University of Cape Town

Department of Astronomy

Characterising radio sources in the Small Magellanic Cloud using citizen science and optical spectroscopy

---

Dissertation presented for the degree of Master of Science in the Department of Astronomy

Student:
Gideon de Beer

Supervisor:
Dr. Vanessa McBride

Date: June 10, 2025

Plagiarism declaration

I, Gideon de Beer, know the meaning of plagiarism and declare that all of the work in the document, save for that which is properly acknowledged, is my own.

Abstract

For several years, the Small Magellanic Cloud (SMC) has been extensively studied in multiple wavelengths, revealing rich populations of radio sources within and beyond the nearby dwarf galaxy. With radio interferometers' ever-increasing sensitivity and resolution, greater source populations will be discovered. Two regions of the SMC dwarf galaxy were observed using MeerKAT, with 10873 radio sources identified across them. In this dissertation, we characterise the radio source population of these two regions. This was approached using citizen science and optical spectroscopic follow-up. From the 10873 radio sources, 1429 were selected as subjects for classification in the citizen science project. This selection was based on the radio signal-to-noise ratio. The citizen science approach used information from the radio images, combined with optical cutouts from the SuperCOSMOS Sky Survey. The results show a high level of agreement among volunteer classifiers on questions concerning basic morphological characteristics. Around 63 per cent of all classifications show complete agreement between at least 5 classifiers in determining whether the displayed source was a radio point or an extended source and whether there is a likely optical counterpart aligned with the radio emission for the 1429 subjects. 41 radio sources were selected from the original 10873 for spectroscopy, by cross-matching the radio source list to the GAIA optical catalogue and making a magnitude cut. Optical spectroscopic follow-up of 23 sources from these 41 was conducted in November 2023. The reduced spectra from the observations were analysed to identify the optical sources, determine whether the optical and radio emissions were coincident, and identify what mechanism might be causing the radio emission if the optical source is considered likely to be the optical counterpart to the radio source. Characterisation revealed 13 stars, 3 eclipsing binaries and 2 nebulae. Between the two analysis methods, with their own independently selected source sample pools, 6 sources were found to overlap. The resultant analysis between the methods gave complementary source characterisation results. Approximately 912 radio sources have been characterised in total, at least by radio morphology and the likely presence of an optical counterpart, during this project.

Acknowledgements

I want to thank:

- The SAAO for hosting me during my MSc studies, to everyone there who gave me motivation and inspiration during colloquiums or coffee breaks.
- The Sutherland staff for all their support during observation periods in 2023-2024.
- My supervisor, Dr. Vanessa McBride, for her guidance and support during this endeavour. 2
- The University of Cape Town staff & students for their invites to events and learning opportunities. I wish to thank all those who contributed to my project during the Zooniverse data-gathering process too.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including from the National Science Foundation, NASA, the Institute of Museum and Library Services, UKRI, a Global Impact Award from Google, and the Alfred P. Sloan Foundation.

The Zooniverse classification output was stripped of any identifiers that could link any classification to who made that classification. No data that could personally identify any volunteers were stored or used and all volunteers participated in this project knowing that they would be participating anonymously. Registration to the Zooniverse platform is voluntary and was not a requirement to be able to classify on the Zooniverse project created for this MSc dissertation work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and background

This MSc research project aims to characterise radio sources in the direction of the Small Magellanic Cloud, especially those with extended morphology such as supernova remnants, super-bubbles, planetary nebulae, AGN and background galaxies. This characterisation was done through the use of citizen science and optical spectroscopic follow-up of potential optical counterparts.

A recent survey of the Small Magellanic Cloud (SMC) was carried out using ASKAP, a precursor of SKA, and revealed over 8000 sources within and behind the SMC (Joseph et al. 2019). In May 2019, the SMC was observed using MeerKAT in the L-band with a central frequency of 1284 MHz, producing 2 deep radio images observed for 8 hours each. These MeerKAT images (see Figure 1.1) are centred around young high-mass X-ray binaries within their natal supernova remnants (SNR). With these MeerKAT observations having a sensitivity of $<$20 μJy/Beam[1], a wide field of view (3° by 3°), and a resolution $\sim$4 times greater than that of previous ASKAP observations (6.5 by 5.9 arcsec beam-size), it was expected that a wealth of radio point and extended sources would be uncovered.

This chapter reviews the literature and background information concerning the Small Magellanic Cloud, the relevant classes of radio sources, and the methods of characterising astronomical sources, alongside relevant details about citizen science and multi-wavelength astronomy as it pertains to this dissertation. Chapter 2 contains details on the citizen science platform and how it was used to gather data through classifications, as well as the methods used to analyse the collected data into results. Chapter 3 expands on the follow-up observation of select optical sources very close ($<$1 arcsec) to radio source locations; these were treated as potential optical counterparts to the radio sources. During analysis, these possibilities were further explored through spectral analysis. Chapter 4 provides a summary discussion of the results from Chapters 2 and 3, a section on potential future work, and the conclusions section.

1.1 The MeerKAT telescope

The MeerKAT telescope array (see Figure 1.2) is a precursor for the Square Kilometre Array (SKA) interferometer project (Dewdney et al. 2009). MeerKAT is located near Carnarvon in the Karoo region of the Northern Cape Province in South Africa. There were 2 prototype telescopes built before the MeerKAT interferometer was commissioned, and construction had started. These two are the single-dish 15-m XDM built at HartRAO (Hartebeesthoek Radio Astronomy Observatory), and the 7-dish interferometer called the KAT-7 array in Carnarvon. The Karoo is arid and surrounded by mountainous regions far from telecommunication towers, which naturally shields the site from interference caused by earth-bound communication technologies. Further shielding from Electromagnetic Interference (EMI) is provided to the power grid, electronics and other infrastructure by having most of the buildings or rooms housing sensitive equipment built below

[1] See the official MeerKAT specifications here.

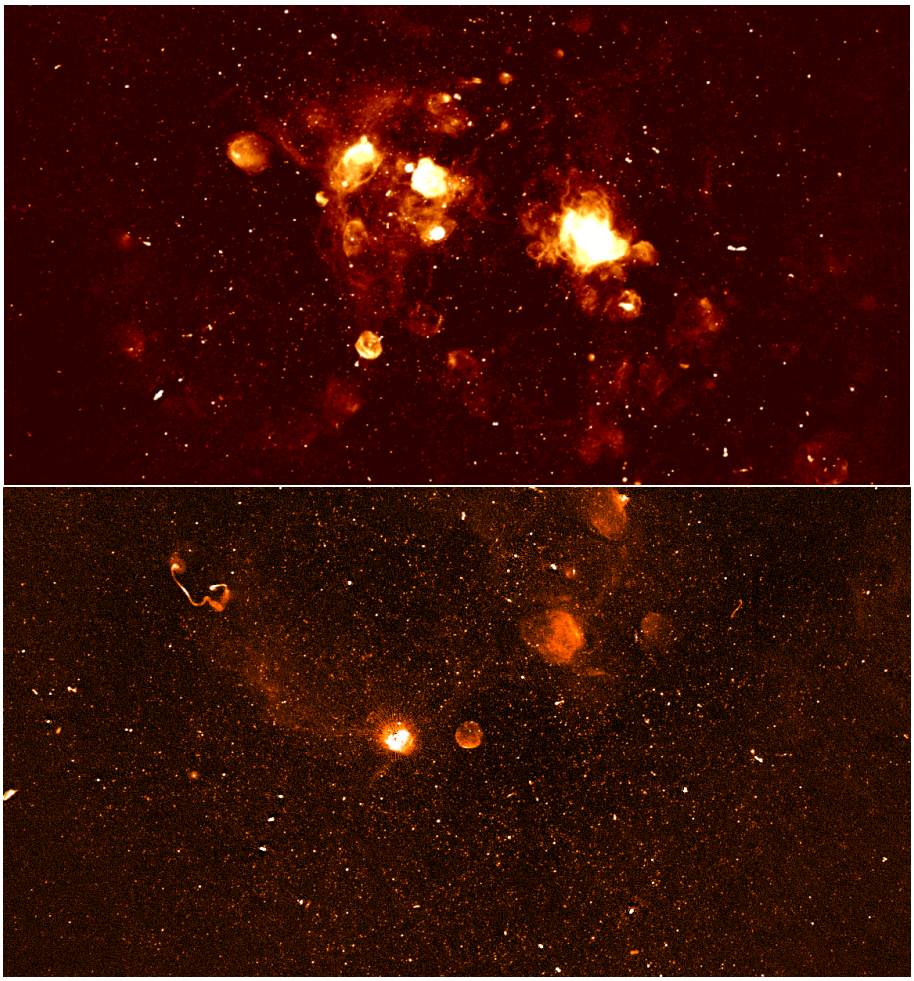


Figure 1.1: MeerKAT image cut-outs (oriented North-up, East-left with angular sizes 1.6° by 0.9°) of the 2 fields centred around high-mass X-ray binaries, SXP1323 (top, Carpano et al. (2017), central coordinates RA (J2000) = $01^h03^m37^s$, Dec (J2000) = -72°01′33″) and SXP1062 (bottom, Haberl et al. (2011), central coordinates RA (J2000) = $01^h27^m46^s$, Dec (J2000) = -73°32′56″). The images are from MeerKAT observations, and data reduction was done by Ian Heywood. It illustrates the different types of radio emissions from this section of the SMC, including point sources, bubbles, background galaxies and pockets of gas. Image credit to Ian Heywood.

the natural ground surface. Many redundancies were built into the power and data lines to prevent any damage to the equipment, dishes and science data. MeerKAT has 64 array elements that cover the frequency range of 0.58 - 3.5 GHz between the UHF (560-1060 MHz), L (881-1670 MHz), and S (1750-3500 MHz) bands for science observations. The spread of these dishes means that the longest baseline for the MeerKAT interferometer is around 8 km (Jonas & the MeerKAT Team 2018; Heywood et al. 2022). In the context of observations in the SMC, the long baselines and a large number of dishes facilitate the identification and study of various radio sources within and beyond the nearby dwarf galaxy, including stars, nebulae, supernova remnants, and background radio galaxies. This has been demonstrated in Cotton et al. (2024) where observations of the SMC in the L-band produced continuum data with 8 arcsec resolution and an RMS noise level of $\sim$11 μJy/beam.

A recent project called the MeerKAT Galaxy Cluster Legacy Survey (MGCLS) looked at 115 galaxy clusters and published its first legacy product data release (DR1) in Knowles et al. (2022). This project also produced a compact source catalogue of $\sim$626000 sources found in the fields of view of the science observa-

tions. This shows what can be achieved when combining large fields of view with the enhanced sensitivity of MeerKAT. Observations for MGCLS were done with MeerKAT (while always utilising at least 59 of the 64 antennas per observation), the task required approximately a year to complete after starting in June 2018. This MGCLS DR1 is a testament to the power of MeerKAT and has stoked excitement in the radio astronomy community to see what the SKA will be capable of.

Figure 1.2: An aerial image of the MeerKAT radio telescope array near Carnarvon. Image credit to South African Radio Astronomy Observatory (SARAO).

1.2 The Small Magellanic Cloud dwarf galaxy

The SMC is a satellite galaxy around 62 kpc away (Graczyk et al. 2013) inside our local group. Over the years, it has become very well studied at multiple wavelengths. A significant portion of sources detected in radio wavelengths are not inside the SMC but are sources in the background or foreground. Many background sources also show extended emission and interesting morphology in the radio images from MeerKAT (Joseph et al. 2019).

In the article by Graczyk et al. (2013), the distance to the SMC was calculated from the observed mean distance modulus of 18.965 ± 0.025 (statistical) ± 0.048 (systematic) in units of magnitude. This value was refined over years of observation of late-type eclipsing binary stars within both Magellanic clouds. From the distance modulus, the corresponding mean distance to the SMC was calculated as 62.1 ± 1.9 kpc. The distance to the SMC has made it very popular for researching supernova remnants (SNR) and other sources showing extended radio emission since the extended radio emission from these kinds of sources is often spatially resolvable in the SMC and the Large Magellanic Cloud (LMC) when using modern telescopes and instruments (Maggi et al. 2019). Most sources within the SMC are also well suited for population and evolution studies due to their relatively close proximity to our galaxy. This is because galaxies with higher neutral hydrogen (H I) mass, or low metallicity as with the SMC, also tend to exhibit higher star formation rates and a lower surface density of stellar mass (Zhou et al. 2018). The SMC also attracts attention to those studying galaxy dynamics and stellar population evolution due to the relatively lower metallicity in its interstellar medium compared to our solar neighbourhood (around 5 times lower). This, in combination with the low amounts of gas & dust between us and the SMC, offers a diverse collection of sources found in the same region of the sky that the SMC occupies (see Cotton et al. 2024 and the map of H I column density $\frac{HI atoms}{cm^2}$ in Appendix C taken with the Australia Telescope Compact Array instrument).

The Sydney University Molonglo Sky Survey (SUMMS) used the Molonglo Observatory Synthesis Telescope (MOST) at 843 MHz to create a radio continuum survey of the southern sky. This survey has a reported RMS noise of 1.27 to 1.9 mJy/beam depending on declination and a resolution ∼45 arcsec (Mauch

et al. 2003). In the paper by Joseph et al. (2019), the RMS noise in the ASKAP continuum data of the SMC using 12 of the dishes available is reported to be between 186 (960 MHz) and 165 μJy/beam (1320 MHz) with a resolution between $\sim$30 (960 MHz) and $\sim$15 arcsec (1320 MHz). Comparing these two projects by their RMS and resolution, it is clear that for radio source population studies focused on certain locations (like the SMC), it is preferable to have the higher sensitivity (or lower RMS) and resolution offered by interferometers. This is because interferometers minimise RMS noise and have greater resolving power when compared to surveys like SUMSS, which used the Molonglo Observatory Synthesis Telescope (MOST). In Wong et al. (2011), the total number of radio point sources detected using MOST telescope data in the SMC is 1689. In the ASKAP EMU survey of the SMC, there are between 4489 and 5954 total radio point sources reported (Joseph et al. 2019). In the MeerKAT data used in this work, there are 10873 radio point sources detected by the source-finding software across the two sections of the SMC that were observed. The sensitivity and resolution offered by MeerKAT make this dataset ideal for population studies, as this will aid efforts to constrain source populations within the field of the SMC. Studying source populations in the SMC is also important for comparative studies, where insights gained from this relatively close-by dwarf galaxy can aid other studies concerning more distant galaxies. These insights can also be used to better understand galactic sources that we may not be able to study so directly due to things like dust obscuration.

1.3 Radio emission & source descriptions

The varying classification conventions for radio sources with extended morphology point to underlying physical mechanisms responsible for the observed structures. This section explores radio emission mechanisms and how they can be connected to observed morphology to aid in source classification. Earlier radio surveys employed instruments that lacked the sensitivity and resolving power of modern radio interferometers. This is why the larger-scale morphology components were studied and characterised first. As sensitivity increased, more pieces of the morphology puzzle could be added (Fraknoi et al. 2022). The FIRST and NVSS radio surveys, produced using the Very Large Array interferometer, provided large catalogues of multiple extended source types over large areas of the sky. Numerous publications have used these surveys to pinpoint the sky coordinates of unidentified sources that show extended radio emissions. Many of these would later be identified and categorised using various methods, see Urry & Padovani (1995); Amirkhanyan et al. (2015):

- Stars and star systems all across the HR diagram have been found to radiate in the radio (Güdel 2002).
- Some active galactic nuclei (AGNs) show extended radio emission like jets and lobes (Miraghaei & Best 2017).
- Supernova explosions frequently spawn an outward expanding remnant (supernova remnant, SNR) or bubble, usually visible in all wavebands from radio to X-ray, with the initial event temporarily outshining the host galaxy (Maggi et al. 2019; Fraknoi et al. 2022).
- Pulsar wind nebulae (PWNe) are formed through a mechanism similar to that of the jets of AGN in that particles are accelerated to tremendous speeds through rotating magnetic fields around compact objects. Powerful magnetic fields from these kinds of pulsars expel winds of particles at velocities close to the speed of light (Gaensler et al. 2003). These crash into the surrounding medium, creating PWNe that shine in radio and other radiation bands.
- Planetary nebulae are formed from the gas and dust layers cast off by dying stars during their last few end-of-life processes. Although these planetary nebulae are not visible/resolvable in distant galaxies, they have been spotted in the Magellanic clouds and as Galactic sources (Westerlund & Smith 1964; Asher et al. 2024).

1.3.1 Radio emission mechanisms

The non-thermal radio emission is primarily associated with shock waves or intense magnetic field activity, where particles are accelerated to high velocities. In contrast, thermal radio emission is produced by objects with large and very hot emission surfaces. Bremsstrahlung, or the free-free emission mechanism, originates from hot ionised thermal plasma and is detectable across the electromagnetic spectrum. This is the mech-

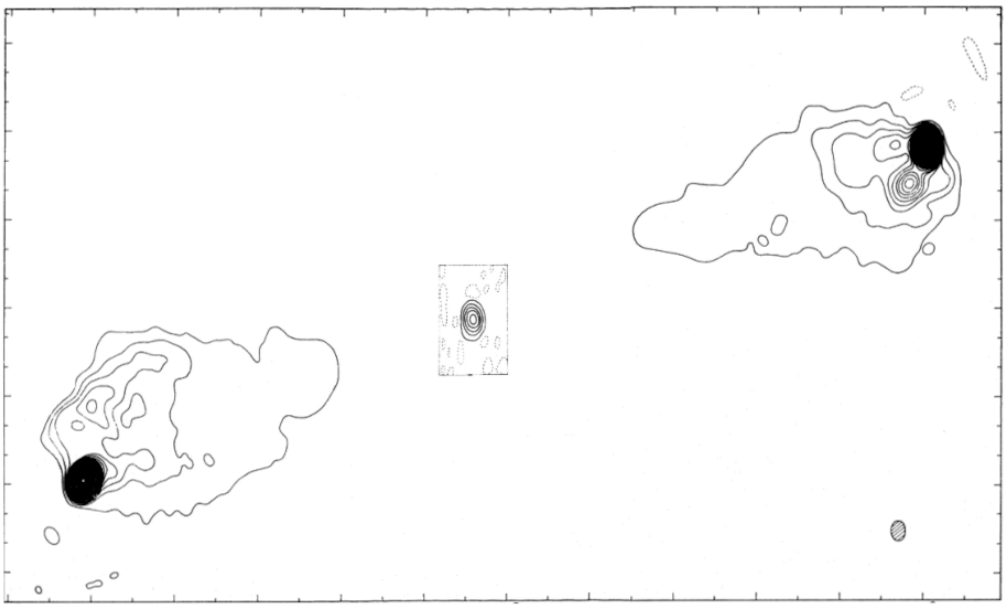


Figure 1.3: Early radio intensity map from an observation of Cygnus A (3C 405), observed with the 5km telescope (Hargrave & Ryle 1974), showing the extended radio galaxy lobe morphology of this radio galaxy. Credit: Hargrave & Ryle (1974).

anism by which free electrons are scattered off ions without being captured. Gyromagnetic (non-thermal mechanism) emission occurs when electrons are caught in strong magnetic fields (Güdel 2002). Depending on the Lorentz factor $\gamma = \frac{1}{\sqrt{1-v^2/c^2}}$ (which is a measure of how fast a particle is moving, v, as compared to the speed of light, c), this includes cyclotron emission from non-relativistic or thermal electrons ($\gamma \sim 1$), gyro-synchrotron emission from mildly relativistic electrons ($\gamma \sim 2$ *to* 3), and synchrotron emission from fully relativistic electrons ($\gamma >> 1$).

1.3.2 Stellar sources

Radio observations of stellar sources serve as an effective diagnostic tool for astronomers, enabling them to investigate and explain certain phenomena like outbursts or interstellar medium (ISM) interactions. Our sun is a source of both thermal and non-thermal radio emission. There are also radio emissions detected from both cool and hot stars (spanning from pre-main-sequence to post-main-sequence age stars), cataclysmic variables, and ultra-cool dwarfs among others (Matthews 2019). Stellar coronae and winds can become sufficiently ionised by hot stars to produce radio emissions through the bremsstrahlung mechanism. Bremsstrahlung emission can also arise from stellar winds as massive stars shed their outer layers, creating a shock wave that interacts with the ISM. Stars have magnetic fields in and around them (Stewart 1989), and they are observed to have varying levels of chromospheric activity leading to non-thermal radio emission. Magnetic fields with opposing directions, or polarities, can collide and destroy one another to create coronal mass ejections (CMEs), detected in radio observations as a variability in the star's luminosity over time. In the case of our star, it is important to be able to predict when a CME will occur and if our planet will be in the path of those particle waves. In binary star systems, the winds generated by the stars can collide with each other, creating regions of gyromagnetic and free-free emission at the impact zones between the stars. Studying these emissions gives insights into stellar evolution, and the influence these star systems have on their surrounding environments (Dulk 1985; Güdel 2002; Matthews 2013).

Radio emission has been detected from stars classified across the parameter space of Hertzsprung-Russell diagrams (HR diagrams, see Figure 1.4). The hot and massive super giant to the giant region is mostly occupied by Wolf-Rayet (WR) stars, along with the very luminous wind-shedding stars and some binary

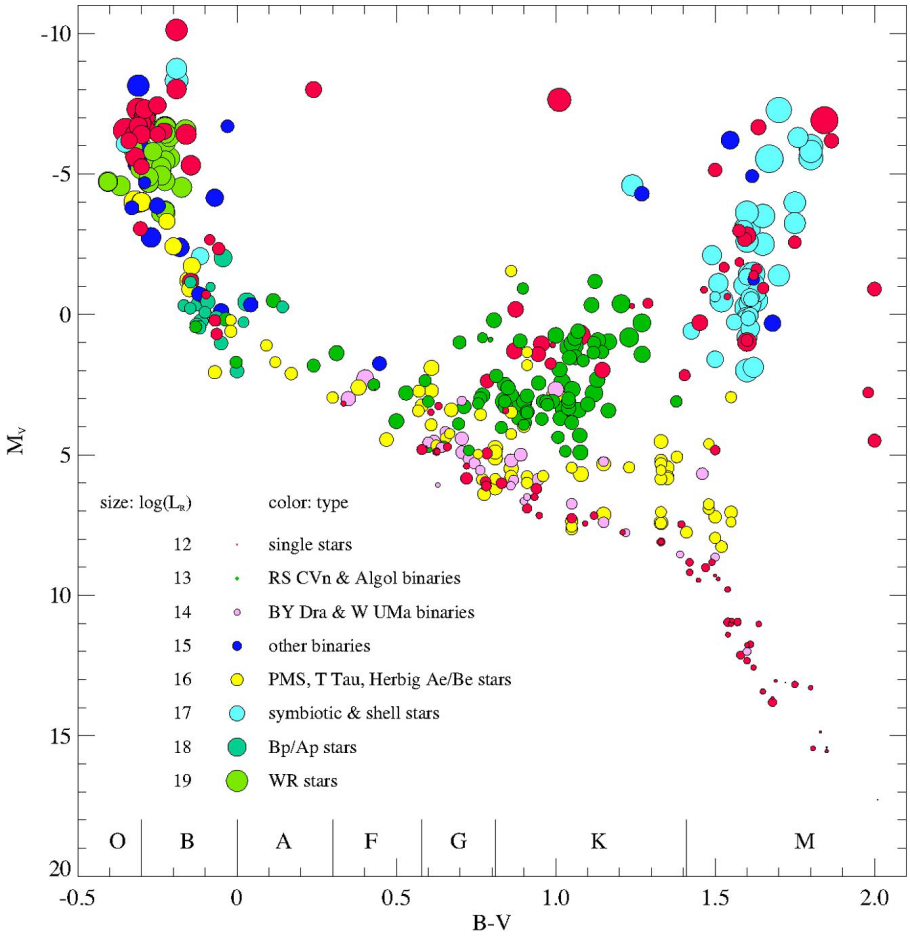


Figure 1.4: HR diagram showing the distribution of stars that are detected in radio. This indicates there are radio-emitting stars across the main-sequence and the pre-main-sequence regions. Image credit: Güdel (2002).

systems. These star systems produce highly ionised optically thick thermal emissions. A large percentage (more than 50 per cent) of massive stars are formed in binary or multiple star systems (Palate et al. 2013). The HR diagram's cooler middle section of sub-giants and giants is dominated by RS CVn and Algol-type star systems, along with F, G, and K-type stars. These systems have a mix of thermal and non-thermal emission mechanisms depending on factors like surface temperature, being in a multi-star system, or the presence and behaviour of an accretion disc. The M-dwarfs dominate the coolest section of the HR diagram, where the cooler stars, symbiotic stars, and some binaries are also found. This was most surprising when it was discovered since cool dwarfs were thought to have no strong magnetic fields due to low amounts or the absence of magnetic dynamo action in stars with mostly convective cores. The stars in this part of the HR diagram emit mostly through non-thermal mechanisms with a range of luminosity variabilities depending on the type of star system. Radio studies of these stars/star systems contribute to the accuracy of evolutionary models as well as tests for theoretical models concerning physical processes (Güdel 2002; Matthews 2019).

1.3.3 AGN

Astronomers noticed the recurring, symmetric, bipolar structure of radio galaxy lobes in early maps of radio emission (see Figure1.3), which suggested that the material emitting in radio must have originated from the centres of the radio galaxies between the lobes. Many early models tried to explain how these lobes were

forming and being sustained, but those models could not account for the shape or continuous activity of AGN jets. The idea of an AGN linked to a radio galaxy displaying radio jets of synchrotron radiation was later accepted after the investigation of a model where the continuous supply of energy to these these radio sources is facilitated through accretion of mass onto a black hole to keep those relativistic jets "switched on" to explain the mechanics, size and shape of radio lobes (De Young 2002).

In terms of classification, Fanaroff & Riley introduced a morphology-based classification system for radio galaxies in 1974, separating them into two primary categories (Miraghaei & Best 2017):

- FR type I are observed to exhibit a jet structure that becomes increasingly diffuse as they extend further from the central black hole. They generally have lower radio luminosities when compared to the type II galaxies and are more often found in dense environments like galaxy clusters, contributing to disruption in their jet structure.

- FR type II radio galaxies have a tightly collimated jet structure extending from the AGN core and ends in bright, lobed hotspots. They are typically more luminous than the FR type I galaxies and are often found in less dense environments, allowing their jets to extend across vast distances without significant disruption.

1.3.3.1 Blazars/Quasars as sub-classes of AGN

Blazars and Quasars are two closely related subclasses of AGN. Blazars are a radio-loud subclass (Giroletti et al. 2007; Dogra et al. 2025) having the AGN jet pointed directly at the observer, making them extraordinarily bright, with much variability observed across all wavelengths (Jovanovic et al. 2023). Quasars, being a mostly radio-quiet subclass of AGN, have jets pointed in the direction of the observer but not directly at them. Both exhibit extremely energetic emissions at multiple wavelengths. As sub-classes of AGN, their emissions are powered by the accretion of matter onto supermassive black holes at the centres of their host galaxies. These types of AGN are usually identified by their X-ray, optical, infrared and radio signatures along with red-shift measurements (Urry & Padovani 1995).

Blazars have characteristic non-thermal radiation as the major component of all their radio emissions. Electrons and other subatomic particles are flung into the jets and spiral along the strong magnetic field lines, radiating as synchrotron radiation (Dogra et al. 2025). The study of blazars provides astronomers with knowledge of what the conditions are like near supermassive black holes, in addition to shedding light on the behaviour and physics of relativistic jets produced from the accretion of matter onto the central black hole (Urry & Padovani 1995).

Quasars are named after the phrase "quasi-stellar radio sources" since the cores of quasars were visually mistaken for stars in earlier (predominantly optical) astronomy. When these objects that looked like stars in optical photographs were observed, their spectroscopic signatures were unlike those observed in stars. Particularly, they found very broad and strong emission lines at wavelengths that were not yet identified. With the advent of radio astronomy and more advanced observation techniques, astronomers noticed that in the cores of some of these "stars" there are compact radio sources and these are often in the centre of double-lobed extended radio emission (which became known as the jet and/or lobe structure of AGN, Peterson 1997). Only around 10 per cent of quasars are considered to be radio-loud based on the ratio between their radio and optical flux ($R = f_{5\,GHz}/f_{4400\,\text{Å}} > 10$ to be considered "radio loud", Chiaberge & Marconi 2011). The source of their radio emission could come from multiple mechanisms like star formation of the host galaxy, hot accretion-disc coronae and low-power jets that are very compact (Liao et al. 2024). Studies on quasars have improved our understanding of how galaxies evolved from the early universe. Their distance from us and their luminosity in some wavelengths make them very valuable for research into the history and cosmology of our universe (Peterson 1997; Jovanovic et al. 2023).

1.3.4 SNR and large-scale Nebulae

A SNR is what remains after the death and core collapse of massive stars followed by a supernova (SN) explosion (see Figures 1.5, and 1.6). The death of a high-mass star (initial mass of $\gtrsim$8 solar masses or $M_\odot$)

leaves behind a dense core (usually a neutron star if the star's initial mass was $\gtrsim$8 $M_\odot$ and $\lesssim$40 $M_\odot$ or a black hole if the initial mass was $\gtrsim$40 $M_\odot$) and extended emission. The SN creates a shock wave which expands outward and collides with the surrounding medium, accelerating particles through magnetic fields and resulting in synchrotron radiation detected by radio observations (Dubner & Giacani 2015). They form large-scale extended shells or bubbles that influence the surrounding medium (Maggi et al. 2019 and references therein).

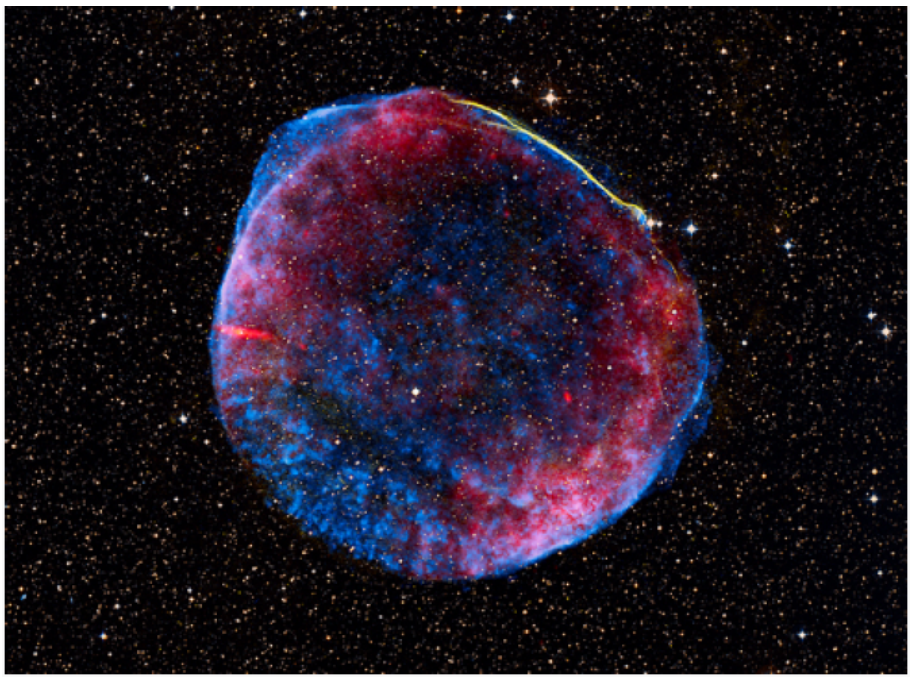


Figure 1.5: SNR of SN1006 with radiation in X-ray in blue, optical in yellow and radio in red. This image shows how the different components of the source morphology from different observation bands paint a more complete picture when combined. Credit: Fraknoi et al. (2022) and sources therein.

This emission is studied in multiple wavelength regimes, and each reveals a different side of a SN's evolution. After a time (on the scale of thousands of years), they form a nebula close to the core. In a recent paper, for example, extended gamma-ray emission was studied around the supernova DA 530, and this high-energy emission extended beyond the extended radio emission shell (Xin & Guo 2023). The SNR SXP1062 (see Figure 1.7) is centred on one of the MeerKAT images used for analysis during this project. The SNR SXP1062 was recently observed with the Southern African Large Telescope (SALT), and its shell expansion velocity was measured to be around $\sim$140 km/s. This result was obtained by analysing high-resolution long-slit spectra obtained from the SALT telescope in Sutherland (Gvaramadze et al. 2021). Results like these tell us about these events' physical evolution and their effect on their surroundings.

SNe can occur when white dwarf stars accrete enough mass from a larger donor star in a binary system; this mass limit is called the Chandrasekhar mass (this limit is also often reached during merger events where 2 stars of white dwarf mass or more collide and merge). Electron degeneracy pressure resists the further shrinking of the core of a white dwarf star; however, their outer shells are still undergoing fusion, and the fusion products sink onto the core, making it more dense. A supernova explosion inevitably follows when the electron degeneracy pressure is overcome, followed by core collapse. This leaves a neutron star, or a black hole if the stellar core mass exceeds 3 solar masses, and a supernova remnant in its wake (Dubner & Giacani 2015; Fraknoi et al. 2022). This type of SN is called a Type Ia SN caused by a runaway thermonuclear reaction in a white dwarf stellar core, and they are identified by the shell structure and presence of silicon lines, along with the absence of hydrogen lines, in spectroscopic observations. There are also the core-collapse SNe of stars that burn up all their fusion fuel, given they have an initial mass greater than 8 solar masses. Type Ib SNe have no hydrogen or silicon in their spectra, and they are helium-rich. Type Ic SNe also have

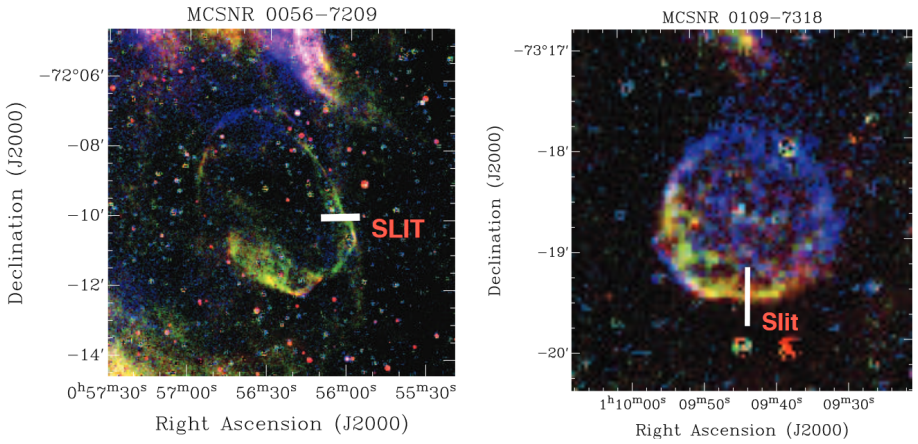


Figure 1.6: Above are images showcasing the shell morphology that ∼80 per cent SNRs have been observed to evolve into (Dubner & Giacani 2015). Magellanic Cloud supernova remnant candidates J0056-7209 left and J0109-7318 on the right, shown as composite images (Red=[SII], Green=H α, Blue=[OIII], respectively) from the Magellanic Cloud Emission Line Survey (MCELS). Credit: (Maggi et al., 2019, Figure 1).

no hydrogen or silicon in their spectra, but they are helium poor. Type II SNe are identified by the signature presence of hydrogen lines in their spectra (Dubner & Giacani 2015). The extended emission from a SNR is visible from radio to X-ray frequencies. When a SN type is identified, we can deduce what type of star it may have been before it underwent core-collapse (Maggi et al. 2019).

1.3.5 Pulsar wind nebulae

Galactic pulsar wind nebulae (PWNe) are identified visually (in multi-wavelength continuum images, but mostly X-ray as seen in Figure 1.8) as compact bubbles. They are formed when the relativistic particles from a pulsar jet shock the surrounding medium after being accelerated through the strong magnetic field of a fast-rotating neutron star (pulsar). PWNe are not likely to be resolved in the SMC as they immediately surround neutron stars; most often, they will be seen in continuum data at the cores of SNRs. Studying the properties of PWNe can provide inferred estimates on the initial spin period and the strength of the magnetic field produced by the pulsars that produce the PWNe. This is especially useful when dealing with pulsars in the Magellanic clouds, as these pulsars are often unresolved or can not be directly observed (Gaensler et al. 2003; Haberl et al. 2012). A famous and well-studied example is the Crab Nebula. See Figure 1.8 which contains a "young pulsar"[2] of around 960 years old which formed in a supernova explosion believed to have been observed and documented by early Chinese cultures in the year 1054 (Fraknoi et al. 2022).

1.3.6 Planetary Nebulae

Planetary nebulae (PNe) are formed around old red-giant stars when they shed their outer layers before contracting and starting the final phase of their lives as white dwarf stars in a nova event. They are most often seen in continuum data as highly symmetrical envelopes or shells composed of ionised gaseous matter that shine for thousands of years while the expelled matter diffuses into the surrounding interstellar medium. These objects are thought to start as non-thermal radio sources in the pre- to early nova phase, then there is a transition to a thermal radio source when electrons in the ionised gas interact with ions to produce radio waves through the bremsstrahlung mechanism. In the SMC, radio continuum studies have revealed on the order of $\tilde{3}5$ confirmed planetary nebulae, with more candidates awaiting further study. Studies of PNe have provided insights into the mass loss mechanism of stars and their influence on the ISM. Over large time scales, they also influence star formation dynamics and gas evolution of galaxies (Asher et al. 2024 and sources within).

[2] "Calculations suggest that the typical lifetime of a pulsar is about 10 million years; after that, the neutron star no longer rotates fast enough to produce significant beams of particles and energy, and is no longer observable." excerpt from (Fraknoi et al., 2022, page 795)

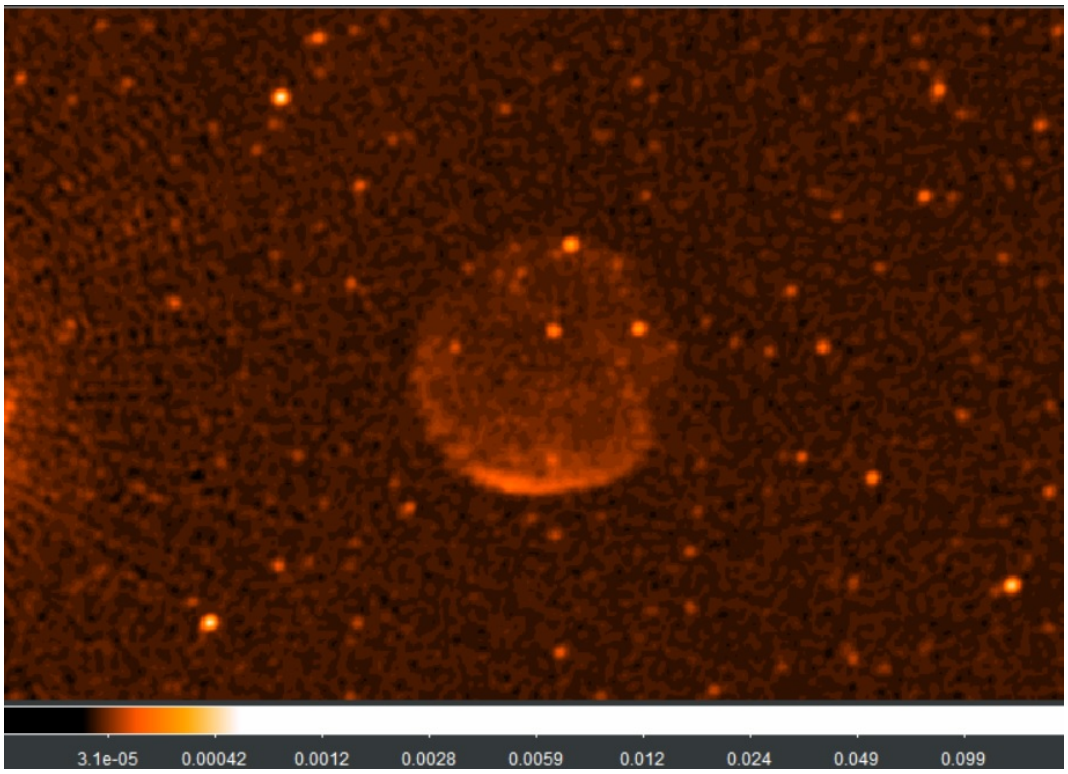


Figure 1.7: SNR SXP1062, imaged with MeerKAT and oriented North-up, East-left. The colour grading is based on radio flux (Jy/beam) and the bubble is ∼2.75 arcmin in diameter, while the rectangular image cutout measures 11.3 by 4.4 arcmin.

A Galactic PN that has become very popular with amateur astronomers and scientists alike is the Ring Nebula due to its proximity and the resolved extended emission structure (see the image on the left in Figure 1.9).

1.4 Methods of identifying astronomical sources and data analysis

The methods of observing celestial objects and analysing what has been observed have evolved over many years. From ancient times when cultures etched stellar constellations into stones, to the Greeks and Romans who laid the foundation for modern astronomy, physics and mathematics by using written records of observations and experiments. Early astronomers kept records of what they called "guest stars" which were too faint to be observed using the naked eye but would flare up and be visible for varying periods. Presently, researchers have access to more advanced techniques and a much broader opportunity pool for collaboration with other researchers across the planet (Fraknoi et al. 2022).

Methods of identifying celestial sources have been predominantly visual (from inspecting morphology) or behavioural (by analysing data gathered using instruments on different types of telescopes), where celestial objects are named and categorised based on their appearance or specific behaviours and characteristics of their emissions. The tools we use for this purpose have progressed from the naked eye before technological advances to early small telescopes and binoculars, and to the modern telescopes we have in this age, on the ground and in space. Many sources require much more than just a visual categorisation from continuum image analysis to fit into a classification category or form part of a catalogue. Spectroscopy is one method of identifying sources that relies on the "fingerprint" of light that is emitted from an astronomical source (Fraknoi et al. 2022). AI algorithms are gaining in popularity as they become better at correctly categorising huge amounts of sources that would take astronomers years to sort through manually (Tang et al. 2022).

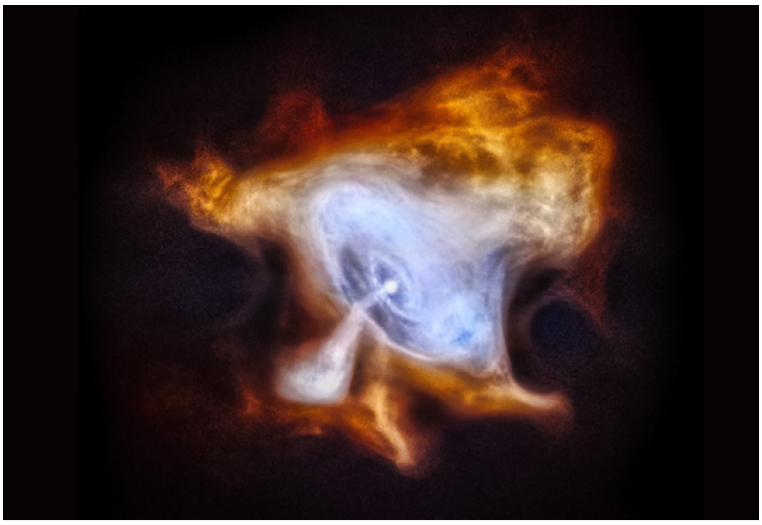


Figure 1.8: X-ray emissions showing the pulsar jet morphology of the PWN in the centre of the Crab Nebula. credit: modification of work by NASA/CXC/SAO in Fraknoi et al. (2022).

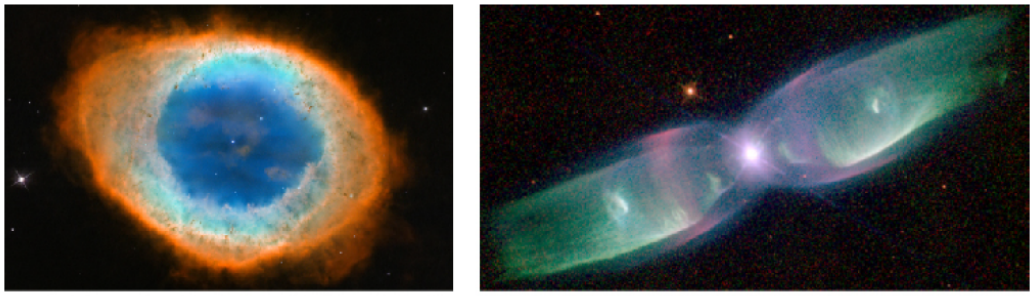


Figure 1.9: Images of PN a)"Ring nebula" and b)"M2-9" taken with the Hubble Space Telescope. In these images, the symmetry in their observed morphology can be seen with the Ring nebula having a shell morphology, and M2-9 having a bi-polar jet morphology. credit: Fraknoi et al. (2022), NASA and the Hubble Heritage Collaboration.

Software packages like AEGEAN (Hancock et al. 2012) and pybdsf[3] allow astronomers to quickly build a catalogue of point sources from continuum images, which can later be further analysed to see if certain sources are real or noise that resembles a source (Hancock et al. 2012, 2018). Citizen science is a relatively new approach that allows almost anyone to contribute to science and helps scientists who need to analyse large amounts of data. Zooniverse is a popular platform for citizen science projects and already has many very successful astronomy-based projects that are busy collecting data or have been completed entirely (Simpson et al. 2012; Heywood et al. 2022; Tang et al. 2022).

1.4.1 Source-finding software packages (AEGEAN)

Source-finding algorithms work by finding islands of pixels in an image that are brighter than the pixels surrounding them, tuned to some threshold. These sources are added to a catalogue along with scientific properties extracted from the image data. An ideal source finder should thus be able to identify all the sources in the images provided to it (it should be complete), while also detecting and extracting genuine sources and avoiding mistakenly identifying noise for a source (so it also needs to be reliably parametrised and set up

[3] https://pybdsf.readthedocs.io/en/latest/

to be accurate). Recently developed source-finding software provides highly complete and reliable results, with only a few sources being missed or incorrectly chosen as a source. This fraction of incorrectly identified sources is around ∼1 per cent (Hancock et al. 2012), and depends on the source profiles being searched for, as well as the S/N of the data being searched through. Although this may be a small percentage, it becomes significant when considering the data output of modern and upcoming radio surveys from instruments like ASKAP, MeerKAT and the future SKA. These instruments can produce surveys with millions of potential sources, meaning that even a ∼1 per cent error rate in false identifications can pose a substantial risk to the trustworthiness of ∼99 per cent of sources identified as detections on that scale when source finding becomes fully automated. The development of AEGEAN focused on correctly identifying this ∼1 per cent of missed or falsely flagged sources where other algorithms would have missed or misidentified them (Hancock et al. 2012, 2018).

The other issue is the data rate of modern radio telescopes, which is so large that there will be almost no chance of reprocessing the data if some real sources, especially transients, are missed. This is because of the time it takes to reduce terabytes of raw data outputs. Sources like transients that are not picked up the first time might never be detected again. AEGEAN was built to improve previous source-finding software packages, including *SExtractor*[4], *IMSAD*[5] and *Salavy*[6]. AEGEAN was developed to improve the completeness and reliability offered by its predecessors in preparation for big radio surveys, with a focus on addressing the shortcomings of these predecessors. According to Hancock et al. 2012, the reliability and completeness of catalogues produced from source finding software will be improved if they can perform constrained multiple Gaussian fitting. AEGEAN can do this demonstrably through the constraining of fitting parameters that are based on the curvature of the input images. This gives AEGEAN the ability to find sources that its predecessors are very good at finding, along with sources that are close to the detection limit of an observation, and individual sources that are all part of the same island, which the preceding algorithms were not good at identifying. AEGEAN has seen further evolution to be able to handle science images that are both larger in dimension and more densely packed with sources than ever before, owing to the increase in sensitivity and large fields of view of telescopes like the SKA precursors' data outputs (Hancock et al. 2018).

1.4.2 Cross-matching between surveys

Cross-matching between surveys has become routine for many astronomers with large survey projects providing online platforms where astronomers can input coordinates or source identifiers to cross-match them to large survey lists. Publications of new catalogues are regularly added to archives like CDS-Vizier, often accompanied by scientific measurements for each source, where available. Source identification and the discovery of counterpart sources can be explored through this method. When a source coordinate coincides with emissions from different bands, multi-wavelength analysis of that source may reveal new characteristics. Many useful tools have been developed with the ability to cross-match source locations between publicly available catalogues and surveys during the analysis of astronomical data. One such tool is called TOPCAT[7], which is a simple yet powerful tool for the handling of large tables.

The GAIA mission has made extensive measurements of parallax linked to most stellar sources in the survey. Parallax measurements can be used to calculate the distance between an observer and an object that was observed with a known parallax. The formula used to convert a parallax measurement to distance is

$$d = \frac{1}{p} \tag{1.1}$$

where distance, d, will be in units of parsec if the measured parallax, p, is in arcsec. This is purely geometric, using the diameter of the Earth's orbit around the sun as a baseline for measuring the parallax angle of a celestial object. It depends on the accuracy of the measured parallax; if the source is too distant,

[4] Source Extractor licence and documentation.
[5] IMage Search And Destroy documentation as part of the "Miriad" software package.
[6] Salavy documentation, software used for radio source finding in ASKAP data.
[7] TOPCAT site and documentation

the parallax can not be measured with enough precision and other methods must be used to determine the distance. When parallax can not be measured with precision, the absolute and apparent magnitude of stars can be used to estimate the distance between the observer and a distant star. Distance to a star can be estimated in this way by using the formula

$$M - m = -5log_{10}(d) + 5 \quad \therefore \quad d = 10^{0.2(m-M+5)} \tag{1.2}$$

where M is the absolute magnitude of a star[8], m is the apparent magnitude as observed from Earth, and d is the distance to the star in units of parsec.

In 2024, a MeerKAT 1.3 GHz SMC survey paper by Cotton et al. was published that covered a field of view of ∼7° x 7°. The catalogue released with the paper contains 108330 sources categorised as point sources and 517 that are categorised as compact extended sources. The SMC has been extensively studied in radio wavelengths over the past 50 years. Discoveries and confirmations of candidate sources are still being made with surveys like these (through cross-matching and further analysis) that offer a resolution of 8 arcsec and low Root Mean Square (RMS) noise of ∼11 μJy/beam (Cotton et al. 2024). Catalogues like these offer a wealth of exciting scientific opportunities and source discoveries through cross-matching.

1.4.3 Spectroscopy

All light is electromagnetic radiation, bundles of information regarding the nature of the matter or process that emitted (or absorbed) the light. During early experiments, it was found that gases were not completely transparent to all wavelengths of light. They absorb certain wavelengths of light, which would disappear from experimental spectra when compared to the control spectrum of light without any obstacles. This is where the "absorption spectrum" of gases was discovered. Other experiments had the gas as the light source by heating it inside a sealed glass container; the light from each heated gas showed distinct bands of dispersed light, and each gas had a distinct "emission spectrum" (see Figure 1.10). After these findings, spectroscopy became a tool for identifying certain gases from their emission and/or absorption lines in astronomical spectroscopic results (Fraknoi et al. 2022).

During the analysis of a spectrum, care must be taken in the identification of real line features that do not originate from a celestial source. For example, certain emission lines show up because of a street lamp nearby, which was left on during the observation. Atmospheric water vapour, airglow effects, and clouds also produce line features visible in spectra (Jaschek & Jaschek 1987). There are also emission lines as a result of the airglow present in the atmosphere that can be enhanced by observing when the moon is above the horizon or when geomagnetic disturbances occur during observations when solar activity is at a high level (Smith 1976; Sunil Krishna & Bag 2013).

Spectra are processed for analysis to extract meaningful information about the properties of astronomical sources, such as elemental composition, temperature, velocity, and distance. This was done through the analysis of the resulting one-dimensional spectra from the observations, enhancing the understanding of the sources based on the continuum shape, along with the positions and (relative) strengths of emission and absorption lines. This technique has proven valuable in astronomy as multi-wavelength analysis on celestial sources can characterise the nature of sources more comprehensively, with each additional wavelength band used to analyse the light from each source. Extended radio sources have also been studied using optical spectroscopy. For example, for AGN (mostly FRI and FRII galaxies that are not part of rich galaxy clusters), it has been found that there is a correlation between the H α emission line luminosity and radio luminosity. Finding this correlation along with clear signs of rotation in the emission lines of a spectrum, which indicate that the intergalactic medium is rotating around a nucleus, would make a strong case for the identification of an AGN (De Young 2002 and references therein).

The significance of conducting follow-up spectroscopic observation lies in establishing a connection between

[8]These can be found in textbooks or papers with tables stating the M value for all spectral types of stars, we used the table available from the Handbook of Space Astronomy and Astrophysics.

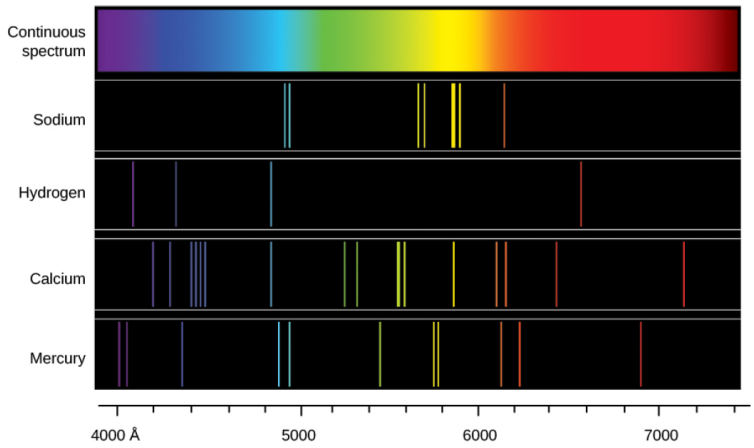


Figure 1.10: The emission line "fingerprint" of sodium, hydrogen, calcium and mercury found when the light they emit as heated substances is dispersed through a prism. Credit: (Fraknoi et al. 2022)

the optical light detected by telescopes and the underlying physical processes responsible for the emission of radio waves (and potentially other wavelengths). By combining the observed optical data with existing radio observations, astronomers can gain deeper insights into various astrophysical phenomena, including different types of galaxies, AGNs, the remnants of supernova explosions and their impact on the surrounding environment, and objects like quasars or star-forming regions. Comparing observed spectra to standardised, verified and labelled spectra of well-studied and classified sources is another useful way to inspect spectra for source characterisation visually. One example of such a spectral database is the Pickles Atlas, it is a stellar spectral flux library spanning 131 flux-calibrated spectra, covering all normal spectral types and luminosity classes between 1150 to 10620 Å, and up to 25000 Å for some of them (Pickles 1998). Follow-up observation is crucial for correlating specific technical observations with broader scientific questions, such as understanding the evolution of celestial objects over varying timescales and deciphering how and why their compositions change over time. The answers to questions like these can shed light on the processes that govern the celestial object's composition, behaviour, and emission mechanisms. Stars can be identified and categorised by the relative absorption intensities of the Balmer lines, the slope of the continuum, and the presence or absence of features such as the sodium doublet, the H α, and Calcium H & K line features during spectral analysis. Other objects like nebulae and star-forming regions are identified by their many, broad, characteristic emission lines (Jaschek & Jaschek 1987; Fraknoi et al. 2022). Spectroscopic follow-up validates (or challenges) what we know about certain objects and how they behave, and improves source classification and the accuracy of astronomical databases, which are then used for future research and scientific discoveries.

In the following chapters, source characterisation results that were produced by analysing a MeerKAT radio continuum dataset using citizen science (Chapter 2) and optical follow-up spectroscopic observation (Chapter 3) are presented.

Chapter 2

Citizen science approach to source classification

Citizen science is the collaboration between the public and scientists in ongoing scientific research. The public is usually asked if a certain pattern or phenomenon is present in a part of the data shown to them, like whether a spiral is visible in an image of a galaxy. All the recorded answers are combined into large tables to form a results database. The project creators can then analyse the results and publish their findings. This approach of giving individuals who may not be professionals in certain fields the opportunity to contribute to data collection, and sometimes the analysis thereof, has gained traction in recent decades. Citizen science has driven technological advances, increased public interest in science, and highlighted the importance of diverse contributions to scientific endeavours (Aristeidou & Herodotou 2020). With the increase in popularity and availability of interactive online tools, many fields, including conservation of biodiversity, weather, archaeology, astronomy, and health, to name a few, have tapped into the use of citizen science.

In general, citizen science has contributed to the global society by playing a role in:

- The enhancement of scientific research through the contributions of citizen scientists to scientific inquiry. In astronomy specifically, new celestial bodies such as exoplanets, comets, and galaxies have been studied through citizen science by sifting through the massive amounts of data gathered by modern telescopes (Masters & Team 2019; Andersson et al. 2023).

- Fostering scientific engagement with the public by opening channels of communication between the public and researchers. Through these channels, the public can learn about topics that interest them while providing valuable data points to scientists. Individuals from the public gain first-hand experience on how the scientific process works. Citizen science can work towards the expansion of access to knowledge and tools, enhancing the accessibility and relevance of science in everyday life (Price & Lee 2013; Jennett et al. 2016).

- Allowing a greater fraction of the public to be scientifically informed by the results that are published using conclusions drawn from citizen science channels. Scientific results are sometimes reported by news outlets whose content is consumed by the public. This cycle can repeat when members of the public are motivated to become citizen scientists (Kloetzer et al. 2016).

- Contributing to the development of skills and education. This is achieved in citizen science projects, where citizen scientists learn a skill used to collect or classify data. These approaches to citizen science help participants develop problem-solving, critical thinking, and data-analysis skills that may enhance their own success in the modern workforce (see Aristeidou & Herodotou 2020 and references therein).

- Addressing global challenges in both the historical and the current sense. Long-term changes in our planet's climate, or the geomagnetic effects in the atmosphere produced by cyclical solar activity, can be studied to predict future changes through the construction of intricate models (Price & Lee 2013; Soni et al. 2023).

- Building community ties with enhanced social cohesion using projects that can be done in localised groups of citizen scientists. Schools are also encouraged, during certain projects, to let their students participate to expand their educational scope.

- Positively influencing policy and decision making when high-impact results are published (Aristeidou et al. 2015; Schade et al. 2021). Citizen science initiatives can also play a role in the decision-making process at the local, national, and global levels. There are large areas of interest, including, for example, biodiversity and forest conservation. Projects in these fields report findings on the wider effects that result from the loss of biodiversity. Pressure can then be placed on the push for policy that will bring about positive changes (Van Brussel & Huyse 2019; Schade et al. 2021).

In Section 2.1, citizen science and the Zooniverse platform through which we realise a citizen science approach to this work are introduced and discussed. The MeerKAT radio source catalogue is introduced in Section 2.2, and we discuss how it was used in connection with the Zooniverse project. The design of the Zooniverse project is outlined in Section 2.3, with a discussion on how classification data points were gathered. The results of this chapter are presented and discussed in Section 2.4.

2.1 The Zooniverse platform

Zooniverse is an online platform for citizen science, with project themes ranging from history to biology and astrophysics. Hundreds of peer-reviewed astronomy papers have used Zooniverse or results from projects published on the platform. We followed two projects that seeded the idea to use citizen science during this study. The first project is called "Galaxy Zoo", which has been running since 2007. The other is a more recently launched project called "Bursts from Space: MeerKAT" by Alex Andersson, and it deals with finding radio transients, the first of its kind on Zooniverse.

Galaxy Zoo (GZ) was the project that led to Zooniverse being created as a citizen science platform. The project has resulted in the morphological classification of more than a million galaxies from multiple telescopes and projects. 2019 marked 12 years of GZ collecting data on galaxy morphologies using images from multiple large survey projects (Masters & Team, 2019). New workflows and datasets have been added as modern telescopes released images for public scrutiny and scientific analysis. Many of the galaxies that were morphologically classified through this Zooniverse project were labelled and subsequently used to train an ensemble of neural networks (Walmsley et al., 2021).

The project by Alex Andersson engaged with more than 1000 volunteers who provided an astounding 89000 individual classifications of possible radio transients within 3 months of launching the project in 2021. The results from the first run of the project produced a large catalogue of radio transients, the largest at that time, after identifying 142 new variable radio sources. This catalogue grew to 168 variable radio sources at the time of publication (Andersson et al., 2023). More recently, the results of this citizen science project were compared to the classifications made by using AI (through machine learning) on the same dataset (Andersson et al., 2024).

The reason for doing citizen science is to leverage collective human effort in the analysis of a dataset, and ultimately for this to yield statistically significant insights. In the case of this dissertation[1], citizen scientists were tasked with analysing optical images, each of which covers 4 by 4 arcmin of the SMC field, centred on radio source coordinates with the radio flux contours overlaid onto the optical image. Following the aims of this MSc work to characterise radio sources in the SMC, there was a large enough sample of images to justify creating a citizen science project to outsource the search for interesting sources in the selected sample. A

[1] Radio whispers from our neighbouring galaxy Zooniverse project, hereafter shortened to "Radio whispers"

collaborative approach like this will not only enhance the reliability of findings but it can also highlight the role that community engagement can play in scientific endeavours.

2.2 MeerKAT radio source catalogue

The radio data used in this project were derived from 2 radio continuum images centred on SNRs SXP1062 (Haberl et al. 2011; Hénault-Brunet et al. 2011) and SXP1323 (Carpano et al. 2017). These SNRs and their associated natal X-ray binaries have been studied in other published papers (see Joseph et al. 2019; Maggi et al. 2019; Gvaramadze et al. 2021 among others). The MeerKAT images are rich with many other radio sources due to the large field of view, sensitivity, and resolution of the MeerKAT telescope. The richness and diversity of these other radio sources were analysed in this Zooniverse project.

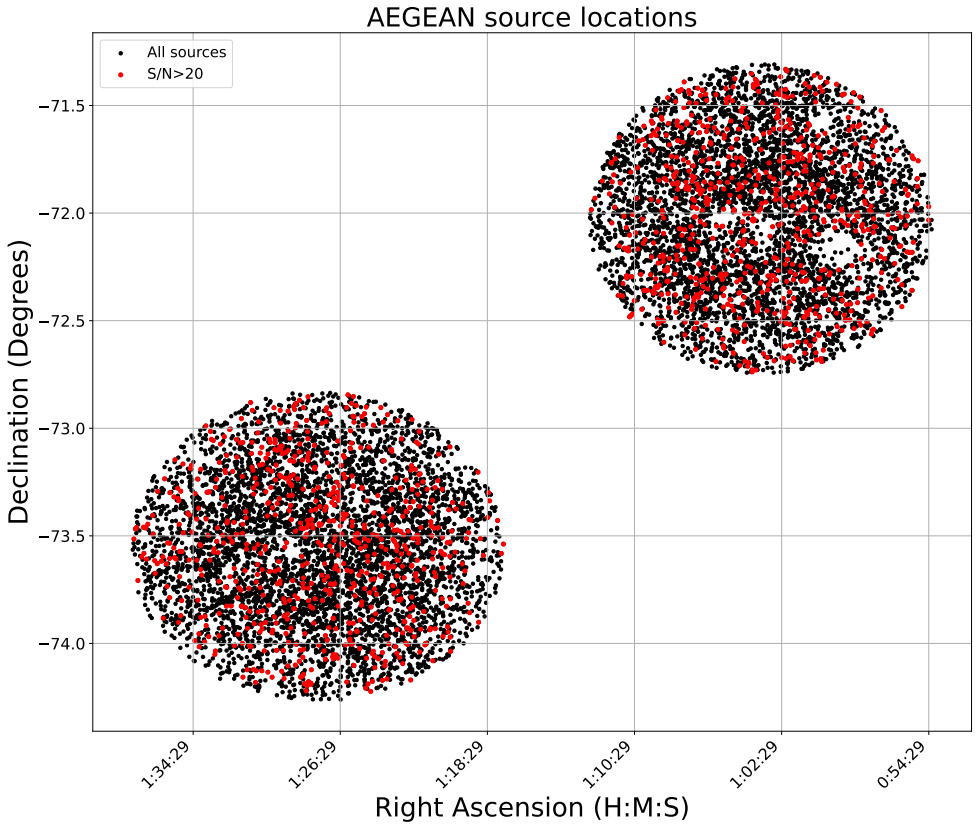


Figure 2.1: Above is a plot indicating the coordinates of all radio sources identified by AEGEAN, within a ∼0.5° radius of the central coordinates, in black dots (10873 total), and those with a radio S/N greater than 20 in red dots (1429 total).

There are a total of 10873 sources from both fields identified by AEGEAN within 0.5 ° of the central SNR sources. To select a subset of the radio source catalogue that would lend itself to a citizen science classification project on the timescale of an MSc, we selected only radio sources with a signal-to-noise (S/N) ratio > 20. This S/N was determined by dividing the peak flux by the RMS background value for each radio source. The sources with a S/N above 10, 20, 50 and 100 were sorted into new tables; these tables contained totals of 3800, 1429, 549 and 285 sources, respectively. In Figure 2.1, the subset of sources with a S/N greater than 20 can be seen in red; this subset was selected for classification through Zooniverse.

2.2.1 Decision tree and workflow design

Based on the objective of characterising the radio sources, the design process of the citizen science project centred around the question "what are the questions we need to ask when a volunteer classifier is looking at each image?". We needed to know what the shape and size of the radio contours tell us about the morphology of the radio source, and whether an optical counterpart to the radio source is present or not. If there is one, we could ask what that looks like to get information about both the radio and optical sources.

- What do the radio contours look like (radio *point* or *extended*)?
- If the radio contours indicate an *extended* radio source, is the extended emission *compact* or *structured*?
- Is there an optical source that seems associated with the radio contours (optical *counterpart*)?
- If there seems to be an optical counterpart to the radio source, what does it look like?

The terms *compact* extended and *structured* extended were used to help classifiers distinguish extended sources whose radio contours contained one or multiple blue ellipses, respectively. Blue ellipses in this data indicate the positions of sources that are to be shown to the classifiers as separate subjects (see Figure 2.5 that illustrates why these terms were used).

We needed an optimisation in the workflow to have a minimum amount of time spent per classification while gathering all the necessary answers for each subject. There are at most 5 answers or data points for all classifications. If all five questions were answered, that source would be a structured extended radio source with a possible optical counterpart associated with the radio source. The minimum number of questions that need to be answered is 2. Subjects with only 2 data points are radio point sources with no optical counterpart. This optimised design can be seen in Figure 2.2, which is a compacted version of the complete decision tree in Appendix A.

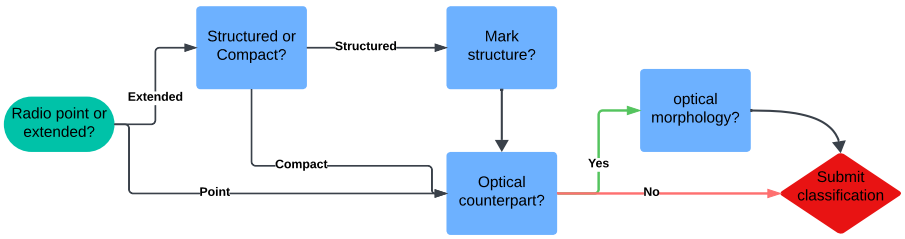


Figure 2.2: This is a simplified version of the complete decision tree (see Appendix A) used to guide the design of the Zooniverse citizen science project workflow for the desired outputs.

The workflows were designed to capture relevant flags for each subject. The first flag indicates whether the radio source is considered a *point* source or an *extended* source based on the radio contours. Volunteer classifiers would select the radio *point* source if the radio contours were of similar size and shape to the beam size (which is indicated visually in the bottom right corner of each image as a white/grey dot as seen in Figure 2.4), or *extended* otherwise. If the radio source is flagged as *extended*, classifiers were asked to indicate whether or not the *extended* source is *compact* (larger than a point source but isolated in terms of radio flux contours) or forms part of a larger extended structure (*structured*). The next flag concerns the presence or absence of an optical *counterpart* when looking at the optical image. If the volunteer classifier indicates that an optical *counterpart* seems associated based on proximity to the radio contour centre, the final flag concerns the morphology of the optical source that is thought to be the optical *counterpart* to the radio emission.

2.3 Designing the online interface

Upon reading all the guides provided by Zooniverse to help project creators along the way, platform-specific terminology came into play. Each image that needs classification is called a "subject". The system of questions and methods of answering them is called a "workflow", and each time a citizen scientist finishes a workflow for a subject, that counts as 1 classification. To make the output data statistically viable for scientific analysis and interpretation, it is recommended that each subject needs to have 5 or more classifications (completed workflows), depending on the nature of the data collected and the size of the workflow, before that subject is "retired". When a subject is retired, it will not be shown to another volunteer for further data collection. This is how the progress of a project is measured, either by the number of classifications submitted or by the number of subjects that have been retired. The design and testing phase spanned from May 2023 and February 2024. The implementation of the decision tree design (see Appendix A) was done during this time. The official data collection phase started in March 2024. We chose to use the list of sources with a S/N of 20 or more and to have 5 classifications on each of the 1429 subjects before retiring it. Thus, 7145 classifications were needed for the 1429 subjects. All subjects were completed on 6 September 2024 (see Figure 2.3 for the monthly classification contributions to the Zooniverse project). During the test phase (2023-05 to 2024-02), 100 images were uploaded as test phase subjects to record test classifications as the workflow design was implemented and modified as needed.

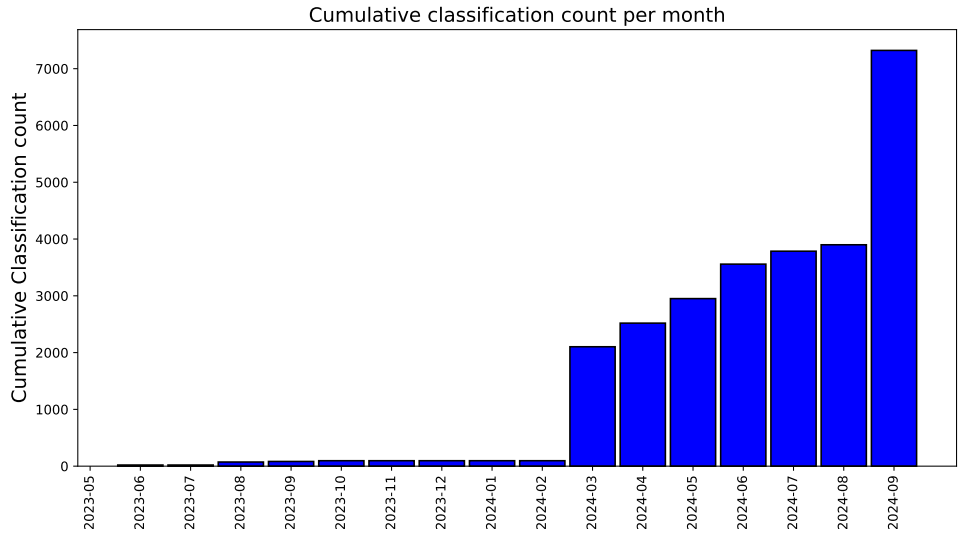


Figure 2.3: The graph shows the cumulative monthly increase in classification count for the duration of the Zooniverse project. The test and development phase spanned 2023/05 to 2024/02. The project was published, and the classification phase spanned 2024/03 to 2024/09.

During the official classification gathering phase, information about each source's radio and optical characteristics (if an optical source is aligned with the radio source and possibly could be a counterpart) had to be gathered. To do this, we required optical images in the field of the SMC. The SuperCOSMOS sky survey (SSS from here on, Peacock et al. 2016) is a survey built from the digitisation of the Schmidt photographic plates from the UK Schmidt Telescope, as well as the Second Palomar Observatory Sky Survey. The Sloan Digital All Sky Survey (SDSS, Eisenstein et al. 2011; Blanton et al. 2017) data was used to calibrate the photometry. This optical survey was chosen as the optical images used/shown during this Zooniverse project for its good resolution (0.7 arcsec pixel size as reported on the SSS website) and wide sky coverage, which includes the entire SMC. This survey conveniently has its own image cutout server, where a list of celestial coordinates can be entered with a cutout image size for each target delivered in Flexible Image Transport System format (FITS). This made the optical and radio target coordinates easy to compare visually, by overlaying the radio flux contours onto the optical images using the SAODS9 software.

2.3.1 Subject preparation for Zooniverse

To create the images that were the subjects of the Zooniverse project, the first task was to generate the flux contour maps from the radio images. The radio flux contour maps were saved from DS9 as separate files, the flux levels for the SXP 1062 field were (0.003, 0.008, 0.2, 0.52, 1.1, 2.2, 8.5, 15, 29, 56, 80, 105, 150, 190) mJy/beam and the levels for the 1323 field were (0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1, 5, 10, 50, 100, 500) mJy/beam. To create the optical images, the cutouts from the SSS server were retrieved from the B-band (blue filter GG385) and R-band (red filter RG630). Composite images were created by importing 2 images and blending them into a single image, where the B-band was allocated the blue colour and the R-band was allocated the red colour. The AEGEAN radio source catalogue relates the approximate size, shape, and orientation of each detected radio source to the parameters of a tilted ellipse. These parameters were used to visually identify each radio source. The radio flux contour maps were overlayed onto the composite images along with source locations and approximate orientation ellipses. The ellipses were colour-coded based on S/N (dotted green for sources with S/N between 5 and 20 and solid blue for sources with S/N greater than 20. See Figure 2.4). The regions with S/N between 5 and 20 were included, as AEGEAN is an algorithm that identifies compact sources (Hancock et al. 2012, 2018) and will identify radio emission regions that are part of an *extended* source's structure as separate sources (this is illustrated in Figure 2.5); because the extended structure of some radio sources can be fainter than the core section of the source (or vice versa). The I-band (near-infrared filter RG715) images from SSS gave a cloudy/obscuring effect to the image after the blue and red bands were merged. To keep more detail in the images, only the B and R-band images were used.

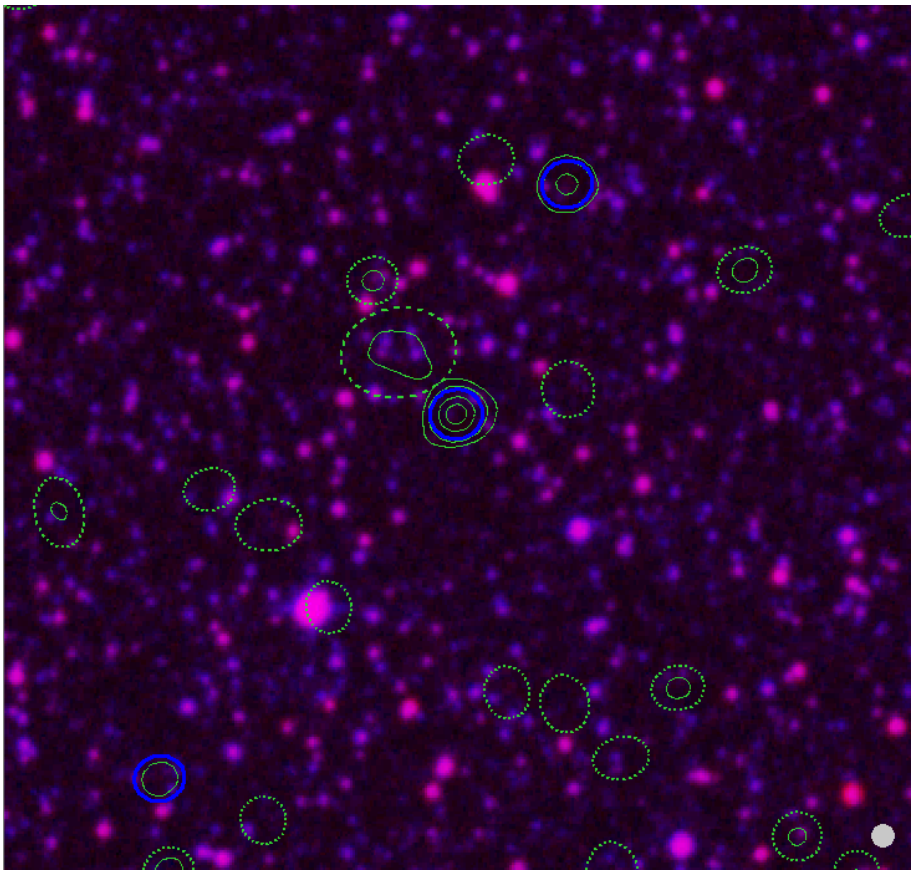


Figure 2.4: Above is one of the subjects as they are displayed on the Zooniverse platform. The images are 4 arcmin by 4 arcmin squares, with North-up and East-left orientation. Radio flux contours are overlayed in the solid green lines. The central source is always indicated by a solid blue ellipse, which identifies it as a source with radio S/N greater than 20. The dotted green ellipses indicate sources with radio S/N between 5 and 20 and are used to show possible extended structures. The radio beam size is displayed in the bottom right corner of the images as a grey dot.

We defined 3 main categories of radio sources in this Zooniverse project. The first is radio *point* sources. These were the most numerous, as expected. The second is *compact* extended radio sources. These are radio sources that are too large to be point sources and do not seem to be connected to larger-scale radio emissions. The last is *structured* extended radio sources with a possible relation to large-scale extended emissions. Examples of these are shown visually in Figure 2.5.

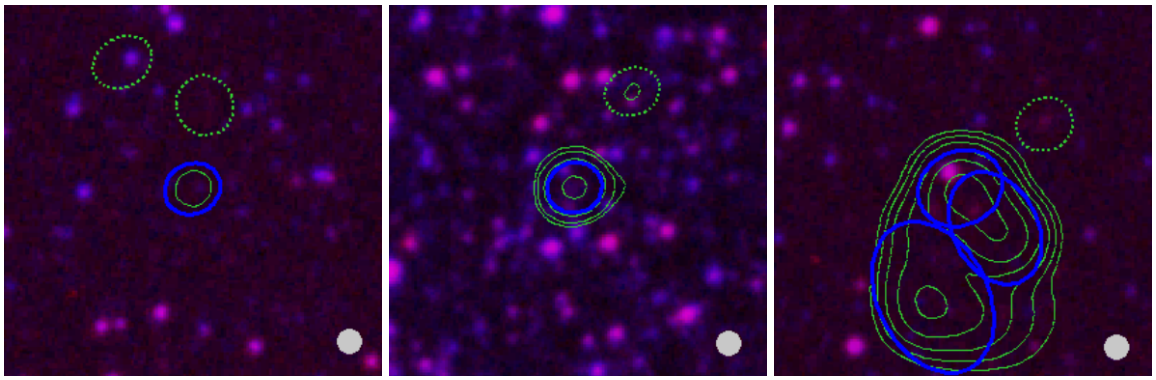


Figure 2.5: The figure above shows a visual example of what should constitute a radio point (left, with a contour of similar shape and size to the grey radio beam size), compact (middle, with a contour having a comparable shape but larger size when compared to the radio beam size), or structured (right, neither the contour shape or size is comparable to the radio beam size) source from the Zooniverse data shown to volunteer classifiers. The images have been cropped to a square showing a ∼1.5 by 1.5 arcmin patch of sky for these 3 images.

2.4 Results and statistics after aggregation

The raw output data tables downloaded from Zooniverse were first stripped of personal identifiers to maintain the anonymity of the volunteer classifiers (and avoid the need for ethical clearance being required to go ahead with the project). Extractors and reducers[2] are scripts that extract raw data from Zooniverse output tables and compile a usable database for analysis of results. This process is termed "aggregation". These extractors and reducers were used to compile a master database from the raw classification data received from Zooniverse for the Radio Whispers project. This raw (not cleaned) master database[3] produced after aggregation spans 29289 rows and 32 columns containing the recorded data of every question answered during each classification. Each row represented one of the four classification task answers recorded for every subject, and each subject (excluding the test phase subjects) had at least 5 classifications.

Overall statistics, including the testing phase:

- The project was active between 2023-05-24 and 2024-09-09.
- 7298 classifications of 1521 subjects (including test phase duplicated subjects).
- 91 classifiers and/or testers.
- Median number of classifications per volunteer: 17.0
- Mean number of classifications per volunteer: 80.2
- A Gini coefficient[4] for classifications of 0.78.
- Based on the 7298 classifications, 2.71 days were spent on classifying.
- Median time spent per classification: 32.3 seconds.
- Mean time spent per classification: 14.3 seconds.

[2] https://aggregation-caesar.zooniverse.org/GUI.html

[3] It is named "question_extractor_extractions.csv" on the GitHub associated with this Zooniverse project

[4] A Gini coefficient is a statistical measure used, in this case, to measure the distribution of work done by volunteers. It is measured from 0 (meaning that all participants made the same number of classifications) to 1 (meaning that 1 person did all the classifications).

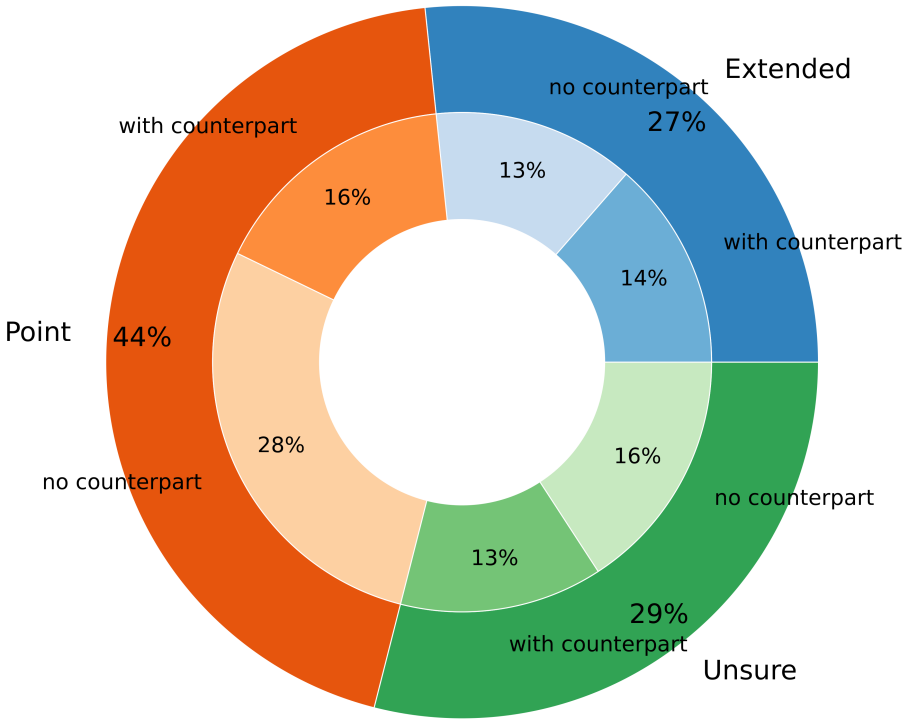


Figure 2.6: This chart indicates what percentage of the total 1429 subjects were classified as a radio-point or extended source, or flagged as uncertain, in the outer ring. The criterion for agreement was set to > 80 per cent, meaning that > 80 per cent of classifiers had to agree on the subject being a point or extended source, or it was flagged as uncertain. The inner ring indicates the percentage of the total 1429 sources belonging to a certain morphological category, while also having an optical counterpart (possibly), or not, using the same agreement criterion as the outer ring (> 80 per cent of classifiers had to agree that a subject has an optical source alignment with the radio emission, or it was flagged as not having it).

The master (not cleaned) database contained many columns that were redundantly kept from previous versions of the workflow design before the project was published after the testing phase. These columns were removed along with columns that contained no data or all 0 values. Each subject occupied 4 lines as the assembly of the output tables was done according to the subject ID number and the task ID, and 4 responses were recorded per classification for each subject. There are technically at most 5 recorded responses per classification, however, the task excluded here is not considered a "Question task", but a "Drawing task" which is handled differently from these 4 during the extraction of Zooniverse result tables. By eliminating the task column from the database in Python, the reduced database[5] spanned 1429 rows and 31 columns (see appendix E for the listed column names with a short description of their content). From this table, useful statistics can be drawn with constraints on agreement between volunteer classifiers. In Figure 2.6 the first broad results are shown for the number of sources classified as radio point or extended sources based on a > 80 per cent agreement criterion between classifiers (71 per cent or 1015 of the 1429 subjects met this agreement criterion). This led to 634 sources being classified as radio point sources, 381 classified as radio extended sources, and 414 having uncertainty in the classification of radio morphology (see Table 2.1 for the final numbers including alignments to possible optical counterparts).

This 80 per cent criterion was chosen after completion of the project. At the time of launch, we chose the retirement limit to be 10 classifications before a subject is removed from rotation. This number proved to be too large, as we required 14290 classifications to complete the project. In June of 2024, it was decided to change this retirement limit to 5 classifications. After completion of the project, we realised that some subjects had already gathered more than 5 classifications at the time of changing the retirement level, however, there

[5]The cleaned and processed table can be found on the GitHub, called REDTEST.csv

were only subjects with 5, 6 or 7 total classifications. The > 80 per cent agreement criterion meant that if the subject had 5 classifications, all 5 classifications had to agree. If the subject had 6 classifications, 5 of the 6 had to agree. If the subject had 7 classifications, 6 had to agree. In this way, the classifications that were accepted all had 5 or 6 independent classifications indicating the same characterisation.

Table 2.1: The final count from the initial results after sorting subjects into either point or extended radio source categories if > 80 per cent of the volunteer classifiers agreed either way, or into the unsure category if there was not > 80 per cent agreement between volunteer classifiers on the radio morphology from the radio flux contours. The number of sources that are also aligned with an optical source (also based on the > 80 per cent agreement criterion between classifiers) is included.

	Point	Extended	Unsure	Totals
Aligned with optical source	231	194	188	613
No obvious optical alignment	403	187	226	816
Totals	634	381	414	1429

The sources flagged as uncertain were excluded from further analysis, and a high-confidence database was constructed for radio point, "compact extended" and "structured extended" sources. The number of point sources is 634 with high confidence. Extended sources were designated "compact extended" if > 80 per cent of the volunteer classifiers agreed that the extended source did not have additional regions of radio flux associated with it, or they were designed as "structured extended" sources if > 80 per cent agreed that they did. This led to a list of 128 "compact extended" and 213 "structured extended" sources (excluding 40 extended sources with uncertainty between compact or structured morphology from further analysis).

Table 2.2: The final high-confidence counts of sources across the 3 final (radio) morphological categories are shown along with a subdivision according to the likely presence of an optical counterpart, according to the data gathered through Zooniverse.

	Point	Compact extended	Structured extended	Totals
Aligned with optical source	231	70	65	366
No obvious optical alignment	403	58	67	528
Totals	634	128	132	894

A final cut was needed on the "structured extended" source list, as sources may have duplicate entries (see Figure 2.8) if one of the associated regions had a S/N high enough to have been included as a subject for classification itself. After manual visual inspections, the high-confidence structured extended source list comprised 132 sources. This means that 894 sources, out of 1429 that were classified (∼63 per cent), show strong agreement on the classification (see Figure 2.7 and Table 2.2).

The results from Zooniverse may not be enough to classify these sources accurately, but the final 6 categories could be linked to potential source types. For instance, radio point sources with optical counterparts are most likely stars in this field. When looking at some of the optical morphologies, they could also be background AGN cores or planetary nebulae in the SMC. Radio point sources with no apparent optical counterpart could be distant radio-loud, optically thick AGN cores. Alternatively, they might be high proper motion stars that have shifted significantly enough on-sky between the time of observation for the SSS-optical survey data in ∼ 1975 and the MeerKAT observations in 2019, causing the old optical position to differ noticeably from the newer radio position. A SIMBAD search by criteria[6] reveals that there are 36 high proper motion stars within a 0.5 ° radius (within which AEGEAN searched for radio sources) of SXP1323 and 35 within the same radius of SXP 1062. Assuming a homogenous distribution of high proper motion stars and a flat circular area of the sky, this leads to an estimated $\frac{35}{\pi(0.5)^2}$ ∼45 high proper motion stars per square degree in either radio field. This shows some of these stars may be radio stars that were recently detected by MeerKAT and have moved out of position since the SSS observations.

[6]The query command used is "region(CIRCLE,01:27:46 -73:32:56,0.5d) & pm > 0.2 & maintype = PM*" for the region around SXP 1062, replacing the coordinates with that of SXP1323 will search around that coordinate.

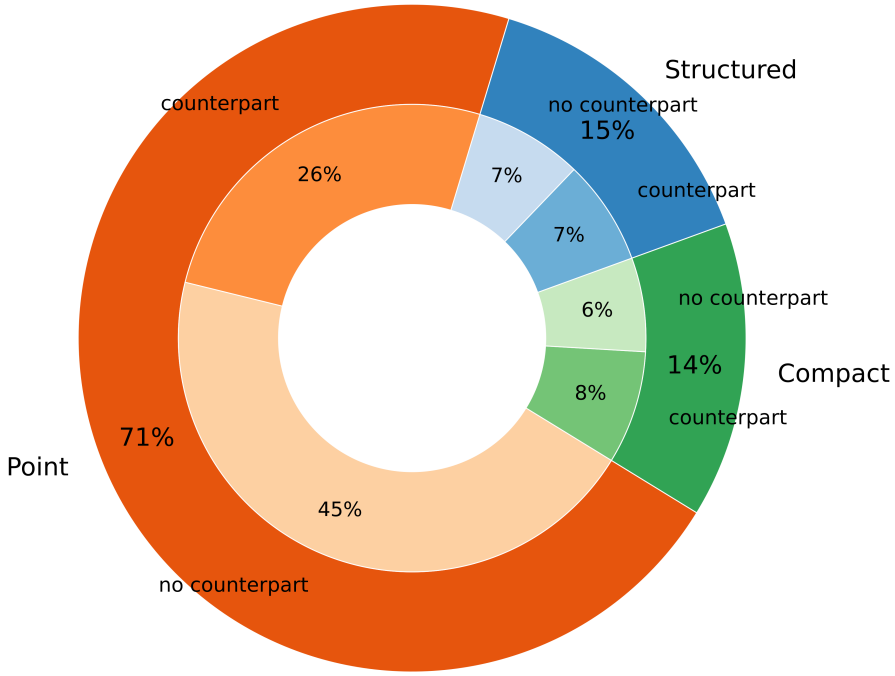


Figure 2.7: The distribution of high-confidence source classifications is shown on the outside ring, and the percentages of those sources that have optical counterparts are again on the inside chart.

Compact extended radio sources with counterparts could be young nebulae, compact binary stars, background AGN, close-by radio stars, or pulsars (Galactic or inside the SMC embedded in a PWN and possibly a SNR shell). Compact extended sources without apparent counterparts could be due to survey coordinate misalignment with a high-proper motion Galactic object, a background AGN that is optically thick or too faint to have been detected with the SSS optical data. They could also include gas/dust obscured young stellar objects (Dent et al. 1998, and references therein), but with the low amount of obscuration between us and the SMC, this is uncertain unless the gas/dust is in the SMC, localised around any YSOs. Finally, the structured extended sources might be AGN with radio lobe structure, supernova remnants between us and the inside of the SMC, or large nebulae from stellar evolution and formation. Structured extended radio sources with no apparent optical counterpart might be background radio galaxies with an optically thick core and possible jet structure not visible in the MeerKAT L-band, or star-forming regions between us and the SMC with no obvious optical point source to call a counterpart. Further observation can be done (or if more data is publicly available from one of the large survey projects) on these sources to get measurements of polarimetry, high-resolution spectroscopy, or spectral energy distributions, which will further narrow down source classifications or identification. In some cases, we just do not know because some source types could fit into more than one category. In Chapter 3, an attempt at this was made with optical spectroscopic follow-up observations of sources from the same base radio source catalogue.

Table 2.3: When identifying point sources using the approximate area of the MeerKAT PSF compared to the area covered by the AEGEAN sources linked to a tilted ellipse, there is a noticeable difference in the number of sources that would have been classified as point sources by AEGEAN compared to the number of high-confidence Zooniverse characterisations.

	AEGEAN+PSF	Zooniverse
Point source	1154	634
Not a point source	275	795

Point sources identified by AEGEAN can easily be separated out of the catalogue since point sources will

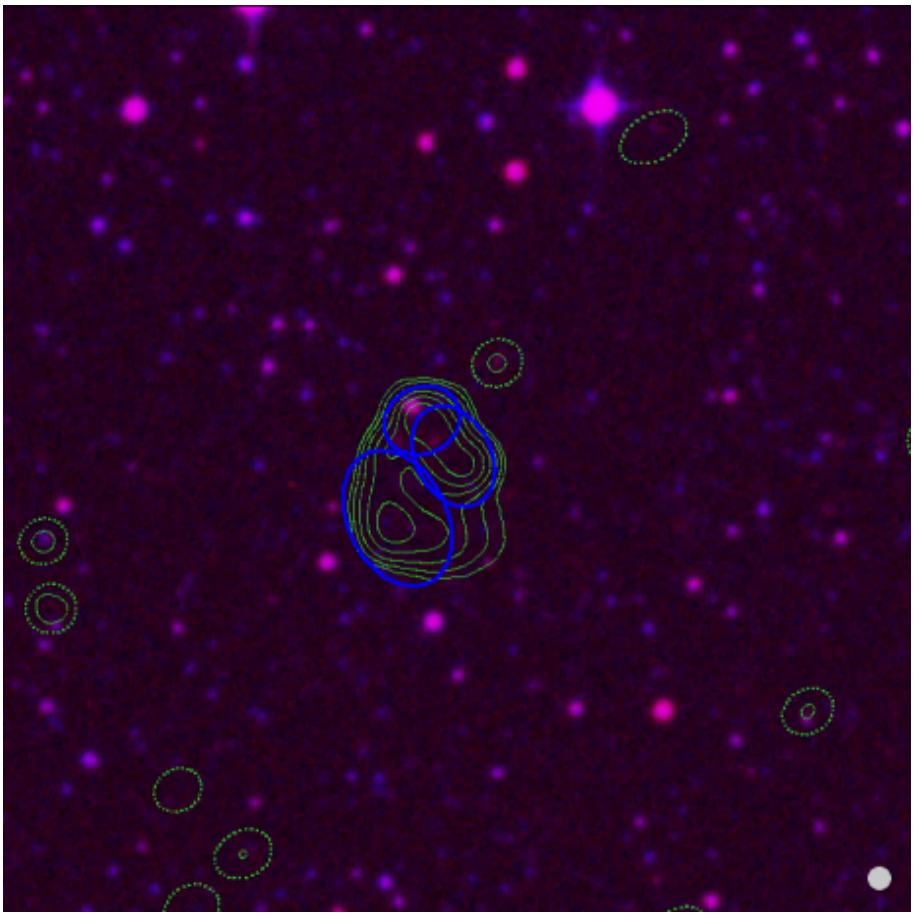


Figure 2.8: An example of one of the images shown to classifiers. It and the other 2 high S/N sources were classified as a structured extended source multiple times since multiple regions have high S/N solid blue ellipses (Note that the solid blue ellipses indicate the locations of sources to be classified in Zooniverse).

look like the point spread function (PSF) of the continuum data (Hancock et al. 2012). AEGEAN data output tables contain parameters for an ellipse linked to each source identified, which indicates the source island's shape and position angle. These ellipses can be compared to the PSF for each source with a justified threshold and compared to those sources identified as point sources during the Zooniverse project to study the accuracy of AEGEAN compared to the Zooniverse classifications.

We used another method to find radio point source numbers using only the MeerKAT point spread function (PSF) and the AEGEAN catalogue source size estimates. Using the PSF and assuming a circular shape with the 6.5 arcsec major axis of the MeerKAT PSF as the radius, and adding 15 per cent on top of that to account for sources slightly bigger than the PSF that would still look like point sources, AEGEAN would have characterised more radio point sources than the Zooniverse results show (see Table 2.3).

For the high-confidence Zooniverse list, AEGEAN would have caught 516 of the 634 high-confidence point sources (81.4 per cent) as point sources using the method described above. The Zooniverse numbers are higher due to the way we defined what a point source would look like in the Zooniverse data (the contours should be inside the AEGEAN ellipse and could be slightly bigger than the PSF indicated on the bottom right-hand corner, so long as the shape is approximately the same).

For the entire sample that went to Zooniverse for characterisation, AEGEAN would have caught 1154 of the 1429 (80.8 per cent) sources as point sources. We believe these are overestimated since many of the smaller ellipses form part of a larger extended structure (or a cluster of ellipses close to each other).

Chapter 3

Optical follow-up observations

Spectroscopy is a method of collecting and analysing light by dispersing light into its constituent wavelengths. A spectrum displays the intensity of light as a function of wavelength, usually having a spatial axis representing the position of a source on the physical slit through which light is passed and a dispersion axis along which light is dispersed to create the spectrum. Each wavelength interval is then measured to find how much of the light's power is contained (or indeed absent) in that wavelength interval. Each atomic element (and molecular binding) has its signature in a spectrum, like hydrogen, which emits photons at different specific wavelengths than anything else like helium or sulphur. Knowing which signature belongs to which element or molecule, we can determine the composition of a source that has undergone spectroscopic analysis. This method creates a spectrum of light that acts as a fingerprint for celestial objects, which can, in turn, be used to help identify a cosmic source after having been observed with a telescope that has a spectrograph instrument mounted on it (Fraknoi et al. 2022). Spectroscopic analysis of celestial sources can reveal characteristics and emission mechanisms of those sources that can narrow down source classifications.

In this chapter, the results of an optical spectroscopic follow-up of high radio S/N sources cross-matched within 1 arcsec to a GAIA survey source with a bright enough magnitude to be visible to the SAAO 1.9-m telescope, since this is the telescope that was used for the observations, are presented. The observation run was done on-site at the Sutherland SAAO site during the nights of the 16th to the 21st of November in 2023. Section 3.1 details how sources were selected for follow-up spectroscopic observation, how sources from this list were prioritised for observation and a summary table of all candidates. Section 3.2 expands on the observation run and steps followed during data reduction. The 1D spectra of all observed sources are shown in Section 3.3 along with individual source analysis results. Section 3.4 contains a summary of the results for this chapter along with problems encountered during analysis, and how these were handled.

3.1 Target selection and spectroscopy using SpUpNIC instrument

AEGEAN identified a total of 10873 radio sources (see subsection 1.4.1 for details on the AEGEAN software) within a 0.5° radius of the central SNR sources. These sources are split into 5546 and 5327 radio sources in the fields centred on SXP1062 and SXP1323, respectively. In TOPCAT[1] the radio source coordinates were cross-matched to the GAIA (Gaia Collaboration et al. 2016) database[2] using a matching distance criterion of 1 arcsec. There were 146 source matches from the SXP1062 field and 484 sources from the SXP1323 field, totalling 630. Most of these sources had optical magnitudes from GAIA (GAIA *g_mean_mag* is the passband that covers a wavelength range from roughly 3300 to 10500 Å) ranging up to ~ 22, which are not bright enough to be detected within the sensitivity range of the South African Astronomical Observatory (SAAO) 1.9-metre telescope in Sutherland. It was decided to make a magnitude cut in the GAIA-matched catalogue

[1] TOPCAT tool website
[2] https://www.esa.int/Science_Exploration/Space_Science/Gaia_overview

and only keep sources with a magnitude (*g_mean_mag*) of less than 18 (see Figure 3.1, and Crause et al. 2016 for magnitude limits of the spectrograph). This reduced the number of potential observation targets to 41 in total; see Table 3.1 for this list and more details. These 41 sources were visually inspected with the Aladin Lite online sky atlas to determine if they have been identified by previous publications linked to SIMBAD (see Wenger et al. 2000 for details about this CDS reference database) and if they have neighbours nearby that could interfere with the observations. Sources that did not have prior identification in the literature were prioritised for observation.

Table 3.1: The table below contains the information of the radio sources with a GAIA match within 1 arcsec of the radio coordinates and a GAIA *g_mean_mag* magnitude less than 18. The RA and Dec coordinates are J2000 coordinates of the radio sources from AEGEAN. The "*g*" column has the GAIA *g_mean_mag* magnitude of the optical source from cross-matching with GAIA. "Observed_date" contains the date of observation. "S/N" contains the Signal/Noise ratio of the 1D spectrum between the wavelengths of 4000 and 7000 Å. "Ang_dist" contains the angular separation between the radio and GAIA coordinates in arcseconds. "Known_ID" has the name of the optical source from SIMBAD if it is identified in published literature. "Counterpart" contains a 1 or 0 to indicate whether it is believed that the optical source of GDBXX is the optical counterpart to the radio source, based on how close the optical source appears visually to the radio source. All observations had single science exposures of 2400 seconds (40 minutes), excluding GDB1, 2 and 5, which were brighter than expected and were exposed for 1800 seconds (30 minutes).

Designation	RA (J2000.0)	Dec (J2000.0)	*g*	Observed_date	S/N	Ang_dist	Known_ID	Counterpart
GDB1	01:10:19.87	-72:03:35.51	10.4	16-11-23	97.6	0.581	FLO 581	0
GDB2	01:01:37.96	-71:54:16.84	11.6	21-11-23	68.9	0.764	UCAC4 091-002858	1
GDB3	01:24:50.77	-73:57:39.85	15.5	21-11-23	25.4	0.76		1
GDB4	01:02:57.70	-72:31:45.78	15.7	21-11-23	36.4	0.982		1
GDB5	01:22:58.56	-73:59:28.59	15.7	17-11-23	65.1	0.774		0
GDB6	00:59:33.74	-72:20:49.80	15.9	19-11-23	26.5	0.844		0
GDB7	01:03:53.91	-72:08:26.83	15.9	21-11-23	42.2	0.319	WBBe NGC 371 16	1
GDB8	01:07:36.22	-72:34:48.52	16.0	21-11-23	45.3	0.829	LHA 115-N 79	1
GDB9	01:00:58.19	-71:35:29.93	16.2	21-11-23	34	0.899	LHA 115-N 71	1
GDB10	01:24:35.08	-73:37:55.46	16.2	21-11-23	45.2	0.906	OGLE SMC-ECL- 8019	1
GDB11	01:07:22.13	-71:49:37.47	16.3			0.954		
GDB12	01:08:40.19	-71:58:49.73	16.3			0.974		
GDB13	01:04:15.92	-71:49:23.44	16.4	17-11-23	26.4	0.923		0
GDB14	00:59:16.19	-72:02:00.80	16.5	17-11-23	11.6	0.994	LHA 115-N 70	1
GDB15	01:04:29.29	-71:45:56.04	16.6			0.856		
GDB16	01:26:40.95	-72:59:39.72	16.6			0.9		
GDB17	01:24:54.52	-73:38:33.42	16.7			0.718		
GDB18	01:08:54.71	-72:12:08.02	17.0			0.831		
GDB19	00:58:42.95	-72:27:17.57	17.0			0.863		
GDB20	01:06:01.61	-72:03:36.29	17.1			0.64		
GDB21	01:19:56.20	-73:21:53.85	17.2	19-11-23	11	0.94		1
GDB22	01:31:10.87	-73:23:18.43	17.2			0.999		
GDB23	01:02:11.99	-71:48:51.07	17.2	18-11-23	15.8	0.925		0
GDB24	00:58:11.94	-71:28:02.00	17.3			0.871		
GDB25	01:27:57.93	-73:22:33.15	17.4	18-11-23	13	0.405		1
GDB26	01:29:36.95	-73:40:35.41	17.4			0.747		
GDB27	00:57:19.13	-72:33:01.99	17.5	18-11-23	10	0.311		0
GDB28	01:24:11.93	-74:02:33.04	17.5			0.936		
GDB29	01:02:01.28	-72:22:19.45	17.6			0.6		
GDB30	01:10:15.23	-72:02:11.83	17.7	18-11-23	15.9	0.41		1
GDB31	01:25:08.32	-73:46:14.09	17.8			0.748		
GDB32	01:06:15.07	-72:19:48.53	17.8	18-11-23	8.9	0.411		0
GDB33	01:32:30.63	-73:12:37.39	17.8			0.586		
GDB34	01:00:06.05	-71:57:06.91	17.8	18-11-23	7.3	0.535		1
GDB35	01:04:32.39	-72:04:43.96	17.8			0.764		
GDB36	01:03:42.68	-72:21:01.13	17.8			0.717		
GDB37	01:03:28.48	-72:29:50.37	17.9	17-11-23	6.6	0.618		0
GDB38	01:05:30.79	-72:17:02.09	17.9	17-11-23	14	0.482		1
GDB39	01:02:08.31	-72:06:48.74	17.9			0.921		
GDB40	00:57:32.81	-72:13:02.79	17.9	16-11-23	5	0.64	AGN Candidate	1
GDB41	01:04:07.21	-71:36:42.18	17.9	16-11-23	25.4	0.87		0

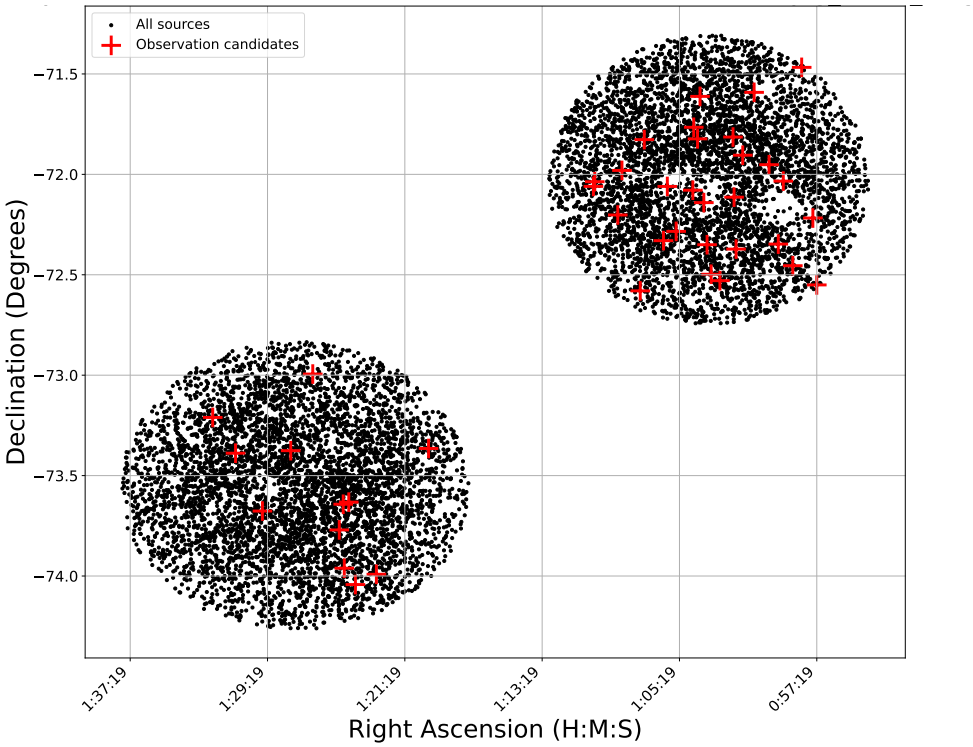


Figure 3.1: Sky location of radio sources identified by AEGEAN (black dots). Red pluses show sources cross-matched to the GAIA survey within 1 arcsec with a magnitude brighter than GAIA *g_mean_mag* = 18.

3.2 Spectroscopic follow-up & data reduction method

During the nights of 16 to 21 November 2023, 23 of the 41 qualifying sources were observed (see Table 3.1) with the SAAO 1.9 m telescope using the SpUpNIC spectrograph[3]. Grating 7 was used for its wide wavelength range, with the grating angled at 16 degrees to cover the wavelength range from 3850 Å to 9400 Å. This grating was chosen because it offers the optimal wavelength coverage, especially the H I, He I and He II lines, for source characterisation and high enough sensitivity to probe faint targets in the selected sample. One night out of the 6 allocated for the observation run was unusable due to complete cloud coverage. All observed science targets had single exposures between 30 and 40 minutes in duration, as most of the sources were faint. Each night, a standard star was observed for 300 seconds to be used during flux calibration. LTT1020[4] was the standard star chosen as it was available for observation each night directly after the flats and biases were taken. LTT1020 is also relatively close to the SMC, making slewing to the first science targets each night quick and efficient. Each science exposure was followed by a Copper-Argon (CuAr) arc lamp exposure before slewing to the next target, the arcs were used for wavelength calibration.

3.2.1 Image trimming to CCD data

These 23 spectroscopic observations were extracted using PyRAF[5] on the servers at the SAAO. The first step in the data reduction process was to trim the spectra to contain only data from the charge-coupled device (CCD) camera. This was accomplished using the task ***ccdproc*** inside the *noao/imred/ccdred* package with only the trim flag active. The *trimsec* variable was set to [2:1982,30:100], the format values are in pixel indices as $[x_{min} : x_{max}, y_{min}, y_{max}]$ where the x-axis measures wavelength, increasing from left to right, and the y-axis indicates the spacial position on the grating.

[3] See Crause et al. 2016 for technical details of the spectrograph
[4] LTT1020 standard star information on ESO website.
[5] Science Software Branch at STScI (2012)

3.2.2 Creating and applying the master flat and master bias files

After all the files had been trimmed to contain only CCD data, the dome flat images of each night were combined (by using median pixel values) into a master flat file for that specific night of observation, and the same for the bias measurements which were median combined into a master bias file for each night. The master flat file was created using the ***flatcombine*** task from the *ccdred* package, replacing the default constant values of the *gain* and *rdnoise* variables with the values recorded in the fits headers by the instrument under *GAINVAL* and *NOISEADU*. The output from this task was then inspected using DS9 to visually check if any problems came from combining the approximately 30 flats for each night. A single flat file was created by combining all viable flats of a night, it was then normalised using the ***response*** task inside the *noao/twodspec/longslit* package to create the master flat file. The master bias files were created by using the ***zerocombine*** task in the *ccdred* package to combine the approximately 30 bias files of each night into that night's master bias. For this task the defaults of the variables *gain* and *rdnoise* were replaced with the fits header values *GAINVAL* and *NOISEADU*. The master bias files were also visually inspected with DS9 to ensure nothing went wrong during the combining process. Thereafter, the ***ccdproc*** task is run with only the *zerocor* and *flatcor* flags active and the names of the master flat and bias are entered into the fields for the *zero* and *flat* variables. This applies the flat and bias corrections to all science spectrum images.

3.2.3 Wavelength calibration

Next in the reduction recipe, the x-axis was calibrated to be in units of wavelength in angstroms (Å). For this, the CuAr arc image taken directly after each science observation was used to wavelength calibrate that science exposure. The ***identify*** task in the *longslit* package was executed on the first arc of each night. During this task, the emission lines of the arc lamp are marked and given a wavelength coordinate belonging to that emission line. The correct wavelength values for each line is documented and publicly available on the SpUpNIC Wiki[6]. This task then matches the remaining lines from a list of lines for the specific arc lamp over a specific range by identifying peaks close to the correctly marked wavelengths. These line coordinates were double-checked and any wrongly marked coordinates were removed to ensure a more accurate wavelength calibration. The output calibration file for the *identify* task was then used to generate calibration files for all other arcs of each night using the ***reidentify*** and ***fitcoords*** tasks in the *longslit* package with the initial calibration file and the new arc as input. A separate calibration file for each science exposure is the output used in the next step. The final step in this phase was running the ***transform*** task, also from the *long-slit* package, on each science exposure using their respective calibration output files to transform the x-axis in the data from the pixel indices into wavelengths across the dispersion axis.

3.2.4 Cosmic ray removal

Cosmic rays were handled using the L.A.cosmic[7] package. After wavelength calibration, the spectral images went through the cosmic ray cleaning process. The task used was ***lacos_im*** from the *stsdas* package. For each image, this task creates a mask that identifies islands of bright spots or line features with a pixel brightness higher than a specified multiple of standard deviations away from the average brightness around that spot. The cosmic ray locations are identified in each image, and they can be removed by the same task. This was convenient as all cosmic ray handling takes place with one command and the task can be rerun with different values for the detection sigma variable called *sigclip*. Each science image had its own value for *sigclip*, all images went through 4 iterations of cosmic ray detection and elimination to ensure smooth removal of cosmic rays without affecting the science data (or in some cases to minimise the effect on the data where cosmic rays hit inside the aperture to be extracted).

3.2.5 Aperture extraction

The aperture locations of the sources in each image were double-checked, as some of the spectra contained other sources that were dispersed onto the CCD. ***apall*** was the task used on all the science images with the same variable settings (width=5 maxsep=1000 lower=-4 upper=4 line=991 t_niterate=1 background=median

[6] https://topswiki.saao.ac.za/index.php/SPUPNIC
[7] van Dokkum (2001)

weights=variance gain="GAINVAL" readnoise="NOISEADU"). The correct aperture was selected, and the background subtraction area was selected such that it is close to the aperture to be extracted and as wide as possible to obtain a good average for the background without crossing the apertures of other sources dispersed in the image. The trace of the aperture is then fit, and the background-subtracted spectrum is extracted along the trace. The output is the extracted 1D spectrum of the source in the aperture.

3.2.6 Flux calibration

The final step in the data reduction recipe is the flux calibration. In this step, the y-axis is calibrated from counts to flux (in units of erg/s/cm^2/Å), as compared to the response of the telescope and instrument to the standard star observation of each night. SAAO has its own database of extinction coefficients for Sutherland, which can be included in PyRAF. A file named *suth_extinct* was provided and used in the ***standard*** task, from the *onedspec* package, along with information like the exposure time and magnitude of the star as input to create a file used as input to the ***sensfunc*** task. This secondary task outputs a calibration file and a sensitivity function file for calibration to the observed standard star. All the 1D extracted spectra files are then run through the ***calibrate*** task and the output from this task is the final reduced spectra of the observed sources.

3.2.7 Comments on the observing run

Of the 23 sources that were observed, 13 have spectra that closely match the spectra in the Pickles Stellar Flux Library database (Pickles 1998). There are also 5 unidentified sources (when using the identify tool on Aladin), 3 eclipsing binaries, 2 nebulae, and a candidate AGN. All of the 1D spectra (in Figs 3.5 to 3.27) have colour-coded lines for the spectra and labelled vertical bars for spectral emission/absorption line feature locations. The spectra in red are the observed objects, while those in teal, if present, are spectra from the Pickles stellar spectral database.

The most difficult observation was the source designated as GDB37 as it had a brighter source off to the west in the slit of the spectrograph, and when the time came for extracting the aperture of GDB37, there were multiple other faint sources dispersed in the 2D spectrum that were not visible on the acquisition camera on the night of observation. This led to only a small patch above the very bright source being usable as background subtraction during the spectral extraction process (see Figure 3.2).

The most interesting observation was the source designated as GDB40. It is identified as the candidate AGN 2E 238, and by visual inspection of its observed optical spectrum, the MeerKAT radio, and Chandra X-ray morphology (see Figure 3.3) it seems likely to be an AGN, or at least not a foreground star or nebula between the Earth and the inside of the SMC. The radio source is quite luminous and coincides with an X-ray source from the Chandra space telescope that shows extended morphology.

The S/N of the 1D spectra were generated with the Specutils[8] method using Python. As seen in the last column of Table 3.1 and in Figure 3.4, the S/N value tends to be higher for sources of lower apparent magnitude, as expected. There are a few sources that deviate slightly from this trend, most notably GDB 1, 2, 5, 41. This deviation may have been caused by the GAIA magnitude used, which is averaged over a broad wavelength range, because these four objects appeared brighter than their GAIA magnitudes suggested on the night of observation. Further, it is believed that all the sources have an over estimated apparent magnitude, again owing to the broad wavelength range over which the GAIA g_mean_mag is averaged (between 3300 and 10500 Å), and an under estimated S/N owing to the single long exposures instead of having multiple shorter exposures which could have been combined to lower noise and increase signal strength. The observations were all taken in one exposure per source, meaning that cosmic rays that happened to hit the aperture of the source to be extracted had to be dealt with manually. This introduces losses to the data if the location hit on the CCD was over a wavelength range used for identification of the source, and this will decrease the S/N of the spectrum. If the source was observed while the moon was up, this introduced some oxygen air glow atmospheric emission line features of varying strengths present in most of the spectra around the wavelengths of 5577 and/or 6300 Å (Smith 1976; Sunil Krishna & Bag 2013). These line features have interesting shapes;

[8] Specutils S/N function documentation

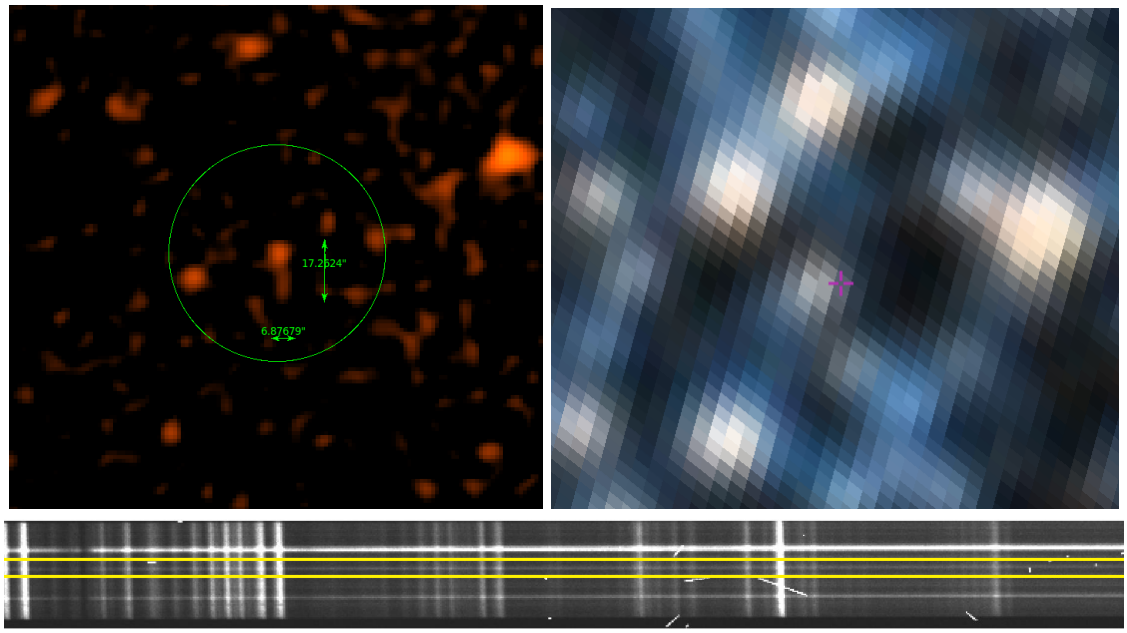


Figure 3.2: Source GDB37 MeerKAT radio view (top left) where the source is measured to be approximately 7 arcsec (East-West) by 17 arcsec (North-South) in size on-sky. The DSS2 optical view on Aladin is displayed in the top right image, and a magenta cross-hair marks GDB37, the physical scale is such that the optical image is as wide as the diameter of the green circle on the left image. The top images are oriented North-up and East-left. A part of the raw 2D spectra from SpUpNIC appears on the bottom in greyscale. The target aperture is the central one beneath the bright source aperture, in between the yellow lines.

some look like inverse P-Cygni features, while others are seemingly double-peaked or have a single small peak. This suggested that the oxygen emission line features might not be purely atmospheric and may be linked to the source being observed, however, this was found to be wrong and all subsequent lines found at these wavelengths (5577 and/or 6300 Å) are considered atmospheric artefacts. This is due to the dissociative recombination of molecular oxygen ions in the upper atmosphere, showing up on spectra as bright skylines that can bleed into faint source apertures and become difficult to remove during background subtraction. These skylines increase in intensity depending on how full the moon is on the night of observation and how close the observed target is to the moon's position on the celestial sphere (Smith 1976), and the observation week was a waxing gibbous moon week. These atmospheric skylines can also be enhanced by geomagnetic disturbances caused by solar activity, or geomagnetic storms in the worst case (Sunil Krishna & Bag 2013). The atmospheric line features, like the broad molecular oxygen absorption features around 6900 and 7550 Å in all spectra, affect the S/N depending on the line strength, along with the emission lines around 5577 and/or 6300 Å. The molecular oxygen in absorption is purely atmospheric and is ignored as it does not originate from the observed source. Some objects, which have spectra closely matched to those of stars, also have an H α emission feature, this can be present because of (or added to by) the H I high-density region glowing from star formation over portions of the SMC galaxy, accretion from certain types of binary systems, or stellar chromospheric activity and hot solar winds ionizing the environment around a star. This is discussed individually for each source later in this chapter.

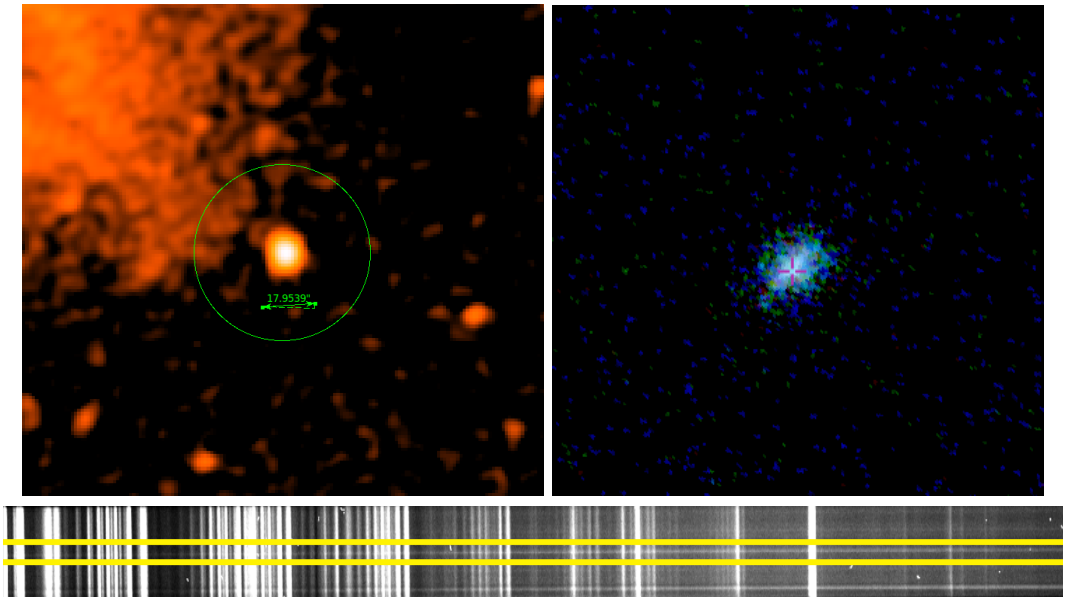


Figure 3.3: Source GDB40 MeerKAT radio view (top left) and the source is around 18 arcsec wide on-sky as indicated by the green distance ruler, Chandra X-ray view on Aladin (top right) where a magenta cross-hair marks GDB40. The top images are oriented North-up and East-left. The raw 2D spectra from SpUpNIC are seen at the bottom in the greyscale. The target aperture is the central one, in between the yellow lines.

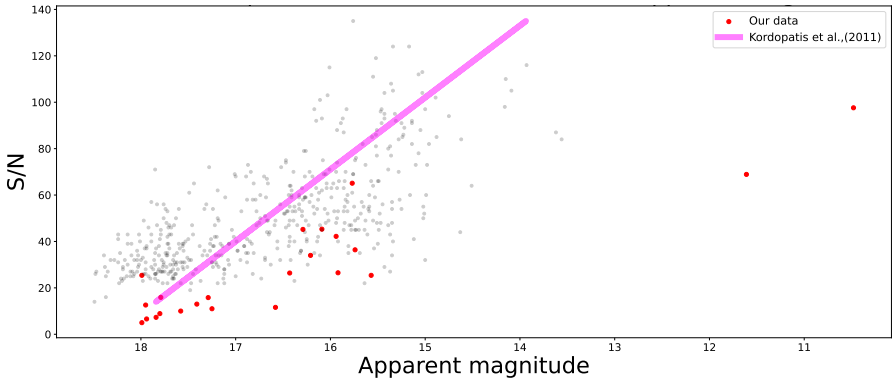


Figure 3.4: Above is a plot of the GAIA g_mean_mag value of optical sources observed vs the S/N of the extracted 1D spectrum represented by the red dots. The expected rise in S/N with a decreasing apparent magnitude is observed. Included in the graph are the Vmag (apparent magnitude) vs observed SNR values in black dots, with a linear best fit in magenta, from Kordopatis et al. 2011 with published spectroscopic data of 479 thick disk stars. There are 2 clear outliers in our data, they are the 2 with the highest S/N (GDB1 and GDB2), both were observed for a shorter amount of time during the observation night because they appeared brighter than their GAIA magnitude had suggested. This likely caused their S/N to be lower than it would have been if they were observed for the same (longer) amount of time as all the other sources.

3.3 Spectroscopic extraction results

Each spectrum of the 23 observed sources is presented and discussed in this section. The spectra in Figures 3.5 to 3.27 all have their y-axis on the logarithmic scale, and the flux values are continuum normalised.

3.3.1 GDB1

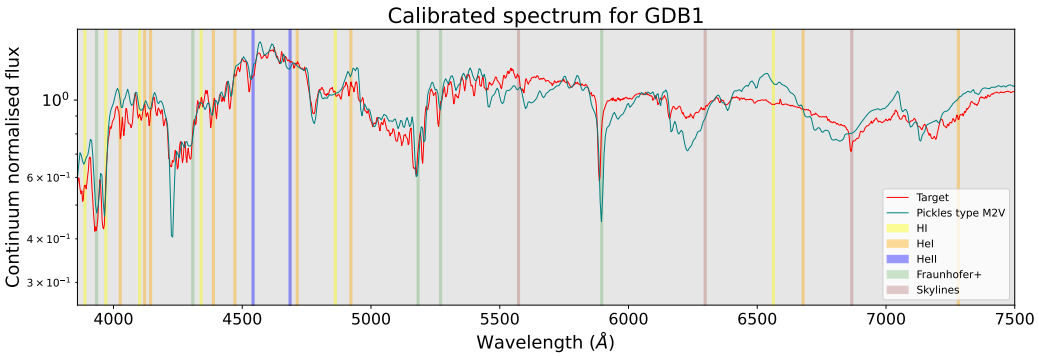


Figure 3.5: Spectrum of GDB1

GDB1 in Figure 3.5 is the brightest observation target, and its spectral features closely resemble the known spectrum of an M2V-type star. This was unsurprising as objects with low optical magnitude are most likely to be stars in our galaxy. From the GAIA magnitude of this star, the distance estimate using Equation 1.2 is ∼12.5 parsecs. It has a significant parallax angle of 21.1361 milliarcseconds (mas) (Gaia Collaboration 2020) corresponding to a distance of 47 parsec using Equation 1.1. GDB1 is a foreground star not in the SMC called "Flo 581" (McGowan et al. 2008), we estimated to be between 12 and 47 parsecs away. The on-sky separation between the radio and optical sources (see Figure 3.29), along with the knowledge that this is a high proper motion M dwarf, which often are radio-emitting stars (see Figure 1.4), suggests this may be the optical counterpart to the radio source. The proper motion of the star is likely the cause of the optical-to-radio source misalignment. The radio source was artificially brightened to reveal the radio morphology in Figure 3.29, with S/N from AEGEAN of 5.2.

3.3.2 GDB2

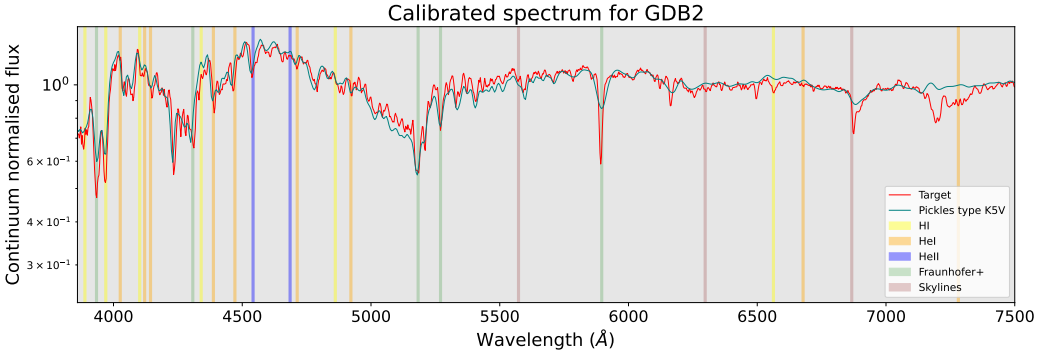


Figure 3.6: Spectrum of GDB2.

The spectrum of source GDB2 shown in Figure 3.6 above traces the Pickles archive spectra for a K5 dwarf star very well. The only notable deviation is the relatively deeper-than-normal absorption feature at the sodium doublet at 5890 Å. GDB2 is identified on SIMBAD as "UCAC4 091-002858" (González-Fernández et al. 2015), a K5V class star. A parallax angle of 11.2 mas (Gaia Collaboration 2020) is measured for this source by GAIA, meaning that this is a Galactic source around 90 parsecs away using Equation 1.1. The

distance estimate from the GAIA magnitude using Equation 1.2 is 73 parsecs, corroborating its identification as a Galactic K5 dwarf star. This optical source is believed to be the optical counterpart to the radio emission after visual inspection (see Figure 3.29), and the radio emission is likely due to chromospheric activity from the stellar surface.

3.3.3 GDB3

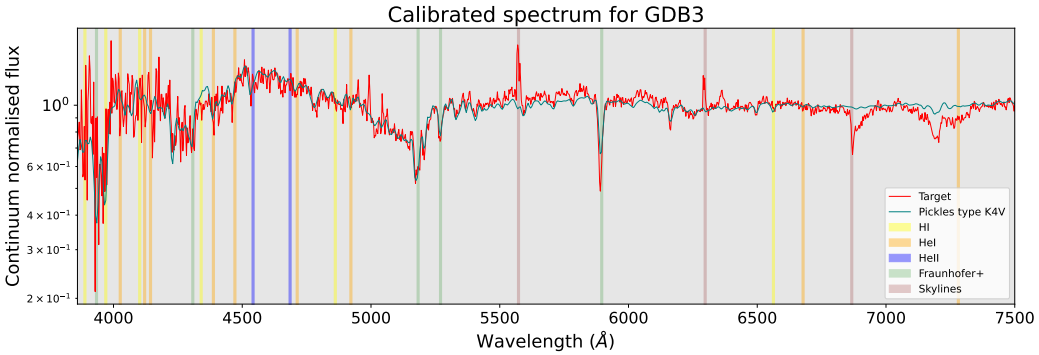


Figure 3.7: Spectrum of GDB3.

GDB3 is an unidentified optical source according to SIMBAD. The spectrum, in Figure 3.7, closely resembles that of a K4 dwarf star; more work can be done to confirm this with more confidence. Using Equation 1.2, the distance to this source is estimated to be ∼450 parsecs, making it a Galactic source. The radio source is directly aligned with the optical source (see Figure 3.29). This optical source is believed to be the optical counterpart to the radio emission. This would mean the radio emission is most likely due to chromospheric activity from the dwarf star.

3.3.4 GDB4

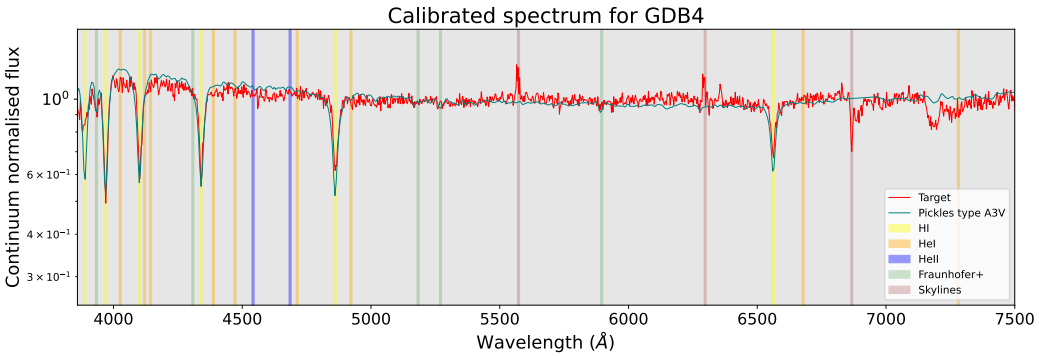


Figure 3.8: Spectrum of GDB4.

GDB4 in Figure 3.8 is an unidentified source according to SIMBAD. This spectrum very closely resembles that of an A3 dwarf star, and the alignment with the radio source is very good. The distance to this source is estimated to be between ∼5.5 and 10 kpc using Equation 1.2, and if correct, it is a Galactic source. This is believed to be the optical counterpart to the radio source (see Figure 3.29, where its brightness was enhanced to reveal the radio source). Some A-type stars do have magnetic fields strong enough to be able to produce gyro-synchrotron radiation, however, they usually have to be fast-spinning, be observed pole-on or be variable stars indicating flare activity (Babcock 1958).

3.3.5 GDB5

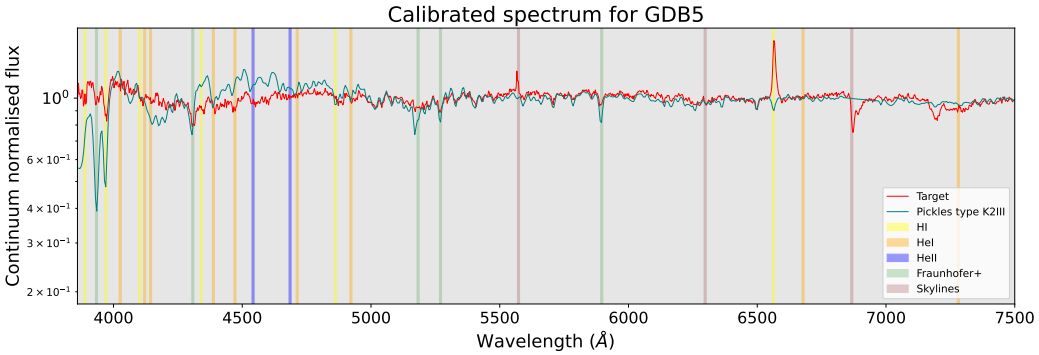


Figure 3.9: Spectrum of GDB5.

The optical source is unidentified on SIMBAD. GDB5 in Figure 3.9 resembles the spectrum of a K-type giant star. However, a large deviation is seen in the presence of H α emission, possibly indicating an active chromosphere and the absence of the expected magnesium and iron Fraunhofer absorption features. Distance estimates using Equation 1.2 would put it at a distance between ∼12 and 15 kpc as a Galactic source. The spectrum shows an emission line at H α, however, it is not situated in an area of high H I density (see Appendix C for the locations of observed targets relative to the H I column density detection around them) and so the H α is believed to originate from the source itself. This source could be the optical counterpart to the radio emission (see Figure 3.29) if the morphology of the radio emission can be linked to the optical source. The FWHM of the H α line is 318 km/s and has an equivalent width of 34 Å.

3.3.6 GDB6

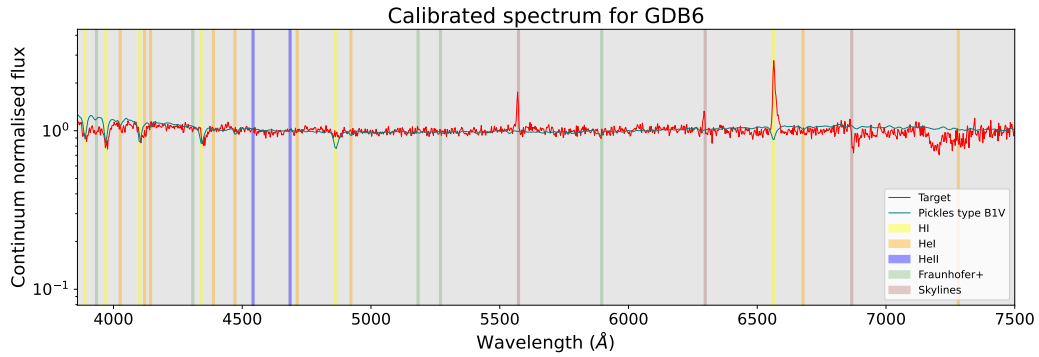


Figure 3.10: Spectrum of GDB6.

GDB6, Figure 3.10, is unidentified on SIMBAD. The H α emission line could be present due to the source being inside a region of high H I density (see Appendix C), or it could originate from the source seeing as the width of the H α and other Balmer lines are similar and indicate an active chromosphere to explain the radio emission from this source. Compared to the Pickles spectrum, the Balmer lines of the spectrum suggest this optical source is likely an early to mid-B type dwarf star. Using this assumption and Equation 1.2, a distance estimate puts this source between 25 and ∼100 kpc, within range of the SMC if the spectral type is different, but also possibly a Galactic source. A better constraint on the apparent magnitude will make the distance estimate more precise. Good alignment between radio and optical sources (see Figure 3.29, note the brightness of the radio image was enhanced for visual purposed) indicates this may be the optical counterpart to the radio emission, but there remains uncertainty due to the optical surroundings being very

densely packed with optical sources. The FWHM of the H α line is 302 km/s and has an equivalent width of 25 Å.

3.3.7 GDB7

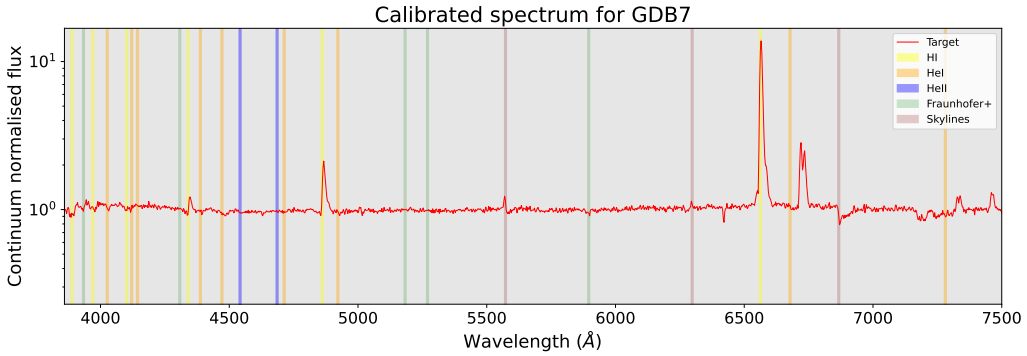


Figure 3.11: Spectrum of GDB7.

GDB7, in Figure 3.11, has a direct alignment with the radio source. From its identification as an eclipsing binary with strong H α emission, and the very good alignment between radio and optical sources, it is believed to be the optical counterpart to the radio source (see Figure 3.29). The integrated radio flux of this source from AEGEAN is 390μJy. The strong H α and H β emission indicate that there is likely an accretion disk with interactions in the binary ionising the surrounding hydrogen in and around it. This emits the radio waves seen as the bright radio source. The optical source is identified on SIMBAD as an emission line star "LIN 422" in the SMC as part of an eclipsing binary (Meyssonnier & Azzopardi 1993). The FWHM of the H α line is 306 km/s in this spectrum with an equivalent width of 27 Å.

3.3.8 GDB8

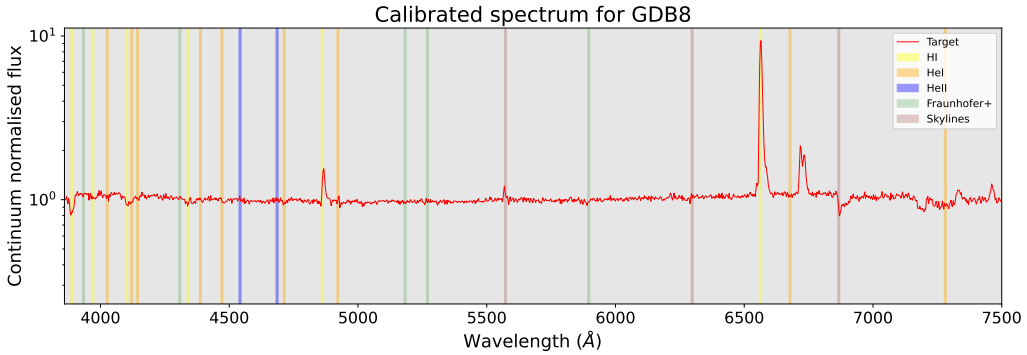


Figure 3.12: Spectrum of GDB8.

SIMBAD has the optical source aligned with GDB8 in Figure 3.12 identified as an emission line star "LHA 115-N 79" as part of an eclipsing binary (no spectral class is given, Meyssonnier & Azzopardi 1993). Some of the Balmer absorption lines indicate that this star may be close in spectral class to a B-dwarf star. This optical source is believed to be the optical counterpart to the radio emission (see Figure 3.29, the radio image was enhanced to reveal the source). The FWHM of the H α line is 312 km/s and it has an equivalent width of 28 Å. Furthermore, the source coordinates are within a high-density H I region, which may be contributing to the H α emission line strength.

3.3.9 GDB9

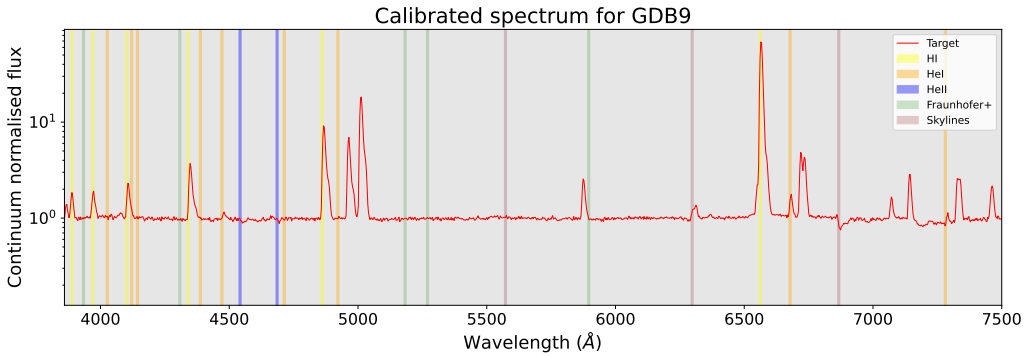


Figure 3.13: Spectrum of GDB9.

The radio emission of GDB9, in Figure 3.13, and the optical emission observed are aligned. The source is inside a region of very high H I density and is identified on SIMBAD as a star-forming region/nebula "LHA 115-N 71" (Livanou et al. 2007) in the SMC. Many nebular lines are present and strong in the spectrum, along with the Balmer emission lines expected from star-forming regions (the two unmarked lines around 5000 Å, are O III emission lines). GDB9 is believed to be the optical counterpart of the radio source (see Figure 3.29). The FWHM of the H α line is 340 km/s and has an equivalent width of 25 Å.

3.3.10 GDB10

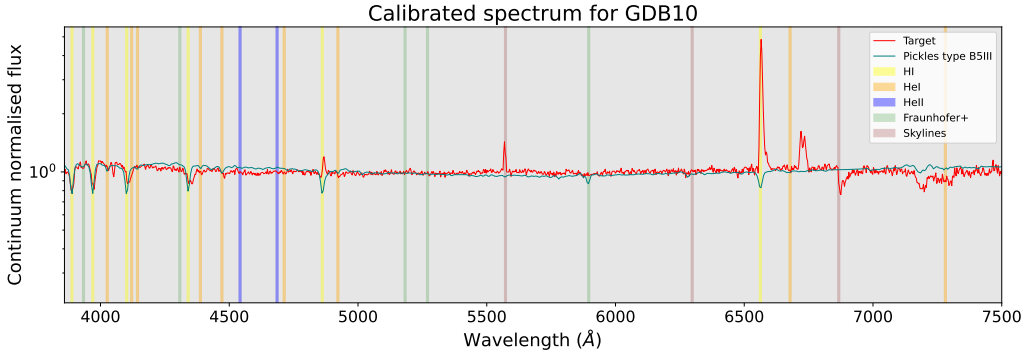


Figure 3.14: Spectrum of GDB10.

The radio and optical sources are directly aligned for GDB10 in Figure 3.14. SIMBAD identifies this optical source as an eclipsing binary "OGLE SMC-ECL- 8019" (Pawlak et al. 2016). Comparing the observed spectrum to that of a B5III type star from Pickles, the observed star is likely a mid-B type giant star. Distance estimates place this source ∼55 kpc away with Equation 1.2, and Pawlak et al. (2016) has the magnitude as Vmag=18.381, placing it beyond the SMC at 120 kpc. Interactions between the stars in the binary may alter the apparent magnitude, and the formula is not meant to be used for binary systems. More work is needed to constrain the distance of this object. GDB10 is believed to be the optical counterpart of the radio emission (see Figure 3.29). This source is located in a region of moderate H I density, which is likely contributing to the emission strength of the H α line due to active star formation in that region, or strong solar wind activity from the binary itself. The FWHM of the H α line is 390 km/s and has an equivalent width of 23.5 Å.

3.3.11 GDB13

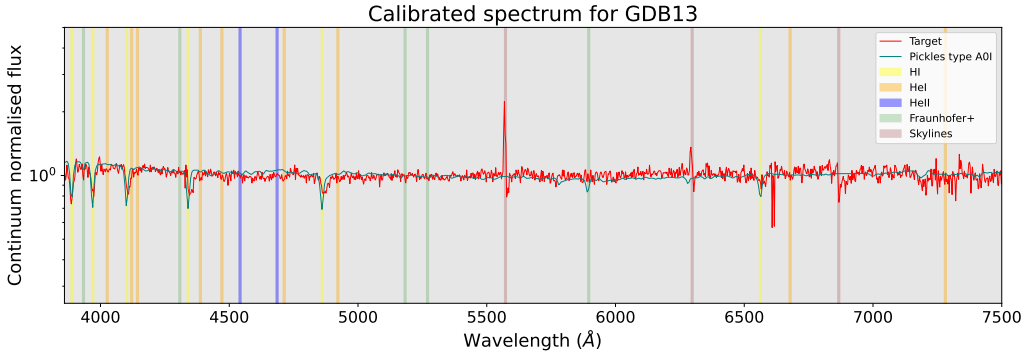


Figure 3.15: Spectrum of GDB13.

The alignment between the radio and optical sources is good for GDB13 in Figure 3.15. The radio source has a S/N of 6.5 in the AEGEAN catalogue (see Figure 3.29, this radio image was enhanced to reveal the source). This optical source is unidentified on SIMBAD, and from the spectrum comparison, it suggests a supergiant star of class A or B. However, distance estimates with its GAIA magnitude and Equation 1.2 place it between 460 and 670 kpc. An A0 supergiant in the SMC would have an apparent magnitude ∼12. It can not be believed that it is a supergiant; more likely an A0II or A0III star. Because of the uncertainty on whether the radio emission is real, it can not be said with confidence that this optical source is the counterpart of the radio emission.

3.3.12 GDB14

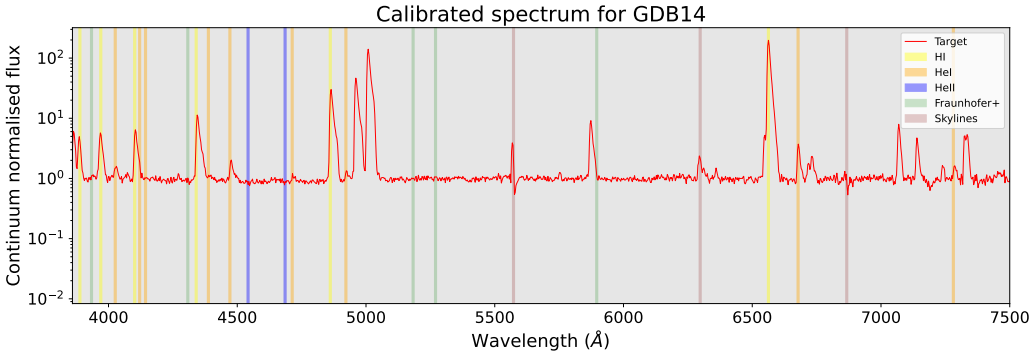


Figure 3.16: Spectrum of GDB14.

There is good alignment between the radio and optical source for GDB14 in Figure 3.16. It is identified as a planetary nebula "SMP SMC 24" (Meyssonnier & Azzopardi 1993, however, the source itself is not reported among the SMP SMC objects in this paper linked to the PN) and is in a region of high H I density. The observed spectrum's strong nebular and Balmer emission lines support this classification (the two unmarked lines around 5000 Å, are O III emission lines). GDB14 is also believed to be the optical counterpart to the radio emission and not a chance alignment (see Figure 3.29). The FWHM of the H α line is 280 km/s and has an equivalent width of 31 Å.

3.3.13 GDB21

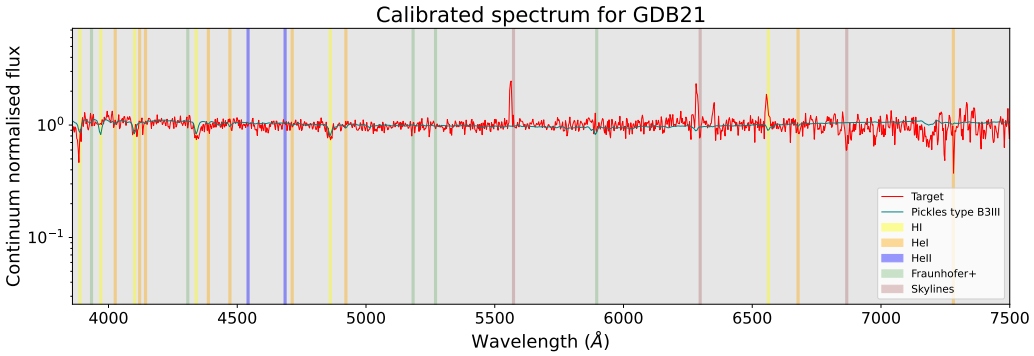


Figure 3.17: Spectrum of GDB21.

The alignment between the radio and optical sources is good for GDB21 in Figure 3.17. The optical source is unidentified on SIMBAD and shows similarities to the spectrum of a B3III star. However, the S/N in the spectrum is not sufficient to constrain the spectral class very well. Estimates place it between 22 and 85 kpc using Equation 1.2. This distance estimate does not constrain the position to being a Galactic source or in the SMC. The source is in the middle of a high H I density region that likely contributes to the strength of the H α emission line. The presence of H α emission could indicate that the radio emission is from an active chromosphere. The optical neighbourhood is not very dense with optical sources. With all this considered, GDB21 is likely the optical counterpart of the radio emission (see Figure 3.29).

3.3.14 GDB23

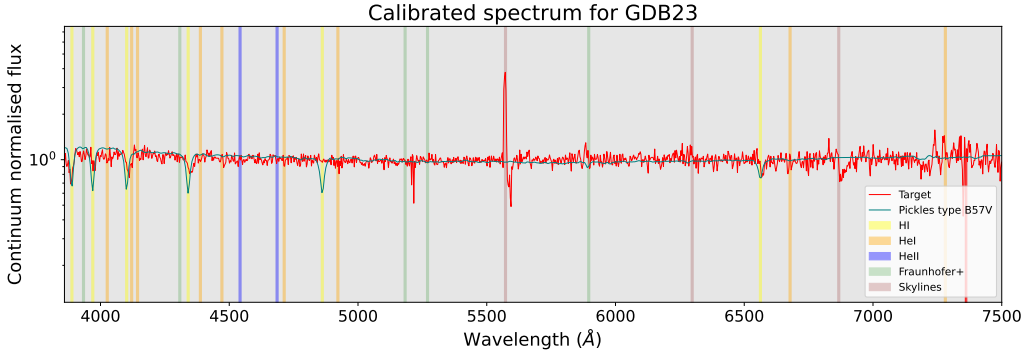


Figure 3.18: Spectrum of GDB23.

GDB23 also has good alignment between the radio and optical sources, its spectrum is in Figure 3.18. The optical source is unidentified and traces the spectrum of a B5 dwarf star quite well, despite the lower S/N ratios expected in optical sources with higher apparent magnitude values. This source is aligned with a region of high H I density but shows no H α emission, so this may be a Galactic B-dwarf at an estimated distance of $\sim$47 parsecs using Equation 1.2 and the GAIA magnitude. This is likely a foreground dwarf star of late B spectral type between us and the SMC. From the direct alignment between radio and optical sources (see Figure 3.29, this radio image was enhanced to better visualise the source), along with the resemblance of the spectrum to a B-dwarf spectrum, it might be the optical counterpart if the source is chromospherically active or has strong enough solar winds to ionise its surrounding medium.

3.3.15 GDB27

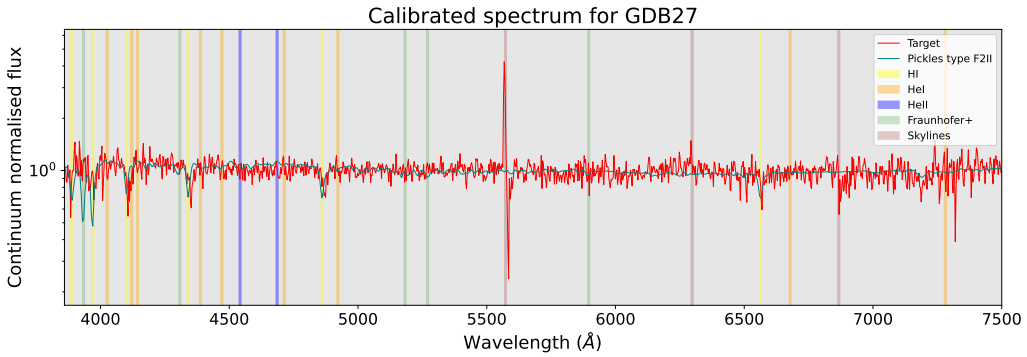


Figure 3.19: Spectrum of GDB27.

Low S/N in the spectrum for GDB27 (Figure 3.19), this unidentified optical source is well aligned with the radio source and is likely a star in the SMC with uncertain spectral class (between an F to G giant to sub-giant star at a distance between ∼85 and 95 kpc using Equation 1.2). A better constrained apparent magnitude could place it close enough to be in the SMC. The Balmer absorption lines seem slightly broadened. It is in line with a region of high H I density in the SMC but has no H α emission. GDB27 is not believed to be the optical counterpart of the radio source due to the slight misalignment as seen in Figure 3.29 (note that in the figure mentioned, the radio image was enhanced to reveal the source for better visualisation), and the very high density of optical sources in the region. The Calcium H&K line is also very weak, this line weakening may have happened during extraction, as the other lines do match up well with the strengths expected from an F2II star.

3.3.16 GDB30

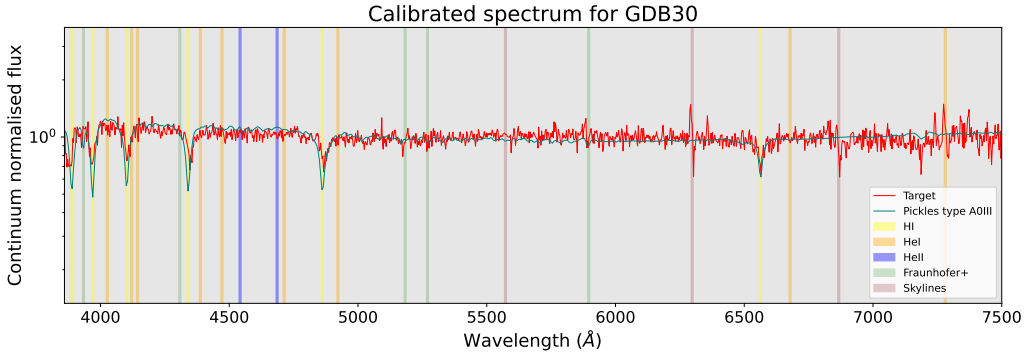


Figure 3.20: Spectrum of GDB30.

The optical source is unidentified for GDB30 in Figure 3.20, likely an early to mid-class A giant or sub-giant (distance estimate between ∼34 and 40 kpc using Equation 1.2) Galactic star when looking at the spectrum comparison and taking into account that no H α emission line is present while the source is aligned with a region of high H I density in the SMC. GDB30 is believed to be the optical counterpart to the radio emission. This could be a binary system when looking at the optical morphology in Figure 3.29 (the radio image was enhanced to better visualise the source).

3.3.17 GDB32

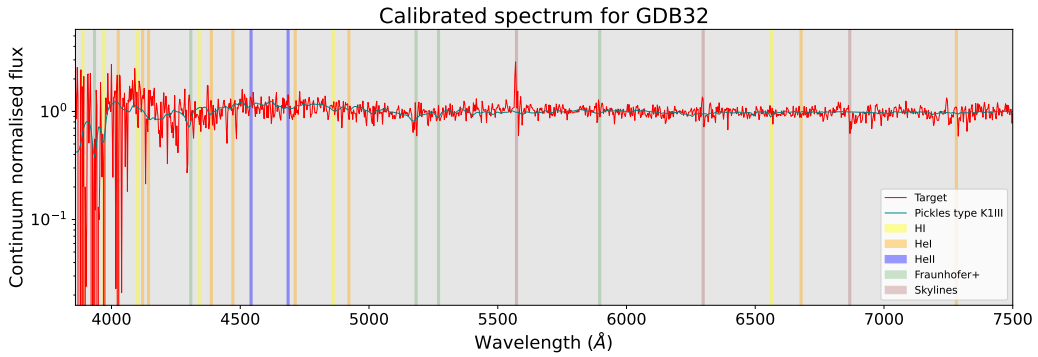


Figure 3.21: Spectrum of GDB32.

The spectrum of GDB32 (Figure 3.21) somewhat follows the shape of a giant K-type star, but there is no clear correlation between emission and absorption features, and the continuum levels are very far apart. As the S/N is relatively low in the observed spectrum, and the radio source image had to be enhanced to be able to see it (AEGEAN has the radio S/N at 5.1 and the source comparison can be seen in Figure 3.29), there is uncertainty on whether this is the optical counterpart of the radio emission.

3.3.18 GDB41

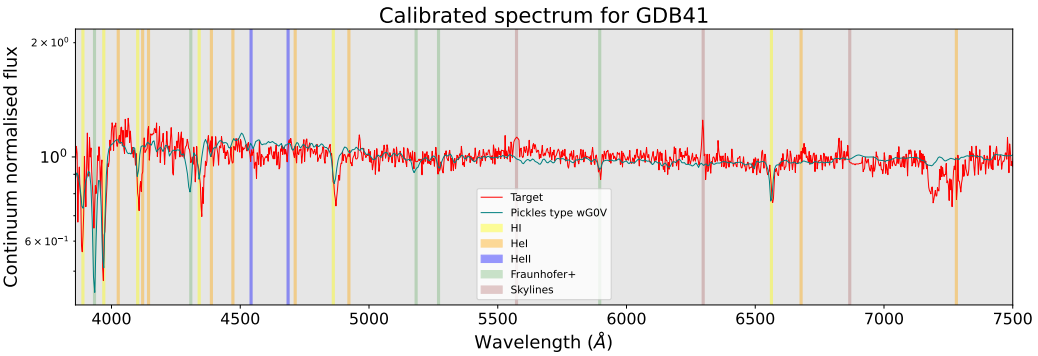


Figure 3.22: Spectrum of GDB41.

For GDB41 (Figure 3.22) the optical source was much easier to find than the finder charts had predicted, and it is not exactly aligned with the radio emission (AEGEAN gave a radio S/N of 5.6) when looking at GDB41 in Figure 3.29 (this radio image also had to be enhanced to show the source). This may be a foreground variable star in our galaxy (the distance is estimated to be $\sim$5 kpc with Equation 1.2) caught during a bright optical flare during the night of observation. The optical S/N is high for the optical source with the highest apparent magnitude (from GAIA) that made the observation cut, and the spectrum has features similar to what is expected of an early G-type dwarf star, with a deviation in the calcium H & K line strength notably present. This is most likely a chance alignment with the radio source unless the proper motion of the optical source is high enough for it to have moved between the dates of the optical and radio observations, causing the apparent mismatch in source locations.

The following are very low S/N spectra of which nothing can be said with confidence

3.3.19 GDB25

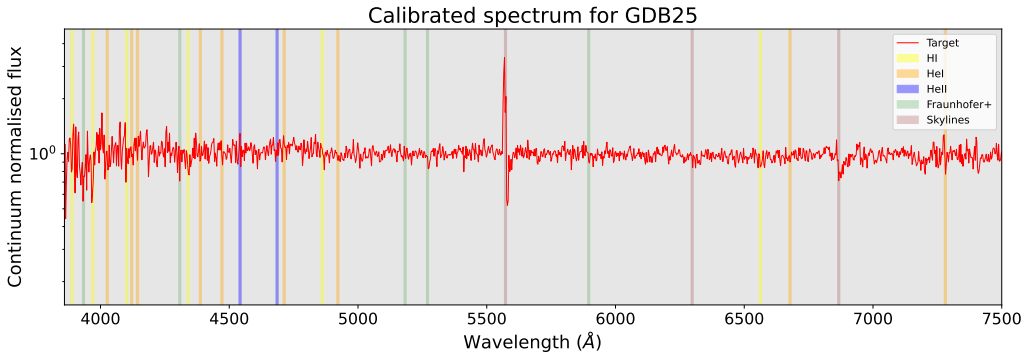


Figure 3.23: Spectrum of GDB25.

The alignment is precise between the radio and optical source for GDB25 in Figure 3.23. However, the S/N is too low to say anything about this source with any confidence. Additionally, the optical neighbourhood has low source density. This source is worth another observation or multi-wavelength analysis. GDB25 is believed to be the optical counterpart to the radio source (see Figure 3.29).

3.3.20 GDB34

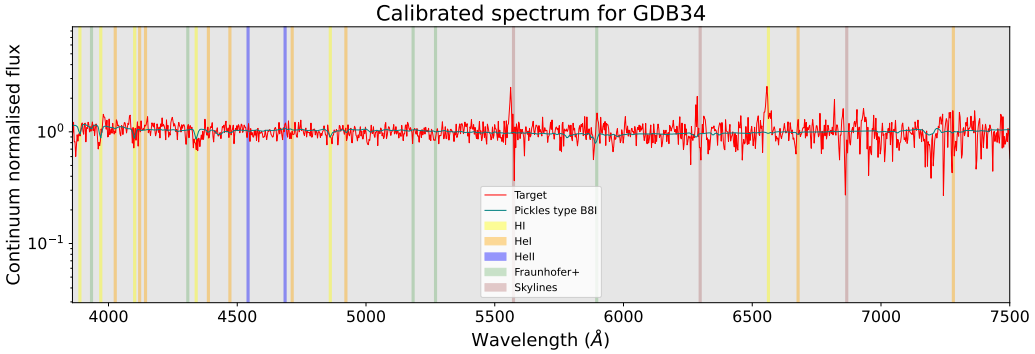


Figure 3.24: Spectrum of GDB34.

The optical S/N in GDB34 (Figure 3.24) is low. While having similar shape and spectral line features when compared to the B8I star spectrum from Pickles, it is unclear what the spectral class might be because of the noise and the low spectral resolution. Whether it is inside the SMC or between us and the SMC is also unclear. What is a bit peculiar is that it has seemingly broadened H α and Hγ line features but is in a low H I density region of the SMC. Due to the direct alignment of the optical and radio sources, it is believed that GDB34 is the optical counterpart of the radio emission. The cause of the radio emission from a hot, large and luminous star around B8I class might be strong solar winds or possible binary interactions.

3.3.21 GDB37

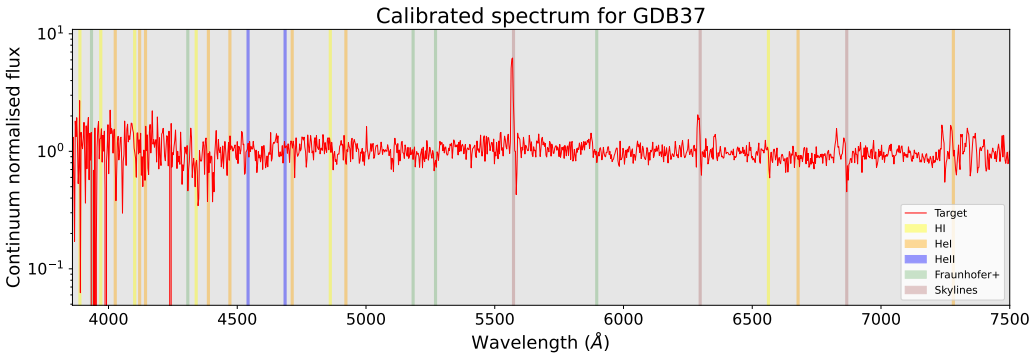


Figure 3.25: Spectrum of GDB37.

The optical source at the GDB37 coordinates (Figure 3.25) is slightly misaligned with the location of the radio source. The optical and radio (AEGEAN radio S/N is 5.8) S/N is relatively lower when compared to the other radio sources in this sample; nothing can be said about this source with any confidence except that it is most likely a chance alignment with a radio source in a very dense optical source field.

3.3.22 GDB38

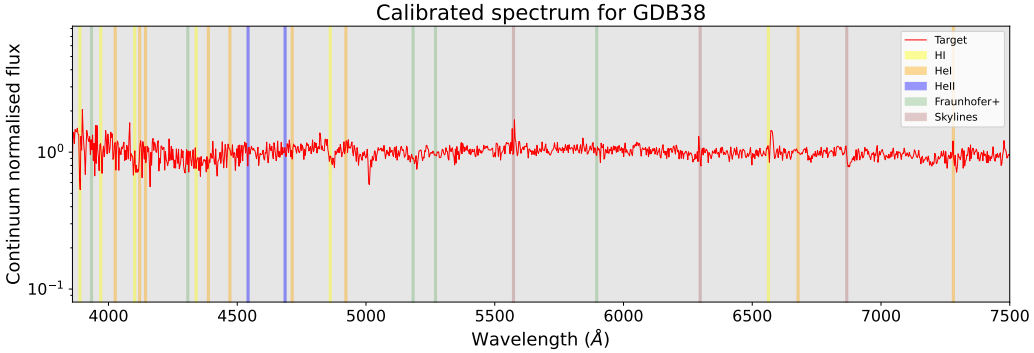


Figure 3.26: Spectrum of GDB38.

GDB38 (Figure 3.26) is an extremely noisy spectrum. A FWHM of the H α line was measured at 302 km/s and has an equivalent width of 16 Å, however, these measurements may be very inaccurate. The optical and radio sources are precisely aligned, with the radio source being noticeably larger than the optical source. Further investigation should be considered on this source to characterise and possibly identify it. GDB38 is believed to be the optical counterpart to the radio source, and due to its similarity to the spectrum of GDB40, and the presence of what looks like radio jet hot spots close to the main radio source, this could be an AGN.

3.3.23 GDB40

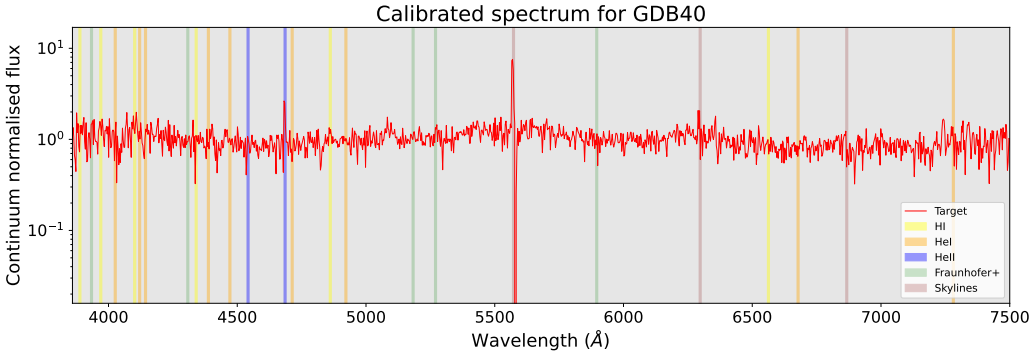


Figure 3.27: Spectrum of GDB40.

GDB40 (Figure 3.27) has a small, faint optical source aligned with a larger, bright radio point source. There is one apparent He II emission line around 4685 Å with no cosmic ray hits indicated nearby. However, the S/N is very poor, and that emission line is narrower than the resolution of the grating (The FWHM of the He II line is 7 arcsec and the resolution is between 11 and 14 arcsec), casting doubt on its believability. This spectrum does not share a similarity with any of the Pickles spectra. It is identified as an AGN candidate (Sturm et al. 2013) on SIMBAD, and the bright radio source can be further evidence of the radio source being an AGN. GDB40 is believed to be the optical counterpart to the candidate AGN radio source.

3.4 Observational conclusions

There are 5 sources with very noisy spectra and low S/N ratios, and nothing concrete could be deduced regarding the characterisation or classification of these sources from the spectra alone. These sources include GDB25, whose strongest feature is the peculiar inverse P-Cygni around 5577 Å, which has too much uncertainty surrounding it to be considered as emission detected from the source. GDB34 has an emission line at Hα, and an absorption line at Hγ that both seem broadened. The spectrum closely matches the Pickles spectrum Balmer lines of a B-type supergiant star within the noise, however, the mean magnitude reported from GAIA is 17.8 while the expected apparent magnitude (using Equation 1.2) of a B supergiant class star is between 12 and 13 using an SMC distance at 65 kpc and absolute magnitudes between -6.9 and -5.7[9]; GDB37 has no identifiable spectral features. From Figure 3.29, it is clear that the radio source is either faint or not real, and the optical source was also close to the magnitude limit for the SAAO 1.9-m telescope used for observation. GDB38 has Hα in emission, the rest of the spectrum is noisy, and the only other useful data point is that the radio source is larger in morphology than the optical source on the sky when comparing between the SSS optical image and the image produced with MeerKAT (see Figure 3.29). It may be a compact extended radio source like an AGN; GDB40 has one emission line around HeII, and it is identified as an AGN candidate on SIMBAD by (Sturm et al. 2013). The spectrum for that source collected in this MSc project is too noisy to say anything with confidence.

Since the SMC has a high density of optical sources, most of which are stars, the likelihood of chance alignments between the radio sources and the cross-matched optical source from GAIA is not negligible. Visual inspection of the alignment between the radio and optical sources was performed and can be seen in Figure 3.29. For each source, it was decided whether a chance alignment was likely or not, based on how close the radio and optical sources appeared visually and how densely the optical neighbourhood was, while also comparing the coordinate locations between the AEGEAN catalogue and GAIA.

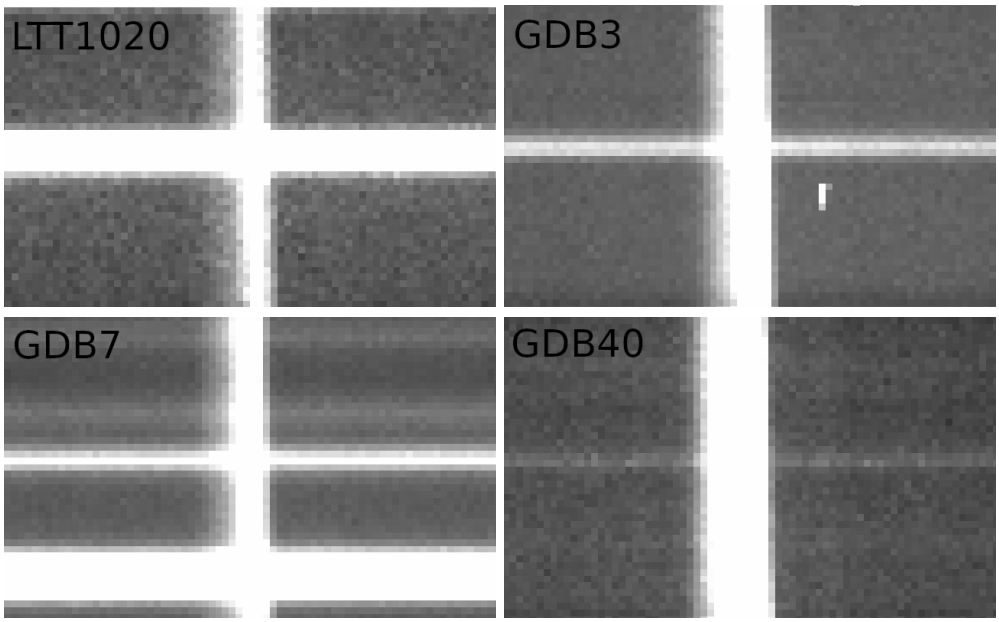


Figure 3.28: Comparison of the 5577 Å atmospheric line (vertical, crossing source apertures) width between 4 spectra. Standard star LTT1020 (top-left, observed for 300 seconds), GDB3 (top-right, observed for 40 minutes), GDB7 (bottom-left, observed for 40 minutes), GDB40 (bottom-right, observed for 40 minutes). All spectra cut-outs are at the same physical scale. Note the increase in skyline width and saturation with decreasing source flux.

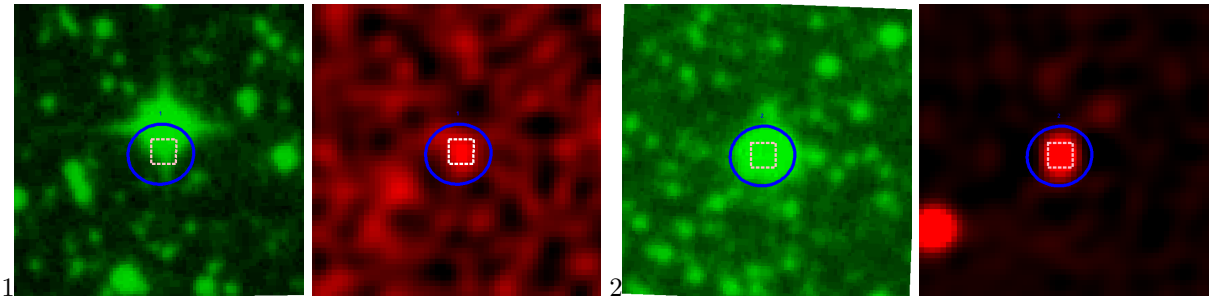
All spectral features around 5577 and 6300 Å are considered to be artefacts. This is due to several factors

[9]Absolute magnitude reference values are from the Handbook of Space Astronomy and Astrophysics

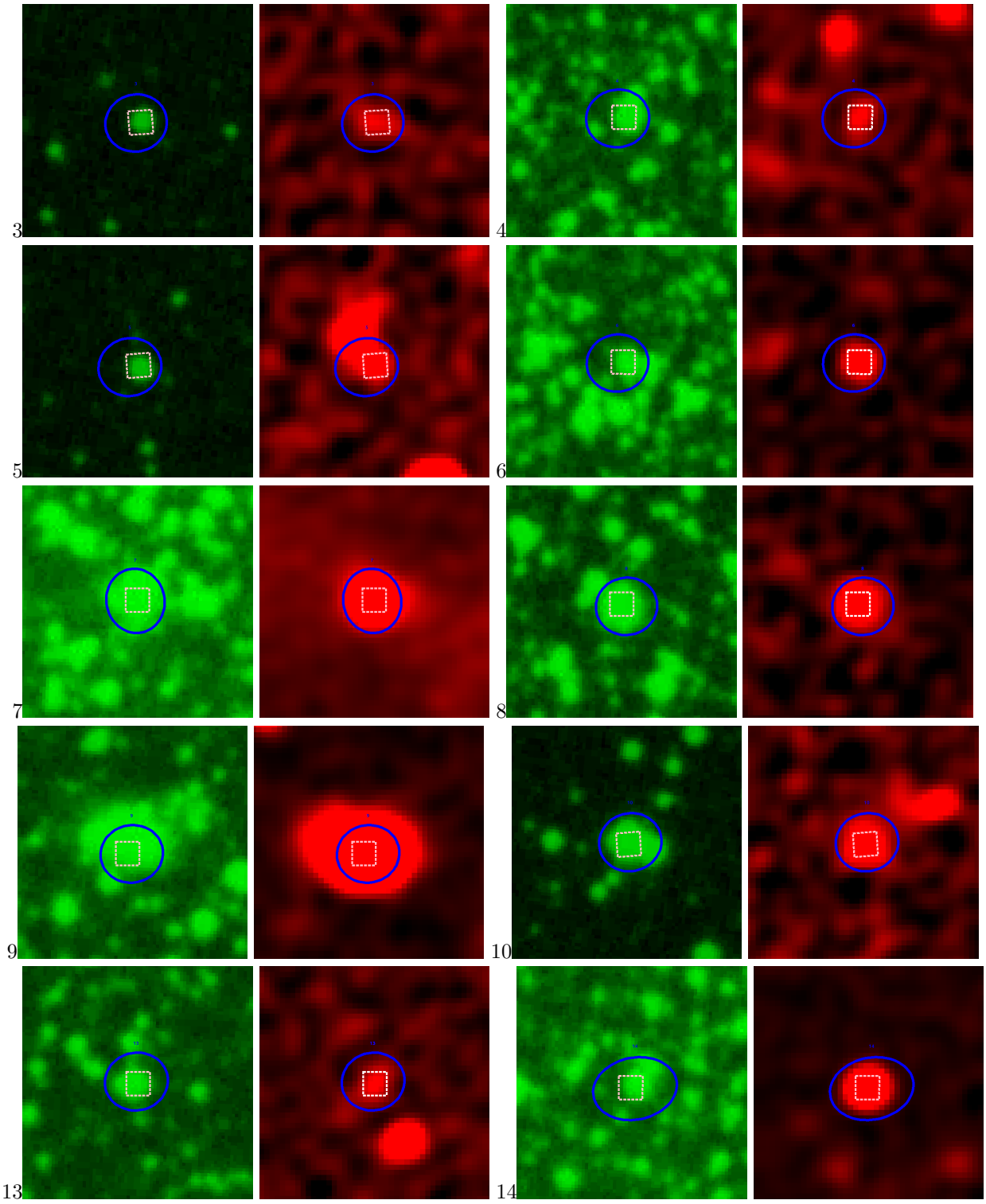
which increase uncertainty on whether the features are from the source in the form of OI and FeI emission lines, or if they are artefacts from bright atmospheric emission lines or background subtraction errors. There is a combination of things causing uncertainty:

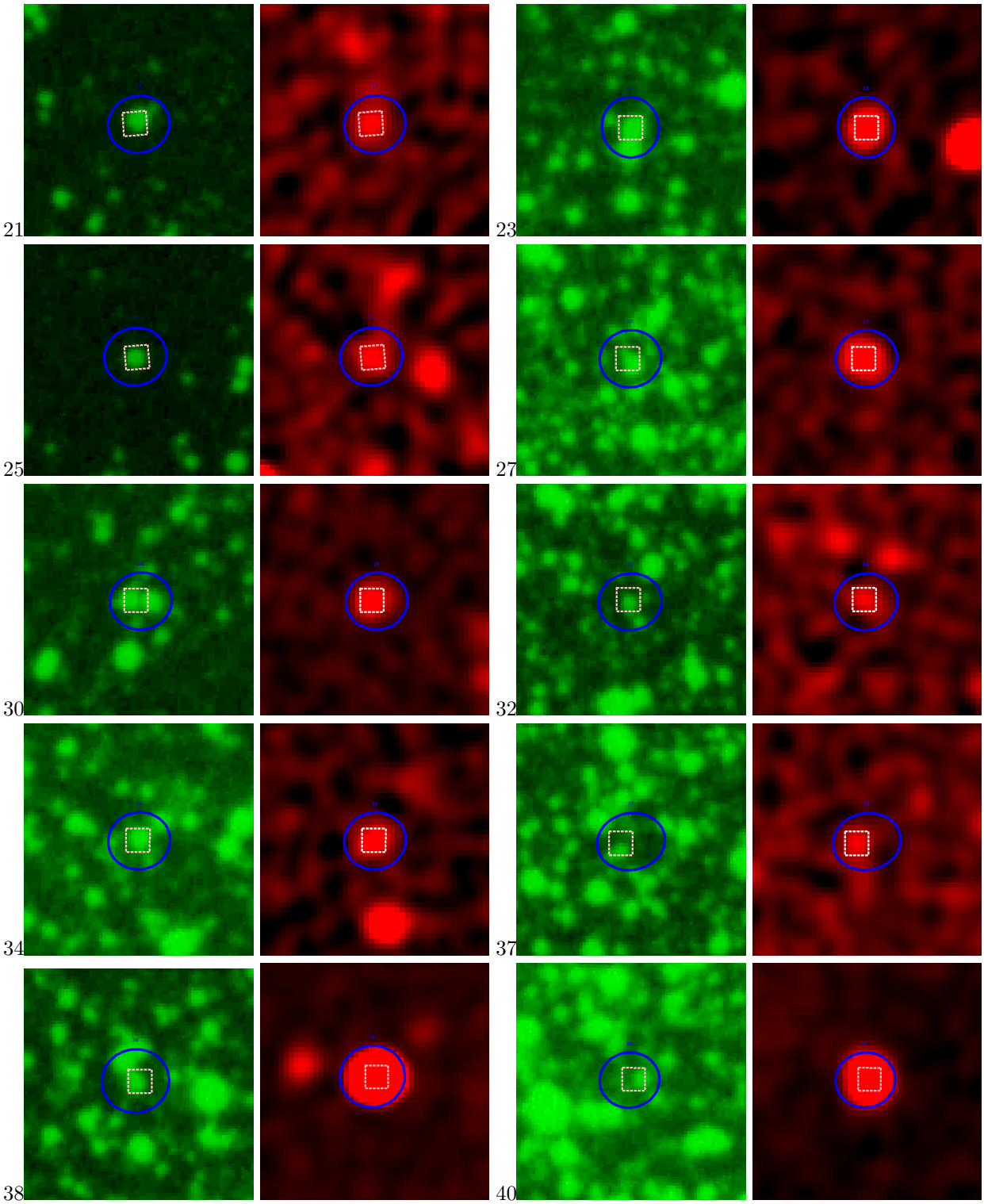
- 5577 and 6300 Å is around the locations of the brightest atmospheric lines (see Figure 3.28).
- There is a bleed-over effect along the dispersion and spatial axis around those atmospheric lines stemming from the difference in brightness from the observed source, vs the atmospheric night glow. This effect can be compounded if the observed source is faint and the moon is bright (Smith 1976). This sometimes led to incorrect subtraction of background pixels and the illusion of inverse P-Cygni features forming in some of the spectra.
- Moderate levels of solar activity during the week of observation could have caused geomagnetic disturbances, which increased atmospheric airglow effects (Sunil Krishna & Bag 2013). The increase may not have been a major factor in the airglow increase as the solar activity indices, as recorded by the NOAA[10], indicate moderate geomagnetic disturbances in the middle latitudes (see Appendix D for a verbatim copy of the NOAA data).

Sources that have identifiable spectral features and higher S/N ratios are classified into 13 stars (GDB 1, 2, 3, 4, 5, 6, 13, 21, 23, 27, 30, 32, 41), 3 eclipsing binaries (GDB 7, 8, 10), and 2 nebulae (GDB 9, 14) based on their spectral features and comparison to Pickles spectra of stars where appropriate. Of the 23 sources observed, 21 had an emission line around 5577 and/or 6300 Å, which are locations close to atmospheric airglow from oxygen, O I and Fe I lines. The separation in rest wavelength between these spectral lines is less than the spectral resolution of the grating used for observation (<5 Å with the spectral resolution being between 11 and 14 Å from FWHM measurements on ARC spectrum peaks). GDB1 and GDB2 did not have any emission around these wavelengths, and they are known Galactic stars identified in SIMBAD. This is likely due to the sources being quite bright, minimising the airglow background effects. These emission features sometimes took the shape of an inverse P-Cygni feature; however, after closer inspection, these are artefacts stemming from the observational conditions and data reduction methods used and not true inverse P-Cygni features. These features could be explained as Fe I emission lines blended with O I emission and atmospheric airglow since the spectral resolution cannot resolve the features individually, combined with the over-subtraction (red of 5577 Å) of the broad, bright atmospheric skyline. This skyline lies across the entire wavelength range where Fe I and O I lines could be detected. It was likely broadened because of the long exposure times of 30 to 40 minutes for each source, causing a dispersion bleed-over effect on the CCD data recorder around bright atmospheric features (see Figure 3.28). 13 of the observed sources (high and low spectrum S/N) are believed to be the optical counterparts to the respective radio sources (see Table 3.1). Some errors that could arise during the wavelength calibration process due to the low spectral resolution made the redshift and spectral line broadening measurements uncertain; hence, the measurements of FWHM and equivalent widths should be taken as ballpark values.



[10] The NOAA data before, during and after the week of observation, are publicly available.


3
4
5
6
7
8
9
10
13
14


21
23
25
27
30
32
34
37
38
40

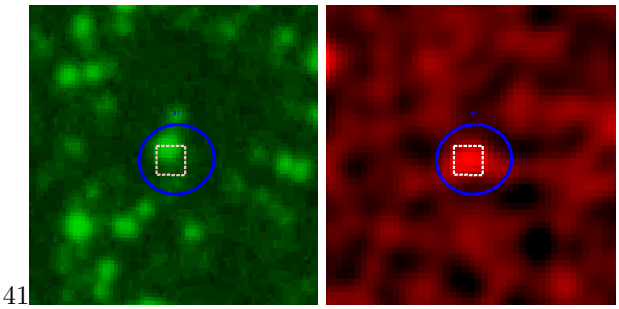


Figure 3.29: This figure shows a side-by-side comparison of the optical (green, from SSS survey) and radio (red, from MeerKAT) images at the AEGEAN radio coordinates (blue ellipses) along with the GAIA cross-matched source coordinates (white/pink boxes) for each of the 23 spectroscopically observed sources. The number labelled to the bottom left of each pair indicates the GDB designation number of the source as given during this observation run. Note that the radio images for GDB1, 4, 6, 8, 13, 23, 27, 30, 32, 34, 37 and 41 were artificially enhanced to better visualise those sources. This was done because the original images were very dark, and the sources could not be seen properly during comparisons.

Chapter 4

Discussion and Conclusions

This chapter presents a summary discussion of the final results, possible future work, and conclusions.

4.1 Results discussion

4.1.1 Citizen science

From the initial 10873 radio sources in the AEGEAN catalogue, 1429 were selected based on having a radio S/N greater than 20 for use as classification subjects in a Zooniverse project[1]. We separated the sources into individual tables for radio "point", "compact extended" and "structured extended" morphology, where > 80 per cent of classifiers agreed on the radio morphology of the source in question. This was based on the radio flux contour shape and size compared to the radio beam size overlaid on the optical image at the AEGEAN coordinates for that source.

The final counts resulted in 894 of 1429 being morphologically characterised as either a radio point source, compact extended source, or structured extended source with high confidence ($\sim$63 per cent of the sample). This number was somewhat higher than expected since the high confidence tables were constructed in such a way that > 80 per cent agreement meant that all 5 classifiers agreed if 5 classifications were made on a subject; 5 or 6 agreed if there were 6 classifications (There were 55 subjects with 6 total classifications in the high confidence tables) on a source; 6 or 7 out of 7 classifiers agreed if there were 7 classifications on a subject (There were 3 subjects in the high confidence tables that had 7 total classifications). This is also why the > 80 per cent criteria was chosen, always to have at least 5 people agree on the answer to a question, and standardise what is meant by "high confidence". The reason for all subjects not having 5 classifications is due to a change in retirement criteria from 10 classifications to 5 in July of 2024, after realising that the project will not be finished in time for results analysis. Zooniverse does not automatically check which subjects have reached the retirement criteria directly after it is changed. The recorded data keeps all classifications from before a retirement criteria change, plus one after another classification is made, triggering a retirement criteria check on that subject. The tables constructed by this analysis resulted in the characterisation of 634 radio point sources, 128 compact extended sources, and 132 structured extended sources (see Chapter 2 for details).

4.1.2 Optical spectroscopic follow-up

From the initial 10873 AEGEAN catalogue of radio sources, 41 were selected based on having a GAIA optical source within 1 arcsec of the AEGEAN coordinates. Additionally, a GAIA g_mean_mag < 18 criterion for the optical source cross-matched was imposed, as constrained by the telescope specifications used for spectroscopic follow-up. 23 of those 41 sources were observed during the allocated observation week, with priority given to those sources with no previous identification from the literature.

[1] Radio Whispers

18 of the 23 observed sources were characterised with confidence. 16 observed sources with higher S/N are classified as stars (including 3 eclipsing binary sources), based on the similarity of spectral features to a Pickles spectrum. 2 were characterised as nebulae (or stars inside of a nebula since point sources were observed) from their bright and broad nebular emission lines. 5 observations had relatively lower optical S/N ratios, and no clear characterisations in terms of source type could be deduced with confidence about those sources from the reduced spectra. Speculations were given in the descriptions below each of those 5 spectra in Chapter 3 based on source morphologies and alignment between optical and radio source locations. With higher-resolution spectroscopy and higher signal-to-noise, these sources can be more accurately classified (see Chapter 3 for full details on these results).

4.1.3 Source overlap between citizen science and spectroscopy results

6 sources were both spectroscopically observed and classified on Zooniverse.

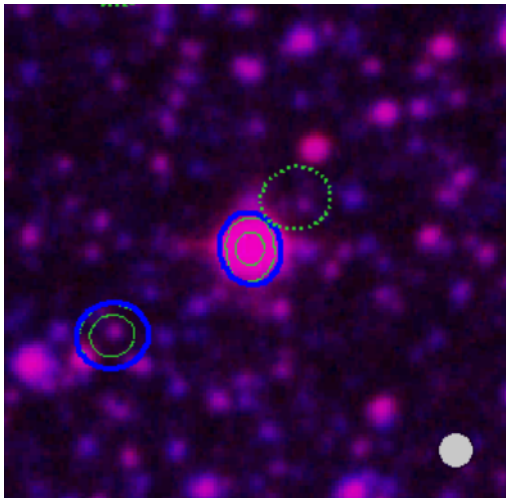


Figure 4.1: A central source crop (∼1.5 by 1.5 arcmin) of the image centred at the coordinates of GDB2, as shown to Zooniverse classifiers, consisting of a composite optical image from the SSS survey with radio contours overlayed in solid green lines, solid blue ellipses indicate the position of high radio S/N targets, dotted green ellipses indicate the location of lower S/N radio sources. The optical spectrum can be seen in Figure 3.6. All Zooniverse images are oriented North-up and East-left, and the radio beam size is shown in the bottom right corner as a grey ellipse.

GDB2 (see Figure 4.1): From spectroscopic follow-up, it is classified as a K-dwarf by spectrum comparison with good alignment between the radio and optical sources. From Zooniverse results, it is included in the high-confidence group classified as a radio point source with an optical counterpart. The results between the two methods agree with each other, indicating that this source is very likely to be a radio-emitting K-dwarf.

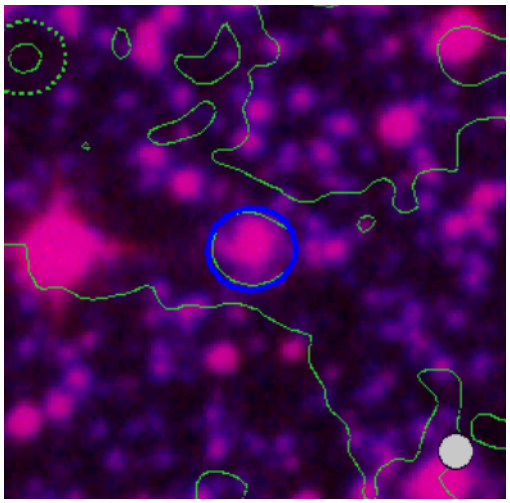


Figure 4.2: A central source crop (∼1.5 by 1.5 arcmin) of the image at the coordinates of GDB7, as shown to Zooniverse classifiers, the spectrum can be seen in Figure 3.11.

GDB7 (see Figure 4.2): From spectroscopic follow-up, it is identified on SIMBAD as an eclipsing binary with good alignment between the radio and optical source. The Zooniverse results show agreement that there is an optical counterpart, but the classifications were split 3 to 2 in favour of radio point vs compact extended.

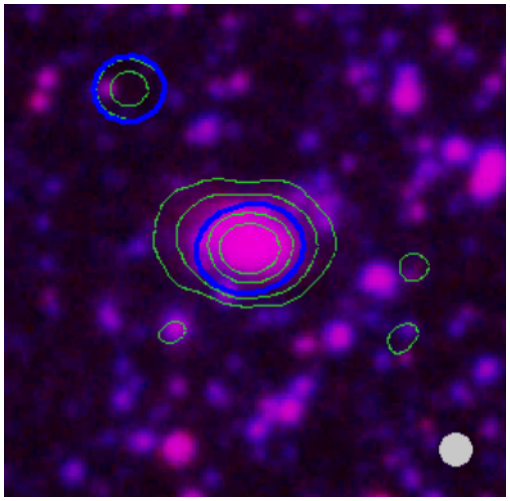


Figure 4.3: A central source crop (∼1.5 by 1.5 arcmin) of the image at the coordinates of GDB9, as shown to Zooniverse classifiers, the spectrum can be seen in Figure 3.13.

GDB9 (see Figure 4.3): From spectroscopic follow-up, it is identified as a nebula/star-forming region on SIMBAD and confirmed by the many broad nebular emission lines in the spectrum, with good alignment between the optical and radio sources. From the Zooniverse results, it is in the high-confidence group classified as a radio compact extended source with an optical counterpart. Seeing nebulae as compact extended radio sources is expected. The results between the two methods support the classification of this source as a nebula.

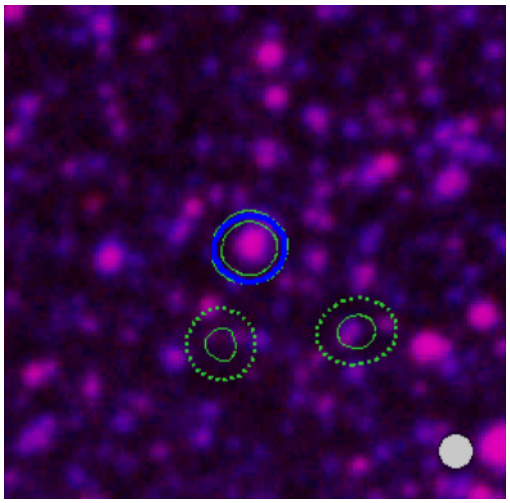


Figure 4.4: A central source crop (∼1.5 by 1.5 arcmin) of the image at the coordinates of GDB14, as shown to Zooniverse classifiers, the spectrum can be seen in Figure 3.16.

GDB14 (see Figure 4.4): From spectroscopic follow-up, it is identified as a planetary nebula, again confirmed by the many broad nebular emission lines present in the spectrum, along with good alignment between the radio and optical sources. This source was not in the high-confidence group in the Zooniverse results, as it was classified 4 to 1 in favour of being a compact extended radio source instead of a radio point source. There was an agreement between classifiers on it having an optical source, thought to be the counterpart to the radio emission. Both methods agree that this source is a planetary nebula as identified on SIMBAD.

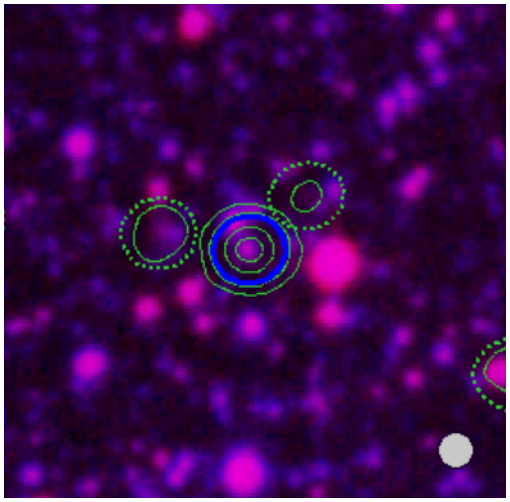


Figure 4.5: A central source crop (∼1.5 by 1.5 arcmin) of the image at the coordinates of GDB38, as shown to Zooniverse classifiers, the spectrum can be seen in Figure 3.26.

GDB38 (see Figure 4.5): This source spectrum is similar to that of GDB40, which is identified in SIMBAD as a candidate AGN, so it could also be an as-yet-unidentified AGN. This is further supported by the extended radio morphology and the presence of what is believed to be an optical counterpart centred inside the radio flux contours found in the Zooniverse results. The Zooniverse results placed this source in the high-confidence group, classified as a structured extended source with an optical counterpart. The agreement between the two methods further strengthens the possibility of this source being an AGN.

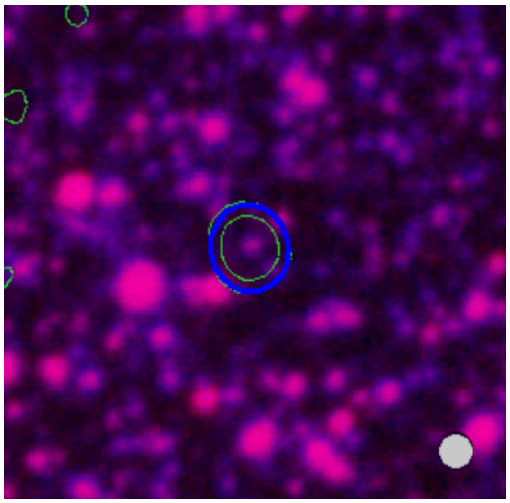


Figure 4.6: A central source crop (∼1.5 by 1.5 arcmin) of the image at the coordinates of GDB40, as shown to Zooniverse classifiers, the spectrum can be seen in Figure 3.27.

GDB40 (see Figure 4.6): This source is identified on SIMBAD as a candidate AGN, with a very noisy spectrum and good alignment between the radio and optical sources. This source was not in the high-confidence group as it has classifications of 4 to 1 in favour of it being a radio point source with an optical counterpart. Since some AGN can be detected as radio point sources (especially if they are faint or far away), the two methods used during this study agree (also evidenced by the compact extended X-ray morphology as seen in Figure 3.3).

4.2 Future work

Improvements in the design of the citizen science project could result in greater accuracy from classifications gathered from a larger audience of volunteer classifiers. Splitting up workflows to focus on specific characterisations with fewer questions asked could be one way of increasing the accuracy of classifications by concentrating on the recognition of one pattern at a time. Using better marketing through public engagement and social media during volunteer recruitment campaigns could also increase classification collection momentum. The marketing of this Zooniverse project fell short of the goal of recruiting many volunteer classifiers from the general public, and so we had to rely mostly on students and staff from the Astronomy department at the University of Cape Town and the SAAO as volunteer classifiers. This may have increased the accuracy of classifications since the pool of classifiers had more knowledge of astronomy compared to the general public.

The Citizen science project only used sources with a radio S/N of greater than 20, meaning only 1429 out of the 10873 radio sources identified by AEGEAN were included for characterisation using this technique. The larger portion of sources may be relatively faint in radio, but most of them are bound to be true detections worth following up on to study sources known to have faint (possibly extended) radio emissions. Future work could include an expansion of the citizen science project to include these lower signal-to-noise sources.

The data outputs from Zooniverse contain classifications concerning what volunteers thought the optical counterpart, if present, looked like. The options were between a point source, a point source surrounded by a diffuse gas-like region, only a diffuse gas-like region, multiple sources inside the radio flux contours making it hard to tell which one might be the actual optical counterpart, or something that is visible but does not fit into any of the previous 4 descriptions. This optical counterpart morphology output was left out of the analysis due to time constraints. The data can be analysed later to further characterise sources in this sample.

The optical survey used for images (SSS survey) shown to volunteer classifiers during the Zooniverse project

is from an all-sky survey built from the digitisation of photographic plates taken from the 1970s (Peacock et al. 2016). Though this survey offered easy access to image cut-outs and decent resolution, SSS lacks the sensitivity and sharper resolution of modern optical telescopes on Earth and in space. Using an optical survey with better sensitivity would lead to more sources in and behind the SMC being identified by citizen scientists while classifying data. Another reason to use a modern survey is to account for objects with high proper motion. Over decades, high proper motion objects can move a significant amount across the sky, causing the position of a source in an old survey image to be different from the position of the same source in a modern survey image. This can lead to optical counterparts being accidentally discarded or incorrectly identified. Using images from multiple recent surveys and catalogues would ensure better alignment between, in this case, radio and optical emissions that potentially originate from the same source.

There are also sources outside the searching radius used with AEGEAN (within 0.5 deg radius of the centres of the 2 MeerKAT images) that seem to show structured extended morphology, like radio lobes of FRI or FRII radio galaxies, found by visual inspection of the MeerKAT images. Due to AEGEAN searching for sources within a radius around the centre of the MeerKAT images, these were not included in the catalogue of radio sources for this project. Since these sources were missed, they could be added manually for characterisation in the future.

Cross-matching to X-ray surveys could narrow down source classification by eliminating sources that are known to emit (or do not emit) in hard or soft X-rays, providing further characterisations of radio sources. Radio-detected sources with morphology indicating they may be background AGN, pulsars in or in front of the SMC, or supernova remnants that also have obvious optical counterparts could be cross-matched to existing X-ray surveys to gain further insight into their behaviours or identification. Multi-wavelength analysis has a way of adding more pieces of the puzzle to any astronomical source being classified.

Redshifts and distances could not be accurately calculated for the sources we followed up spectroscopically during this MSc. High-resolution radio or optical spectroscopy of all sources used in Zooniverse, and those that were observed as follow-up targets, could provide accurate red-shift measurements and distances to further classify these sources as being Galactic, in the SMC or behind the SMC. By extension, this will also exclude certain classifications of what a source may or may not be.

The upcoming SKA radio interferometer will give Astronomers access to data with even more sensitivity and resolution than MeerKAT. This will open multiple new avenues for scientific exploration into the populations of radio sources.

The optical source density in the SMC is very high, making the risk of incorrectly identifying an optical source as the counterpart to radio emissions great and nearly always present. Error margin analysis on how likely it may be for each optical counterpart to be a chance alignment between the radio and optical emission would greatly aid any population study involving the search for multi-wavelength counterparts.

In retrospect, cosmic ray hits on the CCD were compounded by the long exposure times between 30 and 40 minutes per target, making it difficult to completely clean out the data without sometimes affecting the source apertures. In future, more images with lower exposure times should be taken to mitigate cosmic ray hits and increase the trustworthiness of observations.

4.3 Conclusions

The SMC has been studied across multiple wavelengths, revealing rich and diverse source populations distributed within and beyond the SMC. This is especially true for the recent radio surveys from the SKA precursors MeerKAT and ASKAP. Follow-up spectroscopy of individual sources and other means of source characterisation will keep expanding our knowledge of these sources and increase the sample sizes of future studies. The goal of this work was to characterise radio sources observed in the field of the SMC.

The results obtained from spectroscopy and citizen science methods led to the characterisation of ∼912 radio sources from an AEGEAN-generated source catalogue during the pursuit of the goals of this study.

AEGEAN detected 10873 potential radio point sources across 2 MeerKAT images of the SMC observed in 2019.

From spectroscopic follow-up, 13 stars, 3 eclipsing binaries, 2 nebulae and 1 AGN candidate (possibly 2) were characterised from the 23 sources observed. Of the 13 stars, 8 are believed to be the first radio detections from those stars with the optical counterpart also believed to have been observed (GDB 3, 4, 5, 6, 21, 23, 27, and 30). GDB14 was not reported in the paper linked to the SIMBAD identifier "SMP SMC 24" (Meyssonnier & Azzopardi 1993), and this could be the first radio detection of the planetary nebula.

From the citizen science results, 894 sources were morphologically characterised and linked to potential classifications, though no clear classification could be given to all individual sources from the Zooniverse results alone. There were 6 sources characterised using both of the methods (GDB2, a star aligned with a radio point source; GDB7, an eclipsing binary aligned with a radio source with morphology between a point source and compact extended source; GDB9, a nebula aligned with a compact extended radio source; GDB14, a planetary nebula aligned to a compact extended radio source; GDB38, potentially a candidate AGN when comparing the spectrum to that of GDB40, aligned with a radio source almost identical to that of GDB40; GDB40, a candidate AGN aligned with a compact radio source). The overlapping sources were given complementary characterisations when analysed using the two methods.

Employing multiple analysis methods for astronomical data is essential, as it facilitates the classification of emission sources and supports their inclusion in broader population studies. The use of multi-wavelength data is also highly beneficial for confirming classifications and enabling further study of astronomical objects.

Bibliography

Amirkhanyan V. R., Afanasiev V. L., Moiseev A. V., 2015, Astrophysical Bulletin, 70, 45

Andersson A., et al., 2023, Monthly Notices of the Royal Astronomical Society, 523, 2219

Andersson A., et al., 2024, Finding radio transients with anomaly detection and active learning based on volunteer classifications, doi:10.48550/arXiv.2410.01034, http://arxiv.org/abs/2410.01034

Aristeidou M., Herodotou C., 2020, Citizen Science: Theory and Practice, 5, 11

Aristeidou M., Scanlon E., Sharples M., 2015, in Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business. ACM, Graz Austria, pp 1–8, doi:10.1145/2809563.2809567, https://dl.acm.org/doi/10.1145/2809563.2809567

Asher A. D., Filipović M. D., Bojičić I., Alsaberi R. Z. E., Crawford E. J., Leverenz H., 2024, Astrophysics and Space Science, 369, 85

Babcock H. W., 1958, The Astrophysical Journal, 128, 228

Blanton M. R., et al., 2017, The Astronomical Journal

Carpano S., Haberl F., Sturm R., 2017,] 10.48550/ARXIV.1704.07646

Chiaberge M., Marconi A., 2011, Monthly Notices of the Royal Astronomical Society, 416, 917

Cotton W. D., et al., 2024, Monthly Notices of the Royal Astronomical Society, 529, 2443

Crause L. A., et al., 2016, in Evans C. J., Simard L., Takami H., eds, SpUpNIC Spectrograph Upgrade: Newly Improved Cassegrain on the South African Astronomical Observatory's 74-inch telescope. Edinburgh, United Kingdom, p. 990827, doi:10.1117/12.2230818, http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2230818

De Young D., 2002, The physics of extragalactic radio sources. University of Chicago Press, Chicago

Dent W. R. F., Matthews H. E., Ward-Thompson D., 1998, Monthly Notices of the Royal Astronomical Society, 301, 1049

Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L. W., 2009, IEEE Proceedings, 97, 1482

Dogra K., et al., 2025, The Astrophysical Journal Supplement Series, 276, 1

Dubner G., Giacani E., 2015, The Astronomy and Astrophysics Review, 23, 3

Dulk G. A., 1985, Annual Review of Astronomy and Astrophysics, 23, 169

Eisenstein D. J., et al., 2011, The Astronomical Journal

Fraknoi A., Morrison D., Wolff S. C., 2022, Astronomy 2e, 2e edn. OpenStax CNX, United States

Gaensler B. M., Hendrick S. P., Reynolds S. P., Borkowski K. J., 2003, The Astrophysical Journal, 594, L111

Gaia Collaboration 2020, VizieR Online Data Catalog, 1350, I/350

Gaia Collaboration et al., 2016, Astronomy & Astrophysics, 595, A1

Giroletti M., Edwards P., Giovannini G., Perez Torres M. A., 2007, ATNF Proposal, p. M242

González-Fernández C., Dorda R., Negueruela I., Marco A., 2015, Astronomy and Astrophysics, 578, A3

Graczyk D., et al., 2013, The Astrophysical Journal, 780, 59

Gvaramadze V. V., Kniazev A. Y., Gallagher J. S., Oskinova L. M., Chu Y.-H., Gruendl R. A., Katkov I. Y., 2021, Monthly Notices of the Royal Astronomical Society, 503, 3856

Güdel M., 2002, Annual Review of Astronomy and Astrophysics, 40, 217

Haberl F., Sturm R., Filipovic M. D., Pietsch W., Crawford E. J., 2011,] 10.48550/ARXIV.1112.0491

Haberl F., et al., 2012, Astronomy & Astrophysics, 543, A154

Hancock P. J., Murphy T., Gaensler B. M., Hopkins A., Curran J. R., 2012, Monthly Notices of the Royal Astronomical Society, 422, 1812

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

Hargrave P. J., Ryle M., 1974, Monthly Notices of the Royal Astronomical Society, 166, 305

Heywood I., et al., 2022, The Astrophysical Journal, 925, 165

Hénault-Brunet V., et al., 2011,] 10.48550/ARXIV.1110.6404

Jaschek C., Jaschek M., 1987, The Classification of Stars, 1 edn. Cambridge University Press, doi:10.1017/CBO9780511628979, https://www.cambridge.org/core/product/identifier/9780511628979/type/book

Jennett C., et al., 2016, Journal of Science Communication, 15, A05

Jonas J., the MeerKAT Team 2018, in Proceedings of MeerKAT Science: On the Pathway to the SKA — PoS(MeerKAT2016). Sissa Medialab, Stellenbosch, South Africa, p. 001, doi:10.22323/1.277.0001, https://pos.sissa.it/277/001

Joseph T. D., et al., 2019, Monthly Notices of the Royal Astronomical Society, 490, 1202

Jovanovic M. D., Damljanovic G., Taris F., Gupta A. C., Bhatta G., 2023, Monthly Notices of the Royal Astronomical Society, 522, 767

Kloetzer L., Da Costa J., Schneider D. K., 2016, Human Computation, 3, 25

Knowles K., et al., 2022, Astronomy & Astrophysics, 657, A56

Kordopatis G., et al., 2011, Astronomy & Astrophysics, 535, A107

Liao M., Wang J., Ren W., Zhou M., 2024, Monthly Notices of the Royal Astronomical Society, 528, 3696

Livanou E., et al., 2007, The Astronomical Journal, 133, 2179

Maggi P., et al., 2019, Astronomy & Astrophysics, 631, A127

Masters K. L., Team t. G. Z., 2019, Proceedings of the International Astronomical Union, 14, 205

Matthews L. D., 2013, Publications of the Astronomical Society of the Pacific, 125, 313

Matthews L. D., 2019, Publications of the Astronomical Society of the Pacific, 131, 016001

Mauch T., Murphy T., Buttery H. J., Curran J., Hunstead R. W., Piestrzynski B., Robertson J. G., Sadler E. M., 2003, Monthly Notices of the Royal Astronomical Society, 342, 1117

McGowan K. E., et al., 2008, Monthly Notices of the Royal Astronomical Society, 383, 330

Meyssonnier N., Azzopardi M., 1993, Astronomy and Astrophysics Supplement Series, 102, 451

Miraghaei H., Best P. N., 2017, Monthly Notices of the Royal Astronomical Society, p. stx007

Palate M., Rauw G., Koenigsberger G., Moreno E., 2013, Astronomy & Astrophysics, 552, A39

Pawlak M., et al., 2016, Acta Astronomica, 66, 421

Peacock J. A., Hambly N. C., Bilicki M., MacGillivray H. T., Miller L., Read M. A., Tritton S. B., 2016, Monthly Notices of the Royal Astronomical Society, 462, 2085

Peterson B. M., 1997, An introduction to active galactic nuclei. Cambridge University Press, Cambridge

Pickles A., 1998, Publications of the Astronomical Society of the Pacific, 110, 863

Price C. A., Lee H., 2013, Journal of Research in Science Teaching, 50, 773

Schade S., Pelacho M., Van Noordwijk T., Vohland K., Hecker S., Manzoni M., 2021, in Vohland K., Land-Zandstra A., Ceccaroni L., Lemmens R., Perelló J., Ponti M., Samson R., Wagenknecht K., eds, , The Science of Citizen Science. Springer International Publishing, Cham, pp 351–371, doi:10.1007/978-3-030-58278-4˙18, https://link.springer.com/10.1007/978-3-030-58278-4_18

Science Software Branch at STScI 2012, Astrophysics Source Code Library, p. ascl:1207.011

Simpson R. J., et al., 2012, Monthly Notices of the Royal Astronomical Society, 424, 2442

Smith R. E., 1976, Technical report, Post sunset behavior of the 6300 A atomic oxygen airglow emission, https://ui.adsabs.harvard.edu/abs/1976psba.rept.....S. https://ui.adsabs.harvard.edu/abs/1976psba.rept.....S

Soni S., Akhavan-Tafti M., Higgins C., Fung S. F., 2023, 2023, ED34A

Stewart R. T., 1989, Publications of the Astronomical Society of Australia, 8, 60

Sturm R., et al., 2013, Astronomy and Astrophysics, 558, A3

Sunil Krishna M., Bag T., 2013, 2013, SM41A

Tang H., Scaife A. M. M., Wong O. I., Shabala S. S., 2022, Monthly Notices of the Royal Astronomical Society, 510, 4504

Urry C. M., Padovani P., 1995, Publications of the Astronomical Society of the Pacific, 107, 803

Van Brussel S., Huyse H., 2019, Journal of Environmental Planning and Management, 62, 534

Walmsley M., et al., 2021, Monthly Notices of the Royal Astronomical Society, 509, 3966

Wenger M., et al., 2000, Astronomy and Astrophysics Supplement Series, 143, 9

Westerlund B. E., Smith L. F., 1964, Monthly Notices of the Royal Astronomical Society, 127, 449

Wong G., et al., 2011, Serbian Astronomical Journal, pp 103–106

Xin Y., Guo X., 2023, Detection of the extended γ-ray emission around supernova remnant DA 530 with Fermi-LAT, http://arxiv.org/abs/2309.05949

Zhou Z., Wu H., Zhou X., Ma J., 2018, Publications of the Astronomical Society of the Pacific, 130, 094101

van Dokkum P., 2001, Publications of the Astronomical Society of the Pacific, 113, 1420

Appendices

Appendix A

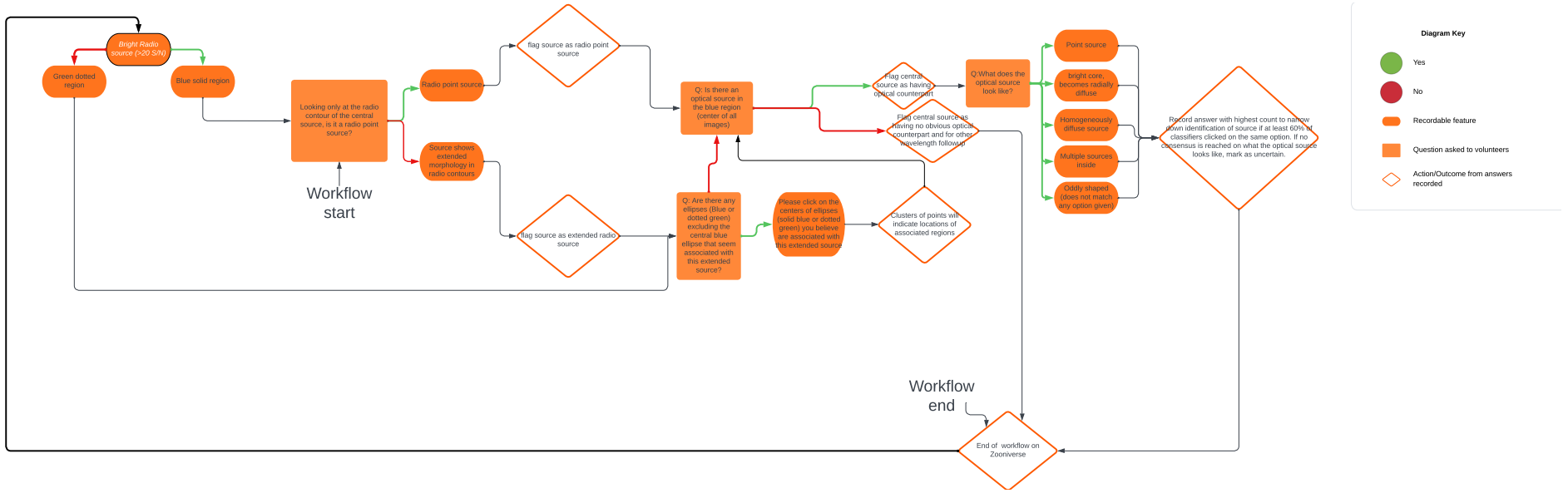


Figure A.1: Above is the full version of the decision tree in the form of a flow chart that was used to design the workflow of the citizen science project built for this work.

Appendix B

Spectroscopically observed targets

Table B.1: Below the sources that were observed in Sutherland are listed with the comments made during the observation run, sorted by the date on which they were observed

Name	**Obs_date**	**Observer_comments**
GDB1	16/11/23	SIMBAD Flo581; High proper motion star, likely a foreground star in the field of the SMC.
GDB40	16/11/23	Identified as AGN candidate on SIMBAD.
GDB41	16/11/23	Most faint source on my list. 40 minute exposure hoping it is enough after finding it on acquisition. No SIMBAD match, but GAIA has it as Gaia DR3 4690582128120828416.
GDB13	17/11/23	Source not identified, Airmass is starting to be a problem.
GDB14	17/11/23	Identified in SIMBAD as LHA 115-N 70 and it is a Planetary nebula, I am going to try and get the targets that are less faint, to have some spectra which have adequate signal to noise to get good extractions. Source was not exactly centre slit, to avoid other source being in the slit.
GDB37	17/11/23	No close matches, dim brown source. A brighter source is also in the slit off to the side. care when making an extraction.
GDB38	17/11/23	Bit easier to find, extraction recommended. Fainter of the 2 sources in slit.
GDB5	17/11/23	Source not identified, This one seems much brighter than Gaia is saying it should be. I double checked that I was on the right target.
GDB23	18/11/23	Unidentified. Looks like there might be emission lines identifiable.
GDB25	18/11/23	Unidentified. Some features visible which might be absorption lines.
GDB27	18/11/23	Unidentified on SIMBAD, blueish source in a busy field. very low signal to noise. most counts are under 700.
GDB30	18/11/23	Target unidentified on SIMBAD, inside a busy field, there is a second source inside the slit which is unavoidable.
GDB32	18/11/23	Target unidentified on SIMBAD. Spectrum has 3 sources in the disperion somehow... will be difficult to get a good extraction.
GDB34	18/11/23	Target acquired. high clouds are making the faint targets even harder to find. Target is unidentified on SIMBAD.
GDB21	19/11/23	Might be an AGN looking at the optical shape on Aladin.

GDB6	19/11/23	Unidentified. Right next to a 2mass source. impossible to not have both in the slit... my target is the brighter of the 2 so I hope it dominates the spectrum.
GDB10	21/11/23	This could be a far IR source or an eclipsing Binary, the sources are on top of each other and it is not clear which one marks the coordinate location.
GDB2	21/11/23	Identified as a star, might be a flare star as it has a radio counterpart. decreasing exposure time as the source is on the brighter side.
GDB3	21/11/23	Unidentified, Also a nice spectrum well worth extracting. 2 cosmic rays close to one end and one about a 3rd of the way from the other end seem to be in the object spectra.
GDB4	21/11/23	Unidentified, Very nice spectrum with some visible features.
GDB7	21/11/23	Identified as an eclipsing binary. expect spectrum of emission line star. the dimmer middle source in the slit is my source I believe. There is a brighter source to the side of the slit. Extract both apertures and take the one with least counts as my source.
GDB8	21/11/23	Identified as emission line star. I can see the emission features but the cosmic rays were not playing nice.
GDB9	21/11/23	Identified as a star forming region.

Appendix C

H I column density map

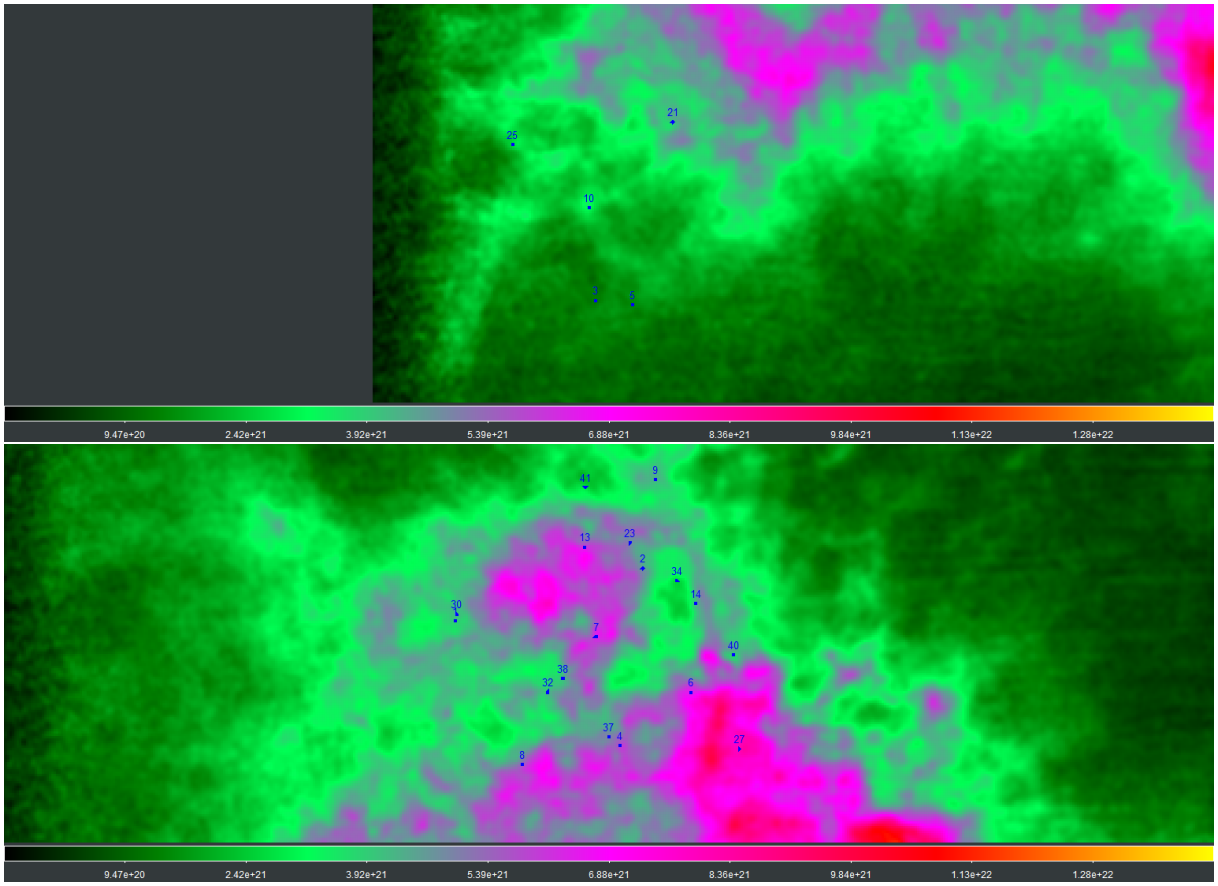


Figure C.1: H I neutral hydrogen column density maps showing the locations of targets observed for its work using the SAAO 1.9-m telescope for optical follow-up of select low optical magnitude, high radio S/N sources found in the SMC.

Appendix D

NOAA solar activity during observing week

In this appendix, a verbatim print of the NOAA data file indicating the levels of solar activity recorded during the week of observing where the optical spectroscopic data was gathered. More on the results in Chapter 3.


```
:Product: Daily Geomagnetic Data     quar_DGD.txt
:Issued: 2130 UT 07 Jan 2024
#
#  Prepared by the U.S. Dept. of Commerce, NOAA, Space Weather Prediction Center
#  Please send comment and suggestions to SWPC.Webmaster@noaa.gov
#
#             Current Quarter Daily Geomagnetic Data
#
#
#               Middle Latitude        High Latitude            Estimated
#              - Fredericksburg -     ---- College ----      --- Planetary ---
#  Date         A     K-indices         A     K-indices        A     K-indices
2023 10 01     9   4 2 2 2 2 2 1 2    10   3 2 2 3 4 1 1 1     9   4.00  2.00  2.00  2.00  2.00  1.00  1.33  1.67
2023 10 02    10   2 3 1 3 3 2 2 2     9   2 2 2 4 2 1 1 2     9   3.00  3.33  1.00  2.33  1.67  1.67  1.67  2.33
2023 10 03     6   1 2 2 1 3 2 1 1    19   2 0 2 4 5 5 3 0     8   1.33  2.67  1.67  1.33  2.33  2.33  2.33  1.67
2023 10 04     8   2 2 2 2 2 1 2 3    19   1 1 5 5 4 2 2 2    10   2.67  2.33  2.67  2.67  2.00  1.67  2.00  3.33
2023 10 05    11   3 4 2 3 2 1 2 2    22   3 4 3 6 4 1 1 2    16   4.33  5.00  2.33  3.00  2.00  1.00  1.33  2.33
2023 10 06     7   2 2 2 2 1 2 1 3    10   2 3 4 1 2 3 0 2     9   2.33  2.67  2.33  1.67  1.33  1.67  1.00  3.33
2023 10 07    -1   1 2-1-1-1-1-1-1     2   1 1 1 0 0 0 1 0     5   1.33  2.00  1.33  0.67  0.33  0.33  1.67  1.67
```

```
2023 10 08    -1 -1-1-1-1-1-1-1-1     7  1 2 3 3 1 1 1 1     7    2.00  2.00  2.33  1.67  0.67  1.00  1.67  2.67
2023 10 09    -1 -1-1-1-1-1-1-1-1    14  0 0 4 4 4 3 3 1     8    2.33  1.33  2.00  1.67  2.00  2.00  3.33  1.67
2023 10 10    -1 -1-1-1-1-1-1-1-1     4  1 2 1 3 0 0 0 0     4    1.00  2.00  1.67  1.67  0.33  0.67  0.67  0.33
2023 10 11     3 -1-1-1-1-1 0 1 2     2  0 0 1 2 1 0 0 0     4    0.67  0.67  1.33  1.33  1.00  1.00  0.67  1.00
2023 10 12     3 -1 1 1 1 2 1 0 1     0  0 0 0 0 0 0 0 0     4    1.33  1.00  1.33  1.00  0.67  1.00  0.67  1.00
2023 10 13    11 -1 1 1 3 2 3 3 3 2  15  0 3 3 3 4 4 3 1    13    1.00  2.67  3.33  2.67  3.00  3.67  2.67  2.33
2023 10 14     6  1 1 1 3 2 1 2 1     5  1 1 1 3 1 0 2 1     8    2.00  1.67  2.00  3.00  1.67  1.00  2.67  1.00
2023 10 15     2  0 0 1 1 1 1 1 1     1  0 0 0 1 1 0 0 0     4    0.67  0.33  1.00  0.67  0.67  0.67  1.00  2.00
2023 10 16     3  2 0 1 1 2 1 0 0     0  0 0 0 0 0 0 0 0     4    2.00  0.33  0.67  1.00  1.33  0.67  0.33  0.33
2023 10 17     2  0 0 0 0 2 1 1 1     0  0 0 0 0 0 0 0 0     3    0.67  0.33  0.67  1.00  1.00  0.33  0.00  0.67
2023 10 18     8  0 2 2 2 3 3 2 2    12  0 1 0 3 5 4 1 1     9    0.33  2.00  1.33  2.33  3.00  3.00  2.33  2.33
2023 10 19     8  2 3 2 3 2 1 1 1    24  2 4 4 6 3 4 2 1    10    2.67  3.67  2.33  2.67  1.67  1.67  1.67  1.00
2023 10 20     8  2 2 2 2 2 1 2 3    10  1 1 3 4 3 1 1 2     8    1.33  2.33  2.33  2.33  2.00  1.00  2.33  2.67
2023 10 21    13  3 3 4 4 2 2 1 1    39  3 5 6 6 4 5 2 2    22    4.00  4.67  5.00  3.33  2.67  3.33  1.67  1.00
2023 10 22     7  2 1 2 2 1 2 3 1    17  1 1 4 5 4 3 2 2     8    2.33  1.67  2.33  2.33  1.33  1.67  3.00  1.67
2023 10 23     2  1 0 0 0 1 1 0 1     1  1 0 0 0 1 1 0 0     3    1.00  0.33  0.33  0.33  1.00  0.67  0.67  0.67
2023 10 24     2  1 0 0 1 1 1 1 1     3  1 0 1 2 0 1 1 1     4    0.33  0.33  0.67  1.33  0.67  1.00  1.33  1.33
2023 10 25     3  1 0 0 1 2 1 1 1     2  0 0 0 1 1 1 1 0     4    0.67  1.00  0.33  0.67  1.00  0.67  1.33  1.67
2023 10 26    18  2 2 2 4 4 3 4 4    29  1 2 1 6 6 5 2 1    23    2.67  2.67  2.00  4.67  4.00  3.33  4.33  4.67
2023 10 27     9  3 4 2 2 2 1 1 1    33  3 5 4 5 3 3 3 6    11    3.33  4.00  2.67  2.33  1.67  1.67  1.33  1.00
2023 10 28    13  1 2 2 4 4 2 2 3    86  6 6 7 7 7 6 3 3    19    1.00  3.00  3.33  4.00  4.33  3.33  3.00  3.67
2023 10 29    21  4 3 4 4 4 3 3 3    49  4 5 6 6 6 5 3 3    28    4.67  3.67  4.67  3.67  4.00  4.00  3.33  4.00
2023 10 30    10  3 2 1 3 3 2 2 2    31  5 2 1 6 6 3 2 2    12    4.00  2.00  1.33  3.33  3.33  1.67  2.33  2.33
2023 10 31     6  2 2 1 2 2 2 1 1    18  2 2 3 5 5 3 1 1     9    2.67  2.33  2.00  2.67  2.67  1.33  1.67  2.00
2023 11 01     6  2 2 1 2 2 2 2 1    13  1 2 3 4 4 3 1 1     9    2.33  2.67  2.33  2.67  2.33  2.00  1.33  1.67
2023 11 02     4  2 2 1 1 1 1 1 1     4  1 1 2 2 2 1 0 1     7    2.67  2.67  2.00  1.67  1.33  1.33  1.33  1.33
2023 11 03     3  1 0 1 1 1 2 1 1     3  0 1 2 2 2 0 0 0     4    1.00  1.00  1.67  1.33  1.00  0.33  0.67  0.67
2023 11 04    12  1 1 2 1 2 4 4 3     8  0 1 2 2 2 2 3 3    14    1.00  1.00  2.00  1.00  1.67  3.00  4.67  4.33
2023 11 05    30  2 2 1 5 5 5 5 4    58  3 2 2 6 5 7 7 4    57    3.00  2.33  1.33  6.00  6.00  7.00  6.33  4.67
2023 11 06    27  3 4 3 4 2 5 5 4    51  3 5 6 7 3 5 5 3    40    4.33  5.67  4.67  4.33  2.67  4.67  5.33  4.67
2023 11 07    11  3 3 2 2 3 3 2 2    25  4 4 2 4 4 5 4 2    18    4.67  4.33  2.33  2.33  2.67  3.33  3.00  2.67
2023 11 08    15  2 4 3 3 2 3 3 3    21  3 3 2 6 3 3 3 2    16    2.67  4.00  3.00  3.33  1.33  2.33  4.00  3.33
2023 11 09     8  2 1 3 2 2 2 2 2    16  2 3 3 5 3 2 3 2    12    2.67  2.33  3.33  2.67  2.00  2.00  2.67  3.00
2023 11 10     7  1 3 2 2 2 2 1 1    13  3 3 4 3 4 2 0 0    10    2.33  3.67  2.67  2.00  2.33  1.67  1.00  2.00
2023 11 11     4  1 1 1 0 2 2 1 1     3  1 1 2 2 1 0 1 0     5    2.00  1.33  1.67  1.00  1.00  1.00  1.00  0.67
2023 11 12    10  0 1 4 3 2 2 1 3    16  0 1 5 5 3 2 1 2    10    0.67  1.33  4.00  2.67  2.33  1.67  1.33  3.00
2023 11 13    15  5 4 1 2 3 2 2 1    11  3 3 2 2 4 3 1 0    16    5.00  3.67  1.67  1.67  3.00  3.00  2.00  1.33
2023 11 14     5  2 1 0 1 2 3 2 0     5  1 1 1 1 1 3 2 0     6    2.33  1.00  0.33  1.00  1.33  3.00  2.33  0.67
2023 11 15    11  2 1 3 2 3 4 2 2    22  1 2 3 4 6 4 2 2    12    2.33  1.67  2.33  2.00  3.33  3.33  3.33  2.67
```

2023	11	16	7	1	1	2	2	2	3	2	2	12	1	1	4	3	4	2	1	2	8	1.67	1.00	2.67	2.00	2.33	2.33	1.33	2.33
2023	11	17	3	1	0	0	1	2	2	1	0	2	0	1	1	2	1	0	0	0	3	0.67	0.33	0.33	0.67	0.67	1.67	1.33	0.33
2023	11	18	2	0	1	0	0	2	1	1	1	1	0	1	1	1	0	0	0	0	3	0.33	1.00	0.67	1.00	0.67	0.33	1.00	1.33
2023	11	19	3	0	0	1	1	2	2	1	1	4	0	0	3	3	1	0	0	0	4	0.67	0.33	1.67	1.67	0.67	0.67	1.33	1.00
2023	11	20	4	1	1	1	1	1	2	2	1	4	0	0	3	2	1	1	0	0	5	1.00	1.00	2.00	2.00	1.67	1.33	1.33	1.00
2023	11	21	15	1	2	4	3	4	3	2	3	31	0	1	5	3	6	6	2	3	18	1.33	2.33	4.33	3.33	3.67	3.67	3.00	3.67
2023	11	22	17	3	3	4	4	3	3	3	2	60	3	6	7	6	5	6	4	2	30	4.00	5.00	5.00	4.33	3.33	4.00	4.00	2.67
2023	11	23	4	2	2	2	1	1	1	0	1	8	1	1	4	3	3	1	0	0	7	2.67	2.33	2.67	1.33	1.67	0.67	0.67	0.67
2023	11	24	5	1	2	1	1	2	1	1	2	6	0	1	0	1	3	3	2	2	7	1.67	1.67	1.33	1.33	2.00	1.67	2.33	2.67
2023	11	25	18	2	2	3	3	4	4	4	3	61	2	2	7	6	6	6	6	3	38	3.00	3.00	4.67	4.67	5.00	4.67	6.00	4.00
2023	11	26	-1	4	2	2	-1	-1	-1	-1	-1	12	3	1	4	4	3	1	1	1	10	4.00	2.33	3.33	2.00	2.00	0.67	1.33	1.67
2023	11	27	2	-1	-1	-1	-1	-1	0	1	1	9	0	0	4	2	4	2	1	0	7	1.33	1.67	2.33	1.33	2.00	2.33	1.33	2.00
2023	11	28	6	1	1	2	2	2	2	2	1	11	1	0	2	5	2	3	2	1	7	2.00	2.00	2.00	2.00	1.67	2.00	2.00	1.67
2023	11	29	5	0	2	2	1	1	2	2	1	8	0	1	3	3	4	1	1	0	5	1.00	2.00	2.00	1.33	1.67	1.00	2.00	0.67
2023	11	30	4	2	1	1	1	1	2	1	0	6	0	2	3	2	3	0	0	0	5	1.67	2.00	1.33	1.33	1.67	1.00	1.33	0.33
2023	12	01	30	4	3	3	6	5	3	3	4	57	3	3	4	7	6	6	6	3	56	4.33	4.00	4.00	7.00	6.33	4.67	5.00	5.00
2023	12	02	11	5	3	2	1	1	2	1	1	8	4	3	2	2	0	2	0	1	14	5.33	4.00	2.00	1.00	1.00	1.67	1.33	1.67
2023	12	03	10	2	3	3	1	2	3	2	2	8	2	2	3	2	1	2	2	2	11	2.67	3.00	3.00	1.67	2.33	3.33	2.00	2.33
2023	12	04	9	2	2	3	2	2	3	2	1	14	1	1	4	4	4	3	2	1	9	2.33	2.33	3.00	2.67	2.33	2.33	2.00	1.33
2023	12	05	9	3	3	2	2	3	1	1	2	32	1	4	2	6	6	4	3	3	15	3.33	3.33	2.67	3.33	3.00	3.00	2.33	3.00
2023	12	06	7	2	2	2	2	2	2	2	1	25	3	2	5	5	5	3	3	2	10	2.67	2.67	2.67	2.33	2.67	2.67	2.00	2.00
2023	12	07	4	1	0	1	1	1	2	2	2	4	1	1	2	2	2	0	1	0	5	1.33	0.67	1.33	1.00	0.67	0.67	2.00	2.33
2023	12	08	4	1	1	1	1	1	2	1	1	3	0	0	1	1	0	2	2	1	5	1.00	0.33	1.33	1.33	0.67	1.67	1.67	1.33
2023	12	09	2	0	0	0	1	1	2	0	1	0	0	0	1	0	0	0	0	0	3	0.33	0.33	1.00	0.67	0.33	0.67	0.67	0.33
2023	12	10	4	0	1	1	1	2	2	1	1	2	0	1	2	0	1	0	0	0	4	0.67	1.33	1.00	0.33	1.00	1.00	1.33	1.00
2023	12	11	3	0	2	1	0	2	1	1	0	0	0	0	0	0	0	0	0	0	3	0.67	1.33	1.00	0.33	0.67	0.67	0.33	0.33
2023	12	12	8	0	1	1	1	4	3	2	1	18	0	0	1	1	6	5	2	0	10	0.33	2.00	1.33	1.00	4.00	3.67	2.33	1.33
2023	12	13	7	0	2	2	1	2	2	2	3	17	0	1	3	5	4	4	3	1	8	0.67	2.00	1.67	1.67	2.00	2.67	2.33	3.00
2023	12	14	13	3	3	3	3	3	3	1	2	30	4	4	5	6	4	3	2	2	16	4.00	4.00	3.33	3.67	2.67	2.00	1.33	2.33
2023	12	15	8	1	1	1	2	3	3	2	2	25	1	1	2	5	6	4	3	3	12	2.00	1.33	1.33	2.67	3.33	3.00	3.33	3.33
2023	12	16	10	3	3	3	2	2	2	2	2	13	3	3	3	3	4	1	1	2	14	3.33	4.00	3.00	1.67	2.33	1.00	3.00	3.00
2023	12	17	32	3	4	4	5	5	5	3	4	88	3	5	6	7	7	8	4	3	36	3.67	4.33	4.33	4.67	5.33	5.33	3.33	4.33
2023	12	18	16	2	3	5	3	3	2	2	3	40	2	3	6	6	6	4	3	3	28	2.67	3.33	6.00	4.00	3.33	3.00	2.67	4.67
2023	12	19	10	3	3	2	1	2	3	2	2	11	3	4	3	1	1	2	2	2	12	3.67	3.67	2.67	1.67	1.67	2.33	2.00	2.67
2023	12	20	7	3	2	2	1	2	2	2	1	21	4	3	4	5	4	3	2	1	11	4.00	3.33	2.33	2.00	2.33	2.33	1.67	1.00
2023	12	21	3	2	1	0	2	1	1	1	0	9	1	1	1	5	2	2	1	0	5	2.67	0.67	0.67	2.33	0.67	1.00	0.67	0.33
2023	12	22	3	0	1	1	1	1	2	1	0	1	0	1	1	1	0	0	0	0	4	0.67	1.33	1.67	0.67	0.33	1.00	0.67	0.67
2023	12	23	5	0	1	3	1	2	2	1	1	12	0	0	3	1	4	5	1	1	7	0.67	1.00	2.67	1.33	2.00	2.67	1.00	1.33
2023	12	24	7	2	2	2	3	2	2	1	1	14	0	3	3	5	4	1	1	0	9	2.67	2.67	2.67	3.00	2.00	1.67	1.33	1.00

2023	12	25	2	0	1	1	0	1	2	1	0	0	0	0	0	0	0	0	1	0	4	0.33	1.00	1.00	0.67	0.33	1.33	1.67	0.67
2023	12	26	4	0	1	1	2	2	2	1	0	4	0	0	0	2	3	1	1	0	5	0.33	1.67	1.67	2.00	1.67	1.33	1.33	0.33
2023	12	27	4	2	1	1	1	1	2	0	1	2	0	0	2	1	0	1	0	0	4	1.67	1.33	1.33	1.67	0.67	1.00	0.67	0.67
2023	12	28	2	0	1	1	0	1	1	0	2	1	0	0	1	1	0	0	0	0	3	0.33	0.33	1.33	0.33	0.33	0.67	0.33	2.00
2023	12	29	6	1	1	1	3	1	2	2	2	8	0	0	2	5	1	0	0	1	6	1.00	1.67	1.67	2.33	1.67	1.00	1.67	2.33
2023	12	30	4	1	2	1	1	1	2	1	0	2	0	1	1	0	1	1	0	0	5	1.00	2.33	1.00	1.00	1.00	1.33	1.00	0.33
2023	12	31	2	0	1	0	0	1	2	1	1	0	0	0	0	0	0	0	0	0	3	0.33	1.00	0.67	0.67	0.67	1.00	1.33	1.33

Appendix E

Zooniverse data columns with short descriptions

Here are the column names and a short description of what data they contain. The columns are from the REDTEST.csv file on the GitHub associated to this work.

- The first column has no name, but it is an index column added by Python during the table creation process.
- **subject_id**: The identification number of the subject image from the Zooniverse database.
- **data.yes**: # of responses indicating the radio source seems to have an associated optical source).
- **data.no**: # of responses indicating the radio source does not seem to have an associated optical source.

- **data.the-radio-source-is-showing-extended-morphology**: # of responses indicating the radio source contours show extended morphology beyond the beam size
- **data.yes-there-seem-to-be-associated-regions-close-to-the-central-source**: # of responses that indicate a structured extended radio source morphology.
- **data.yes-it-looks-like-a-radio-point-source**: # of responses indicating a radio point source morphology from the contours.
- **data.no-nearby-marked-regions-seem-associated-with-the-central-source**: # of responses indicating a compact extended morphology.
- **multi-source**: Python generated column recording the # of responses that indicate the associated optical source could be one of many close-by sources.
- **homogeneous**: Python generated column recording the # of responses that indicate the associated optical source looks like a homogeneously luminous gas-like region.
- **point-star**: Python generated column recording the # of responses that indicate the associated optical source looks like a star or dot.
- **oddORnot-sure**: Python generated column recording the # of responses that indicate the associated optical source has an odd shape that does not match the other options.
- **core-in-diffuse**: Python generated column recording the # of responses that indicate the associated optical source looks like a bright core surrounded by a diffuse gassy region.
- **Sum**: This was an internal confirmation value to check if the number of associated optical sources matches the number of responses regarding the optical source morphology.
- **Total_Class**: Total # of classifications for the subject image.
- **multiFrac**: Fraction of classifications that indicated more than one optical source could be the associated optical counterpart to the radio emission.
- **homoFrac**: Fraction of classifications that indicated a homogeneously diffuse optical counterpart.

- **pointFrac**: Fraction of classifications that indicated a star or point-like optical counterpart.
- **unsureOddFrac**: Fraction of classifications that indicated and oddly shaped or uncertain optical counterpart morphology.
- **coreFrac**: Fraction of classifications that indicated an optical morphology with a bright core and diffuse gas-like emission around it.
- **ext_radFrac**: Fraction of classifications that indicated an extended radio source morphology.
- **point_radFrac**: Fraction of classifications that indicate a radio point source morphology.
- **flag_rad_point**: 1 if > 80 per cent agreement between classifiers that the radio source is a point source.

- **flag_rad_extended**: 1 if > 80 per cent agreement between classifiers that the radio source is an extended source.
- **flag_rad_unsure**: 1 if there is < 80 per cent agreement on either the radio point or extended flags.
- **flag_optical_counterpart**: 1 if > 80 per cent agreement between classifiers that an associated optical source is present.
- **metadata**: contains a ".json" string recording the file name of the subject as uploaded by the project creator.
- **locations**: contains a ".json" string with the image URL inside.
- **classifications_count**: Zooniverse recorded count of classification for the subject.
- **Filename**: Extracted file name as uploaded by the project creator from the "metadata" column.
- **LCurl**: URL extracted from the "locations" column.