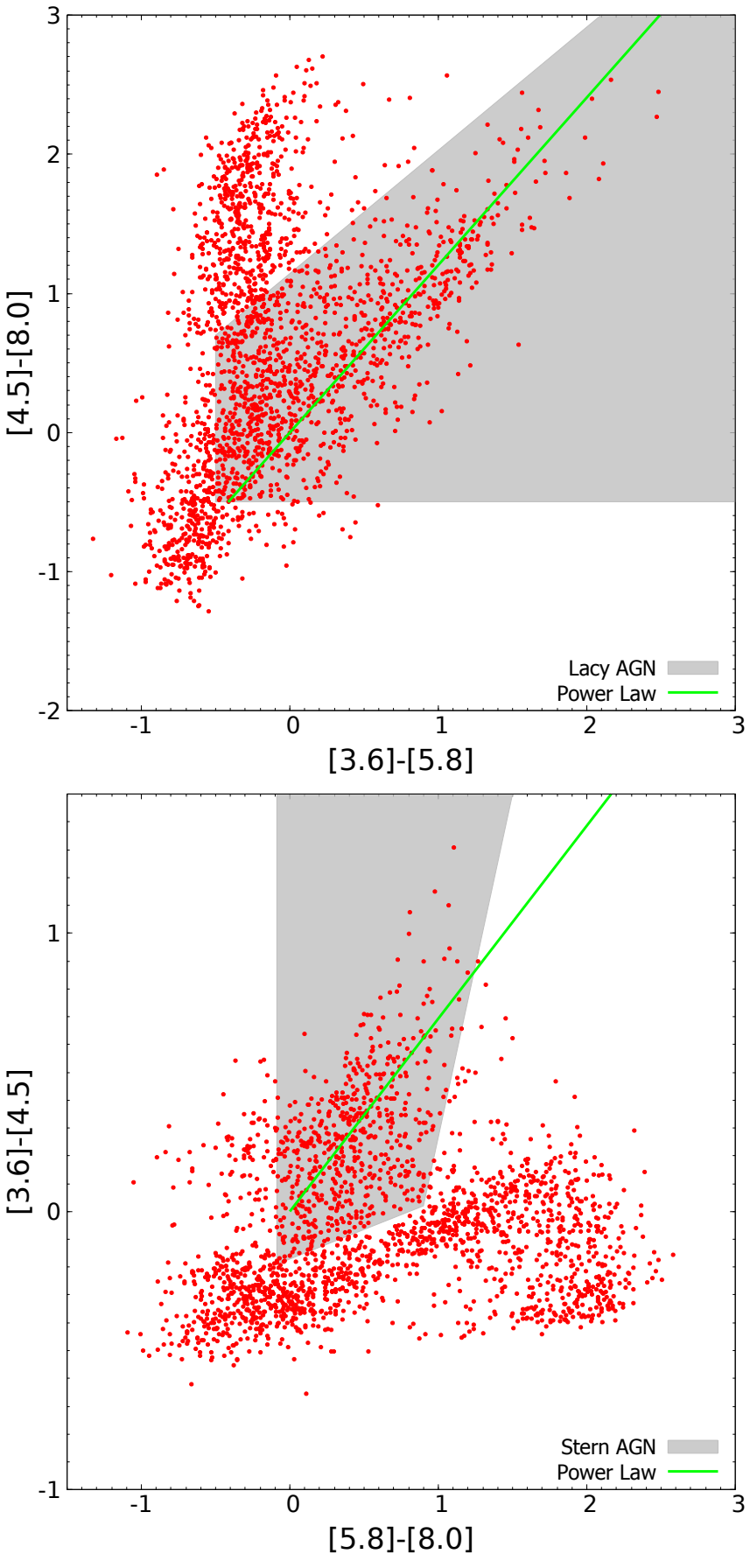


FIGURE 6.14: The same colour-colour diagrams of the FUSION *Spitzer* counterparts to the ATLAS sources (as determined in Section 3.1.6) as in Figures 6.5 and 6.4 with detections in all four IRAC bands. The grey shaded area showing the location of the appropriate AGN selection region. The green line is the theoretical power-law for the ratio of the IR frequencies.

In summary it is apparent that for this sample using the Lacy AGN selection colour-colour criteria the correlation between the AGN candidate IR flux ratios to the power-law locus is very strong. Using the Stern colour-colour criteria no strong correlation for the AGN candidate IR flux ratios to the power-law locus is apparent.

6.3 Radio Luminosity vs Redshift

To investigate the validity "completeness" of the ATLAS DR3 sources that have been cross-identified by LRPY with Fusion, in this section I will look at there Radio Luminosity against redshift. This work has produced extra matches than from the earlier ATLAS DR1 and the work of Mao et al. (2012), and with the OzDES matches provides additional redshift vales. The extra XID's obtained is 986 from CDFS and 929 from ELAIS, this is not the same as the total number of XID's from both fields as only it is possible to only take those matches where there exists a complete set of Spitzer IR fluxes at $S_{3.6\mu m}$; $S_{4.5\mu m}$; $S_{5.8\mu m}$ and $S_{8.0\mu m}$.

The luminosity data used in this section was generated using the formula from Hogg (1999) implemented by Benjamin Weiner (Weiner, 2018) who produced a FORTRAN program called "DISTCALC", see Equation 6.3. The terms in this equation are $L_{1.4}$ the luminosity value at 1.4 GHz, $S_{1.4}$ the radio flux value at 1.4 GHz, D_L the luminosity distance in Mpc and α_{radio} the spectral index for 1.4 GHz which was taken as -0.75. This code was re-written in Perl [2] by myself and in addition communicates with the MySQL database used for this work (Weston, 2012) (use was made of the Astro::Cosmology Perl Library). For this work a Hubble constant of 70 km s^{-1} Mpc^{-1} is used, and matter and cosmological constant density parameters of $\Omega_M = 0.27$ and $\Omega_\Lambda = 0.73$, and assume the convention for spectral index, α , where $S \propto \nu^{\alpha}$.

$$L_{1.4} = \frac{4\pi S_{1.4} D_L^2}{(1+z)^{(1+\alpha_{\rm radio})}} \tag{6.9}$$

In Figure 6.15 the radio luminosity ($L_{1.4}$) of these sources is plotted against redshift. In Figure 6.16 is the same plot but for the range $0 < z < 1.0$ as ATLAS was configure for a depth of a limit of $z \leq 1$ (Aim for ATLAS (Norris et al) , to trace radio luminosity function to a high ($z \simeq 1$) redshift). For reference there is overlaid is the limit of 0.15mJy beam^{-1} with a solid line, this figure was used by Mao et al. (2012) and came from Mauch

[2] Perl Language and Community : https://www.perl.org/

and Sadler (2007) which is included here for comparison with this earlier work (for latter figures in this section I mark the σ and 5σ ATLAS DR3 survey limits).

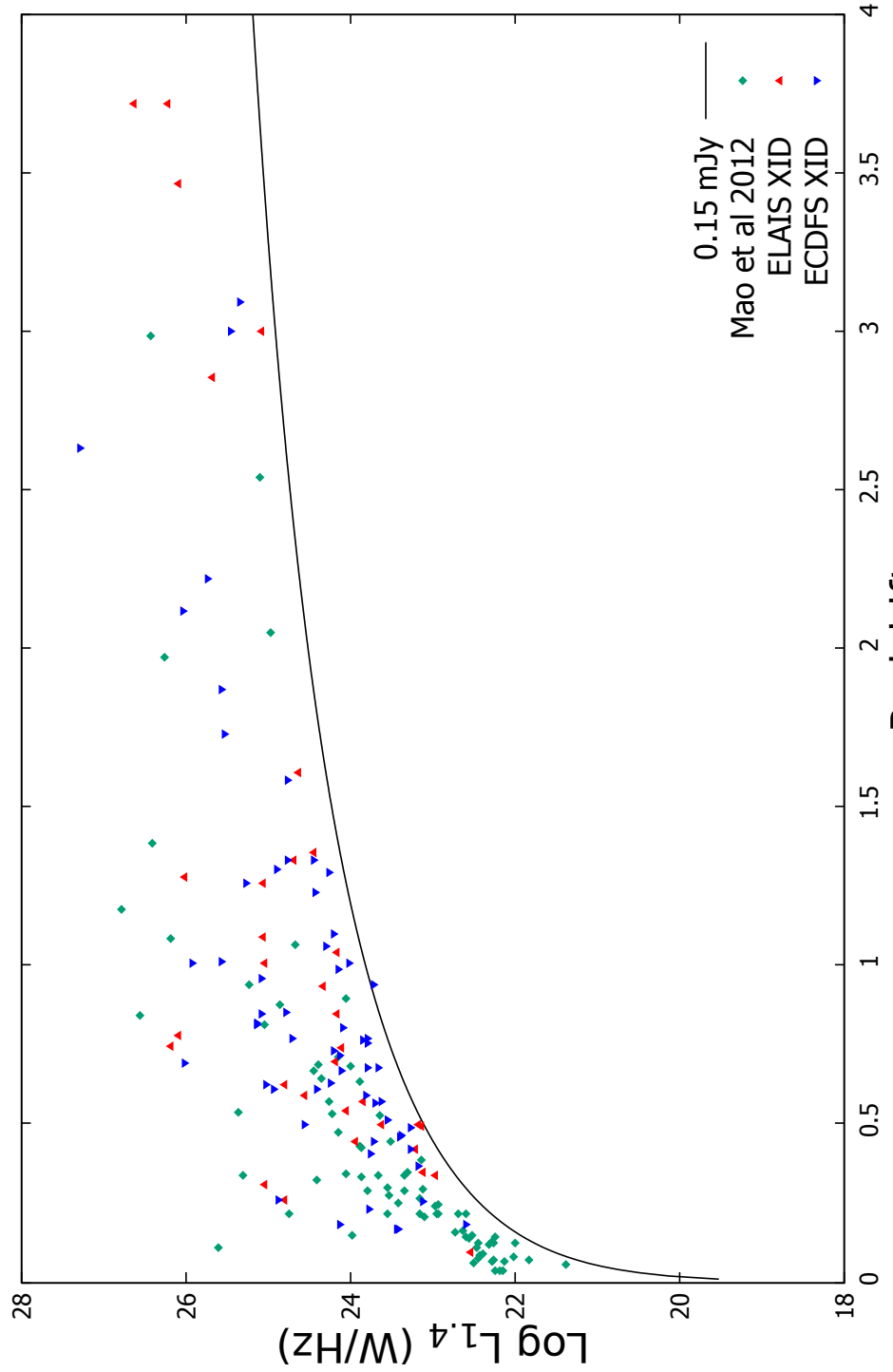


FIGURE 6.15: Radio Luminosity vs Redshift: I have plotted the data from Mao et al. (2012) then overlaid the XID's from this work. The solid line corresponds to a flux limit of 0.15mJy beam^{-1}.

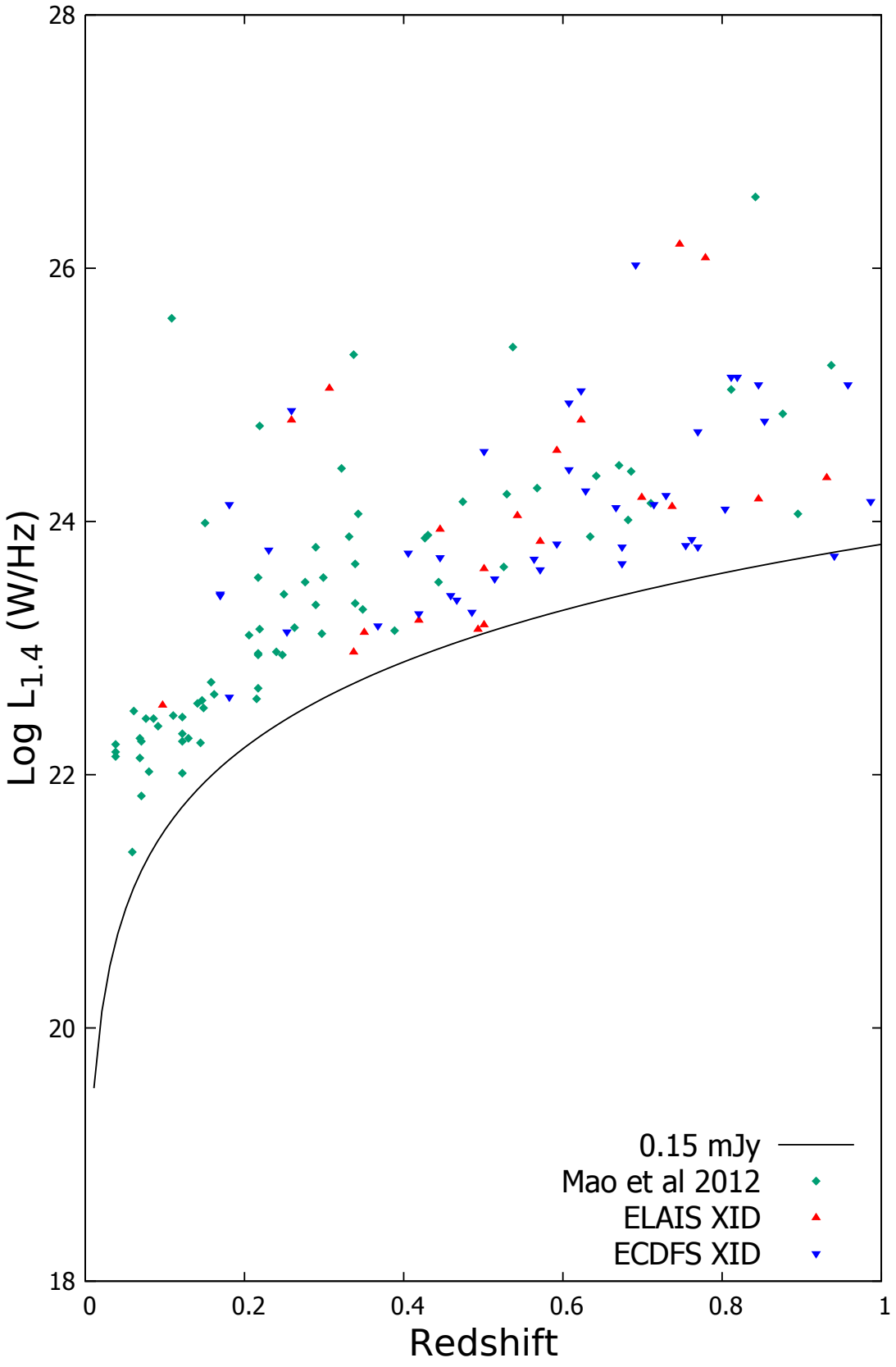


FIGURE 6.16: Radio Luminosity vs Redshift: I have plotted the data from Mao et al. (2012) then overlaid the XID's from this work. In this figure the Redshift has been restricted to $0.0 < z < 1.0$ which was the target limit for the ATLAS survey. The solid line corresponds to a flux limit of $0.15\text{mJy}\,\text{beam}^{-1}$.

The next step was to take the OzDES survey data and undertake a nearest neighbour 1" match with the XID's from this work of the ATLAS DR3 radio sources. In Figure 6.17 are plotted the nearest neighbour OzDES sources (with a redshift value) to the Fusion sources cross identified, this provided more redshifts than where available from Fusion. Also highlighted are the ATLAS DR3 survey limits from Franzen et al. (2015); CDFS $\sigma = 14.9\mu$Jy, $5\sigma = 74.9\mu$Jy, ELAIS $\sigma = 17\mu$Jy, $5\sigma = 85\mu$Jy, are marked with different various types of dashed lines for these limits.

There are a total of 1404 redshifts, of which 136 (9.6%) have a redshift greater than 1. So in the top plot of Figure 6.17 we show the range from $0 < z < 4$, and the bottom plot we have zoomed into the regions between $0 < z < 1$ where 90.4% of the sources with OzDES redshifts reside consistent with the ATLAS survey targeted space.

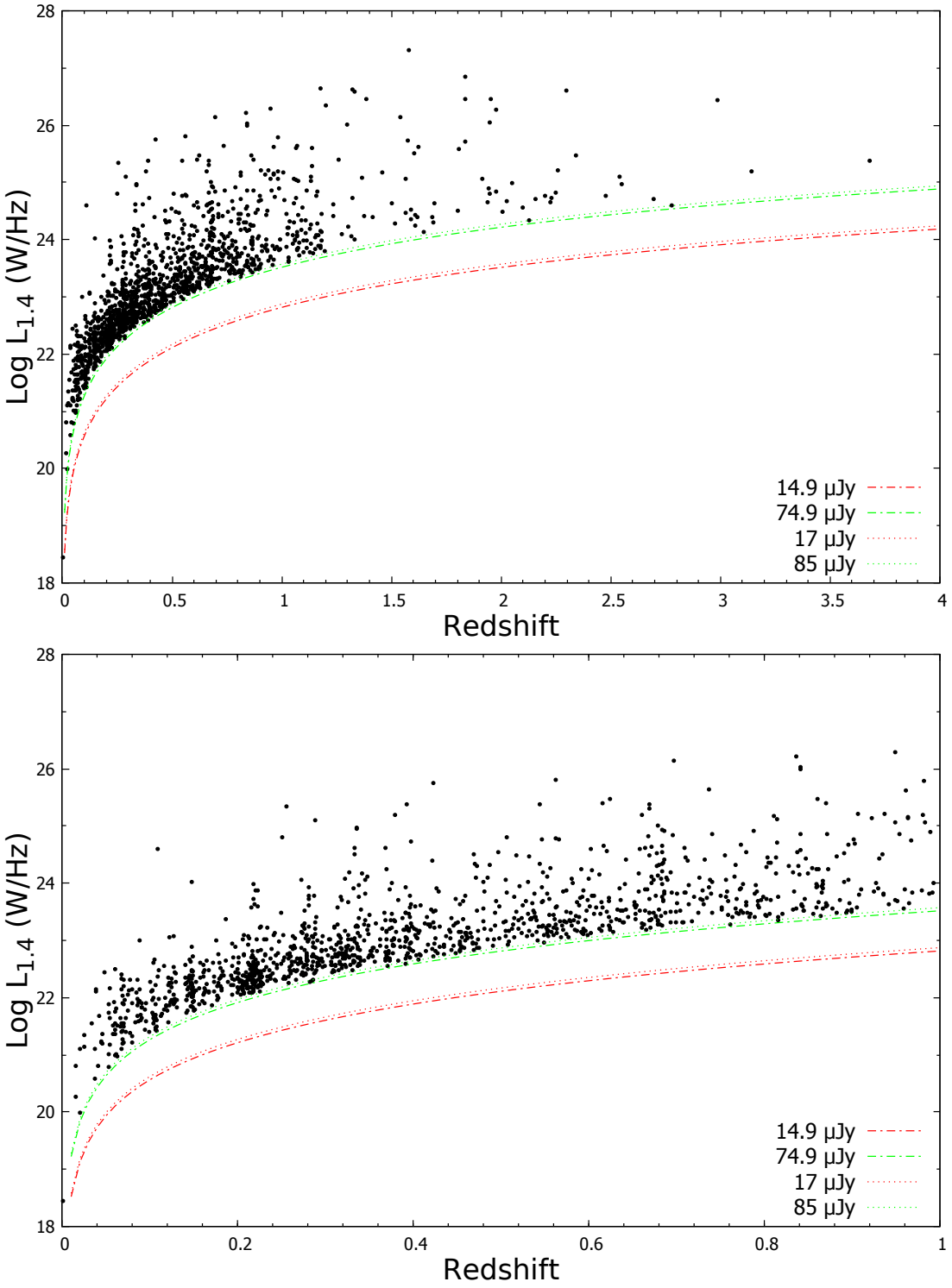


FIGURE 6.17: The top plot is for all OzDES to Fusion (LRPY XID) sources, $0 < z < 4$. The bottom plot is for the range $0 < z < 1$. In both the detection limits taken from (Franzen et al., 2015) for ATLAS DR3 are shown; CDFS $\sigma = 14.9\mu$Jy red dash-dot-dash , $5\sigma = 74.9\mu$Jy green dash-dot-dash, ELAIS $\sigma = 17\mu$Jy red dot-dot-dot, $5\sigma = 85\mu$Jy green dot-dot-dot.

6.4 Comparing Radio and Mid-Infrared Luminosities

Condon et al. (1991) demonstrated a correlation between radio and Far InfraRed (FIR) luminosities. In figure 6.18 we compare the Radio luminosity to the lower two bands of the MIR 3.6 μm (the top plot) and 4.5 μm (the bottom plot). The plots are composed in the following manner: the abscissa is the log of the MIR luminosity (in units of solar luminosity $L_{\rm sun} \equiv 3.828 \times 10^{26}$ W and the ordinate is the log of the radio spectral luminosity at $S_{1.4\rm GHz}$ (W Hz^{-1}). By using the solar luminosity an indication of mass is obtained. In addition I have split the sources into two groups by separating them about $z = 0.3$, the blues dots (●) are for those galaxies that have $z < 0.3$ and the red dots (●) are for the galaxies with $z > 0.3$. It can see a that a similar correlation to Condon et al. (1991) exists, but at the higher redshifts ($z > 0.3$) there is a greater scatter in the there luminosity ratios.

For the luminosity ratio calculations I have corrected the data as we are using only part of the spectrum and $L_\odot$ is for the complete spectrum of emission from the sun. Thus a correction for this by a term ν_{IR} is made due to the part of the spectrum we are observing, which is derived thus :

$$\frac{L_{IR}}{L_\odot}\Big(\nu\Big) = \frac{L_{IR} \times \nu_{IR}}{L_\odot} \tag{6.10}$$

Taking the spectral band information for Spitzer from IRSA (2019) I have selected the corrections as $\nu_{3.6\,\mu\rm m} = 8.327 \times 10^{13}$ and $\nu_{4.5\,\mu\rm m} = 6.66 \times 10^{13}$.

Figure 6.19 contains the same plots as Figure 6.18, but here the galxies are differentiated by the Stern (top plot) and Lacy (bottom plot) selection criteria (these are indicated with red dots ●) in the plots. It can be seen that most of the Stern/Lacy AGNs correlate with the previous Figure 6.18 where I highlighted those galaxies with $z > 0.3$. The top plot in this figure is for $3.6/,\mu$m and the bottom plot is for 4.5 μm. For both plots there is only one AGN outlier (●) in the bottom left hand corner.

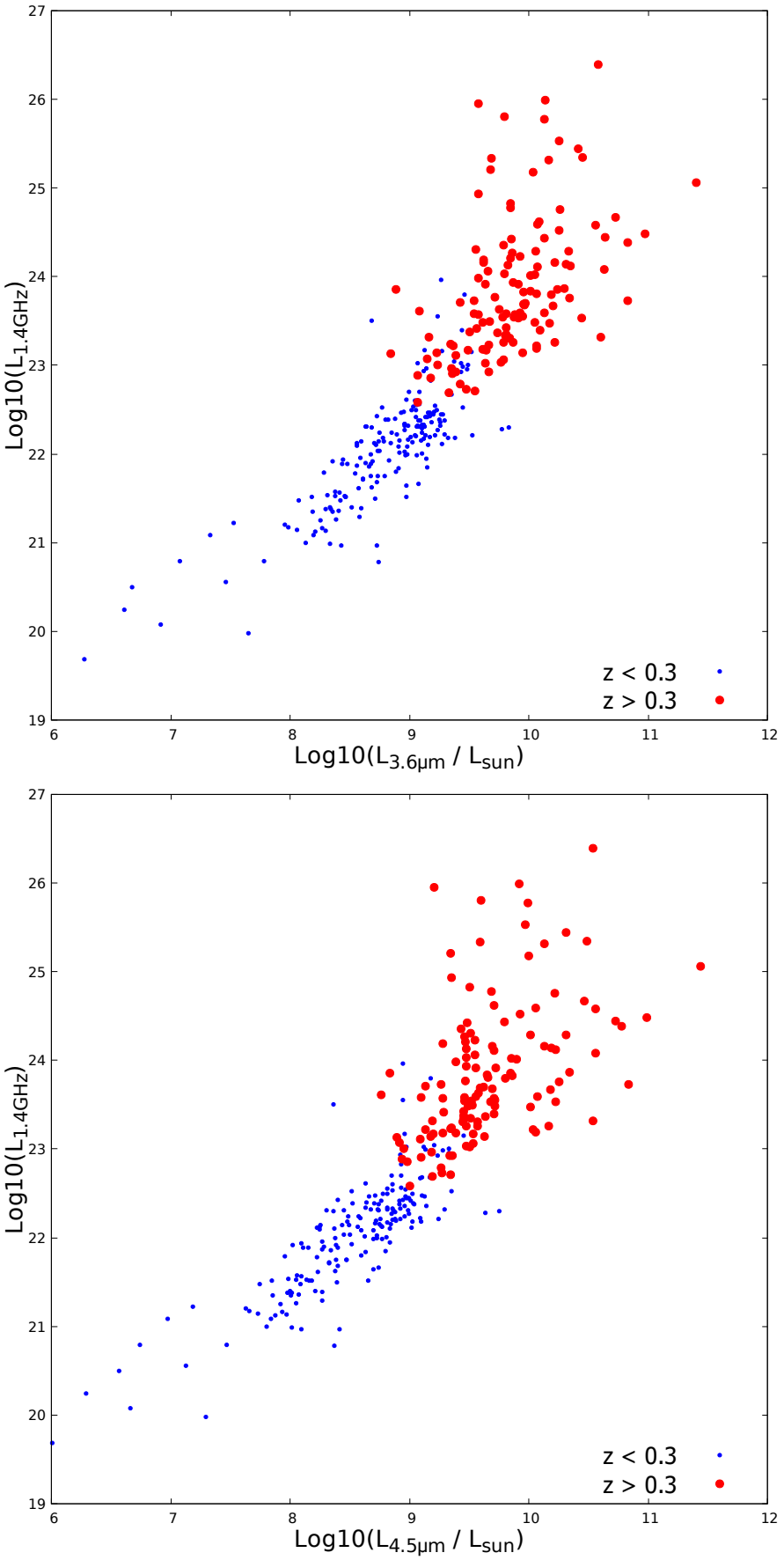


FIGURE 6.18: The correlation of radio to MIR luminosities for the galaxies cross identified in this work. The top plot is for 3.6 μm MIR to Radio $S_{1.4\mathrm{GHz}}$ and the bottom plot is for 4.5 μm to Radio $S_{1.4\mathrm{GHz}}$. The galaxies with $z > 0.3$ are marked with larger red dots and those galaxies with a $z < 0.3$ are marked with blue dots.

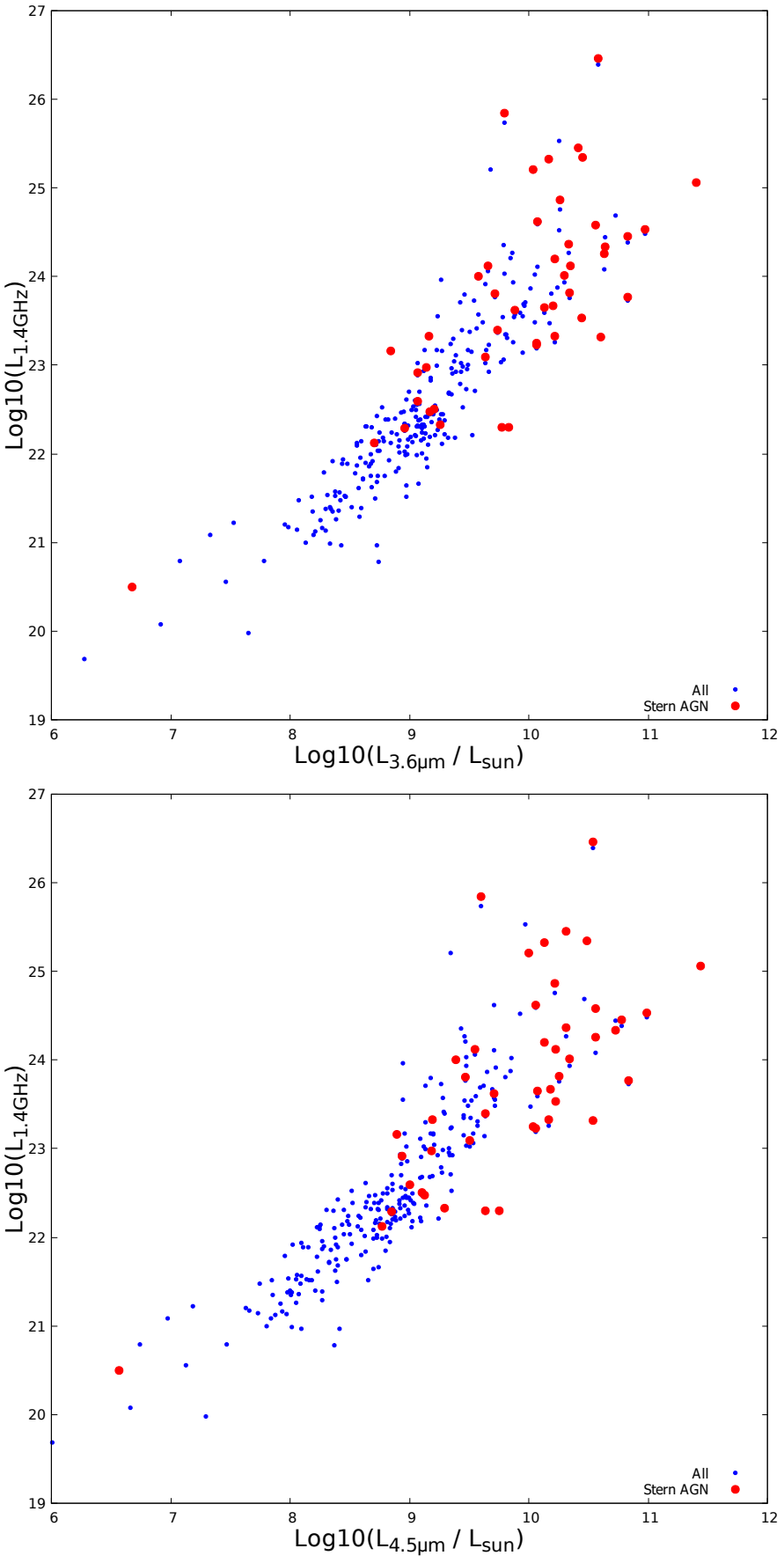


FIGURE 6.19: The correlation of radio to MIR luminosities for the galaxies cross identified in this work. Also the galaxies lying within the Stern MIR colour-colour wedge (see Section 6.1.1) have been identified by the larger red dots . The top plot is for 3.6 μm MIR and the bottom plot is for 4.5 μm.

Is was observed by Condon et al. (1991) for a set of optically selected spiral and irregular galaxies that the FIR to Radio luminosities had a correlation, this had a relationship where the slope $m = 1.02 \pm 0.02$ which was close to unity but for infrared-selected galaxies a slope $m = 1.11 \pm 0.02$ existed. For this dataset of MIR to Radio luminosities a correlation is also seen and I will investigate this in a bit more detail.

In the top plot of Figure 6.20 looking at the Lacy AGN for $3.6\,\mu$m , I have fitted two straight lines to the data. The blue dashed line is a fit to all the points we have $m = 1.32 \pm 0.043$ and $c = 10.493 \pm 0.39$, the red dotted line is a fit only to the Lacy AGN points we see $m = 1.004 \pm 0.155$ and $c = 13.794 \pm 0.1.55$.

In the bottom plot Figure 6.20 looking at the Lacy AGN for $4.5\,\mu$m, I have also fitted two straight lines to this data. The blue dashed line which is a fit to all the points we have $m = 1.265 \pm 0.043$ and $c = 11.323 \pm 0.376$, for the red dotted line which is a fit only to the Lacy AGN points we see $m = 0.980 \pm 0.153$ and $c = 14.092 \pm 0.1.522$.

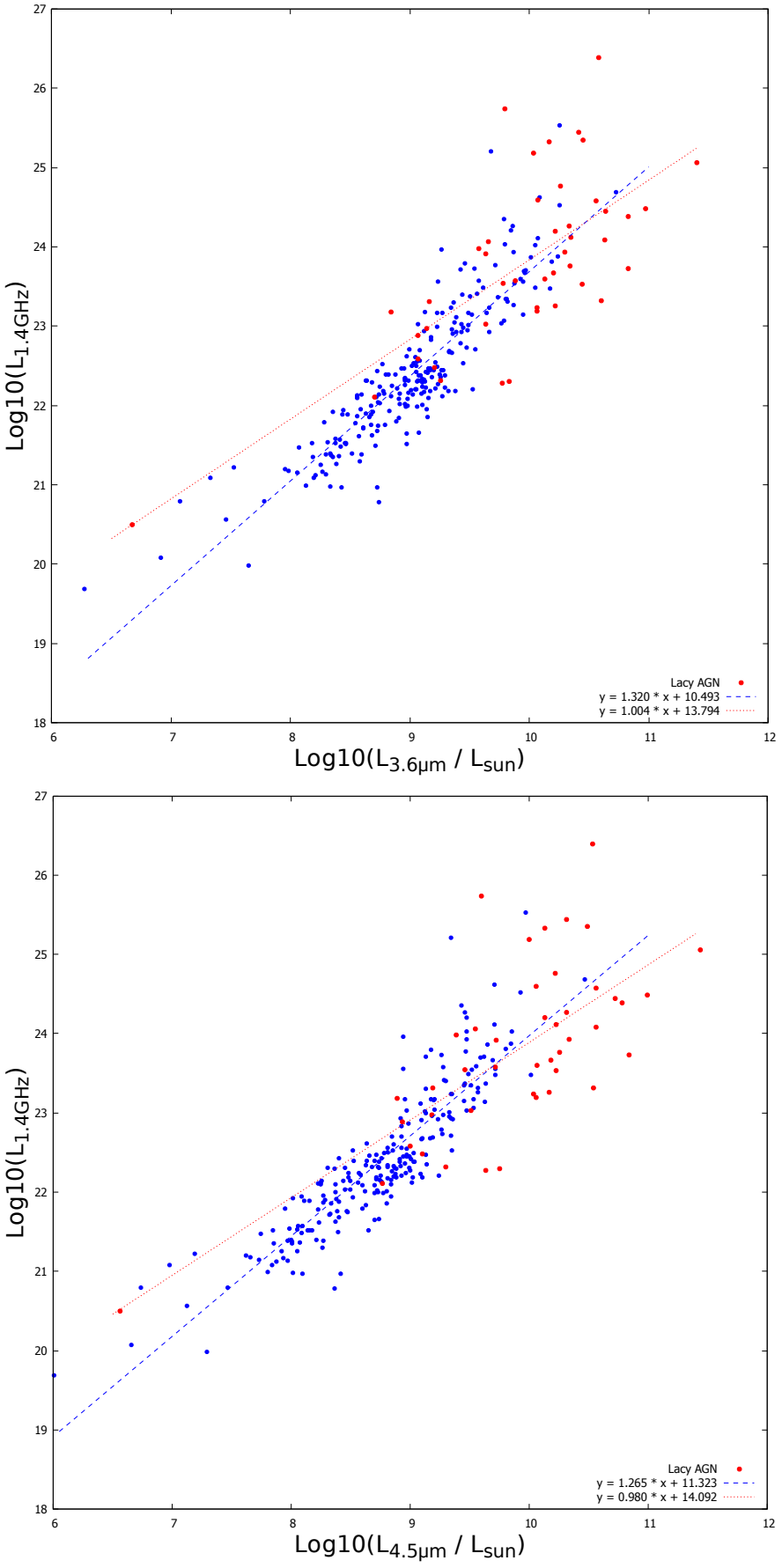


FIGURE 6.20: The correlation of radio to MIR luminosities for the galaxies cross identified in this work. Also the galaxies lying within the Lacy MIR colour-colour wedge (see Section 6.1.1) have been identified by red dots. The top plot is for 3.6 μm MIR and the bottom plot is for 4.5 μm.

6.5 Summary

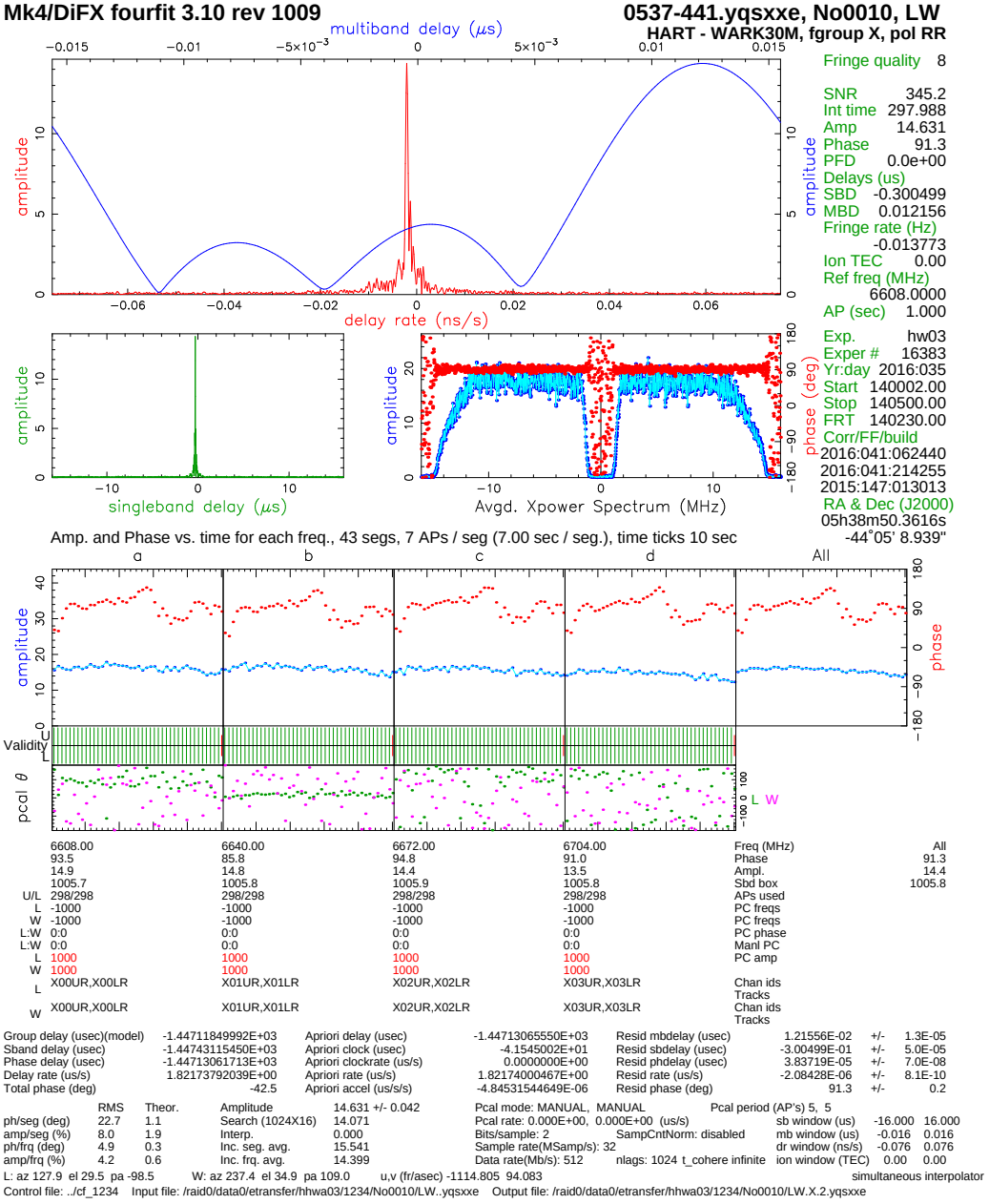
At the beginning of this Chapter in Section 6.1, I used the MIR flux values from the cross matched sources between the ATLAS DR3 radio and Fusion IR catalogues to investigate which of the sources are AGN or SFGs. For the Richards AGN selection criteria it can be seen that the evolutionary tracks of M82 and NGC4429 cross over into the AGN selection area between $0 < z < 1$ and again latter in their eariler life $1 < z < 2$, so for this datase the Richards selection criteria is not as good as Stern or Lacy. What was interesting for this dataset when plotting the Radio $S_{1.4GHz}$ versus the $S_{8.0\mu\,\mathrm{m}}$ values in Section 6.1.4 there is quite clearly two branches of objects. When looking at these objects when classified by the Stern and Lacy selection criteria, it is quite clear that the AGN candidates predominantly lie within the left hand branch, the Lacy AGN candidates being the more numerous. Thus providing another selection method in addition to Stern and Lacy which can be further used to filter SFGs from AGNs.

During an initial review of the dataset it was noticed that there appeared to be a strong correlation to a power law relationship. In Section 6.2 I have investigated this further with some reference to prior work. This strong correlation is first demonstrated for the complete Fusion dataset seen in the AGN arm of the $[3.6]-[5.8]$ vs $[4.5]-[8.0]$ plot for the LACY AGN selection wedge in the top plot of Figure 6.13. The correlation is not very strong in the Stern AGN wedge of the $[5.8]-[8.0]$ vs $[3.6]-[4.5]$ plot. This is also demonstrated in the plot of the Fusion counterparts to the ATLAS sources from this work in Figure 6.14, as can be seen this correlation is far stronger for the Lacy AGN candidates than the Stern AGN candidates.

In section 6.3 I looked at the completness of the XIDs as I had additional matches than from earlier work with ATLAS DR1 and also additional redshifts from the OzDES data. No discrepencies are seen such as candidates below the survey limit.

Finally in section 6.4 I have looked at the relationship between the Radio and IR luminosities, as prior work had provided existance of a correlation. What was found is that the higher redshift sources ($z > 0.3$) lie within the top right quadrant of the plots, when

the Stern and Lacy AGN are identified we see the majority of these AGN also lie within this same quadrant. Looking at the AGN selected sources we see a slope close to unity at both MIR wavelengths, for $3.6\,\mu$m the slope is $m = 1.004 \pm 0.155$ and at $4.5\,\mu$m it was found to be $m = 0.980 \pm 0.153$.



Another major milestone which I was involved in was the fist VLBI with Hartebeestock, SA in C-Band using the then newly converted 30m antenna. The schedule was produced by myself and the data was correlated here in New Zealand using DiFX (Deller et al., 2011) by myself 2015. Here is the plot from *fourfit* for scan 10 on 0537-441 showing the fringe obtained. Credit: Stuart Weston

Chapter 7

Summary, Conclusions and Further Work

In Chapter 1 I explained how the issue of matching radio sources to objects in visible wavelengths due to the differences in resolution was a hurdle for the early radio astronomers, and how through different observational techniques such as interferometry this was overcome. Then with the advent of larger surveys from instruments such as ATCA, VLA and space based instruments such as *Spitzer* and the resultant larger catalogues the number of objects started to became an issue. With the next generation of radio telescopes such as ASKAP and MeerKAT starting surveys as I write this concluding chapter, the era of very large multi-wavelength surveys with millions of sources has arrived and automated machine techniques and algorithms will have to be used. So to conclude this thesis in my final chapter I will start by summarising the work I have done and the refinements made to the Likelihood Ratio and how they might be applied to expand the usefulness of the algorithm in identifying other objects of interest, not just cross matching. I will also review and summarise how my application of the LRPY algorithm to the ATLAS and Fusion catalogues allowed me to investigate the cross matched sources with multi-wavelength (Radio and MIR) information. Finally I discuss how this method might be used for these large surveys that have started to be generated by the next generation of survey instruments, also the future SKA instruments to come.

7.1 Summary

Chapter 2 I started by looking at the surveys used for this thesis, beginning with the Radio Survey catalogue ATLAS DR3 followed by the Fusion IR catalogue which is used for the cross identification in Section 2.1. The OzDES survey is briefly discussed as this provided additional redshift data which was not available for ATLAS DR1, and the FIRST survey and catalogue which is used later to test some of the refinements to the LR Algorithm.

I followed this in Section 2.2 comparing some of the different methods employed to date for cross-identification between catalogues such as a simple Nearest Neighbour followed by the Poisson Probability and finally the Likelihood Ratio. It concludes with a brief overview for some of the more recent techniques being explored by others.

In Chapter 3 I discuss in detail the implementation of the LR method for this thesis in Section 3.1. A different approach to previous work by others for the selection of matches has been presented in Section 3.1.6, which uses a combination of the Reliability and LR values. I noticed during this research that there are some single radio sources that have multiple IR components with similar Reliability values (centered about $0.3 < R < 0.7$). This result had been noticed and reported in prior work that maybe there was perhaps some local clustering of IR sources merging to form one apparent source of radio emission, this is explored in a latter chapter and confirms that this is in fact the case for some of these identified cases.

The LR method works on a principle of matching one Radio source to one IR candidate, but we have some situations where the Radio source has a more complex morphology such as bright side lobes and multiple components. The problem then is which IR source (Galaxy) is the true one. So I investigated a possible adaptation of my LR algorithm to accommodate Radio sources with two components being matched to one IR candidate.

Also presented in Sections 3.2 and 3.3 is how the algorithm was implemented using a Relational Database and Python code.

Finally in Section 3.4 is presented the resultant catalogue of matched sources between ATLAS DR3 and Fusion with a detailed description, as published in (Weston et al., 2018). No selection has been made, all matches are provided even if they do not match the selection criteria used for my later work and analysis in this thesis, this allows others to use the catalogue and define their own selection criteria or other extensions and modifications in the post processing, depending on their scientific requirements.

Then in Chapter 4 I conducted a comparison of the two techniques, nearest neighbour and Poisson probability against the likelihood ratio. I found taking a nearest neighbour approach for those with ATLAS sources with n = 1 Fusion candidates in Table 4.3 we see only 78% pass the likelihood ratio reliability selection criteria in this work. Another failing of the nearest neighbour method is that it can not distinguish between multiple possible candidates within the search radius, it just selects the candidate closest which might not be the ”true” candidate.

I then looked at the Poisson probability method and highlight that in its current form it fails to distinguish between multiple possible candidates within the search radius, a highly probable scenario given the depth and density of the two catalogues being cross matched. In comparison by using the reliability in the likelihood ratio method I am able to assign a value to each of the possible candidates and make a quantitative selection based on these values. Also the P_{not} values do not provide any extra information which in the likelihood ratio Reliability I have used in Chapter 5 for the identification of more complex radio source morphology or a radio source resulting from a merging of the radio emission due to the resolution from multiple candidates.

It was shown earlier in the thesis that there exists a symmetry of data point ”pairs” around $R \approx 0.5 \pm 0.3$, so in Chapter 5 Section 5.1 I investigated these possible pairs in more detail. From matching with OzDES to obtain their redshift z, we found the majority (32 from 34 pairs, each with a spec-z) of these IRD's have a very similar redshift, taking the average of the Δz for these pairs we have $\overline{\Delta z} = 0.00245$. I was able, for two of the IRD candidates with low values of Δz, to obtain HST images and corresponding FIRST radio images, these are quite clearly interacting pairs of galaxies with disturbed morphology and tidal tails. This highlights a small sub-group (34) within

this dataset that could have been missed and in the much larger surveys to come would result in a significant number of potentially interesting objects being missed.

In Chapter 5 Section 5.2 I also investigated to see if the algorithm could be extended to identify radio doubles. I found that it is able to select a group of sources of which over half are probably radio doubles, so this could be a useful method to filter from a larger catalogue a subset to review in more detail perhaps visually by the investigator. Currently the algorithm is unable to identify more complex radio morphology such as triples or even more complex sources by identifying 2 or more possible doubles. Such filters will be important for the larger surveys to come as the number of objects (tens of millions) in the catalogues and the survey areas (tens of thousand square degrees) covered will make it impractical for investigators to do this manually.

Finally in Chapter 6 Section 6.1, I used the MIR flux values from the cross matched sources between the ATLAS DR3 radio and Fusion IR catalogues to investigate which of the sources are AGN or SFGs. For the Richards AGN selection criteria it can be seen that the evolutionary tracks of M82 and NGC4429 cross over into the AGN selection area between $0 < z < 1$ and again latter in their earlier life $1 < z < 2$, so for this data set the Richards selection criteria is not as good as Stern or Lacy. What was interesting for this dataset when plotting the Radio $S_{1.4GHz}$ versus the $S_{8.0\mu\,\mathrm{m}}$ values in Section 6.1.3 there is quite clearly two branches of objects. When looking at these objects when classified by the Stern and Lacy selection criteria, it is quite clear that the AGN candidates predominantly lie within the left hand branch, the Lacy AGN candidates being the more numerous. Thus providing another selection method in addition to Stern and Lacy which can be further used to filter SFGs from AGNs.

During an initial review of the dataset it was noticed that there appeared to be a strong correlation to a power law relationship, and in Section 6.2 I have investigated this further. This strong correlation is first demonstrated for the complete Fusion dataset seen in the AGN arm of the $[3.6]-[5.8]$ vs $[4.5]-[8.0]$ plot for the Lacy AGN selection wedge. The correlation was found to be not as obvious in the Stern AGN wedge of the $[5.8]-[8.0]$ vs $[3.6]-[4.5]$ plot. The strong correlation for the Lacy AGN candidates

is also apparent in the plot of the Fusion counterparts to the ATLAS sources from this thesis, and for the Stern AGN candidates the correlation is week.

In Section 6.3 I looked at the completeness of the XIDs as I had additional matches than from earlier work with ATLAS DR1 and also additional redshifts from the OzDES data. No discrepancies are seen such as candidates below the survey limit.

I conclude Chapter 6 in Section 6.4, by looking at the relationship between the Radio and IR luminosities. What was found is that the higher redshift sources ($z > 0.3$) lie within the top right quadrant of the plots, and when the Stern and Lacy AGN are identified we see the majority of these AGN also lie within this same quadrant. Looking at the AGN selected sources we see a slope close to unity at both MIR wavelengths, for 3.6 μm the slope is $m = 1.004 \pm 0.155$ and at 4.5 μm it was found to be $m = 0.980 \pm 0.153$.

7.2 Conclusions

We have used LRPY to cross-match the ATLAS DR3 radio survey to the Fusion catalogues with a search radius of $6''$. Setting the possible false detection rate of 5 %, and using the new resultant likelihood ratio and Reliability cut-off selection criteria described in Section 4.1, rather than simply a reliability threshold, we obtain 2222 (82%) matches in the CDFS field and 1626 (83%) in the ELAIS-S1 field. Of these matches, we obtain a subset with detections in all four IRAC bands consisting of 2133 sources (1243 for CDFS and 890 for ELAIS-S1). A much smaller subset has redshifts consisting of 295 sources (186 CDFS and 109 for ELAIS-S1). Hence, from this work we present a new catalogue listing ATLAS DR3 radio sources with their *Spitzer* Data Fusion counterparts including the likelihood and reliability figure to allow the reader to use their own selection criteria as required.

We have identified a subset of 64 Fusion doubles (38 in CDFS and 26 in ELAIS-S1), i.e. radio sources with two *Spitzer* Data Fusion candidates meeting our selection criteria in Section 5.1. From these pairs we find 22 with a redshift for each member; we find 20 of these have a $\Delta z/z < 0.01$ and we identify them as potentially interacting galaxies

contributing to the one radio source. Two pairs are confirmed as interacting galaxies from deep *HST* imaging.

Taking the available Fusion colour-colour information for the possible matches we present their characteristics with respect to the Stern and Lacy AGN selection criteria. For the two fields we identify 848 AGN radio sources using the Lacy selection criteria which is $\approx 42\,\%$ of the candidates, and 533 if using the Stern criteria which is $\approx 27\,\%$. Also, we examine the radio to $3.6\,\mu$m flux density ratio as a function of redshift to search for radio-loud AGN. We find a cluster of objects at $z < 0.3$ and flux ratio 0.2 to 2 which we surmise are SFG, but with a lower relative SFR to stellar mass than the modelled M82 track. We propose a cut off where the flux ratio is greater than one third of the value of the RL-AGN track. Taking the union of all three AGN selection criteria we identify $956 \approx 48\,\%$ possible AGN.

In Section 6.2 I investigated the power law relationship for the MIR colour-colour plots. It was found that there exists a strong correlation for the complete Fusion dataset seen in the AGN arm of the $[3.6]-[5.8]$ vs $[4.5]-[8.0]$ plot for the LACY AGN selection wedge. It was also found that this correlation is not as prominent in the Stern AGN wedge of the $[5.8]-[8.0]$ vs $[3.6]-[4.5]$ plot. This correlation is also demonstrated for the Fusion counterparts to the ATLAS sources from this thesis and again it is seen that this correlation is far stronger for the Lacy AGN candidates than the Stern AGN candidates.

In section 6.3 I looked at the completeness of the XIDs as I had additional matches than from earlier work with ATLAS DR1 and also additional redshifts from the OzDES data. No discrepancies are seen such as candidates below the survey limit.

Finally in section 6.4 I investigated the relationship between the Radio and IR luminosities, as prior work had provided existence of a correlation. What was found is that the higher redshift sources ($z > 0.3$) lie within the top right quadrant of the plots, when the Stern and Lacy AGN are identified we see the majority of these AGN also lie within this same quadrant. Looking at the AGN selected sources we see a slope close to unity at both MIR wavelengths, for $3.6\,\mu$m the slope is $m = 1.004 \pm 0.155$ and at $4.5\,\mu$m it was found to be $m = 0.980 \pm 0.153$.

I have developed an algorithm in Python using the Likelihood Ratio to allow for the cross-matching of a lower resolution radio survey to a higher source density and higher resolution catalogue. The algorithm has been extended to identify potential multiple matches to a lower resolution source which I called Infrared Doubles, also I have tried to adapt the algorithm to also try and identify complex sources in the lower resolution survey. From the resultant cross matches I have then undertaken some analysis of their properties using the Radio, Infrared and Redshift information now available. I have made this code publicly available to the community and I hope it will be useful.

As already mentioned in the introduction radio surveys are going deeper (deep in this context means $\approx 10\,\mu$Jy or better), also with the next generation of instruments and surveys (see Figure 1.8) these surveys will cover much larger areas. For example ATLAS covered 7 square degrees at 1.4 GHz to a sensitivity of $\approx 15\,\mu$Jy, EMU with ASKASP will survey tens of thousands of square degrees and try to achieve a sensitivity of $\approx 10\,\mu$Jy or better. This increase in sensitivity has resulted in surveys going from being RL AGN dominated (which are very radio bright even at high redshift) to now finding star forming galaxies (SFG) out to $z \approx 1$. The SKA1 for 1.4 GHz will see another order of magnitude increase in sensitivity to an expected $\approx 2\,\mu$Jy/beam, this will push well into the realm of RQ AGNs and probe the entire AGN population, also powerful star bursts to even higher redshifts and AGNs to the edge of the visible universe $z > 7$ (Prandoni and Seymour, 2015). This is opening up a new area of parameter space looking at the formation and evolution of early galaxies.

7.3 Further Work

The extension to the LRPY algorithm in Section 5.2.1 for identifying radio doubles requires more development. It is currently able to select a subset of which over half are probably radio doubles, so I currently have a useful method to filter from a larger catalogue a more manageable sub-set to review in more detail perhaps visually by the investigator. The algorithm is currently unable to identify more complex radio morphology such as triples or even more complex sources. But with further refinement, such

as also searching for a common source to two or more doubles, the algorithm could be tuned to also select possible triples. Some such algorithm will be important for the larger surveys to come as the number of objects (10's of millions) in the catalogues and the survey areas (10's of thousand square degrees) covered will make it impractical for investigators to do manually. Already for The Galactic and Extra-Galactic all-Sky MWA Survey (GLEAM) White et al. (2019) (paper in preparation) we have already applied the Radio Doubles technique from LRPY, where we found 47 potential 'doubles' which upon visual inspection 18 were confirmed as true 'doubles' and one as a 'triple' for those GLEAM components brighter than 4Jy. For those sources fainter than 4Jy the method was not as reliable, indicating that further refinement of this method is required.

The LRPY algorithm needs to be applied to other catalogues, one such possibility is the ASKAP commissioning observations of the GAMA 23 field (Leahy et al., 2019) as there are other wavelength catalogues for the same field available. There is also the upcoming ASKAP/EMU survey with an estimated 7×10^7 faint radio sources to be cross matched and this method could be one of several complimentary methods to try.

We are now entering an interesting period of very large multi-wavelength surveys generating correspondingly large datasets, this will enable some interesting science to be investigated and serendipitous discoveries into galaxy properties and evolution.

Sunset at the Compact Array, the instrument that the ATLAS survey was conducted with. Credit: Stuart Weston

Appendix A

List of Publications

The following list constitutes the peer reviewed papers and work that I have published or contributed to during the course of my PhD and employment at the Institute of Radio Astronomy and Space Research, AUT University:

S.D. Weston, T. Natusch, S. Gulyaev, Radio Astronomy and eVLBI using KAREN, Proceedings of Electronics New Zealand Conference 2010, pp. 141 - 145, Hamilton, New Zealand, November 2010.

Sergei Gulyaev, Tim Natusch, **Stuart Weston**, Neville Palmer, David Collett, Warkworth 12-m VLBI Station: WARK12M, in International VLBI Service for Geodesy and Astrometry 2010 Annual Report, edited by D. Behrend and K.Baver, NASA/TP-2011-215880, 2011.

Stuart Weston, Timothy Natusch, Sergei Gulyaev, Radio Astronomy Data Transfer and eVLBI using KAREN, The XXX General Assembly and Scientific Symposium of the International Union of Radio Science, 2011

Stuart Weston, Hiroshi Takiguchi, Tim Natusch, Sergei Gulyaev, Warkworth 12-m VLBI Station: WARK12M, in International VLBI Service for Geodesy and Astrometry 2011 Annual Report, edited by K. D. Baver and D. Behrend, NASA/TP-2012-217505, 2012.

Stuart Weston, Hiroshi Takiguchi, Tim Natusch, Lewis Woodburn, Sergei Gulyaev, Warkworth 12-m VLBI Station: WARK12M, in International VLBI Service for Geodesy and Astrometry 2012 Annual Report, edited by K. D. Baver, D. Behrend, and K. Armstrong, NASA/TP-2013-217511, 2013.

Hiroshi Takiguchi, **Stuart Weston**, Tim Natusch, Sergei Gulyaev, Warkworth Radio Astronomical Observatory New Zealand, in International VLBI Service for Geodesy and Astrometry 2013 Annual Report, edited by K. D. Baver, D. Behrend, and K. L. Armstrong, NASA/TP-2014-217522, 2014.

Woodburn, Lewis; Natusch, Tim; **Weston, Stuart**; Thomasson, Peter; Godwin, Mark; Granet, Christophe; Gulyaev, Sergei (2015). Conversion of a New Zealand 30-Metre Telecommunications Antenna into a Radio Telescope. Publications of the Astronomical Society of Australia, Volume 32, id.e017 14 pp.

Petrov, Leonid; Natusch, Tim; **Weston, Stuart**; McCallum, Jamie; Ellingsen, Simon; Gulyaev, Sergei (2015). First scientific VLBI observations using New Zealand 30 metre radio telescope WARK30M. Publications of the Astronomical Society of the Pacific, Volume 127, No. 952, pp. 516-522.

M. Kadler; F. Krauß; K. Mannheim; R. Ojha; C. Müller; R. Schulz; G. Anton; W. Baumgartner; T. Beuchert; S. Buson; B. Carpenter; T. Eberl; P. G. Edwards; D. Eisenacher Glawion; D. Elsässer; N. Gehrels; C. Gräfe; S. Gulyaev; H. Hase; S. Horiuchi; C. W. James; A. Kappes; A. Kappes; U. Katz; A. Kreikenbohm; M. Kreter; I. Kreykenbohm; M. Langejahn; K. Leiter; E. Litzinger; F. Longo; J. E. J. Lovell; J. McEnery; T. Natusch; C. Phillips; C. Plötz; J. Quick; E. Ros; F. W. Stecker; T. Steinbring; J. Stevens; D. J. Thompson; J. Trüstedt; A. K. Tzioumis; **S. Weston**; J. Wilms; J. A. Zensus;
Coincidence of a high-fluence blazar outburst with a PeV-energy neutrino event
Nature Physics (2016).

Krauß, F.; Wilms, J.; Kadler, M.; Ojha, R.; Schulz, R.; Trüstedt, J.; Edwards, P. G.; Stevens, J.; Ros, E.; Baumgartner, W.; Beuchert, T.; Blanchard, J.; Buson, S.; Carpenter, B.; Dauser, T.; Falkner, S.; Gehrels, N.; Gräfe, C.; Gulyaev, S.; Hase, H.; Horiuchi, S.; Kreikenbohm, A.; Kreykenbohm, I.; Langejahn, M.; Leiter, K.; Lovell, J. E. J.; Müller,

C.; Natusch, T.; Nesci, R.; Pursimo, T.; Phillips, C.; Plötz, C.; Quick, J.; Tzioumis, A. K.; **Weston, S.** ;
The TANAMI Multiwavelength Program: Dynamic spectral energy distributions of southern blazars
Astronomy & Astrophysics, Volume 591 (2016).

C. Müller, M. Kadler, R. Ojha, R. Schulz, J. Trüstedt, P. G. Edwards, E. Ros, B. Carpenter, R. Angioni, J. Blanchard, M. Böck, P. R. Burd, M. Dörr, M. S. Dutka, T. Eberl, S. Gulyaev, H. Hase, S. Horiuchi, U. Katz, F. Krauß, J. E. J. Lovell, T. Natusch, R. Nesci, C. Phillips, C. Plötz, T. Pursimo, J. F. H. Quick, J. Stevens, D. J. Thompson, S. J. Tingay, A. K. Tzioumis, **S. Weston**, J. Wilms and J. A. Zensus;
TANAMI: Tracking Active Galactic Nuclei with Austral Milliarcsecond Interferometry. II. Additional sources
Astronomy & Astrophysics, Volume 610 (2018).

S. D. Weston; N. Seymour; S. Gulyaev; R. P. Norris; J. Banfield; M. Vaccari; A. M. Hopkins; T. M. O. Franzen;
Automated cross-identifying radio to infrared surveys using the lrpy algorithm: a case study
Monthly Notices of the Royal Astronomical Society, Volume 473, Issue 4, 1 February 2018

An image of the magnificent starburst galaxy, Messier 82 (M82) used in this Thesis in Section 6.1.1 to model the evolutionary tracks of a SFG in the MIR. This mosaic image is the sharpest wide-angle view ever obtained of M82 by the HST (Gallagher et al., 2006). The galaxy is remarkable for its bright blue disk, webs of shredded clouds, and fiery-looking plumes of glowing hydrogen blasting out of its central regions. Credit : NASA, ESA, and The Hubble Heritage Team (STScI/AURA)

Appendix B

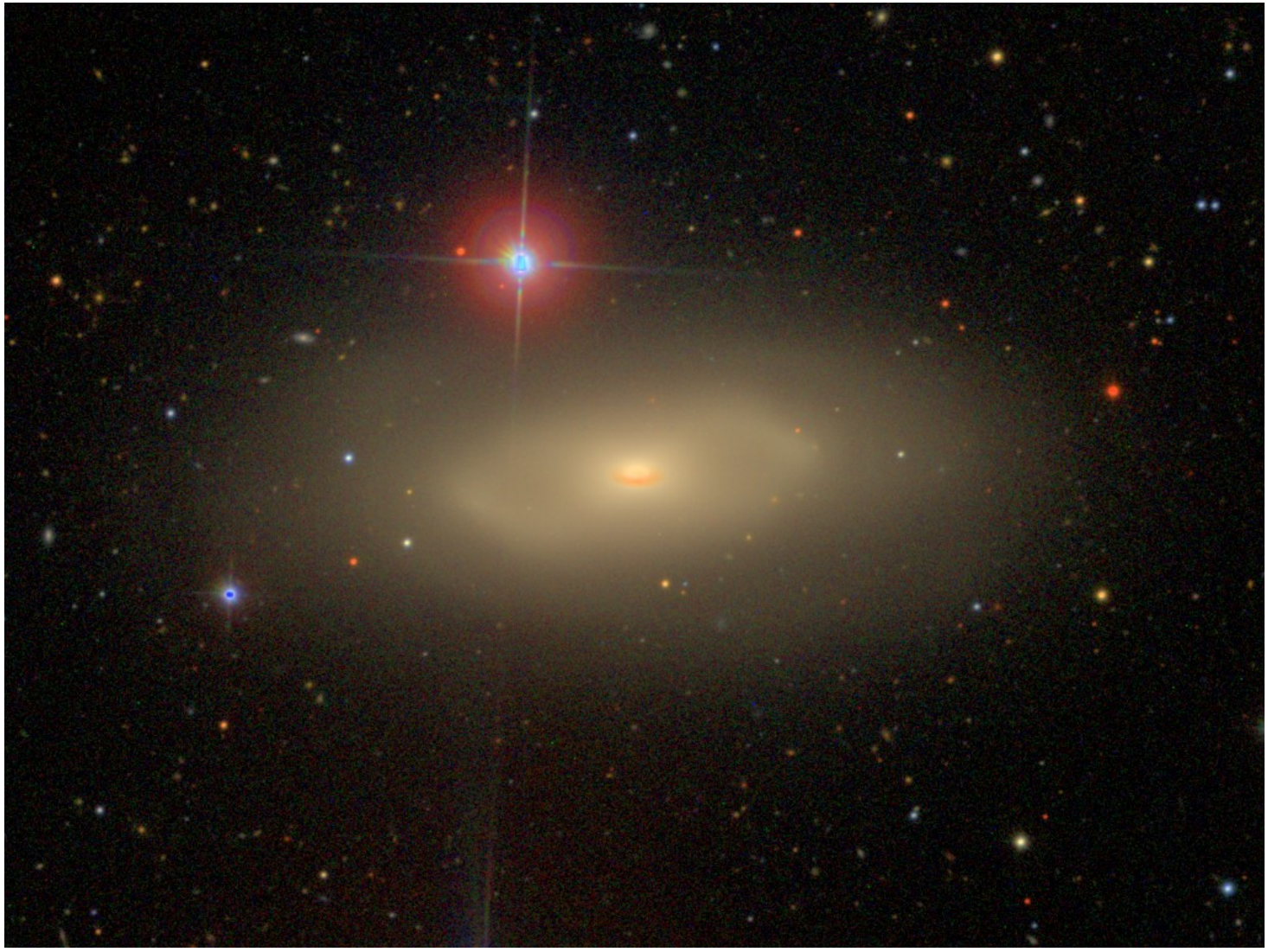
Electronic

Attached to this document is a DVD organised with folders for each Chapter.

The folders contain copy's of the figures in higher resolution, the data used to produce the figures and the appropriate scripts or code to make the figures.

The folder "Chapter 3" contains a full ascii copy of the produced XID catalogue from this thesis. In addition included is a copy of the code used for the XID plus code used for other tasks such as for the Luminosity work in section, although this is also available from *github*.

Also for Chapters 3,4 & 5 are a large collection of various postage stamp images of the XID radio sources, IRD and Radio Doubles which are in the corresponding folders for the chapters.



NGC4429 : A lenticular galaxy with very little star formation, this was also used in Section 6.1.1 to show evolution in MIR (Blanton and Hogg, 2006). Credit : David W. Hogg, Michael R. Blanton, and the Sloan Digital Sky Survey Collaboration.

Appendix C

Conference Posters

What follows are a selection of posters presented by myself at various conferences about some of the work and research I have been undertaking while a PhD student and employee of the Institute of Radio Astronomy and Space Research, AUT University.

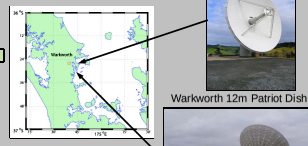
Radio Astronomy and eVLBI using KAREN

S.D.Weston*, T. Natusch and S. Gulyaev
Institute for Radio Astronomy and Space Research, AUT University
*stuart.weston@aut.ac.nz

Abstract

Kiwi Advanced Research and Education Network (KAREN) has been used to transfer large volumes of radio astronomical data between the AUT Radio Astronomical Observatory at Warkworth, New Zealand.

Here we report on the current status of connectivity international organisations with which we are collaborating and conducting observations and on the results of testing different data transfer protocols. We investigate new UDP protocols such as "tsunami" and UDT and show that the UDT protocol is a more efficient protocol. We report on our initial steps towards real-time Very Long Baseline Interferometry (eVLBI) and the attempt to directly stream data from the radio telescope receiving system to the correlation center without intermediate buffering/recording.



Warkworth 12m Patriot Dish

Warkworth 30m Dish ex Telecom

Background

With the connection of KAREN to the AUT Radio Telescope we wish to optimize the use of KAREN for transferring large volumes of observational data to our partners in Australia, the United States and Europe and for conducting eVLBI.

Our research and collaboration can be conditionally broken into three major areas:

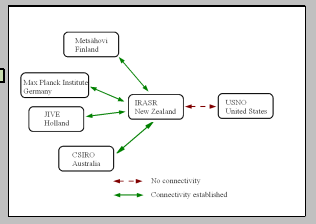
1. Observation and navigation of inter-planetary missions and spacecraft
2. International VLBI Service (IVS) for Geodesy and Astrometry
3. Astrophysical VLBI observations with the Australian Long Baseline Array of extragalatic and galactic radio sources

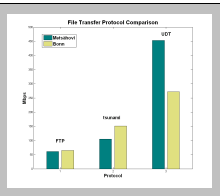
Point to point with no hops FTP is efficient, but as the number of hops in the route increases and the incidence of lost packets and collisions increases the TCP congestion avoidance algorithm becomes a severe limitation to the throughput that can be achieved.

We have compared the use of FTP over TCP/IP for sending these files to Australia, as well as new file transfer mechanisms which are being used by our partners such as tsunami [1] and UDT [2] (UDP-based Data Transfer) via the network protocol UDP.

[1] J. Wagner *Tsunami UDP Protocol* [Online]. http://tsunami-udp.sourceforge.net/ (2010, Sep. 23)

[2] *UDT UDP-based Data Transfer* [Online]. Available: http://udt.sourceforge.net/index.html (2010, Sep. 23)





Results

Over June and July 2010 tests where conducted transferring actual actual VLBI file (16 bit) produced by the AUT Radio Telescope during our VLBI observations:

Route	Protocol	Data	Time (s)	Throughput (Mbps)
AUT - Ilbrek	ftp	4.5G	90.36	8.5
AUT - Ilbrek	Tsunami	4.5G	36.66	116
AUT - Ilbrek	UDT	4.5G	19.20	2.75

Route	Protocol	Data	Time (s)	Throughput (Mbps)
AUT - Metsähovi	ftp	4.5G	76.58	9.5
AUT - Metsähovi	Tsunami	4.5G	49.79	140
AUT - Metsähovi	UDT	4.5G	31.57	4.53

Please see the histogram bottom center of this poster to see a comparison of these results.

Another set of experiments was conducted in September 2010 aiming to test the tsunami protocol for streaming VLBI data directly from the radio telescope receiving system via KAREN to Metsähovi. This test is an important step towards real-time eVLBI. When streaming data from Warkworth to Metsähovi many thousands of lost packets occurred and a sustainable rate of 350 Mbps was achieved. This is significantly lower than the rate of 512Mbps required for the real-time eVLBI streaming of 8bit data to Metsähovi.

Conclusion

It was clearly demonstrated that the use of the UDT protocol for radio astronomical data transfer has a number of advantages compared to the protocols currently used for VLBI and eVLBI. In particular, UDT has some advantages over tsunami:

- UDT is a better citizen on the network leaving bandwidth for TCP and other UDP protocols, which is very important on a shared network such as KAREN.
- UDT has an application programming interface (API) allowing easy integration with existing or future applications.

We have found KAREN to be a very useful tool for transmitting data to our international partners, and our Institute will be extending its use over the coming months. The next challenge will be to establish the IVS regular observational sessions and transmit the even larger volumes of data to the correlation centers, such as Bonn and USNO.

Of future benefit to our work to stream data real-time to the correlation centers would be the ability to reserve bandwidth as a logical pipe within the KAREN bandwidth for the duration of an experiment.

ACKNOWLEDGEMENTS: we wish to thank Donald Clark, Bill Choquette, John Raine and Chris Litter from REANNZ/KAREN, and Ben Yeldon from AUT's ICT Department for their support and assistance. We thank Guifre Molera of the Metsähovi Radio Observatory, Simone Bernhart of the Max Planck Institute for Radio Astronomy at Bonn, Paul Boven from JIVE, and Hiroshi Takeuchi of JAXA, Japan, who all assisted us in establishing connectivity to their sites. We are grateful to all our VLBI colleagues in Australia, in particular to Steven Tingay and the International Centre for Radio Astronomy Research,Curtin University of Technology and Tasso Tzioumis at CSIR.

FIGURE C.1: Poster for the Electronics New Zealand Conference (ENZCON) 2010; Hamilton, New Zealand

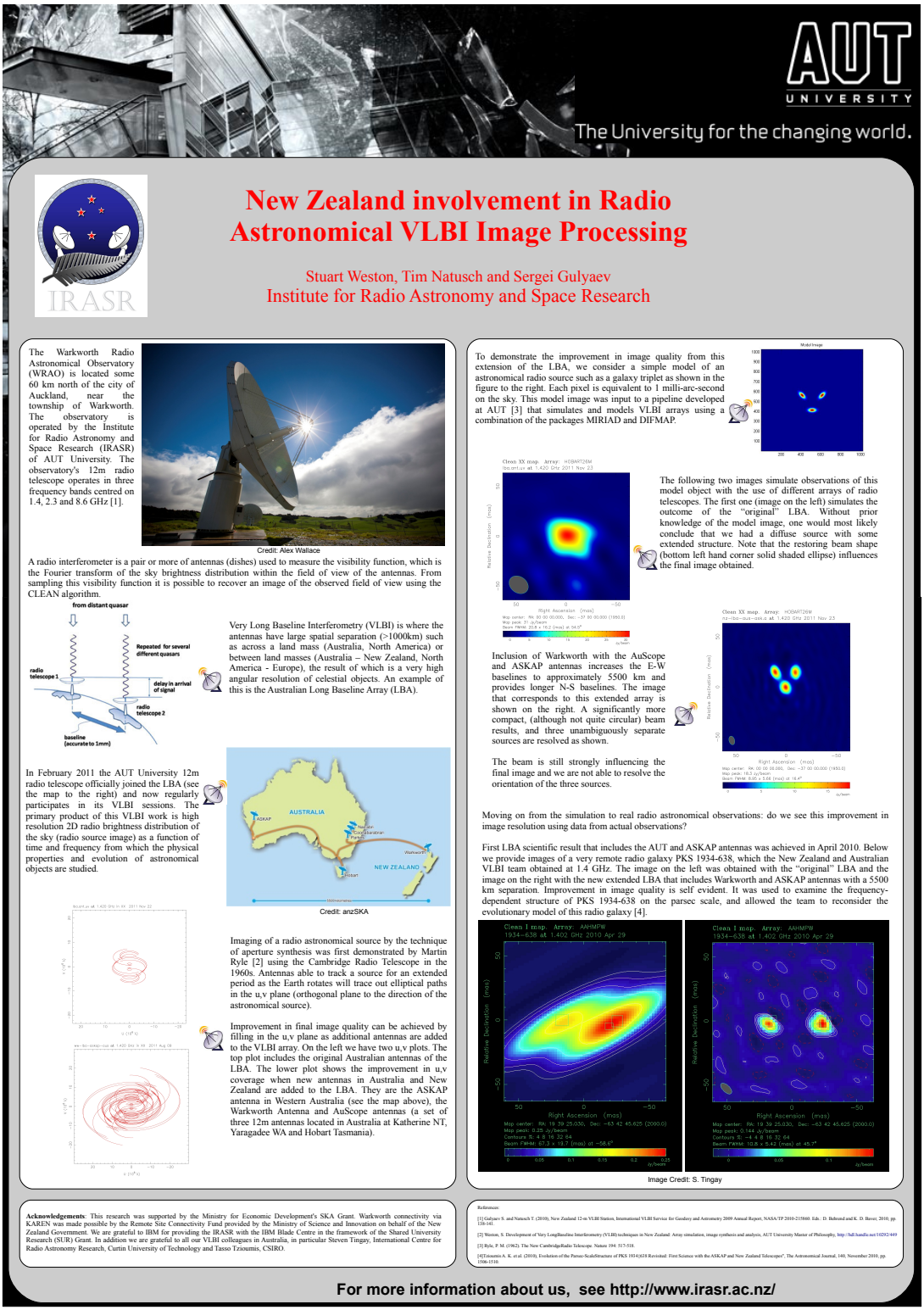


FIGURE C.2: Poster for the Image and Vision Computing New Zealand Conference (IVCNZ) 2011, Auckland, New Zealand.

The Little Kiwi Dish that Could

Stuart Weston (sweston@aut.ac.nz) and Jordan Alexander (jalexand@aut.ac.nz)

Institute for Radio Astronomy & Space Research, AUT University, Auckland

Can we do single dish science ?

With a 12m antenna can we do any science apart from just VLBI as the last remote element on the end of a long baseline array.

Abstract

We report on the first single dish experiments using the Warkworth 12m antenna with a digital baseband converter (DBBC) and Mark5B recorder operated by AUT University. The radio recombination lines of hydrogen H90α, H91α, H92α, helium He91α and carbon C91α were detected from OrionA, and H90α, H91α, H92α from M17 and NGC3603 at 8GHz. Ratios of integrated line intensities agree with published helium abundances.

Background

The Warkworth Radio Astronomical Observatory (WRAO) is located some 60 km north of the city of Auckland, near the township of Warkworth. The observatory is operated by the Institute for Radio Astronomy and Space Research (IRASR) of AUT University. The observatory's 12m radio telescope operates in three frequency bands centred on 1.4, 2.3 and 8.6 GHz [1]. We regularly participate in Australian Long Baseline array observations and also have an active schedule observing for the International VLBI Service for Geodesy & Astrometry.

So what's our advantage ?

In between IVS and LBA observations we have a lot of time to integrate and integrate and integrate ....zzzzzzz

So what equipment do we have that's different ?

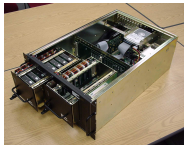
The Digital Baseband Converter (DBBC) replaces the VLBI terminal previously used with a complete and compact system that can be used with any VSI compliant recorder or data transport.

DBBC

The radio telescope at Warkworth is equipped with a DBBC [2] which replaces the VLBI terminal used elsewhere with a complete and compact system that can be used with any VSI compliant recorder or data transport. It consists of four modules each with four RF inputs. These can receive input in the ranges 0.01-512, 512-1024, 1024-1536, 1536-2048 MHz and 2048-2100 MHz.

Mk5

The other piece of important equipment is a Mark5B VLBI VSI-H compliant data recorder. This is used to record the data received from the DBBC at 2 bit resolution.



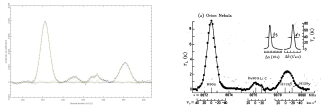
Did we see anything ?

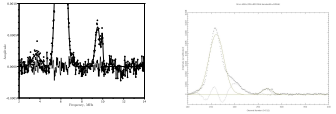
Judge for yourself

Let's also compare with published results.

Orion A

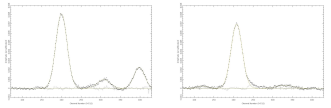
The plots show the integrated line profile for the recombination lines within X-band for a 16MHz band with gaussian fit (*solid line*), data (*crosses*), and the residual (*dashed line*). The recombination lines from left to right are our unprocessed H90α and for comparison on the right we include H90α from McGee [3]. The second line is left H154ϵ, H91α, He91α and C91α after processing and baseline fitting and right is our unprocessed H92α.





M17

We also observed M17 and again found the strong Hydrogen Alpha line. We haven't yet integrated longer for any additional weaker lines. The recombination lines from left to right are H90α and H91α.

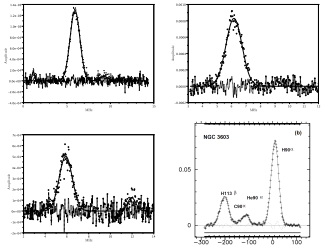


INSTITUTE FOR RADIO ASTRONOMY & SPACE RESEARCH

AN INSTITUTE OF AUT UNIVERSITY

NGC3603

Pree et al [5] detected H90α within this HII region using the ATCA. We present data for the additional lines after processing and baseline fitting H91α, H92α within this HII region. Integrated line profiles for H90α (top left), H91α (top right) and H92α (bottom left) are all 16MHz bands with a resolution of 512. Bottom right is H90α from Pree et al for comparison:



Conclusion

Scepticism has existed that a 12m dish with an un-cooled receiver designed for IVS would be able to detect radio recombination lines. We believe that it has been demonstrated that RRLs in X-band have been detected, with some unpublished. Thus from this work our 12m dish is capable of doing some single dish science which we hope to continue and expand.

References

[1] S. Gulyaev,T. Natusch, 2009, International VLBI Service for Geodesy and Astrometry 2009 Annual Report,D. Behrend & K. D. Baver, 138

[2] Gino et al, 2010, DBBC VLBI2010, IVS 2010 General Meeting Proceedings, D. Behrend, D. Baver, 'http://ivscc.gsfc.nasa.gov/publications/gm2010/tuccari1.pdf'

[3] R. X. McGee, L. M. NewtonR. A. Batchelor, Recombination lines near 8.9 GHz of strong sources in the southern Milky Way, Australian Journal of Physics, vol. 28, Apr. 1975, p. 185-207

[4] V. Pankonin, P. Thomasson, J. Barnulin,1977, A Survey of Radio Recombination Lines from HI Regions and Associated HII Regions, A&A, 54, 335-344

[5] C. G. Pree, M. C. Nyswander, W. M. Goss, NGC 3576 AND NGC 3603: TWO LUMINOUS SOUTHERN H II REGIONS OBSERVED AT HIGH RESOLUTION WITH THE AUSTRALIA TELESCOPE COMPACT ARRAY, 1999, ApJ, 117, 2002

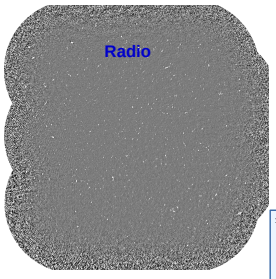
FIGURE C.3: Poster for the Astronomical Society of Australia Annual Scientific Meeting 2012; Sydney, NSW, Australia

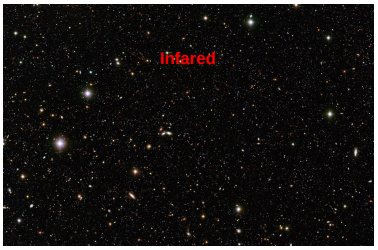
INSTITUTE FOR
RADIO ASTRONOMY & SPACE RESEARCH
AN INSTITUTE OF AUT UNIVERSITY

AUT

Survey Cross Identification

Stuart Weston
sweston@aut.ac.nz
AUT University





1

The old Visual Method

Mk1 eye of an experienced astronomer

2

PROBLEM

New Surveys many orders of magnitude greater in number of objects :

ATLAS DR1 had several hundred sources

ATLAS DR3 has ~ 5000 components

EMU ~ 70 million over a much larger area

3

Solution - Automation

Simple Nearest Neighbour

BUT

Only uses one piece of information – angular separation

4

But we have other information in the surveys:

Radio Flux, Infared Flux
Survey Source/Object Density

Can these assist in matching objects between two different catalogues ?

5

Let's look at the Likelihood Ratio

$$L = \frac{q(m) f(r)}{n(m)}$$

The ratio of probability distribution of true counterparts with magnitude m – q(m) with the probability distribution of positional errors f(r) by the surface density of unrelated background/foreground objects per unit magnitude/flux n(m)

6

Reliability

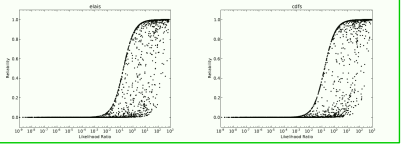
Each radio source can have one or more possible candidates, so how to select the most probable candidate. Therefore we use a reliability measure to compare the LR values :

$$Rel_j = \frac{(LR)_j}{\sum_{i=1}^{N}(LR)_i + (1 - Q_0)}$$

Q_0 is the probability that the true non radio candidate is above the detection limit.

7

Cross ID between ATLAS DR3 & Fusion *Spitzer* Infared



8

Does it work ?

CID	SWIRE_ID	r_arcsec	LR	Reliability
EI0003	220870	0.6828	521.836	0.99942
EI0003	220796	5.7643	0.10152	0.00019

CID	SWIRE_ID	r_arcsec	LR	Reliability
EI0017	112359	2.15206	241.4844	0.5822
EI0017	112417	1.08245	172.3513	0.4155

CID	SWIRE_ID	r_arcsec	LR	Reliability
EI0009C2	162771	5.0414	4.30099	0.52005
EI0009C2	162835	4.10418	3.769	0.45574
EI0009C2	162803	8.312	0.00016	0.00002

Yes !

9

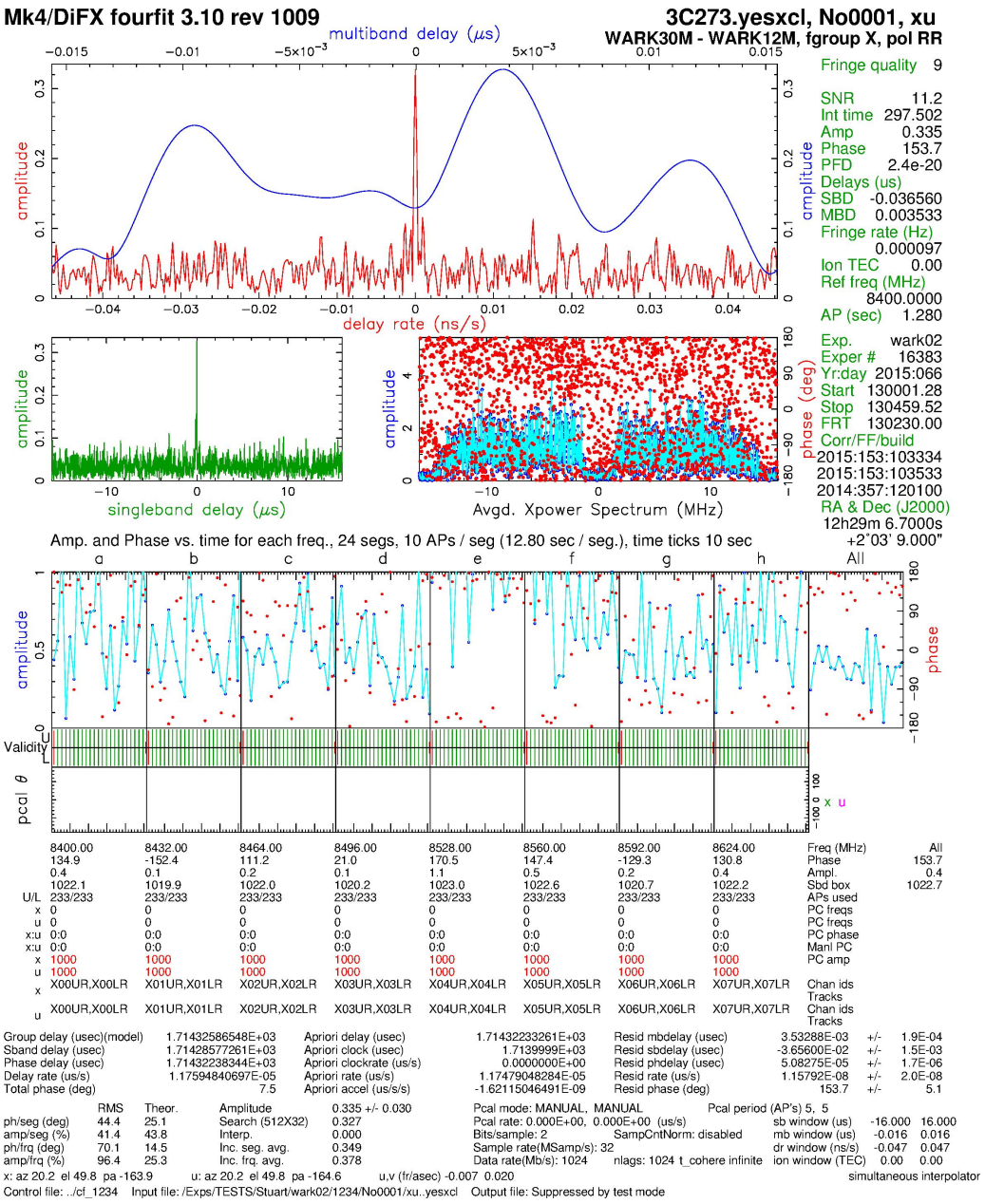
Further Work

Refine the algorithm
Can we identify radio doubles ?
IR doubles ?
Verify against other Catalogues

Stuart Weston
Email: sweston@aut.ac.nz
http://www.atnf.csiro.au/people/Stuart.Weston/index.html

Institute for Radio Astronomy & Space Research (IRASR)
School of Computer & Mathematical Sciences
Faculty of Design and Creative Technologies
AUT University, New Zealand.
http://www.irasr.aut.ac.nz

FIGURE C.4: Poster for the Astronomical Society of Australia Annual Scientific Meeting 2015; Fremantle, WA, Australia. This was awarded Runner Up Best Student Poster.



Another major milestone which I was involved in was the fist 2-element interfermetry between the two Warkworth antennas in X-Band where the schedule was produced by myself and the data was correlated here in New Zealand using DiFX (Deller et al., 2011) by myself 2015. Here is the plot from *fourfit* for scan No0001 on 3C273 showing the fringe obtained. Credit: Stuart Weston

Bibliography

Alexander, F. E. S. (1946). The sun's radio energy. *Radio & Electronics*, 1:16–17.

Alger, M. J., Banfield, J. K., Ong, C. S., Rudnick, L., Wong, O. I., Wolf, C., Andernach, H., Norris, R. P., and Shabala, S. S. (2018). Radio Galaxy Zoo: machine learning for radio source host galaxy cross-identification. *Mon. Not. R. Astron. Soc.*, 478:5547–5563.

Alonso-Herrero, A., Donley, J. L., Rieke, G. H., Rigby, J. R., and Pérez-González, P. G. (2007). The Quest for Obscured AGN at Cosmological Distances: Infrared Power-Law Galaxies. In Figueras, F., Girart, J. M., Hernanz, M., and Jordi, C., editors, *Highlights of Spanish Astrophysics IV*, page 143.

An, F. X., Stach, S. M., Smail, I., Swinbank, A. M., Almaini, O., Simpson, C., Hartley, W., Maltby, D. T., Ivison, R. J., Arumugam, V., Wardlow, J. L., Cooke, E. A., Gullberg, B., Thomson, A. P., Chen, C.-C., Simpson, J. M., Geach, J. E., Scott, D., Dunlop, J. S., Farrah, D., van der Werf, P., Blain, A. W., Conselice, C., Michałowski, M., Chapman, S. C., and Coppin, K. E. K. (2018). A Machine-learning Method for Identifying Multiwavelength Counterparts of Submillimeter Galaxies: Training and Testing Using AS2UDS and ALESS. *Astrophys. J.*, 862:101.

Argue, A. N., de Vegt, C., Elsmore, B., Fanselow, J., Harrington, R., Hemenway, P., Johnston, K. J., Kuhr, H., Kumkova, I., Niell, A. E., Walter, H., and Witzel, A. (1984). A catalog of selected compact radio sources for the construction of an extragalactic radio/optical reference frame. *Astronomy and Astrophysics*, 130:191–199.

Banfield, J. K., Wong, O. I., Willett, K. W., Norris, R. P., Rudnick, L., Shabala, S. S., Simmons, B. D., Snyder, C., Garon, A., Seymour, N., Middelberg, E., Andernach, H.,

Lintott, C. J., Jacob, K., Kapińska, A. D., Mao, M. Y., Masters, K. L., Jarvis, M. J., Schawinski, K., Paget, E., Simpson, R., Klöckner, H.-R., Bamford, S., Burchell, T., Chow, K. E., Cotter, G., Fortson, L., Heywood, I., Jones, T. W., Kaviraj, S., López-Sánchez, Á. R., Maksym, W. P., Polsterer, K., Borden, K., Hollow, R. P., and Whyte, L. (2015). Radio Galaxy Zoo: host galaxies and radio morphologies derived from visual inspection. *Mon. Not. R. Astron. Soc.*, 453:2326–2340.

Baum, W. A. (1962). Photoelectric Magnitudes and Red-Shifts. In McVittie, G. C., editor, *Problems of Extra-Galactic Research*, volume 15 of *IAU Symposium*, page 390.

Becker, R. H., White, R. L., and Helfand, D. J. (1994). The VLA's FIRST Survey. In Crabtree, D. R., Hanisch, R. J., and Barnes, J., editors, *Astronomical Data Analysis Software and Systems III*, volume 61 of *Astronomical Society of the Pacific Conference Series*, page 165.

Becker, R. H., White, R. L., and Helfand, D. J. (1995). The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters. *Astrophys. J.*, 450:559.

Bennett, A. S. (1962). The preparation of the revised 3C catalogue of radio sources. *Mon. Not. R. Astron. Soc.*, 125:75.

Blanton, M. R. and Hogg, D. W. (2006). Sdss images of selected rc3 galaxies. https://cosmo.nyu.edu/hogg/rc3/. [last visited 2/07/2019].

Bock, D. C.-J., Large, M. I., and Sadler, E. M. (1999). SUMSS: A Wide-Field Radio Imaging Survey of the Southern Sky. I. Science Goals, Survey Design, and Instrumentation. *Astronomical Journal*, 117:1578–1593.

Bolton, J. G. and Stanley, G. J. (1948). Variable Source of Radio Frequency Radiation in the Constellation of Cygnus. *Nature*, 161:312–313.

Bolton, J. G., Stanley, G. J., and Slee, O. B. (1949). Positions of Three Discrete Sources of Galactic Radio-Frequency Radiation. *Nature*, 164:101–102.

Bonzini, M., Mainieri, V., Padovani, P., Kellermann, K. I., Miller, N., Rosati, P., Tozzi, P., Vattakunnel, S., Balestra, I., Brandt, W. N., Luo, B., and Xue, Y. Q. (2012). The

Sub-mJy Radio Population of the E-CDFS: Optical and Infrared Counterpart Identification. *Astrophys. J., Suppl. Ser.*, 203:17.

Brown, R. H. (1968a). Measurement of Stellar Diameters. *Ann. Rev. Astron. Astrophys.*, 6:13.

Brown, R. H. (1968b). Stellar Interferometer at Narrabri Observatory. *Nature*, 218:637–641.

Budavári, T. and Szalay, A. S. (2008). Probabilistic Cross-Identification of Astronomical Sources. *Astrophys. J.*, 679:301–309.

Caltech (2015). The cosmic evolution survey (cosmos). http://cosmos.astro.caltech.edu/. [last visited 13/08/2018].

Carroll, B. W. and Ostlie, D. A. (2006). *An Introduction to Modern Astrophysics and Cosmology*.

Ching, J. H. Y., Sadler, E. M., Croom, S. M., Johnston, H. M., Pracy, M. B., Couch, W. J., Hopkins, A. M., Jurek, R. J., and Pimbblet, K. A. (2017). The Large Area Radio Galaxy Evolution Spectroscopic Survey (LARGESS): survey design, data catalogue and GAMA/WiggleZ spectroscopy. *Mon. Not. R. Astron. Soc.*, 464:1306–1332.

Ciliegi, P., G. Zamorani, M. Bondi, L. Pozzetti, M. Bolzonella, L. Gregorini, B. Garilli, A. Iovino, H. J. McCracken, Y. Mellier, M. Radovich, H. R. de Ruiter, P. Parma, D. Bottini, V. Le Brun, O. Le Fèvre, D. Maccagni, J. P. Picat, R. Scaramella, M. Scodeggio, L. Tresse, G. Vettolani, A. Zanichelli, C. Adami, M. Arnaboldi, S. Arnouts, S. Bardelli, A. Cappi, S. Charlot, T. Contini, S. Foucaud, P. Franzetti, L. Guzzo, O. Ilbert, B. Marano, C. Marinoni, G. Mathez, A. Mazure, B. Meneux, R. Merighi, P. Merluzzi, S. Paltani, A. Pollo, E. Zucca, A. Bongiorno, G. Busarello, I. Gavignaud, R. Pellò, V. Ripepi, and D. Rizzo (2005). The VVDS-VLA deep field. *Astron. Astrophys.*, 441(3):879–891.

Ciliegi, P., Zamorani, G., Hasinger, G., Lehmann, I., Szokoly, G., and Wilson, G. (2003). A deep vla survey at 6 cm in the lockman hole. *Astron. Astrophys.*, 398:901.

Condon, J. J., Anderson, M. L., and Helou, G. (1991). Correlations between the far-infrared, radio, and blue luminosities of spiral galaxies. *Astrophys. J.*, 376:95–103.

Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., and Broderick, J. J. (1998). The NRAO VLA Sky Survey. *Astron. J.*, 115:1693–1716.

Coney, S. (2013). *On The Radar*. Keyhole Press & PPH. The Story of Piha's World War 2 radar station.

Dale, D. A. and Helou, G. (2002). The Infrared Spectral Energy Distribution of Normal Star-forming Galaxies: Calibration at Far-Infrared and Submillimeter Wavelengths. *Astrophys. J.*, 576:159–168.

D'Andrea, C. and OzDES (2014). OzDES: 100 Nights of AAT Spectroscopy on DES Sources. In *American Astronomical Society Meeting Abstracts #223*, volume 223 of *American Astronomical Society Meeting Abstracts*, page 254.15.

de Ruiter, H. R., Willis, A. G., and Arp, H. C. (1977). A Westerbork 1415 MHz survey of background radio sources. II - Optical identifications with deep IIIA-J plates. *Astron. Astrophys. Suppl. Ser.*, 28:211–293.

Deller, A. T., Brisken, W. F., Phillips, C. J., Morgan, J., Alef, W., Cappallo, R., Middelberg, E., Romney, J., Rottmann, H., Tingay, S. J., and Wayth, R. (2011). DiFX-2: A More Flexible, Efficient, Robust, and Powerful Software Correlator. *Publ. Astron. Soc. Pac.*, 123:275.

Devriendt, J. E. G., Guiderdoni, B., and Sadat, R. (1999). Galaxy modelling. I. Spectral energy distributions from far-UV to sub-mm wavelengths. *Astron. Astrophys.*, 350:381–398.

Dickinson, M., Giavalisco, M., and GOODS Team (2003). The Great Observatories Origins Deep Survey. In Bender, R. and Renzini, A., editors, *The Mass of Galaxies at Low and High Redshift*, page 324.

Donley, J. L., Koekemoer, A. M., Brusa, M., Capak, P., Cardamone, C. N., Civano, F., Ilbert, O., Impey, C. D., Kartaltepe, J. S., Miyaji, T., Salvato, M., Sanders, D. B.,

Trump, J. R., and Zamorani, G. (2012). Identifying Luminous Active Galactic Nuclei in Deep Surveys: Revised IRAC Selection Criteria. *Astrophys. J.*, 748:142.

Downes, A. J. B., Peacock, J. A., Savage, A., and Carrie, D. R. (1986). The Parkes selected regions - Powerful radio galaxies and quasars at high redshifts. *Mon. Not. R. Astron. Soc.*, 218:31–62.

Dunlop, J. S., Peacock, J. A., Savage, A., Lilly, S. J., Heasley, J. N., and Simon, A. J. B. (1989). The Parkes Selected Regions - Deep optical and infrared observations of radio galaxies and quasars at high redshifts. *Mon. Not. R. Astron. Soc.*, 238:1171–1231.

Edge, D. O., Shakeshaft, J. R., McAdam, W. B., Baldwin, J. E., and Archer, S. (1959). A survey of radio sources at a frequency of 159 Mc/s. *Mem. R. Astron. Soc.*, 68:37–60.

Eisenhardt, P. R., Stern, D., Brodwin, M., Fazio, G. G., Rieke, G. H., Rieke, M. J., Werner, M. W., Wright, E. L., Allen, L. E., Arendt, R. G., Ashby, M. L. N., Barmby, P., Forrest, W. J., Hora, J. L., Huang, J.-S., Huchra, J., Pahre, M. A., Pipher, J. L., Reach, W. T., Smith, H. A., Stauffer, J. R., Wang, Z., Willner, S. P., Brown, M. J. I., Dey, A., Jannuzi, B. T., and Tiede, G. P. (2004). The Infrared Array Camera (IRAC) Shallow Survey. *Astrophys. J., Suppl. Ser.*, 154:48–53.

Elvis, M., Wilkes, B. J., McDowell, J. C., Green, R. F., Bechtold, J., Willner, S. P., Oey, M. S., Polomski, E., and Cutri, R. (1994). Atlas of quasar energy distributions. *Astrophys. J., Suppl. Ser.*, 95:1–68.

ESA, editor (1997). *The HIPPARCOS and TYCHO catalogues. Astrometric and photometric star catalogues derived from the ESA HIPPARCOS Space Astrometry Mission*, volume 1200 of *ESA Special Publication*.

Fan, D., Budavári, T., Norris, R. P., and Hopkins, A. M. (2015). Matching radio catalogues with realistic geometry: application to SWIRE and ATLAS. *Mon. Not. R. Astron. Soc.*, 451:1299–1305.

Fazio, G. G., Hora, J. L., Allen, L. E., Ashby, M. L. N., Barmby, P., Deutsch, L. K., Huang, J.-S., Kleiner, S., Marengo, M., Megeath, S. T., Melnick, G. J., Pahre, M. A.,

Patten, B. M., Polizotti, J., Smith, H. A., Taylor, R. S., Wang, Z., Willner, S. P., Hoffmann, W. F., Pipher, J. L., Forrest, W. J., McMurty, C. W., McCreight, C. R., McKelvey, M. E., McMurray, R. E., Koch, D. G., Moseley, S. H., Arendt, R. G., Mentzell, J. E., Marx, C. T., Losch, P., Mayman, P., Eichhorn, W., Krebs, D., Jhabvala, M., Gezari, D. Y., Fixsen, D. J., Flores, J., Shakoorzadeh, K., Jungo, R., Hakun, C., Workman, L., Karpati, G., Kichak, R., Whitley, R., Mann, S., Tollestrup, E. V., Eisenhardt, P., Stern, D., Gorjian, V., Bhattacharya, B., Carey, S., Nelson, B. O., Glaccum, W. J., Lacy, M., Lowrance, P. J., Laine, S., Reach, W. T., Stauffer, J. A., Surace, J. A., Wilson, G., Wright, E. L., Hoffman, A., Domingo, G., and Cohen, M. (2004). The Infrared Array Camera (IRAC) for the Spitzer Space Telescope. *Astrophys. J., Suppl. Ser.*, 154:10–17.

Fleuren, S., Sutherland, W., Dunne, L., Smith, D. J. B., Maddox, S. J., González-Nuevo, J., Findlay, J., Auld, R., Baes, M., Bond, N. A., Bonfield, D. G., Bourne, N., Cooray, A., Buttiglione, S., Cava, A., Dariush, A., De Zotti, G., Driver, S. P., Dye, S., Eales, S., Fritz, J., Gunawardhana, M. L. P., Hopwood, R., Ibar, E., Ivison, R. J., Jarvis, M. J., Kelvin, L., Lapi, A., Liske, J., Michałowski, M. J., Negrello, M., Pascale, E., Pohlen, M., Prescott, M., Rigby, E. E., Robotham, A., Scott, D., Temi, P., Thompson, M. A., Valiante, E., and van der Werf, P. (2012). Herschel-ATLAS: VISTA VIKING near-infrared counterparts in the Phase 1 GAMA 9-h data. *Mon. Not. R. Astron. Soc.*, 423:2407.

Franzen, T. M. O., Banfield, J. K., Hales, C. A., Hopkins, A., Norris, R. P., Seymour, N., Chow, K. E., Herzog, A., Huynh, M. T., Lenc, E., Mao, M. Y., and Middelberg, E. (2015). ATLAS - I. Third release of 1.4 GHz mosaics and component catalogues. *Mon. Not. R. Astron. Soc.*, 453:4020–4036.

Fricke, W., Schwan, H., Lederle, T., Bastian, U., Bien, R., Burkhardt, G., Du Mont, B., Hering, R., Jährling, R., Jahreiß, H., Röser, S., Schwerdtfeger, H.-M., and Walter, H. G. (1988). Fifth fundamental catalogue (FK5). Part 1: The basic fundamental stars. *Veroeffentlichungen des Astronomischen Rechen-Instituts Heidelberg*, 32:1–106.

Gallagher, J., Mountain, M., and Puxley, P. (2006). Happy sweet sixteen, hubble telescope! – starburst galaxy m82. https://hubblesite.org/contents/media/images/2006/14/1876-Image.html. [last visited 2/07/2019].

Giacconi, R., Rosati, P., Tozzi, P., Nonino, M., Hasinger, G., Norman, C., Bergeron, J., Borgani, S., Gilli, R., Gilmozzi, R., and Zheng, W. (2001). First Results from the XRay and Optical Survey of the Chandra Deep Field South. *Astrophys. J.*, 551:624–634.

Hardcastle, M. J. et al. (2010). Herschel-atlas: far-infrared properties of radio-selected galaxies. *Mon. Not. R. Astron. Soc.*, 409:122–131.

Harris, M. (2017). Rocks, radio and radar: Elizabeth Alexander and the DSIR Radio Development Laboratory, 1942–1945. *Journal of the Royal Society of New Zealand*, 47(2):193–198.

Hazard, C., Jauncey, D., Goss, W. M., and Herald, D. (2015). The sequence of events that led to the 1963 publications in Nature of 3C273, the first quasar and the first extragalactic radio jet. In Massaro, F., Cheung, C. C., Lopez, E., and Siemiginowska, A., editors, *Extragalactic Jets from Every Angle*, volume 313 of *IAU Symposium*, pages 183–189.

Hazard, C., Mackey, M. B., and Shimmins, A. J. (1963). Investigation of the Radio Source 3C 273 By The Method of Lunar Occultations. *Nature*, 197:1037–1039.

Helou, G., Lu, N. Y., Werner, M. W., Malhotra, S., and Silbermann, N. (2000). The Mid-Infrared Spectra of Normal Galaxies. *Astrophys. J., Lett.*, 532:L21–L24.

Hogg, D. W. (1999). Distance measures in cosmology. *ArXiv Astrophysics e-prints*.

Hubble, E. (1929). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173.

Hubble, E. P. (1925). Cepheids in spiral nebulae. *Observatory*, 48:139–142.

Huynh, M. T., Jackson, C. A., Norris, R. P., and Prandoni, I. (2005). Radio Observations of the Hubble Deep Field-South Region. II. The 1.4 GHz Catalog and Source Counts. *Astron. J.*, 130:1373.

IRSA (2019). Irac instrument handbook. https://irsa.ipac.caltech.edu/data/SPITZER/docs/irac/iracinstrumenthandbook/6/. [last visited 2/03/2019].

Ivison, R. J., Greve, T. R., Dunlop, J. S., Peacock, J. A., Egami, E., Smail, I., Ibar, E., van Kampen, E., Aretxaga, I., Babbedge, T., Biggs, A. D., Blain, A. W., Chapman, S. C., Clements, D. L., Coppin, K., Farrah, D., Halpern, M., Hughes, D. H., Jarvis, M. J., Jenness, T., Jones, J. R., Mortier, A. M. J., Oliver, S., Papovich, C., Pérez-González, P. G., Pope, A., Rawlings, S., Rieke, G. H., Rowan-Robinson, M., Savage, R. S., Scott, D., Seigar, M., Serjeant, S., Simpson, C., Stevens, J. A., Vaccari, M., Wagg, J., and Willott, C. J. (2007). The SCUBA HAlf Degree Extragalactic Survey - III. Identification of radio and mid-infrared counterparts to submillimetre galaxies. *Mon. Not. R. Astron. Soc.*, 380:199–228.

Jansky, K. G. (1933). Radio Waves from Outside the Solar System. *Nature*, 132:66.

Johnston, S., Bailes, M., Bartel, N., Baugh, C., Bietenholz, M., Blake, C., Braun, R., Brown, J., Chatterjee, S., Darling, J., Deller, A., Dodson, R., Edwards, P. G., Ekers, R., Ellingsen, S., Feain, I., Gaensler, B. M., Haverkorn, M., Hobbs, G., Hopkins, A., Jackson, C., James, C., Joncas, G., Kaspi, V., Kilborn, V., Koribalski, B., Kothes, R., Landecker, T. L., Lenc, E., Lovell, J., Macquart, J.-P., Manchester, R., Matthews, D., McClure-Griffiths, N. M., Norris, R., Pen, U.-L., Phillips, C., Power, C., Protheroe, R., Sadler, E., Schmidt, B., Stairs, I., Staveley-Smith, L., Stil, J., Taylor, R., Tingay, S., Tzioumis, A., Walker, M., Wall, J., and Wolleben, M. (2007). Science with the Australian Square Kilometre Array Pathfinder. *Proc. Astron. Soc. Aust.*, 24:174–188.

Jonas, J. L. (2009). MeerKAT - The South African Array With Composite Dishes and Wide-Band Single Pixel Feeds. *IEEE Proceedings*, 97:1522–1530.

Joye, W. A. and Mandel, E. (2003). New Features of SAOImage DS9. In Payne, H. E., Jedrzejewski, R. I., and Hook, R. N., editors, *Astronomical Data Analysis Software and Systems XII*, volume 295 of *Astronomical Society of the Pacific Conference Series*, page 489.

Lacy, M., Storrie-Lombardi, L. J., Sajina, A., Appleton, P. N., Armus, L., Chapman, S. C., Choi, P. I., Fadda, D., Fang, F., Frayer, D. T., Heinrichsen, I., Helou, G.,

Im, M., Marleau, F. R., Masci, F., Shupe, D. L., Soifer, B. T., Surace, J., Teplitz, H. I., Wilson, G., and Yan, L. (2004). Obscured and Unobscured Active Galactic Nuclei in the Spitzer Space Telescope First Look Survey. *Astrophys. J., Suppl. Ser.*, 154(1):166–169.

Leahy, D. A., Hopkins, A. M., Norris, R. P., Marvil, J., Collier, J. D., Taylor, E. N., Allison, J. R., Anderson, C., Bell, M., and Bilicki, M. (2019). ASKAP Commissioning Observations of the GAMA 23 Field. *Proc. Astron. Soc. Aust.*, page arXiv:1905.00831.

Lovell, A. C. B. and Banwell, C. J. (1946). Abnormal Solar Radiation on 72 Megacycles. *Nature*, 158:517–518.

Magliocchetti, M., Maddox, S. J., Lahav, O., and Wall, J. V. (1998). Variance and skewness in the FIRST survey. *Mon. Not. R. Astron. Soc.*, 300:257–268.

Mao, M. Y., Sharp, R., Norris, R. P., Hopkins, A. M., Seymour, N., Lovell, J. E. J., Middelberg, E., Randall, K. E., Sadler, E. M., Saikia, D. J., Shabala, S. S., and Zinn, P.-C. (2012). The Australia Telescope Large Area Survey: spectroscopic catalogue and radio luminosity functions. *Mon. Not. R. Astron. Soc.*, 426:3334–3348.

Mauch, T., Murphy, T., Buttery, H. J., Curran, J., Hunstead, R. W., Piestrzynski, B., Robertson, J. G., and Sadler, E. M. (2003). SUMSS: a wide-field radio imaging survey of the southern sky - II. The source catalogue. *Mon. Not. R. Astron. Soc.*, 342:1117–1130.

Mauch, T. and Sadler, E. M. (2007). Radio sources in the 6dFGS: local luminosity functions at 1.4GHz for star-forming galaxies and radio-loud AGN. *Mon. Not. R. Astron. Soc.*, 375:931–950.

Michelson, A. A. and Pease, F. G. (1921). Measurement of the Diameter of α Orionis with the Interferometer. *Astrophys. J.*, 53.

Middelberg, E., Norris, R. P., Cornwell, T. J., Voronkov, M. A., Siana, B. D., Boyle, B. J., Ciliegi, P., Jackson, C. A., Huynh, M. T., Berta, S., Rubele, S., Lonsdale, C. J.,

Ivison, R. J., Smail, I., and Oliver, S. J. (2007). Deep ATLAS Radio Observations of the ELAIS-S1/Spitzer Wide-Area Infrared Extragalctic field. *ArXiv e-prints*.

Mignard, F., Klioner, S., Lindegren, L., Hernandez, J., Bastian, U., and Bombrun, A. (2018). Gaia Data Release 2: The Celestial reference frame (Gaia-CRF2). *ArXiv e-prints*.

Miller, N. A., Bonzini, M., Fomalont, E. B., Kellermann, K. I., Mainieri, V., Padovani, P., Rosati, P., Tozzi, P., and Vattakunnel, S. (2013). The Very Large Array 1.4 GHz Survey of the Extended Chandra Deep Field South: Second Data Release. *Astrophys. J., Suppl. Ser.*, 205:13.

Mingo, B., Watson, M. G., Rosen, S. R., Hardcastle, M. J., Ruiz, A., Blain, A., Carrera, F. J., Mateos, S., Pineau, F.-X., and Stewart, G. C. (2016). The MIXR sample: AGN activity versus star formation across the cross-correlation of WISE, 3XMM, and FIRST/NVSS. *Mon. Not. R. Astron. Soc.*, 462:2631–2667.

Norris, R. P. (2017). Extragalactic radio continuum surveys and the transformation of radio astronomy. *Nature Astronomy*, 1:671–678.

Norris, R. P., Afonso, J., Appleton, P. N., Boyle, B. J., Ciliegi, P., Croom, S. M., Huynh, M. T., Jackson, C. A., Koekemoer, A. M., Lonsdale, C. J., Middelberg, E., Mobasher, B., Oliver, S. J., Polletta, M., Siana, B. D., Smail, I., and Voronkov, M. A. (2006). Deep ATLAS Radio Observations of the Chandra Deep Field-South/Spitzer Wide-Area Infrared Extragalactic Field. *Astrophys. J.*, 132:2409.

Norris, R. P., Hopkins, A. M., Afonso, J., et al. (2011). EMU: Evolutionary Map of the Universe. *Proc. Astron. Soc. Aust.*, 28:215–248.

Norris, R. P., Seymour, N., Sharp, R., and Mao, M. (2013). The ATLAS spare fibre program for OzDES. https://cdcvs.fnal.gov/redmine/attachments/download/9085/ATLAS-sparefibres.pdf. [last visited 12/06/2019].

Norris, R. P., Tingay, S., Phillips, C., Middelberg, E., Deller, A., and Appleton, P. N. (2007). Very long baseline interferometry detection of an Infrared-Faint Radio Source. *Mon. Not. R. Astron. Soc.*, 378:1434.

Nothnagel, A., Artz, T., Behrend, D., and Malkin, Z. (2017). International VLBI Service for Geodesy and Astrometry – Delivering high-quality products and embarking on observations of the next generation. *Journal of Geodesy*, 91:711–721.

OzDES (2016). The dark energy survey collaboration. http://www.mso.anu.edu.au/ozdes/index.html/.

Pawsey, J. L., Payne-Scott, R., and McCready, L. L. (1946). Radio-Frequency Energy from the Sun. *Nature*, 157:158–159.

Peeters, E., Spoon, H. W. W., and Tielens, A. G. G. M. (2004). Polycyclic Aromatic Hydrocarbons as a Tracer of Star Formation? *Astrophys. J.*, 613:986–1003.

Prandoni, I. and Seymour, N. (2015). Revealing the Physics and Evolution of Galaxies and Galaxy Clusters with SKA Continuum Surveys. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 67.

Reber, G. (1944). Cosmic Static. *Astrophys. J.*, 100:279.

Renzini, A., Cesarsky, C., Cristiani, S., da Costa, L., Fosbury, R., Hook, R., Leibundgut, B., Rosati, P., and Vandame, B. (2003). ESO for GOODS' Sake. In Bender, R. and Renzini, A., editors, *The Mass of Galaxies at Low and High Redshift*, page 332.

Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., Hall, P. B., Gallagher, S. C., Hines, D. C., Fan, X., Papovich, C., Vanden Berk, D. E., Trammell, G. B., Schneider, D. P., Vestergaard, M., York, D. G., Jester, S., Anderson, S. F., Budavári, T., and Szalay, A. S. (2006). Spectral Energy Distributions and Multiwavelength Selection of Type 1 Quasars. *Astrophys. J., Suppl. Ser.*, 166:470.

Richter, G. A. (1975). Search for optical identifications in the 5C3-radio survey. II - Statistical treatment and results. *Astronomische Nachrichten*, 296:65–81.

Robertson, H. P. (1929). On the Foundations of Relativistic Cosmology. *Proceedings of the National Academy of Science*, 15:822–829.

Rowan Robinson, M., Oliver, S., Efstathiou, A., et al. (1999). The European large area ISO survey: ELAIS. In Cox, P. and Kessler, M., editors, *The Universe as Seen by ISO*, volume 427 of *ESA Special Publication*, page 1011.

Rowan-Robinson, M., Roseboom, I. G., Vaccari, M., Amblard, A., Arumugam, V., Auld, R., Aussel, H., Babbedge, T., Blain, A., Bock, J., Boselli, A., Brisbin, D., Buat, V., Burgarella, D., Castro-Rodriguez, N., Cava, A., Chanial, P., Clements, D. L., Conley, A., Conversi, L., Cooray, A., Dowell, C. D., Dwek, E., Dye, S., Eales, S., Elbaz, D., Farrah, D., Fox, M., Franceschini, A., Gear, W., Glenn, J., Solares, E. A. G., Griffin, M., Halpern, M., Hatziminaoglou, E., Huang, J., Ibar, E., Isaak, K., Ivison, R. J., Lagache, G., Levenson, L., Lu, N., Madden, S., Maffei, B., Mainetti, G., Marchetti, L., Mortier, A. M. J., Nguyen, H. T., O'Halloran, B., Oliver, S. J., Omont, A., Page, M. J., Panuzzo, P., Papageorgiou, A., Patel, H., Pearson, C. P., Fournon, I. P., Pohlen, M., Rawlings, J. I., Raymond, G., Rigopoulou, D., Rizzo, D., Schulz, B., Scott, D., Seymour, N., Shupe, D. L., Smith, A. J., Stevens, J. A., Symeonidis, M., Trichas, M., Tugwell, K. E., Valtchanov, I., Vigroux, L., Wang, L., Ward, R., Wright, G., Xu, C. K., and Zemcov, M. (2010). Cold dust and young starbursts: spectral energy distributions of herschel spire sources from the hermes survey. *Mon. Not. R. Astron. Soc.*, 409(1):2–11.

Ryle, M. (1952). A New Radio Interferometer and Its Application to the Observation of Weak Radio Stars. *Proceedings of the Royal Society of London Series A*, 211:351–375.

Ryle, M., Smith, F. G., and Elsmore, B. (1950). A preliminary survey of the radio stars in the Northern Hemisphere. *Mon. Not. R. Astron. Soc.*, 110:508.

Ryle, M. and Vonberg, D. D. (1946). Solar Radiation on 175 Mc./s. *Nature*, 158:339–340.

Sadler, E. M., Jackson, C. A., Cannon, R. D., McIntyre, V. J., Murphy, T., Bland-Hawthorn, J., Bridges, T., Cole, S., Colless, M., Collins, C., Couch, W., Dalton, G., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Norberg, P., Peacock, J. A., Peterson, B. A., Sutherland, W., and Taylor, K. (2002). Radio sources in the 2dF Galaxy Redshift Survey - II. Local radio luminosity functions for AGN and star-forming galaxies at 1.4 GHz. *Mon. Not. R. Astron. Soc.*, 329:227–245.

Salvato, M., Buchner, J., Budavári, T., Dwelly, T., Merloni, A., Brusa, M., Rau, A., Fotopoulou, S., and Nandra, K. (2018). Finding counterparts for all-sky X-ray surveys with NWAY: a Bayesian algorithm for cross-matching multiple catalogues. *Mon. Not. R. Astron. Soc.*, 473:4937–4955.

Schmidt, B. P., Suntzeff, N. B., Phillips, M. M., Schommer, R. A., Clocchiatti, A., Kirshner, R. P., Garnavich, P., Challis, P., Leibundgut, B., Spyromilio, J., Riess, A. G., Filippenko, A. V., Hamuy, M., Smith, R. C., Hogan, C., Stubbs, C., Diercks, A., Reiss, D., Gilliland, R., Tonry, J., Maza, J., Dressler, A., Walsh, J., and Ciardullo, R. (1998). The High-Z Supernova Search: Measuring Cosmic Deceleration and Global Curvature of the Universe Using Type IA Supernovae. *Astrophys. J.*, 507:46–63.

Schmidt, M. (1963). 3C 273 : A Star-Like Object with Large Red-Shift. *Nature*, 197:1040.

Scoville, N., Aussel, H., Brusa, M., Capak, P., Carollo, C. M., Elvis, M., Giavalisco, M., Guzzo, L., Hasinger, G., Impey, C., Kneib, J.-P., LeFevre, O., Lilly, S. J., Mobasher, B., Renzini, A., Rich, R. M., Sanders, D. B., Schinnerer, E., Schminovich, D., Shopbell, P., Taniguchi, Y., and Tyson, N. D. (2007). The Cosmic Evolution Survey (COSMOS): Overview. *Astrophysical Journal, Supplement*, 172:1–8.

Seymour, N., Dwelly, T., Moss, D., McHardy, I., Zoghbi, A., Rieke, G., Page, M., Hopkins, A., and Loaring, N. (2008). The star formation history of the Universe as revealed by deep radio observations. *Mon. Not. R. Astron. Soc.*, 386(3):1695–1708.

Seymour, N., Stern, D., De Breuck, C., Vernet, J., Rettura, A., Dickinson, M., Dey, A., Eisenhardt, P., Fosbury, R., Lacy, M., McCarthy, P., Miley, G., Rocca-Volmerange, B., Röttgering, H., Stanford, S. A., Teplitz, H., van Breugel, W., and Zirm, A. (2007). The Massive Hosts of Radio Galaxies across Cosmic Time. *Astrophys. J., Suppl. Ser.*, 171:353–375.

Slipher, V. M. (1921). Two Nebulae with Unparalleled Velocities. *Popular Astronomy*, 29:128.

Slipher, V. M. (1922). Further Notes on Spectrographic Observations of Nebulae and Clusters. In *Publications of the American Astronomical Society*, volume 4 of *Publications of the American Astronomical Society*, page 284.

Smith, D. J. B., Dunne, L., Maddox, S. J., and Eales, S. (2011). Herschel-ATLAS: counterparts from the ultraviolet-near-infrared in the science demonstration phase catalogue. *Mon. Not. R. Astron. Soc.*, 416:857–872.

Smith, G. A., Saunders, W., Bridges, T., Churilov, V., Lankshear, A., Dawson, J., Correll, D., Waller, L., Haynes, R., and Frost, G. (2004). *AAOmega: a multipurpose fiber-fed spectrograph for the AAT*, volume 5492 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 410–420.

Stern, D., Eisenhardt, P., Gorjian, V., Kochanek, C. S., Caldwell, N., Eisenstein, D., Brodwin, M., Brown, M. J. I., Cool, R., Dey, A., Green, P., Jannuzi, B. T., Murray, S. S., Pahre, M. A., and Willner, S. P. (2005). Mid-Infrared Selection of Active Galaxies. *Astrophys. J.*, 631:163.

Sutherland, W. and Saunders, W. (1992). On the likelihood ratio for source identification. *Mon. Not. R. Astron. Soc.*, 259:413.

Taylor, M. B. (2005). TOPCAT & STIL: Starlink Table/VOTable Processing Software. In Shopbell, P., Britton, M., and Ebert, R., editors, *Astronomical Data Analysis Software and Systems XIV*, volume 347 of *Astronomical Society of the Pacific Conference Series*, page 29.

Vaccari, M. (2015). The Spitzer Data Fusion: Contents, Construction and Applications to Galaxy Evolution Studies. In *The Many Facets of Extragalactic Radio Surveys: Towards New Scientific Challenges*, page 27.

Vaccari, M. (2016). HELP: The Herschel Extragalactic Legacy Project and The Coming of Age of Multi-wavelength Astrophysics. *The Universe of Digital Sky Surveys*, 42:71.

Vaccari, M., Marchetti, L., and Gonzalez-Solares, E. (2010a). Spitzer Data Fusion : A Spitzer-Selected Multi-Wavelength Catalog. http://mattiavaccari.net/df/. [Online; accessed 18-August-2014].

Vaccari, M., Marchetti, L., and Gonzalez-Solares, E. (2010b). Spitzer Data Fusion : A Spitzer-Selected Multi-Wavelength Catalog. http://mattiavaccari.net/df/. [Online; accessed 18-August-2014].

Vanden Berk, D. E., Richards, G. T., Bauer, A., Strauss, M. A., Schneider, D. P., Heckman, T. M., York, D. G., Hall, P. B., Fan, X., Knapp, G. R., Anderson, S. F., Annis, J., Bahcall, N. A., Bernardi, M., Briggs, J. W., Brinkmann, J., Brunner, R., Burles, S., Carey, L., Castander, F. J., Connolly, A. J., Crocker, J. H., Csabai, I., Doi, M., Finkbeiner, D., Friedman, S., Frieman, J. A., Fukugita, M., Gunn, J. E., Hennessy, G. S., Ivezić, Ž., Kent, S., Kunszt, P. Z., Lamb, D. Q., Leger, R. F., Long, D. C., Loveday, J., Lupton, R. H., Meiksin, A., Merelli, A., Munn, J. A., Newberg, H. J., Newcomb, M., Nichol, R. C., Owen, R., Pier, J. R., Pope, A., Rockosi, C. M., Schlegel, D. J., Siegmund, W. A., Smee, S., Snir, Y., Stoughton, C., Stubbs, C., SubbaRao, M., Szalay, A. S., Szokoly, G. P., Tremonti, C., Uomoto, A., Waddell, P., Yanny, B., and Zheng, W. (2001). Composite Quasar Spectra from the Sloan Digital Sky Survey. *Astronomical Journal*, 122:549–564.

Villard, R. (2011). Hubble racks up 10,000 science papers. http://hubblesite.org/newscenter/archive/releases/2011/40/image/a/. [last visited 24/02/2014].

Weiner, B. (2018). DISTCALC - program to calculate cosmological distances. http://mingus.as.arizona.edu/~bjw/software/. [last visited 24/09/2018].

Werner, M. W., Uchida, K. I., Sellgren, K., Marengo, M., Gordon, K. D., Morris, P. W., Houck, J. R., and Stansberry, J. A. (2004). New Infrared Emission Features and Spectral Variations in NGC 7023. *Astrophysical Journal, Supplement*, 154(1):309–314.

Weston, S. D. (2012). Cosmological distance calculator. https://github.com/stuartdweston/DistCalc.

Weston, S. D. (2018). Likelihood ratio in python. https://github.com/sdweston/LikelihoodRatio/tree/master/version-0.5-swire.

Weston, S. D., Seymour, N., Gulyaev, S., Norris, R. P., Banfield, J., Vaccari, M., Hopkins, A. M., and Franzen, T. M. O. (2018). Automated cross-identifying radio to infrared surveys using the LRPY algorithm: a case study. *Mon. Not. R. Astron. Soc.*, 473:4523–4537.

White, R. L. (2019). FIRST survey catalogs. http://http://sundog.stsci.edu/first/catalogs.html.

White, S. V., Franzen, T. M. O. Riseley, C., Wong, I. ., Kapińska, A. D., Hurley-Walker, N., Wu, C., Callingham, J., Thorat, K., Hancock, P., Hunstead, R. W., Seymour, N., Swan, J., Wayth, R., Morgan, J., Chhetri, R., Jackson, C., Weston, S. D., and Mauch, T. (2019). The GLEAM 4-Jy (G4Jy) Sample: I. Definition and host-galaxy identification. In Preparation.

Williams, T., Kelley, C., and many others (2019). Gnuplot 5.0: an interactive plotting program. https://sourceforge.net/projects/gnuplot/files/gnuplot/5.0.7/.

Woodburn, L., Natusch, T., Weston, S., Thomasson, P., Godwin, M., Granet, C., and Gulyaev, S. (2015). Conversion of a New Zealand 30-Metre Telecommunications Antenna into a Radio Telescope. *Proc. Astron. Soc. Aust.*, 32:17.

Xu, C. K., Lonsdale, C. J., Rowan-Robinson, M., Smith, H. E., Shupe, D., Surace, J., and SWIRE Team (2002a). The SIRTF Wide-area InfraRed Extragalactic Survey (SWIRE) — Exploring the dark side of galaxy evolution. In *American Astronomical Society Meeting Abstracts*, volume 34 of *Bull. Am. Astron. Soc.*, page 1191.

Xu, C. K., Lonsdale, C. J., Rowan-Robinson, M., Smith, H. E., Shupe, D., Surace, J., and SWIRE Team (2002b). The SIRTF Wide-area InfraRed Extragalactic Survey (SWIRE) — Exploring the dark side of galaxy evolution. In *American Astronomical Society Meeting Abstracts*, volume 34 of *Bulletin of the American Astronomical Society*, page 1191.

Yuan, F., Lidman, C., Davis, T. M., Childress, M., Abdalla, F. B., Banerji, M., Buckley-Geer, E., Carnero Rosell, A., Carollo, D., Castander, F. J., D'Andrea, C. B., Diehl, H. T., Cunha, C. E., Foley, R. J., Frieman, J., Glazebrook, K., Gschwend, J., Hinton, S., Jouvel, S., Kessler, R., Kim, A. G., King, A. L., Kuehn, K., Kuhlmann, S., Lewis, G. F., Lin, H., Martini, P., McMahon, R. G., Mould, J., Nichol, R. C., Norris, R. P., O'Neill, C. R., Ostrovski, F., Papadopoulos, A., Parkinson, D., Reed, S., Romer, A. K., Rooney, P. J., Rozo, E., Rykoff, E. S., Sako, M., Scalzo, R., Schmidt, B. P., Scolnic, D., Seymour, N., Sharp, R., Sobreira, F., Sullivan, M., Thomas, R. C., Tucker, D., Uddin, S. A., Wechsler, R. H., Wester, W., Wilcox, H., Zhang, B., Abbott, T., Allam, S., Bauer, A. H., Benoit-Lévy, A., Bertin, E., Brooks, D., Burke, D. L., Carrasco Kind, M., Covarrubias, R., Crocce, M., da Costa, L. N., DePoy, D. L., Desai, S., Doel, P., Eifler, T. F., Evrard, A. E., Fausti Neto, A., Flaugher, B., Fosalba, P., Gaztanaga, E., Gerdes, D., Gruen, D., Gruendl, R. A., Honscheid, K., James, D., Kuropatkin, N., Lahav, O., Li, T. S., Maia, M. A. G., Makler, M., Marshall, J., Miller, C. J., Miquel, R., Ogando, R., Plazas, A. A., Roodman, A., Sanchez, E., Scarpine, V., Schubnell, M., Sevilla-Noarbe, I., Smith, R. C., Soares-Santos, M., Suchyta, E., Swanson, M. E. C., Tarle, G., Thaler, J., and Walker, A. R. (2015). OzDES multifibre spectroscopy for the Dark Energy Survey: first-year operation and results. *Mon. Not. R. Astron. Soc.*, 452(3):3047–3063.

Sunset at the Warkworth 30m (Station code : Wa) with the moon. Credit: Stuart Weston