Radio and infrared surveys for active galactic nuclei behind the Magellanic Clouds

Clara Marie Pennock

Doctor of Philosophy

Department of Physics, Keele University

June 2023

Abstract

I present an analysis of a new 120 deg^2 radio continuum image of the Large Magellanic Cloud (LMC) at 888 MHz with a bandwidth of 288 MHz and beam size of $13\farcs9 \times 12\farcs1$, from the Australian Square Kilometre Array Pathfinder (ASKAP). I constructed a catalogue of 54,612 sources reaching down to < 0.2 mJy and explore the sources by cross-matching with surveys at other wavelengths. I find sources are predominantly extragalactic, display synchrotron emission associated with AGN, and star-forming galaxies become more prominent below 3 mJy compared to AGN.

I employ machine learning to separate the stellar from the extragalactic in the Magellanic Clouds. The t-SNE algorithm is used with multi-wavelength data from Gaia EDR3, VISTA survey of the Magellanic Clouds (VMC), AllWISE and ASKAP to cluster similar radio sources together. This separates AGN, galaxies, blazars and stellar sources. The probabilistic random forest classifier is trained on known sources with data from optical to mid-IR. This yielded accuracies of 0.93 ± 0.01 (SMC) and 0.91 ± 0.01 (LMC) when tested on known sources. I classify the 31,169,627 sources in the VMC SMC field to find that classes distribute across colour-colour plots and the SMC field as expected, except for in the highest density regions where there is an over-density of AGN due to blending and photometry mismatches.

Following the discovery of SAGE0536AGN ($z \sim 0.14$), with the strongest 10-μm silicate emission ever observed for an AGN, I discovered SAGE0534AGN ($z \sim 1.01$), a similar AGN but with less extreme silicate emission. Both originally mistaken as evolved stars in the Magellanic Clouds. Lack of star-formation implies we are seeing the central engine of the AGN without contribution from the host galaxy. They could be a key link in galaxy evolution. I searched for more of these sources using the SMC t-SNE clusters to find they are grouped with AGN ($0.13 < z < 1.23$) separated from the rest, suggesting a rare class. Their host galaxies appear to be either in or transitioning into the green valley, where AGN properties, such as the torus width, X-ray luminosity, radio loudness/spectral index and Eddington ratio, appear to be tracing the transition.

Acknowledgements

My sincere thanks go to Jacco van Loon for all of his supervision, assistance and support as I undertook this research.

I give thanks to Keele University and to all of the institutions which provide research funding here.

I give thanks to my parents and brother, whose love, support and faith in me have been invaluable in keeping motivated, especially during the Covid pandemic.

And lastly, to my friends. Especially, Zelekha and Ian, who have provided many laughs over the last few years.

Clara Marie Pennock
Keele Astrophysics Group
December 2022

Contents

Abstract . . . . . **i**
Acknowledgements . . . . . **ii**
1 Introduction **1**
1.1 AGN 1
1.1.1 Multi-wavelength view 4
1.1.1.1 X-ray 4
1.1.1.2 UV and optical 6
1.1.1.3 IR 7
1.1.1.4 Radio 9
1.1.2 Variability 12
1.1.3 Relation between supermassive black holes and their host galaxies 13
1.1.4 AGN feedback and galaxy evolution 15
1.1.5 An extreme AGN sample inspired by SAGE0536AGN 17
1.2 Magellanic Clouds 18
1.3 Machine learning 20
1.3.1 Supervised Learning 20
1.3.2 Unsupervised Machine Learning 21
1.4 Scientific questions and thesis outline 22
2 Data **24**
2.1 Photometric Surveys 24
2.1.1 VISTA near infrared imaging 26
2.1.2 ASKAP 28
2.1.2.1 ASKAP ESP SMC image 29
2.1.2.2 ASKAP ESP LMC image 30
2.1.3 Magellanic specific surveys 32
2.1.3.1 Optical SMASH 37
2.1.3.2 Infrared SAGE and HERITAGE 37
2.1.3.3 X-ray XMM-Newton 38
2.1.4 All-sky surveys 38
2.1.4.1 Optical Gaia all-sky survey 38
2.1.4.2 Infrared WISE 39
2.2 Spectra 40
2.2.1 SAAO 1.9m 42
2.2.2 SALT 43
2.2.3 Literature 45
2.3 Training sets for machine learning 45
2.3.1 SAGE-spec 47

2.3.2 Extragalactic . . . 47
2.3.3 Galactic and Magellanic . . . 48
3 Methodology . . . 49
3.1 ASKAP ESP LMC Source Detection . . . 49
3.1.1 Testing of Aegean . . . 51
3.1.2 Source lists . . . 54
3.1.3 Comparison with previous radio catalogues . . . 56
3.1.3.1 SUMSS 843 MHz . . . 60
3.1.3.2 ASKAP SMC ESP . . . 63
3.2 Galfit . . . 63
3.2.1 Models . . . 66
3.2.2 Method . . . 67
3.3 CIGALE . . . 70
3.3.1 Models . . . 70
3.3.1.1 AGN models . . . 71
3.3.1.2 Galaxy component models . . . 73
3.3.2 Inputs . . . 74
3.4 Machine learning . . . 75
3.4.1 t-SNE . . . 75
3.4.2 Machine learning classifier . . . 81
3.4.2.1 Decision Trees . . . 81
3.4.2.2 Random Forests . . . 83
3.4.2.3 Probabilistic Random Forest . . . 84
3.4.3 Creating the multi-wavelength dataset . . . 85
3.4.4 Training . . . 88
3.4.4.1 Inputs . . . 88
3.4.4.2 PRF parameters . . . 88
3.4.4.3 Determining the number of trees . . . 89
3.4.4.4 Dataset configuration test . . . 89
3.4.5 Emission-line stars – help or hindrance . . . 92
3.4.5.1 Testing classifier on the other Cloud . . . 95
3.4.5.2 Balanced vs imbalanced datasets . . . 96
3.4.5.3 Final data configuration . . . 99
3.4.5.4 Feature importance . . . 103
4 The extragalactic radio sky behind the LMC – a new deep and detailed map from ASKAP . . . 106
4.1 Magellanic and Galactic sources . . . 106
4.2 Background sources . . . 108
4.2.1 Radio galaxies . . . 108
4.2.2 Comparison with the GLEAM 4-Jy sample . . . 111

4.2.3 IR properties of radio sources . . . 115
4.2.3.1 Spectroscopic AGN . . . 116
4.3 Spectral indices . . . 124
4.4 Conclusions . . . 125
5 Search for more AGN like SAGE0536AGN . . . 129
5.1 The sample . . . 129
5.1.1 VMC light curves . . . 133
5.1.2 Optical line identifications and spectral analysis . . . 135
5.1.2.1 Black hole masses . . . 139
5.1.3 CIGALE modelling . . . 141
5.1.3.1 CIGALE models for SAGE0536AGN and SAGE0534AGN 141
5.1.3.2 CIGALE models of the t-SNE sample . . . 141
5.1.4 X-ray observations . . . 144
5.1.5 The host galaxies . . . 145
5.2 Discussion . . . 150
5.2.1 Source 8's true nature . . . 150
5.2.2 IR properties and selection criteria . . . 150
5.2.3 Green valley . . . 153
5.2.4 Radio analysis . . . 156
5.2.4.1 Radio morphology . . . 156
5.2.4.2 Spectral Indices . . . 157
5.2.4.3 Radio properties across the green valley . . . 159
5.2.5 AGN dust properties . . . 162
5.2.5.1 Variability . . . 162
5.2.5.2 CIGALE model components . . . 164
5.2.5.3 Silicate 9.7μm dust . . . 166
5.2.5.4 Comparison to other silicate emitting AGN . . . 166
5.3 Conclusions . . . 170
6 Machine learning identification and classification of sources in the VISTA survey of the Magellanic Clouds . . . 172
6.1 t-SNE analysis of the Magellanic Clouds . . . 172
6.1.1 Comparing the SMC and LMC t-SNE maps . . . 172
6.1.1.1 Spectroscopically confirmed classifications . . . 174
6.1.1.2 Photometry survey measurements . . . 176
6.1.1.3 Gaia classifications . . . 181
6.2 PRF classifications of the SMC field . . . 183
6.2.1 Stellar sources . . . 186
6.2.2 t-SNE radio sources . . . 188
6.2.3 PRF on whole SMC . . . 190
6.2.3.1 Extragalactic in the radio . . . 190

6.2.3.2 PRF classification distributions across the SMC field . 193
6.2.3.3 IR selections of PRF classified sources 197
6.3 Conclusions 204
7 Conclusions and outlook 209
7.1 Summary 209
7.2 Future work/improvements 214
Publications 216
A Sources spectroscopically observed 218
B IRAF 226
C PRF Feature Importances 228
D ASKAP Magellanic and Galactic sources 233
D.1 Planetary Nebulæ and other emission-line objects 233
D.2 Young stellar objects 237
D.3 Novæ and supernovæ 239
D.4 Supergiants 240
D.5 X-ray binaries 241
D.6 Galactic active stars 243
E Supplementary data for search for more AGN like SAGE0536AGN 247
E.1 Photometry 247
E.2 Additional spectra 247
E.3 CIGALE best fits 247

List of Figures

1.1 The unified model of AGNs. 3
1.2 Radio contours of a FR I galaxy and a FR II galaxy. 11
1.3 Relation between black hole mass M_{BH} and stellar velocity dispersion σ. 14
2.1 Optical layout of the VISTA telescope. 27
2.2 ASKAP ESP image of the LMC at 888 MHz. 31
2.3 Region around 30 Doradus from the ASKAP ESP image of the LMC. . 33
2.4 LHA 120-N 119 from the ASKAP ESP image of the LMC. 34
2.5 Area of the sky covered by the VMC, EMU ASKAP, SMASH and SAGE surveys of the SMC. 35
2.6 Area of the sky covered by the VMC, EMU ASKAP, SMASH and SAGE surveys of the LMC. 36
2.7 Basic optical diagram of a spectrometer. 41
2.8 Optical diagram of a grating that can be used in spectroscopy. 41
3.1 RMS map of the 888 MHz ASKAP observations. 50
3.2 Regions AEGEAN was tested on for the ASKAP ESP image of the LMC at 888 MHz. 52
3.3 How % Real Detections varies with detection threshold. 53
3.4 Density plots of the GOLD, SILVER and BRONZE source lists. 57
3.5 Histogram of integrated flux density at 888 MHz illustrating the GOLD, SILVER, and BRONZE populations. 58
3.6 Comparison between the integrated flux densities calculated for the ASKAP 888 MHz image, and those from calibrated archival data of SUMSS MOST 843 MHz data. 61
3.7 Figure of positional difference (MOST – ASKAP) of the 2949 radio sources from the ASKAP (888 MHz) catalogue and the SUMSS (843 MHz) catalogue. 62
3.8 Integrated flux density histograms of ASKAP LMC (888 MHz), ASKAP SMC (960 MHz & 1320 MHz) and SUMSS LMC (843 MHz). 64
3.9 Comparison between the integrated flux densities calculated for the ASKAP images, of the LMC and SMC at 888 MHz and 960 MHz, respectively. 65
3.10 Example of Galfit modelling taken from Peng et al. (2010). 68
3.11 Annotated schematic of the SKIRTOR AGN model from Stalevski et al. (2012). 72
3.12 t-SNE maps created using different perplexities and from a combination of VMC, Gaia EDR3, AllWISE and ASKAP data. 79
3.13 t-SNE map created from a combination of VMC, Gaia EDR3, AllWISE and ASKAP data, with known AGN overplotted. 80

3.14 Decision Tree example: A non-parametric model. 82
3.15 Random Forest example: A model consisting of multiple decision trees. 84
3.16 Histogram showing the differences between W1 (blue) and W2 (orange) magnitudes for the unWISE and AllWISE surveys. 87
3.17 Plot of score (average accuracy) against number of trees used in training classifier for the SMC, LMC and MC datasets with colour features included and not included. 90
3.18 Confusion matrix of classifier trained 75% of the SMC dataset and tested on the remaining 25%. 91
3.19 Confusion matrix of classifier trained 75% of the LMC dataset and tested on the remaining 25%. 92
3.20 Confusion matrix of classifier trained 75% of the MC dataset and tested on the remaining 25%. 93
3.21 Average prediction probability per class of the LMC-trained classifier used on emission-line stars. 94
3.22 Confusion matrix of classifier trained on entire SMC dataset and tested on the entire LMC dataset. 96
3.23 Confusion matrix of classifier trained on entire LMC dataset and tested on the entire SMC dataset. 97
3.24 Confusion matrix of the classifier trained and tested on the upsampled SMC dataset. 99
3.25 Confusion matrix of the classifier trained and tested on the upsampled LMC dataset. 100
3.26 Confusion matrix of final classifier trained and tested on the SMC dataset. 101
3.27 Confusion matrix of final classifier trained and tested on the LMC dataset. 102
3.28 Top 15 important features for the SMC and LMC classifier. 105
4.1 Four examples of radio emission from galaxies. 109
4.2 I - Comparison between ASKAP, SUMSS and GLEAM. 112
4.3 II - Comparison between ASKAP, SUMSS and GLEAM. 113
4.4 AllWISE colour–colour diagrams of the ASKAP and non-ASKAP detected sources in the LMC field. 117
4.5 Normalised histograms comparing integrated flux densities and redshifts of spectroscopically observed radio AGN vs. all radio AGN. 119
4.6 AllWISE colour–colour diagram of ASKAP and non-ASKAP detected sources that have been spectroscopically observed. 120
4.7 AllWISE colour–colour diagrams of ASKAP and non-ASKAP detected AGN, where the colour indicates redshift. 121
4.8 LMC AllWISE colour–colour diagram comparing the populations of the Gold, Silver, Bright and Faint source lists. 123
4.9 Distribution of spectral indices calculated from ATCA (1.4 GHz) and ASKAP (888 MHz) integrated flux densities. 126

5.1 t-SNE map created from a combination of VMC, Gaia EDR3, AllWISE and ASKAP data with a zoom in on area containing the dusty AGN. . 131
5.2 Spectrum of carbon star Source 5 where main spectral features are annotated. 132
5.3 Light curves from the VMC survey of all sources in our t-SNE selected sample. 134
5.4 Spectra of t-SNE selected sample, excluding Source 5 (carbon star, spectrum shown in online appendix) and 8 (no spectrum available. 138
5.5 CIGALE best fits of SAGE0536AGN and SAGE0534AGN. 142
5.6 GALFIT models of SAGE0536AGN, Source 10 and Source 16. 149
5.7 The spectrum of Source 8 observed with SALT. 150
5.8 AllWISE colour–colour diagram of the sample. 151
5.9 VISTA colour–colour diagram of the sample. 153
5.10 The position of the t-SNE selected sample on a SFR versus stellar mass diagram. How R and X-ray luminosity changes with SFR. How Eddington ratio and the torus and polar dust emission fractions of the AGN change with M_*. 155
5.11 ASKAP 1320 MHz radio flux of Sources 6, 15 and 16 shown as contours on top of the VMC K_s band images. 157
5.12 Spectral indices of the sample compared to other spectroscopically observed AGN behind the Magellanic Clouds. 159
5.13 Variation of spectral indices, α (top), and radio loudness (bottom) of the sample as they transition from star-forming to green-valley galaxies. It should be noted that many of the plotted points are lower limits in the x-direction. 160
5.14 The comparison of variability with Eddington ratio and optical depth, τ, of the t-SNE selected sample. 163
5.15 Accretion disk fraction versus torus disk fraction for the t-SNE selected sample. 165
5.16 Separate components of the CIGALE models for the t-SNE sample at rest-wavelength. 167
5.17 Silicate strength versus WISE band 4 – IRAS 60-μm (in magnitudes). . 169
6.1 t-SNE map of the LMC. 173
6.2 t-SNE map of the SMC. The highlighted regions show the general area of certain classifications of sources. 174
6.3 t-SNE map of the LMC. The high-lighted regions show the general area of certain classifications of sources 175
6.4 Comparison of the distribution of proper motions of the sources in the t-SNE maps of the SMC (left) and LMC (right). 177

6.5 Comparison of the distribution of Gaia DR3 colour G minus the average of G_{RP} and G_{BP} of the sources in the t-SNE maps of the SMC (left) and LMC (right). 178
6.6 Comparison of the distribution of VMC PSF colour $J - K_s$ of the sources in the t-SNE maps of the SMC (left) and LMC (right). 179
6.7 Comparison of the distribution of VMC PSF J-band sharpness value of the sources in the t-SNE maps of the SMC (left) and LMC (right). . . 180
6.8 Comparisons of the distribution of AllWISE colour W1 − W2 of the sources in the t-SNE maps of the SMC (left) and LMC (right). 180
6.9 Comparison of the distribution of Gaia DR3 probabilities of a sources being a QSO in the t-SNE maps of the SMC (left) and LMC (right). . 182
6.10 Comparison of the distribution of Gaia DR3 probabilities of a sources being a galaxy in the t-SNE maps of the SMC (left) and LMC (right). 183
6.11 Comparison of the distribution of Gaia DR3 probabilities of a source being a stellar source in the t-SNE maps of the SMC (left) and LMC (right). 184
6.12 PRF classifications of the sources with class probability >50% in the SMC t-SNE map. 189
6.13 Radio ASKAP 960 MHz flux density distribution and the Ratio of AGN to galaxy counts of the predicted AGN and galaxy sources in the SMC. 191
6.14 Radio ASKAP 1320 MHz flux density distribution and the Ratio of AGN to galaxy counts of the predicted AGN and galaxy sources in the SMC. 192
6.15 Distributions of the SMC sources classified as AGN (Upper panel) and Galaxies (Lower panel). 195
6.16 Distributions of the SMC sources classified as RGB (Upper panel) and OB stars (Lower panel). 196
6.17 Distributions of the SMC sources classified as AGB (Upper panel) and post-AGB/RGB stars (Lower panel). 198
6.18 Distributions of the SMC sources classified as PNe (Upper panel) and high proper-motion stars (Lower panel). 199
6.19 Distributions of the SMC sources classified as RSGs (Upper panel) and HII/YSOs (Lower panel). 200
6.20 AllWISE colour–colour diagram comparisons of the sources classified as AGN and galaxies sources classified as stellar in the SMC field. 202
6.21 AllWISE colour–colour diagram of sources classified as AGN that have and have not been detected in the radio. 203
6.22 VMC colour–colour diagram of sources classified as AGN that have and have not been detected in the radio. 205
6.23 VMC colour–magnitude diagram of sources classified as AGN that have and have not been detected in the radio. 206

C.1 Feature importances of the SMC (left) and LMC (right) probabilistic random forest classifier. Features ranked from 1st to 60th. 229
C.2 Feature importances of the SMC (left) and LMC (right) probabilistic random forest classifier. Features ranked from 61st to 120th. 230
C.3 Feature importances of the SMC (left) and LMC (right) probabilistic random forest classifier. Features ranked from 121st to 180th. 231
C.4 Feature importances of the SMC (left) and LMC (right) probabilistic random forest classifier. Features ranked from 181st to 237th. 232
D.1 Cumulative 888 MHz flux distributions of planetary nebulæ, post-AGB objects and emission-line stars. 236
D.2 X-ray binary LMC X-1 seen in the focus of a radio arc in the 888 MHz ASKAP LMC image. 243
D.3 LMC AB Dor, a hierarchical young active system detected in the 888 MHz ASKAP LMC image where emission coming from the main component and the B component contribute. 246
E.1 Spectrum of SAGE0534AGN observed with SALT under program 2021-2-SCI-017 (PI: Joy Anih). 249
E.2 Spectrum of SAGE0534AGN observed with SAAO 1.9m telescope with SpUpNIC. 250
E.3 CIGALE best fits of Sources without X-ray observations. 251
E.4 (cont.) CIGALE best fits of Sources without X-ray observations. 252
E.5 CIGALE best fits for sources with X-ray observations. 253

List of Tables

1.1 The main classes of AGN and their properties. 5
2.1 Surveys used in this work. 25
2.2 Sources used for machine learning training. 46
3.1 Example of the catalogue created from the LMC ASKAP ESP image. . 55
3.2 Table comparing existing radio surveys of the LMC with the survey presented here. 59
3.3 Modules and parameter values used to model the sample in CIGALE. . . 76
3.4 Parameters taken from various surveys to act as features in the PRF algorithm. 86
4.1 Classifications of known AGN that have been detected with ASKAP and those that have not. 116
5.1 Sample of similar sources identified through a t-SNE analysis. 130
5.2 t-SNE selected sample variability amplitudes and mean magnitudes in the K_s band. 133
5.3 Table of sources investigated/observed in this work. 136
5.4 Table of sources investigated in this work. 137
5.5 Black hole masses calculated. 140
5.6 AGN properties calculated with CIGALE. 143
5.7 The average, error-weighted 0.2 – 12 keV X-ray flux and the corresponding luminosity and their identifications from Maitra et al. (2018). . . . 145
5.8 The functions fitted to SAGE0536AGN, Source 10 and Source 16 in GALFIT and their parameters. 147
5.9 Radio luminosities and spectral indices for the t-SNE selected sample from ASKAP (888 MHz for LMC, 960 MHz for SMC). 158
6.1 Gaia classification probabilities of QSO ($P_{\rm QSO}$), galaxy ($P_{\rm Gal}$) and stellar sources ($P_{\rm SS}$) for the t-SNE selected AGN sample. 182
6.2 Excerpt from the table predicted class probabilities by the PRF for the SMC dataset. 185
6.3 Classifications of sources from Ruffle et al. (2015) that were not used in the training of the PRF classifier. 187
6.4 Distribution of the classifications of sources in the SMC field. 193
A.1 Sources spectroscopically observed using the SAAO 1.9m telescope during observing runs in 2019 and 2021. 219
A.2 SALT observations made over the course of this work. 224
A.3 Coordinates of the SALT observations made over the course of this work. 225
D.1 Example of ASKAP 888 MHz radio detections within 10 arcsec of sources at one time classed as potential PNe. 234

D.2 ASKAP 888 MHz radio detections within 10 arcsec of sources from the Herschel spectroscopy of the YSO sample (Oliveira et al., 2019). 238
E.1 Table format of the table of photometry of all sources in t-SNE selected sample. 248

1 Introduction

Galaxies are massive, gravitationally bound systems of stars, stellar remnants, dust, gas and dark matter that come in many different forms ranging from red elliptical quiescent galaxies to blue active spiral-shaped galaxies. All these galaxies have at their centre a supermassive black hole which, when undergoing accretion, becomes an active galactic nucleus. These Active Galactic Nuclei (AGN) can produce emission across the entire electromagnetic spectrum, and are some of the most luminous objects in the universe and have the potential to affect the entirety of the galaxy it resides in.

1.1 AGN

AGN preside in the centre of some galaxies, resulting from the accretion of gas by a supermassive black hole ($> 10^6$ solar masses). The mass of supermassive black holes is known to correlate with the mass of the galaxy bulge, implying the formation and evolution of bulges and supermassive black holes are intertwined (Magorrian et al., 1998; Ferrarese & Merritt, 2000; Gebhardt et al., 2000; Tremaine et al., 2002). It is thought that AGN play a significant role in galaxy evolution by creating large outflows that quench (Kormendy & Ho, 2013) and/or trigger bursts of star formation (Xue et al., 2010; Mullaney et al., 2011; Rosario et al., 2015; Cowley et al., 2016), making them an ideal laboratory for studying the evolution and formation of galaxies.

AGN can also be used to great effect as anchor points for an absolute astrometric reference system for determination of the proper-motions of close-by sources (Cioni et al., 2014). Due to their nature of being visible at great distances, this would mean that from our point of view they would be fixed points on the sky to compare the motions of the objects closer to us.

The first hints of the existence of AGN were the observations of quasars in the early 1960s. Quasars (quasi-stellar radio sources) are a subset of AGN (see reviews by Peterson (1997); Ho (2008); Padovani et al. (2017)) and they are systems in which the

host galaxy is not resolved and the AGN light completely dominates the detected emission. Though originally associated with radio emission, the detection in the radio is no longer required for a source to be identified as a quasar. These quasars were star-like objects but extremely distant (Schmidt, 1963), implying that they had luminosities exceeding 10^3 times the typical luminosities of galaxies. In many cases, these sources featured jets spanning from sub-pc to kpc, some of which showed (apparent) superluminal motion, which can be explained by a jet pointing towards us and moving with velocities close to the speed of light, c (Rees, 1966). Another feature of these sources was that they showcased time variability on time-scales of the order of weeks. All these pieces of evidence pointed towards the possibility that these quasars were powered by SMBHs (Salpeter, 1964; Lynden-Bell, 1969). Since then, a plethora of systems that fall under the category of AGN have been observed at all wavelengths. Furthermore, systems which were observed before the first quasars, such as Seyfert galaxies (Seyfert, 1943) were re-catalogued as AGN.

AGN emit across the entire electromagnetic spectrum. The diversity of observed AGN can be explained by a small number of physical parameters, such as the mass of the central supermassive black hole (SMBH), the rate of gas accretion onto the black hole, the orientation of the accretion disk with respect to our line-of-sight, the degree of obscuration of the nucleus by dust, and the presence or absence of jets. This is called the unified model of AGN (Antonucci, 1993; Urry & Padovani, 1995) a diagram of which can be seen in Figure 1.1. The AGN consists of an accreting black hole surrounded by an accretion disk which feeds the black hole. The accretion subsequently also fuels the disk where potential energy is converted into heat that we observe as emission in the UV and optical, with decreasing temperatures away the central black hole. The accretion disk and black hole are then surrounded by a cooler dusty torus which is capable of obscuring the central engine (and therefore obscuring the UV and optical emission) of the AGN depending on the orientation of the AGN with respect to our line of sight. This model is however an oversimplification of the observed variety of AGN evolving through cosmic time (see Cowley et al., 2016, 2018).

Emission from hot dust is associated with the torus of gas and dust surrounding

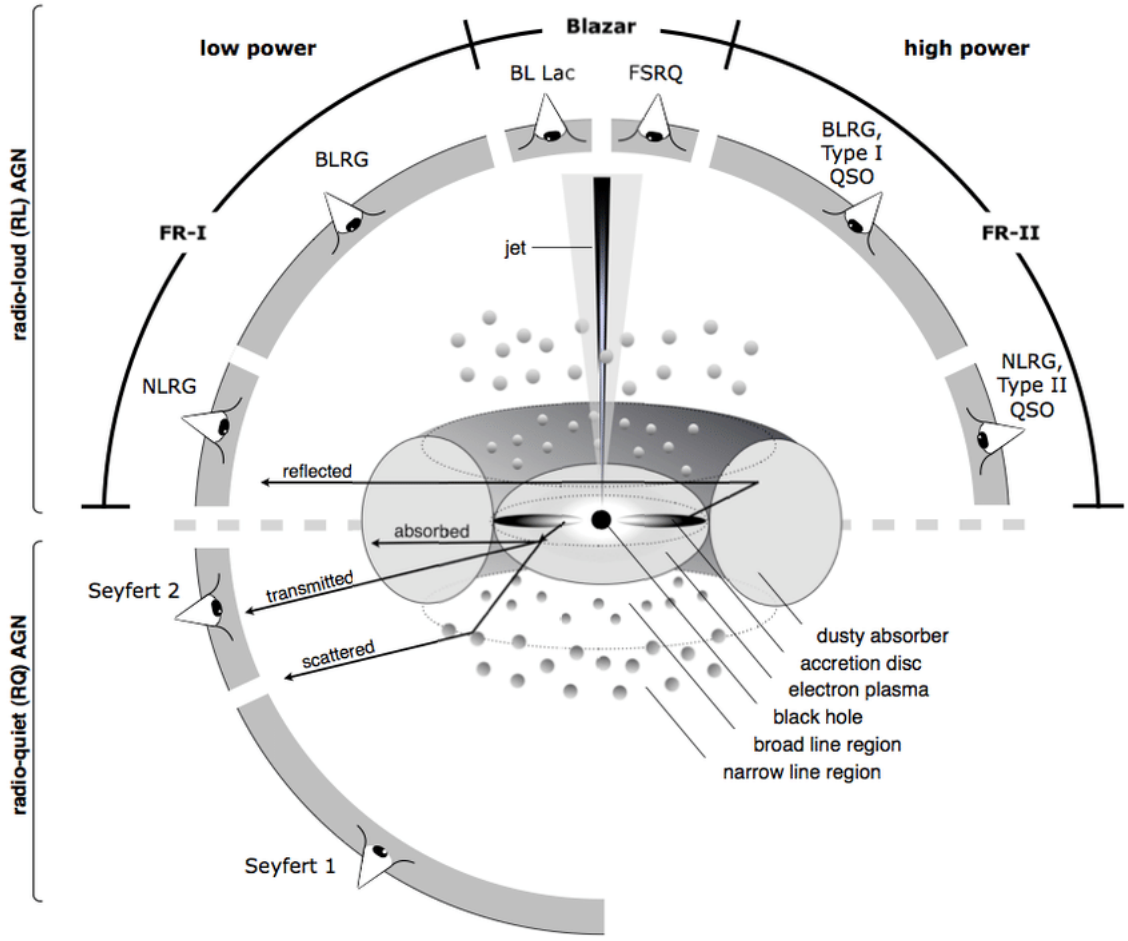

Figure 1.1: The unified model of AGNs. Credit: Beckmann & Shrader (2012).

the central engine of the AGN and most often observed in the mid-infrared (mid-IR) (Antonucci, 1982, 1984; Sanders et al., 1988). The distribution (smooth, clumpy or polar) and kinematics (static, inflowing or outflowing) of this hot dust are however still uncertain. For instance, at parsec scales in the polar regions there exist grains, thought to be irradiated by the AGN almost directly (Raban et al., 2009; Hönig et al., 2012, 2013; Tristram et al., 2014; Asmus et al., 2016; López-Gonzaga et al., 2016; Leftley et al., 2018; Hönig, 2019), which may be associated with an AGN-driven outflow (Schartmann et al., 2014). The properties of these grains observed in the torus and

polar regions appear to be different from those observed in the interstellar medium (ISM), with a dearth of smaller grains such as small graphite grains and/or polycyclic aromatic hydrocarbon (PAH) nanoparticles, indicated by the absence of a 2175 Å bump (Czerny et al., 2004; Gaskell et al., 2004; Gaskell & Benker, 2007), whilst retaining larger grains such as silicate.

1.1.1 Multi-wavelength view

With the ability to identify AGN in every band comes a plethora of different names for AGN, dependent on what band they were first observed in. See Table 1.1 for an overview of the most well-known names and a short description of each.

1.1.1.1 X-ray

The source of X-rays in AGN is thought to be inverse Compton scattering of the accretion disk photons to X-ray energies via the accretion disk corona. Inverse Compton scattering is when low energy photons become high energy photons by transfer of energy from ultra-relativistic electrons. Lower energies of X-rays are due to thermal emission from the inner regions of the accretion disk (Padovani et al., 2017). This X-ray emission then interacts with matter in the nuclear region via processes such as reflection and scattering.

X-ray emission is often separated into soft (0.5 – 2 keV) and hard (2 – 10 keV) X-ray bands. Hard X-ray surveys appear to be near complete (Padovani et al., 2017), as X-ray emission almost always implies the presence of an AGN, except at low luminosities ($L_X < 10^{40}$ erg s^{-1}). This is believed to be due to hard X-rays being able to penetrate through large densities of gas and dust and because X-ray emission from the host galaxies is generally weak when compared to the emission from AGN. This means that X-rays are sensitive to both unobscured and moderately obscured AGNs.

Table 1.1: The main classes of AGN and their properties. For a longer list of AGN classes see Table 1 in Padovani et al. (2017).

Class	Meaning/Properties
Quasar	Quasi-stellar radio source. Very distant and bright AGN. The radio detection is no longer required.
Seyfert 1	Galaxies with an AGN that showcase emission lines with FWHM $> 1,000$ km s^{-1}
Seyfert 2	Galaxies with an AGN that showcase emission lines with FWHM $< 1,000$ km s^{-1}
QSO	Quasar-like. Non-radio source. Interchangeable with Quasar.
QSO2	High power Seyfert 2
RQ AGN	Radio quiet AGN
RL AGN	Radio loud AGN
Jetted AGN	AGN with strong relativistic jets
Non-jetted AGN	AGN without strong relativistic jets
Type 1	Seyfert 1 and Quasars
Type 2	Seyfert 2 and QSO2
FR I	Fanaroff-Riley class I jetted radio core-brightened source
FR II	Fanaroff-Riley class II jetted radio edge-brightened source
Blazar	AGN whose relativistic jets point towards the Earth. They produce both synchrotron and inverse Compton γ rays and are extremely variable over short timescales.

Contaminants in the X-ray band are mainly within the lower X-ray luminosity AGNs, where the host galaxy can start to contaminate the AGN emission. The main source of this emission in the host galaxy is from X-ray binaries. Another source is thermal emission from hot gas either from the host galaxy itself or a galaxy cluster (e.g. Boroson, 2011). These can, however, be selected against by observing in a higher energy (hard) X-ray band where these processes are not as significant.

1.1.1.2 UV and optical

Accretion onto a SMBH gives rise to high X-ray to optical luminosities, a characteristic rest-frame UV/optical power-law continuum, that is distinguishable from the continuum of non-active galaxies, as well as bump in the SED aptly called the "big blue bump". Optical photometric surveys are able to identify more AGN than those at other wavelengths, however, this comes with a bias towards the brighter and unobscured versions of AGN, mainly quasars and type 1 galaxies. This is because in type 2 galaxies the optical (and UV) emission from the accretion disk is absorbed by the dusty torus and re-emitted into the IR.

The main contaminants of the optical sources, despite AGN being bright in the optical when not obscured, are stars. This is mainly due to stars far outnumbering the numbers of AGN, as well as quasars being point-like just like stars. However, this is dependent on the area of sky. Stars are mainly contaminants when looking through the disk of our Milky Way or though close by galaxies, such as the Magellanic Clouds.

Optical spectroscopy is relied upon to provide confirmation that a source is an AGN, and to determine its redshift. Unlike spectra of stars or galaxies, AGN spectra cannot be described in terms of blackbody emission at a single temperature, or as a composite over a small range in temperature. The presence of narrow emission lines in the spectra of galaxies and their ratios are indicative of the presence of an AGN (e.g. Feltre et al., 2016). Broad emission lines come from cold material close to the central SMBH. They are broad because the material is revolving around the SMBH at high speed causing a range of Doppler shifts of the emitted photons (Peterson, 1997).

Narrow optical emission lines come from cold material further away from the SMBH, where they are less affected by the SMBH's gravity, and so produce narrower lines.

The information provided by spectral lines is particularly rich. SMBH mass scaling relations can be used to estimate the masses of SMBHs powering type 1 quasars (e.g. Vestergaard & Peterson, 2006)). This requires emission line widths, the assumption that this width is dominated by gravitational effects, having an estimate of the characteristic radius of the emitting gas assuming $R \propto L^{\sim 0.5}$ (Bentz et al., 2009) and lastly to calibrate this information against the few dozen objects for which there are "reverberation mapping" (Peterson, 1993) data.

In the optical band there are multiple emission lines that can be used to estimate the masses of SMBHs (e.g. Vestergaard & Peterson, 2006; Trakhtenbrot & Netzer, 2012; Baron & Ménard, 2019), depending on the redshift of the AGN. The closest AGN display Hα emission lines. For local AGN, which includes AGN up to $z \sim 0.7$, optical spectra display Hβ emission lines. Mg II lines have been used to extend the relation to SMBH mass up to $z \sim$1.9. However, for higher redshifts either other emission lines have to be used, or the spectrum could be observed at near-IR wavelengths.

1.1.1.3 IR

Studies into AGN in the infrared (IR) band have had a powerful impact on the understanding of the evolution of AGN over cosmic time, their place in galaxy evolution and the overall structure of AGN. The infrared regime is often separated into near-IR (NIR: 1 – 3 μm), the mid-IR (MIR: 3 – 50 μm), and the far-IR (FIR: 50 – 500 μm).

As seen in Figure 1.1, AGN models predict the presence of a dusty torus surrounding the accretion disk on scales larger than the broad line region. Emission from the accretion disk is absorbed by this dust and then reprocessed into the IR, dominating the spectral energy distribution (SED) from the near-IR to the mid-IR. This IR emission played an important role in the unification of AGN, as it established, through polarisation studies, that the difference between type 1 and 2 AGN is an effect of the orientation of this dusty torus with respect to our line of sight.

At rest-frame near-IR wavelengths, the AGN emission is at the crossover point between decreasing accretion disk emission and the beginning of the increase in emission from the dusty torus (Padovani et al., 2017). The mid-IR is where the dusty torus emission dominates. Star formation can be a contaminant in the IR regime; this is because the AGN host galaxies also emit dust emission with luminosities to rival that of an AGN, however this emission is more likely to dominate the far-IR (e.g. Hatziminaoglou et al., 2010). Therefore, mid-IR is optimal for AGN identification. Compared to optical wavebands, it is less sensitive to the obscuration of the central engine of the AGN by dust, as dust opacity is lower at longer wavelengths, which makes mid-IR better for identifying obscured AGN. Although, this does not apply to AGN with increasing redshift (Assef et al., 2011).

Selection criteria for IR photometry observation rely on colours to separate AGN from stars and inactive galaxies, as AGN are expected to be significantly redder in the mid-IR bands (Stern et al., 2005a). The most popular selection criteria were developed for data from satellites such as the Spitzer Space Telescope (Werner et al., 2004a) (for example, see Stern et al., 2005a) and the Wide-field Infrared Survey Explorer (WISE, Wright et al., 2010) (for examples see Stern et al., 2012; Mateos et al., 2012; Assef et al., 2013).

The selection criteria are not perfect as there are a number of different sources that can mimic the colours of AGN in the IR bands. At $z \sim 0.2$, star-forming galaxies with powerful polycyclic aromatic hydrocarbon (PAH) emission can be red enough in colour to be confused with AGN in some selection criteria (e.g. Stern et al., 2005a; Assef et al., 2010). They are typically a minor contaminant, but they affect both deep and shallow observations. A prominent contaminant in the deeper surveys are massive galaxies that have $z > 1$. This is because at rest-frame 1.6 μm there is a stellar bump, which at $z > 1$ is redshifted into the mid-IR range, making the colour of the massive galaxies red enough to be confused with AGN.

Contaminants, however, are not limited to only extragalactic sources. Brown dwarfs, while only accounting for a small fraction of contaminants, could affect the identification of high-redshift ($z > 5$) AGN (Stern et al., 2007). Young stellar objects

can also showcase colours that can be confused with AGN (Koenig et al., 2012). There are also populations of stars that can be confused with AGN, such as carbon and evolved stars.

Near-IR and mid-IR photometry are very important for the identification of the earliest of quasars in the Universe, which are not observable in the optical, due to the Lyα forest absorption and the Lyman break being redshifted into the optical at these redshifts. For example, 77 quasars with $z > 5.6$ were recently discovered (63 of which were newly discovered) by using a combination of photometric near-IR and mid-IR observations (Bañados et al., 2016) from 2MASS (Skrutskie et al., 2006), UKIDSS (Lawrence et al., 2007), VHS (McMahon et al., 2013) and WISE (Wright et al., 2010) to select them for follow-up spectroscopic observations.

IR spectroscopy has become a powerful tool in identifying AGN. This is because it allows the detection of AGN even when the spectrum is dominated by host galaxy emission. This is because of strong spectral features such as the well known silicate feature at 9.7 μm, and Hα, Hβ, etc. The number of objects studied through spectroscopy is however limited due to the significantly longer integration times compared to photometry, and ground based observations are generally limited to the brightest targets which are generally $z < 1$, due to the Earth's atmosphere.

1.1.1.4 Radio

The first radio surveys of the sky, such as 3C (Edge et al., 1959) and Ohio (Ehman et al., 1970), were the cause for the discovery of the first quasars. Most of the sources that were detected at a high Galactic latitude were identified with resolved galaxies. Radio-selected AGN are typically observed from $\lambda \sim 30$ m (10 MHz) down to $\lambda \sim 1$ cm (30 GHz) (Padovani et al., 2017). Synchrotron emission is the dominant process in the radio band. It is the radiation caused by charged particles gyrating at relativistic velocities through magnetic fields. This emission is usually parameterised by a power law:

$$S_\nu \propto \nu^{-\alpha} \tag{1.1}$$

where S_ν is the flux density at frequency ν, and α is the spectral index, which is often used to characterise the radio sources and can be calculated:

$$\alpha = \log_{10}(S_1 - S_2)/\log_{10}(\nu_1 - \nu_2) \tag{1.2}$$

where S_1 and S_2 are measures of flux density at frequencies ν_1 and ν_2 respectively. The main sources of synchrotron emission in galaxies are the processes related to the SMBH and supernova remnants, resulting in two galaxy populations, star forming galaxies and AGN. Star forming galaxies, however, are generally weaker sources of radio emission and only become more prominent at the faintest radio flux densities. A class separation that is based on the spectrum produced by multiple radio continuum frequency observations are the steep ($\alpha \sim 0.7$) and flat ($\alpha \sim 0$) radio spectrum sources (Peacock & Gull, 1981), as well as sources with positive α that are generally due to thermal emission and are fainter than the steep and flat spectrum sources.

The morphology of radio sources is generally divided into two classes, extended and compact (Peterson, 1997). Extended sources are spatially resolved and usually consist of two lobes of radio emission located roughly symmetrically on either side of the centre of the galaxy. These lobes can extend up to lengths as large as megaparsecs. Compact sources are unresolved at $\sim 1''$ and their position generally coincides with the centre of a galaxy. Compact sources are generally associated with flat spectrum sources and extended sources are generally associated with steep spectrum sources. The steep spectrum is caused when the plasma ejected from an active AGN loses energy through synchrotron radiation (and inverse Compton scattering). A consequence of this is that high energy particles that radiate the most at high radio frequencies lose their energy fastest and the low energy particles that radiate the most at low radio frequencies lose their energy the slowest, leading to a steep radio spectrum that evolves over time even after the AGN has switched off.

Extended sources can be further divided into two separate classes based on luminosity classes called FR I (type I) and FR II (type II) (Fanaroff & Riley, 1974). FR I are weaker radio sources which are brighter in the centre, with decreasing surface brightness towards the outer edges. FR II however, contrast this by being more lu-

minous and being limb-brightened, often showing regions of enhanced emission either within the jet structure or at the edges of it. See Figure 1.2 for examples of the FR I and FR II galaxies.

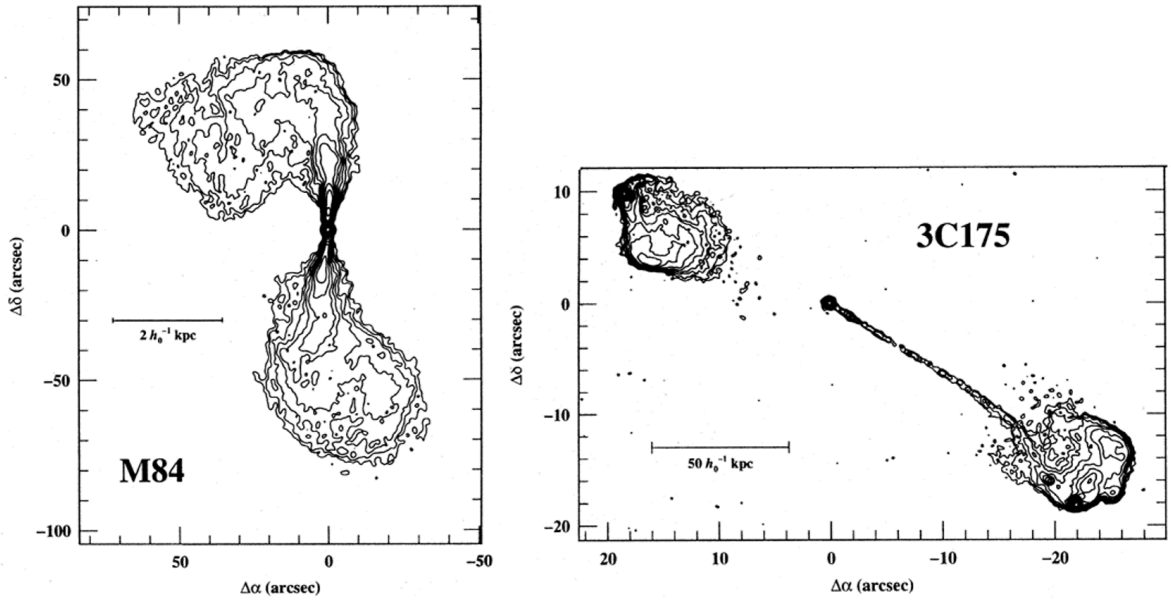

Figure 1.2: Radio contours of a FR I galaxy (left) and a FR II galaxy (right) from Peterson (1997). From this figure it can be seen that the FR I galaxy is brighter in the centre (contours closer together) and that the FR II galaxy is brighter at the outer edge of the lobes.

Another popular radio AGN classification is into radio loud (RL) and radio quiet (RQ). This classification has undergone some changes since it was first introduced as a distinction between quasars dependent on some threshold value in either radio flux density or luminosity (Peacock et al., 1986), or the ratio of radio to optical flux density or luminosity (Schmidt 1970). Padovani et al. (2017) have since argued that RL and RQ AGN represent intrinsically different objects, where RL AGN emit a large fraction of their energy in the radio and are associated with powerful relativistic jets, whereas RQ AGN emit multi-wavelength energy dominated by thermal emission related to the accretion disk. The major physical difference between these classes is the presence of

(or lack there of) strong relativistic jets, which imply that the two classes reach widely different photon energies.

The last well known radio classification is the low-excitation galaxies (LEG) and high-excitation galaxies (HEG) (Hine & Longair, 1979). In general, this classification is based on optical spectroscopy properties, where objects with and without high-excitation emission lines in their optical spectra are referred to as HEGs and LEGs, respectively. Unlike most radio classifications, this also applies to AGN selected in other bands which have a corresponding optical spectroscopy observation. Specifically, quasars and Seyferts have been identified as belonging to the HEG class, while LINERs and absorption line galaxies are classified as LEGs (Padovani et al., 2017).

There is some overlap between the classes, for example, almost all FR I AGN are also LEGs, while most FR II are HEGs. However, whilst FR I HEGs are rare, FR II LEGs are common and can even outnumber the FR II HEGs in deep surveys (e.g. Hine & Longair, 1979; Baldi et al., 2010; Croston et al., 2018). Therefore, the classifications are not one-to one.

Star-formation is the main contaminant to overall radio luminosity. In the local Universe LEGS are found in the red sequence of galaxies, where little to no star-formation is taking place. HEGs, however, occupy the blue sequence of galaxies (Smolčić et al., 2009). Bluer galaxies implies higher star-forming rates, meaning that any radio emission observed could be either partially or entirely due to synchrotron radiation produced by supernova remnants throughout the host galaxy as opposed to the central SMBH. This means that to be truly sure of an AGN classification, properties that are seen in other wavebands must be relied upon.

1.1.2 Variability

AGN that display variability in luminosity have been found in every waveband. Some varied over time scales of a few months, others were found to vary significantly on time scales as short as a few days (e.g. Smith & Hoffleit, 1963). This meant that the radiation must come from a region of order light days in size. When these variable

AGN were first discovered, before AGN were realised to exist, this was perceived as a major problem, since a source with size comparable to the Solar System was emitting hundreds of times as much energy as an entire galaxy.

It has been found that AGN variability is erratic and aperiodic over a range of timescales from minutes to years (Padovani et al., 2017). The speed of this variability is dependent on the waveband it is being observed in, and therefore means it is dependent on what part of the AGN it is being emitted from. The minimum timescale of variability provides us with an estimate of the component size of the emitting source (e.g. Terrell, 1967).

1.1.3 Relation between supermassive black holes and their host galaxies

A black hole has three properties in general relativity: mass, angular momentum and electric charge. Electric charge is unimportant as it just oscillates around zero in the manner of a negative feedback loop. Angular momentum measurements are few and far between (Risalti et al., 2013) due to the difficulties in measuring it directly. However, angular momentum can be measured indirectly as it results in a Kerr metric which allows gas to orbit at much closer range, thus increasing inner disk temperature, that then becomes measurable in X-rays (e.g. Martocchia et al., 2000; Reynolds, 2021). This leaves the mass of the SMBH, M_{BH}, to be used for correlations. Several strong correlations between SMBH masses and the properties of their host galaxies have been observed. These relations imply that the SMBH influences the evolution of its host galaxy or that they co-evolve together.

The most notable relation is the M-σ relation, which connects the SMBH mass to the characteristic velocity of the stars in the host galaxy. The characteristic velocity is the velocity with which stars move within the spheroidal part of the galaxy. For spiral galaxies this is the bulge component only, for an elliptical galaxy the entire galaxy is

taken into consideration (Ferrarese & Merritt, 2000). The M-σ relation is given by:

$$\frac{M_{BH}}{10^8 M_\odot} \simeq A \left(\frac{\sigma}{200\,\mathrm{km\,s^{-1}}} \right)^{\beta}, \tag{1.3}$$

where $A \sim 3$ and $\beta \sim 4 - 6$. The value of β depends on the sample of galaxies being investigated. See Figure 1.3 for a graphical view of this relation.

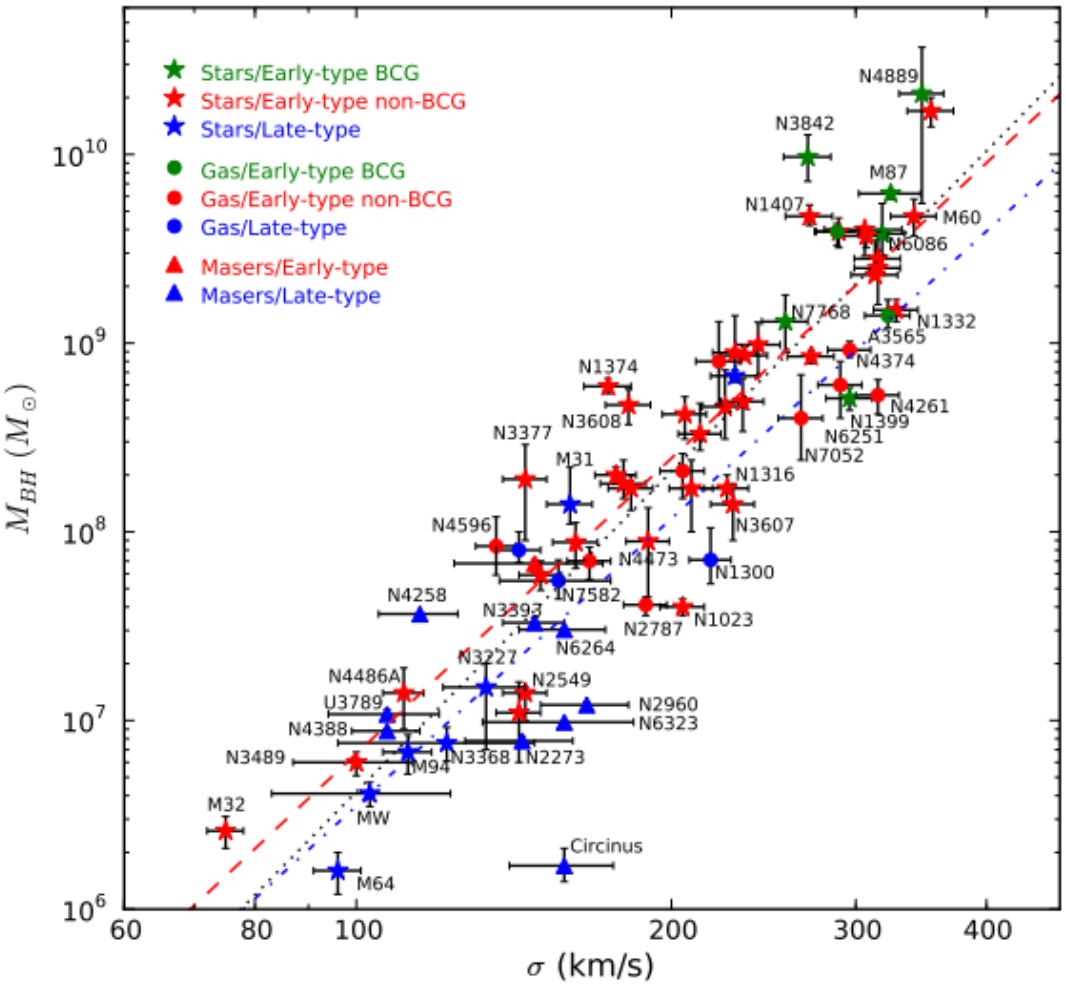

Figure 1.3: Graph showing the relation between black hole mass M_{BH} and stellar velocity dispersion σ. Graph taken from McConnell & Ma (2013) whose sample was a mixture of late and early type galaxies.

The gravitational influence of the SMBH extends only to the inner parts of the

galaxy ($\sim$ several tens of parsecs), whereas the rest of the galaxy can extend to kiloparsecs or more. This gave rise to the belief that the gravity of the SMBH cannot affect the rest of the galaxy, yet this observed relation states otherwise. It implies that the host galaxy is affected by the mass of the SMBH, or alternatively the growth of the SMBH is governed by the same processes that govern the growth of the galaxy.

A number of mechanisms have been proposed to attempt to explain the origin of the M-σ and other observed correlations. They can be separated into roughly three categories: no-connection explanation, where both the galaxies and SMBHs grow via mergers (Peng, 2007); explanations based on the assumption that the host galaxy controls the rate at which the gas is fed to the SMBH, where this rate depends on the gas velocities in the galactic spheroid (Angles-Alcázar, 2015); and feedback mechanisms, based on the energy release by the accretion of matter onto the central SMBH, which has the potential to affect the whole galaxy.

1.1.4 AGN feedback and galaxy evolution

In the context of galaxy evolution, the main importance of the effect of an AGN on its host galaxy is its regulatory effect on its star formation rate, capable of both stopping and starting star formation. AGN feedback has been observed in the form of radiative winds (or outflows) that are generated from the accretion disks of the AGN, or from radio jets that heat, expel or shock their surroundings (Fabian, 2011).

The Eddington luminosity, $L_{\rm edd}$, is the maximum luminosity an object (assuming spherical symmetry) can maintain if radiative forces (from the accretion) are balanced by gravitational forces (of the supermassive blackhole). The Eddington luminosity is found by equating the radiation force and the gravitational force and for an AGN is defined as $L_{\rm edd} = 1.25 \times 10^{38}\ M_{\rm BH}$ erg s^{-1}, where $M_{\rm BH}$ is the mass of the supermassive blackhole. From this we can see that the Eddington luminosity is proportional to the mass of the object providing the luminosity. The accretion rate can be represented by the Eddington ratio, which is defined as $L_{\rm bol}({\rm AGN})/L_{\rm edd}$, where $L_{\rm bol}$ is the observed AGN luminosity. The higher the observed luminosity compared to the Eddington

luminosity, the higher the accretion rate. If an AGN exceeds its Eddington luminosity the gravitational forces are overpowered by the radiative forces, and the AGN is unable to hold on to the material in the outer layers, resulting in a radiation-driven outflow that can then go on to affect its host galaxy.

AGN driven outflows have been observed from pc to kpc scales, traced mostly using the ionised gas, observed from optical to X-ray emission absorption lines (King & Pounds, 2015). They are a common feature of accreting systems, likely created in AGN near the central black hole by a coupling between the magnetic field and accreting material (Tchekhovskoy, 2015). AGN jets transport energy and mass from sub-parsec central regions to Mpc-scale lobes (e.g. Fabian, 2011; Blandford et al., 2019), transporting a huge amount of energy into interstellar and intergalactic space that is capable of suppressing star-formation activity. More recent observations from new or improved telescopes show that gas outflows are multi-phased, i.e. different phases of the gas, detectable at different wavelengths, take part in the outflows.

One type of observed outflows are the Ultra-Fast Outflows (UFOs; Tombesi et al., 2011). They are highly ionised outflows with $v \sim 0.1 - 0.25$ c. They are traced by highly blue-shifted X-ray absorption lines in the iron K band (Pounds et al., 2003). They have been shown to be present in 35% of Seyfert galaxies, and indirect arguments indicate that their location is at scales of $\sim 0.0003 - 0.03$ pc, the inner regions of the AGN outflows.

The more obvious feedback mechanism of AGNs are radio jets which are capable of acting on larger scales than outflows, up to tens of kilo-parsecs. Combining radio observations with X-ray observations has revealed bubbles/cavities in the hot gas, seen in the X-ray, are filled with radio lobes, suggesting the hot gas is displaced by the radio jets emitted by the AGN. This energy injected into the gas surrounding the galaxy has been linked with suppressing gas cooling and therefore quenching star formation (see McNamara & Nulsen (2007) for reviews). In low-powered AGN, radio plasma is considered to be the dominant source of energy to prevent the gaseous atmosphere from cooling back into the galaxy.

Radio jets appear to be capable of producing more extreme gas outflows than

accretion-disk winds/outflows. However, the relative importance of each feedback mechanism, and the dependence on redshift, remains an open area of research.

1.1.5 An extreme AGN sample inspired by SAGE0536AGN

The originator of the strongest 10 μm silicate emission of any known AGN, is the hot dust near the SMBH of SAGE1C J053634.78−722658.5 (hereafter referred to as 'SAGE0536AGN') that was discovered serendipitously behind the Large Magellanic Cloud (LMC) by Hony et al. (2011) in the Spitzer Space Telescope Survey of the Agents of Galaxy Evolution Spectroscopic follow-up of IR sources seen towards the LMC (SAGE-Spec: Kemper et al., 2010; Woods et al., 2011). It lies behind the LMC and was found to be a type 1 AGN with a negligible amount of far-IR emission meaning a lack of star formation, confirmed by spectra obtained with the Southern African Large Telescope (SALT) (van Loon & Sansom, 2015). Finding more of these could provide valuable insight into this stage of galaxy/AGN evolution.

My new spectroscopic observations using the South African Astronomical Observatory (SAAO) 1.9m telescope, reveals SSTISAGE1C J053444.17−673750.1 is one such source that shows similarities to SAGE0536AGN. This source has also been referred to as 4XMM J053444.1−673751, 2MASS J05344418−6737501, SHP LMC 256 or [KWV2015] J053444.17−673750.1 (identifier for post-AGB star candidate), in this paper it shall be referred to as 'SAGE0534AGN'. Both of these sources have been confused as evolved stars, have silicate emission and a lack of star formation. Can more be found? Are they an unusual type, or a short and therefore rarely seen stage of galaxy/evolution?

As these sources mimic evolved stars in the Magellanic Clouds, I therefore needed to adopt a more systematic approach in finding more of them. Unsupervised machine learning has been used to great effect to cluster objects together and reveal patterns in large datasets (e.g. Lochner et al., 2016; Anders et al., 2018; Reis et al., 2018b; Zhang et al., 2020). This can be used to find objects with similar properties to those already discovered, such as SAGE0536AGN and SAGE0534AGN. I am searching for

more examples of these sources behind the Magellanic Clouds.

1.2 Magellanic Clouds

AGN are most readily identified within combinations of multi-wavelength photometric survey data. The Magellanic Clouds span $\sim$ 100 sq. degrees on the sky that have been studied, in parts or as a whole, in the UV (e.g. Thilker et al., 2014), optical (e.g. Gaia, SMASH; Gaia Collaboration et al., 2021; Nidever et al., 2017), IR (e.g. SAGE, AllWISE; Lacy et al., 2004; Cutri et al., 2021), radio (e.g. MOST, ATCA; Mauch et al., 2003b; Murphy et al., 2010) and X-ray (XMM–Newton; Sturm et al., 2013), which makes them an ideal location to search for AGN behind them. Furthermore, they are also located away from the Galactic Plane and Galactic Centre, reducing source confusion in the radio band and extinction at UV/optical/near-IR wavelengths. The combination of all these data has great potential for discovery of the more unusual and extreme cases of AGN, such as SAGE0536AGN. The new and deeper surveys towards the Magellanic Clouds, such as the near-IR VISTA Magellanic Clouds (VMC; Cioni et al., 2011) and radio Evolutionary Map of the Universe all-sky (EMU; Joseph et al., 2019; Pennock et al., 2021) surveys greatly enhance such attempts.

At infrared frequencies we observe thermal emission from stars as well as from dust that has absorbed emission at shorter wavelengths and re-emitted this emission into the IR, such as in areas of star-formation. We also observe red-shifted emission from distant objects such as AGN and galaxies. The Magellanic Clouds have been observed multiple times over the years as part of all-sky surveys such as the near-IR 2MASS (Skrutskie et al., 2006) and mid-IR AllWISE (Cutri et al., 2013) surveys. Furthermore, there have been Magellanic Cloud specific surveys, where depth and angular resolution are an improvement over all-sky surveys. The include surveys observed with Spitzer Space Telescope in the mid-IR as part of the Spitzer Agents of Galaxy Evolution (SAGE) survey of the LMC (Meixner et al., 2006) and SMC (Gordon et al., 2011), and Herschel Space Observatory in the far-IR as part of the HERschel Inventory of

The Agents of Galaxy Evolution (HERITAGE; Meixner et al., 2010). The most recent Magellanic specific survey is the near-IR VISTA Magellanic Clouds (VMC; Cioni et al., 2011) survey which showcases a great improvement in depth and angular resolution compared to previous IR surveys, and has detected stars encompassing most phases of evolution such as main sequence stars, sub-giants, upper and lower red giant branch (RGB) stars, red clump stars, RR Lyræ and Cepheid variables, asymptotic giant branch (AGB) stars, post-AGB stars, planetary nebulæ (PNe) and supernova remnants (SNRs) populations (e.g. Gullieuszik et al., 2012; Ripepi et al., 2015; Zivkov et al., 2018; Groenewegen et al., 2019; Zivkov et al., 2020; Groenewegen et al., 2020; Cusano et al., 2021; Choudhury et al., 2021) that can be used to help assess the age and metallicity within the Magellanic systems. This survey has also had success in discovering background extragalactic sources (Cioni et al., 2013; Ivanov et al., 2016; Pennock et al., 2022).

At radio frequencies, we typically observe free–free emission from ionized gas in H II regions and planetary nebulæ (PNe) and synchrotron emission from supernova remnants (SNRs), as well as from active galactic nuclei (AGN) and star-forming galaxies in the background. These radio emitting sources have been observed multiple times over the years by telescopes such as MOST (Molongo Observatory Synthesis Telescope; e.g. Large et al., 1981; Mills, 1985; Mauch et al., 2003b), ATCA (Australia Telescope Compact Array; e.g. Kim et al., 1998; Dickel et al., 2005; Hughes et al., 2007; Murphy et al., 2010), MWA (Murchison Widefield Array; e.g. For et al., 2018) and Parkes (Griffith & Wright, 1993; Filipovic et al., 1995, 1996, 1998a,b,c,d; Kim et al., 2003). A new generation of radio telescopes, including the Australian Square Kilometre Array Pathfinder (ASKAP; Johnston et al., 2008; Hotan et al., 2021) can improve upon the resolution, sensitivity and speed of these observations. ASKAP studies of the Small Magellanic Cloud (SMC) and LMC (Joseph et al., 2019; Pennock et al., 2021) provide greater details on SNRs, PNe, (super)bubbles and their environments, young stellar objects (YSOs), symbiotic (accreting compact object) binaries, AGN and star-forming galaxies. Despite the increased stellar confusion in the direction of the Magellanic Clouds, AGN and galaxies dominate the source counts at radio frequencies.

Whilst much of the Magellanic Clouds are transparent enough to see through, looking for AGN in the direction of the Magellanic Clouds has some challenges. For example, looking in the direction of the Magellanic Clouds leads to an increase in stellar contamination, more prominent at certain wavelengths than others, as well as complex background and extinction (especially in star forming regions), that must be separated from background sources. Therefore, a more systematic approach in finding them needs to be adopted.

1.3 Machine learning

Machine learning algorithms are a use of artificial intelligence to automate tasks, such as identification and classification, on large sets of data that would otherwise prove time consuming. They can also replace subjective approaches that depend on user choices by objective approaches that are data driven. In the field of astrophysics, due to the increase in data size and complexity coming from the latest and future surveys, machine learning has gained in popularity to detect, characterise and classify objects within these datasets.

Machine learning algorithms are usually divided into two types, supervised and unsupervised.

1.3.1 Supervised Learning

Supervised machine learning algorithms predict classifications/values based on example data with known classifications/values. They do this by analysing a known dataset, the training set, and producing a model from this dataset, which can then be used to make predictions of the output of an unseen dataset. A disadvantage of supervised learning is that it is only as good as the data it is trained upon, and is therefore not best suited to finding new or unusual objects.

The most popular supervised machine learning algorithms used in astronomy

are support vector machines, ensemble methods and neural networks. Support vector machines take a dataset with N features, then finds a hyper-plane (boundary of (N−1) dimensions) in the N-dimensional space that best separates the given classes (e.g. Qu et al., 2003; Huertas-Company et al., 2008; Kim et al., 2011; Małek et al., 2013; Cheng et al., 2020). Ensemble methods combine several supervised learning techniques into a single model, resulting in an overall improved performance compared to using a single algorithm (e.g. Carliles et al., 2010; Pichara et al., 2012; Reis et al., 2018b; Vavilova et al., 2021). Artificial neural networks are inspired by the biological neural networks that make up the human brain, and consist of layers of connected neurons (or nodes) where each neuron receives the weighted inputs from all the neurons in the previous layer and applies a function to it, then passes the output onto all the neurons in the the next layer. Their flexible, non-linear structure allows these algorithms to perform a variety of supervised machine learning tasks such as classification and regression (e.g. Brescia et al., 2013, 2014; Ellison et al., 2016; Bilicki et al., 2018; Kovačević et al., 2020), as well as a variety of unsupervised machine learning tasks.

1.3.2 Unsupervised Machine Learning

Unlike supervised machine learning, unsupervised machine learning does not require prior knowledge in the form of a training dataset. The input to these algorithms are a list of objects with measured properties (features). These algorithms can be used to detect clusters, complex relations, outliers or reduce the dimension of datasets.

Clustering algorithms group objects in a dataset, such that objects in the same group/cluster are more similar to each other than to objects in other groups. Dimensionality reduction algorithms reduce the number of features in the original dataset, either by selecting a subset of the features that best describe the dataset, or by constructing a new set of features that provide a good description of the dataset. Both clustering and dimensionality reduction algorithms can be used to visualize and interpret complex high-dimensional datasets and uncover hidden trends and patterns in large datasets (e.g. Lochner et al., 2016; Anders et al., 2018; Reis et al., 2018b; Zhang

et al., 2020).

1.4 Scientific questions and thesis outline

The Magellanic Clouds are a well documented part of sky with surveys that cover from X-ray to radio. Finding more AGN within such a data rich resource would allow us to see the larger picture of AGNs instead of only seeing them in the partial views from the wavebands in which they were discovered. This means that we can better compare them over cosmic time, whilst also comparing them with non-AGN galaxies. Here are some of the scientific questions that will be explored:

- What is the role of AGN in galaxy evolution?
- How does IR emission relate to radio emission in AGN?
- How does the timing of the switching off of star formation in galaxies relate to the timing of the switching off of the AGN?
- How does AGN dust content change over the life cycle of an AGN?
- How do extreme silicate emitting AGN fit into the life cycle of an AGN and the evolution of its host galaxy?

In Chapter 2 the data, both photometric and spectroscopic, used throughout this thesis will be outlined in order to familiarise the reader with the benefits and limitations these data provide. In Chapter 3 the methods and techniques used to analyse and extract information from the data will be introduced and described.

Chapter 4 contains the exploration of the objects found in the radio ASKAP 888 MHz image of the LMC and the point source catalogue I created from it (the details of the catalogue creation are shown in Chapter 3). First, I briefly discuss the Magellanic and Galactic foreground sources. Then, I discuss the extragalactic sources seen beyond the LMC and lastly, I determine and discuss the spectral indices of sources in the new

catalogue. This data will be used in later chapters as part of a dataset to be run through an unsupervised machine learning algorithm to find more radio detected AGN and to compare radio properties against IR properties of AGN.

In Chapter 5, I use the unsupervised machine learning algorithm, t-SNE (described in Chapter 3), on the photometric survey data of the SMC to discover a sample of AGN similar to the extreme silicate emitter SAGE0536AGN. In Section 5.1 I describe the light curves and spectra, calculation of black hole masses, spectral energy distribution (SED) fitting with CIGALE, X-ray observations and modelling of those sources where we can see their host galaxies with GALFIT. In Section 5.2 I discuss the selection techniques of AGN (Section 5.2.2) and where this sample and sources mistaken for AGN fall within them. This is followed by a discussion of the sample galaxies' identity as either star-forming, quiescent or green valley galaxy and how their properties change as they transition from star-forming to green valley. The radio properties and how they link to the evolutionary stage of the sample are then discussed followed by a discussion of the AGN dust and its effect on observed properties such as variability and the 10 μm silicate emission.

In Chapter 6 I further expand on the work done in previous chapters. Instead of looking at only a small sample of sources, as in Chapter 5, I look at the entirety of the Magellanic Clouds VMC fields. The t-SNE algorithm is used on the radio sources of the Magellanic Clouds (where the LMC sources are from the work in Chapter 4) and the distribution of different classes are analysed by comparing with photometric properties and previous classifications made as part of the Gaia DR3 release. Following this I explore the results of using the supervised machine learning algorithm, probabilistic random forest, used on the SMC VMC field photometric data. The sources in the direction of the SMC are separated into classes, the distribution and properties of which are then further explored.

Chapter 7 discusses the results and conclusions from the previous chapters and provides possible avenues for future work.

2 Data

This chapter is split into three sections. The first section describes the photometric survey data used in this work; the data are chosen as they are the current highest quality datasets that also cover the entirety, or large parts, of the Magellanic Clouds. The second section describes the spectroscopic data, obtained by me on observing runs or from past literature, that are used to classify and characterise specific objects in this research. The last section is on the data sets of spectroscopically classified stellar and extragalactic sources in the direction of the Magellanic Clouds used to make up the training data for the machine learning classifier.

2.1 Photometric Surveys

Photometric surveys record the intensity of light emitted over a specific wavelength range, dependent on the filters used, over a large area of the sky. This allows for the observation of thousands to millions of sources, depending on total area and depth of the observation. The downside of photometric surveys is that the wavelength resolution is only as good as how narrow the filter bandpasses are, meaning individual spectral features are generally not observed like they would be in spectroscopic observations.

AGN SEDs typically display a distinct broad bump in the infrared, that represents the thermal emission from warm to hot dust in the torus, which is heated by the emission from the accretion disk. However, star-formation from galaxies can also emit in the IR, though this tends to be at longer wavelengths. Obscured and unobscured AGN can both be detected in the IR making it an ideal wavelength regime to search for AGN within. Radio emission also shares this feature of being able to detect both obscured and unobscured AGN, though not all AGN emit strongly in the radio.

The photometric survey data used in this research was selected as the current best quality data with the best coverage of the Magellanic Clouds. See Table 2.1 for the photometric surveys used throughout this work.

Table 2.1: Surveys used in this work.

Survey	Bands	Sensitivity	AGN features
VMC	$YJK_{\rm s}$	22, 22, 21.5 mag (Vega)	Hot dust from inner torus
EMU ASKAP	888, 960, 1320 MHz	58, 186, 165 μJy beam^{-1}	Jet/lobes
XMM-Newton	0.2 – 12.0 keV	$\sim 10^{-14}$ erg cm^{-2} s^{-1}	Corona
Gaia DR3	G, G_{BP}, G_{RP}	$\sim$ 21 mag (Vega)	Accretion disk
SMASH	$ugriz$	23.9, 24.8, 24.5, 24.2, 23.5 mag (Vega)	Accretion disk
SAGE	3.4, 4.5, 5.8, 8.0 μm	18.3, 17.7, 15.7 and 14.5 mag (Vega)	Torus
AllWISE/unWISE	W1, W2, W3, W4	16.9, 15.95, 11.50, 8.05 mag (Vega)	Torus

2.1.1 VISTA near infrared imaging

The Visible and Infrared Survey Telescope for Astronomy (VISTA; Dalton et al., 2006) telescope is a 4.1-m near-infrared optimised telescope, which is combined with the VISTA InfraRed CAMera (VIRCAM) and equipped with a large array of 16 infrared detectors that fill about a 1.5 square degree field. It is capable of observing with broad band filters $ZYJHK_{\rm s}$ and narrow band filters at 0.98, 0.99 and 1.13 μm and designed to work for seeing as good as $\sim 0\farcs34$.

A simple diagram of VISTA can be seen in Figure 2.1. After the incoming light reflects off the primary and then secondary mirrors, the light enters through a 95-cm entrance window and passes through three corrector lenses and the filter wheel, to reach the detectors. The 16 detectors of VIRCAM produce 16 images of the sky, with equal sized gaps between the 16 images, which are called a pawprint. In order to produce a contiguous area of sky, multiple offsetted pawprints are combined to fill in the gaps. The combined image is called a tile, and is obtained with a minimum of 6 exposures, which gives a ~ 1.64 deg^2 field of view (FOV).

The VISTA Survey of the Magellanic Clouds (VMC; Cioni et al., 2011) is a near-IR deep, multi-epoch and wide-field study of the Magellanic Clouds, covering an area of about 184 deg^2. VISTA observations for the VMC main survey started in November 2009 and ended in October 2018. It has a spatial resolution of $1\farcs0$ – $1\farcs1$, $0\farcs9$ – $1\farcs0$ and $0\farcs8$ – $0\farcs9$ in the $YJK_{\rm s}$ filters, respectively, where the two values specified for seeing indicate seeing for crowded and uncrowded regions, respectively. It also reaches a sensitivity of about 22, 22 and 21.5 mag (Vega) in the $YJK_{\rm s}$ bands, respectively. Its depth and coverage can be compared to the VISTA Deep Extragalactic Observations (VIDEO; Jarvis et al., 2013) survey, which was specifically designed to study galaxy and cluster/structure evolution up to z 4 in a 12 deg^2 area, reaching depths of about 24.5, 24.4 and 23.8 mag (Vega) in the $YJK_{\rm s}$ bands, respectively. The VMC data provide an opportunity to expand on the effort of the VIDEO survey and cover more volume and better overcome cosmic variance, and has already proven successful in discovering more AGN (e.g. Ivanov et al., 2016). This however comes with the caveat

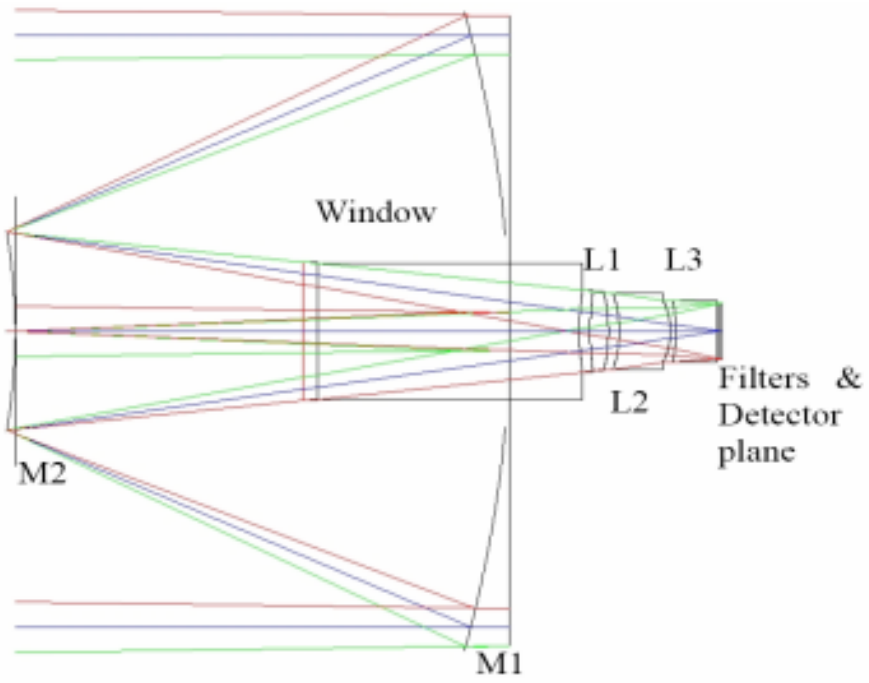

Figure 2.1: Optical layout of the VISTA telescope from the Very Large Telescope Paranal Science Operations VIRCAM/VISTA User Manual[a]. M1 and M2 are the telescope primary and secondary mirrors. The camera's entrance window, the three lenses L1, L2, and L3, the filter and the detector planes are also marked.

[a] `https://www.eso.org/sci/facilities/paranal/instruments/vircam/doc/VIS-MAN-ESO-06000-0002_v108.pdf`

of increased stellar confusion with the presence of the LMC and SMC.

From the images of the Magellanic Clouds the photometry can be measured. Aperture photometry magnitudes are obtained by summing the observed flux within a given radius from the centre of an object (the aperture), then subtracting the total contribution of sky background within the aperture, estimated as an average from each pawprint, leaving only the flux from the object. PSF photometry is best used in crowded regions, such as the central regions of the Magellanic Clouds, and assumes that point sources can be modelled by a point spread function (PSF). The PSF is modelled on isolated point sources (stars), and is then used to fit and measure the flux of other sources across the image.

The catalogues created from the VMC survey provide both aperture and PSF photometry, where PSF photometry reaches sources on average 0.3 magnitudes fainter than aperture photometry. The PSF catalogue is created as described in Rubele et al. (2015). The magnitudes in each band are calculated from deep tile images, which are a combination of single exposure images from different epochs. Due to the PSF photometry's increased depth and ability to distinguish sources in crowded regions, it is the PSF photometry that is used in this work.

2.1.2 ASKAP

Radio telescopes operating at medium ($\sim$ 1 GHz) to high frequency consist of an antenna, a parabolic dish that reflects the radio waves to a receiver, and amplifier, that boosts the very weak radio signal to a measurable level, which is then directed to a recorder to keep a record of the signal. The size of the parabolic dish determines the size of the collecting area. The larger the collecting area the fainter the source that can be detected and better the angular resolution. A telescope array is a group of telescopes arranged so that when combined they function similarly to one giant telescope, allowing for higher resolution data with greater sensitivity to faint signals than a single telescope.

The Australian Square Kilometre Array Pathfinder (ASKAP; Johnston et al.,

2008; Hotan et al., 2021) is an array of 36 12-m diameter dish antennas spread across $\sim$ 6 square kilometres, and therefore an effective maximum baseline of $\sim$ 6 km (where the longer the baseline, the better the resolution and ability to resolve fine structures) corresponding to a point spread function of $\sim 10''$ ($\sim$ 1 GHz).

ASKAP's antennas feature three-axis movement and use 'phased array feeds' (or 'radio cameras') with many separate, simultaneous beams for detecting radio waves rather than 'single pixel feeds' that previous telescopes use, allowing ASKAP to survey large areas of sky with unprecedented sensitivity and speed. Each dish antenna has a phased array feed (PAF) that can be used to form 36 primary beams, giving each dish a wide FOV of 30 square degrees each. Each PAF is comprised of 188 receiving elements sensitive to radio frequencies from 700 – 1800 MHz. The wide field of the array combined with high-speed digital processing systems and supercomputing facilities allow for rapid survey observations.

The Evolutionary Map of the Universe (EMU; Norris et al. (2011)) is a wide-field radio continuum survey which uses the ASKAP telescope. EMU's primary goal is to make a deep (rms $\sim$ 10 μJy/beam) radio continuum survey of the Southern sky, extending as far North as $+30^\circ$ declination, with a resolution of $10''$. It is expected to catalogue about 70 million galaxies, including AGNs up to the edge of the visible Universe.

2.1.2.1 ASKAP ESP SMC image

Two radio continuum images from the ASKAP survey in the direction of the SMC were taken as part of the EMU Early Science Project (ESP) survey of the Magellanic Clouds (Joseph et al., 2019). The two source lists that were produced from these images by Joseph et al. (2019) contain radio continuum sources observed at 960 MHz (4489 sources) and 1320 MHz (5954 sources) with a bandwidth of 192 MHz and beam sizes of $30'' \times 30''$ and $16\farcs3 \times 15\farcs1$, respectively. The median RMS noise values were 186 μJy beam^{-1} (960 MHz) and 165 μJy beam^{-1} (1320 MHz). The observations of the SMC were made with only 33 per cent and 44 per cent (for 960 MHz and 1320

MHz respectively) of the full ASKAP antenna configuration and 66 per cent of the final bandwidth that was available in the final array, with which the LMC was observed.

2.1.2.2 ASKAP ESP LMC image

The LMC was observed at 888 MHz with a bandwidth of 288 MHz taken on 2019 April 20 using ASKAP's full array of 36 antennas (scheduling block 8532). The LMC was observed as part of the ASKAP commissioning and early science (ACES, project code AS033) verification (DeBoer et al., 2009; Hotan et al., 2014; McConnell et al., 2016) in order to investigate issues that were found in higher-frequency higher-spectral-resolution Galactic-ASKAP (GASKAP; Dickey et al., 2013) survey observations, as well as to test the rapid processing with ASKAPsoft (Whiting, 2020). The observations cover a total field of view of 120 deg^2, with a total exposure time of ∼12h40m. It is compiled by four pointings (∼3h10m each) with three interleaves[1] each to result in more uniform depth across the field – effectively 12 pointings. The three interleaves overlap by ∼0.5° to improve the uniformity of sensitivity across the field. The ASKAP ESP image of the LMC can be seen in Figure 2.2. Figures 2.3 and 2.4 show close ups of well known LMC regions 30 Doradus (the Tarantula Nebula) and LHA 120-N 119, an H II region near the kinematic centre of the LMC. The largest angular scales that can be recovered in this survey are 25 – 50′ (McConnell et al., 2020). The LMC exhibits structure on scales larger than can be recovered by the shortest baselines available with ASKAP (22 m). The missing short-spacing data results in negative bowls around some regions, for example, this is particularly evident around the extended bright structure in 30 Doradus.

Although the image noise levels are near to expected values in most areas, a common feature of early ASKAP data has been the presence of some low-level artefacts (at ∼ 1 per cent peak brightness) very close to bright sources (a few hundred mJy and

[1]interleaves are overlapping pointings where the telescope slews between them at a more rapid cadence.

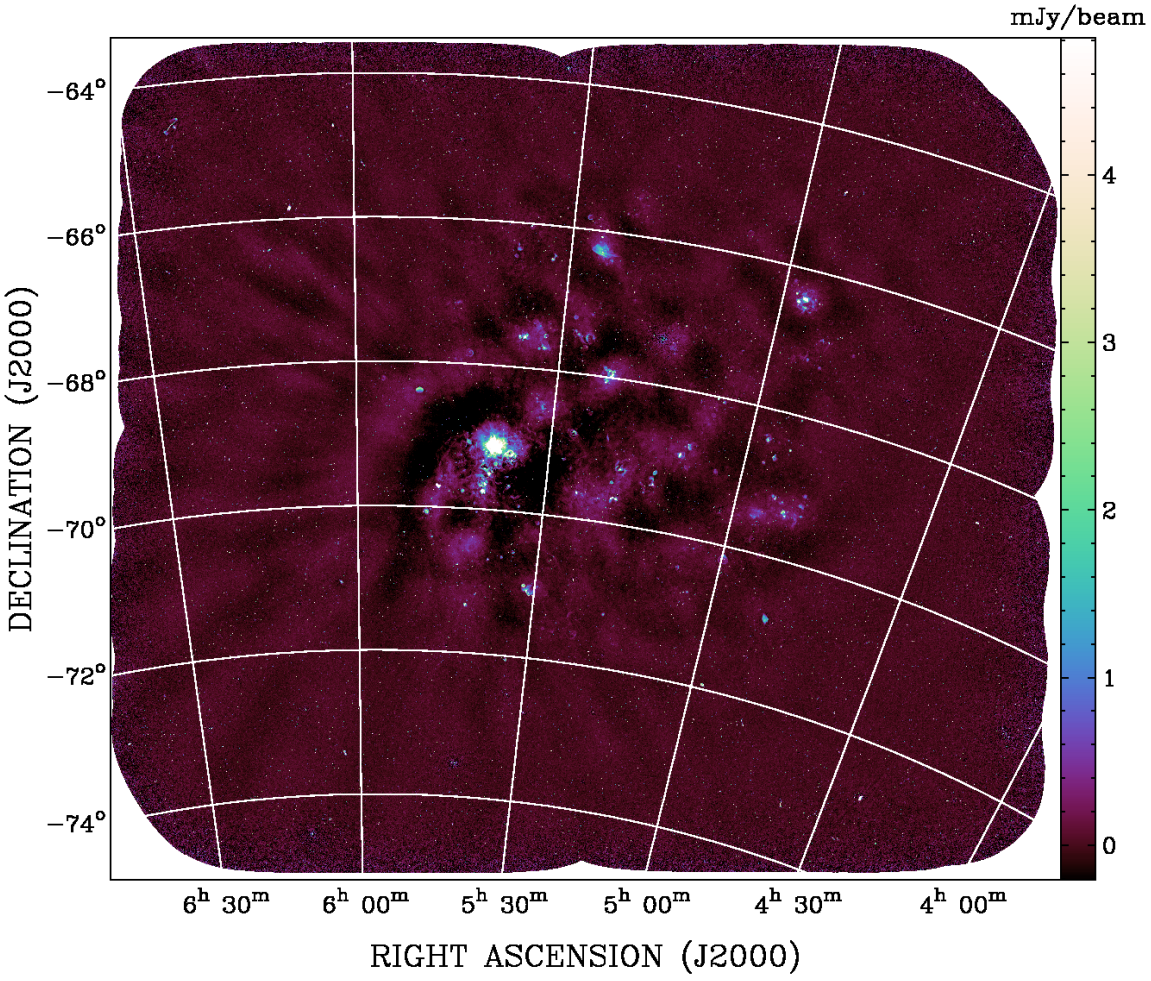

Figure 2.2: ASKAP ESP image of the LMC at 888 MHz. The beam size is $13\farcs9 \times 12\farcs1$

above). These artefacts generally appear as radial stripes and rapidly fade away from the source.

Data processing was performed using ASKAPsoft (Whiting, 2020) by the ASKAP operations team and the resulting images are available on the CSIRO ASKAP Science Data Archive [2]. Bandpass calibration was done using observations of PKS B1934−638, which establishes the flux scale and frequency-dependent complex gains for each of ASKAP's 36 beams. The bandpass-calibrated measurement set for each beam is imaged independently using a phase-only self-calibration approach. The multi-scale, multi-frequency Clean algorithm was used to deconvolve the array response, with two Taylor terms representing spectral behaviour. Images for all beams were combined into a single field using a linear mosaic, correcting for the primary beam response.

The absolute flux density calibration uncertainty is tied to PKS B1934−638 (ATNF Technical Document Ser. 39.3/040[3]), which differs by no more than 1 – 2 per cent compared to the Baars scale (Baars et al., 1977). Additional uncertainties are introduced by PSF variation in a mosaic, generally $\sim$ 1 – 2 per cent for long-track observations such as this one. In later observations this is alleviated by convolving each beam image to the smallest possible common PSF (McConnell et al., 2020). Uncertainties are also introduced by the assumption that all beams have a consistent Gaussian primary beam-shape. In reality holography has shown that they are somewhat non-Gaussian (particularly edge and corner beams of the phased-array instantaneous field) differing from a Gaussian shape by up to 8 per cent (McConnell et al., 2020). Therefore, the overall absolute flux density calibration uncertainty is 8 per cent.

2.1.3 Magellanic specific surveys

There have been many observations made of the Magellanic Clouds, such as the optical Survey of the Magellanic Stellar History (SMASH; Nidever et al., 2017), Spitzer SAGE

[2]https://research.csiro.au/casda/
[3]https://www.atnf.csiro.au/observers/memos/d96783 1.pdf

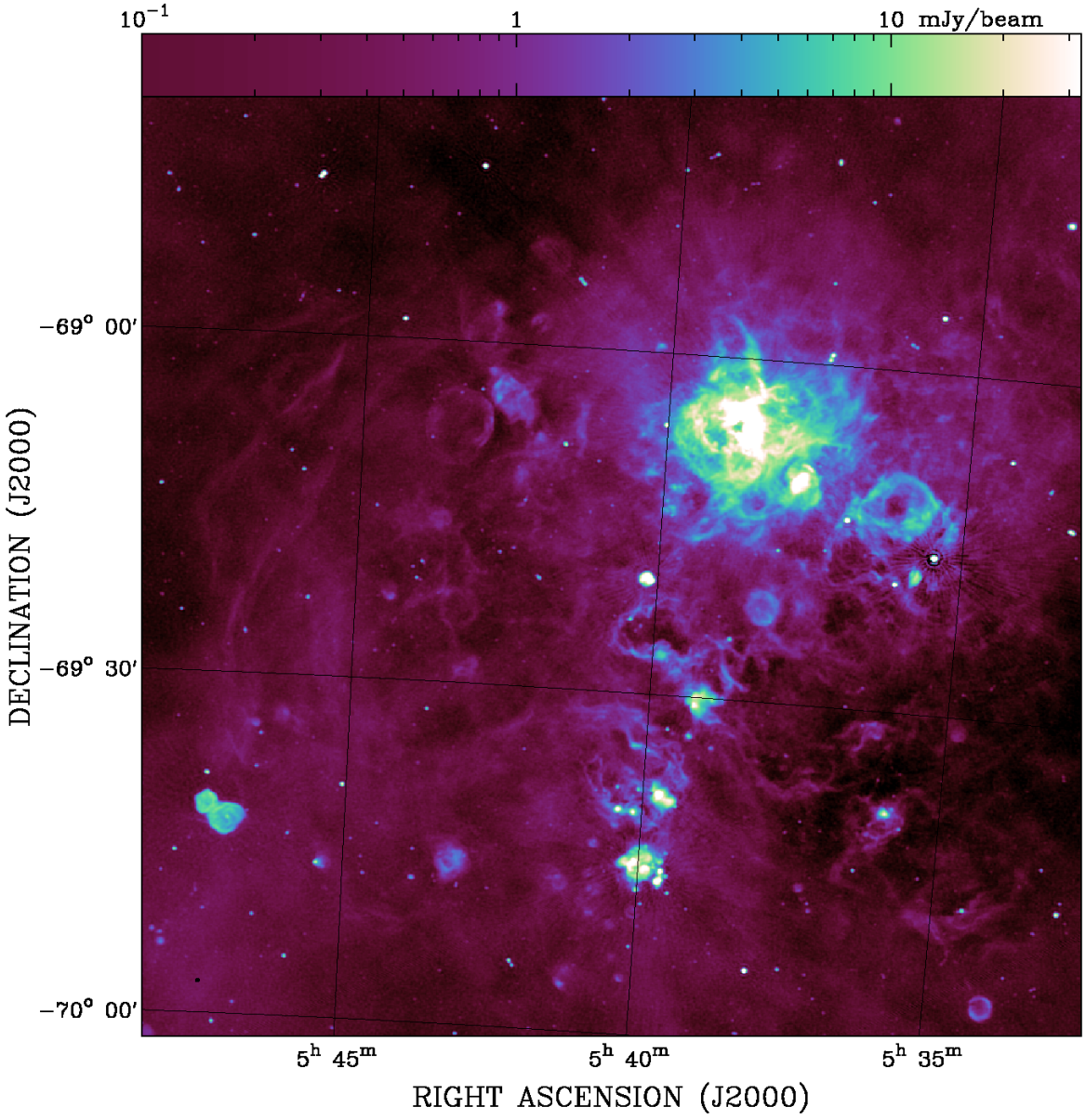

Figure 2.3: Region around 30 Doradus from the ASKAP ESP image of the LMC.

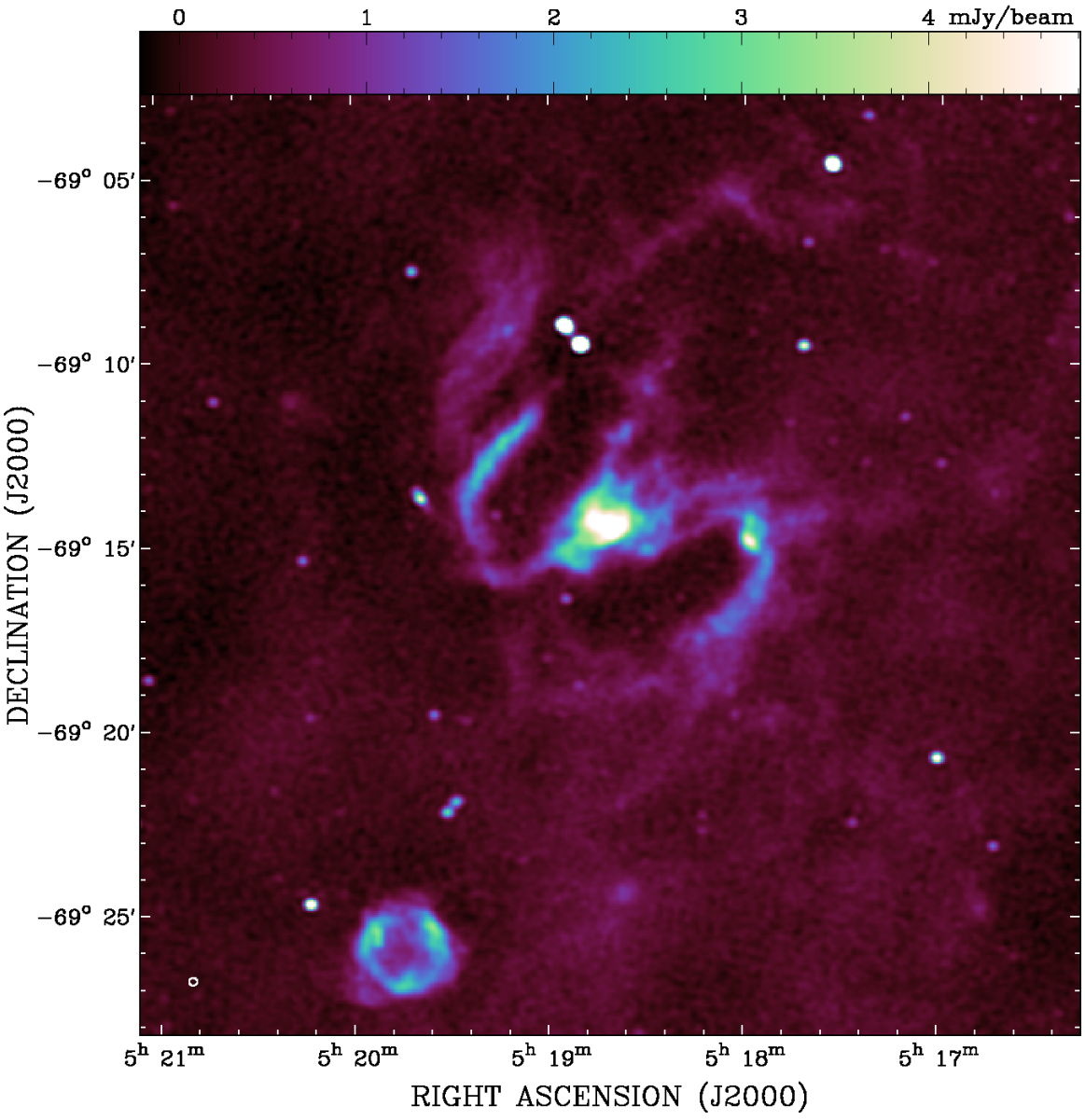

Figure 2.4: LHA 120-N 119 from the ASKAP ESP image of the LMC. N 119 is an H II region close to the kinematic centre of the LMC, and has a pronounced spiral shape that is reminiscent of a barred spiral galaxy.

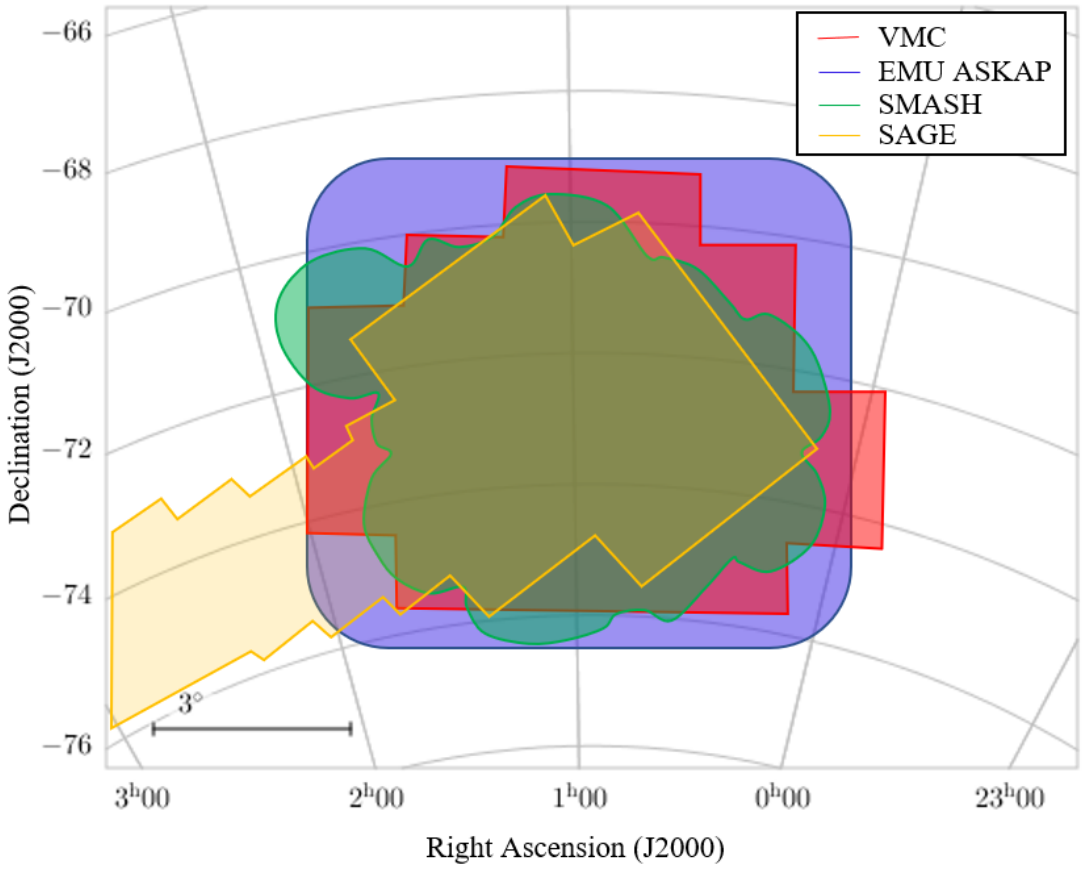

Figure 2.5: Area of the sky covered by the VMC, EMU ASKAP, SMASH and SAGE surveys of the SMC.

(Lacy et al., 2004), Herschel HERITAGE (Meixner et al., 2010) and XMM-Newton (Sturm et al., 2013) X-ray imaging surveys. Figures 2.5 and 2.6 show the comparisons between the area covered by the VMC, EMU ASKAP, SMASH and SAGE surveys of the SMC and LMC, respectively.

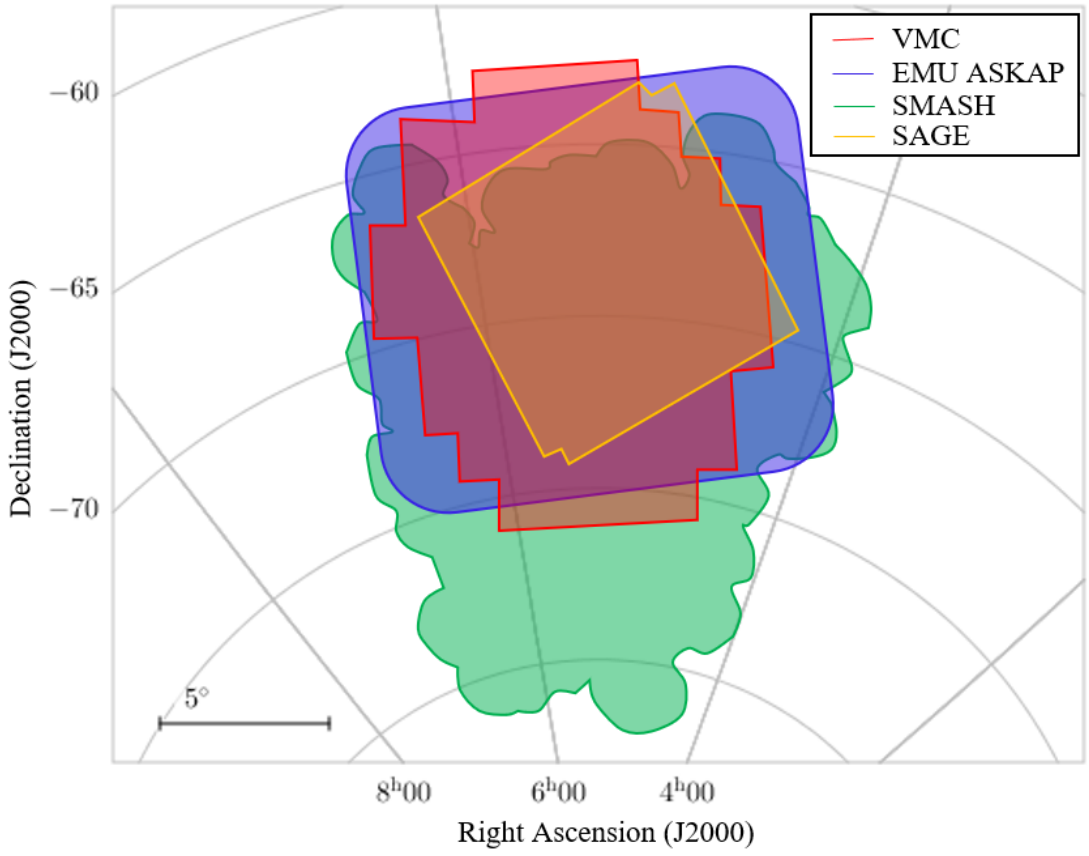

Figure 2.6: Area of the sky covered by the VMC, EMU ASKAP, SMASH and SAGE surveys of the LMC.

2.1.3.1 Optical SMASH

The Dark Energy Camera (DECam; Schumacher et al., 2010) on NOAO's 4-m Blanco telescope was used as part of the Survey of the Magellanic Stellar History (SMASH; Nidever et al., 2017) to map 480 square degrees of sky to depths of $ugriz \sim 23.9, 24.8, 24.5, 24.2, 23.5$ mag (Vega) at seeing of $1\farcs22$, $1\farcs13$, $1\farcs01$, $0\farcs95$, $0\farcs90$. Its main goal was to identify broadly distributed, low surface brightness stellar populations associated with the stellar halos and tidal debris of the Magellanic Clouds. The catalogue contains $\sim$ 360 million objects in 197 fields.

2.1.3.2 Infrared SAGE and HERITAGE

Spitzer Space Telescope (SST; Werner et al., 2004b) is an infrared space telescope, with a primary mirror diameter of 85 cm, launched in 2003 and ending its operations in 2020. Spitzer carried three instruments on board, the Infrared Array Camera (IRAC; Hora et al., 2008) which operated simultaneously at 4 broad bands centred at 3.6, 4.5, 5.8 and 8.0 μm, the Infrared Spectrograph (IRS; Houck et al., 2004) which operated at wavelengths 5.3 – 40 μm, and the Multiband Imaging Photometer for Spitzer (MIPS; Rieke et al., 2004) which operated from the mid to far-infrared at broad bands centred at 24, 70 and 160 μm.

The Magellanic Clouds were observed by Spitzer as part of the Spitzer Agents of Galaxy Evolution (SAGE) survey of the LMC (Meixner et al., 2006) and SMC (Gordon et al., 2011) which map 49 deg^2 and 30 deg^2 respectively. It produced a list of about 8.4 million sources taken with IRAC filters 3.6, 4.5, 5.8, 8.0 μm on the SST with an angular resolution of $2''$. The faint limits for SAGE are 18.3, 17.7, 15.7 and 14.5 mags respectively.

The Herschel Space Observatory (Pilbratt et al., 2010) was a 3.5-m infrared telescope that was active from 2009 to 2013 and was sensitive to the far infrared and submillimetre wavebands (55 — 672 μm). HERschel Inventory of The Agents of Galaxy Evolution (HERITAGE; Meixner et al., 2010) used the Herschel's Photodetector Ar-

ray Camera and Spectrometer (PACS, 100 and 160 μm; Poglitsch et al., 2010) and the Spectral and Photometric Imaging REceiver (SPIRE, 250, 350, 500 μm; Griffin et al., 2010) bands to image the LMC, SMC and Magellanic Bridge. This survey is complementary to the SAGE survey.

2.1.3.3 X-ray XMM-Newton

The XMM-Newton space observatory (Jansen et al., 2001) consists of three X-ray telescopes designed to make long uninterrupted exposures over a large collecting area ($30'$) over a wide energy band, to provide highly sensitive observations. This allows for observations in the 0.2 – 12.0 keV band with an angular resolution of $5'' - 6''$.

An SMC-survey point-source catalogue was created from archival XMM-Newton data with additional newer data from XMM-Newton (Sturm et al., 2013), which covers 5.6 deg^2, including the bar and eastern wing of the SMC. The catalogue contains 3053 unique X-ray sources with a median position uncertainty of $1\farcs3$ down to a flux limit of $\sim 10^{-14}$ erg cm^{-2} s^{-1}. The majority of the sources are expected to be AGN. One limitation of this survey is that it only covers the central part of the SMC, and therefore does not cover the same breadth as the VMC survey.

2.1.4 All-sky surveys

All-sky surveys used in this work include optical Gaia DR3 survey (Gaia Collaboration et al., 2021) photometry and astrometry, as well as mid-IR WISE (Wright et al., 2010; Cutri et al., 2013; Schlafly et al., 2019).

2.1.4.1 Optical Gaia all-sky survey

The Gaia mission (Gaia Collaboration et al., 2016) telescope was launched on 19 December 2013, with the aim of measuring the 3D spatial and velocity distribution of stars, as well as determining their astrophysical properties. The Gaia on-board system

is designed to detect point-like sources, but can detect extragalactic sources if their central region is sufficiently bright and compact. The latest data release, DR3 (Gaia Collaboration et al., 2022), is based on 34 months of Gaia operations. The catalogue provides celestial positions, proper-motions, parallaxes, and broad band photometry in the wide G (centred on 650 nm), blue-enhanced G_{BP} (centred on 360 nm), and red-enhanced G_{RP} (centred on 750 nm) pass-bands. This data release also includes class probabilities (QSO, galaxy or stellar source) for 1.5 billion sources.

2.1.4.2 Infrared WISE

Wide-field Infrared Survey Explorer (WISE Wright et al., 2010), is a telescope launched in 2009 to repeatedly map the entire sky in infrared. WISE mapped the whole sky in four bands W1, W2, W3, W4 centred at 3.4 μm, 4.6 μm, 12 μm, and 22 μm, respectively, using a 40-cm telescope feeding arrays with a total of 4 million pixels. The sensitivities of W1, W2, W3 and W4 correspond to Vega magnitudes of 16.5, 15.5, 11.2, and 7.9, respectively, in the all-sky WISE survey.

The AllWISE (Cutri et al., 2013) programme extended the work of the WISE survey mission by combining data from the cryogenic and post-cryogenic survey phases to form the most comprehensive view of the mid-infrared sky currently available. The AllWISE Catalogue is more sensitive than the all-sky WISE survey in W1 and W2 because images from two WISE survey phases were combined, doubling the depth-of-coverage in those bands. W3 and W4 measurements remain unchanged from the All-Sky Release because no additional data were included in those bands. The catalogue contains accurate positions, apparent motion measurements, four-band fluxes and flux variability statistics for over 747 million objects detected on the deep AllWISE images.

Further adding to the WISE mission, is the unWISE (Schlafly et al., 2019) catalogue, which used the deep unWISE coadded images built from five years of publicly available WISE imaging, as well as improved modelling of crowded regions. This resulted in a catalogue of $\sim$ 2 billion unique objects detected in the W1 and/or W2 channels, reaching depths $\sim$ 0.7 mag fainter than those achieved by AllWISE.

2.2 Spectra

Spectroscopy is the technique of splitting light into its constituent wavelengths, in the same way as a prism disperses light into a rainbow of colours. The absorption and emission of electromagnetic radiation occurs at specific wavelengths, which is linked to the quantised energy levels of electrons in specific atoms and molecules. Spectroscopy, with good enough resolution, allows for these 'lines' of absorption or emission to be observed. The lines not only reveal the presence of specific atoms/molecules in the astronomical object, but also the motion of the object with respect to the Earth, and therefore its distance from Earth in terms of redshift, as well as the motion of different parts of the astronomical object itself, such as an outflow from an AGN.

Spectroscopy is obtained using a spectrometer. The basic components of a spectrograph are the entrance slit, a collimator, a grating or prism, a camera (a lens or mirror), and a detector. A basic diagram of a spectrometer can be seen in Figure 2.7. The width of the slit is chosen to balance between flux (wider slit) and spectral resolution (narrower slit), where the maximum slit width, when observing a point source, should be no more than the "seeing" width. Low resolution spectroscopy is often good enough to detect and identify the strong emission lines from AGN, so a wider slit closer to the "seeing" width can be used.

A prism can split light into its various wavelengths as the index of refraction of the prism has a wavelength dependence. Another method of separating light into its multiple components is via the use of slits, such as in Young's Double Slit experiment, which use diffraction to separate the wavelengths. A diffraction grating is a collection of multiple slits that act to disperse light and are often used in astronomical spectrographs. The slits of the grating are equally spaced parallel grooves on a surface, where the distance between adjacent grooves and the groove angle determine the grating's dispersion and efficiency. The blaze angle, the angle at which maximum throughput is achieved, influences what wavelength range the grating is best optimised for. Changing the angle of the grating would shift the wavelength range being observed, but would also decrease throughput. See Figure 2.8 for a basic diagram of a grating. Gratings

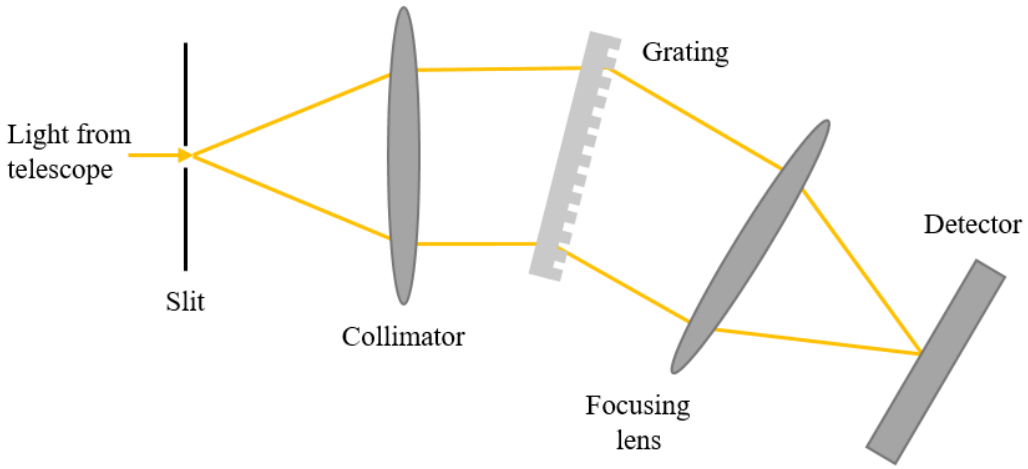

Figure 2.7: Basic optical diagram of a spectrometer.

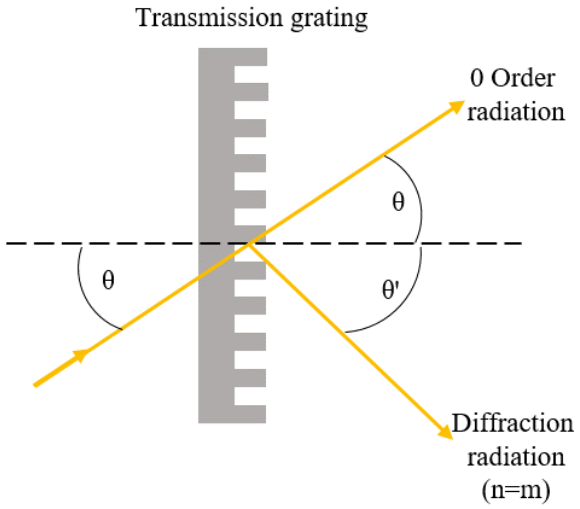

Figure 2.8: Optical diagram of a grating that can be used in spectroscopy.

are generally an improvement over prisms as they are more efficient, do not suffer from the absorption effects and provide a linear dispersion of wavelengths.

After passing through the grating, the camera (lens or mirror) collects the

spectrally-dispersed beams, which are still collimated, and make them converge so that the spectrum is imaged onto the detector, forming a spectrum where the direction along the slit is the spatial axis, and the direction along the spectrum is the dispersion axis. The detector is usually a charged couple device (CCD), which is a two–dimensional array of independent semi–conductors, known as pixels.

There are a plethora of lines that are associated with AGN. Some of these lines can be found in emission or absorption depending on the orientation of the AGN with respect to our line of sight. Typically when we are seeing an AGN face-on, we see lines in emission and when we see an AGN edge-on, we see some lines disappear whilst others remain, due to the dusty torus absorbing the emission from the central engine of the AGN. Notable lines in the optical range (3000 – 10,000 Å) that can be used to identify and characterise AGN by are Hα ($z \sim 0$ – 0.52), Hβ ($z \sim 0$ – 1.05), [Mg II] ($z \sim 0.07$ – 2.57).

Besides the astronomically relevant lines the astronomical objects, there are also absorption and emission lines either from space or from the Earth's atmosphere to contend with.

2.2.1 SAAO 1.9m

I observed 174 new optical spectra (see A.1 for full list) at the South African Astronomical Observatory (SAAO) 1.9-m telescope with SpUpNIC (Spectrograph Upgrade: Newly Improved Cassegrain; Crause et al., 2019) during observing runs in 2019 and 2021. Grating 7 (grating angle of 16°) and the order blocking 'BG38' filter were used, delivering a resolving power $R = \frac{\lambda}{\Delta\lambda} \sim 500$ over a wavelength range of 3800 Å – 9000 Å. Dome-flats and bias images were taken at the beginning of each night. The CuAr lamp was used for wavelength calibration. Three 600 s (300 s for sources brighter than $\sim$ 16 mag) exposures were obtained for each source. The standard stars (EG 21, Feige 110 or LTT 1020; Hamuy et al., 1994) were observed on the same night under the same conditions for 30 s.

I processed the data using the standard IRAF[4] tools (Tody, 1986, 1993). Basic CCD reduction was performed, this included trimming the 2D spectrum, removing bias, flat-fielding and cosmic ray removal. The arc spectra emission lines of the CuAr lamp were identified and used to wavelength calibrate the science images. The standard stars were used to calibrate the flux of the science images. Then finally the 1D spectrum of the sources on the slit were extracted. The steps and IRAF functions used can be seen in Appendix B.

2.2.2 SALT

The Southern African Large Telescope (SALT) (Buckley et al., 2006) is located in Sutherland, South Africa. SALT's primary mirror is an 11-m diameter segmented mirror, made up of 91 1.2 m hexagonal segments. Due to the primary mirror being fixed at an angle of 37° with respect to the zenith, thus limiting the declination across which observations can be made, and science observations being performed with a 'tracker' that follows targets across the sky, the telescope's effective diameter is closer to 7 – 9 m. SALT was used to observe AGN candidates that had the potential to be similar to SAGE0536AGN. I used the Robert Stobie Spectrograph (RSS; Burgh et al., 2003; Kobulnicky et al., 2003), a combination of three CCD detectors with total 3172 × 2052 pixels and spatial resolution of $0\farcs1267$ per pixel. I used the long-slit with width $1\farcs5$ or $1\farcs25$, grating PG0300 or PG0900 and an Argon arc lamp. In some cases the default blaze angle was changed to incline the grating, which sacrificed some efficiency, to tune the wavelength range to, for example, avoid emission lines landing in a CCD gap. Initial processing (basic CCD data reductions) was done automatically by the SALT pipeline (Crawford et al., 2010). I processed (cosmic ray removal, wavelength calibration and source extraction) these data also using the standard IRAF tools (Tody, 1986, 1993).

[4]IRAF is distributed by the National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

The focus of the SALT telescope observations was to find and characterise more AGN-dust dominated galaxies and characterise extragalactic radio sources with interesting extended emission. The observations were done under programmes 2019-1-SCI-032 (PI: Clara Marie Pennock), 2019-2-SCI-041 (PI: Jacco van Loon), 2019-2-SCI-045 (PI: Jacco van Loon), 2020-1-SCI-028 (PI: Clara Marie Pennock), 2020-2-SCI-025 (PI: Clara Marie Pennock), 2021-1-SCI-029 (PI: Jacco van Loon), 2021-1-SCI-032 (PI: Jacco van Loon), 2021-2-SCI-018 (PI: Jacco van Loon), 2021-2-SCI-017 (PI: Joy Anih), 2022-1-SCI-022 (PI: Jacco van Loon) and 2022-1-SCI-023 (PI: Joy Anih). The list of sources observed by SALT can be seen in Tables A.2 and A.3.

Prior to this study SAGE0536AGN had been observed with SALT by van Loon & Sansom (2015). Further observations of SAGE0536AGN were obtained with SALT RSS in 2017 (programme 2017-1-SCI-001) but were unfortunately affected by focus issues. These spectra covered ∼534 to 623 nm, with PG2300 grating, including Hβ, Mg b and Fe λ5335 spectral features. Two of the five exposures (observed on 20/10/2017) were of sufficiently good quality and high spectral resolution to attempt kinematic measurements. Using Python PPXF[5] and INDO-US star spectral templates (Valdes et al., 2004) the measured velocity dispersion was $\sim 202 \pm 15$ km s^{-1} (Sansom et al., *in prep.*), with overall errors from PPXF uncertainty and spectral resolution uncertainty added in quadrature. This measurement was within the central $\sim 1''$ along the major axis of SAGE0536AGN and is larger than previously found, $s \sim 123 \pm 15$ km s^{-1}, in van Loon & Sansom (2015). This may be because of the focus problems with the 2017 data but could also result from measurement in a better spectral range, less affected by a particular (NaD) spectral feature and along the major axis. IFU data would be needed to more accurately determine the kinematics across SAGE0536AGN.

[5] https://www-astro.physics.ox.ac.uk/∼cappellari/software/#ppxf

2.2.3 Literature

Some of the sources studied had already been spectroscopically observed prior to this study. The spectra of these sources were studied in Chapter 5, and the sample list for which can be found in Table 5.1.

Three sources had been observed as part of the Magellanic Quasars Survey (MQS, Kozłowski et al., 2013) and one source as part of a search for variability-selected quasars in the Magellanic Field (Geha et al., 2003).

Lastly, another source had been observed with European Southern Observatory's 3.6-m telescope with EFOSC2 as part of a survey to find polarized quasars (see Kishimoto et al., 2008, Kishimoto et al. *in prep.*). For all frames, the CCD was read out with 2×2 binning, giving a spatial sampling of $0\farcs316$ per pixel. The grism Gr#1 was used at a dispersion of 13 Å per pixel (after the binning). The target was observed with a $1\farcs5$ slit width, giving a spectral resolution of ~ 60 Å. The data were reduced in a standard manner. The averaged bias frame was subtracted, and each frame was flat-fielded. The wavelengths were calibrated using arc frames, and the spectra were extracted within a $2\farcs8$ window and flux-calibrated.

2.3 Training sets for machine learning

To train a supervised machine learning classifier a training set of known sources is required. The sources I observed with SALT and SAAO were added to these training sets. The total number of sources for each class can be seen in Table 2.2.

Using spectroscopically observed sources, however, introduces bias into the training sample, since the sources observed tend to be chosen based on colour cuts that similar previously observed objects conform to. Furthermore, there is also a bias in magnitude, as the faintest sources would be too faint for spectroscopy. This therefore leaves the rarer/unusual and fainter versions of each class to not be observed.

Table 2.2: The number of sources for each class and the region of the Clouds they were spectroscopically observed in, as well as the references of the literature they originated from. (1) From own observations using SALT or SAAO's 1.9m telescope; (2) Spitzerspec surveys (Ruffle et al., 2015; Jones et al., 2017); (3) From a Simbad (Wenger et al., 2000) search of ' PM' stars in the VMC footprint that had listed spectral type and reference given.

Class	SMC	LMC	References
AGN	306	639	(1); (2); Flesch (2019b); Kozłowski et al. (2012); Kozłowski et al. (2013); Geha et al. (2003); Esquej et al. (2013); Ivanov et al. (2016); Ivanov et al. *in prep.*
Galaxies	124	430	(1); (2); Jones et al. (2009)
OB stars	417	1073	(2); Walborn et al. (2014); Evans et al. (2015a,b); Lamb et al. (2016); Grin et al. (2017); Roman-Duval et al. (2019); Dorigo Jones et al. (2020)
Em-line stars	5	563	(1); (2); Reid & Parker (2012)
RGB stars	519	489	(2); Cole et al. (2005); Neugent et al. (2020) Parisi et al. (2009, 2010, 2022); De Bortoli et al. (2022)
H II/YSOs	86	459	(2); Seale et al. (2009); Oliveira et al. (2011, 2013); Oliveira et al. (2019); van Gelder et al. (2020)
PNe	53	50	(2); Shaw et al. (2001)
AGB stars	165	221	(2); van Loon et al. (1998); Groenewegen & Blommaert (1998); van Loon et al. (1999a,b, 2005, 2006, 2008); Kamath et al. (2014)
RSG stars	44	70	(2); Neugent et al. (2020)
pAGB/RGB stars	46	33	(2); van Loon et al. (2008); Kamath et al. (2014)
PM stars	78	303	(3)

2.3.1 SAGE-spec

The Infrared Spectrograph onboard the Spitzer Space Telescope was used to observe the LMC and SMC in low and high resolution modes for the wavelength range of 5 – 38 μm. The resolving power varies between 60 – 130 for low resolution mode, and high resolution mode has a resolving power of $\sim$ 600.

All the spectra obtained by Spitzer within the SAGE footprint were looked at as part of the SAGE-Spec project. In the SMC (Ruffle et al., 2015), this survey found 58 asymptotic giant branch (AGB) stars, 51 young stellar objects (YSOs), 4 post-AGB objects, 22 red supergiants (RSGs), 27 stars (of which 23 are dusty OB stars), 24 planetary nebulæ (PNe), 10 Wolf-Rayet (WR) stars, 3 H II regions, 3 R Coronæ Borealis (R CrB) stars, 1 blue supergiant and 6 other objects.

In the LMC (Jones et al., 2017), this survey observed $\sim$ 800 sources in the SAGE LMC footprint, the majority of which are YSO and H II regions and (post-)asymptotic giant branch stars, PNe and massive stars. Also observed were two supernova remnants, a nova and several background galaxies.

2.3.2 Extragalactic

There are 657 spectroscopically observed AGN, from the milliquas catalogue of Flesch (2019a,b), in the field of the VMC footprint of the LMC. The highest contributions are from Kozłowski et al. (2012, 2013), Geha et al. (2003), Esquej et al. (2013) and Ivanov et al. (2016), contributing 547, 24, 23 and 10 objects, respectively.

There are 240 spectroscopically observed AGN, from the milliquas catalogue of Flesch (2019a,b), in the field of the VMC footprint of the SMC. The highest contributions are from Kozłowski et al. (2011, 2013) and Ivanov et al. (2016), contributing 194 and 10 objects, respectively.

Galaxies were taken from the 6dFGS survey (Jones et al., 2009). The observations for this survey were carried out using the Six Degree Field (6dF) fibre-fed multi-object spectrograph at the UK Schmidt Telescope (UKST) over 2001 May to 2006 January

(Jones et al., 2005). Target fields covered the $\sim 17\,000$ deg^2 of southern sky more than 10° from the Galactic Plane. This survey data, however, comes with the caveat that it is limited to the brightest (it is complete to total extrapolated 2MASS magnitude limits of (J, H, K) = (13.75, 12.95, 12.65) mag) and closest (median redshift of whole survey is $z \sim 0.053$) of galaxies.

2.3.3 Galactic and Magellanic

For the training sample, sources that are often mistaken with AGN were needed, whilst also including other sources that are more distinct from AGN that are prevalent through the Magellanic Clouds. There have been many spectroscopic surveys of the LMC and SMC, generally looking at specific types of stellar objects that can be found within the Magellanic Clouds. This led to the accumulation of 1490 OB stars, 1008 RGB stars, 568 emission-line stars, 545 YSO or compact H II regions, 386 AGB stars, 114 RSGs, 103 PNe, 79 pAGB/pRGB and 382 high proper-motion/foreground stars (references: see Table 2.2). A large part of these sources were from the SAGEspec surveys (Ruffle et al., 2015; Jones et al., 2017), which observed 209 and 862 sources in the direction of the SMC and LMC, respectively.

The more mundane stars (e.g. main-sequence M stars), however, tend not to get spectroscopically observed on purpose, this therefore leads to a lack of these for training. They are, however, distinct from AGN, with a lack of emission in the IR, so should not be mistaken for extragalactic sources, and be more associated with the stellar sources based on proper-motions and colours.

3 Methodology

This chapter is split into 4 sections. The first section is about the testing and use of a source detection algorithm, Aegean on the ASKAP LMC image and comparing it to a previous observation of the LMC with the SUMSS telescope, and the EMU ASKAP observation of the SMC. The second section describes the usage of Galfit to model the light profiles of AGN with resolved host galaxies. The third section describes the usage of CIGALE spectral energy distribution (SED) fitter to model the photometric SEDs of a sample of sources. The final section describes the two machine learning algorithms that were used to classify sources.

3.1 ASKAP ESP LMC Source Detection

I created the 888 MHz LMC ASKAP source catalog using the Aegean source finder algorithm (Hancock et al., 2012; Hancock et al., 2018). Variable noise is present across the image, from the combination of the multiple beams and artefacts from bright sources. To ensure accurate source thresholds the variable noise must be parameterised before source finding. The Bane (Background and Noise Estimation) routine in the AegeanTools software was used, with default parameters, to create root mean squared noise (RMS) and background level maps[1] (see Figure 3.1). The maps were then used, alongside the LMC image, with Aegean, using default parameters to create initial source lists. A cursory visual inspection of the sources was carried out to verify detections from the initial source list.

[1]It does this via a sliding boxcar approach where it calculates the background and noise properties of all pixels within a box centred on a given grid point, the resulting maps having the same pixel scale as the input images.

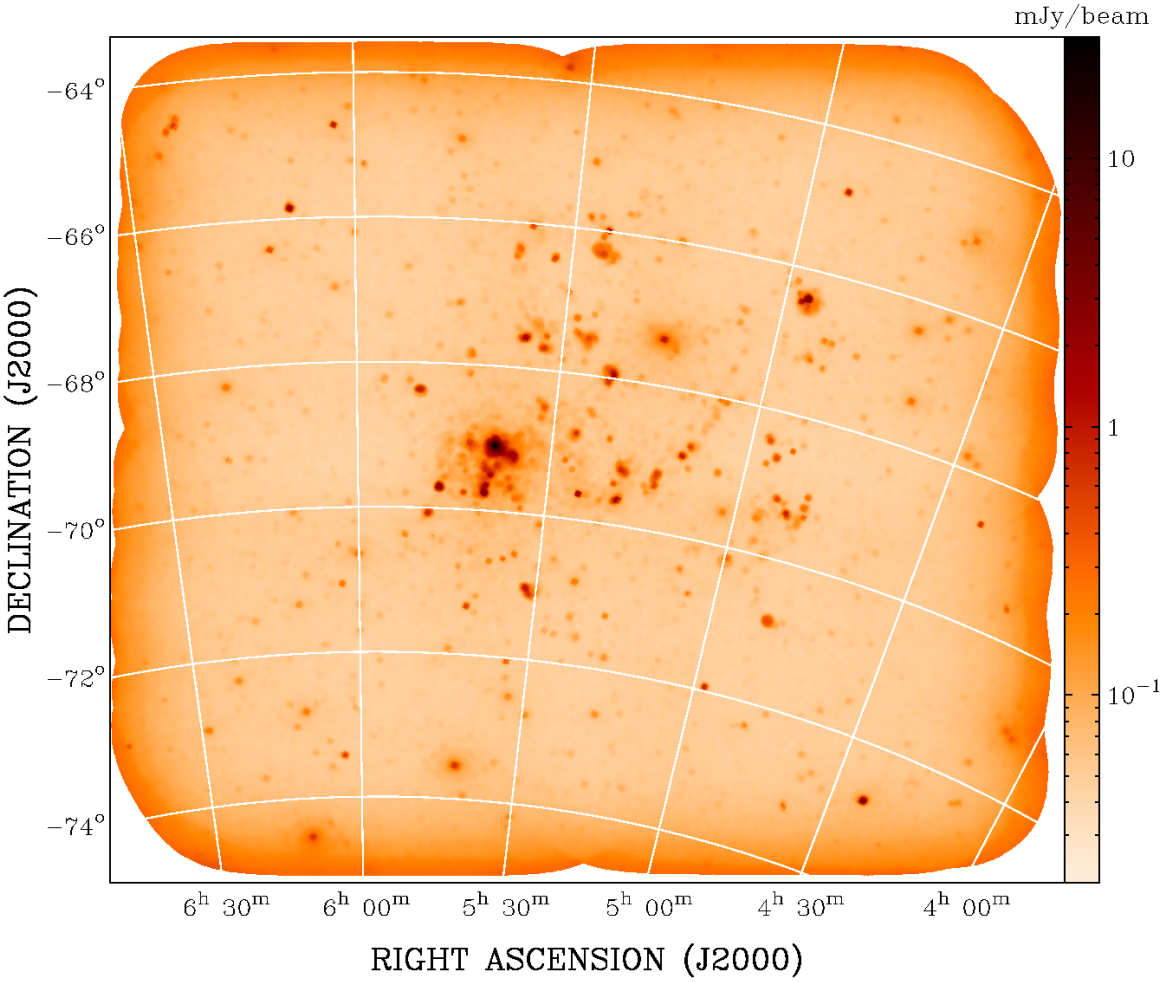

Figure 3.1: RMS map of the 888 MHz ASKAP observations, produced by BANE with the default parameters. The image is of the same pixel scale as in Figure 2.2. Higher RMS levels are found at the edge of the field (due to fewer beams present and because of the primary beam shape of which the sensitivity falls off) and around the brighter sources.

3.1.1 Testing of Aegean

To ascertain the best AEGEAN parameters, I used the algorithm on subsamples of the image with different values of SEEDCLIP (the number of σ above the local RMS for sources to be deemed detected, also known as the detection threshold) to find the best SEEDCLIP value that obtains the least false positive detections without sacrificing too many true positive detections. To quantify the false detection rate it can be assumed that the noise in the image follows a symmetric, Gaussian distribution. This symmetry means that the negative image (inverted) has the same noise properties as the non-inverted image, meaning large noise troughs in the original image are now detected as sources. Hence, the number of detected sources should be approximately equal to the number arising from the false positive noise spikes in the true image. From this I constrained the "real" detections in an image by investigating how the false detections vary with SEEDCLIP and then picked out the optimum value that successfully extracts sources with minimal contamination from noise (Hale et al., 2019). The percentage of real detections can therefore be calculated using

$$\%RealDetections = 100 \times \frac{N_+ - N_-}{N_+} \tag{3.1}$$

where N_+ and N_- are the total number of sources detected with AEGEAN on the original and inverted images, respectively.

The subsamples AEGEAN were tested on can be seen in Figure 3.2. A rectangular region (area is 5000 x 14000 pixels) was chosen such that at one end there was an area of bright, extended radio emission (the centre of the LMC), and at the other end is an area of mostly isolated point sources (edge of ASKAP LMC image). This rectangle was then split into 7 segments. To compare the difference in performance for using AEGEAN on high and low source density regions at different values of SEEDCLIP, AEGEAN was used on the first segment, M1 (LMC centre), then Equation 3.1 was used to calculate the %Real Detections. This is repeated for different values of SEEDCLIP. Then the area being tested on was expanded to the second segment to form a larger segment, M2, and AEGEAN was used again and %Real Detections was calculated for varying values

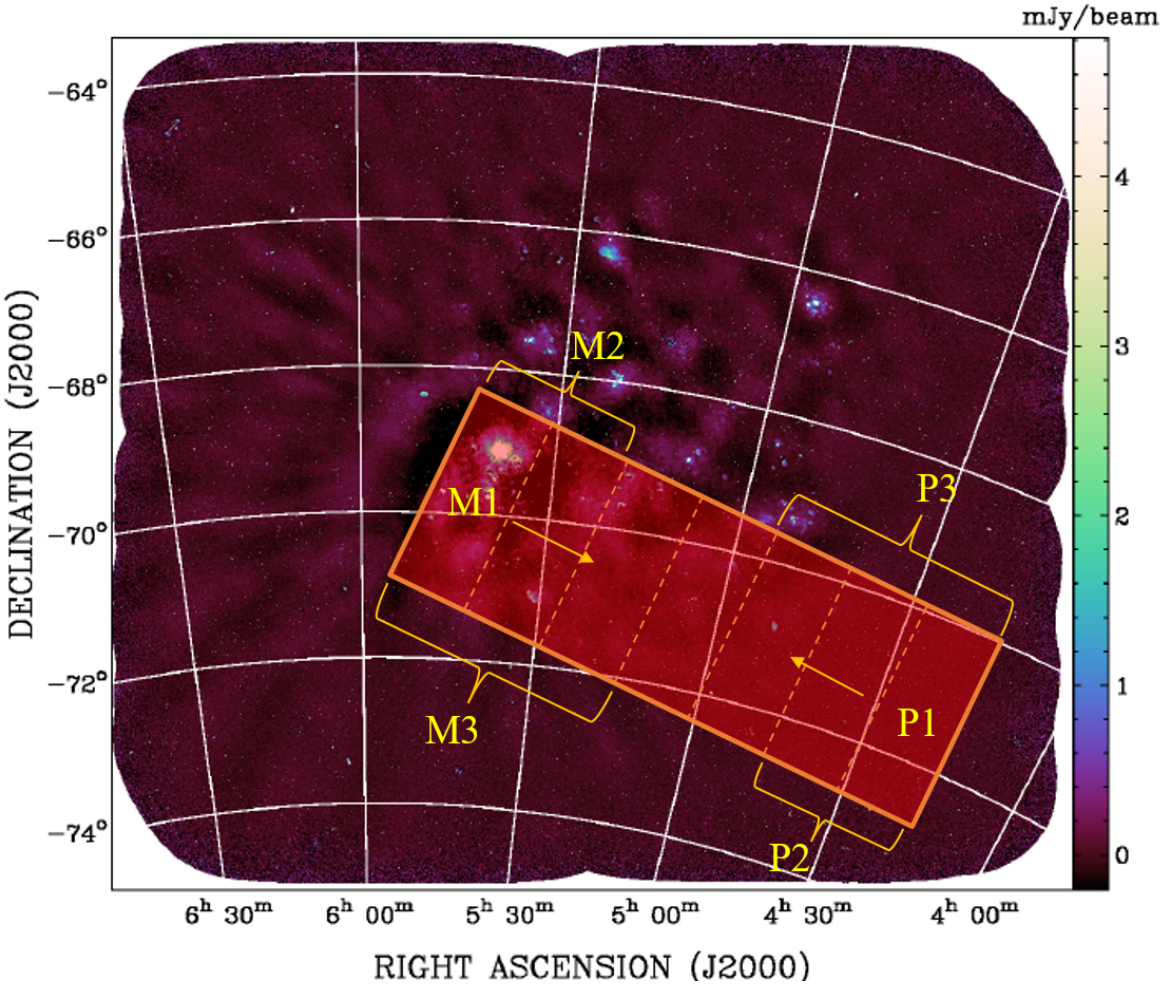

Figure 3.2: Regions AEGEAN was tested on for the ASKAP ESP image of the LMC at 888 MHz. AEGEAN was tested by starting on a region at the centre of the LMC where there is extended emission, M1, and a region away from the LMC where there are mostly point sources, P1. Then the image size is extended (indicated by the arrow) to P2 and M2, and AEGEAN was tested on the larger images. This is repeated for both starting points until the full region is tested on as outlined in the figure.

of SEEDCLIP. This is repeated until the entire region outlined in Figure 3.2 is tested on. Starting from the other end of the region, P1, this process is then repeated.

How the detection threshold (SEEDCLIP) affects the %Real Detections for different areas of the ASKAP LMC image can be seen in Figure 3.3, where the left graph shows

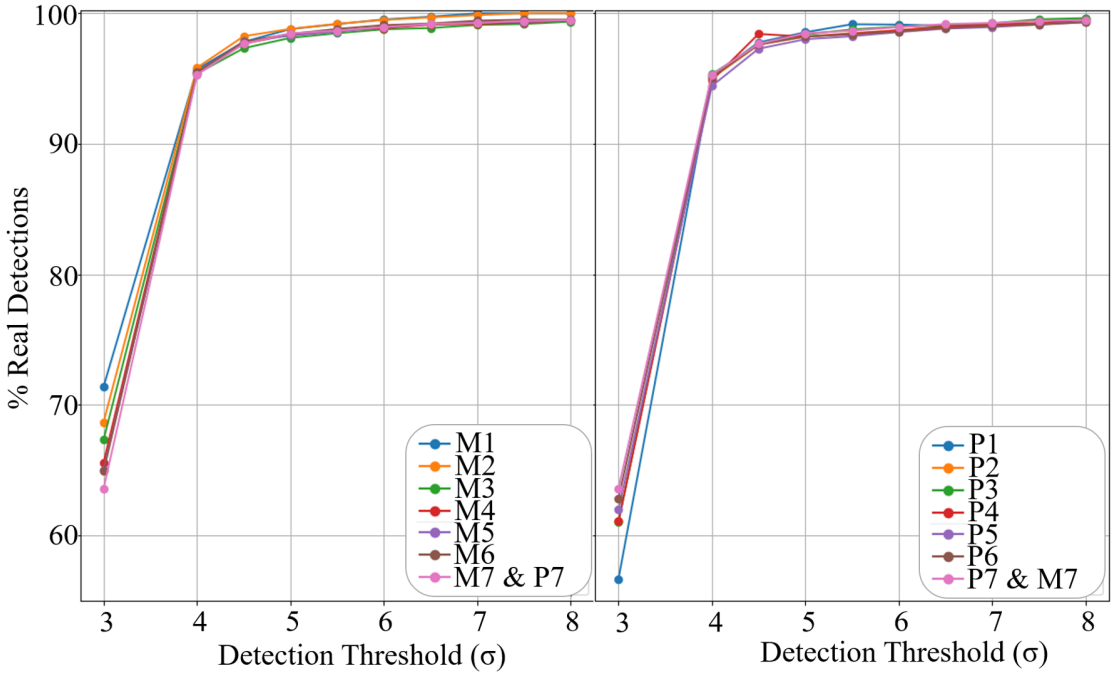

Figure 3.3: How % Real Detections varies with detection threshold for sources in the centre of the LMC (left) and sources far from the centre of the LMC (right), where the pink line in both graphs represent the same test area. The regions names are explained in Figure 3.2. This shows that the % Real Detections plateaus at 5σ for all image sizes irrespective of whether they are in a region of bright highly extended sources or a region of mostly isolated point sources.

the test starting from the centre of the LMC and the right graph shows the test starting from the edge of the ASKAP LMC image. From this it can be seen that the % Real Detections plateaus at $\sim 5\sigma$ for all tests. It should be noted that despite the lower %Real Detections for the P1 start (most obvious at 3σ), that the number of sources detected is higher than for the M1 start, where the large bright extended sources would cause the flux corresponding to the detection threshold to be higher, so only the brighter, more likely real points sources would be detected. Whereas further away from the LMC, fainter sources can be picked up, and at a lower detection threshold (such as 3σ) the variations in background are more likely to be detected as point sources.

Overall, testing proved that for a large variety of regions, containing from compact to large extended sources, the default value of 5σ was appropriate.

3.1.2 Source lists

In total, I found 54,612 sources in the ASKAP 888 MHz image. I separated this source list into a GOLD (30,866), SILVER (22,080) and BRONZE (1666) source list. Some sample records of the GOLD source list can be found in Table 3.1. The PSF was generated during image restoration, which means that the semi-major axis ($13\farcs9$), semi-minor axis ($12\farcs1$) and position angle ($-84\fdg4$) of the PSF are constant for all sources. The sizes of the sources were determined during source finding. All sources have a designated 'Island' number, and if multiple sources have the same 'Island' number then they are considered components of the island and will have different 'Source' numbers. These components of the islands can be spread across the different source lists. The isolated (one component in an island) sources make up $\sim$ 90 per cent of the entire catalogue. Note that I find isolated sources have more reliable flux density errors than sources that are components of islands, which tend to have unreasonably small flux density errors calculated by AEGEAN.

To remove multiple detections of large/bright extended sources, a threshold of the local RMS was used. This is because in the area of these large sources the value of local RMS is higher than elsewhere. A cut of RMS < 0.12 mJy/beam was used and a cursory visual inspection was undertaken of the sources removed (and not removed) to verify the effectiveness of this cut. This cut also had the effect of removing sources close to/surrounded by bright extended sources, as well as the sources along the outermost edges of the image due to the poorer sensitivity in this part of the image.

The GOLD standard source list required an integrated flux density > 0.5 mJy in order to give a spatially uniform distribution, so that the fainter sources, that are not uniform in spatial distribution, are separated into a SILVER source list (see Figure 3.4 (left & centre)). This value was chosen because this is where the flux density distribution of the ASKAP radio sources turns over, see Figure 3.5. The BRONZE source

Table 3.1: Example of the catalogue created from the LMC ASKAP ESP image. In the second part of the "EMU ID" the sources are labelled either as "ES" or "EC", where "E" stands for early science data, "S" stands for an isolated source and "C" stands for a component of a source. The σ indicates the source-finding fitting error. "Peak" indicates the peak flux density per beam and error. "Integrated" indicates the integrated flux density, which is not fitted directly but is calculated from a, b and peak flux density and the synthesised beam size. The a is the fitted semi-major axis, b is the fitted semi-minor axis and PA is the fitted position angle. "Island" is the number of the island from which the source was fit and the "Source" is the number within that island, starting from zero for components of a source (EC) and always zero for isolated sources (ES). "Background" is the flux density per beam measured from the background map created with BANE at the source position. "Local RMS" is the flux density per beam measured from the rms map created with BANE at the source position. "Residual Flux" mean and standard deviation are the mean and standard deviation of the residual flux remaining in the island after fitted Gaussian is subtracted. Flags are 0 for a good fit, 2 indicates an error occurred during the fitting process (e.g. the fit did not converge), 4 indicates a component was forced to have the shape of the local point spread function.

Source List	EMU ID	RA (J2000) $h\ m\ s$	Dec (J2000) $\circ\ \prime\ \prime\prime$	...
GOLD	EMU ES J063120.8−741322	06:31:20.8	−74:13:23	...
GOLD	EMU ES J063111.1−741304	06:31:11.3	−74:13:04	...
GOLD	EMU ES J062722.6−742301	06:27:22.6	−74:23:01	...
...				

— peak —		— integrated —		— a —		— b —		...
F (mJy beam^{-1})	σ (mJy beam^{-1})	F (mJy)	σ (mJy)	($''$)	σ ($''$)	($''$)	σ ($''$)	
6.20	0.11	7.60	0.14	15.49	0.07	13.51	0.06	...
1.65	0.08	4.56	0.25	27.27	0.39	17.28	0.23	...
1.46	0.11	1.67	0.14	16.18	0.32	12.05	0.22	...
...								

— PA —		Island	Source	Background	Local RMS	Residual Flux		flags
$\circ$	σ ($^\circ$)			(mJy beam^{-1})	(mJy beam^{-1})	mean (mJy)	σ (mJy)	
−87.07	0.10	113	0	0.009	0.114	−0.038	0.146	0
−87.08	0.08	115	0	0.007	0.111	0.007	0.189	0
−86.66	0.19	117	0	0.015	0.118	−0.013	0.075	0
...								

list was created, which consists of the sources where the local RMS is > 0.12 mJy/beam, which indicated noisy regions, either at the edges of the field or around bright sources. Each source was individually visually inspected using SAOImageDS9 (Joye & Mandel, 2003), where a source was kept if it was clearly visible above the background noise and was point-like or slightly resolved. This reclaimed some of the sources near the edge of the image where the high noise removed them from the Silver and Gold lists, as well as the sources that were near bright extended emissions, including well-known bright sources such as the supernova SN 1987A. The sources that were not recovered are those located in complex regions or blended with other sources that would require higher angular resolution to disentangle. The flux densities of a few sources in noisy or confused regions may be unreliable. Therefore, it is advised that for these sources the user inspects the image and ascertain whether to trust the given flux density values. Of the 4072 sources in areas where the local RMS is > 0.12 mJy/beam, 1666 were placed in the Bronze source list, the density map of which is presented in Figure 3.4 (right). Note that the local RMS cut intrinsically leads to incompleteness in the Gold sample at the brightest flux densities, as the brightest sources cause a higher local RMS, thus leading to their inclusion in the Bronze source list instead of the Gold source list.

3.1.3 Comparison with previous radio catalogues

The ASKAP ESP source lists were compared with various radio continuum catalogues that cover the LMC. The number counts for our survey were compared with those from the previous ASKAP ESP catalogue of the SMC (Joseph et al., 2019), for which the full ASKAP array was not available. See Table 3.2 for a comparison of previous radio surveys of the LMC.

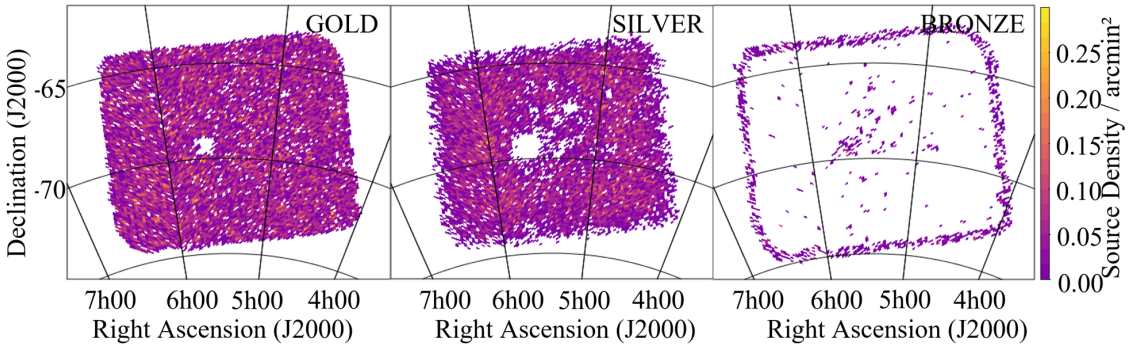

Figure 3.4: Density plots of the Gold (left), Silver (centre) and Bronze source lists (right). This shows that the Gold source list is uniform across the LMC field, except in the area of 30 Doradus, which is seen as a blank space. The Silver source list is not as uniform and showcases gaps where the brighter Magellanic sources are present. The Bronze source list covers the edges of the LMC field and the areas of the brightest sources, where the local RMS is greatest.

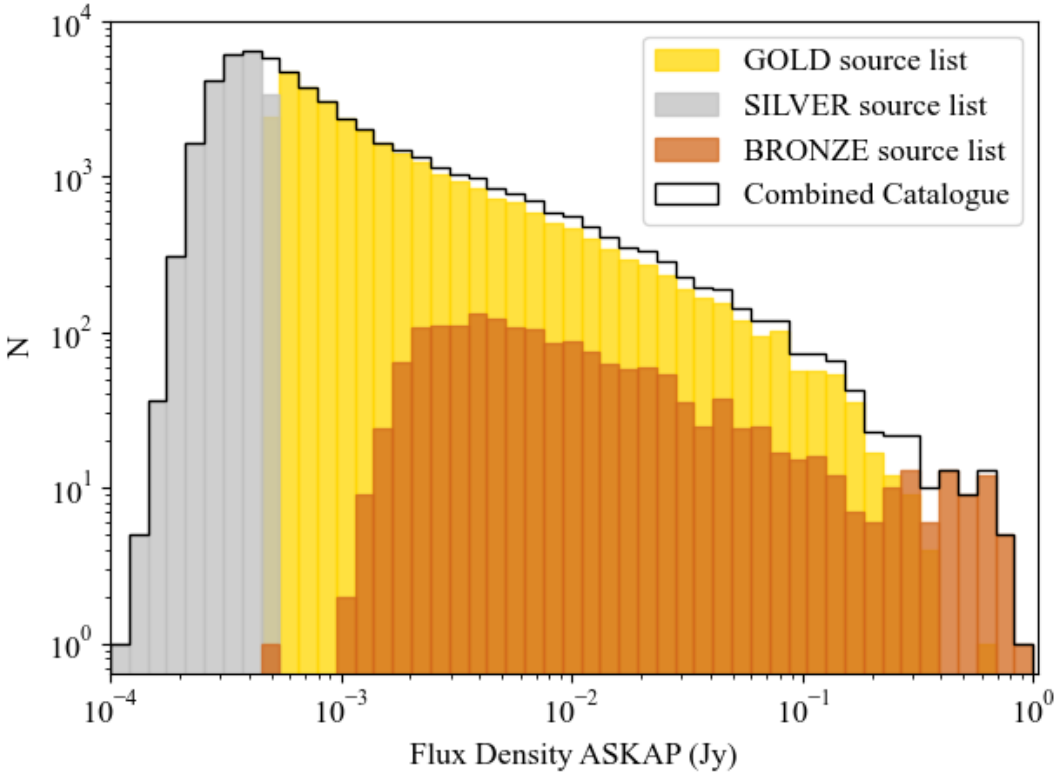

Figure 3.5: Histogram of integrated flux density at 888 MHz illustrating the Gold, Silver, and Bronze populations.

Table 3.2: Table comparing existing radio surveys of the LMC with the survey presented here. References: (1) This work. (2) Filipović et al. (2021), (3) Hughes et al. (2007), (4) Dickel et al. (2005), (5) Mauch et al. (2003b), (6) Griffith & Wright (1993), (7) Mills (1985).

Survey	Telescope	Frequency GHz	RMS mJy beam^{-1}	Beam Size $''$	Bandwidth MHz	Source Count	Reference
EMU ESP	ASKAP	0.888	0.0577	13.9 × 12.1	256	53547	(1)
ASKAP-Beta	ASKAP	0.843	0.71	61 × 53	1	1995	(2)
1.384 GHz ATCA	ATCA	1.384	0.5	40	128	6623	(3)
4.8 / 8.64 GHz ATCA	ATCA	4.8 / 8.64	0.28 / 0.5	33 / 20	128	801 / 419	(4)
SUMSS	MOST	0.843	~ 1	43	3	3829	(5)
PMN Survey	Parkes	4.85	~ 4.2	126	600	410	(6)
843 MHz MOST	MOST	0.843	0.3 – 0.4	43	3	2162	(7)

3.1.3.1 SUMSS 843 MHz

The MOST telescope was used to carry out the Sydney University Molonglo Sky Survey (SUMSS; Mauch et al., 2003b,a) at 843 MHz with a beam size of $45'' \times 45''$ to create a catalogue that covers the whole sky South of DEC $= -30$ deg, including the LMC. The positions in the catalogue are accurate to within 1 – 2 arcsec for the brighter sources (> 20 mJy beam^{-1}) and always better than 10 arcsec. This catalogue therefore makes for a suitable comparison for the new 888 MHz ASKAP catalogue.

The new ASKAP 888 MHz source lists (Gold and Silver, 53,547 sources) were cross-matched with the SUMSS 843 MHz catalogue (3829 sources), with a search radius of 10 arcsec, yielding 3211 matches. The flux densities are compared in Figure 3.6; the dashed blue line indicates the one-to-one relation and the red line indicates the line of best fit. The sources that show higher than expected flux density from ASKAP compared to SUMSS were found to be near bright point sources with nearby artefacts which, when coinciding with a fainter source, increases their total flux density. The sources that show lower than expected flux density from ASKAP compared to SUMSS were found to be extended sources, that were seen as one source in SUMSS but multiple components in ASKAP, hence the lower flux density. The correlation between the calculated integrated flux densities of ASKAP 888 MHz and SUMSS 843 MHz source lists is described by the Pearson Correlation Coefficient, $r = 0.967 \pm 0.004$, showing a strong positive correlation and that LMC ASKAP is well flux calibrated.

In Figure 3.6 the red line is slightly "above" the density ridge of the source distribution, so could be interpreted by the expectation that the small frequency difference from 843 to 888 MHz causes on average a lower flux density at 888 MHz. For $\alpha = -0.7$ (typical of synchrotron emission) this would be $(888\text{MHz}/843\text{MHz})^{-0.7} = 0.96$, whereas the true mode of the integrated flux density ratio is (0.946 ± 0.005), which is only a little smaller than expected.

We compare the positional differences (ΔRA and ΔDEC) between our new ASKAP catalogue and the previous SUMSS catalogue at 843 MHz in Figure 3.7. This figure shows that the positional offsets are well centred, with the mean offsets calcu-

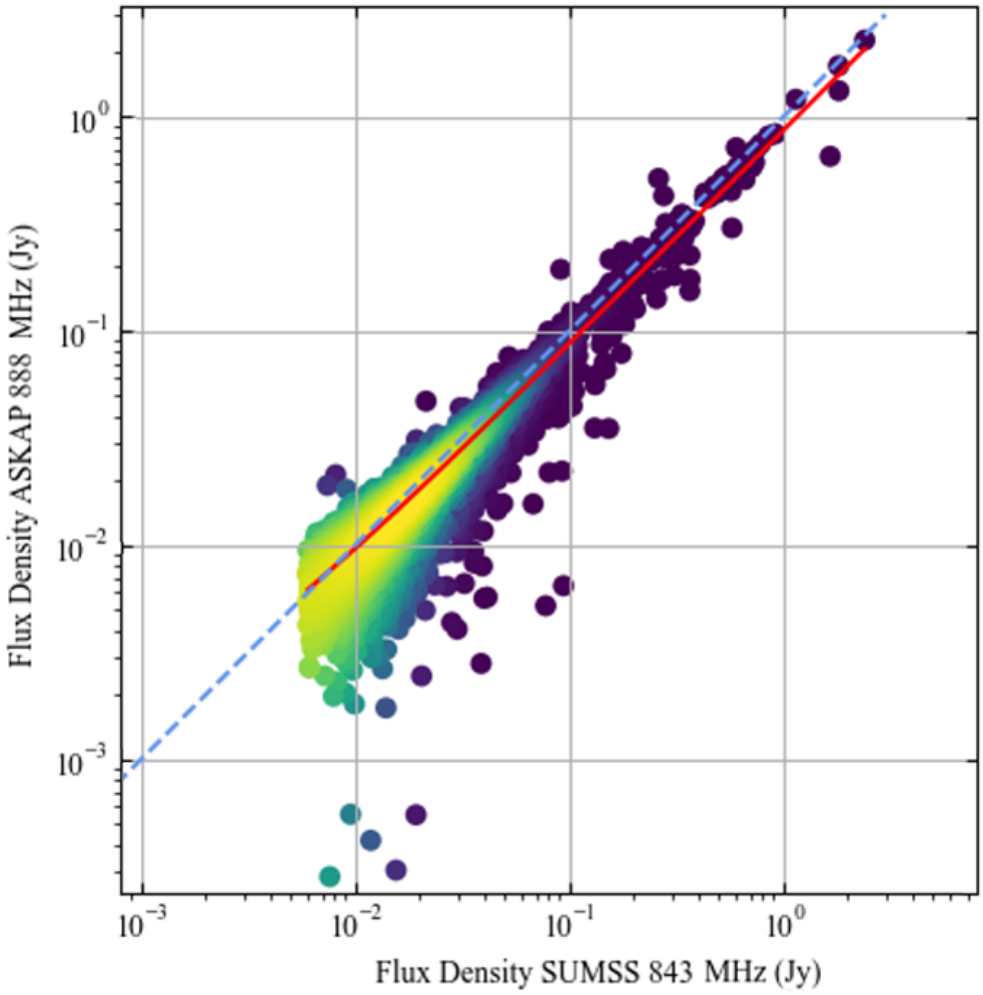

Figure 3.6: Comparison between the integrated flux densities calculated for the ASKAP 888 MHz image, and those from calibrated archival data of SUMSS MOST 843 MHz data. The blue dashed line indicates the one-to-one relation and the red line is the line of best fit. This shows a tight correlation. The outliers in the lower part of the figure are the single SUMSS sources that were resolved by ASKAP into multiple components.

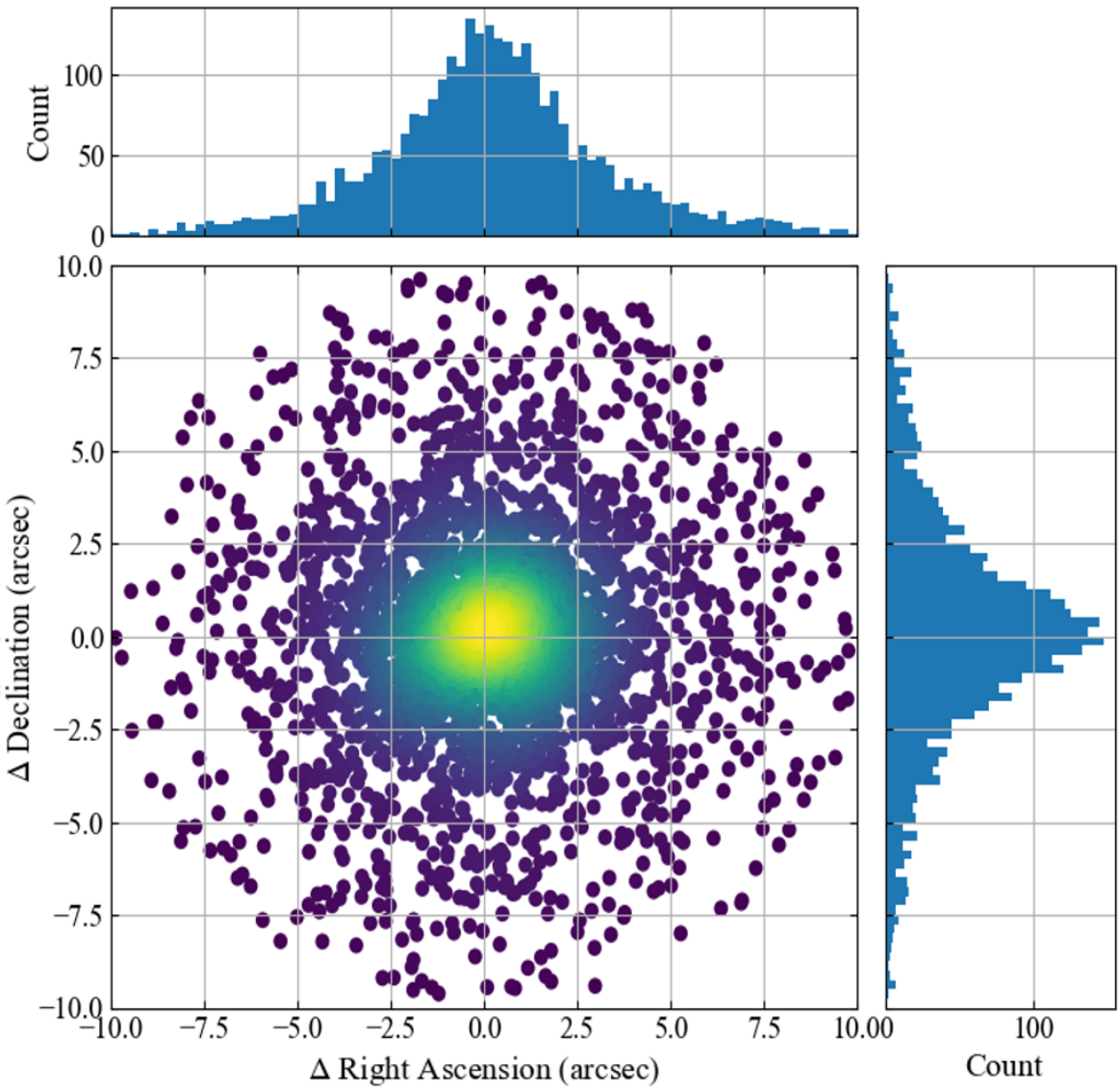

Figure 3.7: Figure of positional difference (MOST – ASKAP) of the 2949 radio sources from the ASKAP (888 MHz) catalogue and the SUMSS (843 MHz) catalogue.

lated to be $(0.189 \pm 0.054)''$ and $(0.088 \pm 0.054)''$ for RA and DEC, respectively. Overall, the ASKAP LMC astrometry is well-calibrated.

3.1.3.2 ASKAP SMC ESP

Two radio continuum images from the ASKAP survey in the direction of the SMC were taken as part of the EMU Early Science Project (ESP) survey of the Magellanic Clouds (Joseph et al., 2019). The 888 MHz observation of the LMC covers four times the area of the ASKAP SMC ESP observations.

Figure 3.8 shows the number of sources as a function of flux density for ASKAP LMC (888 MHz), ASKAP SMC (960 MHz and 1320 MHz) and SUMSS LMC (843 MHz). As expected, the number of sources is lower and the detection limit is higher in the ASKAP SMC catalogue compared to the ASKAP LMC catalogue. This is due to the smaller survey area, the differences in sensitivity for each image and the observing frequency.

The comparison of the SMC (960 MHz) sources with the LMC sources (888 MHz) where the SMC is scaled up to the same area size as the LMC (see Figure 3.9), shows increased sensitivity with the array fully operational and that within the flux limits of the SMC both LMC and SMC ASKAP fields show a similar source density. In contrast to the SMC observation, the LMC observation goes deep enough to surpass the upturn in the luminosity function and probes the radio-faint population (for more details see section 4.2). Overall, this shows the improvement of using the fully operational ASKAP array. Lastly, the LMC and SMC (the foreground sources) do not fill the area (at least not at high source density) of the ASKAP images, which means that the Clouds do not contribute a significant fraction to the predominantly extragalactic source density.

3.2 Galfit

Galfit (Peng et al., 2002) is a well-known software used for galaxy decomposition and by using it I hoped to shed some light on the structure of the sources with resolved host galaxies. It uses parametric functions to model objects as they appear in 2D images, i.e. modelling their light distributions. It can be used to determine the global

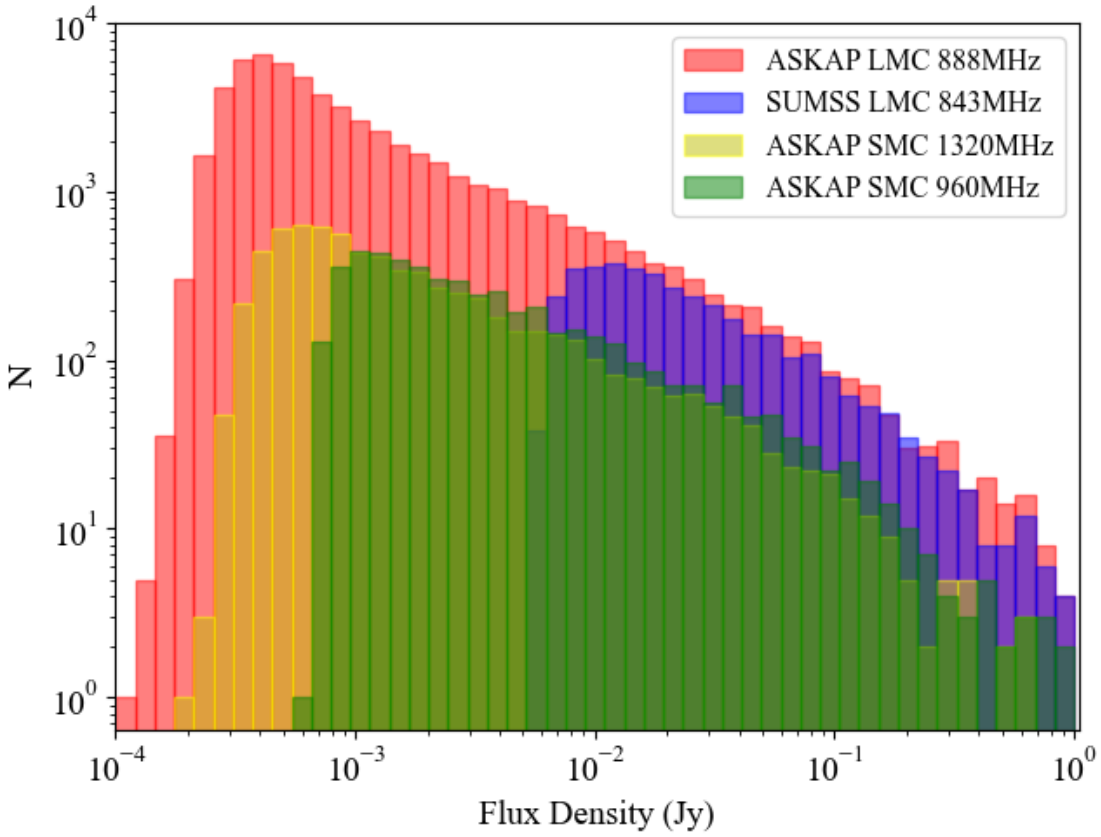

Figure 3.8: Integrated flux density histograms of ASKAP LMC (888 MHz), ASKAP SMC (960 MHz & 1320 MHz) and SUMSS LMC (843 MHz). The new ASKAP LMC image shows a marked improvement in sensitivity for a radio survey of the LMC and compared to the ASKAP SMC survey.

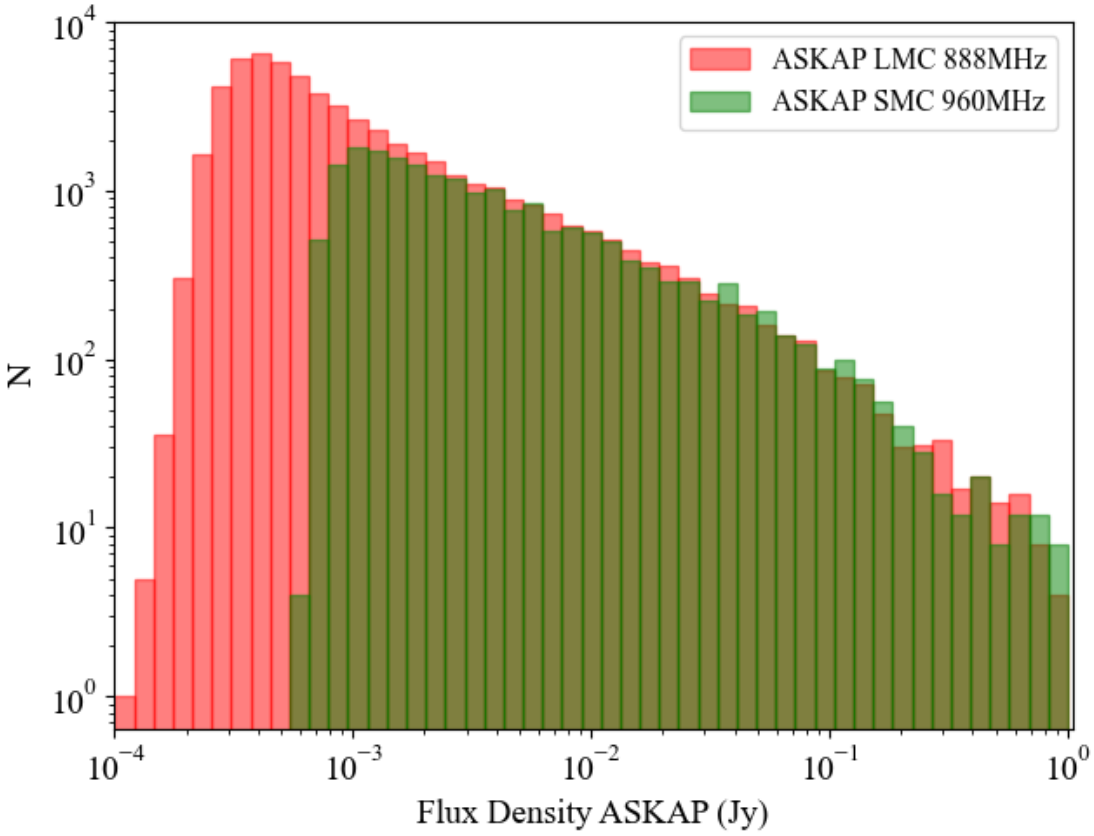

Figure 3.9: Comparison between the integrated flux densities calculated for the ASKAP images, of the LMC and SMC at 888 MHz and 960 MHz, respectively. The SMC numbers have been scaled to the same as the LMC image for better comparison.

morphology or to dissect a galaxy into its separate components such as bulge, disk, bar, etc.

GALFIT is a non-linear least-squares fitting algorithm that can fit complex images with multiple components and a large number of parameters. It determines goodness of fit by calculating χ^2 and then computes the adjustment of the parameters for the next iteration, and then continues to iterate until χ^2 no longer decreases significantly.

3.2.1 Models

GALFIT uses radial profile functions to model galaxies, where a radial profile describes the intensity fall-off of a model away from the peak. In general, early-type galaxies have a steep radial profile, whilst late-type galaxies have shallower radial profiles near the centre.

The Sérsic profile is a power law that has been frequently used to study galaxy morphology, and has the functional form of

$$\Sigma(r) = \Sigma_e \exp[-\kappa((r/r_e)^{1/n} - 1)] \tag{3.2}$$

where Σ_e is the pixel surface brightness at r_e, the effective radius. n is the concentration parameter, also known as the Sérsic index, and κ is a variable that is dependent on n. When n is small, it has a shallow inner profile and steep truncation towards larger radii. When n is large, it has a steep inner profile with an extended outer wing. Galaxy bulges can be described by a classical de Vaucouleurs profile, which is a special case of a Sérsic profile with $n = 4$. The disk of a galaxy can be described by the Exponential Disk profile, which is a Sérsic profile with $n = 1$. Another special case of a Sérsic profile is the Gaussian profile which is when $n = 0.5$.

Another profile is the Modified Ferrer profile, which has a nearly flat core and a sharp outer truncation, and is used mostly to fit galaxy bars and lenses. However, since a Sérsic profile with $n < 0.5$ has a similar shape, the Sérsic profile is often used instead.

Other profiles include the Empirical (Modified) King profile, used to fit globular clusters, and the Moffat profile, which describes the PSF of the Hubble Space Telescope WFPC2 images well, but otherwise is used less frequently for galaxy fitting than the Sérsic profile. The Nuker profile is used to fit central light distributions of nearby galaxies and the Edge-On Disk profile is used to describe a galaxy that is seen edge-on. Lastly, the PSF profile is used to fit unresolved sources, but can be replaced with a Gaussian profile with a small width.

Overall, the Sérsic profile and its counterparts at specific values of n are best used to describe the images fit in this work, as most of the other profiles can be replaced with a Sérsic profile and since the sources are all mostly face-on sources there is no need for the Edge-On Disk profile. An example of Galfit use can be seen in Figure 3.10, which is from Peng et al. (2010), where the three separate components of the model can be seen, all of which are Sérsic profiles with different values of n.

For describing galaxies with a spiral shape a coordinate rotation can be applied to a profile. The Power-law - Hyperbolic Tangent (α-tanh, see the appendices of Peng et al. (2002) for the full functional form of this rotation function), can account for a bar shape to a specified radius, after which the profile spirals outwards, and can describe tightly and loosely wound spirals.

3.2.2 Method

The method of decomposing the AGN was split into two main parts. First the PSF of the images had to be found. This was an input requirement needed to use Galfit. This was done in IRAF via identifying stars in the image and then using psf noao.digiphot.daophot to create the psf image.

After the PSF image was generated, it could be inputted into Galfit along with the original science image in order for Galfit to estimate the composition of the galaxy. Another input into Galfit is a text file with the models to be fitted, along with the parameters of each model that will be varied by Galfit until a solution is reached. After the inputs were put through Galfit, the program iteratively varies the

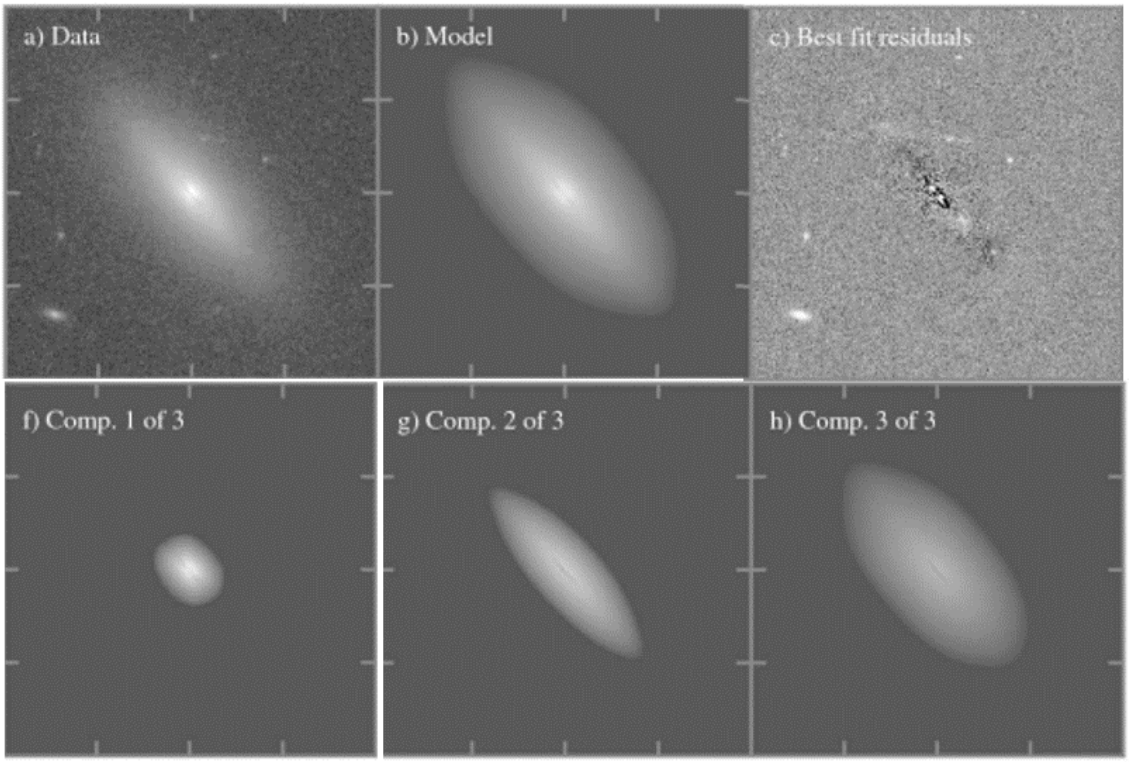

Figure 3.10: Example of Galfit modelling taken from Peng et al. (2010). The original image of the galaxy is shown in a), whilst b) shows the best fit model and c) shows the residuals leftover from subtracting the model from the original image. f), g) and h) show the three models used create the full model in b). f) is the bulge component, g) is the edge-on disk component and h) is the extended halo component.

parameters of each of the models to be fitted until the best fit model is found. If a model cannot be found then the program times out and gives an error message. When successful three outputs are given: the original image, the model and the residuals.

If there is a star close to the galaxy being fitted, then it causes GALFIT to wrongly estimate the location of the galaxy. Therefore any close-by stars are removed by using GALFIT to model the star with a PSF model and subtracting the model from the image to remove the star.

It was found that starting with more than one model with no parameters fixed led to GALFIT timing out and giving an error message. Therefore, one model was added to freely iterate over. The parameters of this model would then be fixed and another model would be added to freely iterate over. And so on until a good approximation of the true source was found. After this, the parameters would no longer be fixed, but their values would be close enough to the correct values that GALFIT would not time out.

The first attempt at using GALFIT to model was done with a single Sérsic profile. The image, PSF image and the list of parameters for the models to be fitted were then inputted into GALFIT. The centre of the object was initially estimated from the image and then the co-ordinates were inputted as a fixed parameter. Then GALFIT was run on the image to get the nearest approximation of the object. Next, to account for extended emission from the object an exponential function model was added. The first model's parameters were fixed and the parameters of exponential function model were allowed to iterate freely over its parameters to get a rough approximation of the extended emission. After this, if the residuals showed a circular area of light at the centre of the object, another Sérsic profile model would then be added to the program, the previous models parameters were fixed whilst the parameters of the new model were allowed to be iterated over.

GALFIT was then allowed to look for a solution with no fixed parameters, but provided with initial parameters from which to start its search from (which were found previously). This then provided a relatively good solution, with only a few residuals left over from the fitting.

3.3 CIGALE

Code Investigating GALaxy Emission (CIGALE; Noll et al., 2009; Boquien et al., 2019; Yang et al., 2020, 2022), is a versatile Python code for studying the evolution of galaxies by modelling the X-ray to radio spectrum of galaxies and estimating their physical properties such as star formation rate, attenuation, dust luminosity, stellar mass and characteristics of an active nucleus. It does this by comparing modelled galaxy SEDs to observed ones. Recently, Yang et al. (2022) developed an updated version of CIGALE, which allows for the modelling of the X-ray emission to account for X-ray fluxes in the fits of the SED. This version also includes a more recent AGN model, with a clumpy two-phase torus model derived from a radiative-transfer method (SKIRTOR model; Stalevski et al., 2012, 2016). This model also accounts for the presence of AGN polar dust extinction that has been observed in type 1 AGN (Gandhi & Hoenig, 2015). Furthermore, the radio models now account for radio emission from an AGN, not just star formation as it did previously. It is this version of the code that I used in this work.

3.3.1 Models

The CIGALE fit is made of a maximum of eight modules for which a selection of models can be picked, each corresponding to a unique physical process or component. These modules include the computation of: the star formation history (SFH) of the galaxy; the stellar spectrum from the SFH and simple stellar population (SSP) models; nebular emission lines and continuum from the Lyman continuum production; the attenuation of the stellar and nebular emission; dust emission in the mid-IR and far-IR from the galaxy; the emission from an AGN; the radio emission from star-formation and/or an AGN; the X-ray emission from the galaxy and/or an AGN. Note that for the radio, X-ray and nebular emission modules there is only one possible model each.

3.3.1.1 AGN models

The first AGN model of CIGALE is from Fritz et al. (2006) and assumes that the dusty torus is a smooth structure. However, more recent theoretical and observational works find that the torus is mainly made of dusty clumps (e.g. Nikutta et al., 2009; Stalevski et al., 2012). To account for this a newer model was added.

The SKIRTOR model is a library of AGN dusty torus emission models that were calculated with SKIRT, a radiative transfer code based on a Monte Carlo technique. In this model the dust distribution of the torus is modelled as a two phase medium. This medium consists of a large number of high-density clumps embedded in a smooth dusty component of low density. The advantage of this model is that it can produce both attenuated silicate features and pronounced near-IR emission at the same time, which both smooth and clumpy models find challenging. Since SKIRTOR's creation evidence, both simulated (Roseboom et al., 2013) and observational (Ponti et al., 2013; Markowitz et al., 2014; Leighly et al., 2015), has shown that the dusty torus is a multi-phase structure. It is the SKIRTOR model that is used in this work.

For the AGN model, modelled as a two phase torus (Stalevski et al., 2012, 2016), I set the extinction law of the polar dust to the SMC values (Prevot et al., 1984), the temperature to 100 K (e.g. Buat et al., 2021) and the emissivity index of the polar dust to 1.6 (Casey, 2012), default values that have been shown to work well for most AGN. All other parameters have been allowed to vary, which include: AGN fraction, the fraction of IR luminosity from the object that is due to the AGN; torus optical depth at 9.7 μm, τ; torus radial density parameter, pl and torus density angular parameter, q, such that $\rho \propto r^{-pl} e^{-q|\cos(\theta)|}$, where ρ is the torus density and r is the radius of the torus; opening angle, oa, which is the angle between the equatorial plane and edge of the torus; inclination angle, i, which is the viewing angle, where $i = 0°$ is face-on and $i = 90°$ is edge-on; the ratio between the maximum ($R_{\rm out}$) and minimum radii ($R_{\rm in}$) of the torus, R; the extinction caused by polar dust, $E(B-V)$. A schematic diagram outlining some of these parameters can be seen in Figure 3.11.

The radio module is also included as all the sources have radio observations, as the

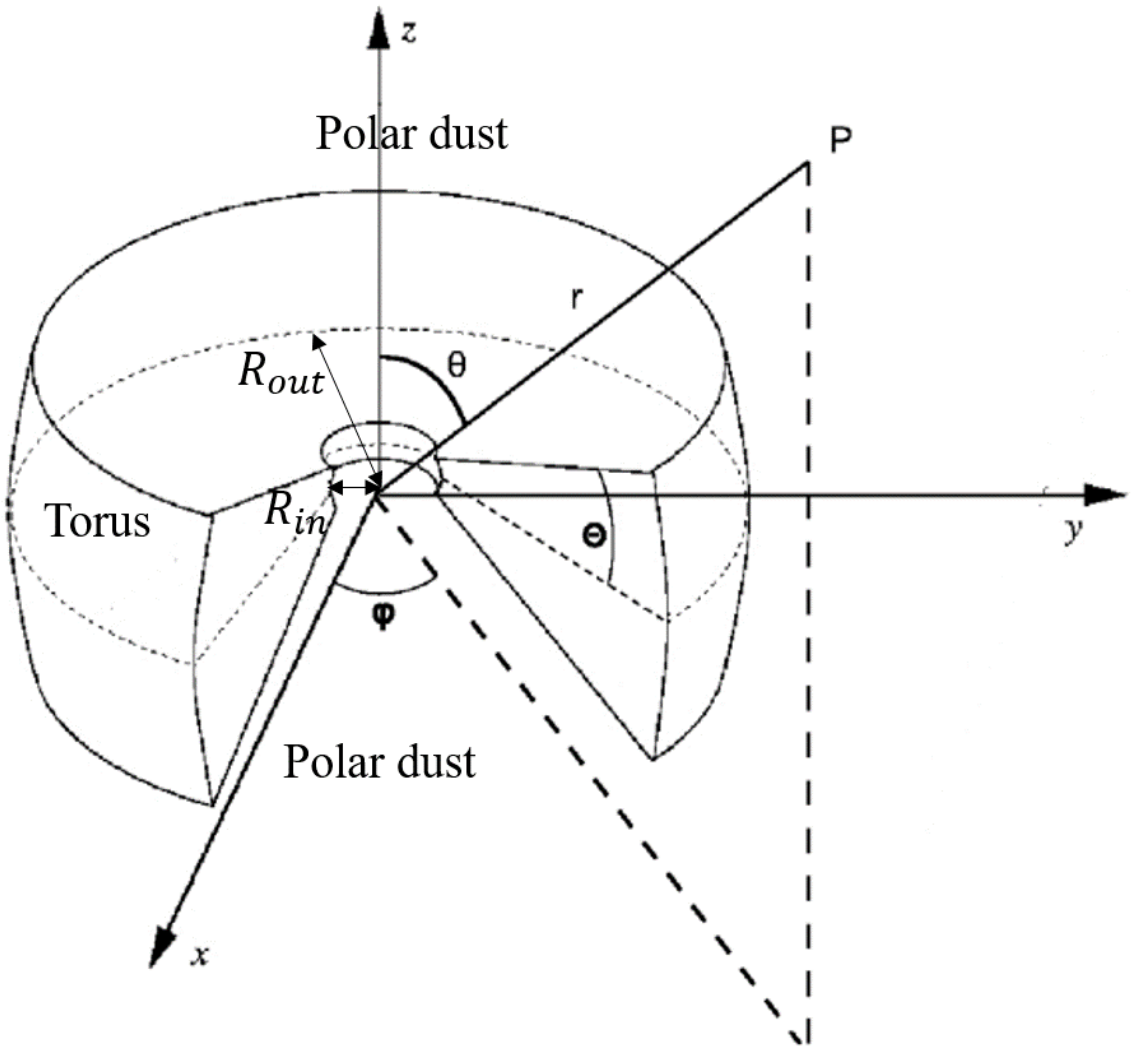

Figure 3.11: Annotated schematic of the SKIRTOR AGN model from Stalevski et al. (2012). Note that Θ represents the half opening angle in this diagram. Optional polar dust regions above and below the plane of the torus have been added to this model in the CIGALE version of this model.

recent update to CIGALE (Yang et al., 2022) now models radio emission from an AGN. Where there is only one radio observation, the spectral index, α, is set to the default of -0.7, typical of synchrotron emission. Where the sources have X-ray observations the X-ray module was implemented and allowed to vary over its parameters.

3.3.1.2 Galaxy component models

The first of the galaxy component models is the star formation history (SFH) module. The model used here is the delayed SFH with optional exponential burst which provides efficient modelling of early–type and late–type galaxies. The parameters that were allowed to vary were the e-folding time of the main stellar population, $\tau_{\rm main}$, from 100 – 4000 Myr, and the age of the main stellar population in the galaxy, t, from 100 – 6000, which provided a wide range of galaxy ages to test. The rest of the parameters were to do with the optional exponential burst in star-formation, which were set to default, and therefore not included in the model.

After computing the SFH, the second step is the simple stellar population module that combines a simple stellar population library with the SFH to compute the intrinsic stellar spectrum. The stellar spectra of the young and old stellar populations are computed separately in this module. There were two options for this module, both with the same input parameters. For this I selected the standard Bruzual & Charlot (2003) model. This model has three parameters: the initial mass function (IMF), which was set to 0 (corresponding to the IMF model by Chabrier (2003)) after many tests of the CIGALE algorithm showed that it was favoured over the other model; the metallicity, which was allowed to vary over the full allowed range from 0.0001 – 0.05 (note that only specific values are allowed); and the age of separation between the old and young stellar populations, which was allowed to vary from 1 – 10 Myr (where 10 Myr is the default value).

The modified dust attenuation law from Calzetti et al. (2000) is our third module, which controls the UV attenuation with the colour excess E(B−V), and also the power-law slope (δ) that modifies the attenuation curve. For this the only parameter changed

was the extinction law for the foreground reddening to use for attenuating the emission line flux, which was set to 'LMC' for sources behind the LMC and 'SMC' for sources behind the SMC.

We included the nebular emission module. Nebular emission is caused by ionisation of gas, which re-emits the energy in the form of continuum and emission lines that extend into the radio regime. However, nebular emission generally contributes little to broadband fluxes of quiescent star-forming galaxies, therefore I kept default parameters of the ionisation parameter (-3), fraction of Lyman continuum photons escaping the galaxy (0) and those absorbed by dust (0) and lastly the line width (300 km/s).

The module to model the dust emission in the SED uses a modified blackbody spectrum following Dale et al. (2014), which was chosen for its simplicity (the other models would have created a larger parameter space to iterate over and higher computation costs). This model has two parameters, the AGN fraction of the dust emission which was set to 0 (as the AGN emission is accounted for in a different model), and value β in $dM_{\rm dust} \propto U^{-\beta} dU$, where $dM_{\rm dust}$, is the dust mass and U is the radiation field intensity. β was allowed to vary across the full breadth of the allowed values, 0.0625 – 4. Where there were no observations (or were all upper limits) for the far-IR regime (such as in SAGE0536AGN, which has shown no far-IR emission), this model was left out as it could not be fit well.

For the galaxy components of the radio and X-ray models the default parameters were kept as the majority of the radio and X-ray emission is expected to be from the AGN and not the host galaxy.

3.3.2 Inputs

The known redshifts and photometry from SMASH, Gaia EDR3, VMC, SAGE, All-WISE and HERITAGE (Meixner et al., 2013) were used to model the SEDs of 17 objects (see Chapter 5 for details) that are known and candidate AGN similar to SAGE0536AGN, discovered using the t-SNE machine learning algorithm (see Section 3.4.1). Not all sources had far-IR fluxes, due to either being outside of the HERITAGE

survey field of the Magellanic Clouds or the fluxes being too faint. Where far-IR fluxes were not found in images an upper limit on the flux was measured from the HERITAGE images. For Source 8, where there was no spectroscopically determined redshift, the photometric redshift, measured by Flesch (2015, 2021) was used.

Models used and the parameters that were varied over the fit are shown in Table 3.3. Each AGN was initially fit without extragalactic dust model and where the models did not fit in the far-IR and showed $f_{\rm AGN} < 0.99$, where $f_{\rm AGN}$ is the fraction of the total dust that is due to the AGN, the extragalactic dust model was then added, which is the case for five of the AGN.

3.4 Machine learning

3.4.1 t-SNE

In order to find sources similar to SAGE0536AGN and SAGE0534AGN in a large dataset with ill defined properties, I employ machine learning, which has been used to great effect to separate sources into different classes (e.g. Lochner et al., 2016; Anders et al., 2018; Reis et al., 2018b; Zhang et al., 2020).

t-SNE (t-distributed stochastic neighbour embedding; van der Maaten & Hinton, 2008) is an unsupervised machine learning dimensionality reduction algorithm. It can visualise any high–dimensional dataset by projecting each data-point onto a low–dimensional map, which reveals local as well as global structure of the data at many different scales. t-SNE has been shown to be adept at separating sources into different classes with no prior information about the source nature (e.g. Steinhardt et al., 2020).

t-SNE uses hyperparameters (perplexity, early exaggeration, learning rate and number of steps) and is a non-linear and non-deterministic (output is different each time it is used) technique. It embeds the points from a higher dimension into a lower dimension whilst trying to maintain the neighbourhood of that point, preserving the

Table 3.3: Modules and parameter values used to model the sample in CIGALE. For the parameter values not listed the default values were used.

Parameter	Model/Values	Description
Star formation history	delayed SFH with optional exponential burst	
$\tau_{\rm main}$	100 – 4000	e-folding time of main stellar population model (Myr).
t	100 – 6000	Age of oldest stars in galaxy (Myr).
Simple stellar population	Bruzual & Charlot (2003)	
IMF	0	Initial Mass Function from Chabrier (2003)
Metallicity	0.0001, 0.01, 0.02, 0.05	Metallicity, where solar metallicity ~ 0.02.
Separation age	1, 5, 10	Separation between young and old star populations (Myr).
Galactic dust attenuation	Modified Calzetti et al. (2000) attenuation law	
$E(B-V)$	0.4	Colour excess of nebular lines (mag).
Ext. law em. lines	LMC, SMC	Extinction law for attenuating emission lines flux (Pei, 1992).
Galactic dust emission	Dale et al. (2014)	
β	0.0625 – 4	Slope in $dM_{\rm dust} \propto U^{-\beta} dU$
AGN	SKIRTOR UV-to-IR, from Stalevski et al. (2012, 2016)	
τ	3, 5, 7, 9, 11	Optical depth at 9.7 μm.
pl	0, 0.5, 1, 1.5	Torus radial density parameter, such that $\rho \propto r^{-pl} e^{-q\lvert\cos(\theta)\rvert}$, where ρ is the torus density and r is the radius of the torus.
q	0, 0.5, 1, 1.5	Torus density angular parameter.
oa	10, 20, 30, 40, 50, 60, 70, 80	Angle between the equatorial plane and edge of the torus.
R	10, 20, 30	Ratio of the outer to inner radii of the dust torus, $R_{\rm out}/R_{\rm in}$.
i	0, 10, 20, 30, 40, 50, 60, 70, 80, 90	Viewing angle where face-on: $i = 0°$, edge-on: $i = 90°$

Parameter	Model/Values	Description
$f_{\rm AGN}$	0.6, 0.7, 0.8, 0.9, 0.999	AGN fraction, $f_{\rm AGN} = \frac{L_{\rm dust,AGN}}{L_{\rm dust,AGN}+L_{\rm dust,galaxy}}$, where $L_{\rm dust,AGN}$ and $L_{\rm dust,galaxy}$ are AGN and galaxy dust luminosity integrated over all IR wavelengths, respectively.
δ	-0.36 – 0.36	Power-law modifying the optical slope of the disk.
$\rm Law_{polar}$	SMC	Extinction law of polar dust.
$E(B-V)_{\rm polar}$	0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0	Polar-dust colour excess (mag).
$T_{\rm polar}$	100 K	Temperature of polar dust.
$\rm Emissivity_{polar}$	1.6	Emissivity index of polar dust (see equation (10) of Yang et al., 2020).
X-ray		
Γ	1.5 – 2.0	Photon index, Γ, of the AGN intrinsic X-ray spectrum.
α_{ox}	-1.9, -1.8, -1.7, -1.6, -1.5, -1.4	UV/X-ray slope calculated at $i = 30°$.
Radio		
$\alpha_{\rm SF}$	0.8	Slope of the power-law synchrotron emission related to SF, which is a free power-law slope.
$\alpha_{\rm AGN}$	0.01 – 2	Slope of the power-law AGN radio emission, defined as $L_{\nu,\rm AGN} \propto \nu^{-\alpha_{\rm AGN}}$.
$R_{\rm AGN}$	0.1 – 300	Radio-loudness parameter, defined as $L_{\nu,\rm 5GHz}/L_{\nu,2500\mathring{\rm A}}$, where $L_{\nu,\rm 5GHz}$ and $L_{\nu,2500\mathring{\rm A}}$ are the monochromatic AGN luminosities per frequency at rest-frame 5 GHz and 2500 Å.

local structure of the data. More specifically, the t-SNE technique minimizes the divergence between a probability distribution that measures pairwise similarities of the high-dimensional data and a probability distribution that measures pairwise similarities of the low-dimensional points in the embedding. Unlike the linear Principal Component Analysis (PCA) algorithm, t-SNE cannot preserve global structure (variance) but can preserve the local structure, allowing fine structures to be found, which PCA is incapable of.

I searched for SAGE0536AGN and SAGE0534AGN analogues to further explore this AGN class. I used the t-SNE algorithm on a clean dataset (no error/missing values) of 1359 sources that was the combination of VMC, Gaia EDR3 (Gaia Collaboration et al., 2021), AllWISE (Cutri et al., 2021) and EMU ASKAP 960 MHz (Joseph et al., 2019) and 888 MHz (Pennock et al., 2021) photometry, colours and astrometry in the area of the SMC. Where there were no proper-motions, usually an indication of an extragalactic extended source, the proper-motion value was assumed and set to 0. Surveys of the SMC have also been performed in the X-ray (e.g. Sturm et al., 2013), UV (GALEX; Martin et al., 2005) and mid to far-IR (SAGE/HERITAGE; Meixner et al., 2006, 2013). These were not used because they lack the same coverage of the Magellanic Clouds as the VMC survey, as well as having missing values for many of observed sources, which would have caused the sample to be explored to be reduced significantly. We focus on the SMC because the VMC Point-Spread Function (PSF) photometry and ASKAP radio survey were available for the SMC first.

The perplexity parameter of t-SNE is related to the number of nearest neighbours that is used in the algorithm, where larger datasets usually require a larger perplexity. Perplexity was chosen by creating multiple maps and choosing the value of perplexity that created the most obvious clustering. Figure 3.12 shows what a range of perplexity values does to the outputted t-SNE map. A low perplexity value leads to too much separation between sources, whilst a large perplexity leads to little separation between sources. A value of around perplexity = 50, seemed to show good clustering.

This technique reduced the high-dimensional dataset down to two dimensions, producing a t-SNE map seen in Figure 3.13.

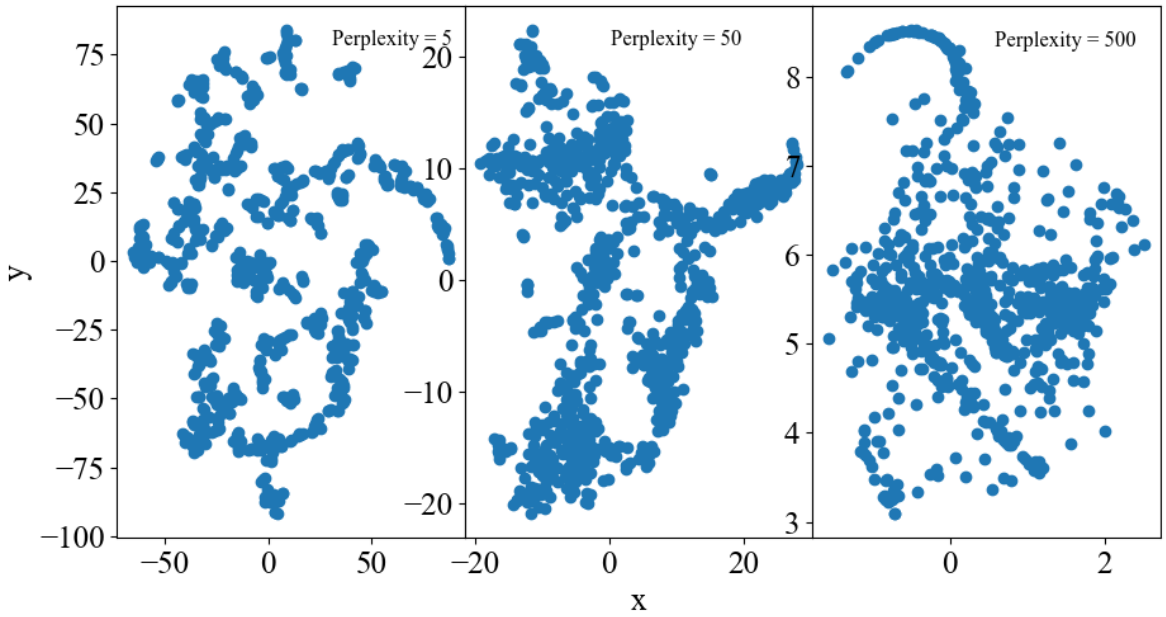

Figure 3.12: t-SNE maps created from a combination of VMC, Gaia EDR3, AllWISE and ASKAP data. Different values of perplexity were used for each map, where (left) used a perplexity of 5, (centre) used a perplexity of 50, and (right) used a perplexity of 500. This shows that a high perplexity gives less distinct clustering and a low perplexity gives too much clustering/separation.

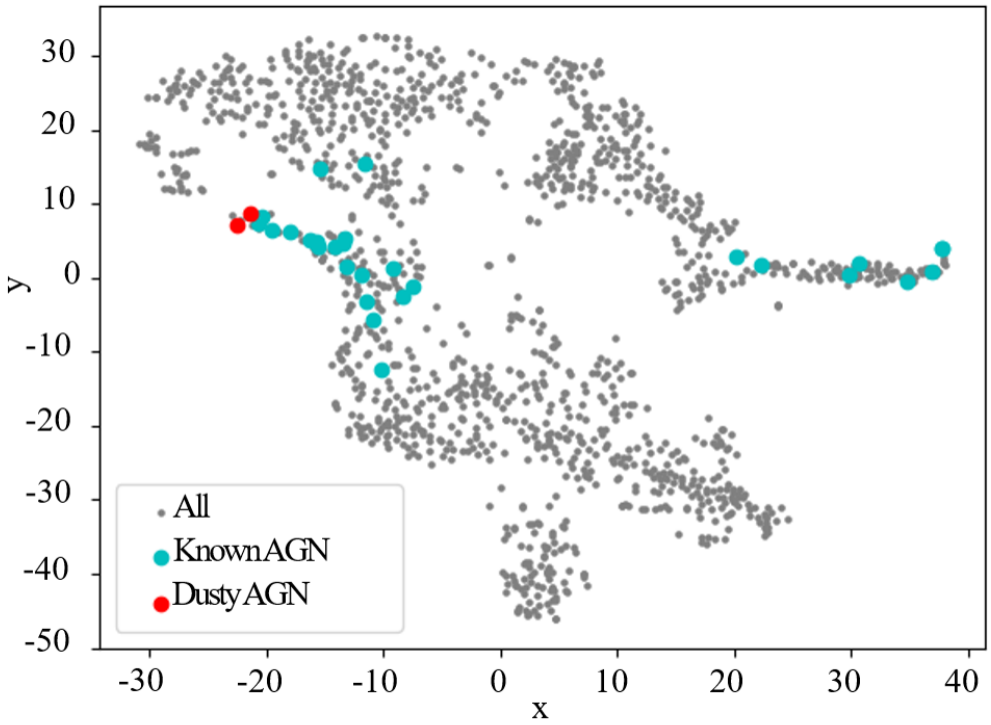

Figure 3.13: t-SNE map created from a combination of VMC, Gaia EDR3, AllWISE and ASKAP data. Blue dots represent known AGN that have been spectroscopically confirmed. Red indicates the dusty AGN SAGE0536AGN and SAGE0534AGN.

3.4.2 Machine learning classifier

3.4.2.1 Decision Trees

Decision trees (James et al., 2017) are a non-parametric supervised machine learning method that can be used for both regression and classification tasks. A decision tree predicts the value/class of a target variable by creating a model that has learnt simple decision rules inferred directly from the data features it is trained upon. It is a model that is easily visualised (see Figure 3.14). It consists of nodes where a condition is given that is either true or false. The answer to this condition leads down a "branch" to the next node and condition, where either another split happens or the output variable is given. In training, the "branches" split when only one target class/value of the training data remains at the end. If there is more than one class it keeps splitting the data until that is no longer the case. This is called the stopping criterion.

An advantage to using decision trees is that they require very little data preparation. Most machine learning algorithms usually require the data to be normalised first, dummy variables to be created and blank values removed. Though most decision trees do not support missing/blank values, this can be worked around by replacing the blank value with a value far away from the range of values in a particular feature variable. Decision trees, unlike other algorithms, can also perform both classification and regression tasks. They are able to handle multi-output problems (more than one output for each single input) and use a white box model, which means results are easier to interpret as they can be explained by a condition that uses Boolean logic, as opposed to a black box model where you can only observe the input and output.

Disadvantages to using decision trees include being biased if some classes dominate in the training set, so the dataset would need to be balanced prior to fitting the decision tree. Over-complex versions can be created which do not generalise the data well and cause overfitting. Overfitting occurs when a machine learning model gives accurate predictions for training data but not for new data. This can be avoided by minimising the number of samples required in a node, and/or setting a maximum depth

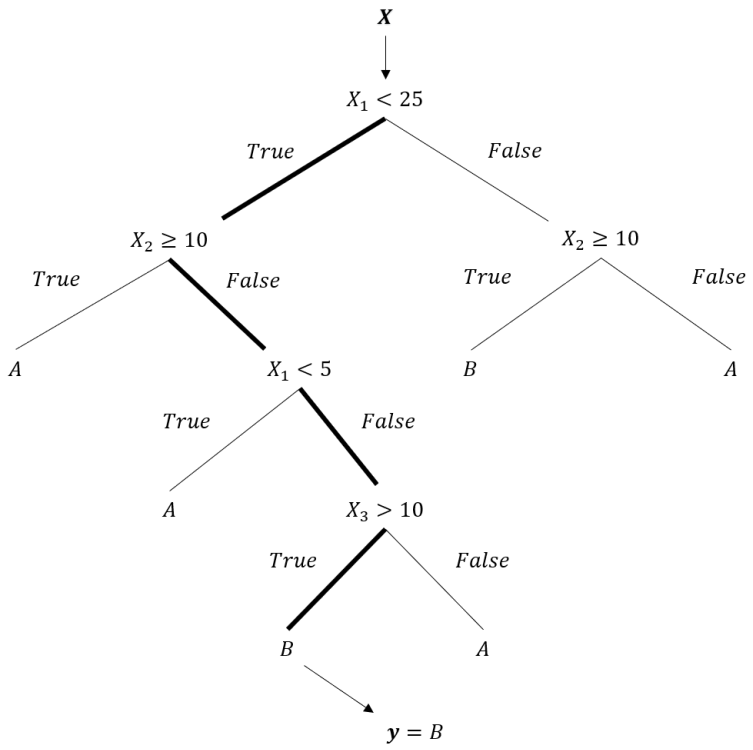

Figure 3.14: Decision Tree example: A non-parametric model. The machine learning algorithm creates the model during the training phase. X represents the input data with features which are used to separate the data. The y represents the output (predicted) class/value. It is described by a tree-like graph, where each split in the tree separates the data based on a condition involving one feature. It can be used for both classification and regression tasks. In training the stop criterion for a "branch" is that only one class remains.

of the tree, so fewer branches are produced, reducing the complexity. They can also be unstable, meaning that small variations in the data, or even choosing a different subset to train upon, could lead to vastly different trees being generated. Most of the disadvantages of decision trees can be mitigated by using decision trees within an ensemble, i.e. multiple decision trees used together.

3.4.2.2 Random Forests

One such ensemble of decision trees is called the random forest algorithm (Breiman, 2001), which can also be used for both classification and regression problems. The algorithm builds several decision trees independently and then averages the predictions of these to obtain the final prediction (see Figure 3.15). This reduces variance over using a single estimator and creates an overall more stable model. It is called a random forest because randomness is injected into the training process of each individual tree via a method called 'bagging'. This method splits up the training set into randomly selected subsets, and each decision tree is then trained on one of those subsets. Furthermore, at each node of the decision tree, only a randomly selected subset of the features is considered.

Examples of where random forests have proven to be particularly effective are in classification tasks such as photometric classification of type 1a supernovæ (Möller et al., 2016), classification of variable stars (Masci et al., 2014) and creating a catalogue of quasars from survey data (Nakoneczny et al., 2019).

Advantages of the random forest include all the advantages of using a decision tree. Also, it constructs a non-linear model during training, which is advantageous as most problems require non-linear solutions. In its simplest form, it has very few free parameters and can handle numerous features and numerous objects. It can produce classification probabilities, showing you how certain the model is for each classification it has made, simply calculated by counting the fraction of trees in a forest that vote for a certain class. It produces feature importance, showcasing which features have more bearing on the classifications it had made. Lastly, it generalises well to unseen datasets due to the inclusion of randomness, and therefore reduces overfitting.

The main disadvantage of a random forest is that it cannot handle measurement uncertainties, but this is true for most machine learning algorithms.

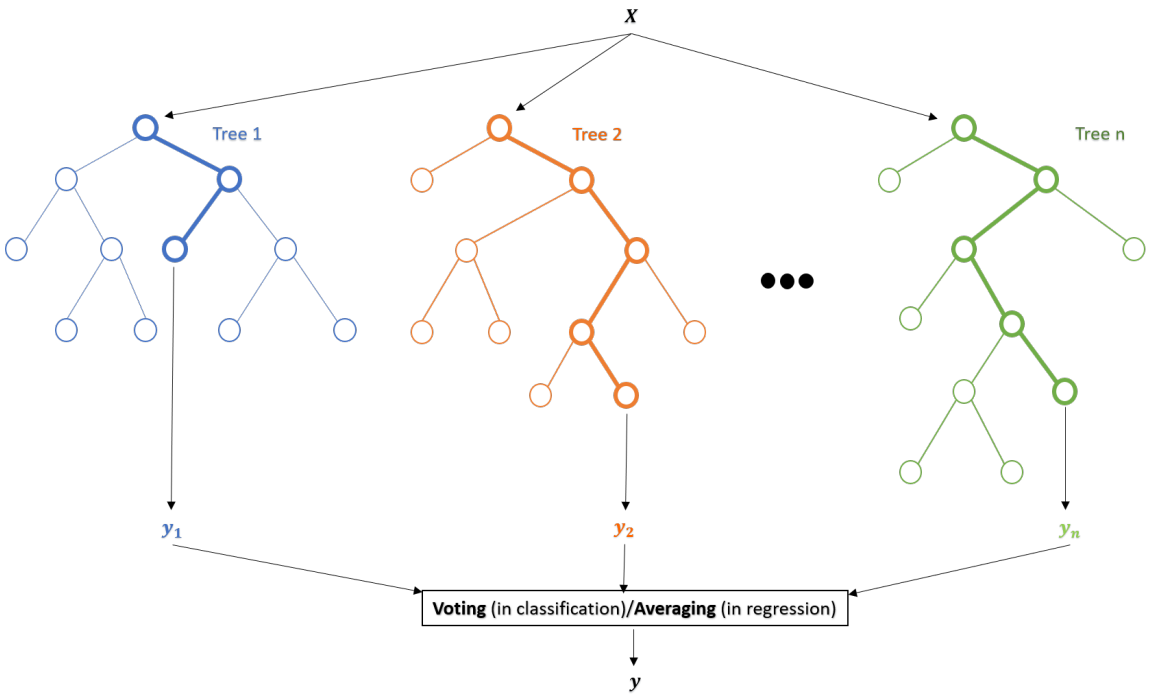

Figure 3.15: Random Forest example: A model consisting of multiple decision trees. The final result is based on the output of all the trees in either a voting system (classification), where the most often predicted class is the output, or via averaging (regression), where the final output is an average of all the values predicted by the decision trees.

3.4.2.3 Probabilistic Random Forest

The Probabilistic Random Forest [2] (PRF; Reis et al., 2018a) is a random forest algorithm that can handle and take into account measurement uncertainties and missing data. Compared to an ordinary random forest it has been proven to provide an up to 10% increase in classification accuracy with noisy features and proven to be more accurate than the original random forest when up to 45% objects in the training set are misclassified. Thus this would prove helpful and would allow for the inclusion of AGN in the training set that have been classified as AGN but not via spectroscopy.

[2]Python code can be found here: `https://github.com/ireis/PRF`

This algorithm was created in Python and requires the Python module Scikit-learn Pedregosa et al. (2011) to run.

3.4.3 Creating the multi-wavelength dataset

The base of the multi-wavelength data set is the near-IR VMC PSF survey catalogue. All coordinate matchings were made to the VMC coordinates. Through the use of TOPCAT (Taylor, 2005) the VMC catalogue was matched with SMASH (Nidever et al., 2017), Gaia DR3 (Gaia Collaboration et al., 2022), SAGE (Meixner et al., 2006; Gordon et al., 2011), UnWISE (Schlafly et al., 2019) and AllWISE. The parameters of different surveys in the dataset can be seen in Table 3.4.

Some of the parameters had to be calculated. For instance, the unWISE catalogue only used fluxes and not Vega magnitudes like the rest of the catalogues. For consistency the fluxes were converted using the method recommended in the notes of the table on CDS. The fluxes in Vega nanomaggies (nMgy; Finkbeiner et al., 2004) were converted to Vega magnitudes using $m = 22.5 - 2.5\log(flux)$. These fluxes showed slight discrepancies with the AllWISE values and a correction was provided (see Schlafly et al., 2019) of subtracting 0.004 mag and 0.032 mag from unWISE W1 and unWISE W2, respectively. The discrepancies, using the SMC training set of sources, between unWISE W1 and AllWISE W1, as well as unWISE W2 and AllWISE W2 can be seen in Figure 3.16. This shows that the differences for both bands is, as expected, centred around 0. Differences beyond this could be explained by variability.

Other parameters that had to be calculated were colours between all photometry bands and their corresponding errors, which were calculated with standard propagation of errors. Note that if, for example, $Y - J$ was calculated, the reverse, $J - Y$, would not be calculated and added as a feature. This led to a total of 237 features.

Different surveys have different ways of designating sources that were unobserved/missing for a particular parameter. For example, in the VMC survey they use an extremely negative value far from the expected range, -9.99999E8, and in SMASH they use the positive number 99 instead. The SAGE survey tends to leave missing

Table 3.4: Parameters taken from various surveys to act as features in the PRF algorithm. For each parameter the error on the values is also taken from the corresponding surveys. New features were created by subtracting each feature from all the other features to create colours. This did not include the sharpness (a measure of the difference between the observed width of the object and the width of the PSF model, where stars should have a sharpness value of 0 and resolved objects shaprness values of >0. Sharpness values < 0 indicate artefacts such as bad pixels or cosmic ray impacts.) and proper-motions in RA and DEC (pmRA and pmDEC, respectively).

Parameter	Units	Survey
Y PSF	mags (Vega)	VMC
J PSF	mags (Vega)	VMC
$K_{\rm s}$ PSF	mags (Vega)	VMC
Y sharp PSF	–	VMC
J sharp PSF	–	VMC
$K_{\rm s}$ sharp PSF	–	VMC
u	mags (Vega)	SMASH
g	mags (Vega)	SMASH
r	mags (Vega)	SMASH
i	mags (Vega)	SMASH
z	mags (Vega)	SMASH
sharp	–	SMASH
pmRA	mas/yr	Gaia DR3
pmDEC	mas/yr	Gaia DR3
G	mags (Vega)	Gaia DR3
$G_{\rm BP}$	mags (Vega)	Gaia DR3
$G_{\rm RP}$	mags (Vega)	Gaia DR3
IRAC 3.6 μm	mags (Vega)	SAGE
IRAC 4.5 μm	mags (Vega)	SAGE
IRAC 5.8 μm	mags (Vega)	SAGE
IRAC 8.0 μm	mags (Vega)	SAGE
unW1	mags (Vega)	unWISE
unW2	mags (Vega)	unWISE
W1	mags (Vega)	AllWISE
W2	mags (Vega)	AllWISE
W3	mags (Vega)	AllWISE
W4	mags (Vega)	AllWISE

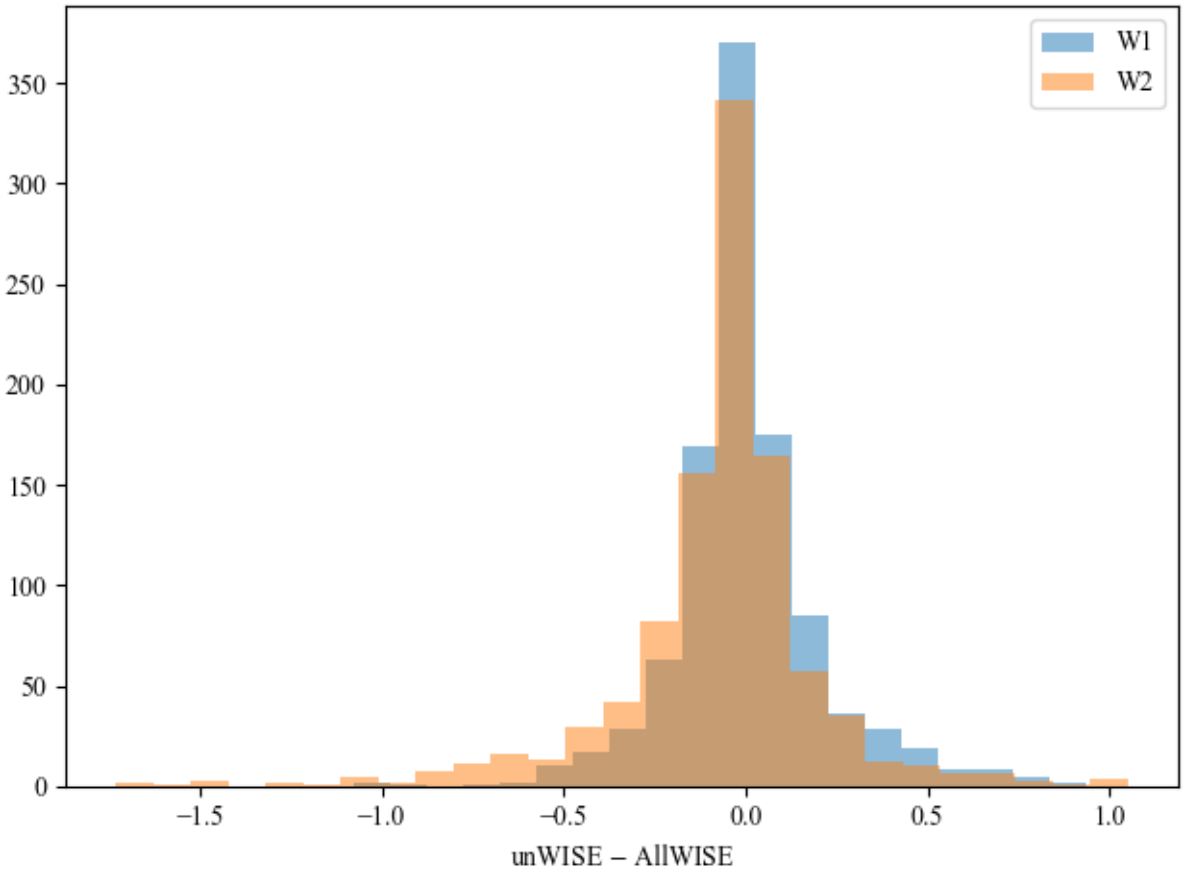

Figure 3.16: Histogram showing the differences between W1 (blue) and W2 (orange) magnitudes for the unWISE and AllWISE surveys.

values blank, but in some instances there are values of 0 instead. Before this dataset can be used with the PRF these designations needed to be homogenised. This was done by setting these designations to "NaN", so the PRF can treat them correctly as missing values.

3.4.4 Training

3.4.4.1 Inputs

The data were put in three configurations, individual 'LMC' and 'SMC' datasets, as well as a further 'MC' (LMC and SMC) data set. Each configuration had two versions, one with no colour features and another with colour features.

For the training the datasets were split into features, X, errors on features, dX, and class, y.

To ascertain the accuracies of the trained classifiers each dataset was split into training and testing sets, where 75% of the data were trained on and 25% were retained to test the classifier on. For each of the training runs the data split was randomised. Note that when testing a machine learning model a training dataset is often split into training, validation and test set. A validation set is used to tune the parameters of the model, whilst a test set is used to test the final model. Both of the datasets are not trained upon. In the interest of not splitting the different sets into too small groupings, and therefore not providing a good overview of how well the trained classifier works, I combined the validation and test sets together for both tuning the model and testing the final model into an overall 'test' set.

3.4.4.2 PRF parameters

The probabilistic random forest classifier has parameters that can be varied. Most parameters are set to default.

There is the probability threshold parameter. In an ideal PRF all objects propagate to all branches of the tree, this requires a higher amount of computation time, and for any given object there may be nodes with small propagation probability. Stopping the propagation at these nodes reduces the run time without decreasing overall importance. The probability threshold parameter, $p_{\rm th}$, determines the probability threshold at which to stop propagating along a branch. In an ideal PRF, $p_{\rm th} = 0$, where all objects propagate along all branches to all terminal nodes. Whereas $p_{\rm th} = 1$ denotes

a classical RF, where each object propagates to only one terminal node. In Reis et al. (2018a) they found that reducing the probability threshold below a value of $p_{\text{th}} = 0.05$ does not improve prediction accuracy, so this is the value that was used in this work.

The number of trees/estimators, n_{trees}, is another input. The value of this is determined in section 3.4.4.3.

3.4.4.3 Determining the number of trees

To determine the optimum number of trees for the PRF, for each of the datasets, the dataset was split into training and testing and then the classifier was trained on this at $n_{\text{trees}} = 1, 5, 10, 25, 50, 100, 200, 500$, and then the score (average class accuracy) was calculated. Next the whole dataset was split randomly again and then the classifier was trained again on the different number of trees. This is done for five iterations for each n_{trees} and then the score is averaged for each value of n_{trees}. The results of this can be seen in Figure 3.17.

From this it can be seen for all dataset configurations that after $n_{\text{trees}} > 100$ the score for each classifier plateaus. Therefore the value of the number of trees will be set to $n_{\text{trees}} = 100$. It can also be seen that the overall accuracy is greater when colour features are included.

3.4.4.4 Dataset configuration test

From the test to determine the optimum number of trees, see Figure 3.17, we can see the overall accuracy of the classifiers. This shows that the SMC-trained classifier tested on SMC data has the greatest overall accuracy and the LMC-trained classifier tested on LMC data has the lowest overall accuracy. The overall accuracy does not tell one how the classifier performs on individual classes, this can be shown through the use of confusion matrices.

A confusion matrix shows the comparison between the true labels (y-axis) vs the predicted labels (x-axis). A perfect classifier would show a value of 1 (100%) in a

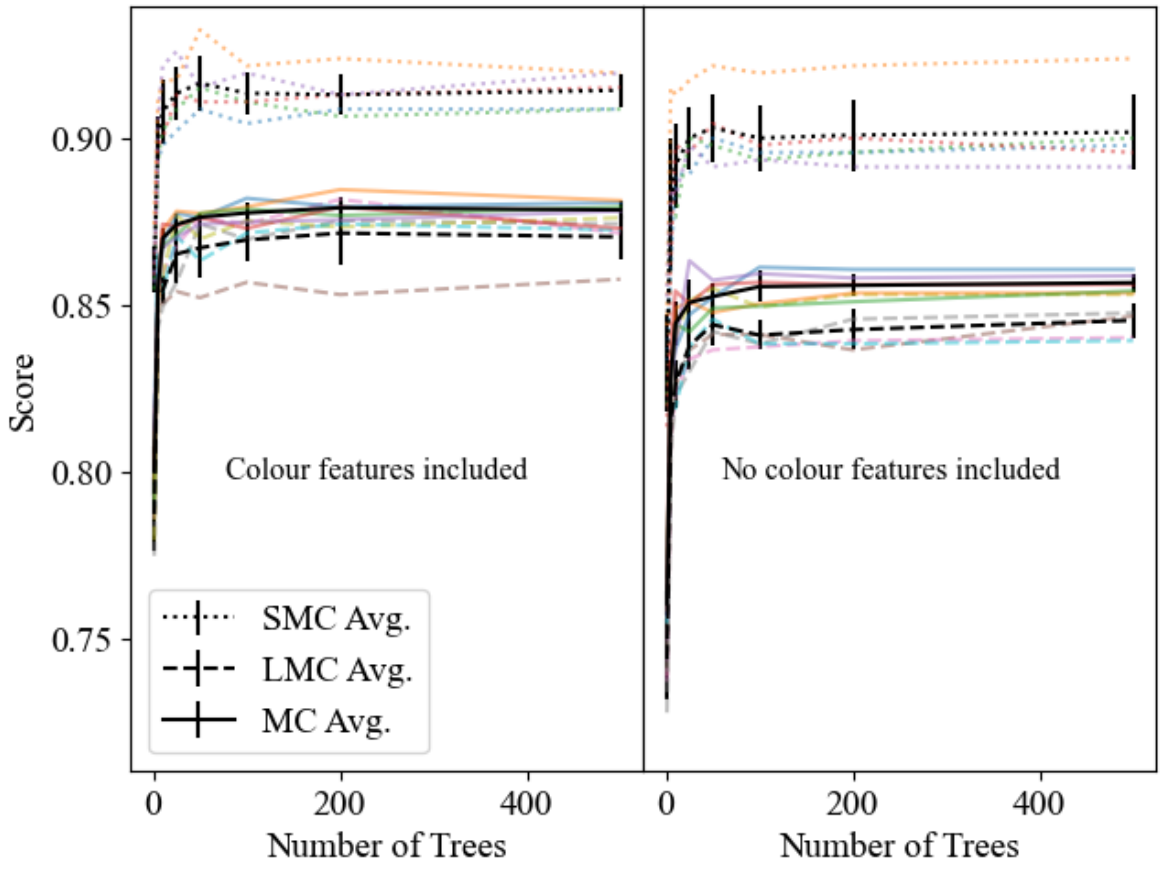

Figure 3.17: Plot of score (average accuracy) against number of trees used in training classifier for the SMC, LMC and MC datasets with colour features included (left) and not included (right). For each value of the number of trees the classifier was trained and tested five times and averaged. The black lines represent the average and the coloured lines represent the individual runs.

diagonal line from top left to bottom right of the confusion matrix, with all the other values as 0 (0%). This would show that all classes have been predicted correctly. The values for an entire row should sum to 1, showing the distribution of class predictions for each class. In Figures 3.18, 3.19 and 3.20 are the confusion matrices of SMC, LMC and MC datasets, respectively.

For extragalactic sources, the recall (the ratio of $\frac{tp}{tp+fn}$, where tp is the number of true positives and fn is the number of false negatives) of AGN is better for the

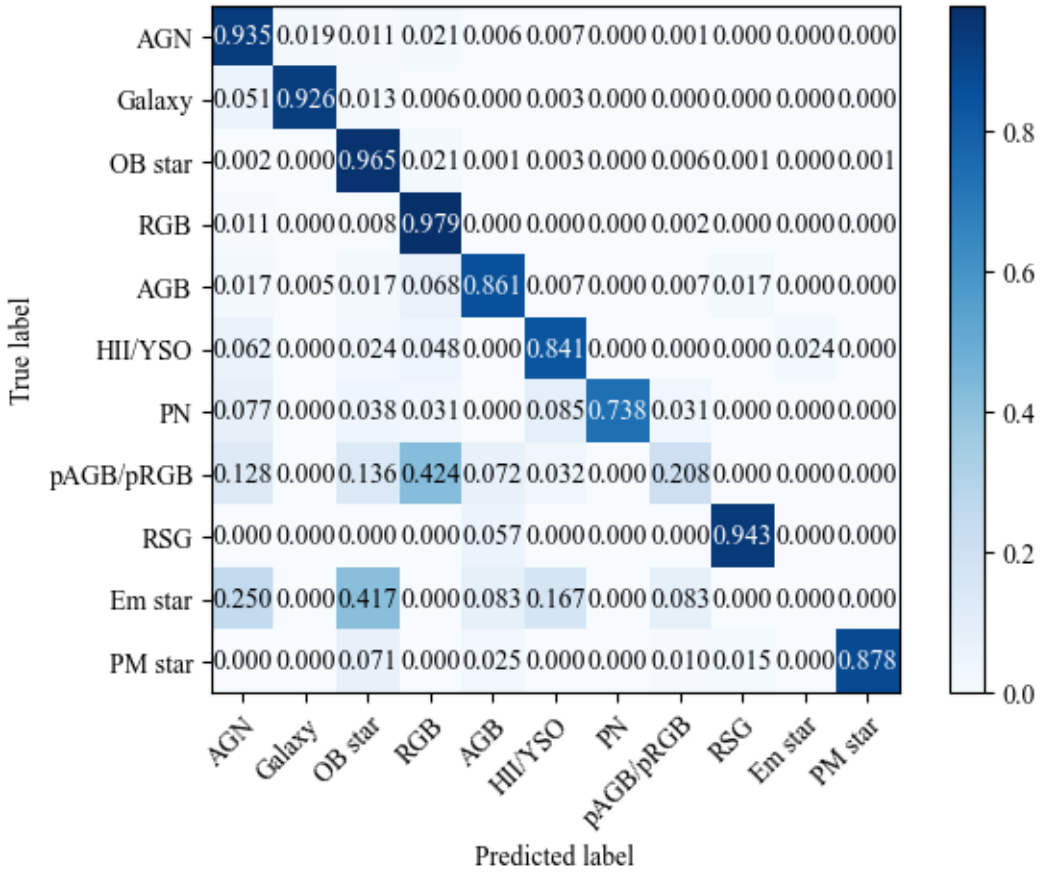

Figure 3.18: Confusion matrix of classifier trained 75% of the SMC dataset and tested on the remaining 25%.

LMC-trained classifier, where recall $\sim$ 0.97, and worst for the SMC where recall is $\sim$ 0.94. The SMC-trained classifier has the most false positives for AGN, this is most likely due to the smaller training set. The recall of the galaxies is marginally greater for the SMC. Most misclassified AGN tend to be classified as galaxies and vice versa, which is expected as there is not a fine line in parameter space between galaxy and AGN, especially at lower redshifts.

For the stellar sources, the SMC, if one discounts the emission line stars for which there were only five to train and test on, the SMC generally did better on classifying

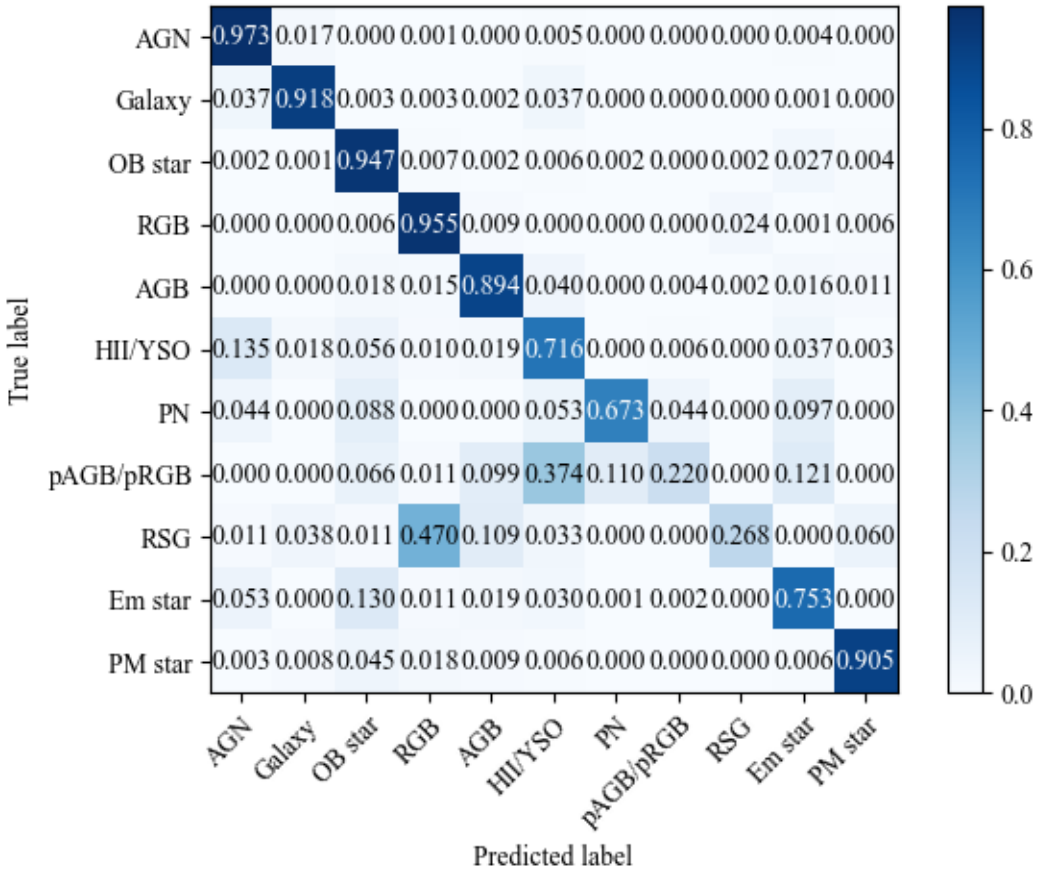

Figure 3.19: Confusion matrix of classifier trained 75% of the LMC dataset and tested on the remaining 25%.

stellar sources. Most confusion amongst the stellar sources tends to be with other classes of stellar sources.

3.4.5 Emission-line stars – help or hindrance

Emission-line stars tend to be a mixed bag of sources, and telling the difference between them and other sources such as OB stars, PNe and YSOs, is not an easy endeavour. The SMC classifier has only five available emission-line sources to train and test on,

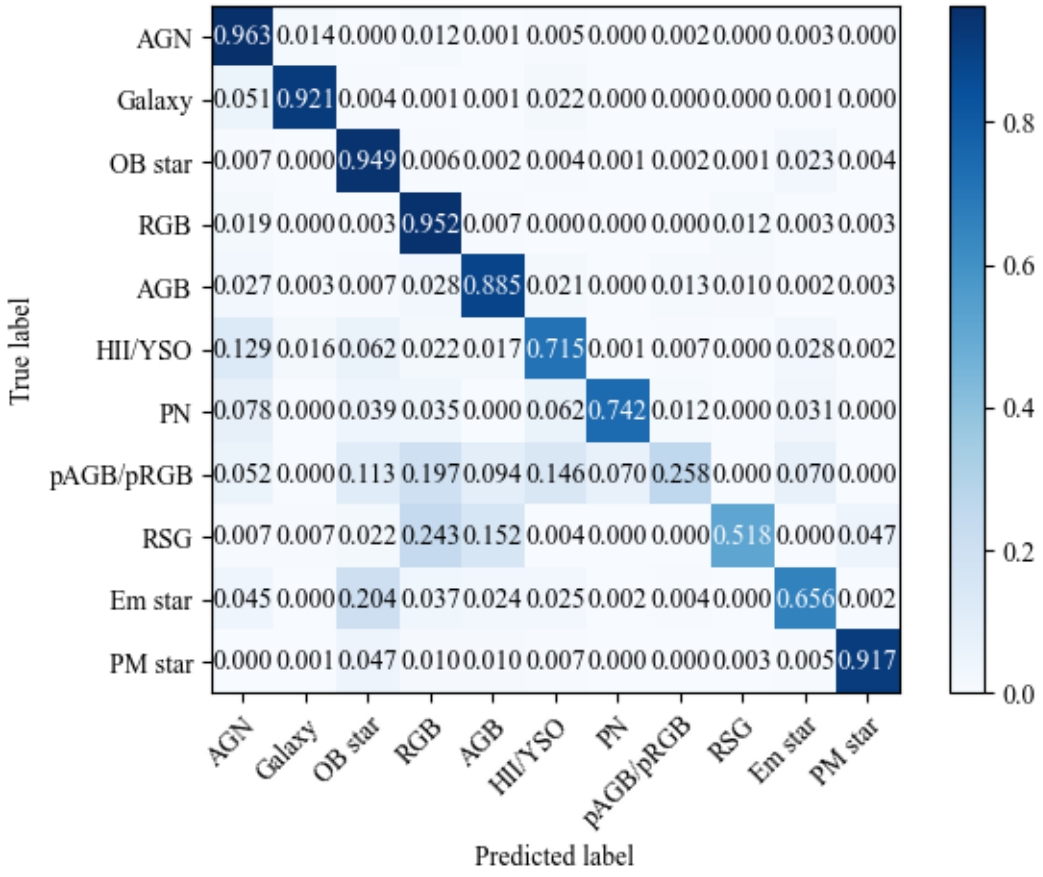

Figure 3.20: Confusion matrix of classifier trained 75% of the MC dataset and tested on the remaining 25%.

so it is no surprise that these sources are all misclassified. However, the rest of the classifications do not seem to suffer for it.

The majority of the emission-line stars are in the LMC dataset, so testing the effect of the removal of emission-line stars from the training set will be done with this dataset. I ran the classifier five times on the LMC dataset, for a dataset with and a dataset without emission-line stars included. Using the remainder of emission-line stars in the dataset after training (in the case of the classifier trained on the dataset without emission-line stars, this would be the full list) I then predicted the average

probability for each class for the all the emission-line stars, averaged over the five runs. This returned Figure 3.21. This shows that when the emission-star class is removed the emission-line stars tend to be classified as OB stars, with only a slight average class probability increase for HII/YSO, PNe and AGN classifications.

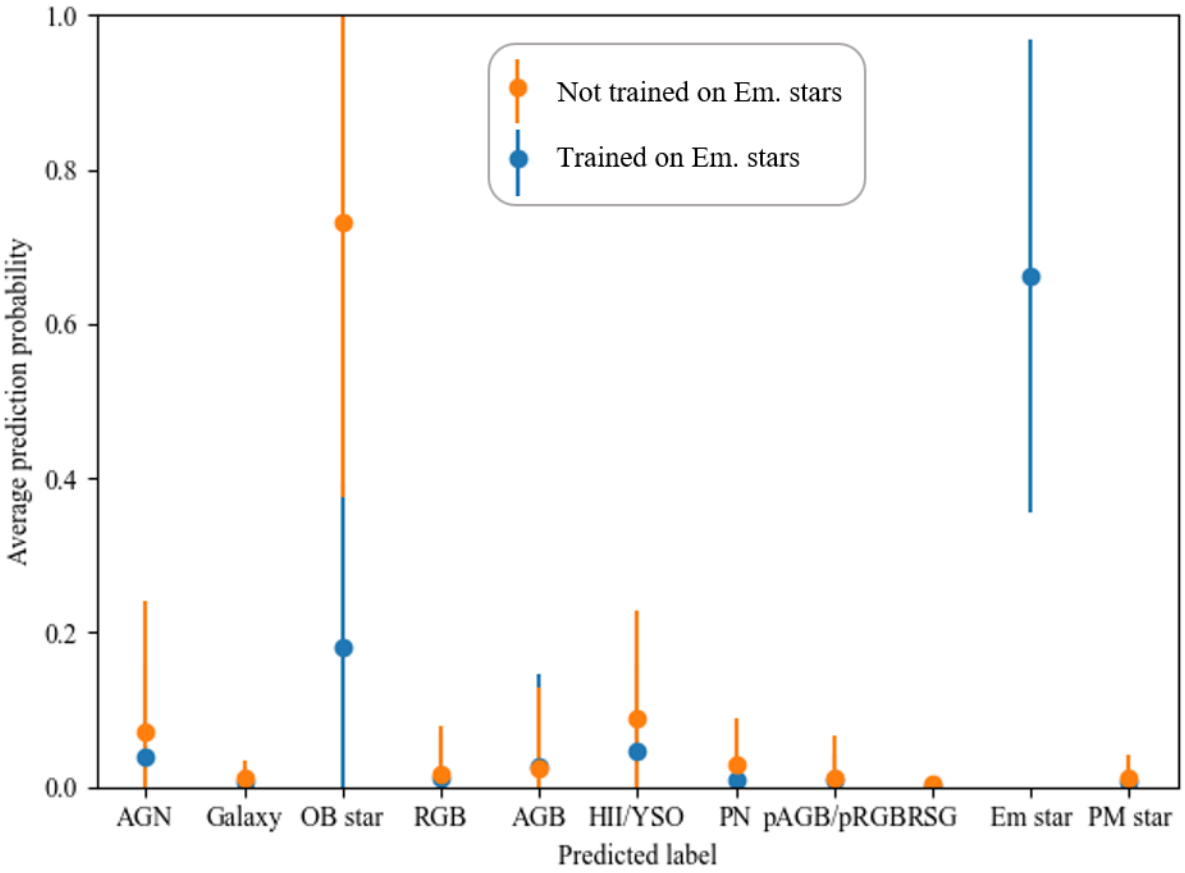

Figure 3.21: Average prediction probability per class of the LMC-trained classifier used on emission-line stars. Blue shows the prediction probabilities of a classifier trained on a dataset that included emission-line stars, whilst orange shows the prediction probabilities of a classifier trained on a dataset that did not include the emission-line stars. This shows that on average, removing emission-line stars does not affect the average probability of these sources predicted as AGN by much.

However, for a source to be classified as a particular class, the probability only

has to be the highest, not above a certain probability threshold. Across the 5 runs for the classifier not trained on emission-line stars there were only 24 out of 562 (emission-line star dataset multiplied by 5) sources that had an AGN class probability > 0.5. Therefore, setting a prediction probability threshold for a source to be classified as a specific class will most likely remove most misclassifications. Therefore, emission-line stars can be removed as they will most likely be predicted to be a stellar class, and among those that are not, the majority can be removed with a prediction probability threshold.

3.4.5.1 Testing classifier on the other Cloud

The SMC-trained classifier was tested on the LMC data, and vice versa, to determine if the sources in one region can be used to ascertain the classification of the sources in the other region. This was done by training the classifier on the entire dataset of one Cloud and then testing on the entire dataset of the other Cloud. This was repeated five times for each cloud, and then the average values of the confusion matrix were then calculated. The confusion matrices can be found in Figures 3.22 and 3.23 for the SMC-trained classifier tested on the LMC data and the LMC-trained classifier tested on the SMC data, respectively.

The overall accuracy of the SMC-trained classifier used on the LMC dataset was $\sim 0.79 \pm 0.10$ and the overall accuracy of the LMC-trained classifier used on the SMC dataset was $\sim 0.68 \pm 0.10$. This follows the trend of the SMC-trained classifier being more accurate than the LMC-trained classifier from Figure 3.17. AGN and galaxies have similar recall for both classifiers, no matter what dataset it is tested upon, which shows that the extragalactic sources can be combined together to be trained upon for both regions. The stellar sources however are not classified as well as they are by the classifiers trained on the same data, most likely due to the different stellar populations of the LMC and SMC.

Note that when emission-line stars were included, the accuracy of the SMC-trained classifier tested on the LMC dataset dropped to ~ 0.69 and the accuracy of

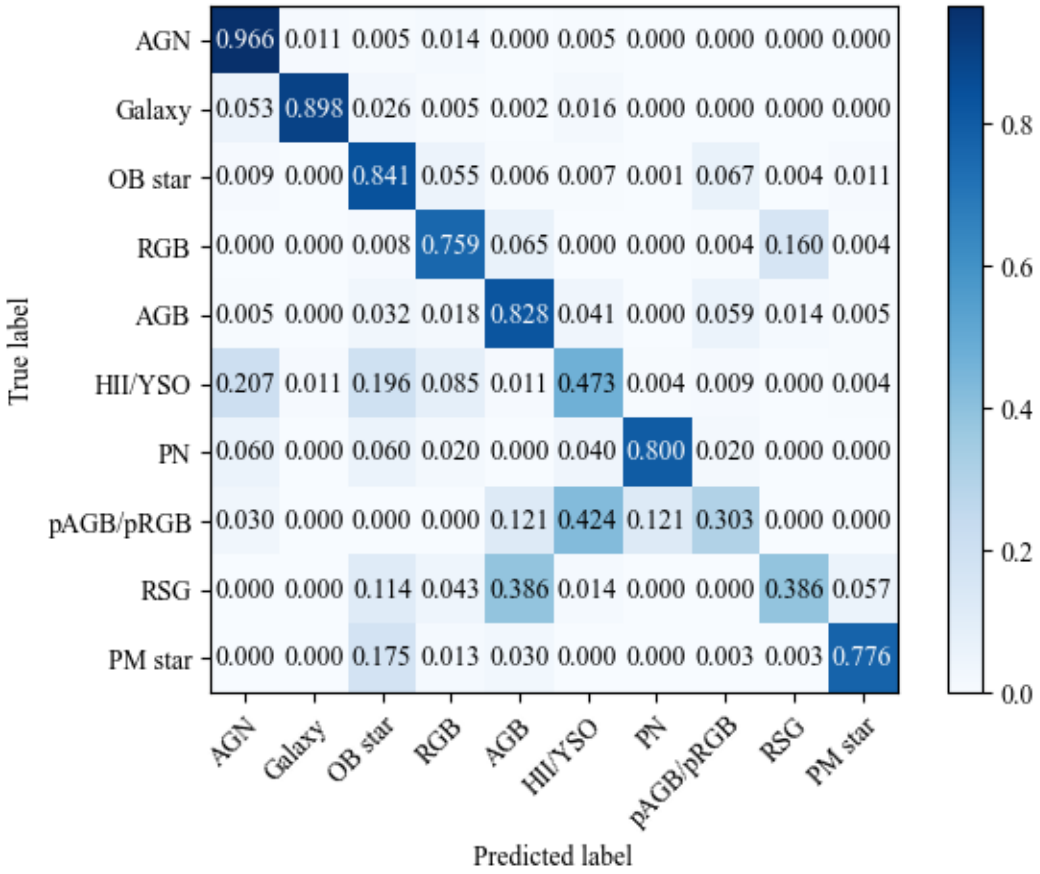

Figure 3.22: Confusion matrix of classifier trained on entire SMC dataset and tested on the entire LMC dataset.

the LMC-trained classifier tested on the SMC dataset dropped to ~ 0.52. This shows that removing emission-line stars was the correct choice overall.

3.4.5.2 Balanced vs imbalanced datasets

A balanced dataset would have all classes roughly equal in size. An imbalanced dataset would have large differences between class sample sizes. This unbalance can cause a poor predictive performance for the minority classes as most machine learning algo-

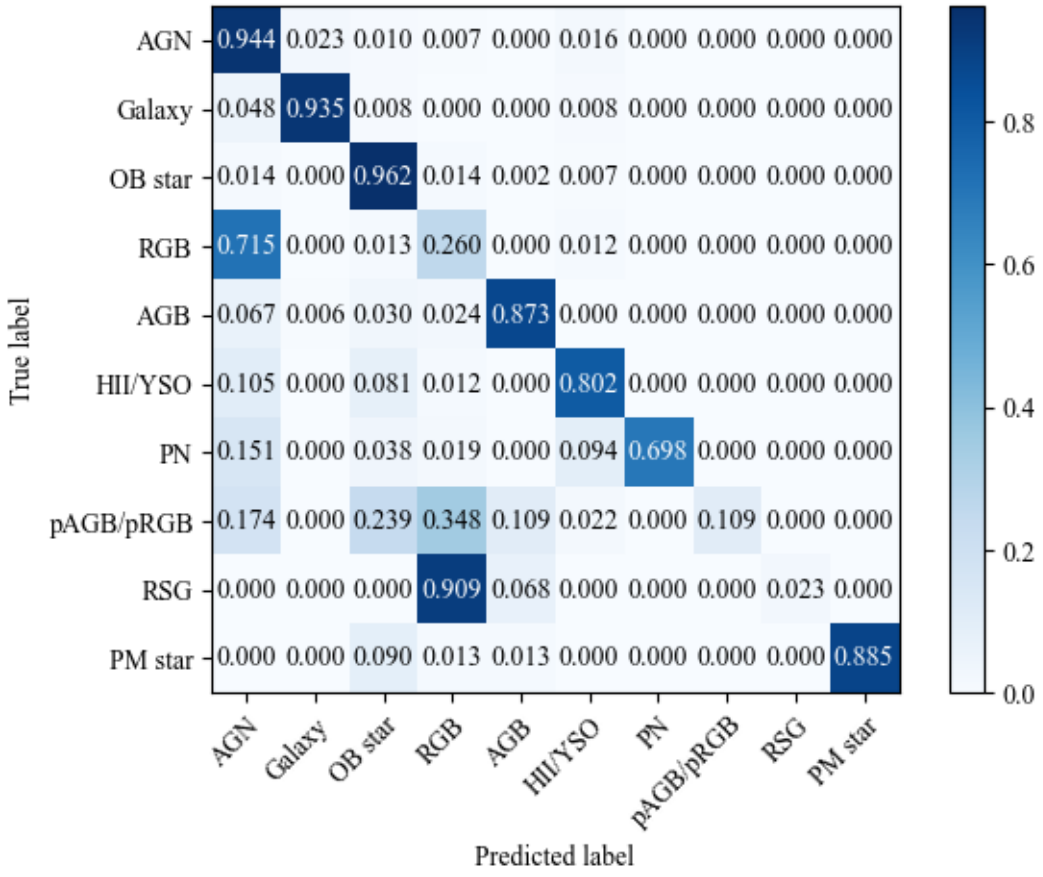

Figure 3.23: Confusion matrix of classifier trained on entire LMC dataset and tested on the entire SMC dataset.

rithms operate under the assumption of an equal sample size for each class.

Ensemble methods such as random forests can mitigate the effects of imbalanced datasets by training each tree on an independently randomly selected subset of the training set and then combining the results of all the trees together. However, for extremely imbalanced datasets, when randomly selecting a subset to train a tree on, if a minority class is too small then only a few or even none at all of the minority class may be selected for a particular tree, meaning there will be trees that have not seen the minority class at all, so will not know how to classify them. This effect can be

mitigated by balancing the dataset.

Balancing the dataset can either be done by downsampling, which reduces all the class sizes to the smallest class size, or upsampling, which increases/augments the minority class with synthetic data so that all the class sizes are the same as the largest class size. Downsampling, however, comes with the caveat of potentially losing important information and therefore will not be used here. Upsampling maintains the same amount of information, so this is what I used.

Upsampling can be done in one of two ways. Either by using machine learning on the minority class to generate synthetic data points based on the real data of the minority class sample; or by randomly copying objects from the minority class sample to increase the sample size. The latter method was used as it maintains that only real data is used whilst balancing the dataset so that each tree will randomly sample sources from each class. To do this the 'resample' function of Python's Scikit-learn module (Pedregosa et al., 2011) was used to upsample all the class samples to the same size as the majority class.

The SMC and LMC training samples were upsampled to the size of the AGN and OB classes, respectively. This was only done after the training sample was split into training and test sets, and only on the training set. This was so that the same objects did not end up in both the training and test set. This was done three times and then averaged. This produced confusion matrices for the SMC and LMC that can be seen in Figures 3.24 and 3.25, respectively. This yielded overall accuracies of 0.89 ± 0.01 and 0.87 ± 0.01 for the SMC and LMC, respectively, averaged over three runs.

Comparing the confusion matrices for the classifiers trained with the upsampled training sets with the classifiers trained on training sets that have not been upsampled (see Figures 3.18 and 3.19) it can be seen that the misclassifications are slightly reduced, and correct classifications improved.

Note that an imbalance in the dataset can also be caused by bias within the classes themselves. In this case, it would be selection bias of the sources being bright enough in the optical to be spectroscopically observed. One way this could be improved upon is by finding suitable, brighter targets in the Milky Way for some classes, for example,

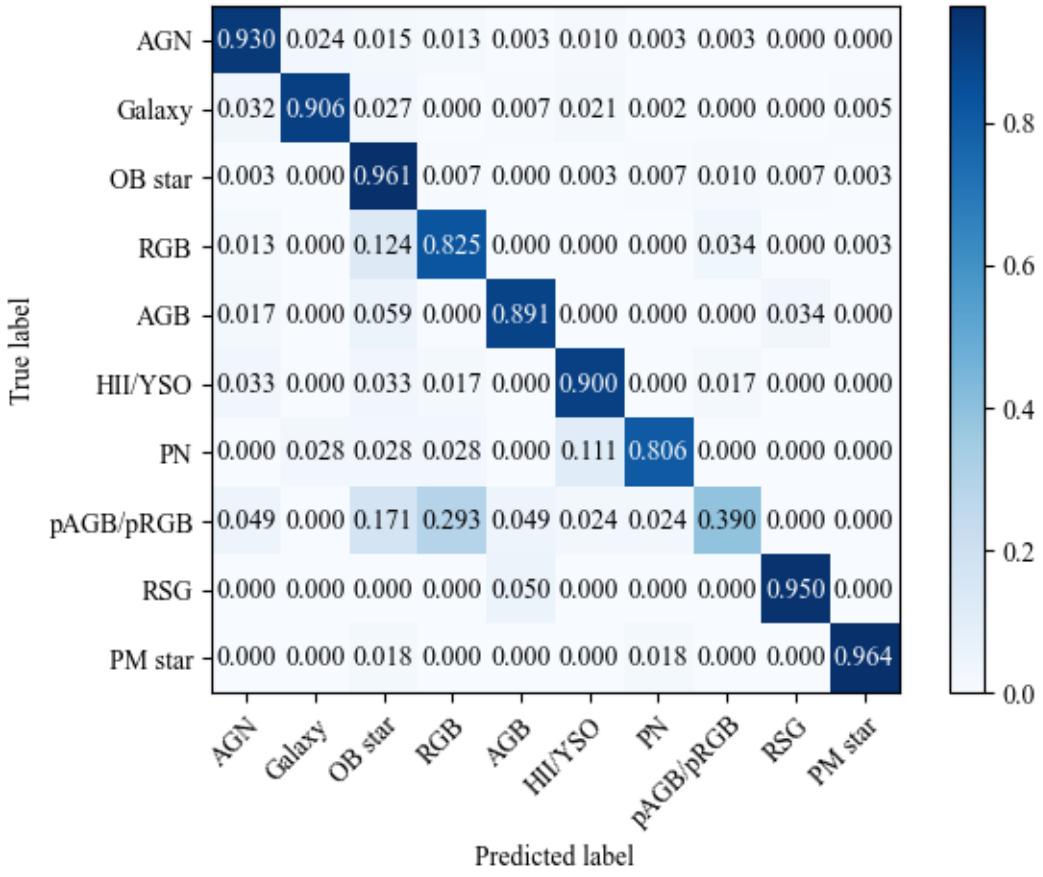

Figure 3.24: Confusion matrix of the classifier trained and tested on the SMC dataset, where each of the class training sets have been upsampled to the size of the class with the largest size. This dataset configuration includes extragalactic sources from both the LMC and SMC, but stellar sources only from the SMC.

RGB stars of SMC metallicity in the globular cluster 47 Tucanae.

3.4.5.3 Final data configuration

The final dataset configuration to be trained upon was for separate classifiers for the LMC and SMC, where both datasets will share extragalactic sources from both regions, whilst keeping stellar sources specific to the Cloud they are from. The PRF has 100

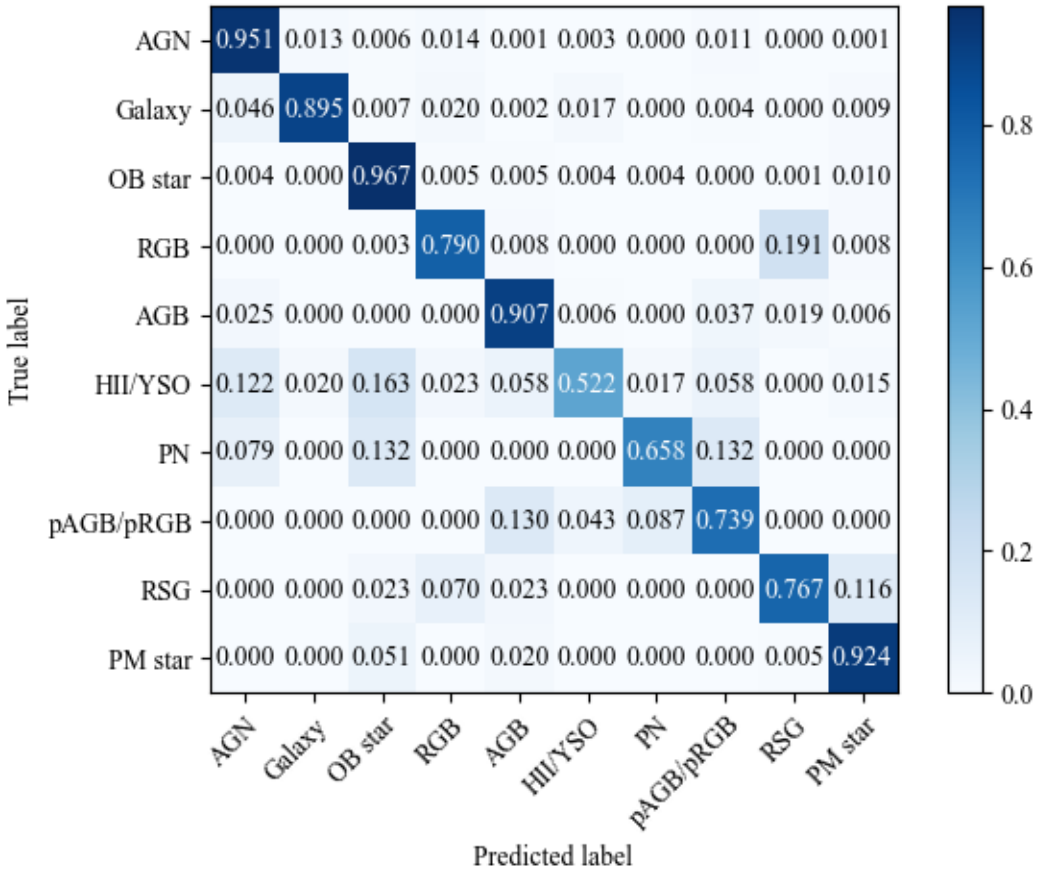

Figure 3.25: Confusion matrix of the classifier trained and tested on the LMC dataset, where each of the class training sets have been upsampled to the size of the class with the largest size. This dataset configuration includes extragalactic sources from both the LMC and SMC, but stellar sources only from the LMC.

trees. The classifier was trained and tested on the dataset 10 times to create 10 confusion matrices. The values were then averaged to create an 'average' confusion matrix. The confusion matrix for the SMC-trained classifier tested on SMC data can be found in Figure 3.26, and the confusion matrix for the LMC-trained classifier tested on LMC data can be found in Figure 3.27.

The overall accuracy of the SMC classifier is found to be 0.90 $\pm$ 0.01, and the overall accuracy of the LMC classifier is found to be 0.87 $\pm$ 0.01. For both classifiers

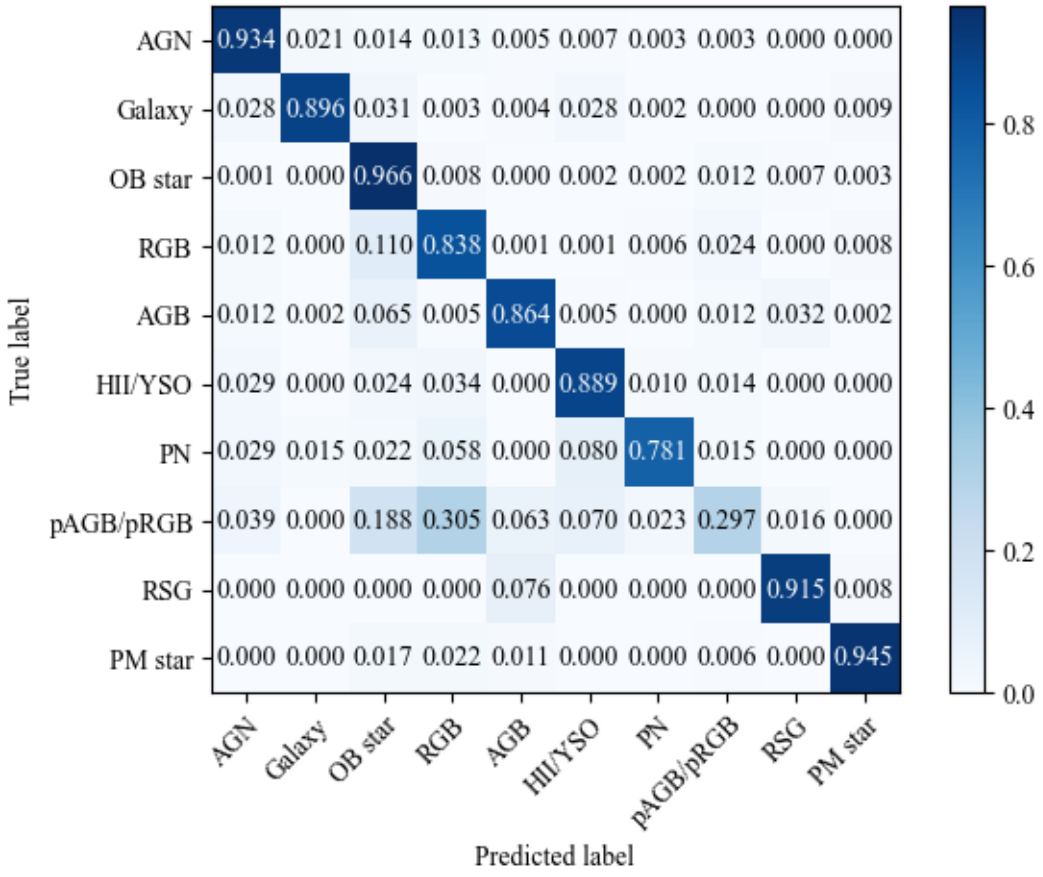

Figure 3.26: Confusion matrix of final classifier trained and tested on the SMC dataset. This dataset configuration includes extragalactic sources from both the LMC and SMC, but stellar sources only from the SMC.

the AGN class has one of the highest recall, $\sim 93\%$ and $\sim 95\%$ of all AGN in the test set are classified correctly for the SMC and LMC, respectively. For the AGN misclassified as other sources, they are most often misclassified as galaxies, which is not unexpected, and a smaller fraction are misclassified as RGB stars, OB stars, post-RGB/AGB stars or HII/YSOs. The precision of the AGN class is not as great, as other sources are misclassified as AGN. For the LMC $\sim 12.4\%$ of HII/YSOSs and $\sim 6.7\%$ of PNe are misclassified as AGN. Whereas for the SMC, 5% of all sources are misclasified as

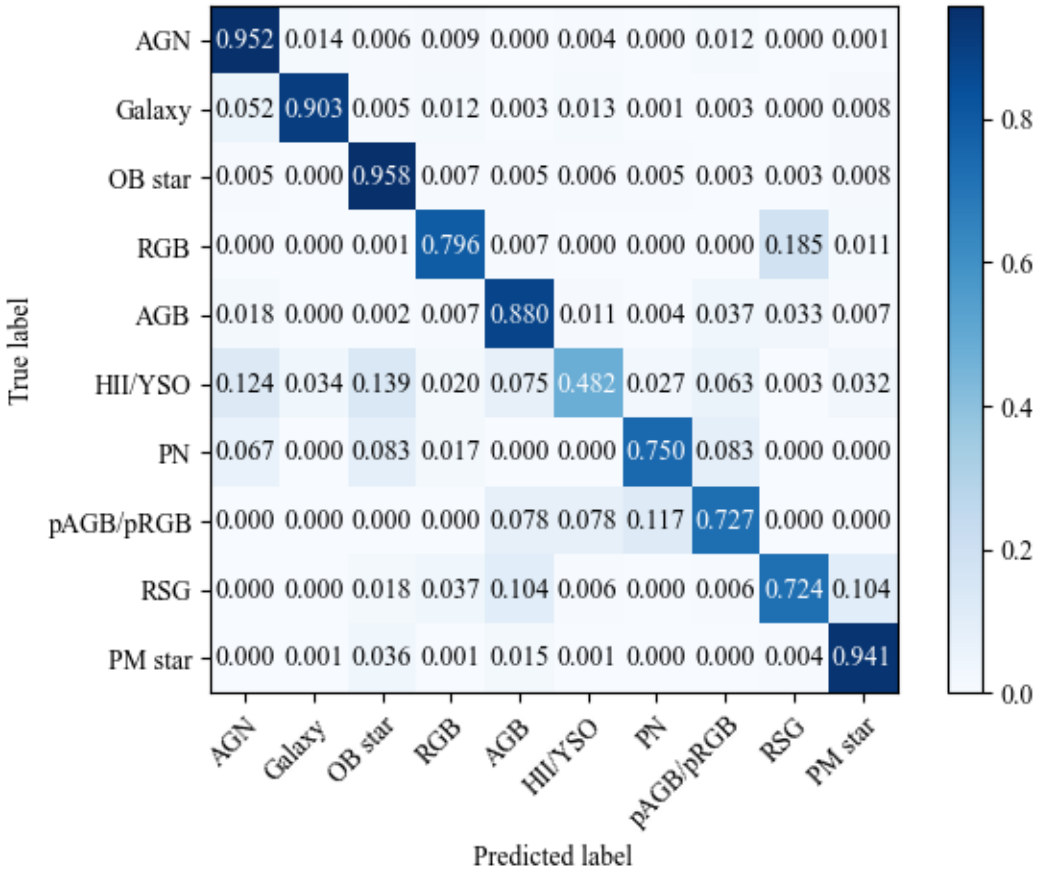

Figure 3.27: Confusion matrix of final classifier trained and tested on the LMC dataset. This dataset configuration includes extragalactic sources from both the LMC and SMC, but stellar sources only from the LMC.

AGN. Overall, the classifiers will get most of the AGN correct, but there will be some interlopers.

For the SMC classifier, the recall of the post-AGB/RGB is the worst, and for the LMC classifier the HII/YSO sources have the worst recall. For the SMC, post-AGB/RGB stars are mostly misclassified as other stellar sources, with only ∼4% misclassified as AGN. For the LMC-trained classifier, however, ∼12.4% of HII/YSO sources are misclassified as AGN. It is not surprising that post-AGB/RGB have the lowest re-

call, since they have one of the lowest number of sources, and therefore do not represent all the possible variations, to train on: 46 and 33 (trained on 75% of this) for the SMC and LMC, respectively. What is unexpected is that the LMC classifiers recall of RSG stars is lower than the SMC classifiers, since the LMC has 70 RSGs to train on compared to the SMC's 44. A similar trend can be seen for the recall of HII/YSO, where the total number of LMC sources is 459 compared to the SMC's 86, and yet the LMC has a recall of ∼48.2% for these sources whilst the SMC has a recall of ∼88.9%. This could potentially be due to a wider variety of sub-classes in a larger pool of sources for the LMC compared to a smaller more easily defined group of sources for the SMC.

The misclassification of stellar sources as AGN, and AGN as stellar sources, reflects what has been found anecdotally in the Magellanic Clouds. Classifications based on photometry have led to stars masquerading as AGN and vice versa in the Magellanic Clouds, such as SAGE0536AGN (Hony et al., 2011) and SAGE0534AGN (see Chapter 5), two AGN which were first thought to be evolved stars in the LMC, and Source 5 and Source 8 from the study of a small sample of AGN in Chapter 5, which were revealed to be stars in the SMC instead. These sources are within the datasets to be trained upon, which increases the likelihood the PRF would classify similar sources correctly, but their small number could lead similar sources to be misclassified instead.

Overfitting of a machine learning model can generally be spotted by using the classifier on the training set, and if the performance is much better than on the test set, then the model is overfitting. For the SMC classifier used on the SMC training set, the average accuracy was 0.92 ± 0.01, and for the LMC classifier used on the LMC training set the average accuracy was 0.88 ± 0.01. Both classifiers only performed marginally better on the training sets compared to the test sets (0.90 ± 0.01 for the SMC and 0.87 ± 0.01 for the LMC), meaning the machine learning model is not overfitting.

3.4.5.4 Feature importance

The PRF algorithm can calculate overall feature importance for the entire classifier. This level of importance is calculated as "mean decrease impurity", which is defined as

the total decrease in node impurity (weighted by the probability of reaching that node, which is approximated by the proportion of samples reaching that node), averaged over all trees over the ensemble (Breiman, 2001). In other words, how well does each feature separate the sample into the expected classes (the decrease in class impurity). The values of the feature importances are then normalised, such that they all sum to one.

The classifiers are trained on the full datasets for SMC and LMC, from which the feature importances were calculated. This was then repeated 10 times and the feature importances were then averaged for each feature. The top 15 ranked feature importances for both the SMC and LMC classifiers can be seen in Figure 3.28, the full list can be seen in Appendix C. For both the SMC and LMC classifiers the top 15, whilst in a different order, only have one feature difference between them.

VMC photometry and colours rank high amongst the feature importance, most likely due to all sources having at least one observation in the $YJK_{\rm s}$ bands. However, it is unlikely that is the sole reason for their high importance, so therefore the colours, photometry and sharpness values are providing good distinction between the sources as well.

Because all of the sources used for training have classifications based on spectroscopy, this means that they tend to be bright enough to be observed with Gaia, hence why the Gaia colours and photometry rank quite highly. SMASH, despite covering a similar wavelength range and increased depth compared to Gaia is not ranked as highly, which could be due to the lack of coverage for some parts of the VMC footprint (see Section 2.1.3). However, despite this high dependence on Gaia photometry and colours, the proper-motions in RA and DEC do not rank nearly as high, at ranks 117 (0.0012 ± 0.0002) and 64 (0.0031 ± 0.0007) for proper-motion in RA for the SMC and LMC, respectively, and at ranks 58 (0.0036 ± 0.0014) and 79 (0.0023 ± 0.0002) for proper-motion in DEC for the SMC and LMC, respectively. This is unexpected since proper-motions would be the most obvious way of separating the high proper-motion stars and extragalactic sources from the Magellanic stellar sources.

The worst features are most likely due to an abundance of missing values for

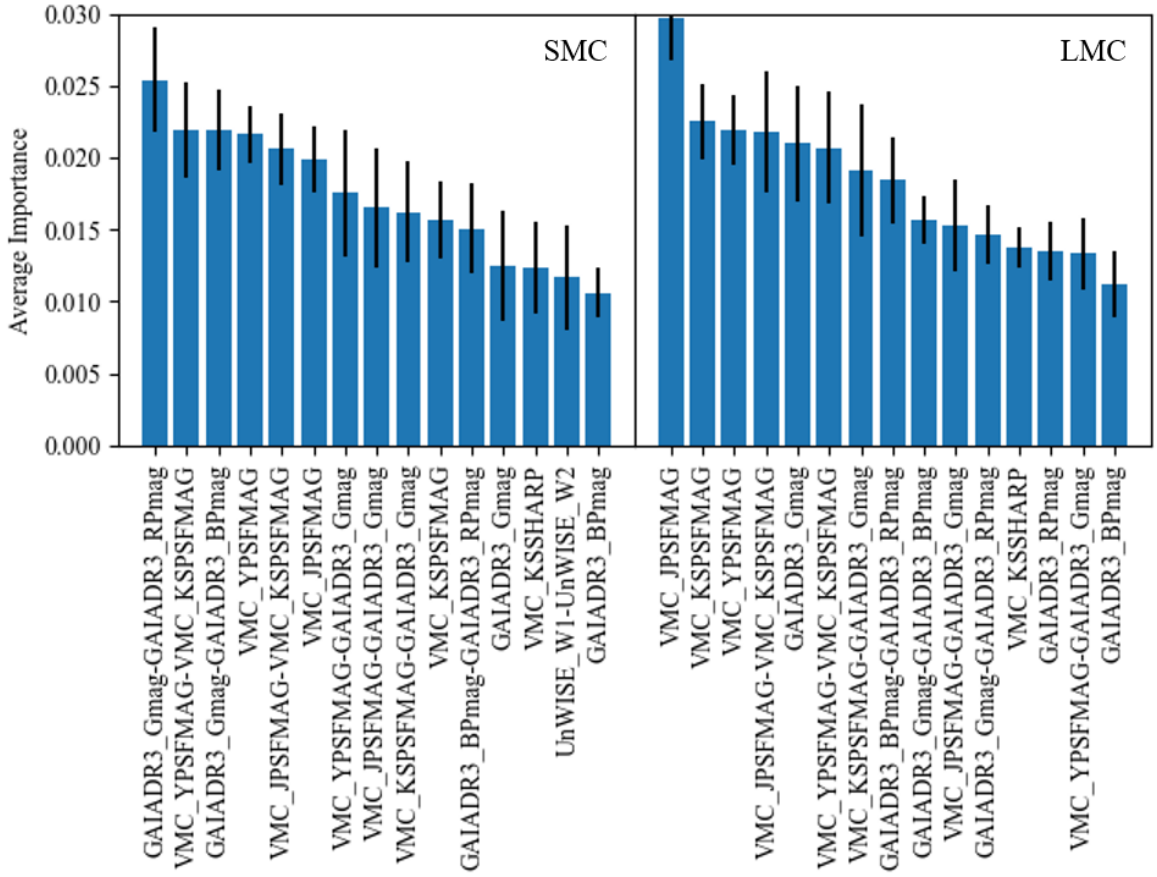

Figure 3.28: Top 15 important features for the SMC (left) and LMC (right) classifier. Note that, while the top 15 features are in a different order, they are almost all the same features for both the SMC and LMC classifier.

these features, brought upon by either a lack of coverage in certain areas and/or a lack of depth, such as for colours, SMASH u − any AllWISE photometry. Leaving these features in should not affect the accuracy of the classifiers as they have been deemed unimportant, and therefore unlikely to be relied upon to make a classification.

4 The extragalactic radio sky behind the LMC – a new deep and detailed map from ASKAP

This chapter describes the sources detected with the AEGEAN source finder algorithm in the EMU ASKAP radio image of the LMC. Section 4.1 briefly describes the foreground and Magellanic sources that obscure the extragalactic sources from our view and how to differentiate between the stellar and the extragalactic. Section 4.2 looks at the sources behind the LMC, the galaxies and AGN, with comparisons between the extended radio emission in ASKAP and other radio surveys taken with other instruments, as well looking at the properties of the spectroscopically confirmed AGN and their spectral indices.

The work in this chapter is published as Pennock et al. (2021) including additional analysis of non-background sources.

4.1 Magellanic and Galactic sources

I discuss a few types of Magellanic and Galactic sources in the appendices (Section D). That work was done mostly by/with scientific collaborators, the full list of which can be seen in the published paper.

These Magellanic and Galactic sources were found by cross-matching with various catalogues using TOPCAT[1] (Taylor, 2005) with an initial search radius of 10″ (to allow for inaccuracies in both the ASKAP and literature positions), after which visual inspection was used to determine whether they are true counterparts. All quoted flux densities from the comparison catalogues are integrated flux densities unless stated otherwise.

[1] http://www.star.bris.ac.uk/∼mbt/topcat/

Examples of radio sources in the LMC are planetary nebulæ (PNe), young stellar objects (YSOs), supernova remnants, neutron stars, emission-line stars and star-forming regions, which, when emitting sufficiently bright and extended radio emission (see Figures in Section 2.1.2.2 and Appendix D), can obscure the extragalactic sources behind them, and partially obscure the sources near them.

These stellar sources can also appear as point-like (unresolved) sources, which can be mistaken for AGN. One way of separating the AGN from the stellar radio sources would be to look at their spectral indices, α, which is defined as $F_\nu \propto \nu^\alpha$, where F_ν is the integrated flux density at frequency ν. A flatter spectral index close to zero indicates free–free emission, and a steep negative spectral index, ~ -0.7, indicates synchrotron emission. Stellar sources are more likely to show free–free emission and AGN are more likely to emit synchrotron emission. In order to use this to differentiate all the point sources in the ASKAP survey this would require another survey of the LMC at a different frequency with a similar (or better) depth/resolution. An example of a future survey that will allow for measurement of spectral indices is the MeerKAT (Jonas, 2009) survey of the LMC, which has already been observed and is currently being processed for the LMC, which uses a much wider band which allows for in-band spectral index measurements.

Another way to differentiate between stellar and AGN radio sources is to cross-match with surveys at different wavelengths, such as the IR and optical. Not all sources will have a counterpart, due to being too faint at other wavelengths, which could point to a distant QSO, where the radio jet is pointing directly towards us. For those with optical counterparts in the Gaia DR3 survey, proper-motions in RA and DEC that indicate a foreground or LMC source can remove the most obvious stellar sources. However, there could be cases of stellar sources having a chance superposition in front of the true extragalactic radio source, though these also could be potentially interesting discoveries of rare radio stars (e.g. Wang et al., 2022). For those with IR counterparts colour–colour and colour–magnitude diagrams can differentiate the most obvious sources (e.g. Lacy et al., 2004; Stern et al., 2005b; Mateos et al., 2012; Donley et al., 2012; Cioni et al., 2013), however confusion can still be found. For an example

of a sample of high confusion sources see Chapter 5.

However, these simple assumptions used to differentiate sources, even about spectral index, may also lead to rejection of background sources. Hence, I am looking at machine learning (next chapters) as well as spectroscopic followup of interesting cases where feasible.

4.2 Background sources

Here I discuss the extragalactic sources seen behind the LMC, which dominate the source counts. The radio emission from these sources is due to synchrotron radiation from relativistic electrons in magnetic fields and free–free emission from H II regions.

4.2.1 Radio galaxies

The ASKAP LMC image features some spectacular radio galaxies, showcasing a clearer morphology of these sources than has been seen before. Figure 4.1 shows a few examples of these radio systems.

Figure 4.1, (a). The NW quadrant shows a Fanaroff–Riley II (FR II) system (Fanaroff & Riley, 1974), which are luminous radio systems brighter near their extremities than near their centres. This is a known radio source, PMN J0449−7219 (Filipovic et al., 1998c) and is seen as three islands (islands 14012, 13945 and 13724) in our ASKAP catalogue. The jets originate from galaxy WISEA J044917.77−721843.7 (a.k.a. 2MASS-6X J04491744−7218434, VISTA J044917.454−721843.85) for which Bilicki et al. (2016) give an extinction-corrected R-band magnitude of 19.62 and a photometric redshift of 0.47. With a radio size of 5′.5, this makes it a giant radio galaxy of projected linear size of ∼1.9 Mpc. A similar source in the SE corner of the image is most likely hosted by the galaxy VISTA J045013.036−722036.25 ($K_{\rm sp}$ = 15.45 mag, a.k.a. WISEA J045013.00−722036.0, 2MASS-6X J04501307−7220362), for which Bilicki et al. (2016) give an extinction-corrected R-band magnitude of 19.68 and a pho-

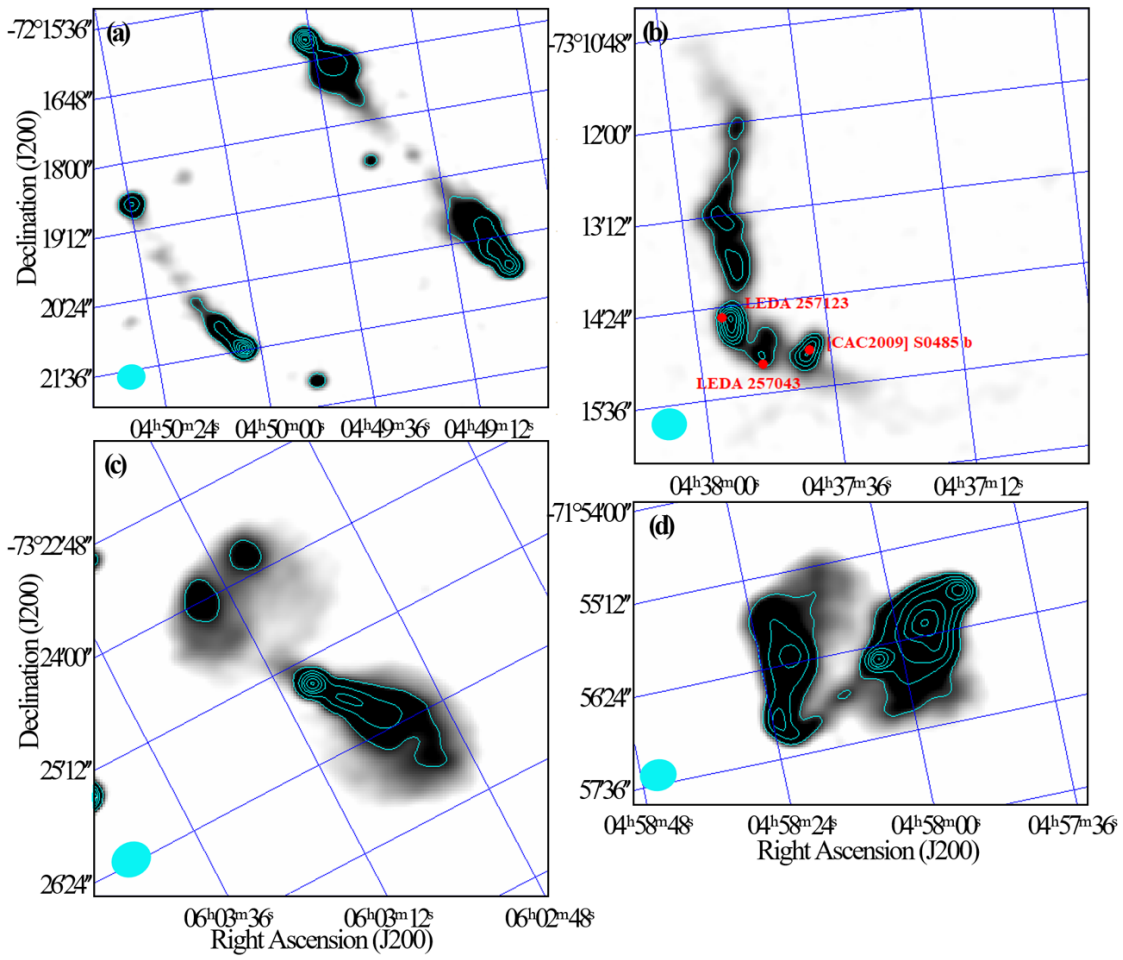

Figure 4.1: Four examples of radio emission from galaxies. The structure of the four radio sources is shown by the cyan contours that are drawn at equally spaced values in brightness. The beam size for ASKAP 888 MHz is $\sim 13\farcs9 \times 12\farcs1$, which is represented by the cyan ellipse. (a) Two examples of jets originating from a less bright central source (FR II systems) where contours span from 0.2 – 29 mJy beam^{-1}. (b) Galaxy cluster at $z \sim 0.06$. Low surface brightness at lower right is 0.25 – 0.45 mJy beam^{-1}. The contours span from 0.2 – 19 mJy beam^{-1}. (c) A jetted system that reveals interesting morphology in the jets. Contours span from 3.7 – 28 mJy beam^{-1}. (d) A central radio source from which two jets originate. These jets are bent as they come into contact with the intra-cluster medium. The contours span from 3 – 19 mJy beam^{-1}.

tometric redshift of 0.35. With a radio size of 3′.4, this makes it another giant radio galaxy of projected linear size of ∼1.0 Mpc. Whilst the radio emission from the ends of the jets can be seen in the SUMSS 843 MHz image, the fainter, central part of this source, where the radio jets most likely originate, has not been seen before.

Figure 4.1, (b). This is likely a wide-angle tailed (WAT) radio galaxy hosted by LEDA 257123 (EMU EC J043753.4−731437), one of the members of galaxy cluster Abell S0485 with a mean spectroscopic redshift of 0.061 based on seven galaxies from the 6dF redshift survey (Jones et al., 2009). Neither the host galaxy of the WAT nor LEDA 257043 (EMU EC J043740.1−731508) or [CAC2009] S0485 b (a.k.a. 2MASX J04373976−7315056), apparently superposed on the SW tail of this WAT, have a spectroscopic redshift, and the host galaxy LEDA 257123 is not among the three brightest galaxies (Coziol et al., 2009). Assuming the cluster redshift for the host, the radio size of the WAT of 6′.2 corresponds to a projected linear size of 440 kpc.

Figure 4.1, (c). ASKAP 888-MHz image reveals that this known radio source (PMN J0603−7325) has an S-shaped morphology indicating precessing jets. The strong radio core is coincident with the galaxy 2MASX J06030694−732529 (EMU EC J060306.4 −732529) for which Bilicki et al. (2016) give a corrected R-band magnitude of 16.32 and a photometric redshift of 0.156. With an angular extent of 3′.5 this corresponds to a projected linear size of 570 kpc.

Figure 4.1, (d). This radio source (PKS 0458−720, PMN J0458−7156, EMU EC J045804.2−715635) shows an even more extreme S-shape likely due to projection along the line of sight. Its radio core coincides with 2MASX J04580433−7156347 (a.k.a. SMSS J045804.30−715635.0, $r_{\rm Petro}$ = 17.07 mag, Wolf et al., 2018) , but no (photometric) redshift is available.

These galaxies provide examples of the potential for exploring the morphologies in greater detail than previously, and for locating the host galaxies of these extragalactic marvels. I provide more in-depth analysis using high-resolution IR images and optical spectroscopy in Chapter 6 and a followup paper (Pennock et al., in prep.).

4.2.2 Comparison with the GLEAM 4-Jy sample

The GaLactic and Extragalactic All-sky MWA 4-Jy sample (G4Jy; White et al., 2020a,b), from the Murchison Widefield Array (MWA), provides flux-densities across the range 72 to 231 MHz at DEC $< 30°$). G4Jy is a complete sample of the brightest radio sources, $F > 4$ Jy, which are expected to be mainly AGN with powerful radio-jets (strong optically-thin synchrotron emission). Six of the sources from this sample lie within the ASKAP LMC field, all of which were detected in the ASKAP observations. These are shown in Figure 4.2 and show the large improvement the ASKAP image offers in resolution, which allows the identification of the optical/infrared counterpart of these radio sources, and while almost all host galaxies are identified in the WISE image, some are very faint. These GLEAM sources, which appear at first glance to be one source, are revealed to be multiple sources by ASKAP. The majority of these sources have the appearance of FR II sources. For the sources G4Jy 482, G4Jy 553 and G4Jy 596, ASKAP shows that they are double sources not resolved in SUMSS.

Figure 4.2, top-left. Contours of G4Jy 595 (a.k.a. GLEAM J060849−655110) from ASKAP reveal that the tear drop shaped radio source seen in GLEAM is in fact two sources. The southern source is an extreme bent tailed source hosted by the galaxy 2MASX J06085498−6552559 (a.k.a. WISEA J060854.94−655255.6) with $z = 0.03752$ (Jones et al., 2009). The northern one is a radio point source hosted by the galaxy ESO 086-G062 (a.k.a. WISEA J060852.00−654350.2), with $z = 0.037$ (Wegner et al., 2003), which appears to be part of a galaxy group that the bent tailed radio source has moved through. In fact, Saulder et al. (2016) lists the group of three galaxies 2MRS 3198 at $z = 0.03774$. The third member galaxy is 2MASX J06090657−6544539 (a.k.a. LEDA 310538) with $z = 0.03828$ (Huchra et al., 2012), which lies $\sim 2'$ SE of ESO 086-G062 and is also detected as a radio point source.

Figure 4.2, top-right. ASKAP reveals that G4Jy 453 (GLEAM J042358 −724601) comprises multiple sources. The western-most ASKAP radio source features an X-shaped radio galaxy known as PKS 0424−728 at a redshift of ∼0.19 (inferred from photometry; Subrahmanyan et al., 1996). With more recent photometric redshifts of

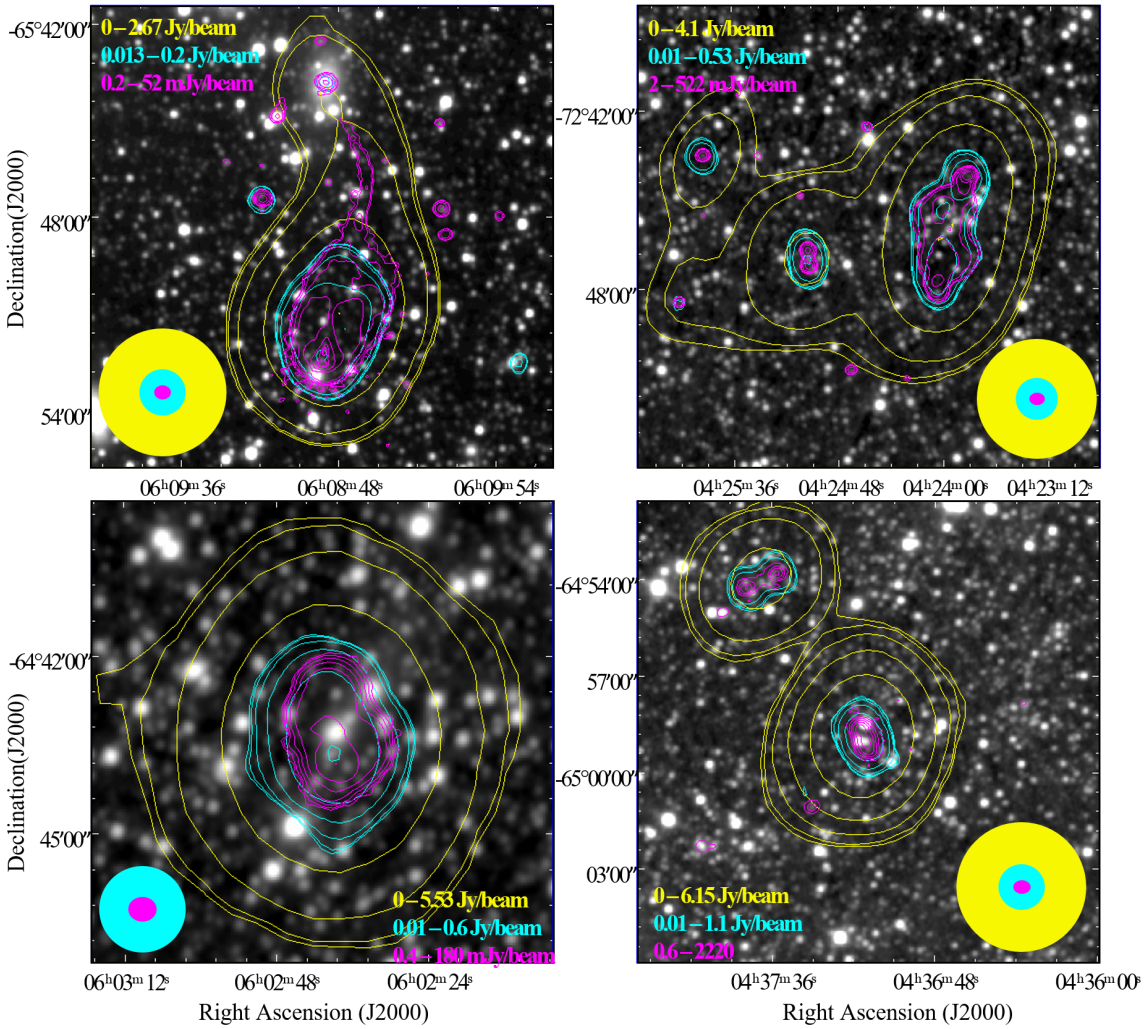

Figure 4.2: Comparison between ASKAP (magenta contours), SUMSS (cyan contours) and GLEAM (yellow contours) that represent logarithmically spaced brightness levels in the ASKAP image, the minimum and maximum of which are annotated in their respective colour. The background image is the WISE W3 band. The beam size for ASKAP 888 MHz is $\sim 13\farcs9 \times 12\farcs1$ and is represented by the magenta ellipse. The beam size for SUMSS 843 MHz is $\sim 43'' \times 43''$ and is represented by the cyan circle. The beam size for GLEAM 72 – 231 MHz is $\sim 2' \times 2'$ and is represented by the yellow circle. The sources are (top-left) G4Jy 595, GLEAM J060849−655110, (top-right) G4Jy 453, GLEAM J042358−724601, (lower-left) G4Jy 587, GLEAM J060239−644324, (lower-right) G4Jy 482, GLEAM J043709−645853.

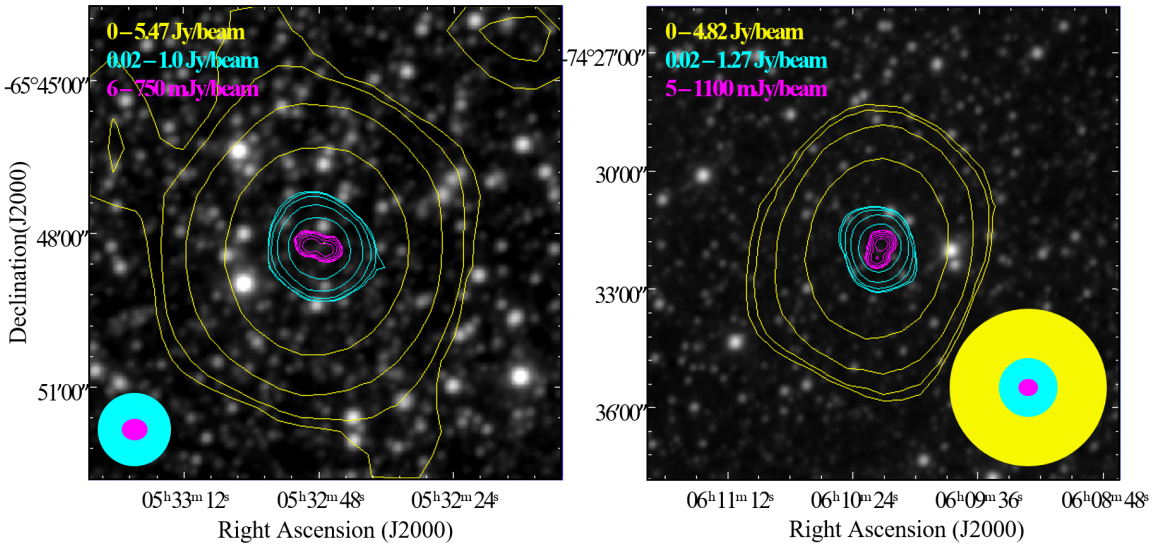

Figure 4.3: Comparison between ASKAP (magenta contours), SUMSS (cyan contours) and GLEAM (yellow contours) that represent logarithmically spaced brightness levels in the ASKAP image, the minimum and maximum of which are annotated in their respective colour. The background image is the WISE W3 band. The beam size for ASKAP 888 MHz is $\sim 13\farcs9 \times 12\farcs1$ and is represented by the magenta ellipse. The beam size for SUMSS 843 MHz is $\sim 43'' \times 43''$ and is represented by the cyan circle. The beam size for GLEAM 72 – 231 MHz is $\sim 2' \times 2'$ and is represented by the yellow circle. The sources are (left) G4Jy 553, GLEAM J053248−634820 and (right) G4Jy 596, GLEAM J061014−743159.

0.229 and 0.214 from Bilicki et al. (2014) and Bilicki et al. (2016) its angular size of 4.′3 gives it a projected linear size of 880 kpc. An unresolved source at 888 MHz is hosted by WISEA J042434.02−724241.3 with an even fainter 888 MHz source $\sim 24''$ SW with no optical/IR counterpart and the X-ray source 1RXS J042432.7−724253 in between them. The central source is revealed as a double radio source of size $\sim 33''$ hosted by WISEA J042501.60−724700.3 with no optical counterpart. Two ASKAP point sources at the E edge of the radio complex can be identified with WISEA J042549.29−724330.8 (NE) and WISEA J042600.47−724828.9 (SE).

Figure 4.2, lower-left. This is G4Jy 587 a.k.a. PKS 0602−647 or GLEAM J060239−644324. At the centre of this radio structure is a pair of galaxies, the brighter of which, LEDA 319866 with $z = 0.045$ (Burgess & Hunstead, 2006), is likely the host. Nearby galaxies with known redshifts show similar redshifts, such as the galaxy NW of the host that lies on the edge of the ASKAP contours, which is 2MASX J06023279−6442344 with $z \sim 0.048$ (photometric; Bilicki et al., 2014).

Figure 4.2, lower-right. The central source is G4Jy 482 (GLEAM J043709 −645853) and the NE source is GLEAM J043738−645400 (not in the G4Jy sample). The former is MRC 0436−650 a.k.a. PKS 0436−650 hosted by the galaxy DES J043708.39−645901.9 ($r = 17.97$ mag, $z_{\rm sp} = 0.360$, Burgess & Hunstead, 2006). The NE double source is hosted by the QSO candidate DES J043738.82−645402.1 ($r = 21.53$ mag, no $z_{\rm phot}$ available, Abbott et al., 2018) a.k.a. CWISE J043738.83−645402.0 (Marocco et al., 2021). The source $\sim 2.'7$ SE of G4Jy 482 is DES J043724.19−650105.6 a.k.a. WISEA J043724.19−650104.9 with high stellarity, i.e. likely a QSO.

Figure 4.3, left panel. The central source is G4Jy 553 (GLEAM J053248 −634820), also known as PKS 0532−638 and MRC 0532−638 or ATPMN J053246.7 −634820 (McConnell et al., 2012), hosted by the QSO candidate SMSS J053247.95 −634818.1 ($g_{\rm Petro} = 19.89$ mag, Wolf et al., 2018) a.k.a. WISEA J053247.87−634816.7 with typical QSO colors ($W12 \sim 1.0$ mag, W23 ~ 3.0 mag), suggesting a redshift of $z > 1$ (see e.g. Fig. 2 of Krogager et al., 2018).

Figure 4.3, right panel. This is G4Jy 596 (GLEAM J061014−743159, PKS 0611 −745), which was identified by White et al. (2020a) with WISEA J061013.90−743201.6

and is not detected in the optical.

4.2.3 IR properties of radio sources

The AllWISE catalogue (Cutri et al., 2013) is a well-explored all-sky survey with photometry from 3.35 to 22.1 μm. This wavelength regime allows for a selection of both obscured and unobscured AGN, which makes it a great catalogue to cross-match to discover extragalactic sources. Selection criteria for different types of AGN/galaxies are well documented (e.g. Wright et al., 2010; Jarrett et al., 2011; Mateos et al., 2012; Stern et al., 2012; Assef et al., 2013; Nikutta et al., 2014) and the criteria from Wright et al. (2010) will be used here to discuss the radio population.

Cross-matching the AllWISE catalogue with the new ASKAP catalogue with a search radius of 5″ and a maximum magnitude error in the W1 (3.4 μm), W2 (4.6 μm) and W3 (12 μm) bands of 0.2 mag yielded a total of 14,333 sources (combined Gold and Silver, which contain 7413 and 6920 sources, respectively). This search radius was used as the density of coincidences drops off sharply, approaching that of chance coincidences (tested by applying a large offset in RA and DEC and then cross-matching again) beyond a search radius of 5″, and coincidences found between 6 – 10″ of stellar sources (including PNe) almost invariably turned out to be implausible associations. As a comparison with ASKAP detected sources, the sources within the LMC field that had no ASKAP detection (no ASKAP emission above 5σ of the local RMS) within 5″ were considered. This amounted to 253,533 AllWISE non-ASKAP detected sources.

I plotted the ASKAP detections and non-detections on the W1 – W2 vs. W2 – W3 colour–colour diagram (Figure 4.4), based on Figure 12 of Wright et al. (2010). The majority of the radio detections are extragalactic sources. The stellar AGB sequence seen in the non-ASKAP detected diagram (right) is clearly not visible in the ASKAP detected diagram (left). The Bronze source list also shows a similar colour–colour diagram to the combined Gold and Silver source lists and therefore the majority of sources are extragalactic.

Dust emission from the LMC ISM can be detected at WISE wavelengths, as well

Table 4.1: Classifications of known AGN that have been detected with ASKAP and those that have not. Q = QSO, type-I broad-line core-dominated; A = AGN, type-I Seyfert/host-dominated; B = BL Lac object; N = narrow-line AGN (NLAGN), type-II Seyfert/host-dominated (includes unknown number of legacy narrow emission line galaxies (NELGs)/emission line galaxies (ELGs)/ low-ionization nuclear emission-line region (LINERs)); R = radio associated ; X = X-ray associated.

	Total	A	Q	N	B	R	X
Radio	202	67	117	5	1	12	48
Non-radio	453	80	368	5	0	0	110

as at radio frequencies, affecting what we observe of background extragalactic sources in these regions. Extinction caused by dust within the LMC would also affect the optical brightness of background AGN and galaxies. The vast majority of the ASKAP field is affected only by low or modest levels of such extinction ($A_V \ll 1$ mag), judging from extinction maps such as those produced by Skowron et al. (2020), which would not cause major difficulties for optical spectroscopy and certainly not lead to any bias between the extragalactic populations with and without such spectroscopy.

4.2.3.1 Spectroscopic AGN

There are 657 spectroscopically observed AGN, from the milliquas catalogue of Flesch (2019a,b), in the field of the ASKAP LMC image. The highest contributions are from Kozłowski et al. (2012, 2013), Geha et al. (2003), Esquej et al. (2013) and Ivanov et al. (2016), contributing 547, 24, 23 and 10 objects, respectively.

This list of spectroscopically observed AGN was cross-matched with the combined Gold and Silver ASKAP LMC source lists with a search radius of $5''$. This gave a total of 190 known AGN with radio detections, only 14 ($< 8\%$) of which had known radio detections prior to the present ASKAP observations. As expected, all known radio associated AGN are in the new ASKAP catalogue. The classifications of the known AGN that were radio detected and non-radio detected are described in Table 4.1.

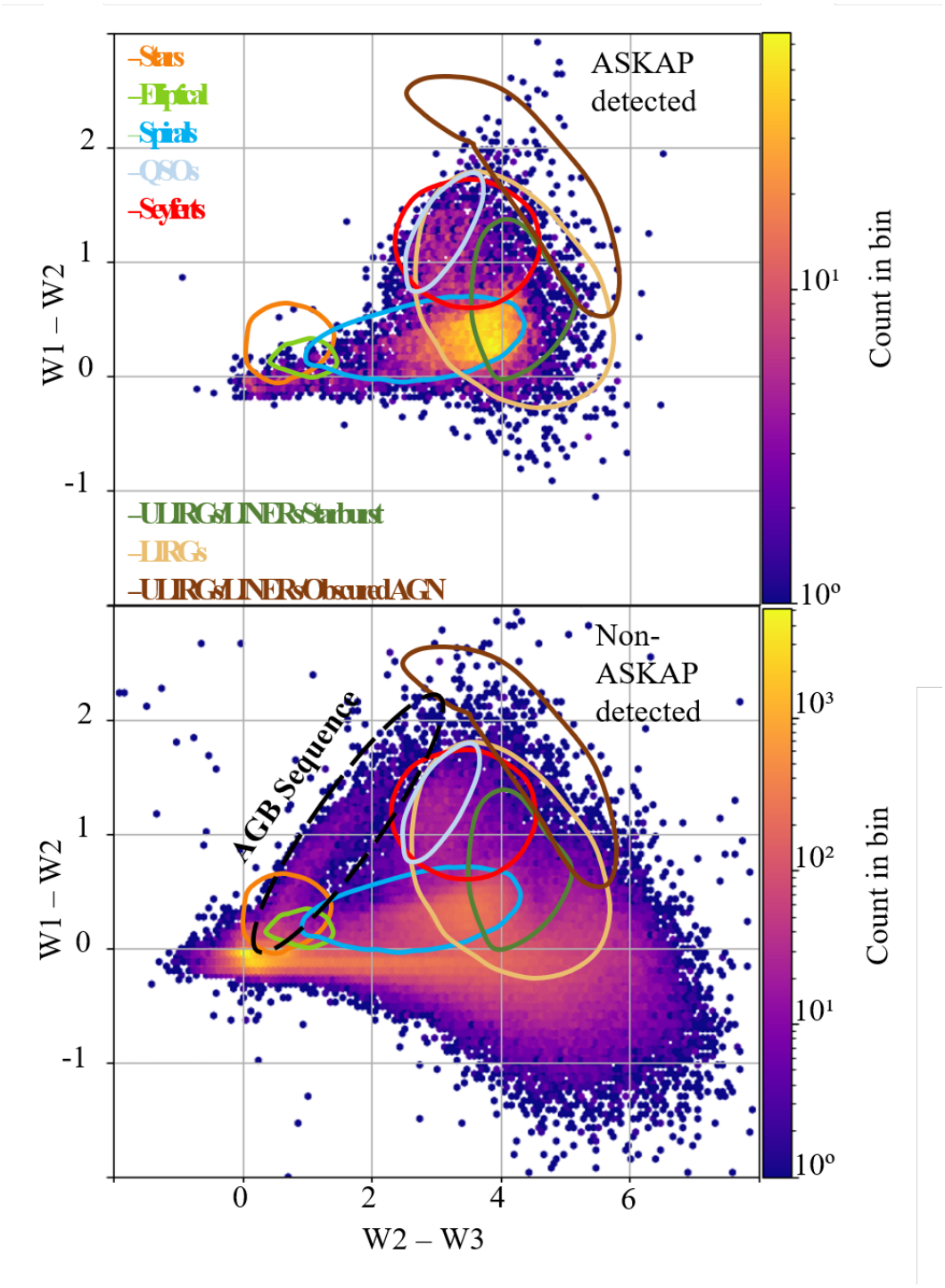

Figure 4.4: (Left) LMC AllWISE colour–colour diagram of the combined Gold and Silver source lists. (Right) LMC AllWISE colour–colour diagram of the non-ASKAP detected sources in the LMC field.

The spectroscopically observed AGN have a selection bias. All had been chosen first and foremost because they are optically bright enough to take a spectrum of, which leads to a surplus of lower redshift AGN. Whether they are bright enough in the optical and a candidate AGN or not is determined from photometric surveys. The AGN has to dominate over the host to be selected as an AGN candidate, based on colour selections. If selected in the optical, optical emission from AGN comes from the accretion disk and the broad line region (BLR), which introduces a bias towards unobscured AGN, typically type 1 AGN. IR provides a selection that is sensitive to both obscured and unobscured AGN, providing a more isotropic selection than in the optical. In this regime the dust obscuring the central AGN re-emits the absorbed optical emission into the IR. Most of the AGN behind the LMC were spectroscopically observed in the optical, even if they were selected in the IR, which lends the bias towards the more unobscured AGN yet again.

The redshifts of the known radio AGN range from $z = 0.001 - 3.46$. The distribution of integrated radio flux densities for spectroscopically observed radio AGN compared to all radio sources (shown in Figure 4.5 left) reveals only a slight bias towards the brightest radio sources having more spectroscopic follow-up, which is not unexpected, since the radio brightness would have made them targets for spectroscopy. The slope of the histogram of flux density of known AGN detected is similar to the slope of the combined Gold and Silver source lists; below ~ 10 mJy the shape of the flux density functions are indistinguishable.

The redshift distributions for radio-bright, radio-faint and non-radio sources (Figure 4.5 right) are similar, with a slight bias towards lower redshift AGN being radio sources. While one might expect a bias towards higher redshift for radio sources compared to the overall distribution of galaxies, observational bias in the spectroscopy most likely negates that. The Silver source list ($F_{888} < 0.5$ mJy) overall redshift distributions seem to be half way between the Gold source list ($F_{888} > 0.5$ mJy) distribution and the non-radio distribution. Non-radio detected spectroscopically observed AGN are generally at higher redshift, this could be because at higher redshift the optical will in fact be the UV restframe, thus making AGN brighter and easier to obtain an optical

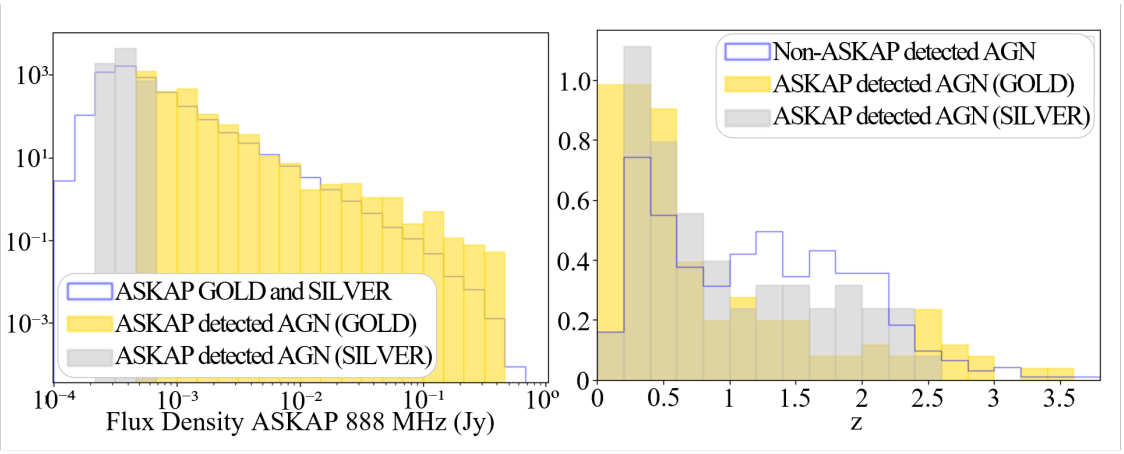

Figure 4.5: (Left) Normalised histogram comparing integrated flux densities of spectroscopically observed radio AGN vs. all radio AGN in the LMC field. This shows that the spectroscopically observed radio AGN tend to brighter flux densities than the overall radio population. (Right) Normalised histogram comparing the redshift of spectroscopically observed radio AGN (separated into Gold and Silver) vs. all spectroscopically observed AGN in the LMC field. This shows that radio-detected spectroscopically observed AGN tend more towards lower redshift compared to non-radio detected spectroscopically observed AGN.

spectrum of. Furthermore, at higher redshift one gets a higher fraction of radio-quiet AGN, which will be detectable in the optical but too faint for ASKAP.

Figure 4.6 compares, using AllWISE colours, the radio and non-radio detected AGN. Radio detected AGN overall appear redder in W2 – W3 than those not detected in radio. A Kolmogorov–Smirnov test of the radio and non-radio detected populations in the W2 – W3 colour band shows that the radio detected population is redder at a significance level of 10% [2].

We compare the redshift distributions of ASKAP and non-ASKAP detected AGN using the same AllWISE colours as before and colouring based on redshift. (Figure 4.7) shows that both the detected and non-detected AGN show a general gradient in redshift as a function of WISE colour, with a few outliers.

[2]the probability of the null hypothesis in the Kolmogorov–Smirnov test that these two samples are

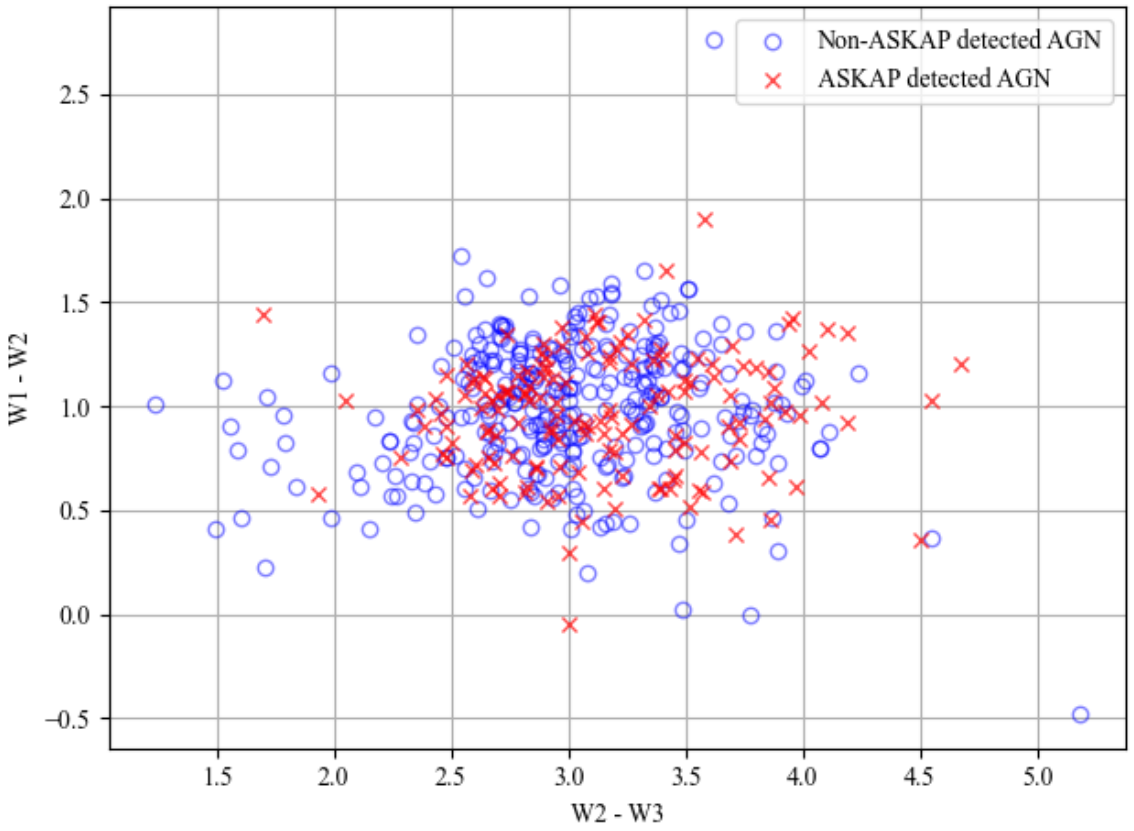

Figure 4.6: AllWISE colour–colour diagram of spectroscopically observed AGN, where those detected in the radio with ASKAP (red crosses) are compared with those not detected with ASKAP (blue circles).

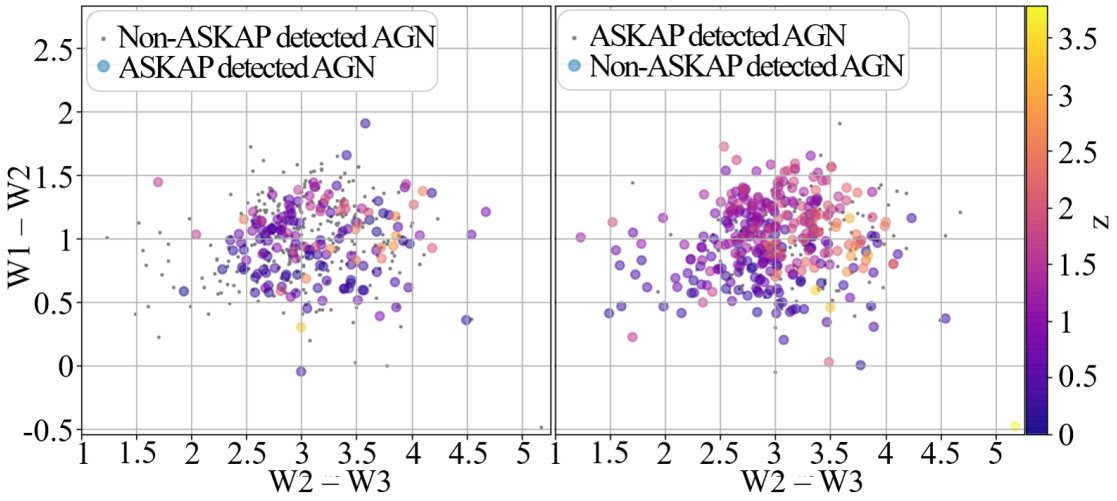

Figure 4.7: (Left) AllWISE colour–colour diagram of ASKAP detected AGN, where the colour indicates redshift. Grey dots indicate the non-ASKAP detected AGN. (Right) AllWISE colour–colour diagram of non-ASKAP detected AGN, where the colour indicates redshift. Grey dots indicate the ASKAP detected AGN. Both the detected and non-detected AGN show a general gradient in redshift as a function of WISE colour, with exceptions.

Figure 4.8 (top-left and right) shows a WISE colour–colour diagram for the ASKAP detected sources and the known AGN in the GOLD and SILVER samples. Using the outlined regions for "Seyferts" to represent the AGN population and "Spirals" to represent the star-forming galaxies population, in the GOLD list ~ 445 out of 4852 ($\sim 9.2\%$) are predicted to be AGN (from AllWISE colours; Wright et al., 2010) and in the SILVER ~ 270 out of 5050 ($\sim 5.4\%$) are predicted to be AGN. As expected, there are more spirals compared to AGN in the SILVER list, since spirals tend to be star-forming and that is where the radio emission comes from, which tends to be fainter than the radio emission from an AGN and therefore a higher fraction of AGN are expected in the GOLD list.

The upturn towards the fainter flux densities at $F_{888} \sim 3$ mJy can be seen in the overall flux distribution (see Figure 3.5) representing the beginning of the faint galaxy population. I thus separated the catalogue into a bright sample ($F_{888} > 3$ mJy) and a faint sample ($F_{888} < 3$ mJy). Cross-matching this with the AllWISE catalogue provides 806 and 13,414 sources, respectively, with a maximum magnitude error in the W1, W2 and W3 bands of 0.2 mag. In Figure 4.8 (lower panels) I plot these two samples on a colour–colour diagram, which shows the distributions of the source types in the bright and faint samples, with regions of expected classifications annotated. From the colours shown in the plots the number of spirals in the faint population is greater than the number of AGN, whereas in the bright sample the ratio of spirals to AGN is lower. As expected the spiral galaxies are more prevalent in the faint population. Cross-matching with the spectroscopically observed AGN list gives 123 radio-faint AGN and 24 radio-bright AGN. The fraction of AGN is larger in the bright sample ($\sim 3\%$) than in the faint sample ($\sim 0.9\%$) as expected.

drawn from the same parent sample.

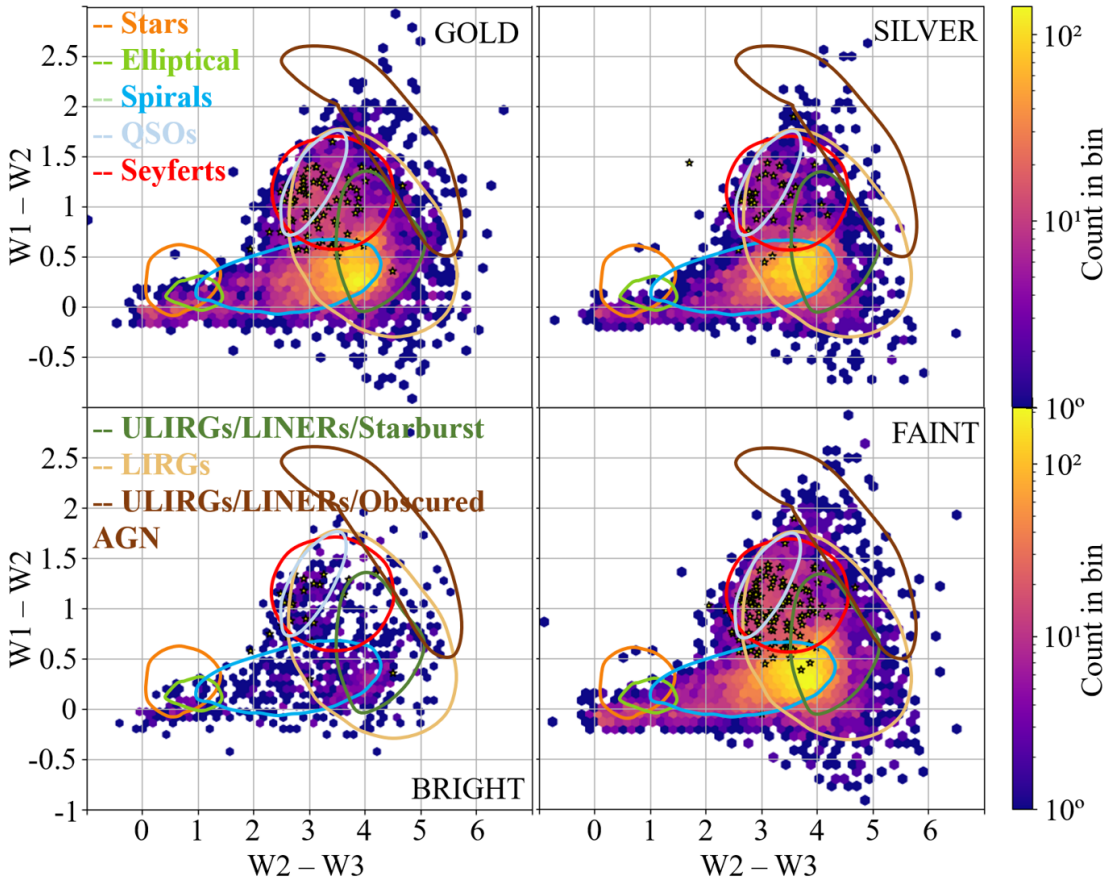

Figure 4.8: LMC AllWISE colour–colour diagram comparing the populations of the Gold (top-left), Silver (top-right), Bright (lower-left; $F_{888} > 3$ mJy) and Faint (lower-right; $F_{888} < 3$ mJy) source lists. The stars represent the spectroscopically observed AGN in their respective source list.

4.3 Spectral indices

We define the spectral index α by $F_\nu \propto \nu^\alpha$, where F_ν is the integrated flux density at frequency ν. A flatter spectral index close to zero indicates free–free emission, and a steep negative spectral index, ~ -0.7, indicates synchrotron emission. Archival ATCA data at 1.384 GHz (Hughes et al., 2007) cover a similar area as the ASKAP image, though with a larger beam size of $40''$. I cross-matched the ASKAP LMC 888 MHz catalogue (Gold, Silver and Bronze) with the ATCA 1.384 GHz source list described in Filipović et al. (2021) with a search radius of $5''$. Limiting the sources to those with a $\leq 1\%$ error in ASKAP and ATCA integrated flux densities (determined by Aegean) provides 1869 sources.

The histogram of spectral indices (Figure 4.9) is dominated by an extragalactic population of synchrotron sources ($\alpha < 0$). This is confirmed by the small sample of 25 spectroscopically confirmed AGN – it is not unusual for some AGN to show a flatter spectrum or even to peak at a few GHz. On the other hand, two YSOs (SAGE 053054.2−683428.3 from Table 4, and [BE74] 615 from Table 3) and an H II region ([RP2006] 1933 from Table 3) have $\alpha \sim 0$ consistent with their free–free emission spectrum. The two "candidate PNe" (SMP LMC 32 and [RP2006] 338, Table D.1) at $\alpha \sim -0.8$ are in fact among the few PN candidates that were rejected on the basis of the large (in these cases $\sim 9''$) distance from the radio position – their synchrotron nature confirms that they are instead background sources.

We suspect that sources with $|\alpha| > 3$ are the result of variability between the epochs at which the ATCA and ASKAP data were obtained. It should be noted that due to the closeness of the two spectral bands, order-of-magnitude changes are not needed to get an extreme value of α. The most extreme case with $\alpha = 5.3$, EMU EC J041753.5−731556 has a spectral index of −0.69 based on SUMSS instead of ASKAP, while EMU EC J054929.5−703545 with $\alpha = 3.2$ has spectral index −1.28 when using SUMSS. The second-most extreme case, EMU EC J052029.2−680051 with $\alpha = 3.9$ sits at the edge of H II nebula LHA 120-N 41 associated with emission-line star AL 139 making the comparison with ATCA data problematic. The other six radio point sources

with $|\alpha| > 3$ generally have faint or no visible optical/IR counterparts in DSS, 2MASS and WISE. These could be extreme radio-variable sources at high redshift. It should be noted that at these frequencies there is the possibility of free–free absorption, which can rapidly absorb the power-law synchrotron emission spectrum towards lower radio frequencies (Clemens et al., 2010). This would cause the radio spectrum to turn-over/peak, which could lead to rapid changes in spectral index, such as those seen here. Another option of a variable radio source could be a variable steep-spectrum pulsar such as the one found recently in the LMC (Wang et al., 2022) using ASKAP data. Lastly, due to the different resolutions of ASKAP, ATCA and SUMSS, extended emission that, for example, would be observed as part of the source in one survey could be resolved out in a higher resolution survey, leading to a lower flux density for the source, and therefore a steep α.

4.4 Conclusions

I present the 888 MHz ASKAP EMU ESP radio continuum survey of the LMC. Our findings can be summarised as follows:

- This new ASKAP survey is a significant improvement (factor of ~ 5 in the median RMS) compared to previous ATCA/MOST surveys of the LMC, and compared to the ASKAP survey of the SMC observed with a partial array (factor of ~ 3 in the median RMS). The improvement in angular resolution has also allowed for greater detail of radio structure to be observed.

- I extracted 30,866, 22,080 and 1666 sources at 888 MHz separated into Gold, Silver and Bronze source lists, respectively, with the majority of these sources detected above the 5σ threshold.

- Exploring the GLEAM 4-Jy sample (White et al., 2020a,b) in the ASKAP field of the LMC revealed six GLEAM sources, showcasing that the ASKAP is an improvement in angular resolution (from $\sim 2'$ to $\leq 14''$), allowing for

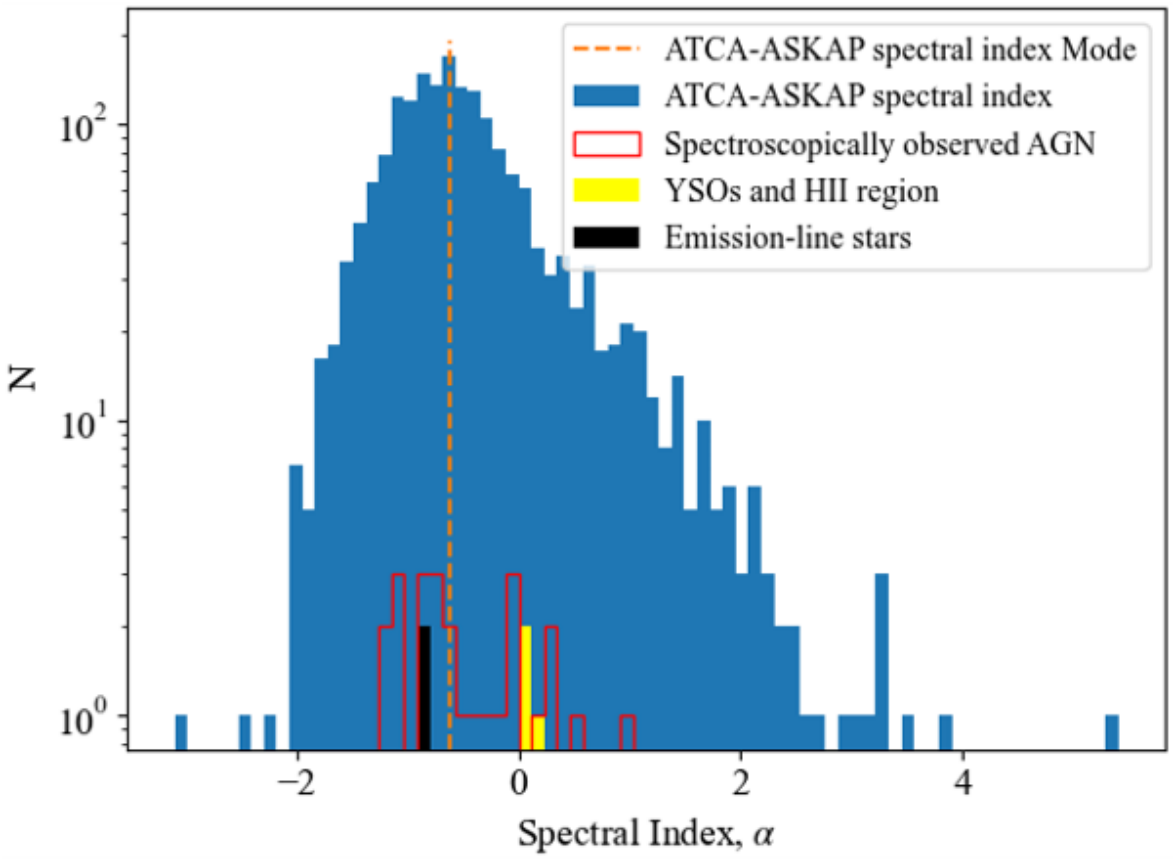

Figure 4.9: Distribution of spectral indices calculated from ATCA (1.4 GHz) and ASKAP (888 MHz) integrated flux densities.

radio sources to be better traced to their optical/IR counterparts and revealing morphology associated with jets and likely interaction with the intra-cluster medium.

- Cross-matching the new LMC ASKAP source catalogue with AllWISE reveals 14,333 sources with photometry errors < 0.2 mag. Comparing the radio and non-radio sources reveals that the majority of the radio sources are extragalactic.

- Cross-matching the new LMC ASKAP source catalogue with the milliquas catalogue (Flesch, 2019b,a) reveals 190 radio detections out of 657 spectroscopically confirmed AGN. More than 92% of these are newly detected here.

- A higher fraction of AGN is found within the bright population ($F_{888} > 3$ mJy) than within the faint population ($F_{888} < 3$ mJy). As expected there is a higher fraction of spirals in the faint population compared to the bright population.

- Cross-matching the new ASKAP 888 MHz data with archival ATCA 1.4 GHz data reveals a spectral index distribution peaking at $\alpha = -0.6$ (where $F_\nu \propto \nu^\alpha$). This corroborates that the majority of sources are extragalactic synchrotron emitters. Extreme spectral indices ($|\alpha| > 3$) can be explained by variability.

The study of the foreground and Magellanic sources can be found in the Appendices, where a more detailed description can be found in Pennock et al. (2021). From this 114 out of 1334 PNe listed in Simbad (Wenger et al., 2000) were found to have a counterpart in ASKAP within $10''$. Only 37 of these detections are likely true PNe, with the remaining sources being emission-line stars, candidate post-AGB objects, YSOs and compact H II regions. Furthermore, the YSOs detected with Herschel (Oliveira et al., 2019) were discovered to show a strong correlation between integrated radio flux densities and gas temperature, with those with $T > 700$ K detected in radio and those with $T < 700$ K not detected in radio. This suggests a strong relation between the line emission from the photo-dissociation regions at the interfaces of

neutral and ionized gas, and the free–free emission from the ionized regions surrounding the nascent, presumed massive O-type stars. Lastly, of 46 SNe listed in Simbad, most associated radio detections could not be unequivocally ascribed to the SN itself and are more likely emission from a background host galaxy. Only SN 1978A and Nova LMC 1988 b are securely detected with ASKAP.

Future work will include more comprehensive analyses of the radio variability. I am employing machine learning to classify a combined multi-wavelength data set and study the radio–IR relation for galaxies and quasars, which can be seen in Chapter 6. The background population seen in the ASKAP image shows the potential for large-scale Faraday rotation and H I absorption measurements throughout the LMC.

5 Search for more AGN like SAGE0536AGN

This chapter explores a small sample of sources found to be similar to SAGE0536AGN, a source dominated by nuclear dust emission. This sample was discovered using the machine learning algorithm, t-SNE, on a combination of photometric VMC, Gaia, AllWISE and ASKAP EMU survey data. The first section describes the sample, the light curves and the spectra newly observed and from the literature, as well as the modelling of the photometric SEDs to estimate properties of the AGN and their host galaxies. In the next section IR, radio and AGN dust properties are discussed, as well as the samples place in the green valley.

This chapter is published as Pennock et al. (2022), but here includes additional analysis as obtained/performed since publication.

5.1 The sample

The sample is chosen from the t-SNE map that resulted from the use of the t-SNE algorithm on a clean dataset of a combination of VMC, Gaia EDR3, WISE and ASKAP data of the SMC, see Figure 5.1 (left). SAGE0536AGN and SAGE0534AGN are shown to be close to each other in an area containing few sources, implying a rare class of AGN. I focused on the group of 18 sources that include SAGE0536AGN and SAGE0534AGN, see red box in Figure 5.1 (right), that are also separate from the large clusters of sources, in order to find more such objects. A list of these sources, with their identifiers and co-ordinates, can be seen in Table 5.1. The other objects in this t-SNE map are to be explored in Chapter 6 that makes use of more than one machine learning technique and a wider range of multi-wavelength data, which will classify the sources in the direction of the SMC and LMC.

The sample (including SAGE0536AGN and SAGE0534AGN) is made up of 18 sources, 16 of which are spectroscopically confirmed extragalactic sources (see next subsection).

Table 5.1: Sample of similar sources identified through a t-SNE analysis.

Source Name	Identifier	RA (J2000)	DEC (J2000)
SAGE0536AGN	SAGE1C J053634.78−722658.5	5:36:34.78	−72:26:58.5
SAGE0534AGN	SAGE1C J053444.17−673750.1	5:34:44.17	−67:37:50.1
1	WISEA J003617.01−743131.3	0:36:16.99	−74:31:31.3
2	WISEA J011337.10−742755.3	1:13:37.08	−74:27:55.3
3	WISEA J003156.88−733113.6	0:31:56.89	−73:31:13.6
4	WISEA J002602.54−724718.0	0:26:02.54	−72:47:18.0
5	OGLE SMC-LPV-7107	0:48:25.71	−72:44:02.8
6	WISEA J011408.02−723243.1	1:14:07.99	−72:32:43.3
7	[MCS2008] 11	0:55:51.51	−73:31:10.0
8	[MA93] 1895	1:22:36.94	−73:10:16.7
9	MQS J012108.42−730713.1	1:21:08.43	−73:07:13.1
10	MQS J011534.10−725049.3	1:15:34.09	−72:50:49.3
11	WISEA J003910.76−713409.9	0:39:10.78	−71:34:09.9
12	[VV2006] J005116.9−721651	0:51:16.95	−72:16:51.5
13	2E 238	0:57:32.75	−72:13:02.3
14	WISEA J013604.46−721315.3	1:36:04.46	−72:13:15.4
15	NAME SMC B0031−7042	0:34:05.26	−70:25:52.3
16	WISEA J004952.56−692956.4	0:49:52.53	−69:29:56.4

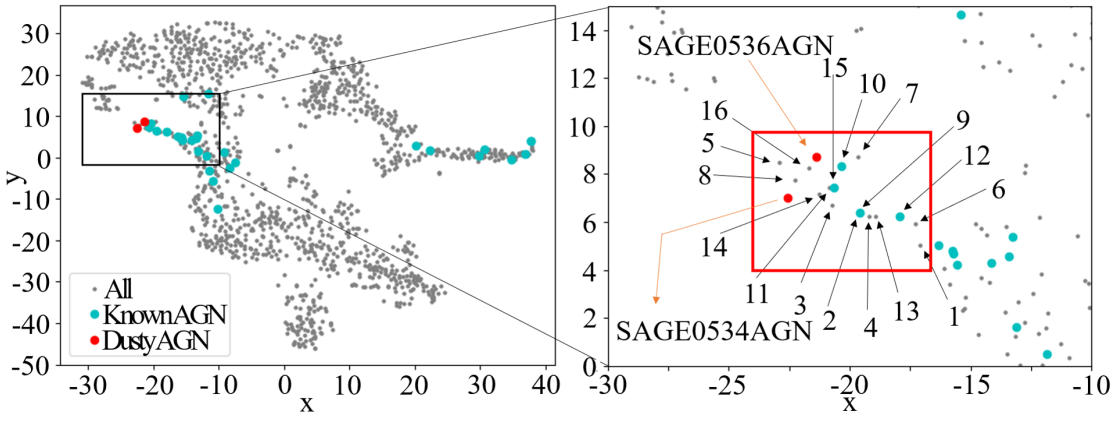

Figure 5.1: (Left) t-SNE map created from a combination of VMC, Gaia EDR3, AllWISE and ASKAP data. Blue dots represent known AGN that have been spectroscopically confirmed. Red indicates the dusty AGN SAGE0536AGN and SAGE0534AGN. (Right) Zoom in on area containing the dusty AGN. The 16 sources (blue and grey dots) within the red box in this plot represent the sample explored in this paper. The numbers correspond to the source names in Table 5.1.

Source 5 is a near-superposition of a carbon star in front of the true extragalactic radio source, which can be seen from the spectrum and the annotated spectral lines (Barnbaum et al., 1996; van Loon et al., 1998), which can be seen in Figure 5.2. The flux of this spectrum has not been calibrated due to no standard star being observed.

Source 8 has been previously identified as a potential Hα emission line star (Meyssonnier & Azzopardi, 1993, identified an emisson line and no underlying continuum), a far-infrared (far-IR) object (Boyer et al., 2011; Srinivasan et al., 2016) and an emission line object (Groenewegen et al., 2020). After this work was published Source 8 was spectroscopically observed and revealed to be an emission-line star in the SMC, but the work presented here in Section 5.1 will continue as originally presented.

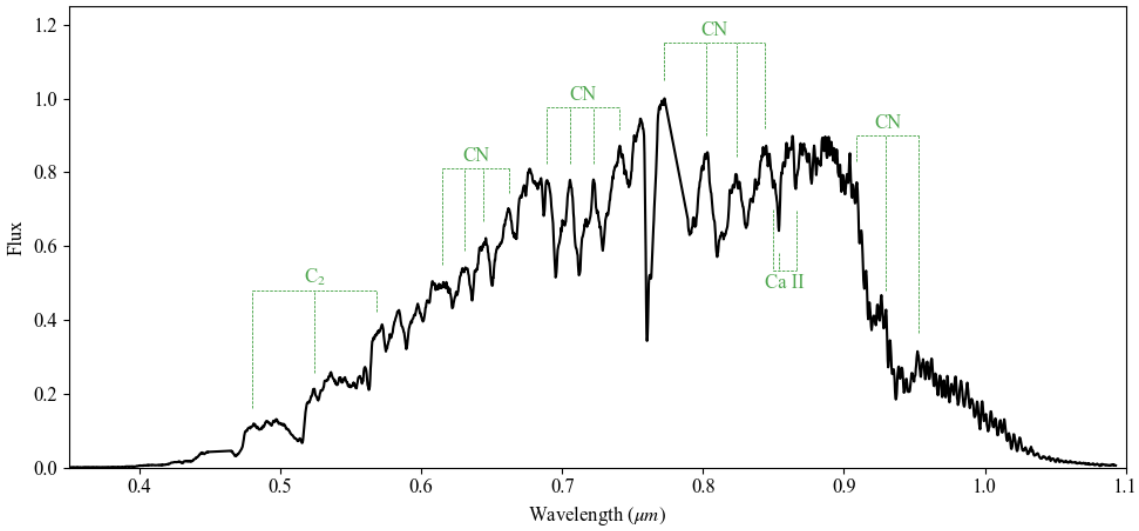

Figure 5.2: Spectrum of carbon star Source 5 where main spectral features are annotated. It was observed with SALT under program 2021-1-SCI-029 (PI: Jacco van Loon). I used the RSS, a combination of three CCD detectors with total 3172 × 2052 pixels and spatial resolution of 0.″1267 per pixel. I used the long-slit with width 1.″5, grating PG0300, filter PC03400 and an Argon arc lamp. The data was processed using the standard IRAF tools. This spectrum is not flux calibrated, the y-axis shows normalised flux.

Table 5.2: t-SNE selected sample variability amplitudes and mean magnitudes in the K_s band.

Source Name	K_s Amplitude (Vega mag)	Mean K_s (Vega mag)
SAGE0536AGN	0.09±0.01	13.49±0.01
SAGE0534AGN	0.05±0.01	13.98±0.03
1	0.18±0.02	15.83±0.05
2	0.16±0.02	15.34±0.05
3	0.02±0.02	14.79±0.02
4	0.02±0.01	15.65±0.03
5	1.20±0.13	11.45±0.36
6	0.03±0.03	15.95±0.04
7	0.16±0.03	15.10±0.06
8	0.55±0.07	13.19±0.14
9	0.09±0.02	15.30±0.03
10	0.17±0.02	14,53±0.05
11	0.22±0.01	14.56±0.07
12	0.25±0.01	15.34±0.10
13	0.27±0.01	15.49±0.08
14	0.26±0.02	14.13±0.32
15	0.09±0.02	14.70±0.03
16	0.12±0.01	13.70±0.08

5.1.1 VMC light curves

The VMC survey is comprised of multi-epoch observations, which allows for the detection of variability. The light curves of the sample can be seen in Figure 5.3. The amplitudes of variation in K_s were calculated between the highest/lowest daily median values.For some of these sources we are not seeing the full amplitude, such as for Source 14, where the source becomes brighter without reaching a noticeable peak. The amplitudes calculated from the VMC light curves in K_s can be seen in Table 5.2.

Source 5 shows large-amplitude, semi-regular variability that corroborates its identity as a carbon star. The variability of the other sources combined with their extragalactic spectroscopic confirmation, confirms the presence of an AGN. However,

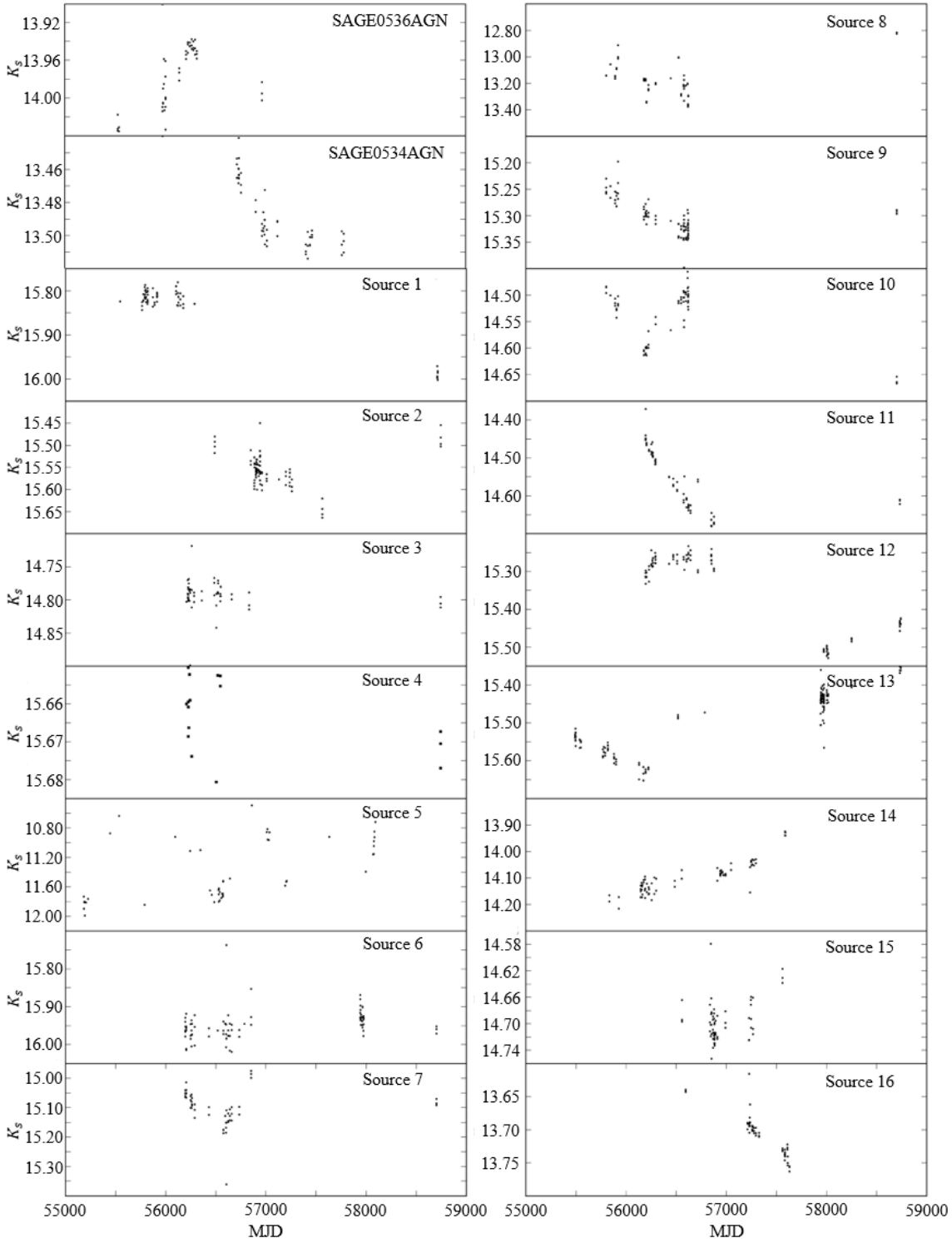

Figure 5.3: Light curves from the VMC survey of all sources in our t-SNE selected sample. Note that typical uncertainties in the K_s band are < 0.01 mag.

Sources 3, 4 and 6 show little variability.

5.1.2 Optical line identifications and spectral analysis

The spectra of SAGE0536AGN, SAGE0534AGN and 14 out of 16 t-SNE sample sources that were observed with SALT, SAAO's 1.9m telescope or with other facilities prior to this study are shown in Figure 5.4. Sources that were observed as part of other surveys are also shown in Figure 5.4. Only Sources 5 (Star) and 8 (was not observed before this work was published) are not shown. The spectra of Source 5 and sources with multiple available spectra that are not shown in Figure 5.4 can be found in Appendix E. Observations details are listed in Table 5.3. Redshifts are listed in Table 5.4. The redshift for Source 8 was estimated from photometry (Flesch, 2015, 2021).

The Full Width at Half Maximum (FWHM) is calculated by modelling the continuum surrounding the emission line and then subtracting the continuum from the spectra. After this the half maximum height of the emission line is calculated from the line profile and then subsequently the width of the emission to get the observed FWHM. The intrinsic FWHM is then calculated from $\mathrm{FWHM_{intrinsic}} = \sqrt{(\mathrm{FWHM_{observed}})^2 - (\mathrm{FWHM_{instrument}})^2}$, where $\mathrm{FWHM_{instrument}}$ is the FWHM of the instrument used to obtain the spectrum.

Table 5.3: Table of sources investigated/observed in this work. (1) This work; (2) van Loon & Sansom (2015); (3) Kozłowski et al. (2013); (4) Flesch (2015, 2021); (5) Geha et al. (2003); (6) Kishimoto et al. (*in prep.*).

Source Name	Telescope	Date Observed	Ref.
SAGE0536AGN	SALT	08-09-2012	(2)
SAGE0534AGN	SAAO 1.9m	16-11-2019	(1)
SAGE0534AGN	SALT	01-11-2021/17-03-2022	(1)
1	SAAO 1.9m	24-11-2019	(1)
2	SAAO 1.9m	31-10-2019	(1)
3	SAAO 1.9m	05-11-2019	(1)
4	SAAO 1.9m	22-11-2019	(1)
5	SALT	17-07-2021	(1)
6	SAAO 1.9m	29-10-2019	(1)
6	SALT	20-07-2022	(1)
7		02-2012 – 01-2013	(3)
8		N/A	(4)
9		02-2012 – 01-2013	(3)
9	SALT	09-09-2022	(1)
10		02-2012 – 01-2013	(3)
11	SAAO 1.9m	02-11-2019	(1)
12		10-1999 – 01-2001	(5)
13	SAAO 1.9m	19-11-2019	(1)
13	SALT	24-07-2021	(1)
14	SAAO 1.9m	30-10-2019	(1)
14	SALT	07-09-2022	(1)
15		22-08-2004	(6)
16	SALT	01-09-2021	(1)

Table 5.4: Table of sources investigated in this work. Redshifts are calculated from spectroscopy, except for source 8 (indicated with an *) for which the redshift was calculated from photometry. Source 5 was found to be a carbon star in the SMC dominating in the optical/IR.

Source Name	z	Emission Lines	— FWHM (km s^{-1}) —		
			Hα	Hβ	Mg II
SAGE0536AGN	0.1428±0.0001	Hα	3900± 450		
SAGE0534AGN	1.009±0.002	Mg II			10310±300
1	0.77±0.01	Mg II			2300±250
2	1.12±0.01	Mg II, C III			5100±550
3	1.2±0.01	Mg II			3800±350
4	1.23±0.02	Mg II, C III			6100±200
5	–				
6	1.06±0.01	Mg II, C II, C III			3250±160
7	0.186±0.005	Hα, Hβ			
8	0.5*				
9	0.985±0.005	Mg II			6280±160
10	0.201±0.005	Hα, Hβ	3050±1000	3700±1000	
11	0.4±0.01	Hβ, Hγ, Mg II		1900± 250	2200±250
12	0.49±0.005	Mg II, Hγ, Hβ, O III			5750±500
13	0.81±0.02	Mg II, Hβ			6300±350
14	0.41±0.01	Mg II, Hγ, Hβ, Hα, O III	3210± 100		
15	0.363±0.005	Mg II, Hγ, Hβ, Hα, O III	5450± 900	6150± 300	6450±500
16	0.125±0.01	Hα, Hβ, O III	2050± 450	2800± 300	

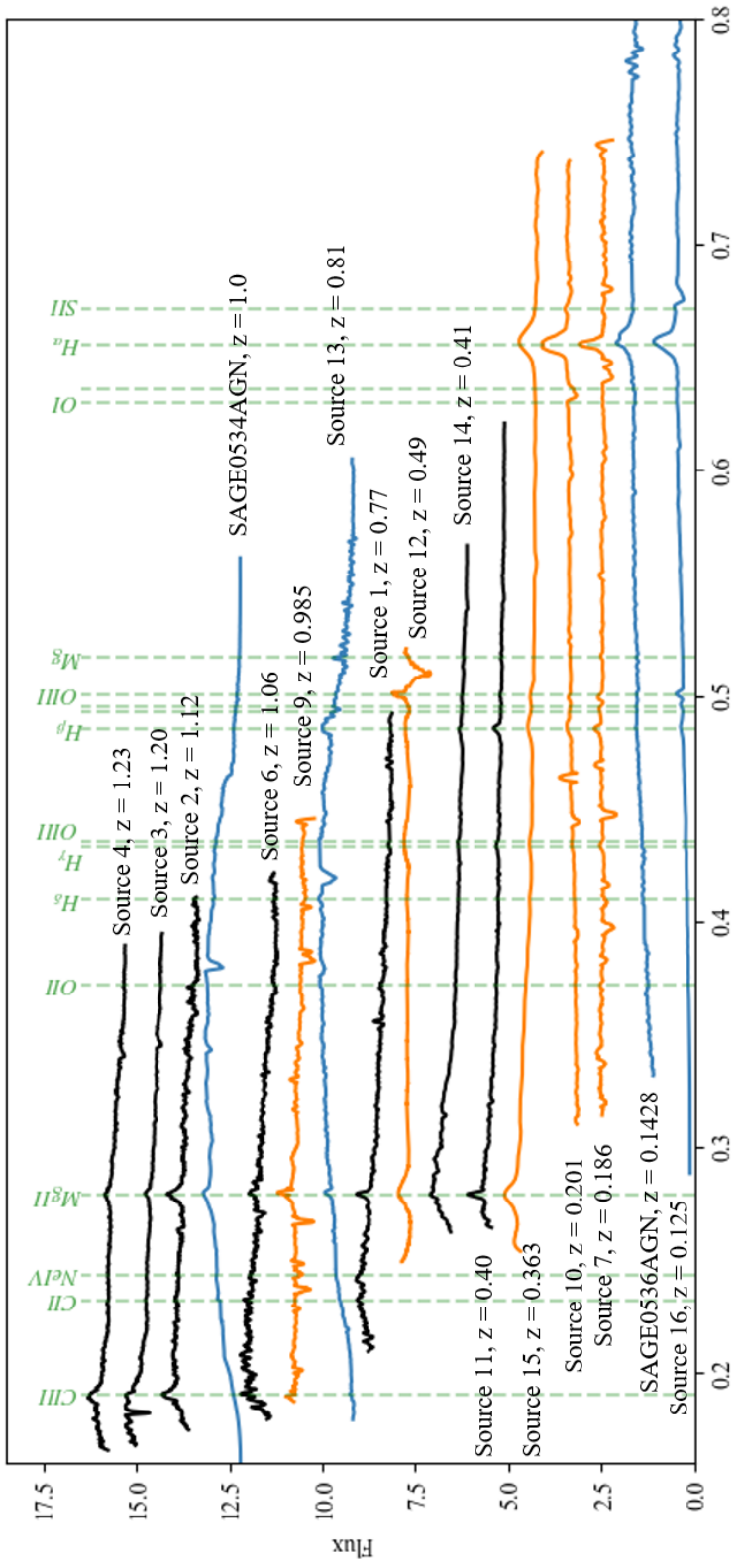

Figure 5.4: Spectra of t-SNE selected sample, excluding Source 5 (carbon star, spectrum shown in online appendix) and 8 (no spectrum available). Spectral flux has been normalised. Black indicates sources that were observed with SAAO 1.9m telescope and blue indicates sources observed with SALT. Orange indicates sources that were observed prior to this study with other facilities.

The continuum of the optical spectra was modelled using Python's Astropy module (Astropy Collaboration et al., 2013, 2018). This facilitated the use of a low order polynomial to model the continuum. The most notable emission line is the Mg II λ2798 line, observed in all but four of the sources.

5.1.2.1 Black hole masses

The sample showcases emission lines from either Mg II λ2798, Hα λ6563 Å or Hβ λ4861 Å. From these we can calculate the black hole masses ($M_{\rm BH}$). The black hole mass was not calculated for Sources 6 and 14 due to their spectra being too noisy, and Sources 1, 7 and 11 had FWHM that were too close to the instrumental FWHM to be disentangled. These measurements could, however, be used to set an upper limit on the black hole mass.

The calibrations used for calculating the black hole masses from the Mg II (Trakhtenbrot & Netzer, 2012), Hβ (Vestergaard & Peterson, 2006) and Hα (Baron & Ménard, 2019) emission lines are:

$$M_{\rm BH} = 10^{6.748} \left(\frac{L_{3000}}{10^{44}{\rm erg/s}}\right)^{0.620} \left(\frac{\rm FWHM_{Mg\,II}}{10^{3}{\rm km/s}}\right)^{2} M_{\odot} \tag{5.1}$$

$$M_{\rm BH} = 10^{6.910} \left(\frac{L_{5100}}{10^{44}{\rm erg/s}}\right)^{0.500} \left(\frac{\rm FWHM_{H\beta}}{10^{3}{\rm km/s}}\right)^{2} M_{\odot} \tag{5.2}$$

$$M_{\rm BH} = \epsilon 10^{6.9} \left(\frac{L_{5100}}{10^{44}{\rm erg/s}}\right)^{0.54} \left(\frac{\rm FWHM_{H\alpha}}{10^{3}{\rm km/s}}\right)^{2.06} M_{\odot} \tag{5.3}$$

where L_{3000} and L_{5100} are the monochromatic continuum luminosities at rest-frame 3000 Å and 5100 Å respectively in erg s^{-1} derived from SED fitting (see Section 5.1.3 for more information). Calculated black hole masses are shown in Table 5.5. The constant $\epsilon = 1.075$ (Reines & Volonteri, 2015) was adopted. Errors on the monochromatic continuum luminosities were calculated by the SED fitter, errors on the FWHM are the standard error of the mean measurement from the line profile and error of the black hole mass is propagated from these two errors.

Table 5.5: Black hole masses calculated using equations 5.1, 5.2 and 5.3. $L_{\rm bol}$(AGN) is the AGN bolometric luminosity calculated during SED fitting (see Section 5.1.3). Eddington Ratio is defined as the $L_{\rm bol}$(AGN)/$L_{\rm edd}$, where the $L_{\rm edd}$=1.25×10^{38}$M_{\rm BH}$ erg s^{-1}. Sources that are not listed here either have a noisy spectrum or emission lines that are smaller or close to the FWHM of the instrument.

Source Name	— $M_{\rm BH}$ ($M_\odot$)—			$L_{\rm bol}$(AGN)	Edd. Ratio
	Hα	Hβ	Mg II	(erg s^{-1})	(%)
SAGE0536AGN	(5.5±1.3)×10^7	-	-	(4.4±0.2)×10^{44}	6.3± 1.5
SAGE0534AGN	-	-	(1.9±0.1)×10^{10}	(2.4±0.1)×10^{47}	9.8± 0.8
2	-	-	(2.9±0.6)×10^9	(9.5±0.5)×10^{46}	26.0± 5.6
3	-	-	(1.8±0.3)×10^9	(1.9±0.1)×10^{47}	80.4±15.5
4	-	-	(3.3±0.2)×10^9	(7.8±0.4)×10^{46}	18.6± 1.6
6	-	-	(5.2±0.5)×10^8	(3.3±0.2)×10^{46}	50.7± 5.6
9	-	-	(2.5±0.1)×10^9	(4.5±0.2)×10^{46}	14.5±1.1
10	(6.7±4.6)×10^7	(9.2±4.7)×10^7	-	(6.4±0.3)×10^{44}	6.4± 3.9
12	-	-	(2.9±0.5)×10^8	(5.1±0.3)×10^{45}	14.2± 2.6
13	-	-	(1.2±0.1)×10^9	(2.8±0.2)×10^{46}	18.0± 2.6
14	(3.4±0.2)×10^8	-	-	(9.9±0.5)×10^{45}	23.1± 1.7
15	(5.7±1.9)×10^8	(5.0±0.5)×10^8	-	(6.5±0.3)×10^{45}	9.7± 3.7
16	(1.7±0.8)×10^7	(3.1±0.7)×10^7	-	(5.7±0.3)×10^{44}	19.1± 8.2

SAGE0536AGN's black hole mass was previously reported as $M_{\rm BH} = (3.5 \pm 0.8) \times 10^8 {\rm M}_\odot$, and $L_{\rm bol} = (5.5 \pm 1.3) \times 10^{45}$ erg s^{-1} (≈ 12 % of the Eddington luminosity (van Loon & Sansom, 2015). In this work the black hole mass of SAGE0536AGN is calculated from the Hα line to be $M_{BH}=(5.5 \pm 1.3)\times 10^7$ M$_\odot$ with an Eddington ratio of $\sim$6 %. This mass combined with the calculated velocity dispersion, $\sigma \sim 202 \pm 15$ km s^{-1}, puts SAGE0536AGN in agreement with the known correlation between velocity dispersion and black hole mass (e.g. Graham, 2008).

5.1.3 Cigale modelling

5.1.3.1 Cigale models for SAGE0536AGN and SAGE0534AGN

CIGALE SED fits of SAGE0534AGN and SAGE0536AGN are shown in Figure E.3 and the calculated parameters can be found in Table 5.6.

The fit of SAGE0536AGN shows that the emission from this object is not solely due to the AGN, as expected from the visible galaxy seen in survey images (see top left in Figure 5.6), $\sim 11\%$ is from the host galaxy. Extinction due to polar dust is the highest for SAGE0536AGN compared to the rest of the sample. Accretion power is smallest for SAGE0536AGN.

The fit of SAGE0534AGN, however, shows that the emission is almost solely due to the AGN. Compared to SAGE0536AGN, SAGE0534AGN shows similar τ, i and R (radial thickness of torus) values. Extinction in polar dust is minimal compared to SAGE0536AGN. The opening angle is expected to be $\sim 40^\circ$ from observations (e.g. Stalevski et al., 2016). SAGE0534AGN has the expected opening angle, whilst SAGE0536AGN has the largest opening angle of the sample, implying a thinner torus.

5.1.3.2 Cigale models of the t-SNE sample

The CIGALE models of the other sources can be seen in Appendix E. The majority of sources show a lack of host galaxy contribution, $f_{\rm AGN} > 70\%$, which implies differences in the dusty torus (shape, density, etc.) are causing the differences. This is shown by

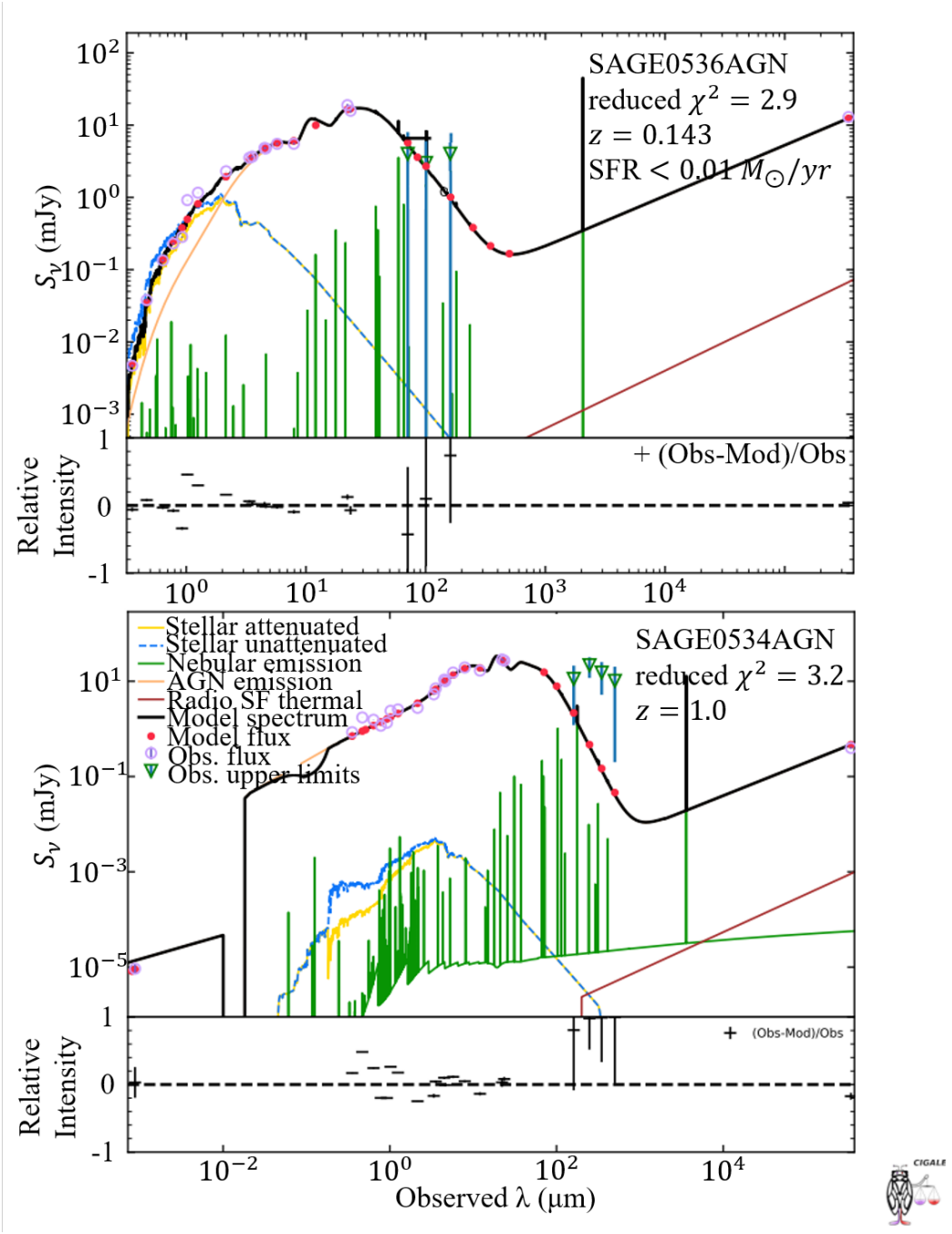

Figure 5.5: CIGALE best fits of SAGE0536AGN (left) and SAGE0534AGN (right). SED fits of the rest of the sample can be found in the Appendix E. Source 8 is included despite being revealed as a star due to its identity being uncertain at time of fitting.

Table 5.6: AGN properties calculated with CIGALE. AGN fraction is the fraction of IR luminosity from the object that is due to the AGN. τ is the torus optical depth at 9.7 μm. The inclination angle, i, is the viewing angle, where $i = 0°$ is face-on and $i = 90°$ is edge-on. R is the ratio between the maximum and minimum radii of the torus. The opening angle, *oa*, is the angle between the equatorial plane and edge of the torus. *pl* is the torus radial density parameter and q is the torus density angular parameter, such that $\rho \propto r^{-pl} e^{-q|\cos(\theta)|}$, where ρ is the torus density and r is the radius of the torus. $E(B-V)$ is the extinction caused by polar dust. Accretion power is the intrinsic AGN disk luminosity averaged over all directions. AGN luminosity is the sum of the observed AGN disk luminosity (some might be extincted) and the observed AGN dust re-emitted luminosity. AGN torus fraction is the fraction of the AGN luminosity that is re-emitted by the torus dust. (Continued on the next page.)

Source	AGN fraction	τ	i (degrees)	R	*oa* (degrees)	*pl*
SAGE0536AGN	0.90±0.01	3.0±0.1	3.4± 4.7	29.1±2.9	79.5±2.2	0.5±0.2
SAGE0534AGN	0.99±0.01	4.0±1.0	4.5± 4.9	26.2±4.9	40.1±1.0	1.2±0.3
Source 1	0.98±0.02	7.4±1.5	10.3± 7.9	23.0±6.5	65.9±7.7	1.2±0.2
Source 2	0.99±0.01	4.1±1.0	21.9± 9.5	16.6±7.5	40.3±1.6	0.6±0.3
Source 3	0.91±0.01	3.3±0.7	5.7± 5.0	29.0±3.1	50.0±0.1	1.4±0.2
Source 4	0.99±0.01	4.7±1.6	10.7± 9.0	27.1±4.6	54.9±5.0	1.2±0.3
Source 6	0.94±0.01	3.7±0.9	24.4± 7.0	29.0±3.0	53.9±4.9	1.2±0.2
Source 7	0.72±0.01	5.8±1.4	49.3± 2.5	24.8±6.3	40.7±2.5	0.9±0.2
Source 8	0.89±0.01	4.6±0.8	21.8± 3.9	28.1±3.9	68.2±3.9	1.4±0.2
Source 9	0.98±0.02	3.8±1.0	4.6± 5.0	26.5±4.8	53.3±4.7	0.4±0.3
Source 10	0.70±0.02	10.1±1.3	2.5± 4.5	29.4±2.4	74.0±6.0	0.6±0.3
Source 11	0.96±0.03	5.4±1.9	21.8± 9.6	28.1±3.9	52.6±4.5	1.2±0.2
Source 12	0.99±0.01	9.1±1.6	19.1±11.5	24.7±6.0	46.7±4.8	0.6±0.3
Source 13	0.93±0.03	8.5±1.6	5.7± 6.4	23.8±6.2	62.7±6.5	0.5±0.2
Source 14	0.98±0.02	5.7±1.3	12.3± 9.7	25.8±5.5	50.6±2.7	0.9±0.3
Source 15	0.90±0.02	7.8±2.2	0.4± 2.0	26.4±5.5	79.5±2.2	0.1±0.2
Source 16	0.95±0.02	5.1±1.6	27.4± 7.0	20.4±8.0	46.0±4.9	0.7±0.6

Source	q	$E(B-V)$	Accretion power (10^{37} W)	AGN luminosity (10^{37} W)	Torus fraction
SAGE0536AGN	1.2±0.3	0.96±0.01	2.7± 0.2	4.4± 0.2	0.72±0.05
SAGE0534AGN	1.0±0.4	0.00±0.01	772.1±38.6	2375.9±118.8	0.17±0.01
Source 1	0.2±0.2	0.10±0.01	69.4± 6.9	171.4± 8.6	0.52±0.04
Source 2	1.2±0.2	0.00±0.01	358.1±29.2	953.6± 47.7	0.17±0.02
Source 3	1.3±0.2	0.04±0.01	738.9±36.9	1850.2± 92.5	0.30±0.02
Source 4	1.0±0.4	0.00±0.01	233.1±17.4	774.8± 38.7	0.28±0.03
Source 6	0.6±0.4	0.30±0.01	124.3±14.1	329.7± 16.5	0.33±0.03
Source 7	1.3±0.2	0.22±0.03	5.7± 0.3	5.9± 0.3	0.43±0.04
Source 8	1.5±0.1	0.08±0.01	117.6±12.4	223.1± 11.2	0.58±0.04
Source 9	1.3±0.2	0.00±0.01	136.5± 6.8	447.7± 22.4	0.24±0.02
Source 10	0.5±0.4	0.30±0.01	2.7± 0.1	6.4± 0.3	0.65±0.05
Source 11	1.1±0.4	0.08±0.01	38.3± 4.9	82.6± 5.2	0.41±0.03
Source 12	0.4±0.4	0.12±0.02	24.0± 3.4	50.6± 2.5	0.40±0.04
Source 13	1.0±0.4	0.10±0.02	119.3±13.5	276.7± 22.9	0.52±0.06
Source 14	0.4±0.4	0.00±0.02	29.9± 2.2	98.6± 4.9	0.28±0.02
Source 15	0.0±0.1	0.00±0.01	16.4± 0.8	64.8± 3.2	0.35±0.03
Source 16	1.0±0.4	0.50±0.07	3.9± 0.5	5.7± 0.3	0.43±0.06

the ranges of the other parameters such as R, τ, pl, q and polar dust extinction. All the sources, except Source 7, show an inclination angle between $0 < i < 45°$, implying the central engine of the AGN is seen for all sources. All the sources, except Source 2, show $R > 20$, implying a sample with a thick torus, some of which may be thicker than the models allow ($10 < R < 30$).

5.1.4 X-ray observations

Seven of these sources have been detected at X-ray energies with the XMM-Newton telescope (Jansen et al., 2001) (Table 5.7).

Comparing with the CIGALE parameters calculated, the overall luminosity of the AGN increases with X-ray luminosity as expected. Also, the X-ray luminosities decrease with both the AGN inclination angle, polar extinction and AGN dust percent-

Table 5.7: The average, error-weighted 0.2 – 12 keV X-ray flux and the corresponding luminosity and their identifications from Maitra et al. (2018).

Source Name	0.2 – 12 keV flux (10^{-13} erg s^{-1} cm^{-2})	Luminosity (10^{36} W)	Designation
6	0.76 ± 0.09	47.27±5.59	J011408.02−723243.1
7	2.58 ± 0.14	2.56±0.14	J005551.53−733110.1
8	0.16 ± 0.07	1.57±0.69	
9	1.13 ± 0.11	58.49±5.69	J012108.43−730713.1
10	0.50 ± 0.20	0.59±0.24	
12	6.94 ± 0.30	64.85±2.80	208.16034.100
13	1.72 ± 0.18	54.83±5.74	J005732.73−721302.1

age, as expected, as when the central engine becomes more obscured, X-ray emission decreases. Overall, from looking at Table 5.6, the X-ray luminosities of the sources with $f_{\rm AGN}$ <90% are in general lower than for $f_{\rm AGN}$ >90% sources.

5.1.5 The host galaxies

The host galaxy of SAGE0536AGN is resolved in VMC images, giving the appearance of a red galaxy. The fits of CIGALE also show that $\sim$ 11% of this object's total dust emission is due to the host galaxy. CIGALE also calculated Sources 7, 8, 10, 15 and 16 to have emission contributed by the host galaxy. Of these, only Sources 10 and 16 have host galaxies that are resolved in the VMC images.

The appearance of the host galaxies provides insight into what step of evolution they are in, be they red dead elliptical, blue and star-forming, or intermediate as a green valley galaxy (Salim, 2014).

The host galaxies were fitted with GALFIT. I used the Sérsic profile function, as varying the Sérsic exponent (which determines the light profile) can match the other available functions in GALFIT. The GALFIT modelling was done using VMC $K_{\rm s}$ band images. The models used are shown in Table 5.8. Each AGN was fit with three Sérsic functions, one for the host galaxy (largest component) and two for the central

component that includes the AGN (smallest component) and the bulge of the host galaxy. The models are shown in Figure 5.6.

Table 5.8: The functions fitted to SAGE0536AGN, Source 10 and Source 16 in GALFIT and their parameters. All sources were fitted with one Sérsic function for the host galaxy and two Sérsic functions for the central AGN. A sky background object was also fitted, the only parameter of that used was sky background at centre of fitting region [ADUs], fitted prior to fitting the other models by setting an estimate of the background and allowing GALFIT to iterate and find the best value for the background. The magnitude is the total K_s Vega magnitude from the VMC survey. R_e is the effective radius in kpc (calculated from the redshift, conversion factors are 2.49 kpc arcsec^{-1} for SAGE0536AGN, 3.34 kpc arcsec^{-1} for Source 10 and 2.26 kpc arcsec^{-1} for Source 16), such that half of the total flux is within R_e. R_e for AGN is not meaningful since the AGN is not resolved. n is the Sérsic exponent. b/a is the axis ratio. The position angle is the angle the major axis, a, is orientated to. To account for the presence of spiral arms the Sérsic components of the host galaxy include PA rotation angle function. The bar radius is the radius where the rotation reaches roughly 20°. The 96% asymptotic radius is the radius at 96% tanh rotation. Rotation is the cumulative coordinate rotation out to the asymptotic radius. The asymptotic spiral arm power law is related to the rotation, $\theta \propto r^a$, where r is the radius and a is the power law.

	SAGE0536AGN			Source 10			Source 16		
Object type	AGN	Bulge	Host	AGN	Bulge	Host	AGN	Bulge	Host
Magnitude	13.1	13.2	12.7	14.9	16.1	14.5	14.6	14.7	14.1
R_e (kpc)	1.3	2.3	7.8	0.4	1.1	8.9	1.5	3.4	11.6
n	0.6	0.6	0.6	0.2	1.7	0.7	0.5	0.7	0.7
b/a	0.9	1.0	0.7	0.6	0.5	0.4	1.0	0.8	0.9
Position angle (degrees)	−6.5	51.7	77.8	−55.3	6.8	86.9	−19.8	−87.9	12.8
PA rotation func.	none	none	none	none	none	power	none	none	power
Bar radius (kpc)	-	-	-	-	-	16.1	-	-	3.9
96 % asymp. radius (pixels)	-	-	-	-	-	106.9	-	-	96.0
Rotation (degrees)	-	-	-	-	-	126.7	-	-	100.4
Asymp. spiral arm power law	-	-	-	-	-	−3.3	-	-	−5.2
Inclination to L.o.S. (degrees)	-	-	-	-	-	0.0	-	-	−47.0
Sky Position Angle (degrees)	-	-	-	-	-	7.7	-	-	67.5

The host galaxies of Sources 10 and 16 display rotation as a function of radius, as is seen in spiral galaxies. GALFIT allows for coordinate rotation in the light profile, and in this case I use the power law spiral function in conjunction with the Sérsic function to account for the spiral arms. The residuals of SAGE0536AGN also suggest the presence of spiral arms, though they were not fitted here.

When the Sérsic exponent, n, is large, it has a steep inner profile (cusp), and a highly extended outer wing. When n is small, it has a shallow inner profile (core) and a steep truncation at large radius. For the host galaxies of these three AGN, SAGE0536AGN has $n = 0.62$, implying a form between a Gaussian function ($n \sim 0.5$) and an exponential disk ($n \sim 1$). Source 10 has $n = 0.71$ and Source 16 has $n = 0.70$, implying host galaxies closer to an exponential disk than SAGE0536AGN.

Combining the AGN and bulge components of each source provides a total magnitude brighter than the host galaxy, as expected. The AGN appears unresolved in the images, so the $R_{\rm e}$ calculated by GALFIT for the Sérsic profile are not true values. A Sérsic exponent, $n \sim 0.5$ gives a Gaussian component, and a Gaussian component with $R_{\rm e} \sim 0.5$ pixels is an alternative for fitting a PSF profile and therefore an unresolved source, such as an AGN. Source 10 AGN has $R_{\rm e} \sim 0.5$ pixels, showing it is as expected an unresolved source. SAGE0536AGN and Source 16 both have $R_{\rm e} \sim 1.5$ pixels, implying the bulge is slightly resolved.

All three galaxies are brighter at redder wavelengths, implying dust and/or lack of star formation. However, these three galaxies also have the appearance of spiral galaxies. This could imply a recent shut-down of star formation and that the galaxy has yet to transition morphologically into an elliptical galaxy. This could mean green-valley galaxies.

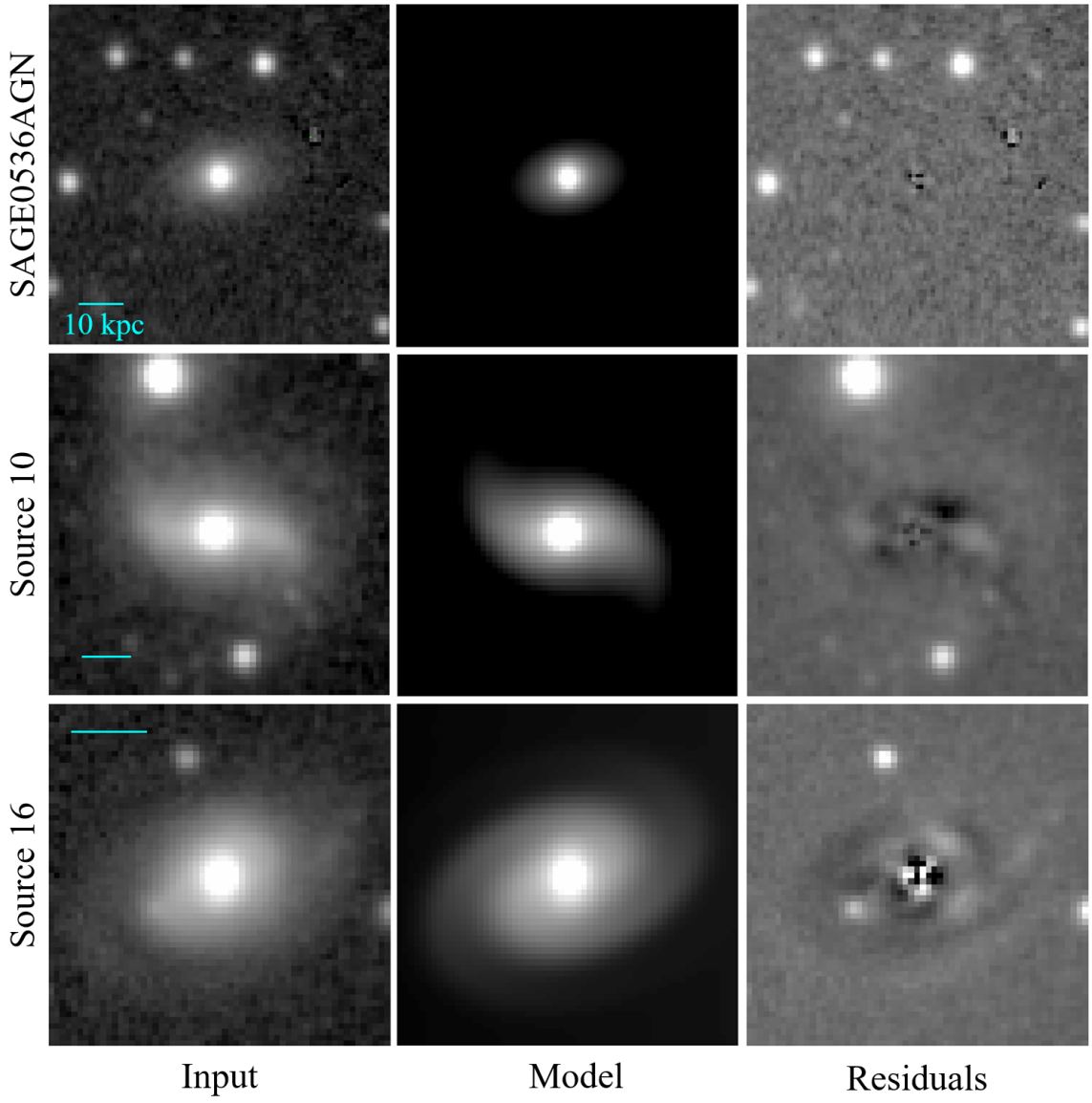

Figure 5.6: GALFIT models of SAGE0536AGN (top), Source 10 (centre) and Source 16 (bottom). The cyan line represents 10 kpc based on the respective redshift of each source. The input images are K_s band VMC images.

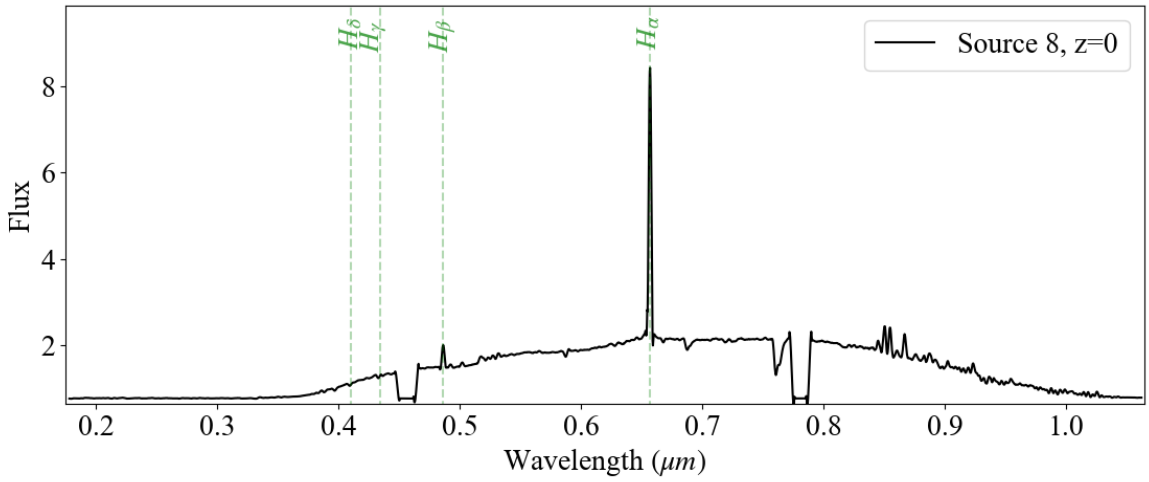

Figure 5.7: The spectrum of Source 8 observed with SALT, showing a redshift of near zero and emission lines at Hα and Hβ.

5.2 Discussion

5.2.1 Source 8's true nature

Since this research was published I observed Source 8 with SALT. The spectrum can be seen in Figure 5.7 , and reveals the source as an emission line star in the SMC. Whether the X-ray and radio emission is from this star, making it active source such as a symbiotic binary, or from a background extragalactic source is still unclear.

Work after this point has been updated to account for the new observations of Source 8.

5.2.2 IR properties and selection criteria

The infrared has proven to be an effective wavelength regime to select AGN in and therefore many selection criteria for AGN have been created for IR wavebands based on previously spectroscopically identified AGN.

AGN selection criteria have been created for Spitzer IRAC and WISE wavebands

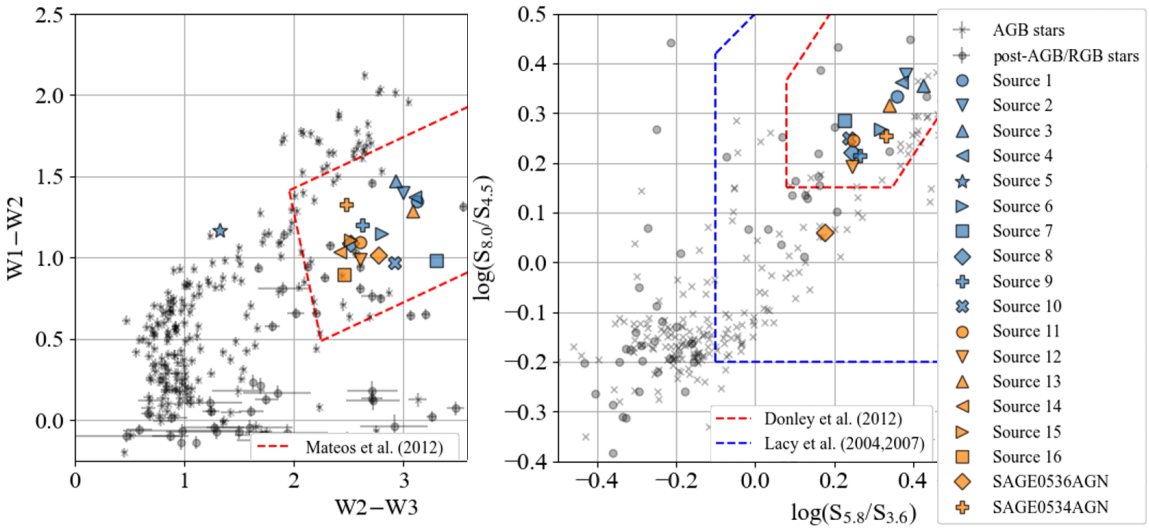

Figure 5.8: (Left) AllWISE colour–colour diagram of the sample. The sample is clumped together in the AGN region, apart from Source 5, a carbon star, which is to the left of AGN region and within the region populated by AGB stars. The red dashed line indicates the Mateos et al. (2012) AGN selection criteria. Spectroscopically observed AGB stars in the Magellanic Clouds have been added to both plots as grey points and show how the two classes can be mistaken for the other. (Right) IRAC colour–colour diagram for the sample. The red and the blue dashed lines indicates the Lacy et al. (2004) and Donley et al. (2012) AGN selection criteria, respectively.

by Lacy et al. (2004), Stern et al. (2005b), Mateos et al. (2012) and Donley et al. (2012). The Donley et al. (2012) wedge, shown in Figure 5.8 (right), was designed to be an improvement on the Lacy et al. (2004) and Stern et al. (2005b) wedges, as it excludes high-redshift star-forming galaxies whilst incorporating the best aspects of the previous AGN selection wedges. All but SAGE0536AGN fall within the Donley et al. (2012) wedge. SAGE0536AGN, however, does still fall within the Lacy et al. (2004) wedge. This implies that sources similar to SAGE0536AGN could potentially be missed by the Donley et al. (2012) wedge.

The Mateos et al. (2012) wedge, shown in the left panel of Figure 5.8 (left) shows that all sources fall well within the expected area in the colour–colour diagram. Also, as expected, Source 5 (the carbon star) falls outside this area. All the sources fall well

within the criteria from Stern et al. (2005b), except SAGE0536AGN which is only just within its bounds.

Cioni et al. (2013) created AGN selection criteria which are shown in Figure 5.9. They separate the colour–colour space into four regions. Regions A and B are where most known AGN are found where point-like AGN dominate region A and AGN with visible host galaxies dominate region B. The average redshift was found to be $z = 1.22 \pm 0.25$ in region A and $z = 0.44 \pm 0.25$ in region B. Region C was found to contain reddened Magellanic sources and region D was found to contain stars and low-confidence AGN. As expected, none of our sources are found in region C. Source 9, however, was found unexpectedly in region D, where stars dominate. The three sources with clear host galaxies, Source 10, 16 and SAGE0536AGN are found, as expected, in region B. All sources that have CIGALE fits with $f_{\rm AGN} < 0.90$ are found in region B and all of the sources found in region A have $f_{\rm AGN} > 0.90$. Sources 11, 13 and 14 are also found in region B, despite the predicted $f_{\rm AGN} > 0.90$. Source 5 is found at $Y - J = 1.4$ and $J - K_{\rm s} = 4.6$ (not on the diagram), when it would be expected to be found in region C or D. From the sample the average redshifts for A and B are, $z \sim 1.02$ and $z \sim 0.35$, respectively, as expected.

AGB and post-AGB/RGB stars, classes known for being confused with AGN, have been added to the plots. These stars have all been spectroscopically observed (Groenewegen & Blommaert, 1998; van Loon et al., 1998, 1999a,b, 2005, 2006, 2008; Kamath et al., 2014) and are all in the Magellanic Clouds. In Figure 5.8 (right) the locus of the AGB stars is outside the two AGN selection criteria, however some are still found within the Lacy wedge, some of which are avoided with the Donley wedge. In Figure 5.8 (left) most of the stars are outside the Mateos wedge. Of those that encroach on the AGN criteria, AGB stars are in the top of the wedge, whilst post-AGB/RGB stars are at the bottom of the wedge. In Figure 5.9 AGB stars can be mostly found in region B, whilst post-AGB/RGB stars can be found in region A. For all colour–colour diagrams AGB and post-AGB/RGB stars can be found amongst the AGN sample. It is known that combining near-IR and mid-IR selection techniques can efficiently select a high number of AGN (Bornancini et al., 2022), and the combination of WISE and

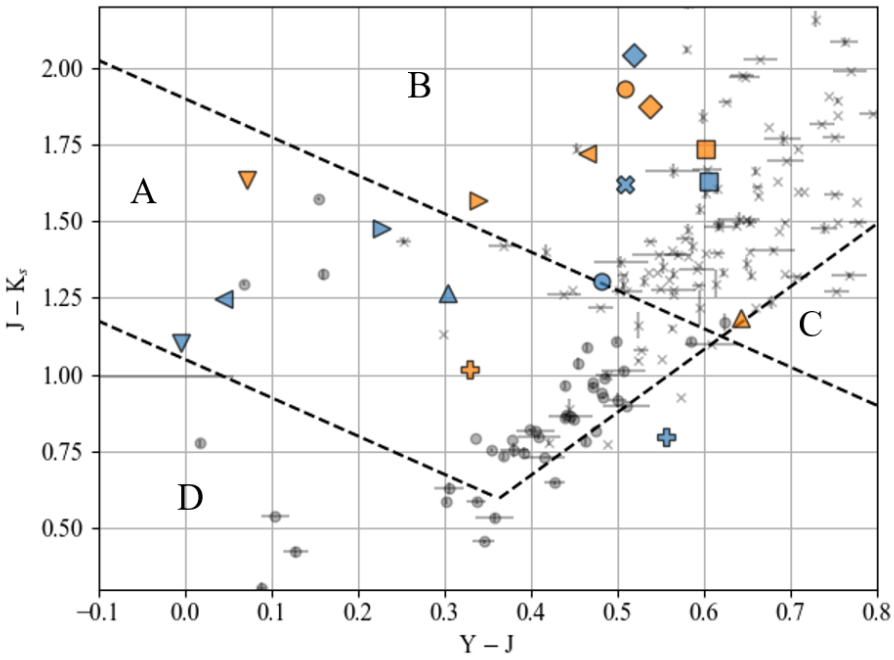

Figure 5.9: VISTA colour–colour diagram of the sample. The regions A, B, C and D were devised by Cioni et al. (2013). Regions A and B are where most known AGN are found and are well matched by the models. Region C is dominated by reddened Magellanic Cloud sources and region D is populated by stars and low confidence AGN. See Figure 5.8 for the sample identifications.

VMC colour selection techniques has the potential to efficiently remove the AGB and post-AGB/RGB sources.

Note that Source 8, despite being revealed as an emission-line star, appears well within the selection criteria for AGN in VISTA, SAGE and AllWISE colour selections.

5.2.3 Green valley

To determine whether the sources of the t-SNE selected sample are blue star-forming, green-valley or quiescent galaxies, they can be plotted on a diagram of star-formation rate (SFR) versus stellar mass of the host galaxy (e.g. Chen et al., 2016; Belfiore

et al., 2018).

The SFR was calculated using the CIGALE best fits (after subtracting the AGN components) by using the correlation between total luminosity between 8 μm and 1000 μm and SFR as shown in Bell (2003):

$$\mathrm{SFR}(M_{\odot}\mathrm{yr}^{-1}) = \begin{cases} 1.57 \times 10^{-10} L_{\mathrm{TIR}}(1 + \sqrt{\frac{10^9}{L_{\mathrm{TIR}}}}) & , L_{\mathrm{TIR}} > 10^{11} \\ 1.17 \times 10^{-10} L_{\mathrm{TIR}}(1 + \sqrt{\frac{10^9}{L_{\mathrm{TIR}}}}) & , L_{\mathrm{TIR}} \leq 10^{11} \end{cases} \tag{5.4}$$

where L_{TIR} is the total luminosity between 8 μm and 1000 μm in solar luminosities. The stellar mass was calculated by using the correlation between black hole mass and stellar mass as described in Häring & Rix (2004):

$$\log(M_{\mathrm{BH}}) = -4.12 + 1.12 \log(M_*) \tag{5.5}$$

where M_* is the stellar mass. The resulting diagram can be seen in Figure 5.10 (centre). The CIGALE output for SFR was not used as most of the sources are dominated by the AGN, meaning the host galaxy, and therefore the SFR, could not be modelled accurately. For these sources the calculated SFR is an upper limit.

This plot shows that 12 out of 17 of the sources are green-valley galaxies. However, the far-IR fluxes from these galaxies are upper limits, meaning the SFR could possibly be lower and therefore in the quiescent region. SAGE0536AGN and Source 16 show spiral arm structure, meaning star formation shut off recently and they are at least more likely to be green-valley galaxies rather than quiescent.

Of the sources in the star-forming region, Source 6 may be a green-valley galaxy as its SFR is an upper limit. Of the other galaxies it is possible that the far-IR emission is not due to star formation. The far-IR from these galaxies could instead be accounted for by dust heated by the AGN beyond the torus. It has been shown that for torus opening angles of 20 – 70° (Zhuang et al., 2018) the AGN emission will heat dust in the narrow-line region (where the polar dust is) if the black hole accretion disk is aligned with the galaxy plane (Baron et al., 2016), or the dust in the host galaxy if the accretion disk and the galaxy plane are misaligned (Viaene et al., 2020).

We compare other observed characteristics of these sources and CIGALE model outputs with the distance along the potential evolutionary sequence, which is defined

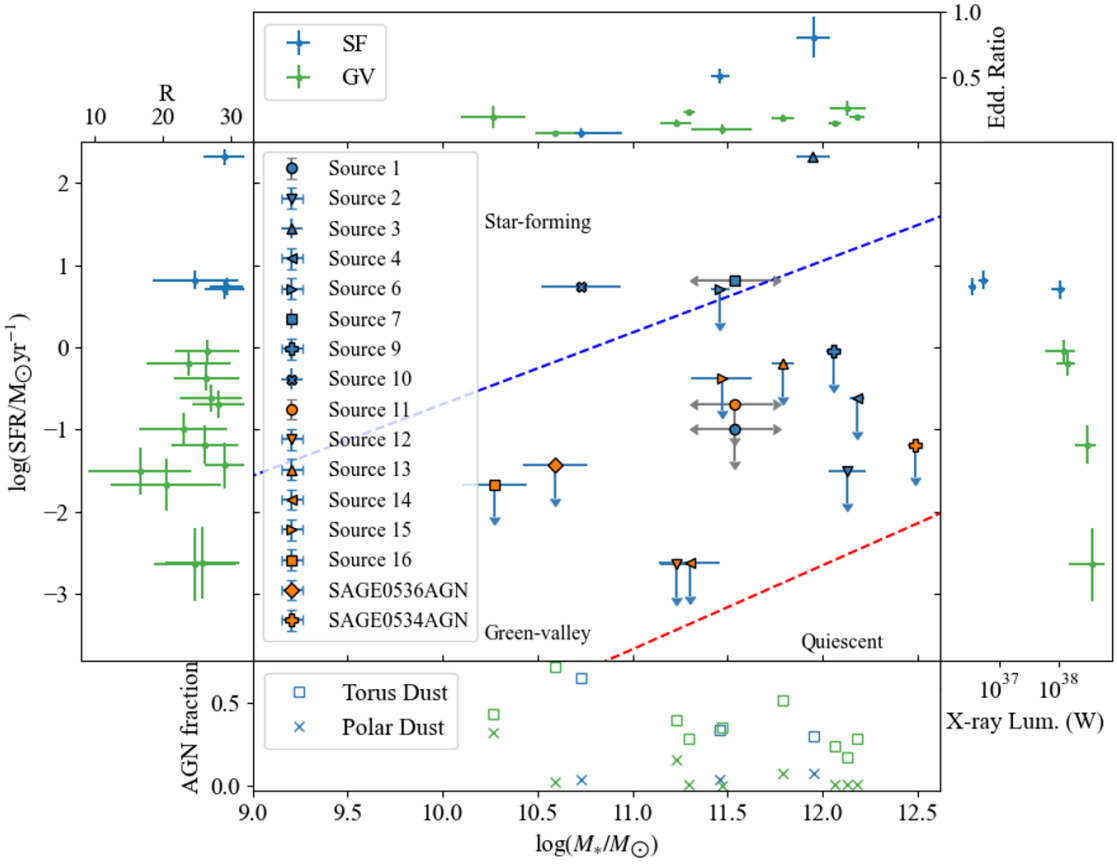

Figure 5.10: (centre) The position of the t-SNE selected sample on a SFR versus stellar mass diagram. The blue and red dashed lines are an approximation of the boundary (at the 1σ level in scatter from the main trend lines of the star-forming (SF) and quiescent galaxies) of the star-forming main sequence and the quiescent sequence, respectively, taken from Chen et al. (2016). Those with blue error bars are sources with known black hole masses to calculate stellar mass from. Those with grey error bars are those with no known black hole mass, set at the average black hole mass of the sample. (top) How Eddington ratio changes with M_*. (left) How R changes with SFR. This shows that green-valley (GV) galaxies tend to have smaller tori than star-forming galaxies. (right) How X-ray luminosity changes with SFR. This shows that higher X-ray luminosities are seen in the green valley. (bottom) How the torus and polar dust emission fractions of the AGN change with M_*. Accretion disk emission is the remaining fraction of AGN not plotted here.

as the distance of the source in the host mass versus SFR plane from the 1σ scatter from the star-forming galaxies main sequence line.

From this it can see that the AGN fraction increases along the potential evolutionary sequence, which is expected as the host galaxy star formation reduces and the AGN becomes more prominent. Also as expected, the BH mass increases along the potential evolutionary sequence.

Source 3 is the furthest above the star-forming main sequence and has the largest Eddington ratio. The upper limit to Eddington ratio seems to decrease along the potential evolutionary sequence, this could imply its running out of fuel at later stages.

X-ray observed AGN are more likely to be found in the green valley than in the star-forming region (e.g. Treister et al., 2009; Pović et al., 2012, 2013). This is corroborated by the most X-ray luminous of the AGN being found in the green valley. X-ray luminosity increases along the potential evolutionary sequence, which could correspond with decreasing R (ratio of outer torus radii to inner torus radii) and torus fraction, implying a thinning torus, which would mean that there is less dust and gas to absorb the X-ray emission. However, in this sample the majority of sources are being seen close to face-on, meaning the torus thickness is less likely to affect the X-ray emission. Source 7 is predicted to have an inclination angle of $49°$, which would mean the torus would be affecting the X-ray emission, and yet it has the second highest X-ray luminosity. Another possible interpretation is that the X-ray emission is tracing the AGN activity, and we are seeing the AGN activity increase as the sources move from the star-forming region to the green valley region.

5.2.4 Radio analysis

5.2.4.1 Radio morphology

From the radio continuum images taken with ASKAP of the SMC and LMC, all 18 sources appear compact (unresolved at ASKAP resolutions) apart from three: Sources 6, 15, 16, which are shown in Figure 5.11. Sources 6 and 16's extended nature could

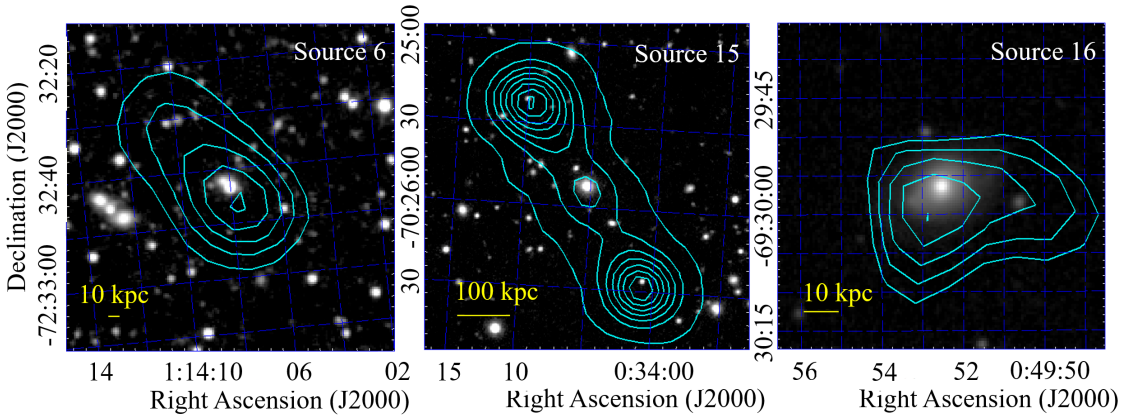

Figure 5.11: ASKAP 1320 MHz radio flux of Sources 6, 15 and 16 shown as contours on top of the VMC K_s band images. Source 6 has five linearly spaced contours from 0.18 – 1.1 mJy. Source 15 has eight linearly spaced contours from 1 – 70 mJy. Source 16 has five linearly spaced contours from 0.3 – 0.72 mJy. Source 6 could potentially be an FR II source, with an offset radio peak and a counter-lobe at the other side. Source 15 is an FR II source. Source 16 radio emission extends towards the West most likely due to the point source to the West of it also being a radio source.

be caused by radio emission from nearby sources blending with the main source.

Source 15's radio lobes would imply we are not observing the source close to face-on. The CIGALE model predicts $i \sim 0.4°$, and therefore very close to face-on. However, the radio image in Figure 5.11 also shows a bright centre to the source. This could mean that the lobes are old relics, since then the source has rotated, and is now emitting a face-on radio jet.

5.2.4.2 Spectral Indices

We define the spectral index α by $F_\nu \propto \nu^\alpha$, where F_ν is the integrated flux density at frequency ν. A flatter spectral index close to zero indicates free–free emission, and a steep negative spectral index of approximately -0.7, indicates synchrotron emission. Table 5.9 shows the radio luminosity and radio MHz spectral indices, α.

Figure 5.12 shows the distribution of spectral indices for our sample compared to

Table 5.9: Radio luminosities and spectral indices for the t-SNE selected sample from ASKAP (888 MHz for LMC, 960 MHz for SMC).

Source Name	Luminosity (10^{23} W)		α
	888/960 MHz	1320 MHz	
SAGE0536AGN	0.5± 0.1	–	–
SAGE0534AGN	0.1± 0.1	–	–
1	0.1± 0.1	0.1± 0.1	−0.34
2	99.7±14.2	69.8± 5.0	−1.05
3	123.1±24.8	77.8± 7.6	−1.46
4	179.8±18.0	142.1± 7.2	−0.69
6	115.1± 9.6	109.9± 5.3	−0.1
7	0.6± 0.2	0.4± 0.1	−1.12
8	6.4± 1.6	3.6± 0.6	−1.73
9	113.9±10.4	89.6± 3.1	−0.69
10	0.9± 0.2	0.9± 0.1	−0.38
11	7.5± 1.7	5.9± 0.5	−0.67
12	18.7± 1.9	11.4± 0.8	−1.58
13	44.6±15.9	35.1± 6.4	−0.93
14	11.0± 1.8	6.5± 0.6	−1.67
15	–	340.3± 4.8	–
16	–	0.6± 0.2	–

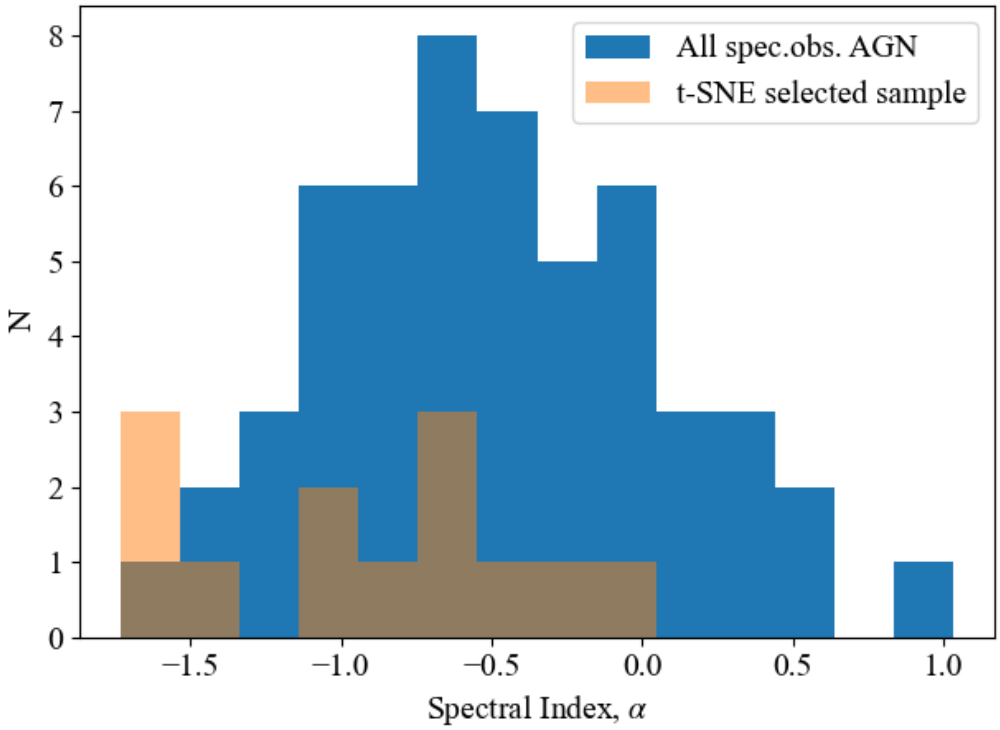

Figure 5.12: Spectral indices of the sample compared to other spectroscopically observed AGN behind the Magellanic Clouds.

the spectroscopically observed AGN behind the Magellanic Clouds. This shows that whilst there is a peak at $\alpha \sim -0.7$, coinciding with the expected value for synchrotron radiation, there is also an unexpected peak at more negative values, where the most extreme radio spectral slope belongs to Source 8.

5.2.4.3 Radio properties across the green valley

We compare the radio properties of our sources with the distance along the potential evolutionary sequence, which is defined as the distance of the source in the host mass versus SFR plane from the 1σ scatter from the star formation main sequence line (Figure 5.13).

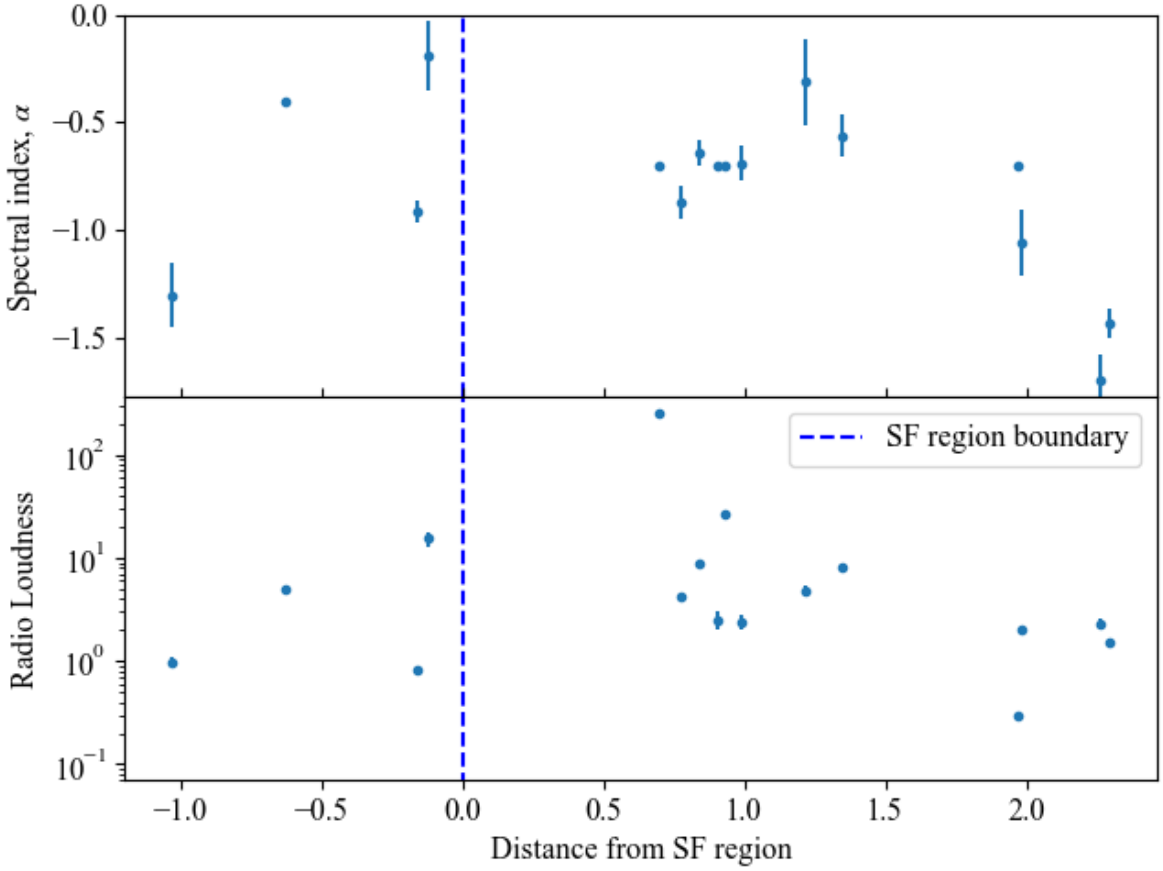

Figure 5.13: Variation of spectral indices, α (top), and radio loudness (bottom) of the sample as they transition from star-forming to green-valley galaxies. It should be noted that many of the plotted points are lower limits in the x-direction.

From Figure 5.13 (top) it can see that the spectral index is steep at the start of the potential evolutionary sequence, is flattest at the beginning of the green-valley section and then steepens again. Removing Source 8 from this graph has improved the trend at the beginning of the potential evolutionary sequence with respect to spectral index, as Source 8 has a more negative index than Source 3 (first source on the left) but was estimated to be slightly further along the potential evolutionary sequence, causing a dip in the trend. Figure 5.13 (bottom) shows that radio loudness, $R_{\rm AGN}$ (see Table 3.3), is lowest at the start of the potential evolutionary sequence, reaches a peak at the beginning of the green-valley section, and then reduces again.

Sources 3 (source on the far left of Figure 5.13) could be a Compact Steep Spectrum (CSS) source, which are young sources that could go on to become large-scale Fanaroff–Riley II (FR II) objects (Fanti et al., 1995), such as Source 15. Furthermore, an observational signature of an AGN "switching-off" is also a steep spectrum ($\alpha < -1.5$). This is due to plasma ejected from the AGN losing energy causing high energy particles that radiate mostly at high radio frequencies to lose their energy fastest, making radio emission strongest at lower frequencies and causing a steep spectrum to be observed. This could imply that the AGN of the sources at the beginning of the potential evolutionary sequence have just switched on, explaining their steep radio spectrum. As the sources transition into the green valley the sources are at their radio loudest and have spectral indices ~ -0.7, implying steady synchrotron emission, after which the AGN, and subsequently the radio emission switches off, causing the radio loudness to decrease and the spectral index to steepen. The overall implication is that the AGN trace the transition from star-forming, across the green valley and into quiescence.

Another interpretation of these results could be that AGN triggered in quiescent galaxies have a different nature to those triggered in star-forming galaxies.

It should be noted that these interpretations are based on visual appearance rather than statistical tests. Statistical tests will be implemented at a later date once data from the LMC had been added to these plots, which will increase the sample size significantly, and from which more certain conclusions can be drawn. Furthermore, once

the LMC sample has been added, an analysis comparing the timescales for evolution of the radio properties with the timescales for colour evolution of galaxies after star-formation ceases would be worth exploring.

5.2.5 AGN dust properties

5.2.5.1 Variability

All the sources show some variability. From those with a known Eddington ratio it can seen that high Eddington ratio sources tend to have little variability, while those with decreasing Eddington ratios tend to include sources with larger variability (Figure 5.14). This relation holds true for black hole mass in place of Eddington ratio, so therefore smaller black holes (with smaller tori) vary the most whilst larger black holes (with larger tori) tend to vary less.

There is a general increase in variability with an increase in optical depth, implying the more emission from the accretion disk that is absorbed by the torus, the greater the variability. The Eddington ratio also decreases with increased optical depth and torus fraction. This implies Eddington ratio increases with less attenuation by the torus, as expected, but also that variability decreases with less attenuation by the torus, implying the torus could be playing a part in the variability of the AGN that we observe. This could mean we are seeing variability in the attenuation of the emission from the accretion disk, instead of the variability of the accretion disk emission, which could mean high variability is caused by a "clumpy" torus moving around the accretion disk causing the amount of attenuation of the emission to increase and decrease.

High Eddington ratio sources vary the least. In general the highest Eddington ratios are at the start of the potential evolutionary sequence, whilst the lowest Eddington ratios are at the end of the green valley. This could imply that the sources at the beginning, where the AGN is just "switching on" and have the greatest amount of fuel and a high accretion rate, have the lowest variability, whilst the AGN that are starting to "switch off" and have the lowest amount of fuel and a lower accretion rate, have the

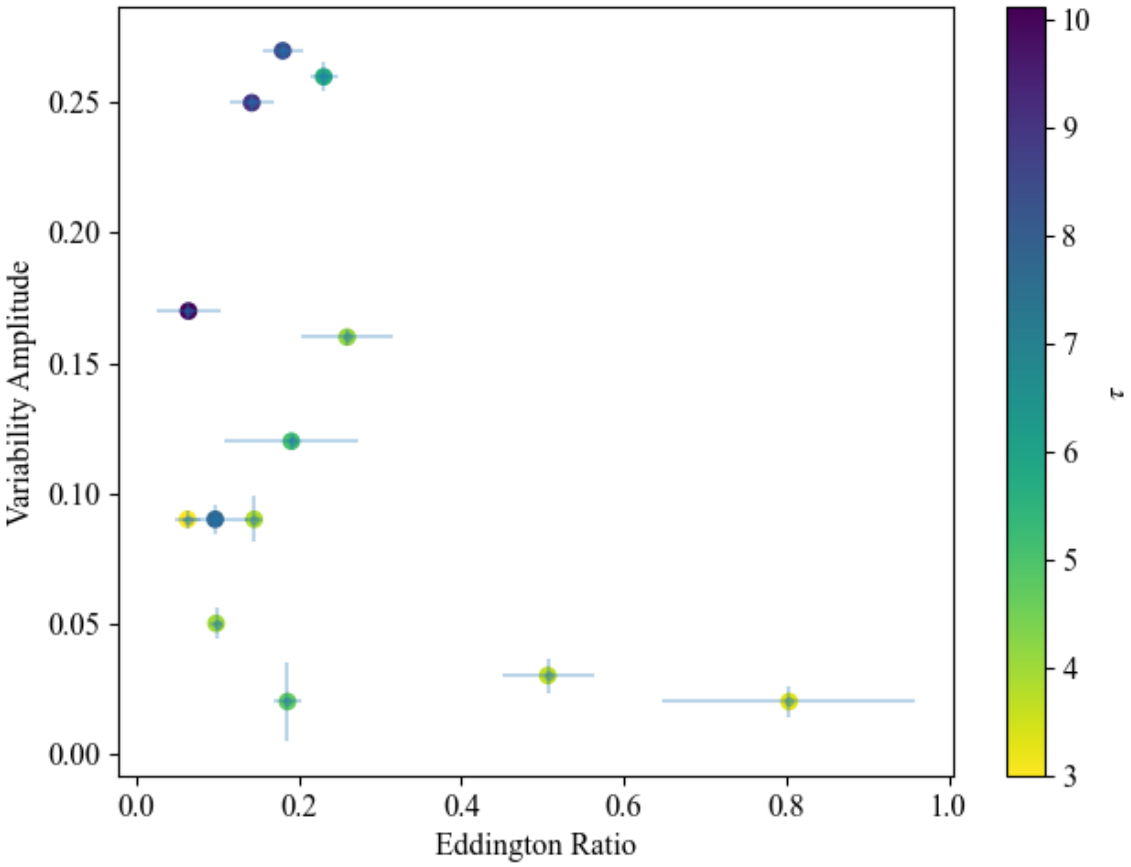

Figure 5.14: The comparison of variability with Eddington ratio and optical depth, τ, of the t-SNE selected sample.

highest variability. This could be due to the erosion of the dusty torus surrounding the AGN leading to a more porous torus and therefore increased variability. This would however be dependent on inclination angle of the AGN.

5.2.5.2 Cigale model components

CIGALE provides the separate models that make up the overall best-fit model. These models can be seen in Figure 5.16. In the optical, the sources where the host galaxy dominates over the AGN disk are SAGE0536AGN and Source 7.

Polar dust contribution varies from source to source. Source 7 has the highest polar dust fraction at $\sim 53\%$. This source also has one of the lowest AGN fractions of $\sim 72\%$. However, the source with the lowest AGN fraction of $\sim 70\%$, only has polar dust fraction of $\sim 3\%$. Those with $< 1\%$ polar fraction can all be found in the green valley. These could be those where no outflows are present to push out the polar dust, and the AGN is starting to turn off.

As expected, we see a negative correlation between accretion disk fraction and torus disk fraction (Figure 5.15), where SAGE0534AGN has the highest disk fraction and SAGE056AGN has the highest torus dust fraction, bracketing the sample. The opening angle also follows this trend, increasing with torus dust fraction (and decreasing with accretion disk fraction). However, Source 16, and to a larger extent, Source 7, veer off from these correlations due to their increased polar dust fraction. Their decrease in opening angle leads to an increase in polar dust fraction, which could be due to the increased space available at the poles with a smaller opening angle, as well as the poles being less obscured by the torus.

In general polar dust fraction increases with decreasing R and oa, and increases with i. Note that Table 5.6 shows that Sources 7 and 16, that have the highest polar dust fraction, also have the highest values for i, which implies a link between the narrow-line region and polar dust fraction. The relation with R could imply the presence of an outflow. This outflow would push out polar dust to become observable, increasing the polar dust fraction, as well as erode the dusty torus, decreasing R, which then in turn

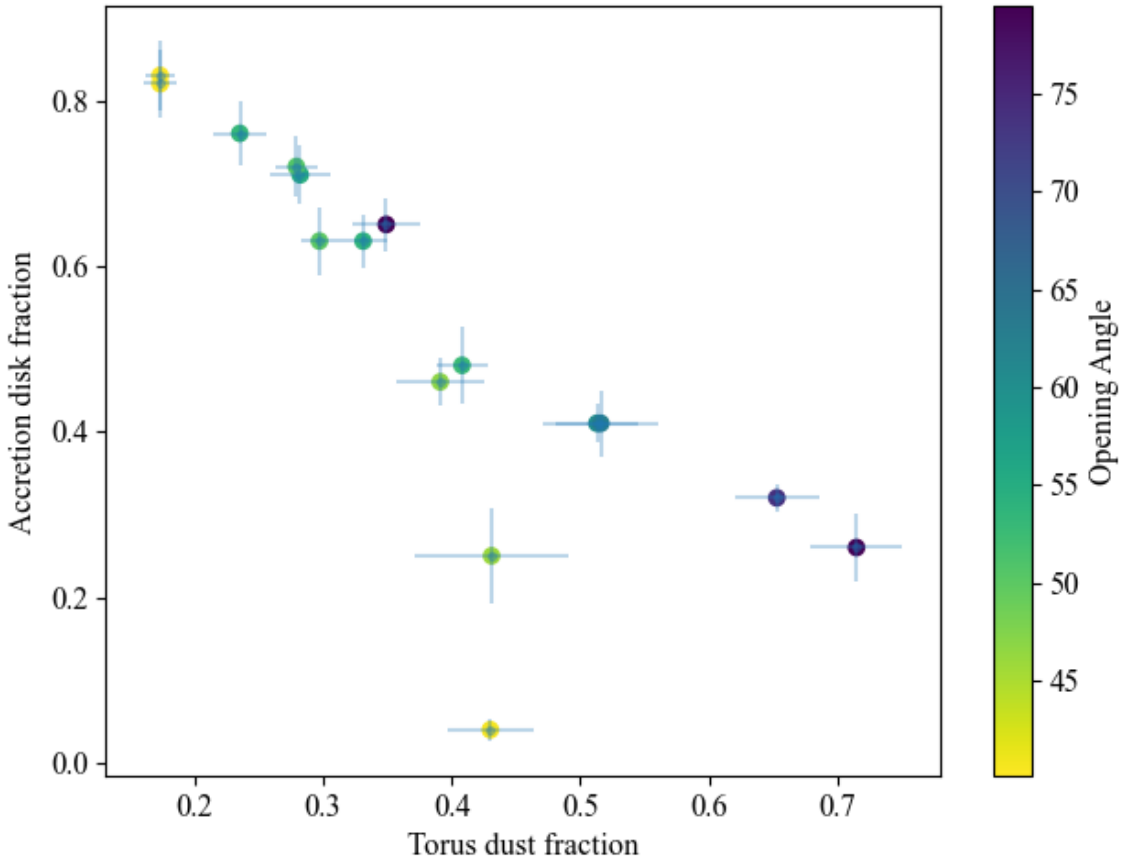

Figure 5.15: Accretion disk fraction versus torus disk fraction for the t-SNE selected sample. The colour bar represents the opening angle, *oa*. Sources 16 and 7 veer off from the negative correlation shown due to increased polar fraction, which also correlates with a decreased opening angle.

reveals more of the polar dust. The polar dust fraction would then also be expected to increase with i: as the accretion disk becomes more obscured by the dusty torus, then the torus dust and polar dust being pushed out by an outflow would become more prominent. As *oa* decreases the view into the centre of the AGN opens, increasing the space over which polar dust can be found, thus increasing polar dust fraction.

5.2.5.3 Silicate 9.7μm dust

The prominence and peak wavelength of the silicate emission of AGN varies. To quantify the strength of this emission it is defined as the silicate peak relative to the continuum, at the wavelength where the silicate feature peaks (Hao et al., 2007):

$$Si_{9.7\mu m} = \ln \frac{f_{9.7\mu m}(\text{peak})}{f_{9.7\mu m}(\text{continuum})} \tag{5.6}$$

The silicate strength of SAGE0534AGN is calculated to be 0.24 $\pm$ 0.04. In comparison, SAGE0536AGN yields a silicate strength of 0.85 $\pm$ 0.13.

CIGALE models the silicate dust as part of the AGN modelling. Calculating the silicate strength for all the sources as done previously with SAGE0536AGN and SAGE0534AGN revealed that silicate strength seems to increase with redshift. However, while the model correctly predicts that the silicate 9.7-μm feature is in emission for SAGE0536AGN and SAGE0534AGN, the model underestimates SAGE0536AGN (low-z source) as $\text{Si}_{9.7\mu\text{m}} \sim 0.54$ and overestimates SAGE0534AGN (high-z source) as $\text{Si}_{9.7\mu\text{m}} \sim 0.72$. This could imply the modelling of the silicate feature strength is not accurate or missing something.

The model predicts that all sources in this sample show silicate 9.7-μm emission. This could mean that the t-SNE selection separates those sources in emission from those in absorption. Confirming this requires follow-up mid-IR spectroscopy.

5.2.5.4 Comparison to other silicate emitting AGN

SAGE0534AGN and SAGE0536AGN are not alone in their emission of silicate features. Comparing the two with more common less extreme versions of silicate emitting AGN

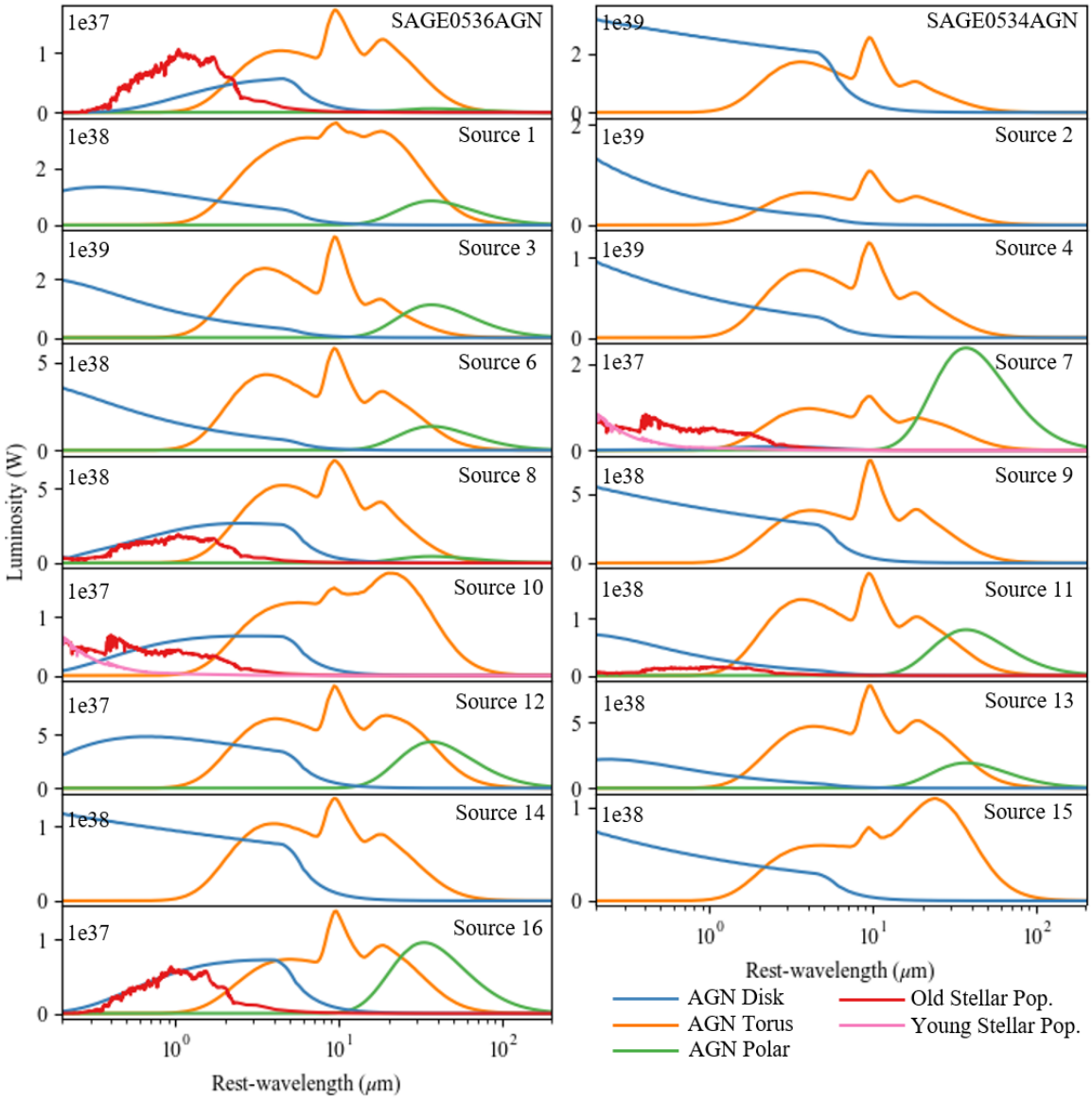

Figure 5.16: Separate components of the CIGALE models for the t-SNE sample at rest-wavelength. The fit of Source 8 is still included here to the comparison to the other models of true AGN.

may lend a clue to how these came about, whether they be extreme versions of an already established class of AGN or exist in a class of their own. To compare, a sample of local ($z < 0.1$) type 1 AGN with silicate emission were taken from Martínez-Paredes et al. (2020). They are a sample of 67 local ($z < 0.1$) type 1 AGN. Another comparison was made with a sample from Dicken et al. (2014) which include 46 2Jy radio galaxies ($0.05 < z < 0.7$) and 17 3CRR FRII radio galaxies ($z < 0.1$) nuclei (AGN) with Spitzer spectra dominated by non-stellar processes. The sources in this sample have silicate strength calculated using Equation 5.6.

The silicate strength of SAGE0534AGN, SAGE0536AGN and the silicate emitting AGN sample were compared with their far-IR colour (WISE 4 (23 μm) – IRAS 60 μm). For the sources in these samples that had no IRAS 60-μm measurements, a limit on the flux was calculated from the IRAS images. For the t-SNE selected sample the IRAS 60-μm magnitudes were estimated from the CIGALE best fit models. This comparison is shown in Figure 5.17. The AGN sample are all to the right from SAGE0536AGN, SAGE0534AGN and the t-SNE selected sample. The higher redshift galaxies tend to be further to the right, but this could be because high-z galaxies are most likely biased towards star-forming galaxies to make them bright enough. The limits on W4 – IRAS60 could suggest there are already interesting sources observed. SAGE0536AGN however, remains apart due to its high silicate strength.

SAGE0536AGN is the strongest 10-μm silicate emitter currently known. In terms of torus properties predicted by CIGALE, SAGE0536AGN has one of the highest values for R, the largest values for *oa* and torus fraction, lowest value for optical depth, τ and one of the lowest values for inclination angles. All this together could have provided the necessary environment for strong silicate emission. In contrast, SAGE0534AGN's silicate strength was overestimated. CIGALE predicts SAGE0534AGN has a similar inclination angle and optical depth to SAGE0536AGN, but the smallest values for *oa* and torus fraction. Both of these sources also have very little polar dust, the presence of which correlates with weak or absent silicate emission (e.g. Tazaki & Ichikawa, 2020). This could imply that the increased silicate emission strength is due to a thicker torus with little to no polar dust to obscure the centre of the AGN. Of the rest of the t-SNE

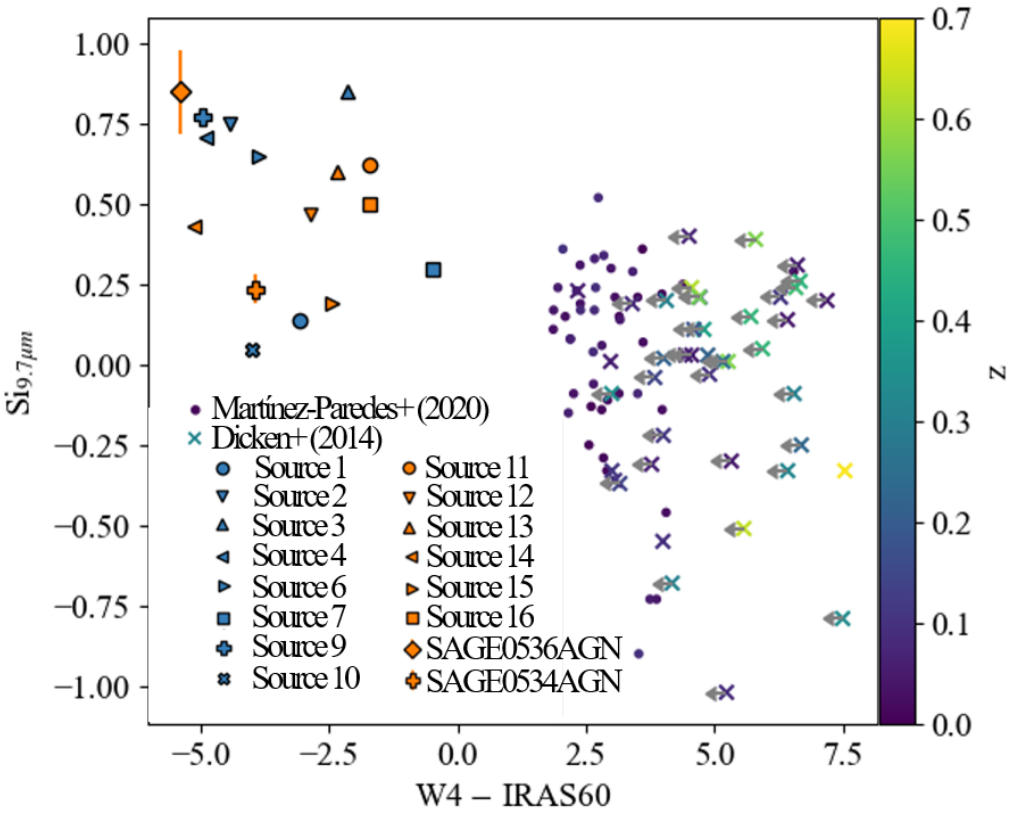

Figure 5.17: Silicate strength versus WISE band 4 – IRAS 60-μm (in magnitudes). The IRAS60 measurements for SAGE0536AGN, SAGE0534AGN and the t-SNE sample are all calculated from the CIGALE model best fit.

selected sample, the closest to SAGE0536AGN in terms of *oa*, torus fraction, polar fraction and inclination angle is Source 15, however the values for τ and R are not as close. This could mean that the silicate emission of this source may rival that of SAGE0536AGN.

5.3 Conclusions

In this work I used unsupervised machine learning, t-SNE, with Gaia EDR3, VMC, AllWISE and EMU ASKAP photometric data, to find sources similar to SAGE0536AGN, the strongest 10-μm silicate emitter known, and SAGE0534AGN, a similar source with weaker silicate emission. This provided 16 sources to add to the sample. I took optical spectra of 15 of these sources and found that all but one were extragalactic in nature. From these spectra I calculated black hole masses and Eddington ratios. I used CIGALE to model the SEDs and characterise these sources, and GALFIT to model the morphology of the three nearest sources.

From this investigation I discovered most of the sources (12 out of 17) are in the green valley transitional phase, with the potential for some of these to be quiescent. I find that as sources move away from the star-forming phase and through the green valley phase the properties of the AGN change, such as the torus depletes, the Eddington ratio decreases, signalling the AGN is running out of fuel, and the X-ray luminosity increases as the material that would absorb it has depleted. Radio properties also change across this potential evolutionary sequence. The radio spectral slope starts off steep in the star-forming phase, before flattening to the expected value of $\alpha \sim -0.7$ at the beginning of the green valley, and then steepening again as the sources move further into the green valley. Radio loudness also follows this trend, starting off quiet in the star-forming phase, becoming loudest at the beginning of the green valley, before quietening again. This implies the "turning on" of the AGN to transition from star-forming to green valley, and then the AGN "turns off" again, before transition to quiescence.

All sources are variable and this variability decreases when there is less attenuation by the torus, implying the torus is playing a part in the variability.

SAGE0536AGN remains the most extreme 10-μm silicate emitter, which is not modelled well with CIGALE, which predicts weaker emission for SAGE0536AGN and stronger emission for SAGE0534AGN. CIGALE predicts all sources have silicate in emission. This needs to be verified by spectroscopic observations in the mid-IR, such as with the James Webb Space Telescope.

6 Machine learning identification and classification of sources in the VISTA survey of the Magellanic Clouds

This chapter expands on the work done in previous chapters, but instead of looking at a small sample of sources, like in Chapter 5, looks at the entirety of the Magellanic Clouds that were observed in the VMC survey. The t-SNE algorithm is used on the radio sources of the Magellanic Clouds, as done in Chapter 4, and the different source classifications are observed. The PRF algorithm (see Section 3.4.2.3) is used on the entire VMC SMC data set cross-matched with other surveys (the PRF will be used on the LMC data set in a future paper).

6.1 t-SNE analysis of the Magellanic Clouds

In Chapter 5 I explored a small sample of sources dominated by nuclear dust emission, discovered using a machine learning algorithm, t-SNE, on a combination of photometric VMC, Gaia, AllWISE and ASKAP EMU survey data. Now that I also have a radio map of the LMC (see Chapter 4) I can proceed to do the same for the LMC. In this section I explore the rest of the SMC t-SNE map, as well as the t-SNE map produced from LMC data using the same photometric surveys as for the SMC. The LMC t-SNE map can can be seen in Figure 6.1.

6.1.1 Comparing the SMC and LMC t-SNE maps

In this section I compare how the t-SNE algorithms cluster the sources in the SMC and LMC t-SNE maps.

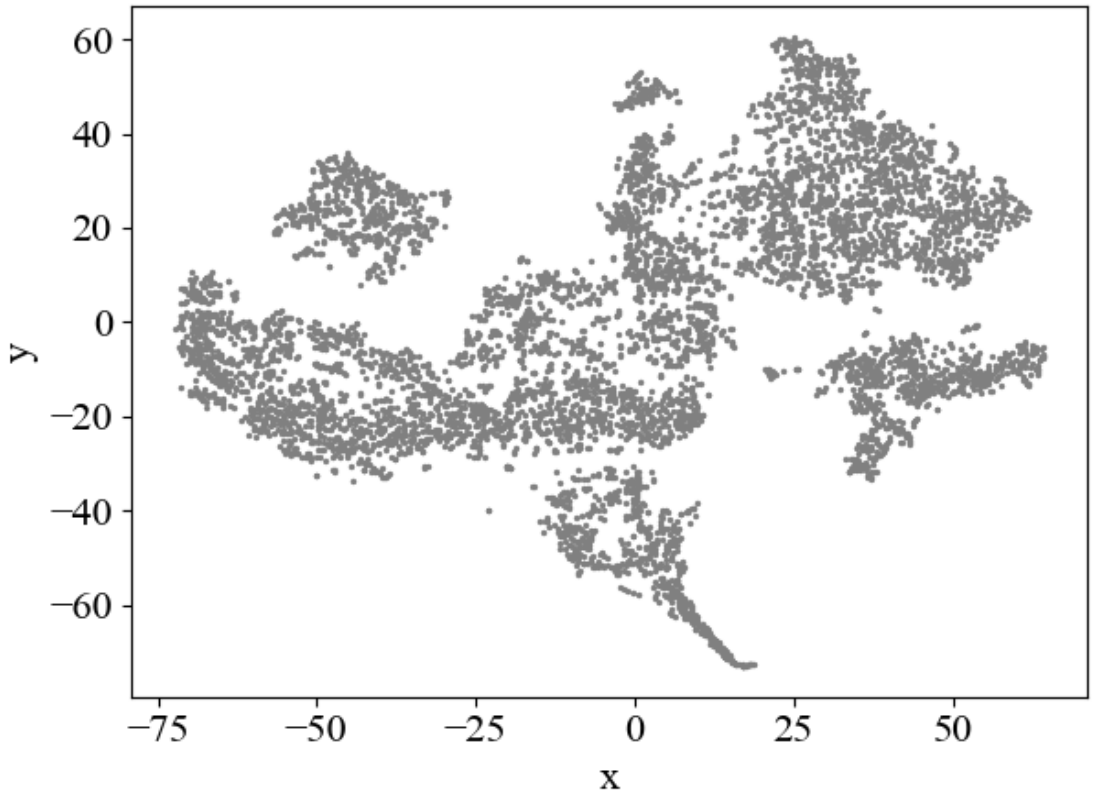

Figure 6.1: t-SNE map of the LMC.

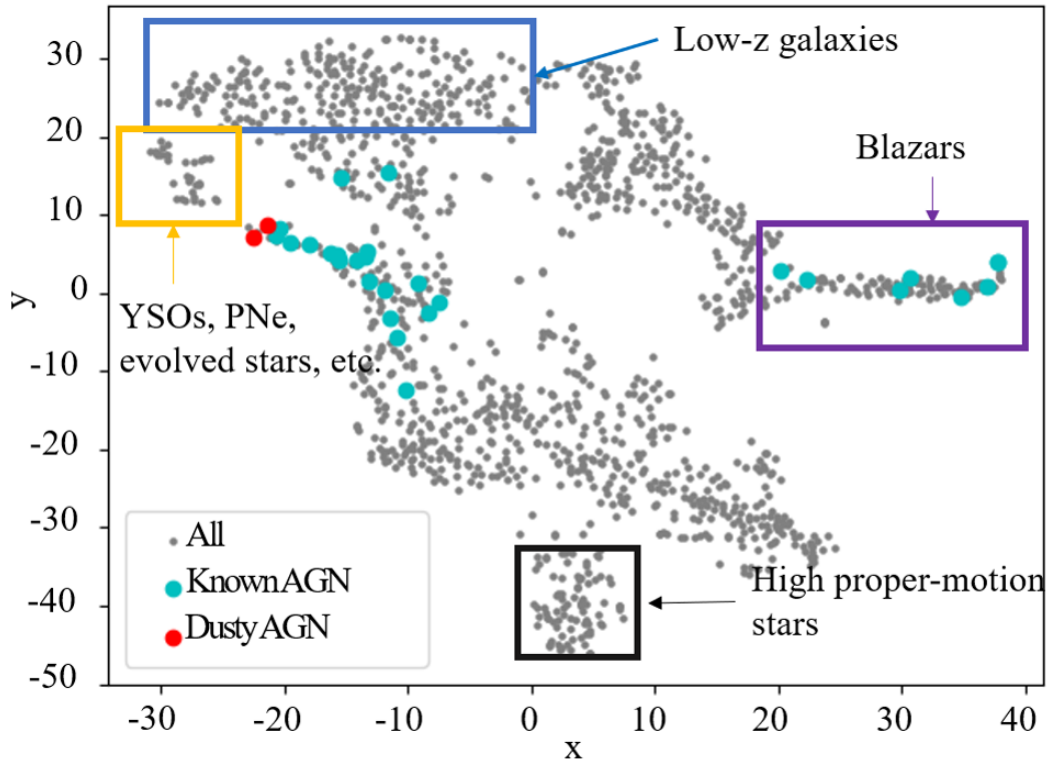

Figure 6.2: t-SNE map of the SMC. The highlighted regions show the general area of certain classifications of sources.

6.1.1.1 Spectroscopically confirmed classifications

Using classifications from the literature we can loosely classify the rest of the t-SNE map of the SMC and LMC. To do this I used classifications from Simbad (Wenger et al., 2000). From the map of SMC and LMC, see Figure 6.2 and Figure 6.3, we can loosely classify regions of sources.

High proper-motion stars are outlined in black and are an isolated group of sources in both maps. These are all most likely stellar sources in front of the true radio sources. This result shows that this technique could be a good way to remove false cross-matches between radio surveys and shorter wavelength photometry surveys.

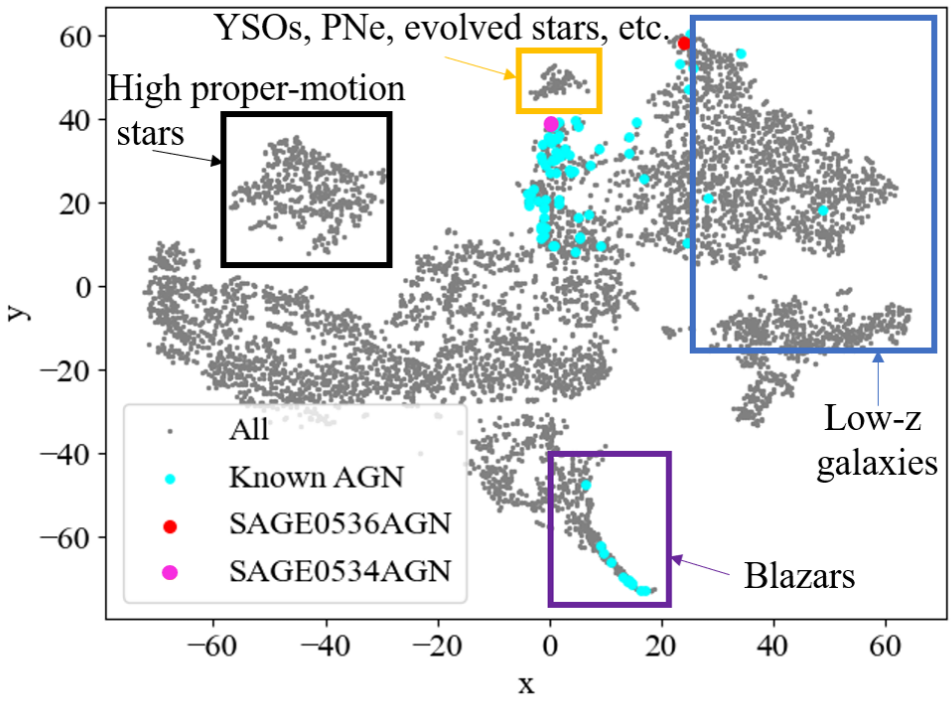

Figure 6.3: t-SNE map of the LMC. The high-lighted regions show the general area of certain classifications of sources. Note that the dusty AGN from Figure 6.2 have been separated here.

Some of these, however, could possibly be one of the rare radio active stars (e.g. Wang et al., 2022).

In the region outlined by orange, there is a small island of YSOs, PNe, evolved/dusty stars, emission-line stars, etc, sources that are often misclassified as AGN and vice versa. Source 5 and Source 8, from an area of transitioning galaxies hosting AGN on the SMC map, and whose true classifications turned out to be emission-line and evolved dusty stars, are found on the edge of the sample closest to the small island of YSOs, PNe, evolved/dusty stars, emission-line stars, etc. If some of the stellar sources can be found on the outskirts of the AGN section of the map, this could imply the opposite, that some transitioning galaxies hosting AGN are on the edge of the small stellar island closest to the sample investigated in this work.

AGN that have been spectroscopically observed can be found roughly along the centre of the map, y $\sim$ 0 for the SMC and x $\sim$ 0 for the LMC. Blazars can be found to the far right of the AGN sequence in the SMC and to the far bottom of the AGN sequence of the LMC map.

Low-z galaxies, with obvious galaxy shape in survey images, can be found from the AGN region to up towards the top of the map in the SMC, and to the far right in the LMC. There is a redshift gradient in both maps of increasing redshift towards the AGN sequence (down and to the left in the SMC and LMC, respectively).

The LMC map, unlike the SMC map, separates SAGE0534AGN from SAGE0536 AGN. These two sources can be found on two separate points that reach towards the island of YSOs, PNe, evolved stars and emission line stars, which is similar to the SMC map. SAGE0534AGN can be found among the AGN sequence and SAGE0536AGN can be found among the low-z galaxies, which makes sense with respect to the redshift gradient in the map.

6.1.1.2 Photometry survey measurements

The Gaia survey provides proper motions measurements, and from the distribution of proper motions across the t-SNE maps shown in Figure 6.4 it can be seen that most

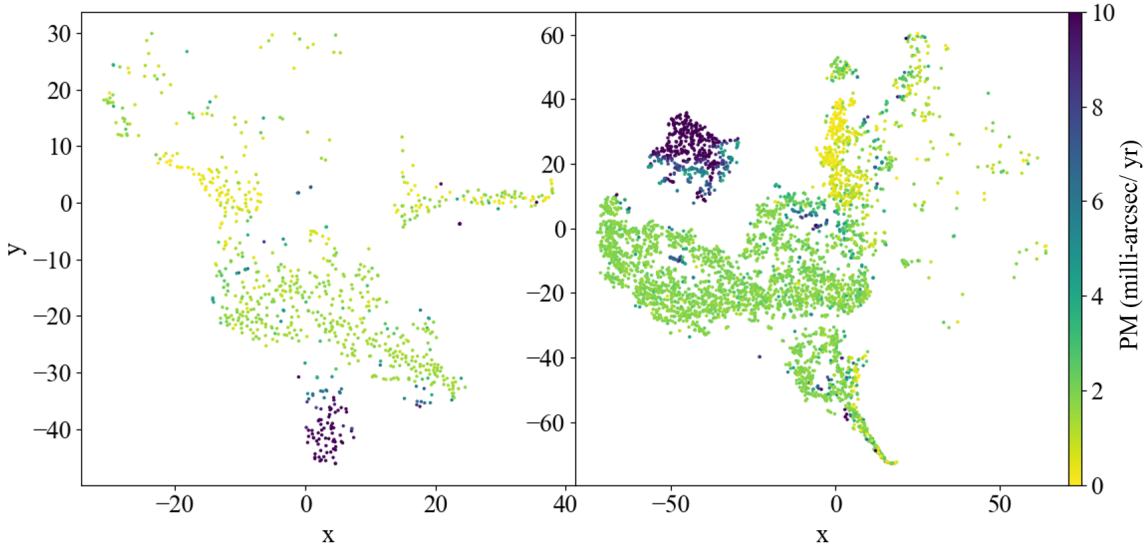

Figure 6.4: Comparison of the distribution of proper motions of the sources in the t-SNE maps of the SMC (left) and LMC (right).

sources show Magellanic proper motions or lower. The lowest proper motions can be matched with known AGN and galaxies, as expected. Where there are no data for the proper motions is where most of the galaxies can be found. Where the galaxies can be found also matches up with the greatest difference between Gaia G and the average of G_{RP} and G_{BP} bands, as seen in Figure 6.5. This shows that the difference between Gaia G and the average of G_{RP} and G_{BP} can be used to locate extended objects. This magnitude difference identifies extended objects because the value of G for a point source is expected to roughly equal the average of G_{RP} and G_{BP}. Therefore, the value would be around zero, which can be seen is true for areas where known stars and higher redshift sources are found. The larger magnitude differences are caused by the way the flux is measured for G (broad pass band filter) compared to G_{RP} and G_{BP} (prism and colour filter, creating low-resolution spectra in broad blue and red pass bands).

High proper-motion (foreground) stars can be found in the isolated bottom and top left islands of the SMC and LMC t-SNE maps, respectively. These sources have the potential to be radioactive stars. Further inspection of these sources in the radio

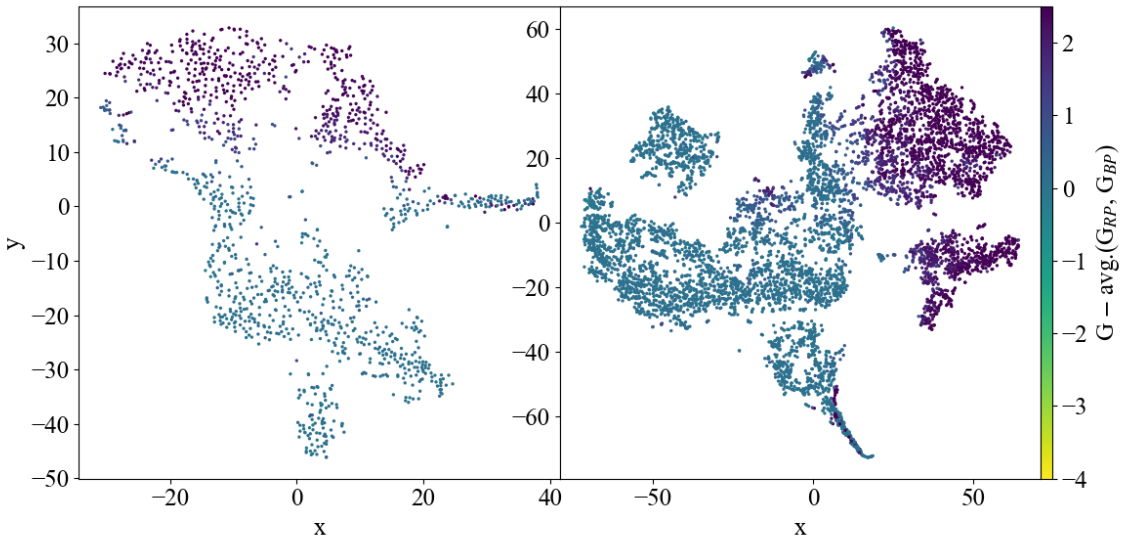

Figure 6.5: Comparison of the distribution of Gaia DR3 colour G minus the average of G_{RP} and G_{BP} of the sources in the t-SNE maps of the SMC (left) and LMC (right).

have shown that they are a mixed bag. Some of the sources match up with artefacts around bright radio sources that have made their way into the radio catalogues, some are so faint they could be part of the background variation and others have matched up with jets/lobes/extended emission of true radio sources. Some do match up with point sources, so have the potential for the radio emission to belong to the foreground star. This requires further investigation as I did identify some foreground cool dwarf systems with radio emission in the LMC ASKAP image (see Appendix D and Pennock et al., 2021). It is possible there lies a large population of radio stars just below or around the sensitivity of the ASKAP survey.

The VISTA parameters that show the greatest difference across class positions on the map are the colour $J - K_{\rm s}$ and the J sharpness parameter. The distribution of $J - K_{\rm s}$ across the maps can be seen in Figure 6.6. The distribution of J sharpness across the maps can be seen in Figure 6.7. The $J - K_{\rm s}$ colour is highest, and therefore reddest, where the extragalactic sources can be found. The J sharp parameter is highest to the

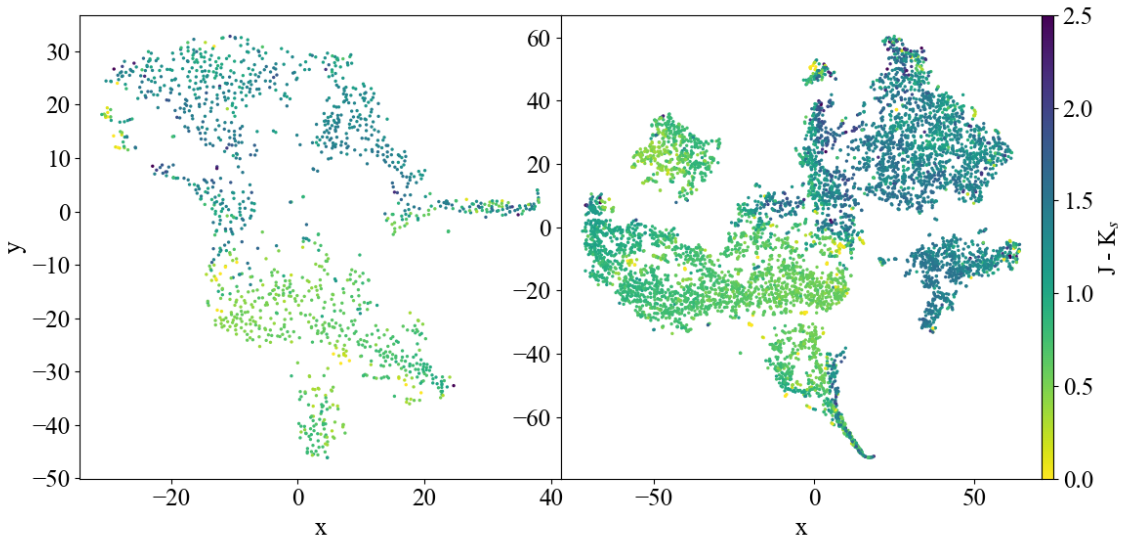

Figure 6.6: Comparison of the distribution of VMC PSF colour $J - K_s$ of the sources in the t-SNE maps of the SMC (left) and LMC (right).

far right where the lowest redshift galaxies, which would be seen as extended in the VMC survey, reside so this is as expected. Note that the distributions of both $J - K_s$ and the J sharpness are very similar, most likely because redder sources are generally associated with extended sources such as galaxies and star-forming regions.

The WISE parameter that shows the greatest difference across class positions on the map is the W1 − W2 colour. This distribution can be found in Figure 6.8. The highest values can be seen where the known AGN are, showing that this WISE colour is good for selecting AGN.

In Figure 6.8, the area highlighted in light green/yellow only stands out in the W1−W2 colour, where this colour is more negative than the rest of the sources, and W2 mag, where they tend to be fainter than the rest of sources. The identity of these sources and why they stand out in this way requires further investigation.

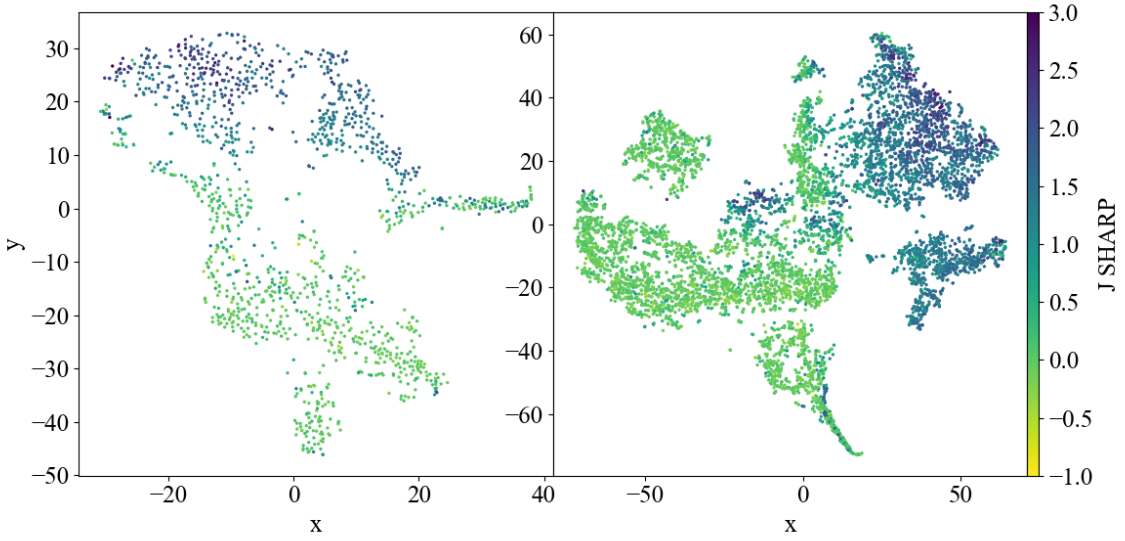

Figure 6.7: Comparison of the distribution of VMC PSF J-band sharpness value of the sources in the t-SNE maps of the SMC (left) and LMC (right).

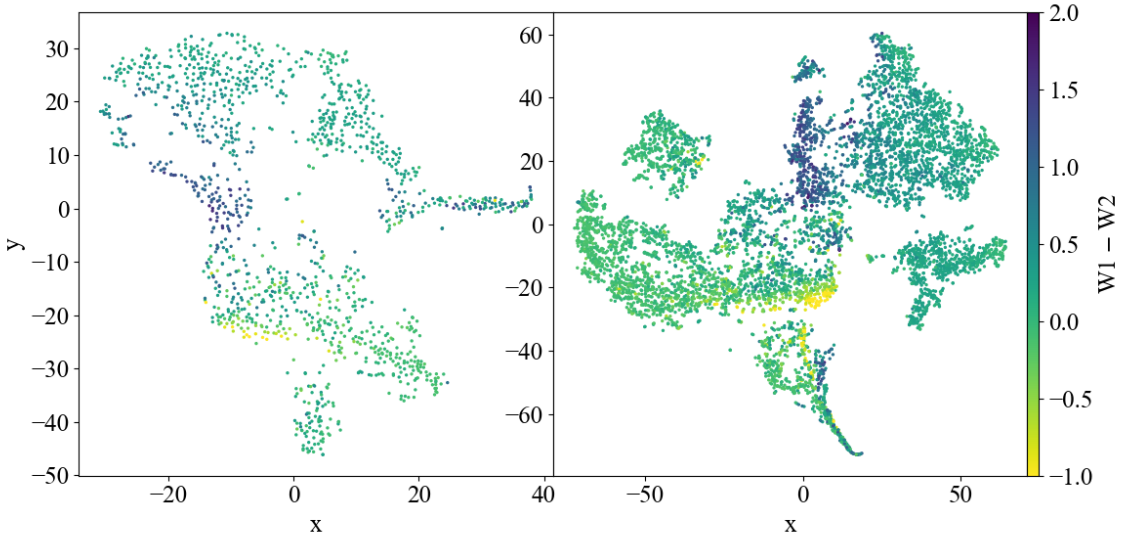

Figure 6.8: Comparisons of the distribution of AllWISE colour W1 − W2 of the sources in the t-SNE maps of the SMC (left) and LMC (right).

6.1.1.3 Gaia classifications

The Gaia DR3 (Gaia Collaboration et al., 2022) catalogue includes three columns that give probabilities that the source is a QSO, a galaxy or a stellar source. When comparing the Gaia classifications to the true classifications of the t-SNE selected sample of dusty AGN (see Table 6.1) we see that the majority of sources are correctly predicted as QSOs, though this also includes Source 5, which was identified as a carbon star. Only two sources were classified by Gaia as stellar sources, Source 8, whose true identity turned out to be an emission-line star, and SAGE0534AGN, an AGN. SAGE0536AGN was strongly predicted by Gaia to be a galaxy, which is not unexpected due to the host galaxy being slightly resolved, but there was also a $\sim 5\%$ chance of being a star. Source 7 was predicted to be $\sim 70\%$ QSO and 30% galaxy, which interestingly matches with the AGN fraction from CIGALE.

Looking at the rest of the SMC t-SNE map, as well as the LMC t-SNE map we can see how the Gaia classification probabilities are distributed across the map. Looking at the distribution of $P_{\rm QSO}$ (see Figure 6.9) it can be seen that the Gaia classifications match up with most of the known AGN, though do not manage to get the AGN that are within the low redshift galaxy region (see Figures 6.2 and 6.3).

Looking at the distribution across the SMC and LMC t-SNE maps of $P_{\rm Gal}$ (see Figure 6.10) it can be seen that this matches up with the areas where known galaxies can be found. Unexpectedly, in the LMC map, there is an area of high probability galaxies in the centre of the t-SNE map (at $x \sim -13$, $y \sim 9$), to the left of the AGN sequence, which also matches up with high values of J sharp, indicating extended sources (see Figure 6.8). These sources require further investigation.

Looking at the distribution across the SMC and LMC t-SNE maps of $P_{\rm SS}$ (see Figure 6.11) it can be seen that stellar source probabilities are highest away from where extragalactic sources are found (below $y \sim -10$ for SMC and to the left of $x \sim 0$ for the LMC; also see Figures 6.2 and 6.3). However, some sources with high probabilities as a stellar source can also be found within the extragalactic areas of the t-SNE maps. For instance, the blazar sequence identified in Figures 6.2 and 6.3 are shown to be

Table 6.1: Gaia classification probabilities of QSO ($P_{\rm QSO}$), galaxy ($P_{\rm Gal}$) and stellar sources ($P_{\rm SS}$) for the t-SNE selected AGN sample.

Source Name	$P_{\rm QSO}$	$P_{\rm Gal}$	$P_{\rm SS}$
SAGE0536AGN	0.0002	0.9471	0.0527
SAGE0534AGN	0.0015	0.0000	0.9985
1	1.0000	0.0000	0.0000
2	1.0000	0.0000	0.0000
3	1.0000	0.0000	0.0000
4	1.0000	0.0000	0.0000
5	0.9952	0.0000	0.0048
6	0.9939	0.0000	0.0061
7	0.7073	0.2911	0.0016
8	0.0001	0.0001	0.9998
9	1.0000	0.0000	0.0000
10	0.9958	0.0033	0.0009
11	1.0000	0.0000	0.0000
12	1.0000	0.0000	0.0000
13	1.0000	0.0000	0.0000
14	1.0000	0.0000	0.0000
15	1.0000	0.0000	0.0000
16	0.8864	0.1136	0.0000

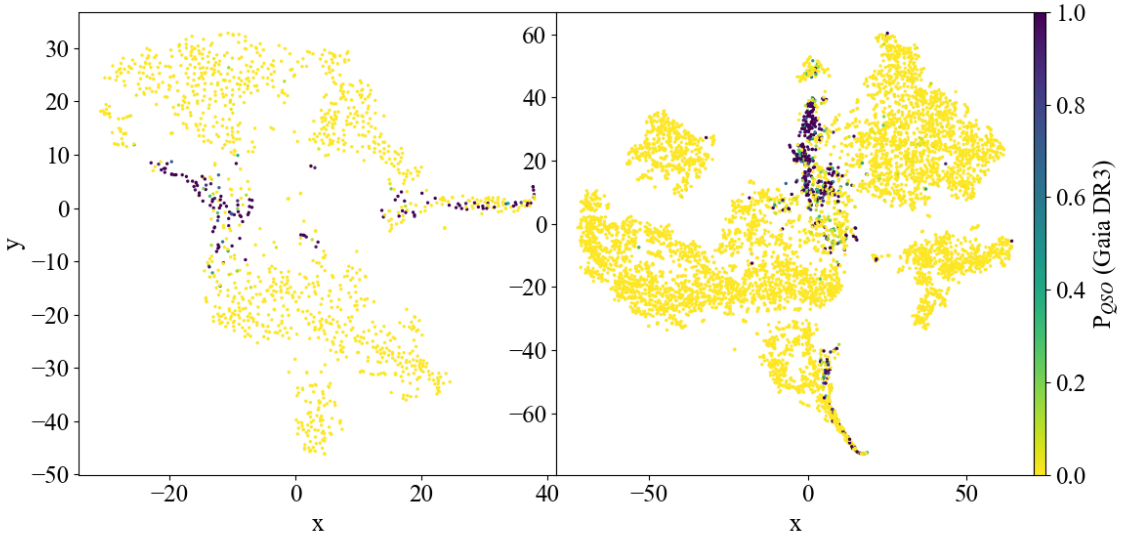

Figure 6.9: Comparison of the distribution of Gaia DR3 probabilities of a sources being a QSO in the t-SNE maps of the SMC (left) and LMC (right).

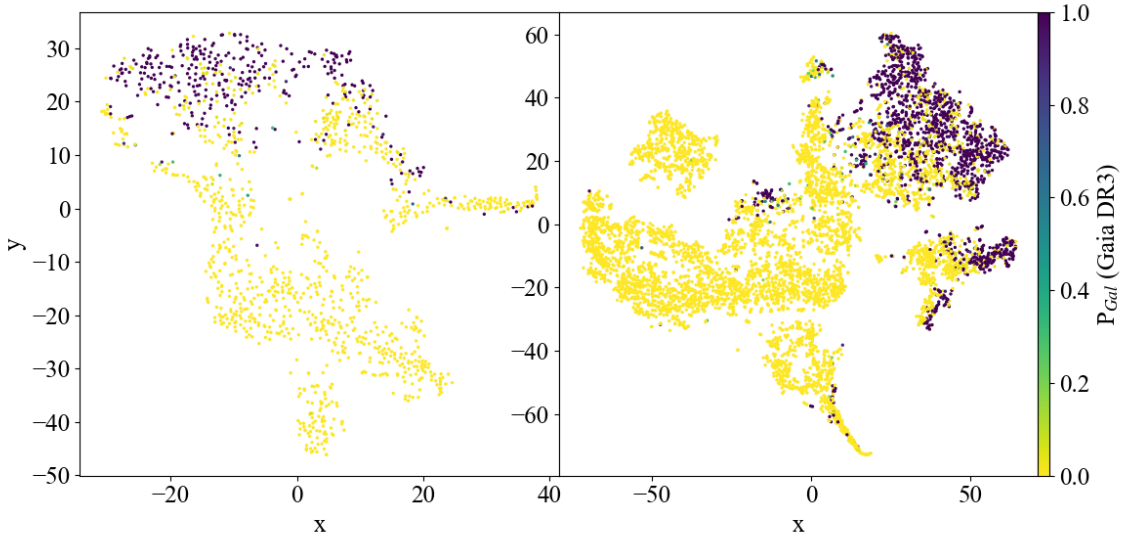

Figure 6.10: Comparison of the distribution of Gaia DR3 probabilities of a sources being a galaxy in the t-SNE maps of the SMC (left) and LMC (right).

predicted as mostly stellar sources in Figure 6.11, though a few are identified as AGN in Figure 6.9.

6.2 PRF classifications of the SMC field

In this section I explore the performance of the PRF classifier trained on the SMC training dataset (see Section 3.4.2) used on the entirety of the SMC VMC dataset. The LMC will be explored in a future paper (Pennock et al. *in prep.*).

The SMC consists of 31,169,627 sources in the VMC catalogue. After removing possible duplicates (identified as sources that are closer than the expected resolution of the survey) between overlapping observations this reduced to 28,598,800 source. Due to the size of the data it was classified in batches of one million at a time. The probabilities for each specific class were calculated by the classifier, and from this the overall predicted class was the class with the highest probability.

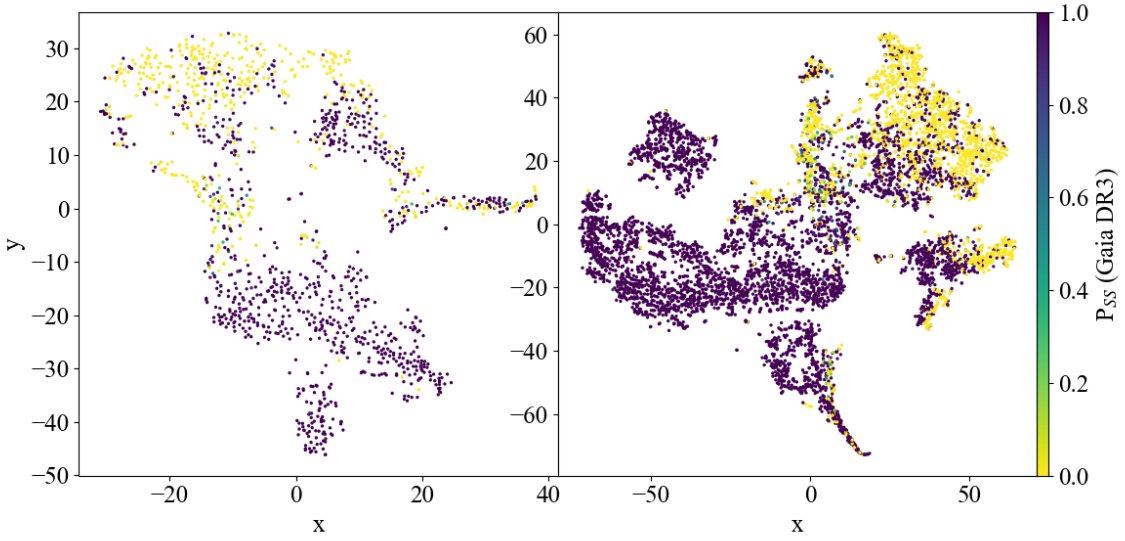

Figure 6.11: Comparison of the distribution of Gaia DR3 probabilities of a source being a stellar source in the t-SNE maps of the SMC (left) and LMC (right).

A sample of the table of probabilities and classes I created can be found in Table 6.2.

Table 6.2: Excerpt from the table predicted class probabilities by the PRF for the SMC dataset. All probabilities are from 0 – 1. The full table will be made available alongside the paper this research will be a part of.

VMC RA (J2000)	VMC DEC (J2000)	AGN Prob.	Galaxy Prob.	OB Prob.	RGB Prob.	AGB Prob.	H II/YSO Prob.	PNe Prob.	pAGB/pRGB Prob.	RSG Prob.	PM Prob.	Class	Class Prob.
00:54:40.34	−72:46:52.6	0.05	0.21	0.21	0.09	0.06	0.00	0.07	0.08	0.11	0.11	Galaxy	0.21
00:35:25.35	−73:27:17.5	0.23	0.10	0.14	0.21	0.04	0.03	0.10	0.09	0.02	0.05	AGN	0.23
01:49:18.32	−74:14:02.2	0.28	0.18	0.12	0.10	0.06	0.05	0.09	0.03	0.03	0.06	AGN	0.28
00:29:13.66	−75:28:48.3	0.32	0.28	0.13	0.10	0.04	0.02	0.00	0.02	0.01	0.05	AGN	0.32
...	...	...	...	...	...	...	...	...	...	...	...	...	...

6.2.1 Stellar sources

One way to ascertain the performance of the classifier in separating extragalactic from stellar sources is to test it on classes it has not seen before. The classifications from SAGE-spec (Ruffle et al., 2015) were used as part of the training set for the classifier. Not all the classes in this dataset were added to the training set as they were deemed too few in number for training purposes and/or not a well defined class (e.g. emission-line stars). The SAGE-spec dataset for the SMC has 18 sources that were not used. The predicted probabilities and classes of these 18 sources were extracted from the full SMC dataset and can be seen in Table 6.3.

Table 6.3: Sources from Ruffle et al. (2015) that were not used in the training of the PRF classifier. The SAGE classes are Wolf-Rayet stars (WR), R Coronæ Borealis variable stars (RCrB), Blue supergiants (BSG), S-type stars (S star), Symbiotic stars (Sym. star), and stars of indiscernible type (star).

Source Name	SAGE Class	RA (J200)	DEC (J200)	Predicted Class	Class Prob.
SMC-WR9	WR	00:54:32.2	−72:44:36	OB	0.98
SMC-WR12	WR	01:02:52.2	−72:06:52	OB	0.94
GSC 09141−05631	WR	00:43:42.2	−73:28:54	OB	0.91
SMC-WR2	WR	00:48:31.0	−73:15:45	OB	0.99
SMC-WR3	WR	00:49:59.3	−73:22:14	OB	0.98
SMC-WR4	WR	00:50:43.4	−73:27:05	OB	0.95
RMC 31	WR	01:03:25.2	−72:06:44	OB	0.45
SMC-WR11	WR	00:52:07.5	−72:35:38	RGB	0.46
MSX SMC 014	RCrB	00:46:16.4	−74:11:13	AGB	0.81
MSX SMC 155	RCrB	00:57:18.2	−72:42:35	AGB	0.46
AzV 404	Star	01:06:29.4	−72:22:09	OB	0.53
BFM 1	S star	00:47:19.3	−72:40:04	AGB	0.88
AzV 456	Star	01:10:55.8	−72:42:57	OB	0.84
AzV 23	Star	00:47:38.9	−73:22:54	OB	0.62
OGLE SMC-SC10 107856	RCrB	01:04:53.1	−72:04:04	OB	0.17
MSX SMC 185	Sym. star	00:54:19.5	−72:29:09	PNe	0.78
HD 5980	WR	00:59:26.7	−72:09:54	OB	0.60
HD 6884	BSG	01:07:18.1	−72:28:04	OB	0.33

Overall, all of the sources were classed as stellar sources. All sources are predicted to have a $< 10\%$ probability of being an AGN and $< 5\%$ probability of being a galaxy. This implies that similar stellar sources of these natures would most likely not be predicted to be extragalactic.

R Coronae Borealis variable stars (RCrB) are often associated with post-AGB stars so it is not surprising that two out of three were classed as AGB stars. Wolf-Rayet stars (WR) can be similar to O type stars so would be expected to be predicted as an OB star, which all but one are. The WR star predicted as an RGB star could possibly be a case of a dusty WR star. The blue supergiant (BSG) star is also predicted to be an OB star, which is not unexpected since OB stars are also blue in colour.

These results are promising for AGN classifications, though the testing of the PRF showed that emission-line stars are still likely interlopers.

6.2.2 t-SNE radio sources

The SMC t-SNE map explored in Section 6.1 and Chapter 5 were cross-matched with the probabilities calculated for the SMC sources with a class probability >50%.

The classifications distribution across the SMC t-SNE map are shown in Figure 6.12. The AGN and galaxy classes shown are those with a secondary extragalactic classification, and the stellar classes shown are those with a secondary stellar classification. This map matches up with expected known sources (as seen in Figure 6.2), where extragalactic sources are in the top half of the map, with more galaxies towards the top and more AGN towards the bottom. The small island to the top left where sources such as YSOs, PNe, dusty evolved and emission-line stars are known to reside, which are often confused with AGN, show a mix of OB (potentially have emission lines), PNe and YSO classes, as expected.

The bottom half of the t-SNE map is dominated by RGB stars. RGB (as well as PM, AGB, OB and RSG) stars are unlikely to have radio emission associated with them. These could be cases of chance superpositions of stars in front of the true radio source, or a case of the true radio source being too faint at shorter wavelengths, causing

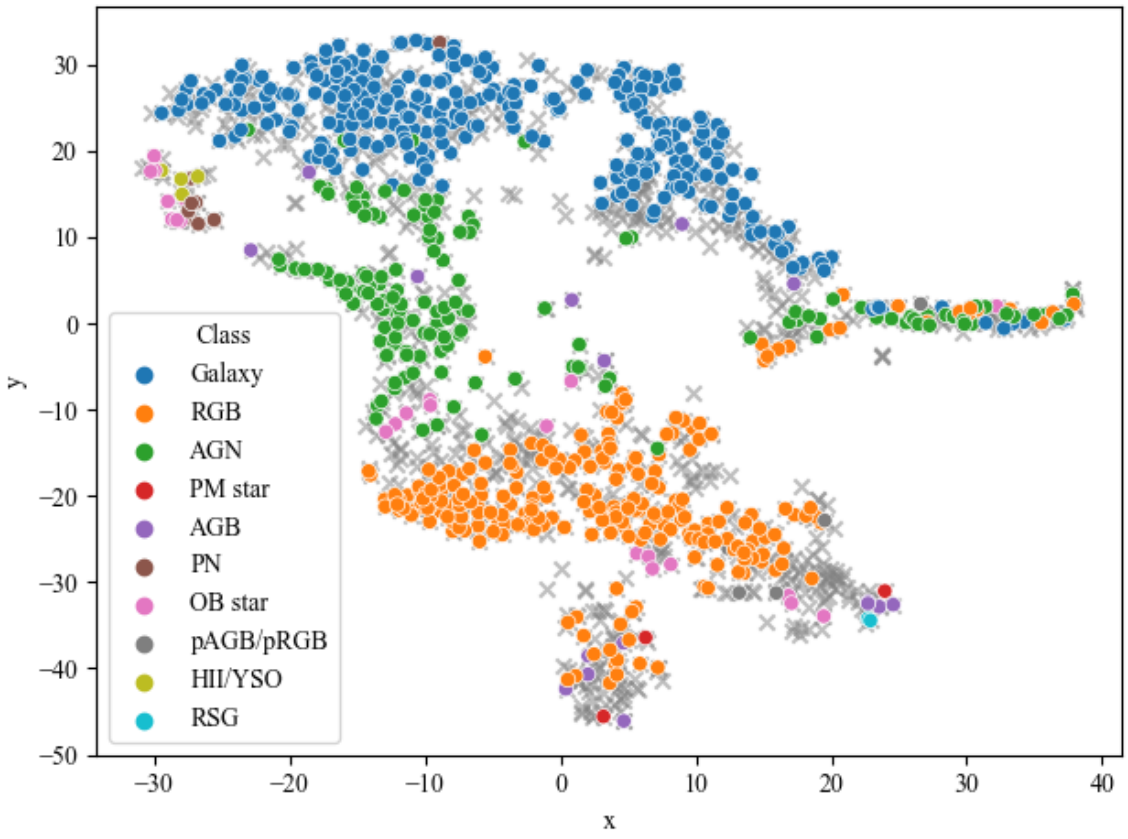

Figure 6.12: PRF classifications of the sources with class probability >50% in the SMC t-SNE map. The grey crosses represent sources with overall class probability <50%.

a mismatch to the nearest (within 5″) visible source. The blazar sequence at x $\sim$ 20 – 40 and y $\sim$ 0 includes RGB stars as well, which is unsurprising since blazars are often mistaken as stars and vice versa.

6.2.3 PRF on whole SMC

The full SMC VMC dataset consists of 28,598,800 sources, where $\sim$ 94% are classed as stellar and $\sim$ 6% are classed as extragalactic. Table 6.4 shows the distribution of classes for the entire SMC field, as well as for sources with class probabilities $>$ 50% and $>$ 75%. This shows that the majority of sources in the SMC field are classified as stars, as expected. OB and RGB stars have the highest number, which could be caused by similar stellar sources being classed as these classes, such as other main sequence stars. The majority of extragalactic sources are expected to be galaxies not hosting an AGN, however, galaxy counts tend to be lower than AGN. This could be explained by the host galaxies used to train the PRF being all low redshift sources, which could mean the higher redshift galaxies are being predicted to be other classes, possibly AGN or even PNe, which has an unexpected high count even at probability $>$ 75%. The PNe class numbers do not reduce to the expected amount (similar to the HII/YSO number) until probability $>$ 80%.

6.2.3.1 Extragalactic in the radio

Separating the extragalactic sources into AGN and galaxies, and cross-matching with the radio ASKAP SMC catalogue with a 5″ search radius, gives 216 and 366 sources for AGN and galaxies, respectively, that have a probability >50%. Looking at the radio flux density distribution at 960 MHz and 1320 MHz, and the ratio of AGN to galaxy counts in Figures 6.13, 6.14, it can be seen there are more AGN compared to galaxies at brighter (>10 mJy) radio flux densities, as expected. It can also be seen that towards fainter radio flux densities galaxies become more favoured.

From my work in Chapter 4, it was seen that there was an upturn towards the

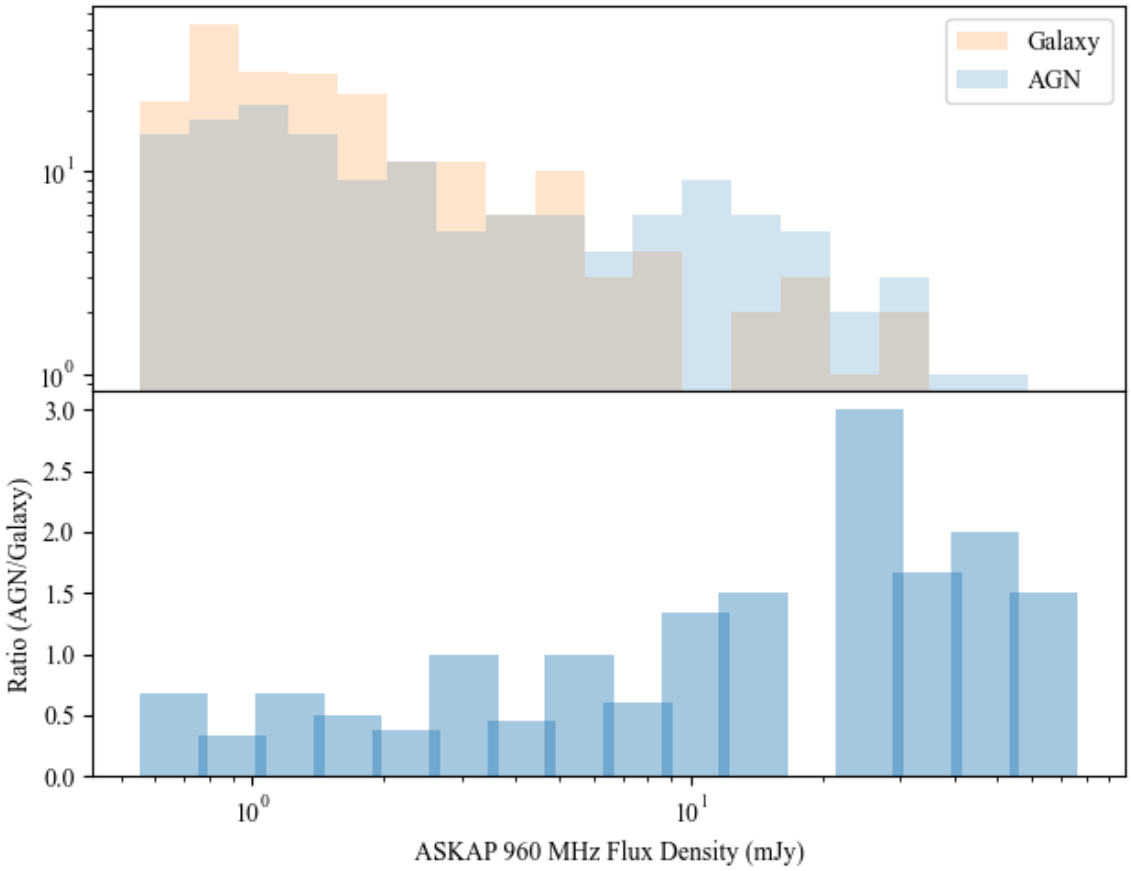

Figure 6.13: (Upper panel) Radio ASKAP 960 MHz flux density distribution of the predicted AGN and galaxy sources in the SMC. (Lower panel) Ratio of AGN to galaxy counts across the observed ASKAP 960 MHz flux density range.

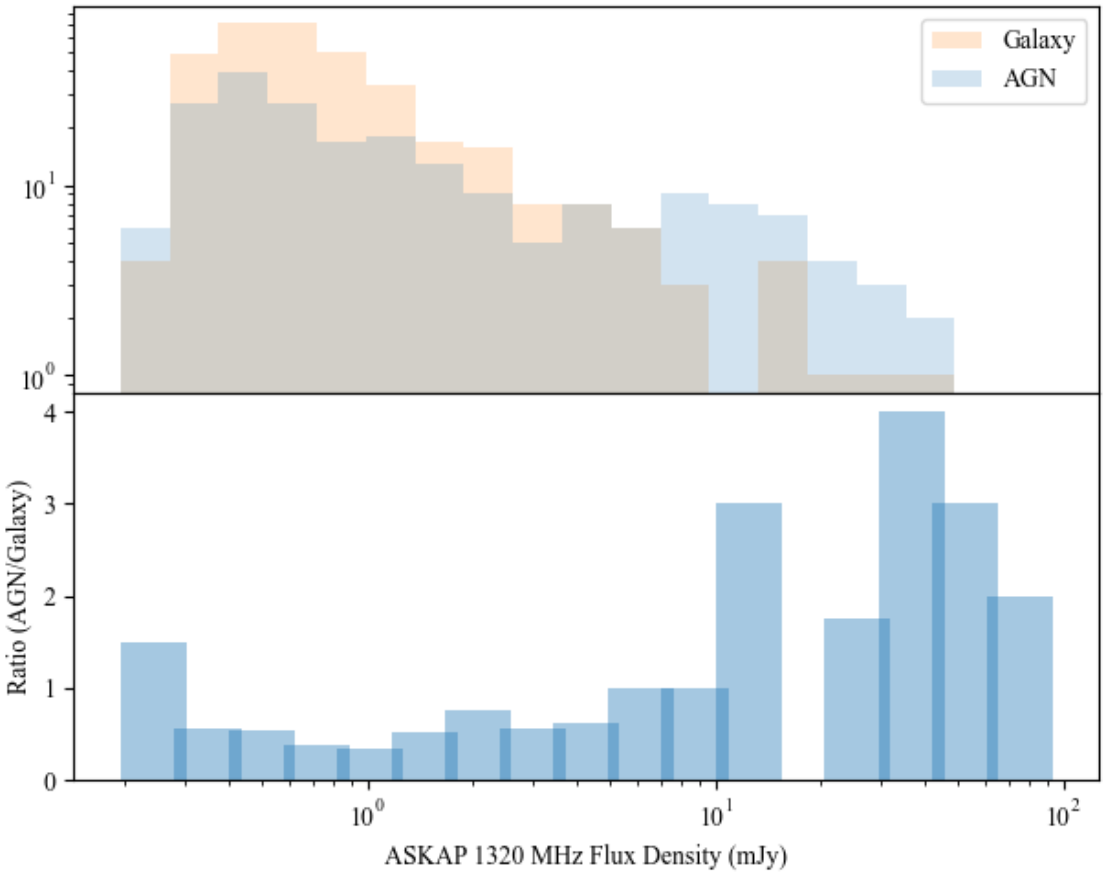

Figure 6.14: (Upper panel) Radio ASKAP 1320 MHz flux density distribution of the predicted AGN and galaxy sources in the SMC. (Lower panel) Ratio of AGN to galaxy counts across the observed ASKAP 1320 MHz flux density range.

Table 6.4: Distribution of the classifications of sources in the SMC field for all sources, sources with class probability > 50% and sources with class probability > 75%.

Class	All	> 50% Class Probability	> 75% Class Probability
All	28,598,800	2,085,277	422,564
AGN	1,716,273	3837	1033
Galaxy	93,803	3333	818
OB	17,191,315	488,647	102,943
RGB	7,089,264	941,945	309,556
PNe	1,827,123	629,697	5511
Post-AGB/RGB	397,674	1585	1
AGB	168,122	8307	1924
HII/YSO	74,317	83	6
PM	36,048	6368	412
RSG	4861	1475	360

fainter flux densities at $F_{888} \sim 3$ mJy representing the beginning of the faint galaxy population. Therefore, it is expected that the number of radio detected galaxies will have increased towards the fainter fluxes. This appears to be what we are seeing, however, to see if the trend continues towards fainter flux densities a comparison with the LMC PRF classification will be needed, as the LMC ASKAP observation at 888 MHz reaches fainter flux densities than the SMC ASKAP observations.It should also be noted that the AGN and galaxies being compared here only represent the more confident classifications of the PRF, a fraction of the true number in this field that will have been detected with ASKAP.

6.2.3.2 PRF classification distributions across the SMC field

The distributions of each of the classes across the SMC field can be seen in Figures 6.15 – 6.19. Note that the most confident class predictions tend to be in the areas where all the photometric surveys overlap (see Figure 2.5 for comparison), and therefore the PRF had the least amount of missing data to classify with. Most of the sources in

these distributions have class probabilities $> 75\%$ in order to show the more confident classifications. For PNe sources the probability threshold is increased to $> 80\%$ because at the 75% threshold 5511 sources are predicted to be PNe, which is unlikely due to their rarity. The probability threshold for post-AGB/RGB stars and HII/YSO sources has been lowered to $> 50\%$ to increase the source count.

The distributions of the extragalactic sources are expected to be homogeneous when not looking through a nearby galaxy. In the presence of the SMC the distribution is expected to be highest away from the centre of the SMC, and decrease as the stellar density increases towards the centre of the SMC, as stellar sources are more likely to be in the way of the background extragalactic sources. The distribution of the sources classed as AGN and galaxies, seen in Figure 6.15, is as expected, the number of sources decreasing towards the centre of the SMC. Note that at the $> 50\%$ probability threshold the distribution becomes more homogeneous for both AGN and galaxies across the SMC field.

The RGB stellar distribution, seen in Figure 6.16 (Upper panel), concentrates nicely in the centre of the SMC, with fewer sources predicted towards the edges of the SMC field. At the very centre of the dense region of RGB stars the density decreases, which matches with where other stellar classes (OB, AGB, post-AGB/RGB and RSG stars) seem to concentrate. The OB stellar distribution, seen in Figure 6.16 (Lower panel), also concentrates nicely in the centre as expected with fewer sources at the edge of the SMC. The SMC is known to have a bar structure with an extension towards the East, which is what we are seeing here with the OB distribution. Emission-line stars were seen to most likely be classed as OB stars and the slight increase in sources towards the lower left might include emission-line stars from the Bridge.

The AGB stellar distribution can be seen in Figure 6.17 (Upper panel), where it can be seen that the AGB sources are centred on the SMC, with the density of sources increasing towards the centre of the SMC as expected. Figure 6.17 (Lower panel) shows that the post-AGB/RGB also generally lie where expected. Like the OB class, the post-AGB/RGB seems to also loosely follow the SMC's bar shape. However, from the confusion matrix of the classifier (see Figure 3.26), whilst little of the other classes

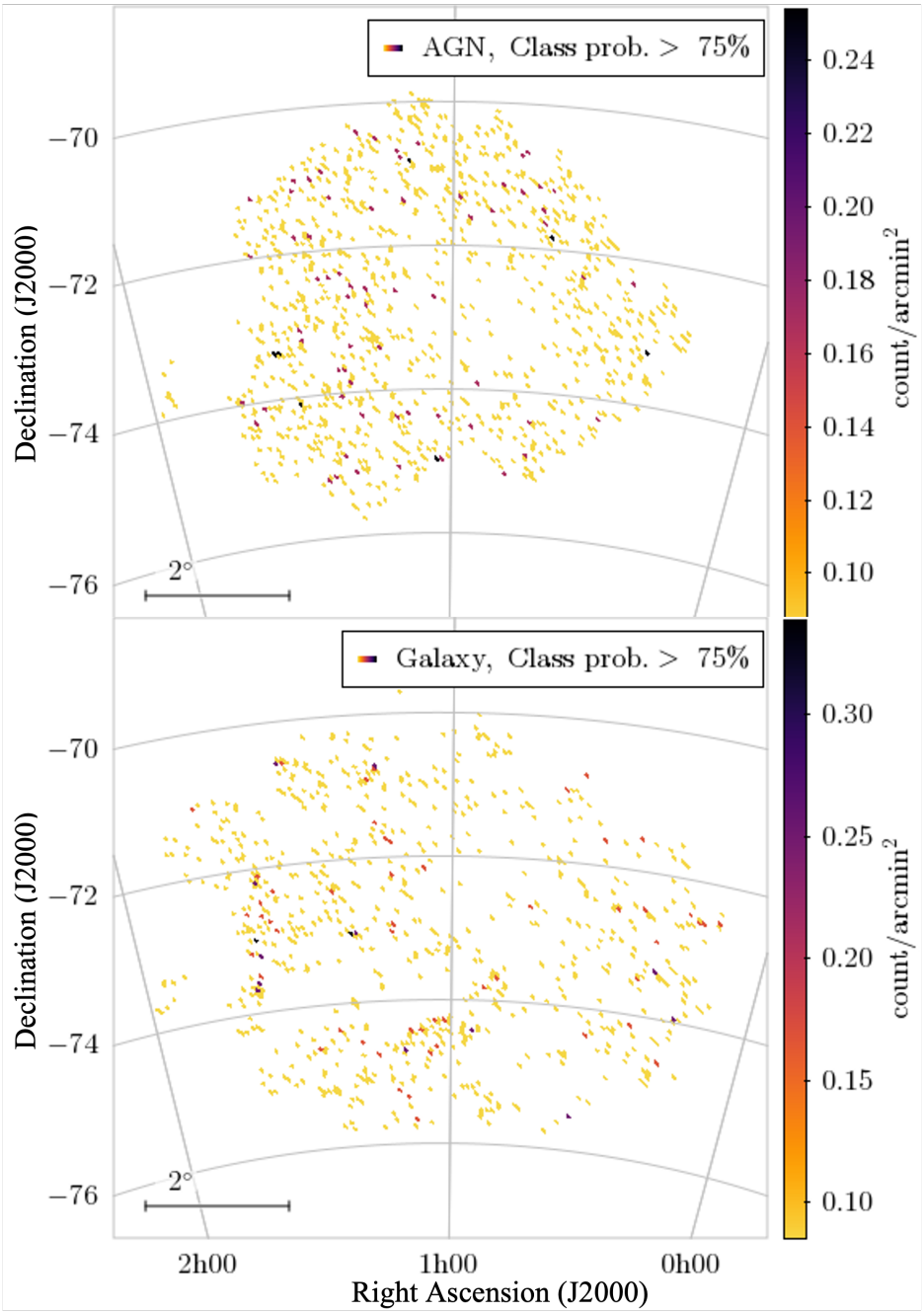

Figure 6.15: Distributions of the SMC sources classified as AGN (Upper panel) and Galaxies (Lower panel).

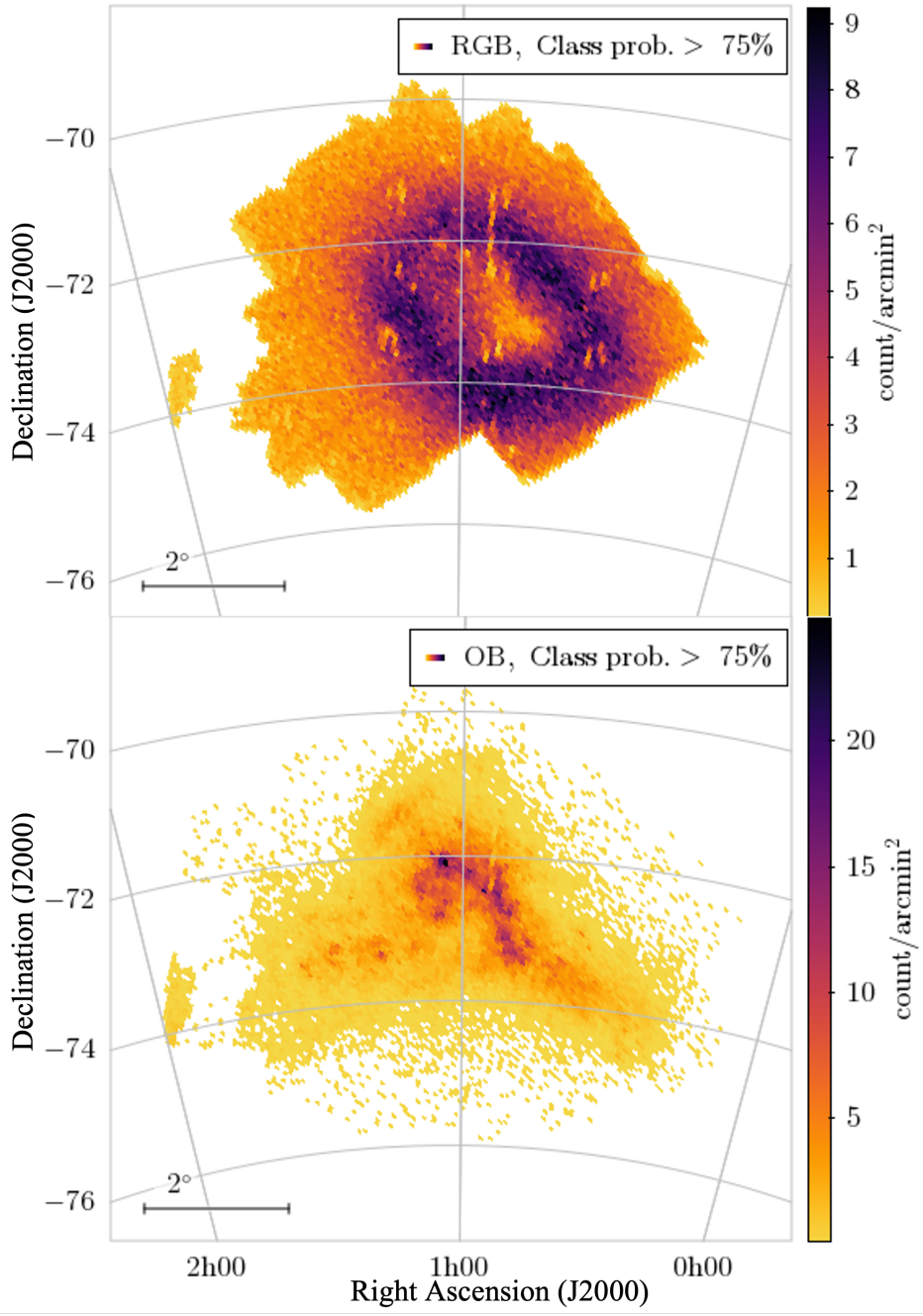

Figure 6.16: Distributions of the SMC sources classified as RGB (Upper panel) and OB stars (Lower panel).

were misclassified as post-AGB/RGB sources, the majority of known post-AGB/RGB sources were misclassified as other classes, mainly RGB stars ($\sim 30\%$) and OB stars ($\sim 19\%$). This means that while the sources seen here are most likely true post-AGB/RGB stars, a lot of the post-AGB/RGB stars will have been misclassified as something else.

For the PNe class a low number of sources being classified is expected due to PNe being hard to classify, as well as the small number of sources in the training set from which the classifier can learn from would restrict and bias the classifier to a small sample of sources. This, however, only becomes true if using a probability threshold of $> 80\%$. Below this threshold the distribution nicely concentrates on the SMC but PNe classed sources can be found across the entirety of the SMC field, which is not expected, so these classifications must be used with more caution. The distribution of PNe, with a probability threshold of $> 80\%$, can be seen in Figure 6.18 (Upper panel). Most of the PNe can be found towards the centre of the SMC as expected. In Figure 6.18 (Lower panel) it can be seen that sources classed as PM stars are roughly evenly distributed across the SMC field as expected.

The distribution for RSG sources can be found in Figure 6.19 (Upper panel) and appears to be tracing the SMC bar and extension towards the East, as well as reaching towards the Bridge. The YSO class is expected to be, just as the PNe, hard to classify, and only comprises a small number of confident classifications. This is found to be true even with a lower probability threshold of $> 50\%$, the distribution for which can be seen in Figure 6.19 (Lower panel), where the sources appear to be tracing the bar and the extension towards the East.

6.2.3.3 IR selections of PRF classified sources

Here I focus on the sources with class probabilities $> 75\%$ ($> 80\%$ for the PNe and $> 50\%$ for the post-AGB/RGB and HII/YSO classes). The classifications of the lower probability sources will be explored at a later date.

The distributions of extragalactic and stellar sources are plotted in AllWISE colour–colour diagrams like those in Chapters 4 and 5 (Wright et al., 2010). The

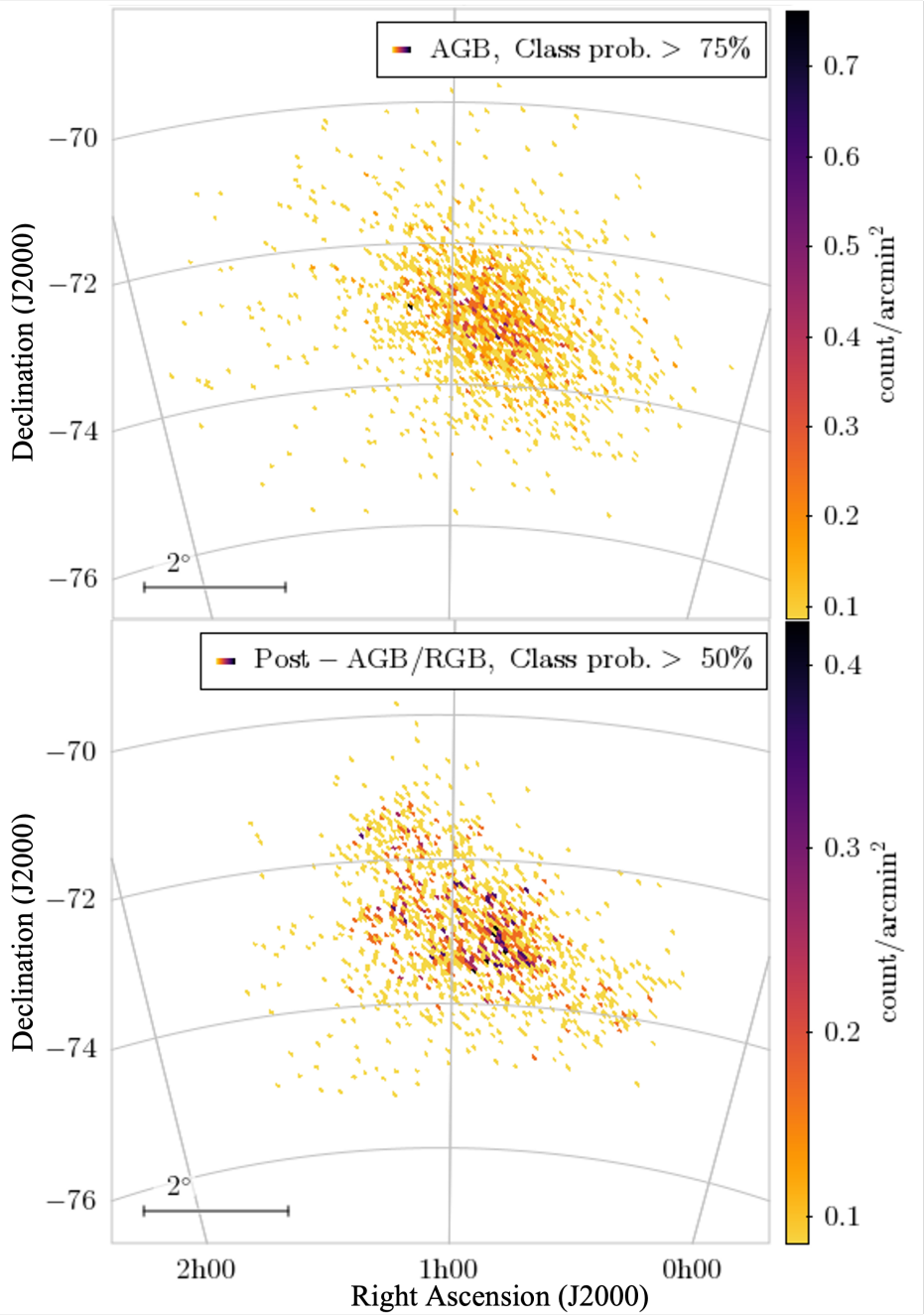

Figure 6.17: Distributions of the SMC sources classified as AGB (Upper panel) and post-AGB/RGB stars (Lower panel).

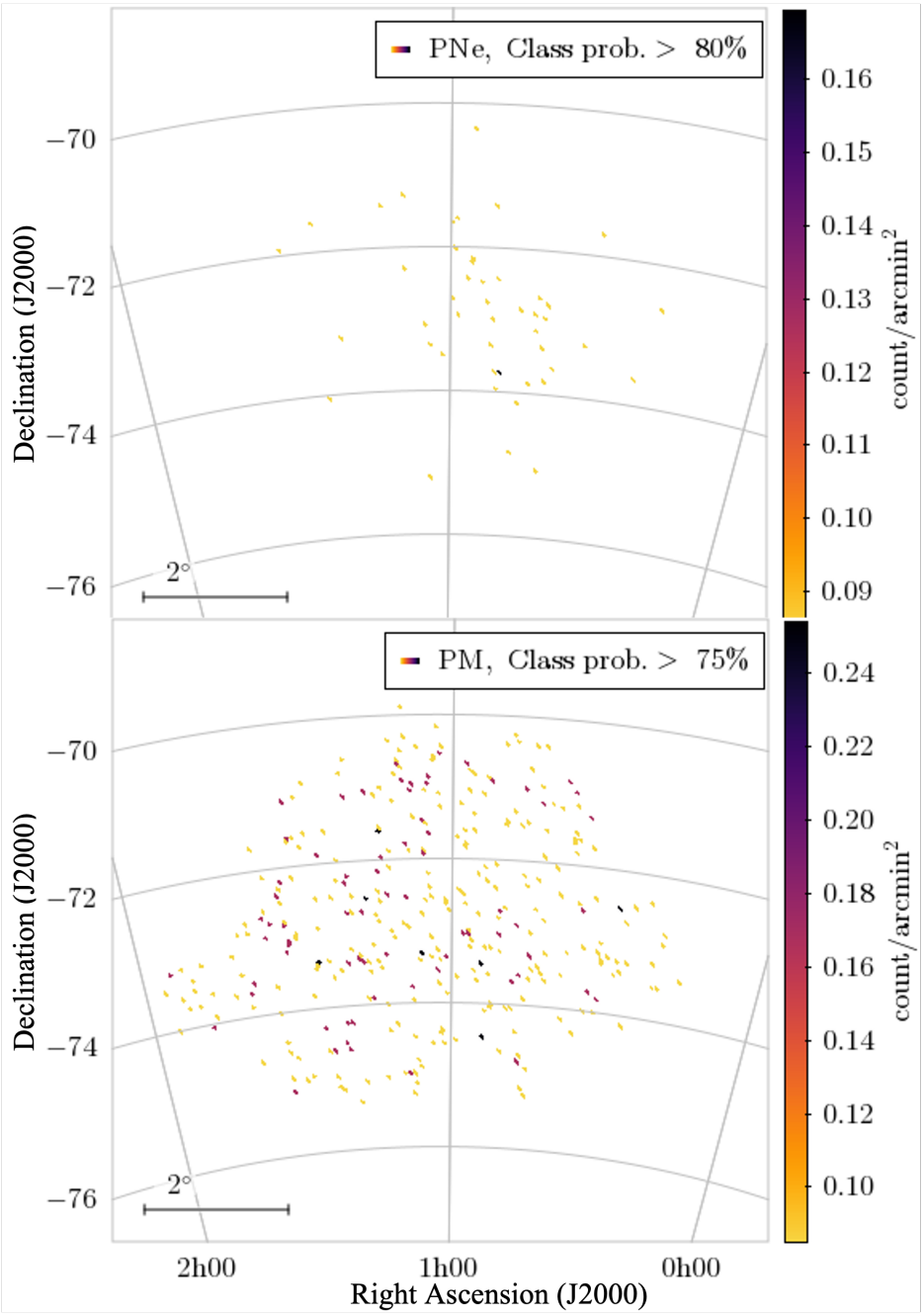

Figure 6.18: Distributions of the SMC sources classified as PNe (Upper panel) and high proper-motion stars (Lower panel).

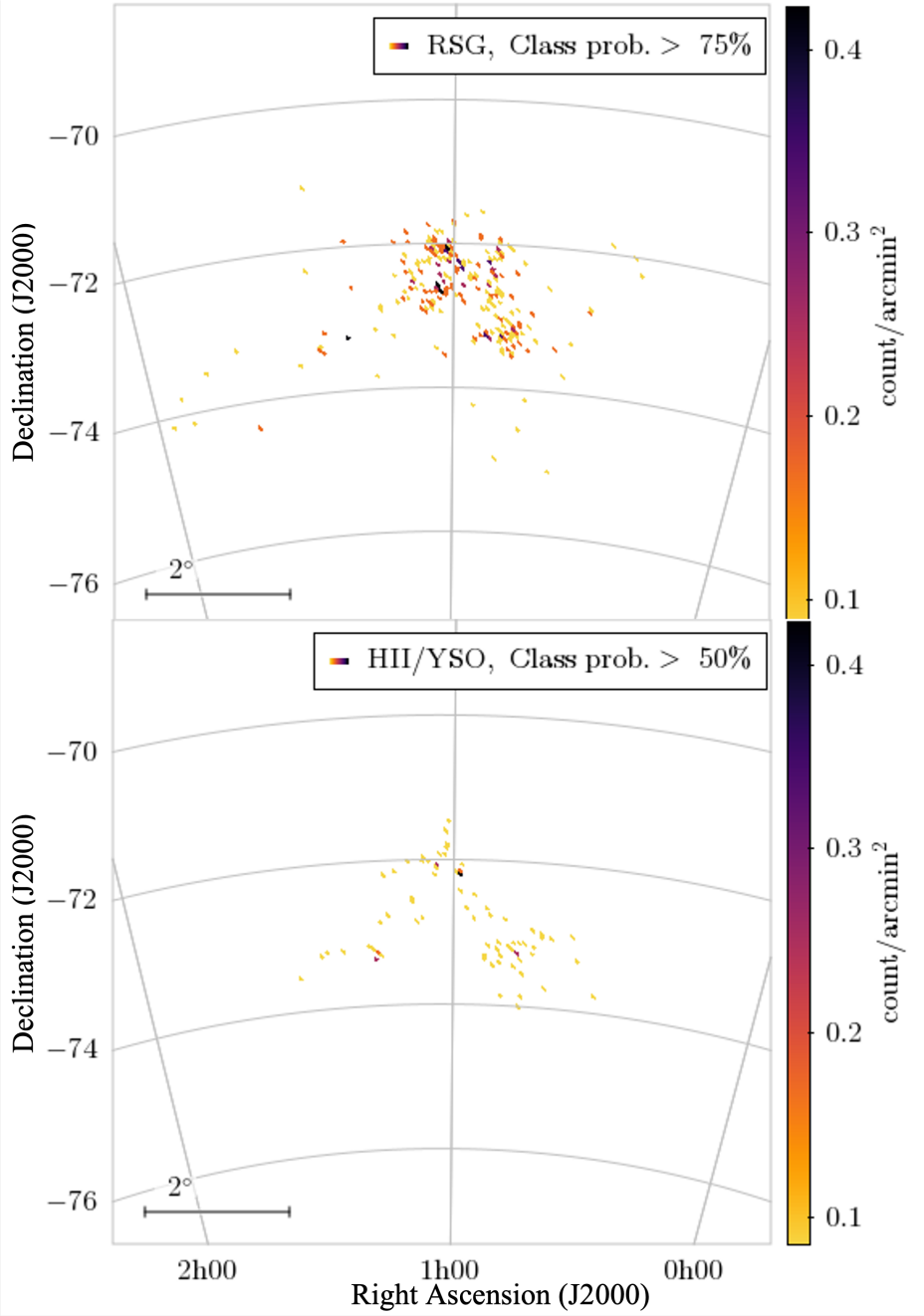

Figure 6.19: Distributions of the SMC sources classified as RSGs (Upper panel) and HII/YSOs (Lower panel).

extragalactic sources can be seen in Figure 6.20 (Upper panel) where both AGN and galaxies occupy expected regions in colour–colour space. The galaxy class occupies the region expected for both star-forming and elliptical galaxies, whilst AGN occupy the region of QSOs and Seyferts.

The distributions of the stellar sources across the AllWISE colour–colour diagram can be seen in Figure 6.20 (Upper panel). The AGB sequence can be clearly seen (similar to what was seen in Chapter 4 (Figure 4.4, right panel) and the PM stars are nicely centred on (0,0) as expected. The populations of RGB, OB, RSG and post-AGB/RGB stars tend to concentrate below W1 – W2 $\sim$ 0, unlike the extragalactic sources which tend to concentrate above W1 – W2 $\sim$ 0. The PNe and YSO are the classes that show the most cross-over with the extragalactic classes, which is not unexpected as they are known to be hard to differentiate from extragalactic sources in colour–colour diagrams. It is noted that sources below W1 – W2 ~ -1 mag are only found within the higher density regions (centre of the SMC). This could be WISE photometry affected by blends, making the longer wavelength, poorer angular resolution data appear brighter.

I cross-matched the sources classed as AGN with the ASKAP SMC 960 and 1320 MHz catalogues and separated them into radio and non-radio detected AGN. This was plotted on an AllWISE colour–colour diagram, seen in Figure 6.21, alongside the t-SNE selected AGN sample from Chapter 5. There is not much difference seen between the radio and non-radio detected AGN. The t-SNE selected AGN are shown to be clearly in the centre of the AGN distribution.

The radio detected AGN, non-radio detected AGN and t-SNE selected AGN are also plotted on VISTA $Y - J$ vs $J - K_{\rm s}$ and $J - K_{\rm s}$ vs $K_{\rm s}$ diagrams in Figures 6.22 and 6.23, respectively. For both plots the overall distributions of radio detected vs non-radio detected are very similar. In Figure 6.22 there is a slight difference between the regions of highest density of sources for the radio and non-radio detected AGN. The highest concentration of radio detected AGN appear to be in region B, whilst non-radio detected AGN concentrate highest in region A. Region A is where the more point-like sources can generally be found, whilst region B is where the extended sources

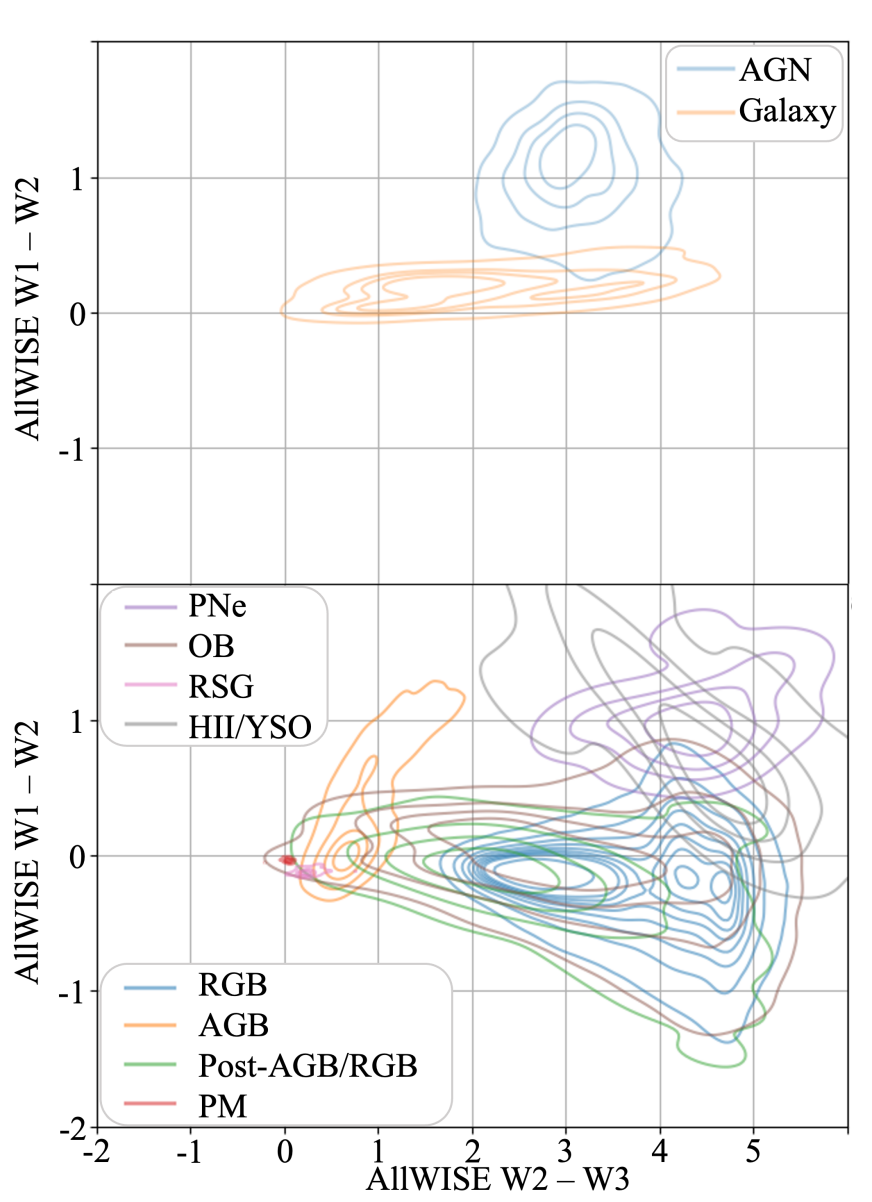

Figure 6.20: AllWISE colour–colour diagrams of the sources classified as AGN and galaxies (Upper panel) and stellar sources (Lower panel) in the SMC field. The contours represent a probability distribution which are in intervals of 0.2.

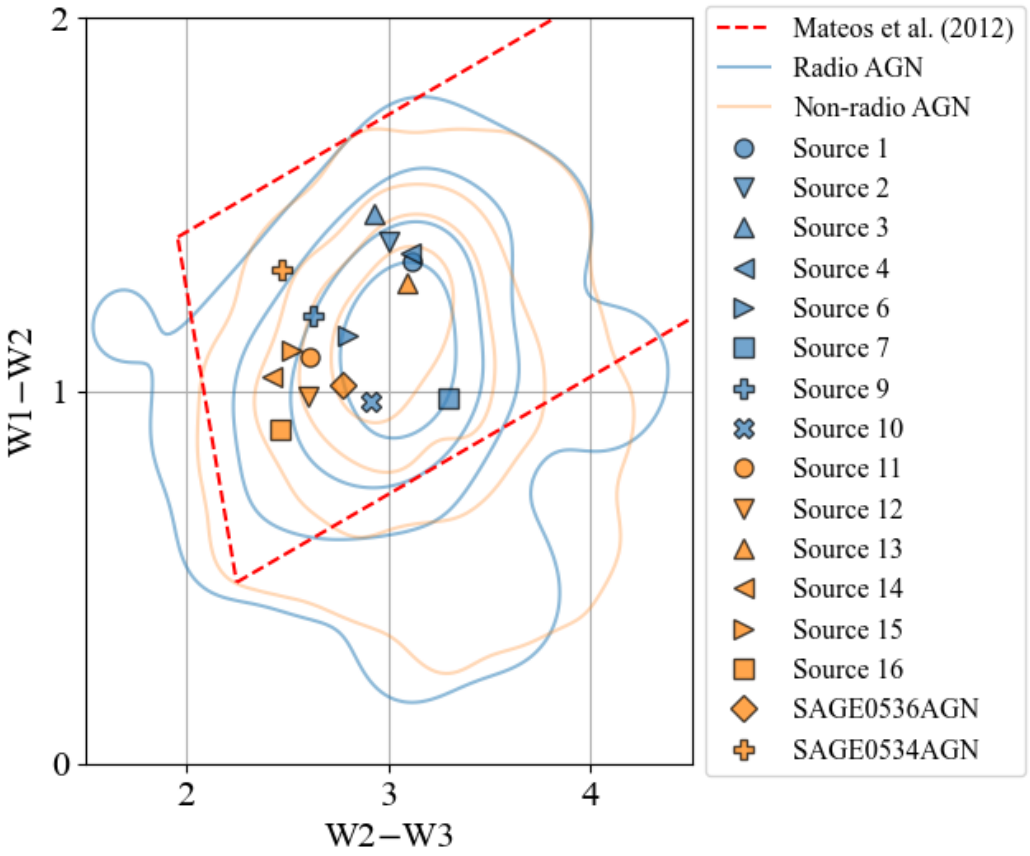

Figure 6.21: AllWISE colour–colour diagram of sources classified as AGN that have (blue) and have not (orange) been detected in the radio. The t-SNE selected AGN sample from Chapter 5 were also plotted. The contours represent a probability distribution which are in intervals of 0.2.

are more likely to be found. The higher concentration of radio detected AGN in region B could be attributed to radio emission coming from the host galaxy instead of, or alongside, radio emission from the AGN, whereas the non-radio detected AGN are not close enough to detect star-formation from their host galaxies, and so would appear more point-like. Radio detected AGN also appear to reach redder/higher values of VISTA $Y - J$, where AGN are known to occupy. However, Cioni et al. (2013) shows that YSOs also can be found here, and it is possible that these radio sources might actually be misclassified YSOs. The-t-SNE selected sample, that are all radio detected AGN, seem to spread out across the distributions for both radio and non-radio detected AGN. Figure 6.23 shows a larger difference between radio and non-radio detected AGN, mostly in the $K_{\rm s}$ band, where radio detected AGN tend to be brighter than non-radio detected AGN. Radio detected AGN also have a slightly higher tendency to be redder in $J - K_{\rm s}$ colour. The t-SNE selected sample appear to follow the trend of the radio detected AGN, though at even brighter values of $K_{\rm s}$ band than the majority of the radio detected population.

The AGN sources show clearer differences between radio and non-radio observed sources in the near-IR bands than in mid-IR bands. The mid-IR emission is dominated by torus emission that can be seen irrespective of the angle of the AGN respective to our line-of-sight. Near-IR emission, however, is more likely related to the hot dust emission from the centre of the dusty torus that can only be seen at angles that are (or close to) face-on. The strongest radio sources also tend to be AGN that are face-on, where the radio jet is heading straight towards us and is Doppler boosted. This could lead to the larger differences seen between radio and non-radio detected AGN in the near-IR than in the mid-IR.

6.3 Conclusions

In this work I used the unsupervised machine learning algorithm t-SNE with Gaia EDR3, VMC, AllWISE and EMU ASKAP photometric data of the SMC and LMC to

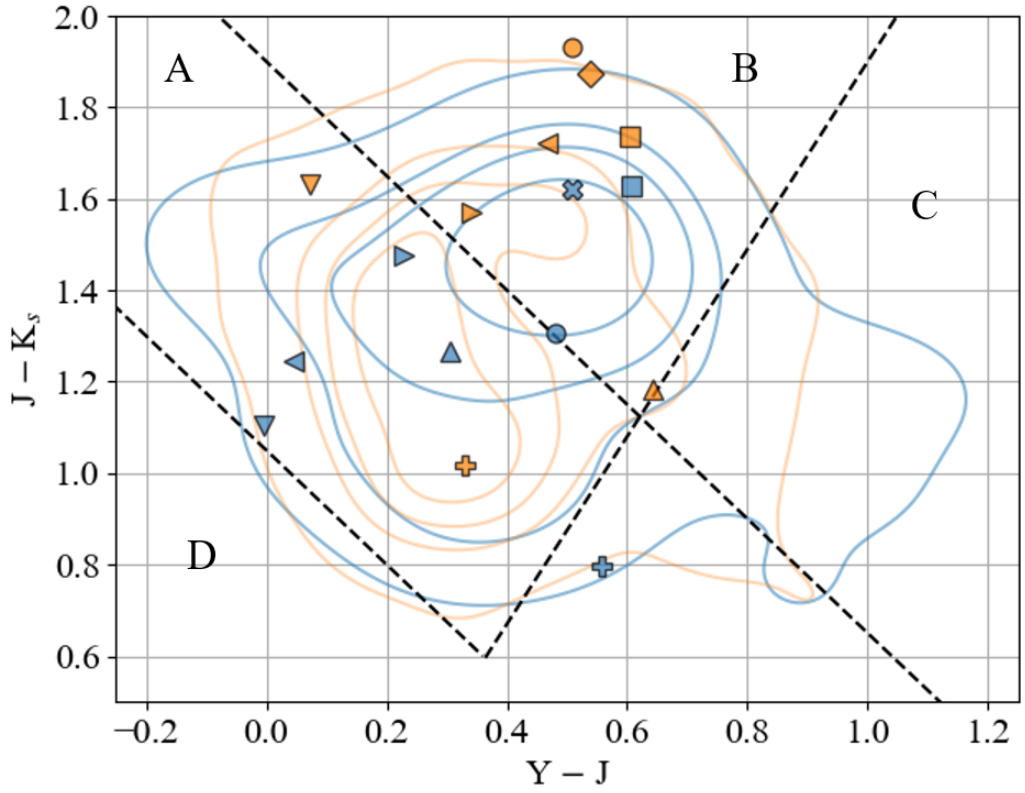

Figure 6.22: VMC colour-colour diagram of sources classified as AGN that have (blue) and have not (orange) been detected in the radio. The t-SNE selected AGN sample from Chapter 5 were also plotted. The contours represent a probability distribution which are in intervals of 0.2. The regions A, B, C and D are as described Figure 5.9.

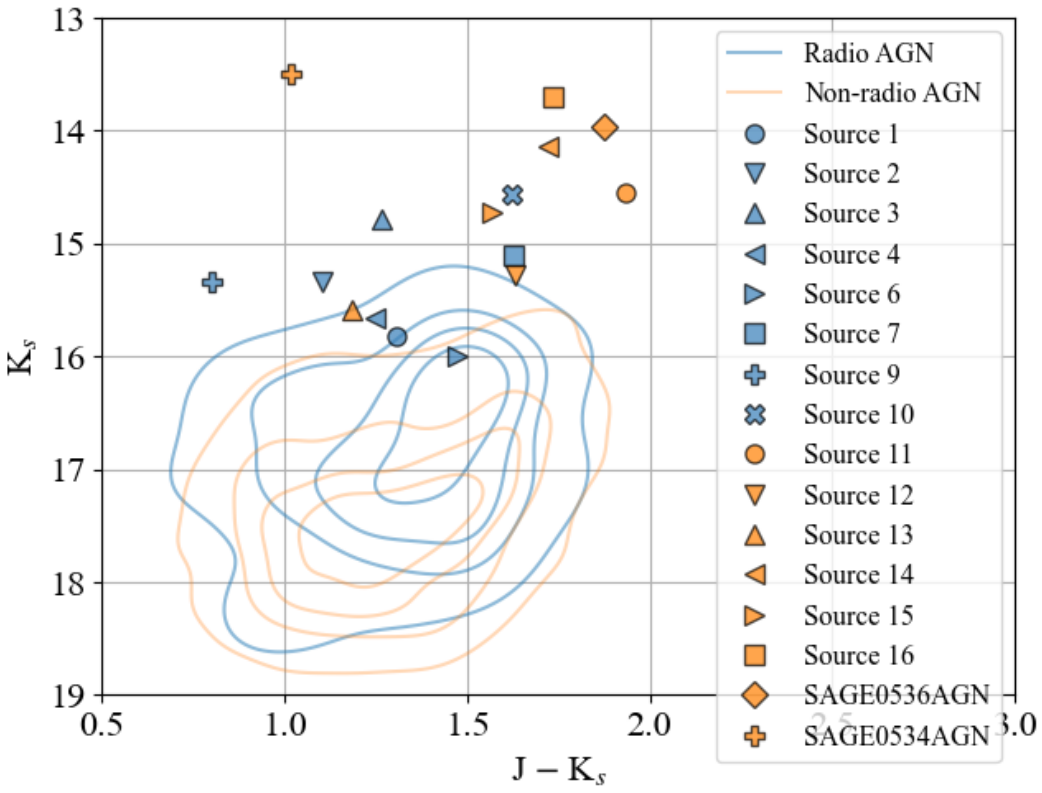

Figure 6.23: VMC colour–magnitude diagram of sources classified as AGN that have (blue) and have not (orange) been detected in the radio. The t-SNE selected AGN sample from Chapter 5 were also plotted. The contours represent a probability distribution which are in intervals of 0.2.

cluster similar sources together on a t-SNE map. In this work:

- I plotted known sources on the generated t-SNE maps of the SMC and LMC to identify clusters of AGN, low-z galaxies, blazars, high proper-motion stars and stellar sources that are often misidentified as AGN such as YSOs, PNe and dusty evolved stars.

- I showed how the properties (from Gaia, VMC and AllWISE) of these sources changed across both the SMC and LMC t-SNE maps, such as Gaia $G-$ the average of $G_{\rm BP}$ and $G_{\rm RP}$, which is found to be highest in the areas where extended objects such as galaxies are found.

- I compared the classifications made as part of the Gaia DR3 survey with the identified classes across the t-SNE maps to find that the Gaia classifications of QSO, galaxies and stellar sources mostly match up with the expected areas on the t-SNE map.

I also used the supervised machine learning algorithm, probabilistic random forest, with Gaia DR3, SMASH, VMC, SAGE, unWISE and AllWISE photometric data of the SMC. I used a sample of sources with known classifications to train the PRF classifier on, after which the classifier was then used on the entire VMC field of the SMC. In this work:

- I created and trained PRF classifiers on a data set of known sources of AGN, galaxies, RGB, OB, AGB, YSOs, PNe, post-AGB/RGB, RSG and PM stars separated into SMC and LMC sources. When trained on 75% of the data and tested on the remaining 25% of the data the classifiers yielded accuracies of 0.90 ± 0.01 and 0.87 ± 0.01 for the SMC and LMC classifiers, respectively.

- I used the PRF classifier on the 28,598,800 sources in the VMC SMC field.

- I tested the classifier on classes of 19 stellar sources that were not used in training and found that all the sources were predicted to be stellar classes. 13/19 were predicted to be OB stars.

- I compared the PRF classifications with the SMC t-SNE map to find that the majority of classifications line up with t-SNE clustering of source classes.
- I displayed the distributions of the different classes across the VMC SMC field that have the most confident classifications. The distributions of the stellar sources concentrate in the area of the SMC as expected. The density of AGN and galaxy sources decreases towards the centre of the SMC as expected.
- IR selections techniques in the near-IR and mid-IR show the majority of PRF classed sources match up with these selections. The t-SNE selected sample from Chapter 4 is shown not to match up strongly with either the radio or non-radio detected AGN.

Future work will include an exploration of the classifications of the LMC, as well as the fainter and the most confused sources in both the SMC and LMC. Other features will also be considered for use in the PRF, for example contextual information which could increase the likelihood of being something based on neighbouring sources. Furthermore, predictions of the photo-z of the extragalactic sources will be made using a machine learning estimator that trains on the same group of extragalactic sources as used in the training of the classifiers for the SMC and LMC.

7 Conclusions and outlook

In this chapter I summarise the main conclusions that have been made in this work and how they relate to each other and in a wider context. Then I outline possible avenues for future work and improvements to better understand the role of the unusual AGN found in this work in the evolution of AGN and their host galaxies.

7.1 Summary

In this thesis I set out to explore the AGN population in the direction of the Magellanic Clouds, which has been well documented by photometric surveys from X-ray to radio. This multi-wavelength view would help to place AGN in the context of galaxy evolution. The Magellanic Clouds have been most recently observed in the near-IR and radio, with great improvements in depth and angular resolution, and it is these wavelengths that I have focused my attention on.

In Chapter 4 I presented the most recent radio continuum survey of the LMC, observed at 888 MHz by ASKAP. This new survey showed an improvement in the depths reached by a factor ~ 5 (in the median RMS) over previous radio surveys of the LMC, such as those done with ATCA and MOST, allowing for fainter and higher redshift extragalactic sources to be detected. ASKAP also shows significant improvement in angular resolution which allowed for greater detail of radio structures, such as radio jets/lobes from AGN and their interactions with the intra-cluster medium, to be observed. The improved angular resolution also allows for radio sources to be better traced to their optical/IR counterparts, so that a multi-wavelength view of different source classes can be obtained.

Cross-matching the catalogue of 54,612 radio point sources with mid-IR AllWISE survey data and comparing with the non-radio detected sources, and using IR selection techniques, revealed that the majority of radio sources are extragalactic. This result is corroborated with the spectral index distribution peaking at $\alpha = -0.6$, indicating the

majority of radio sources (that are detected with both ASKAP and ATCA) are extragalactic synchrotron emitters. Cross-matching the radio catalogue with known AGN in the ASKAP field revealed 190 out of 657 known AGN, $> 92\%$ of which are newly detected here, showcasing the increase of AGN that can be detected with ASKAP. This would mean that radio detections can be used to remove the majority of stellar sources in the Magellanic Clouds optical/IR survey datasets, though with the caveat of also removing the extragalactic sources with no radio detections.

Furthermore, through separating AllWISE detected radio sources into separate extragalactic sub-classes using IR selection techniques, it was found that a higher fraction of AGN is found within the brighter radio population ($F_{888} > 3$ mJy), which is not unexpected as they are some of the brightest sources in the Universe. This also revealed that a higher fraction of spiral galaxies is found in the faint population ($F_{888} < 3$ mJy), implying that at these depths we are potentially starting to see the radio emission due to star formation.

The distribution of radio spectral indices also revealed extreme spectral indices ($|\alpha| > 3$). This can be explained by variability in the radio emission between when ATCA and ASKAP were observed. These sources, if revealed to be extragalactic, could potentially be related to the sample of sources found in Chapter 5, where the more negative spectral indices were an indication of an AGN that had just turned on or off, and the radio activity traced the path of the host galaxy from star forming and into quiescence.

In Chapter 5 I used the unsupervised machine learning algorithm, t-SNE, with Gaia EDR3, VMC, AllWISE and EMU ASKAP photometric data of the SMC VMC field to cluster similar sources together on a t-SNE map. This was done to find sources similar to SAGE0536AGN, the strongest 10-μm silicate emitter known, and SAGE0534AGN, a similar source with weaker silicate emission. Making it necessary that all sources had a radio ASKAP detection removed the majority of stellar sources and allowed me to concentrate on the mostly extragalactic sources left over. This revealed a small sample of 16 sources clustered together with SAGE0536AGN and SAGE0534AGN, two of which were revealed (with optical spectroscopy) to be an

evolved dusty star and an emission-line star, sources often mistaken for AGN and vice versa. The rest of the sources were revealed to be AGN from optical spectroscopy I observed myself or from past literature.

Calculating the black hole masses and Eddington ratios from spectra, and using CIGALE to model the photometric SEDs to better characterise these sources, led to the discovery that 12 out of 16 sources are in the green valley transitional phase. However, this was based on upper limits on star-formation, therefore these sources could be quenched instead. Furthermore, as these sources move away from the star-forming phase and through the green valley phase towards quiescence, the properties of the AGN change, such as the torus depletes and Eddington ratio decreases, signalling the AGN is running out of fuel. The X-ray luminosity also increases as the material that would absorb it has depleted.

Radio properties also change across this potential evolutionary sequence. The radio spectral slope starts off steep in the star-forming phase, before flattening to the expected value of $\alpha \sim -0.7$ for synchrotron radiation at the beginning of the green valley, and then steepening again as the sources move further into the green valley. Radio loudness also follows this trend, starting off quiet in the star-forming phase, becoming loudest at the beginning of the green valley, before quietening again. This implies that the radio activity of the AGN may be tracing the evolution of the host galaxy from star forming to quiescent. As this is a small sample more of these transitioning sources need to be found to ascertain whether this radio trend holds true.

The silicate strength of these sources and how it relates to the evolutionary stage of the galaxy still remains uncertain. This is partly due to lack of observations to detect if these sources have silicate emission, and partly due to not being modelled well by CIGALE, which underpredicts SAGE0536AGN and overpredicts SAGE0534AGN, putting the modelled silicate strength of the rest of the sample into question.

If the remaining AGN properties predicted by CIGALE are assumed to be true then the extreme silicate emission of SAGE0536AGN could be the result of the torus environment. SAGE0536AGN was predicted to have a lack of polar dust and the largest values for the ratio of the outer to inner radii of the dust torus, opening angle

and torus fraction, and one of the lowest values for optical depth and inclination angle. This implies that an AGN seen close to face on with a thick and wide torus, with little to no polar dust to obscure the centre of the AGN, could be the cause of the extreme silicate emission.

In Chapter 6 I expanded on the work done in Chapter 5 by investigating the rest of the sources in the SMC t-SNE map, as well as using the t-SNE algorithm on the LMC (using the same surveys) to create an LMC t-SNE map.

Using previously classified sources, I identified clusters of AGN, low-z galaxies, blazars, high proper-motion stars and stellar sources that are often misidentified as AGN (such as YSOs, PNe and dusty evolved stars) in both the LMC and SMC t-SNE maps. I compared these cluster locations with optical and IR properties and showed that the optical colour, Gaia G minus average of $G_{\rm BP}$ and $G_{\rm RP}$, can be used to find extended sources such as galaxies, similar to how VMC sharpness values can be used to do the same. I also show how higher values of IR colours such as VMC $J - K_{\rm s}$ and AllWISE W1 − W2 match up with areas of extragalactic sources.

As done with the SMC t-SNE map in Chapter 5, I identified the locations of SAGE0536AGN and SAGE0534AGN on the LMC t-SNE map. Just as in the SMC t-SNE map, these sources were found close to the stellar sources most often misidentified with AGN. However, unlike in the SMC map these sources were split up between two branches. One branch was mostly made up of AGN (SAGE0534AGN) and the other branch was mostly (active) galaxies and other extended objects (SAGE0536AGN), such as those associated with star formation in the LMC. The sources nearby to SAGE0536AGN and SAGE0534AGN require further investigation in order to ascertain their place in galaxy evolution.

In order to separate stellar sources without excluding the non-radio detected extragalactic sources, I used a supervised machine learning algorithm called a probabilistic random forest, with Gaia DR3, SMASH, VMC, SAGE, unWISE and AllWISE photometric data of the SMC. This was trained on known classes of stellar (RGB, OB, AGB, YSOs, PNe, post-AGB/RGB, RSG and PM stars) and extragalactic (AGN, galaxies) sources so that the algorithm can find more of these sources in the SMC. A

caveat of this is that the algorithm is biased towards finding sources that have been found before.

The SMC classifier yielded an overall accuracy of 0.89 ± 0.01, with the highest recall for the AGN and galaxy classes. Testing the classifier on stellar classes it has never seen before showed that the all of the sources were classed as stellar, with $< 10\%$ proabailities of being an AGN or galaxy.

Comparing the SMC t-SNE clustering of sources with the PRF classifications, this showed that the two methods – clustering and classifications – match up.

The distributions of the stellar classes across the VMC SMC field concentrate as expected in the area of the SMC, with some sources matching up with the direction to the Magellanic Bridge. The distribution of the source density of the AGN and galaxy sources decreases towards the centre of the SMC as expected.

Using the same IR selection techniques used in Chapters 4 and 5 it can be shown that the majority of PRF classifications of the VMC field match up with expected IR cuts. Though there are some overlap between stellar and extragalactic sources, mainly from the PNe and YSOs, which is expected as they are notoriously hard differentiate from extragalactic sources. Separating the AGN further into radio and non-radio detected shows that the clearest differences between radio and non-radio detected AGN are in the near-IR, where hot dust emission from the centre of the dusty torus is observed at angles that are close to face-on. As the strongest radio sources also tend to be AGN that are face-on, this could have resulted in the population of radio sources that represent some of the brightest VMC $K_{\rm s}$ band sources of the AGN classified sources.

The t-SNE selected sample from Chapter 5 is not shown to strongly match up with either the radio or non-radio detected AGN populations. The brighter magnitudes in the $K_{\rm s}$ band are expected as CIGALE models predict low inclination angles for these AGN, which means that we are seeing more directly the hot dust emission from the inner torus, therefore making for brighter magnitudes in the VMC $K_{\rm s}$ band. However, whilst of a similar VMC $J-K_{\rm s}$ colour, 12/16 of the t-SNE selected sample are brighter in VMC $K_{\rm s}$ band than the majority ($> 80\%$) of the radio detected (and non-radio detected) sources. This marks them still as extreme sources.

7.2 Future work/improvements

Future research would include continuing to search for extreme and unusual AGN, such as SAGE0536AGN, and placing them in the context of galaxy evolution. Finding more of these, as well as other types of unusual AGN, could provide valuable insight into rarely seen stages of galaxy/AGN evolution. Telescopes such as JWST would help in this endeavour as it is capable of spectroscopic observations from 4.9 – 28.8 μm, which would allow for the silicate emission to be observed up to $z \sim 1.8$. JWST is also capable of multi-object spectroscopy in the range 0.6 – 5.3 μm, which would also allow for the detection of multiple higher-z AGN whose emission lines (such as Hα and Hβ) have redshifted into the near-IR. This would allow us to determine where these unusual AGN are placed in galaxy evolution, such as whether they are restricted to green valley galaxies, or can they be seen throughout all stages of galaxy evolution?

Future work would also include further exploring AGN and their role in galaxy evolution by, for example, investigating the relationship between radio loudness/spectral index and the host galaxies state of evolution, to see if this relationship was unique to the small sample of AGN I studied in detail in Chapter 5, or is a more general trend. The unsupervised machine learning algorithm t-SNE has proven that it can be used to find these transitioning galaxies hosting AGN. The t-SNE map of the LMC and the placement of SAGE0536AGN and SAGE0534AGN on this map are currently being used to find more similar AGN in the VMC LMC field. These AGN have been put in the queue to be spectroscopically observed by SALT. These sources will then be analysed and plotted alongside the SMC sources to see if the AGN radio properties continue to trace the evolutionary state from star forming to quiescent. To calculate the radio spectral indices of the sources in direction of the Magellanic Clouds would require a second radio observation of similar or better quality at a higher/lower frequency than the current observation. This investigation will benefit from the recent MeerKAT (Jonas, 2009) observation of the LMC, which is currently being processed. This work could also be further expanded to include AGN away from the Magellanic

Clouds.

In regards to the work done in Chapter 6, future work will include an exploration of the classifications of the LMC, as well as the fainter and the most confused sources in both the SMC and LMC. The X-ray population of sources classed as AGN will also be explored to ascertain a relationship between X-ray, near-IR, mid-IR and radio emission. The double sources found where VMC tiles overlap will also be removed as it causes these sources to not be matched with the full array of photometric measurements that they should have.

Other features will also be considered for use in the PRF. For example, contextual information which could increase the likelihood of being something based on neighbouring sources. Flux measurements from a H I map, such as the one produced by GASKAP (Dickey et al., 2013; Pingel et al., 2022), that trace neutral hydrogen could be used to indicate areas of star formation associated with Magellanic stars. Source density averaged over a specified area around a source could increase the likelihood a source is stellar (high density) or extragalactic (low density).

Furthermore, predictions of the photo-z of the extragalactic sources will be made using a machine learning estimator that trains on the same group of extragalactic sources as used in the training of the classifiers for the SMC and LMC. A comparison between template fitting methods and machine learning regression algorithms will need to be made to ascertain the best method.

The PRF and t-SNE algorithms show great potential for separating the stellar from the extragalactic in complex fields, though confusion still exists in the most stellar dense fields. These algorithms could be adapted for use on similar high density stellar fields, such as for searching for AGN near the equatorial plane of the Milky Way, where the Gaia DR3 proper-motions would be more reliable than they are at the distances of the Magellanic Clouds.

Publications

Refereed

- Bozzetto L. M., et al., 2023, New ASKAP radio supernova remnants and candidates in the Large Magellanic Cloud, MNRAS, 518, 2574

- Bell C. P. M., et al., 2022, The intrinsic reddening of the Magellanic Clouds as traced by background galaxies – III. The Large Magellanic Cloud, MNRAS, 516, 824

- Pennock C. M., et al., 2022, The VMC survey – XLIX. Discovery of a population of quasars dominated by nuclear dust emission behind the Magellanic Clouds, MNRAS, 515, 6046

- Filipović M. D., et al., 2022, Mysterious odd radio circle near the large magellanic cloud - an intergalactic supernova remnant?, MNRAS, 512, 265

- Otani T., et al., 2022, Anomalous Orbital Characteristics of the AQ Col (EC 05217-3914) System, ApJ, 926, 17

- Filipović M., D. et al., 2021, Radio continuum sources behind the Large Magellanic Cloud, MNRAS, 507, 2885

- Pennock C. M., et al., 2021, The ASKAP-EMU Early Science Project: 888 MHz radio continuum survey of the Large Magellanic Cloud, MNRAS, 506, 3540

- Jadhav V. V., Pennock C. M., Subramaniam A., Sagar R., Nayak P. K., et al., 2021, UOCS – III. UVIT catalogue of open clusters with machine learning-based membership using Gaia EDR3 astrometry, MNRAS, 503, 253

- Bell C. P. M., et al., 2020, The intrinsic reddening of the Magellanic Clouds as traced by background galaxies – II. The Small Magellanic Cloud, MNRAS, 499, 993
- Joseph T. D0,. et al., 2019, The ASKAP EMU Early Science Project: radio continuum survey of the Small Magellanic Cloud, MNRAS, 490, 1202
- Bell C. P. M., et al., 2019, The intrinsic reddening of the Magellanic Clouds as traced by background galaxies – I. The bar and outskirts of the Small Magellanic Cloud, MNRAS, 489, 3200
- Kovacs T., et al., 2019, Quasar Candidates behind the Milky Way Disk and M31, RNAAS, 3, 1, 3 (2019).

Non-refereed

- Pennock C. M., van Loon J. , Bell C. P. M. , Filipović M. D., Joseph T. D., Vardoulaki E., 2020, Discovering exotic AGN behind the Magellanic Clouds, Proceedings paper of the IAU symposium "Nuclear Activity in Galaxies Across Cosmic Time" (Ethiopia), IAUS, 356, 335

A Sources spectroscopically observed

I list the sources I spectroscopically observed with SAAO 1.9m telescope in 2019 and the sources observed on my behalf in 2021 in Table A.1.

I list the sources that were spectroscopically observed with SALT in Table A.2 and A.3, where I list the proposals under which they were observed and the classifications that have been made and redshifts that have been measured.

Table A.1: Sources spectroscopically observed using the SAAO 1.9m telescope during observing runs in 2019 and 2021.

Target Name	RA	DEC	Date Observed
C031	01:34:22.9	−73:18:11	23-24/10/2019
C034	00:25:00.0	−72:33:01	25-26/10/2019
C014	01:38:23.4	−72:36:53	25-26/10/2019
C132	01:34:20.5	−73:18:07	25-26/10/2019
C010	00:20:23.9	−73:20:21	25-26/10/2019
C023	04:17:53.6	−73:15:56	25-26/10/2019
C139	20:35:00.3	−62:34:54	26-27/10/2019
C136	00:08:22.9	−73:52:52	26-27/10/2019
MQ120	00:08:37.88	−72:33:46.2	26-27/10/2019
C141	01:46:29.7	−72:48:45	26-27/10/2019
B009	00:58:21.53	−72:25:02.5	26-27/10/2019
A129	00:24:46.29	−74:50:13.5	26-27/10/2019
MQ118	05:32:06.08	−66:30:23	26-27/10/2019
C092	05:25:25.3	−67:29:24	26-27/10/2019
C124	06:10:06	−66:11:33	26-27/10/2019
C122	23:06:53.8	−34:39:09	29-30/10/2019
A056	00:33:24.08	−74:13:57.9	29-30/10/2019
C032	00:26:12.6	−72:37:05	29-30/10/2019
C045	01:03:10.2	−71:51:54	29-30/10/2019
A244	01:14:07.99	−72:32:43.3	29-30/10/2019
C012	01:04:50	−70:21:06	29-30/10/2019
C028	04:23:57.3	−72:47:02	29-30/10/2019
C101	05:27:56	−67:25:35	29-30/10/2019
C137	20:39:42.1	−59:57:32	30-31/10/2019
A288	00:44:54.39	−74:17:50	30-31/10/2019
A060	00:38:57.67	−72:48:57.9	30-31/10/2019
A023	00:50:57.44	−73:12:49.0	30-31/10/2019
A053	01:36:04.46	−72:13:15.4	30-31/10/2019
C138	20:39:51.1	−59:57:48	31/10-1/11/2019
A148	01:15:04.9	−73:28:16.4	31/10-1/11/2019
A075	01:13:37.08	−74:27:55.3	31/10-1/11/2019
C158	03:40:08.52	−12:49:05.9	31/10-1/11/2019
MQ108	05:12:55.5	−72:53:10.6	31/10-1/11/2019
C127	06:10:51.4	−65:24:08	31/10-1/11/2019
C128	06:09:49.9	−65:28:03	31/10-1/11/2019
C042	00:36:01.6	−72:21:12	1-2/11/2019
A133 (red)	00:51:39.93	−72:38:17.6	1-2/11/2019
A213	01:37:07.45	−74:23:39.2	1-2/11/2019

Target Name	RA	DEC	Date Observed
MQ115	04:26:57.52	−68:46:15.6	1-2/11/2019
C026	04:17:45.8	−73:15:08	1-2/11/2019
C030	06:07:55	−65:52:31	1-2/11/2019
C126	06:08:57.8	−65:46:01	1-2/11/2019
C140	20:36:03.1	−58:10:13	2-3/11/2019
A142	00:52:45.52	−72:44:01.1	2-3/11/2019
A152	00:39:10.78	−71:34:09.9	2-3/11/2019
A314	00:31:25.48	−71:50:03	2-3/11/2019
C157	01:53:51.51	−50:31:37.7	2-3/11/2019
C059	05:50:38.5	−69:55:41	2-3/11/2019
C063	06:03:27	−72:48:36	2-3/11/2019
C133	00:06:29.2	−74:01:33	3-4/11/2019
A103	00:39:53.96	−72:24:08.3	3-4/11/2019
A111	01:17:32.64	−74:39:34.9	3-4/11/2019
A124	01:02:42.18	−73:24:41.6	3-4/11/2019
D006	04:24:15.7	−70:34:15.4	3-4/11/2019
D001	05:31:20.6	−71:31:22.7	3-4/11/2019
C130	06:08:49.5	−65:44:41	3-4/11/2019
C134	00:06:45.9	−72:08:01	4-5/11/2019
C147	00:51:16.8	−73:40:02	4-5/11/2019
C009	00:20:43.4	−73:21:26	4-5/11/2019
C025	04:17:59.3	−73:16:58	4-5/11/2019
C125	06:08:42.4	−65:47:14	4-5/11/2019
C029	06:08:52	−65:43:50	4-5/11/2019
C011	00:19:56.9	−73:22:17	5-6/11/2019
A036	00:31:56.89	−73:31:13.6	5-6/11/2019
A045	00:59:34.14	−72:09:43.2	5-6/11/2019
MQ031	01:27:17.35	−71:04:10.2	5-6/11/2019
C058	05:12:05.4	−70:32:04	5-6/11/2019
C062a	05:44:24	−72:51:07	5-6/11/2019
C129	06:11:17	−66:09:22	5-6/11/2019
C135	00:06:48.3	−72:22:52	13-14/11/2019
C148	00:55:53.5	−72:42:01	13-15/11/2019
A039	00:43:12.46	−73:46:46.8	14-15/11/2019
C150	00:57:35.4	−73:46:30	14-15/11/2019
A043	00:47:08.38	−74:30:09.5	14-15/11/2019
C152	00:59:09.7	−74:02:37	14-15/11/2019
C124	06:10:06	−66:11:33	14-15/11/2019, 23-24/11/2019
C174	05:08:11.4	−68:06:11	14-15/11/2019

Target Name	RA	DEC	Date Observed
A144	00:41:25.33	−70:57:43.8	15-16/11/2019
C047	01:13:18.3	−72:25:19	15-16/11/2019
A038	00:58:38.17	−73:34:49.6	15-16/11/2019
C155	01:13:40.6	−71:32:19	15-16/11/2019
C171	04:17:28.3	−68:37:30	15-16/11/2019
C184	05:34:44.18	−67:37:50.1	15-16/11/2019
C162	05:19:42.4	−65:02:16	15-16/11/2019
C177	06:23:36	−64:34:41	15-16/11/2019
A138	00:26:02.54	−72:47:18	16-17/11/2019
A040	01:17:34.7	−72:50:42.5	17-18/11/2019
A030	01:06:00.03	−73:31:25.9	17-18/11/2019
A111	01:17:32.64	−74:39:34.9	17-18/11/2019
C182	04:55:01.5	−64:49:57	17-18/11/2019
C202	05:15:09.8	−63:46:07	17-18/11/2019
C209	05:25:04.4	−64:40:50	17-18/11/2019
A045	00:59:34.14	−72:09:43.2	18-19/11/2019
A021	00:35:18.24	−73:18:26	18-19/11/2019
C201	04:52:49.8	−64:46:23	18-19/11/2019
C186	05:05:24.35	−67:34:35.4	18-19/11/2019
C168	06:08:21.8	−65:11:44	18-19/11/2019
C208	05:25:02.9	−64:40:15	18-19/11/2019
C123	23:52:40.1	−55:35:23	19-21/11/2019
A201	00:57:32.75	−72:13:02.3	19-20/11/2019
MQ077	01:26:10.53	−71:14:10.1	19-20/11/2019
MQ123	00:39:49.40	−66:07:26.3	19-20/11/2019
C187	04:51:38.41	−71:02:06	19-20/11/2019
MQ119	05:20:34.68	−68:35:18.2	19-20/11/2019
C216	06:17:24	−70:40:49	19-20/11/2019
C208	05:25:02.9	−64:40:15	19-20/11/2019
C019	00:52:47.9	−71:15:32	20-21/11/2019
C185	04:38:31.19	−68:12:00.4	20-21/11/2019
C169	06:11:46.5	−65:59:02	20-21/11/2019
C196	04:35:22.5	−64:26:29	20-21/11/2019
MQ112	00:53:27.4	−73:45:48.9	21-22/11/2019
C036	00:24:07.8	−72:41:13	21-22/11/2019
C035	00:26:17.2	−72:33:02	21-22/11/2019
C178	04:33:04.1	−67:52:56	21-22/11/2019
C113	05:27:56.9	−67:24:09	21-22/11/2019
C167	05:14:17.9	−72:20:19	21-22/11/2019, 23-24/11/2019

Target Name	RA	DEC	Date Observed
C061	06:02:12.8	−72:27:36	21-22/11/2019
A201	00:57:32.75	−72:13:02.3	22-23/11/2019
A286	00:21:08.27	−72:17:42	22-23/11/2019
A138	00:26:02.54	−72:47:18	22-23/11/2019
C200	04:51:36	−63:54:56	22-23/11/2019
C176	05:58:43.2	−65:15:53	22-23/11/2019
C199	04:48:14.7	−64:10:45	22-23/11/2019
A062	00:53:56.22	−70:38:04.4	23-24/11/2019
MQ044	00:32:25.73	−73:59:08.3	23-24/11/2019
MQ121	01:02:48.28	−72:06:16	23-24/11/2019
C023	04:17:53.6	−73:15:56	23-24/11/2019
C166	05:30:41.1	−66:05:35	23-24/11/2019
C159	05:06:47.96	−19:36:50.7	23-24/11/2019
A310	00:36:16.99	−74:31:31.3	24-25/11/2019
A089	00:41:05.74	−70:14:34.8	24-25/11/2019
A300	01:02:37.76	−75:05:30.7	24-25/11/2019
C172	04:14:04.9	−69:34:11	24-25/11/2019
C210	05:26:13.5	−64:31:41	24-25/11/2019
C170	04:17:31.5	−68:37:40	24-25/11/2019
C212	05:29:07.3	−63:58:38	24-25/11/2019
L01	04:37:48.48	−73:15:13.0	13-14/01/2021
P11	04:55:06.6	−69:17:02.5	13-14/01/2021
P23	05:41:13.13	−64:11:53.6	13-14/01/2021
P47	05:58:46.46	−74:59:05.2	13-14/01/2021
P12	06:15:04.4	−66:17:16.3	13-14/01/2021
P11 South	04:55:06.6	−69:17:02.5	14-15/01/2021
L05	05:58:49.49	−67:08:00.0	14-15/01/2021
C008 (mouse)	06:08:55.55	−65:52:55.0	14-15/01/2021
P51	04:21:16.16	−67:10:25.2	15-16/01/2021
S28	04:35:02.2	−65:19:48.0	15-16/01/2021
P16	05:33:53.53	−66:43:24.5	15-16/01/2021
C184	05:34:44.44	−67:37:50.1	15-16/01/2021
P14	05:53:42.42	−66:52:43.1	15-16/01/2021
P21	04:36:58.58	−66:22:50.6	16-17/01/2021
L03	04:37:40.40	−73:15:06.0	16-17/01/2021
L02	04:37:55.55	−73:14:32.0	16-17/01/2021
L04	06:03:07.7	−73:25:29.0	16-17/01/2021
C131	06:09:07.7	−65:44:54.0	16-17/01/2021
S62	06:10:41.41	−65:06:59.0	16-17/01/2021
S15	06:13:55.55	−66:03:33.0	16-17/01/2021

Target Name	RA	DEC	Date Observed
P26	04:29:38.38	−65:26:36.5	17-18/01/2021
P05	04:38:51.51	−72:17:12.5	17-18/01/2021
C192	04:53:05.5	−69:41:06.4	17-18/01/2021
C218	05:20:57.57	−70:24:53.9	17-18/01/2021
C219	05:34:49.49	−66:08:57.2	17-18/01/2021
S20	05:50:17.17	−66:02:18.0	17-18/01/2021
S11	05:53:51.51	−67:33:40.0	17-18/01/2021
P39	04:50:09.9	−68:25:33.2	18-19/01/2021
P53	04:50:53.53	−62:38:54.0	18-19/01/2021
P31	05:00:08.8	−73:37:49.5	18-19/01/2021
P38	05:09:53.53	−69:14:36.4	18-19/01/2021
P02	06:15:00.0	−72:56:42.7	18-19/01/2021
P01	06:21:06.6	−74:09:26.1	18-19/01/2021
P34	06:23:16.16	−69:27:36.5	18-19/01/2021
P37	05:05:59.59	−69:39:53.5	19-20/01/2021
P27	05:08:44.44	−64:28:31.6	19-20/01/2021
P24	05:33:58.58	−64:20:24.8	19-20/01/2021
P45	05:43:34.34	−64:22:58.3	19-20/01/2021
P10	05:44:05.5	−68:27:21.4	19-20/01/2021
P20	05:49:13.13	−64:29:29.1	19-20/01/2021
P36	05:51:23.23	−70:03:13.8	19-20/01/2021

Table A.2: SALT observations made over the course of this work.

Name	Proposal	Date observed
LMCtSNE2	2022-1-SCI-022	07,17/10/2022
LMCtSNE1	2022-1-SCI-022	05/10/2022
LMCtSNE3	2022-1-SCI-022	04/10/2022
LMCtSNE4	2022-1-SCI-022	27/09/2022
Source 9	2022-1-SCI-022	09/09/2022
Source 14	2022-1-SCI-022	07/09/2022
Source 6	2022-1-SCI-023	20/07/2022
Source 8	2022-1-SCI-023	20/07/2022
SAGE0534AGN	2021-2-SCI-017	17/03/2022
SAGE0534AGN	2021-2-SCI-018	20/11/2021
OGLE source	2021-2-SCI-017	15/11/2021
C217	2021-2-SCI-017	12/11/2021
C175+FRII	2021-1-SCI-029	23,25/09/2021
C175+PAGB	2021-1-SCI-029	23,24/09/2021
TSNE16	2021-1-SCI-032	01/09/2021
TSNE12	2021-1-SCI-032	01/09/2021
TSNE11	2021-1-SCI-032	01/09/2021
TSNE10	2021-1-SCI-032	01/09/2021
TSNE13	2021-1-SCI-029	14,23/07/2021
Source 5	2021-1-SCI-029	17/07/2021
TSNE15	2021-1-SCI-032	13/06/2021
ER1	2020-2-SCI-025	25/11/2020
ER6	2020-2-SCI-025	09/11/2020
SA12	2020-1-SCI-028	29/06/2020
Helicopter	2019-2-SCI-045	12/01/2020
FuzzyOrange	2019-2-SCI-041	05,07/12/2019
Mouse	2019-2-SCI-045	04/12/2019
Blazar	2019-2-SCI-041	03/11/2019
FuzzyOrange	2019-2-SCI-041	02/11/2019
2MASS 01260512-7055408 (companion)	2019-1-SCI-032	20/09/2019
2MASS 01260512-7055408 (jets)	2019-1-SCI-032	20/09/2019

Table A.3: Coordinates of the SALT observations made over the course of this work.

Name	RA (J2000)	Dec. (J2000)	Notes
LMCtSNE2	06:15:04.01	−66:17:16.4	AGN, $z \sim 0.18$
LMCtSNE1	05:58:48.50	−67:08:00.1	AGN, $z \sim 1.26$
LMCtSNE3	05:33:57.69	−64:20:24.9	AGN, $z \sim 0.063$
LMCtSNE4	05:01:10.84	−73:36:35.0	AGN, $z \sim 1.37$
Source 9	01:21:08.43	−73:07:13.1	AGN, $z \sim 0.99$
Source 14	01:36:04.20	−72:13:15.4	AGN, $z \sim 0.41$
Source 6	01:14:08.00	−72:32:43.4	AGN, $z \sim 1.06$
Source 8	01:22:36.94	−73:10:16.7	Emission-line star
SAGE0534AGN	05:34:44.17	−67:37:50.1	AGN, $z \sim 1.01$
OGLE source	05:49:30.13	−70:35:43.6	Variable M-type star (oxygen-rich AGB)
C217	06:08:43.60	−70:50:07.7	AGN, $z \sim 0.60$
C175+FRII	05:19:00.90	−68:01:58.0	
C175+PAGB	05:19:00.90	−68:01:58.0	
TSNE16	00:49:52.50	−69:29:56.0	
TSNE12	01:13:37.10	−74:27:55.0	
TSNE11	01:35:05.10	−75:07:03.0	
TSNE10	01:27:17.30	−71:04:10.0	
TSNE13	00:57:32.8	−72:13:02.0	
Source 5	00:48:25.70	−72:44:03.0	Carbon star
TSNE15	01:19:13.4	−71:08:51.0	
ER1	00:36:59.04	−71:38:10.9	AGN, $z \sim 0.46$
ER6	01:36:34.37	−71:38:38.7	AGN, $z \sim 0.60$
SA12	00:36:59.25	−71:38:13.6	AGN, $z \sim 0.46$
Helicopter	06:02:54.1	−71:03:10.0	Galaxy, $z \sim 0.08$
FuzzyOrange	01:21:28.50	−73:20:04.0	
Mouse	06:08:54.7	−65:52:55.0	Galaxy, $z \sim 0.037$
Blazar	00:57:16.0	−70:40:46.0	Not a blazar. AGN, $z \sim 0.59$
2MASS 01260512-7055408 (companion)	01:26:05.12	−70:55:40.8	Stars.
2MASS 01260512-7055408 (jets)	01:26:05.12	−70:55:40.8	Star. "Jets" too faint.

B IRAF

Here I outline the commands used in IRAF to reduce the spectroscopy data observed with the SAAO 1.9m and SALT telescopes:

1. NOAO IMRED CCDRED : Library of functions to deal with general CCD reduction.

2. ZEROCOMBINE : Used to combine the bias files into one .fits file.

3. CCDPROC : Used to bias correct the science, arc and flat files. Where 'overscan', 'trim' and 'zerocor' are set to 'yes' and 'darkcor', 'flatcor', 'illumcor' are set to 'no'. 'trim' is optional and is used to remove bad data from the edges of the image, the boundaries to cut are specified with 'biasec' and 'trimsec'.

4. FLATCOMBINE : Used to combine the flat files into one .fits file.

5. RESPONSE : Used to normalise the spectroscopic flat. Requires a function to be fit to the response curve.

6. CCDPROC : Used for flat fielding the science and arc files. Set 'flatcor' to 'yes'.

7. STSDAS LACOS_SPEC : Used to remove the cosmic rays in the science files.

8. TWODSPEC LONGSLIT : Library of functions for dealing with 2D longslit spectroscopy reduction.

9. IDENTIFY : Used to identify the wavelength the emission lines are at in the arcs and their corresponding image coordinates.

10. REIDENTIFY : Used to identify the same features found with IDENTIFY across the entire 2D image, as well as in other arc files.

11. FITCOORDS : Fit a transformation function to the image coordinates of features identified with IDENTIFY and REIDENTIFY and their corresponding wavelengths.

12. TRANSFORM : Use the transformation function to calibrate the wavelength of the science images.

13. SETAIRMASS : Used to set the airmass of the science images to correct airmass. Set 'observa' to 'saao'.

14. STANDARD : Used to assign the correct light profile to the standard star.

15. SENSFUNC : Used to create a response curve from the standard star and its expected light profile.

16. CALIBRATE : Used to flux calibrate the science images using the response curve from the standard star.

17. APALL APEXTRACT : Used to extract the 1D spectrum from the calibrated 2D image.

18. ONEDSPEC wspectext : Used to write the 1D spectroscopy file to a readable .txt file that can be manipulated by other programs, such as Python.

When more than one science image was available the 2D science images could be combined with IMCOMBINE.

Note that SALT spectroscopy data is received with steps 1 – 6 already done by the SALT pipeline (Crawford et al., 2010). The SALT RSS .fits files had two headers, which IRAF cannot handle. Therefore the first step for using the SALT files was to use IMCOPY in IRAF to make a copy with a specified header. Due to no standard stars being observed alongside the SALT observations (to maximise the signal-to-noise obtained for the science observations), steps 13 – 16 were not undertaken.

C PRF Feature Importances

Here I plot (in Figures C.1, C.2, C.3 and C.4) the feature importances, in order of importance, of the SMC and LMC probabilistic random forest machine learning classifiers described in Chapter 3, Section 3.4.5.4.

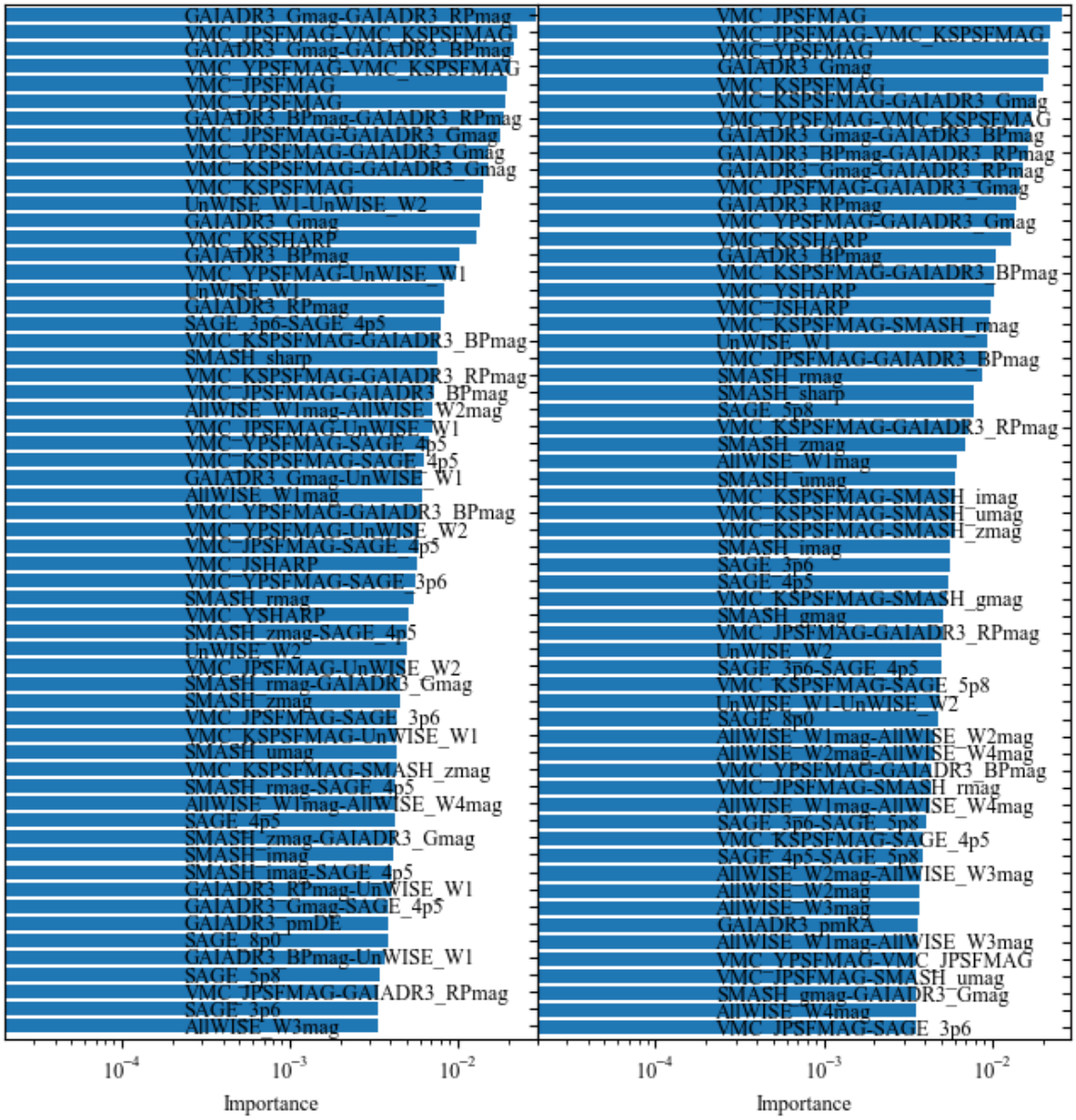

Figure C.1: Feature importances of the SMC (left) and LMC (right) probabilistic random forest classifier. Features ranked from 1st to 60th.

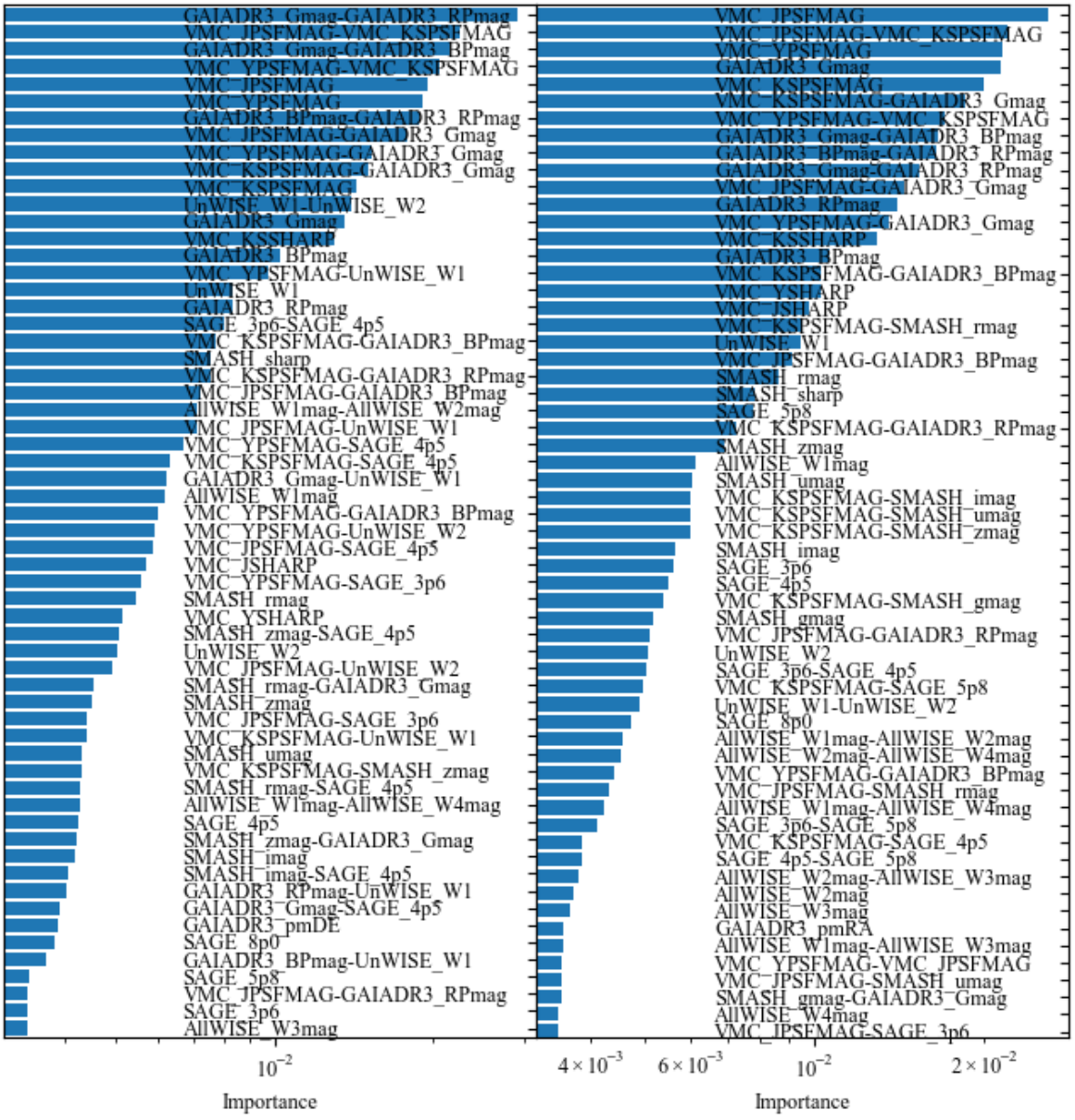

Figure C.2: Feature importances of the SMC (left) and LMC (right) probabilistic random forest classifier. Features ranked from 61st to 120th.

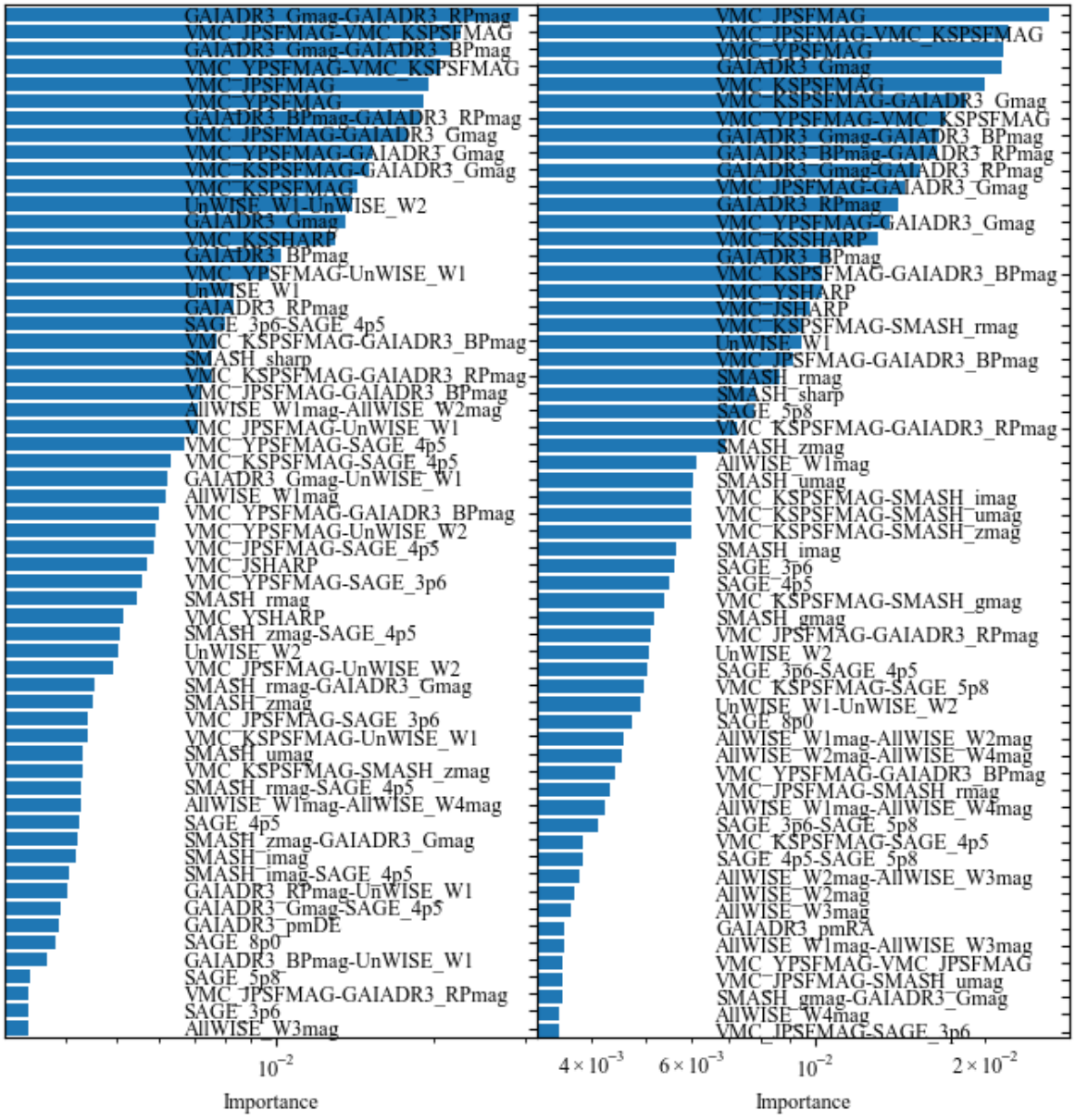

Figure C.3: Feature importances of the SMC (left) and LMC (right) probabilistic random forest classifier. Features ranked from 121st to 180th.

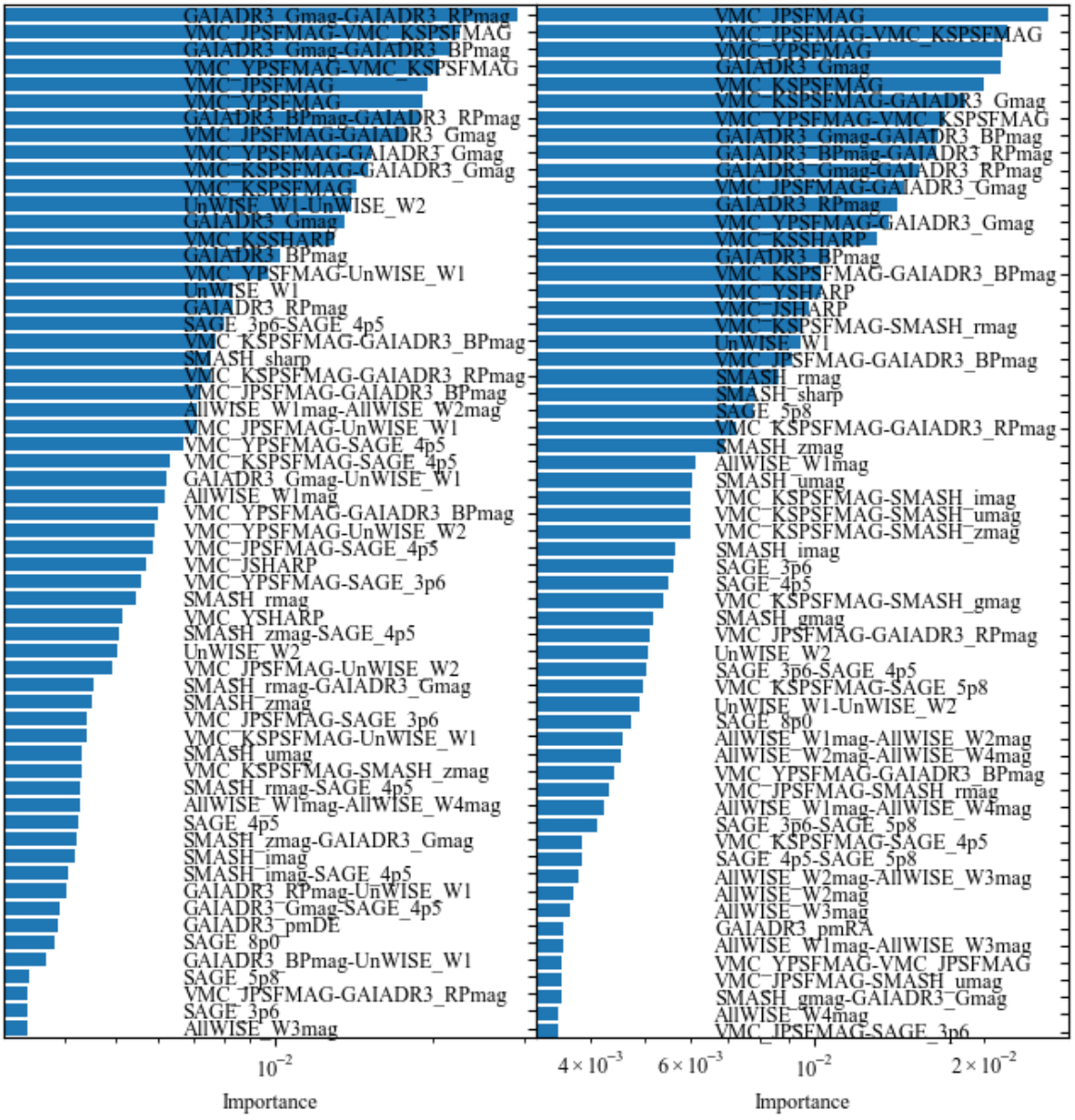

Figure C.4: Feature importances of the SMC (left) and LMC (right) probabilistic random forest classifier. Features ranked from 181st to 237th.

D ASKAP Magellanic and Galactic sources

I discuss a few types of Magellanic and Galactic sources here. These sources were found by cross-matching with various catalogues using TOPCAT[1] (Taylor, 2005) with an initial search radius of $10''$ [2], after which visual inspection was used to determine whether they are true counterparts. All quoted flux densities from the comparison catalogues are integrated flux densities unless stated otherwise.

D.1 Planetary Nebulæ and other emission-line objects

Leverenz et al. (2017) (see also Filipović et al., 2009) presented radio detections of 28 PNe in the LMC, but these were mostly detected at 4.8 GHz (26) with only 14 detected at 1.4 GHz as the radio emission is dominated by free–free emission. Simbad (Wenger et al., 2000) lists 1334 objects within the ASKAP footprint that have at one time been classed as a PN or candidate PN. Of these, 114 have counterparts in our ASKAP catalogue within 10 arcsec. An example of this list of objects is shown in Table D.1 (the full table is available online as a supplementary material to Pennock et al., 2021). Note that in some cases we believe the (candidate) PN is not the radio counterpart, especially for positional differences of > 6 arcsec.

However, only 43 of these have a primary Simbad class of PN (or candidate PN) and two of these turn out not to be PNe (extended radio emission) and another four have highly uncertain radio detections (a noise spike in one case, two cases of blends, and HD 269404 described below), leaving 37 good radio detections of likely PNe. Another 23 are classed as candidate post-asymptotic giant branch (post-AGB)

[1] http://www.star.bris.ac.uk/∼mbt/topcat/

[2] to allow for inaccuracies in both the ASKAP and literature positions.

Table D.1: Example of ASKAP 888 MHz radio detections within 10 arcsec of sources at one time classed as potential PNe. The positions are those of the radio source, and the distance (Δ) is that between the radio source and the coordinates listed in Simbad. The most likely classes are Planetary Nebula (PN), candidate post-AGB object (pAGB), emission-line star (em$*$), Young Stellar Object (YSO) or H II region (H II). See comments where there is doubt. The full table is available online in the supplementary material to Pennock et al. (2021).

Name	EMU ID	RA (J2000) $h\ m\ s$	Dec (J2000) $\circ\ \prime\ \prime\prime$	...
SMP LMC 104	EMU ES J042437.3−694221	04 24 37.4	−69 42 21	...
SMP LMC 6	EMU ES J044738.8−722821	04 47 38.8	−72 28 21	...
SMP LMC 5	EMU ES J044808.4−672606	04 48 08.4	−67 26 07	...
SMP LMC 7	EMU ES J044829.6−690813	04 48 29.6	−69 08 13	...
SMP LMC 10	EMU ES J045108.7−684904	04 51 08.7	−68 49 05	...
Sk −70 10	EMU EC J045325.9−703541	04 53 25.9	−70 35 42	...
...				

Name	— peak —		— integrated —		Δ	Class	Comments
	F (mJy)	σ (mJy)	F (mJy)	σ (mJy)	$\prime\prime$		
SMP LMC 104	0.31	0.06	0.34	0.06	0.9	PN?	
SMP LMC 6	0.80	0.05	0.78	0.05	0.9	PN	
SMP LMC 5	0.39	0.05	0.34	0.04	0.7	pAGB	
SMP LMC 7	0.45	0.05	0.48	0.05	0.5	PN	
SMP LMC 10	0.45	0.05	0.47	0.05	1.3	PN	
Sk −70 10	0.44	0.09	1.45	0.30	3.1	H II	Contains B0 III[e]
...							

objects, of which one has an uncertain radio detection (and is discarded) and another one is classed as a carbon-rich Wolf–Rayet star.

This radio sample of 114 radio sources also includes 31 emission-line stars and (three) YSOs, of which two are discarded because of their large positional differences >9 arcsec, and another 19 compact H II regions or other extended and/or bright radio sources that are likely also H II regions. It is unclear whether there are any post-AGB objects among the emission-line stars.

Among the more interesting sources are HD 269404 = IRAS 05216−6753 ($F_{888} = 3.77 \pm 0.11$ mJy) – still listed in Simbad as a PN but probably a young, massive dust-enshrouded star (van Loon et al., 2001, 2010); and LMC SySt-23 ($F_{888} = 0.37 \pm 0.05$ mJy) – a confirmed D-type symbiotic binary (van Aarle et al., 2011; Akras et al., 2019). One of the most interesting true PNe is SMP LMC 83 ($F_{888} = 1.36 \pm 0.07$ mJy) – with a nitrogen-rich Wolf–Rayet type central star ([WN4.5:]) betraying a massive AGB star progenitor commensurate with it being the radio loudest PN in this sample.

The cumulative histogram of integrated flux densities of the PNe (Figure D.1) rises rapidly towards the faint end. The turnover of the distribution at $F_{888} \sim 0.5$ mJy can be attributed to incompleteness; a similar turnover can be seen in the integrated flux density distribution of candidate post-AGB objects. However, the candidate post-AGB distribution is less steep at the faint side of the turnover, compared to the PNe. This may reflect the evolutionary timescale, with a low-mass post-AGB system evolving more slowly and therefore not ionizing its circumstellar envelope as efficiently. Possibly, the brighter sources descend from massive AGB progenitors and are surrounded by a larger ionized envelope and/or a higher electron density compared to lower-mass systems. The distinction between post-AGB and PN is not sharply defined, though, and some of the post-AGB objects may well be considered to be PNe.

The emission-line star integrated flux densities function (Figure D.1) resembles that of the candidate post-AGB stars at the faint end. At the bright end, however, there is a clear excess of emission-line stars over PNe/post-AGB candidates. Given that the H II sources within this compilation of 114 radio sources also often had emission-line stars associated with them, it is likely that those bright emission-line stars are massive

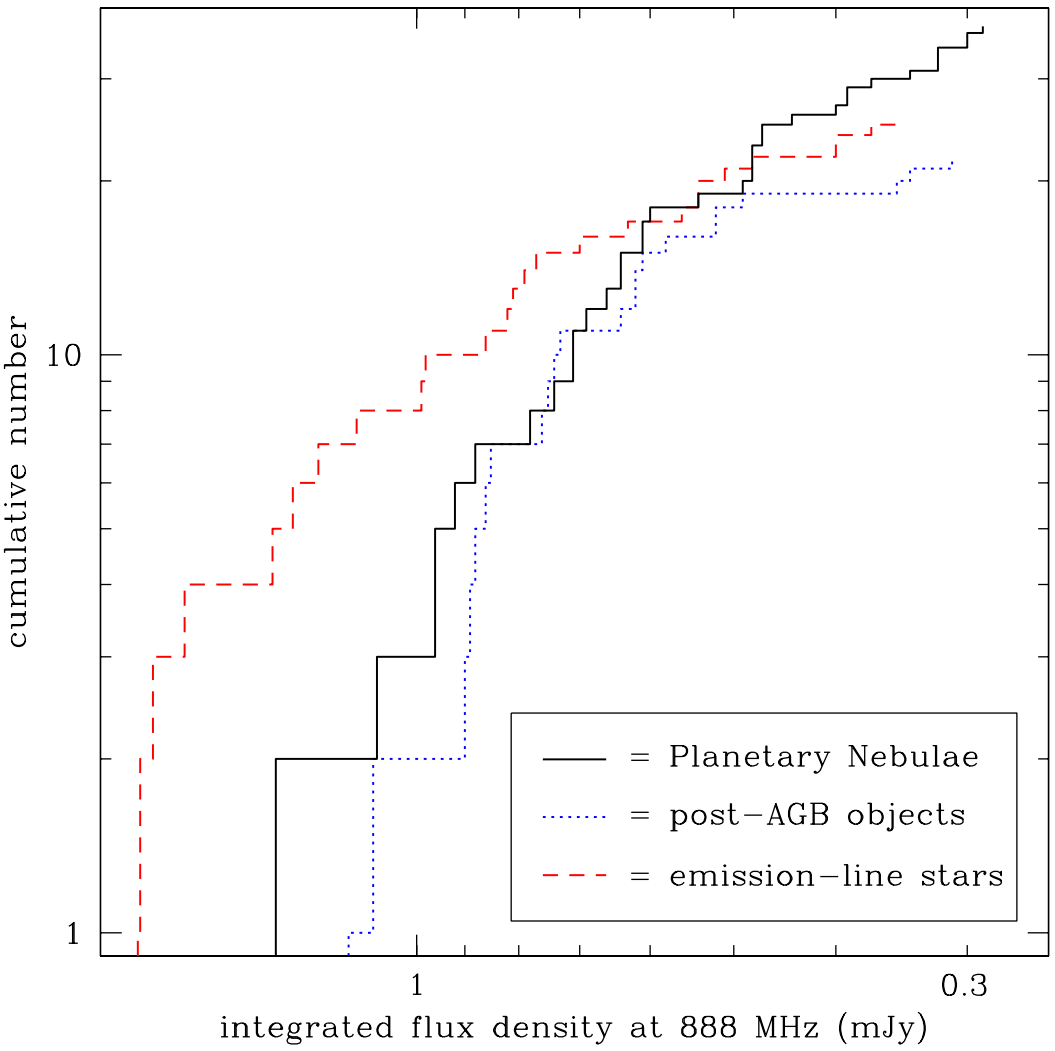

Figure D.1: Cumulative 888 MHz flux distributions of planetary nebulæ, post-AGB objects and emission-line stars.

stars. It is not clear whether these are young or evolved, but one could speculate that young sources are more often associated with extended radio emission from an H II region and evolved stars are more often isolated and thus unresolved. This is somewhat corroborated by the fact that of all stars with supergiant luminosity class (I or II, according to Simbad), five are unresolved by ASKAP and only one is found in extended radio emission, versus a ratio of 20:11 for luminosity classes III–V. Forbidden line emission ("[e]") is marginally more prevalent among unresolved radio sources (15 out of 27) as well, compared to stars associated with extended radio emission (5 out of 11).

D.2 Young stellar objects

YSOs become radio continuum sources once they ionize a sufficient amount of gas and free–free emission yields a relatively flat continuum emission spectrum extending from the near-IR far into the radio regime. They are also found in complex star-forming regions that are often embedded in H II regions that exhibit bright free–free radio emission. As seen in the previous section, some YSOs have been mistaken for PNe. We here examine the ASKAP radio detections of a small sample of particularly well characterised YSOs from the groundbreaking Herschel Space Observatory spectroscopy sample of Oliveira et al. (2019).

Of the seventeen YSOs (within the LMC), three had initially been associated with one compact region of star formation and would be blended in our radio survey. These were not detected; among the remainder, nine were detected (within a search radius of 10 arcsec; all were found within ≤ 6 arcsec) including all three YSOs from the unique star-forming region LHA 120-N 113 in which complex organic molecules were discovered with the Atacama Large (sub-)Millimetre Array (ALMA) by Sewiło et al. (2018). Table D.2 lists their properties. They span a range in radio integrated flux densities from 0.45 to 47 mJy, and all but two are brighter than the PNe distribution (Figure D.1). Because they are among the more extremely luminous, massive and consequently radio brighter

Table D.2: ASKAP 888 MHz radio detections within 10 arcsec of sources from the Herschel spectroscopy of the YSO sample (Oliveira et al., 2019). The names are as in Oliveira et al. (2019), the positions are those of the radio source, and the distance (Δ) is that between the radio source and the YSO coordinates (IR-based).

Name	EMU ID	RA (J2000) $h\ m\ s$	Dec (J2000) $\circ\ \prime\ \prime\prime$
IRAS 04514−6931	EMU ES J045111.7−692647	04 51 11.7	−69 26 48
LHA 120-N 113 YSO-1	EMU EC J051317.6−692223	05 13 17.6	−69 22 24
LHA 120-N 113 YSO-4	EMU EC J051321.4−692239	05 13 21.4	−69 22 40
LHA 120-N 113 YSO-3	EMU EC J051324.7−692245	05 13 24.8	−69 22 46
SAGE 051351.5−672721.9	EMU ES J051351.4−672719	05 13 51.4	−67 27 19
SAGE 052202.7−674702.1	EMU ES J052202.7−674658	05 22 02.7	−67 46 59
SAGE 052212.6−675832.4	EMU EC J052213.3−675834	05 22 13.4	−67 58 34
SAGE 053054.2−683428.3	EMU ES J053054.5−683422	05 30 54.6	−68 34 23
ST 01	EMU EC J053931.0−701216	05 39 31.0	−70 12 16

Name	— peak —		— integrated —		Δ	Comments
	F (mJy)	σ (mJy)	F (mJy)	σ (mJy)	$\prime\prime$	
IRAS 04514−6931	0.50	0.07	0.45	0.06	1.9	
LHA 120-N 113 YSO-1	39.0	0.6	44.5	0.6	1.2	
LHA 120-N 113 YSO-4	40.7	0.6	47.0	0.7	1.9	
LHA 120-N 113 YSO-3	11.2	0.6	14.2	0.7	2.0	
SAGE 051351.5−672721.9	3.37	0.17	4.53	0.24	2.8	
SAGE 052202.7−674702.1	1.20	0.06	3.8	0.2	3.4	extended
SAGE 052212.6−675832.4	30.2	1.1	72.4	2.8	4.7	extended
SAGE 053054.2−683428.3	9.17	0.08	14.90	0.13	6.1	
ST 01	2.43	0.08	3.18	0.11	1.1	

YSOs they appear relatively isolated and unresolved against the ambient emission, except for two clearly extended sources both in the LHA 120-N 44 region.

The most striking result is the sharp-cut division between those YSOs detected with ASKAP and those not, in terms of the measurement of the gas temperature based on far-IR emission lines of CO detected with Herschel (Oliveira et al., 2019, their Table 6). All sources with temperatures $T > 700$ K are detected, and all sources with $T < 700$ K are not. This suggests a strong relation between the line emission from the

photo-dissociation regions at the interfaces of neutral and ionized gas, and the free–free emission from the ionized regions surrounding the nascent, presumed massive O-type stars. The brightest radio source, LHA 120-N 113 YSO-4 did not have a temperature determined from CO but it is one of the two brightest [O III] 88 μm emitters (Oliveira et al., 2019) and thus hosts a highly ionized compact or ultra-compact H II region.

D.3 Novæ and supernovæ

Of 46 novæ listed in Simbad within the ASKAP footprint, two are recovered in our ASKAP catalogue (both in the GOLD list) within $\sim 2''$: Nova LMC 1988 b ($F_{888} = 2.81 \pm 0.06$ mJy) and Nova candidate LHA 120-S 162 ($F_{888} = 0.46 \pm 0.05$ mJy). While the latter was designated Nova LMC 2001 it is a Galactic mid-M type main-sequence emission-line star (Morgan et al., 1992; Shafter, 2013). The former, Nova LMC 1988 b was a fast nova and the first neon nova discovered in the LMC (Sekiguchi et al., 1989). This implies the eruption occurred on a massive O–Ne–Mg white dwarf. Neon novæ are energetic and not rare, so it is not immediately clear why this should have been the only nova detected with flux density > mJy.

SN 1987A is detected as a bright point source, with $F_{888} = 1.1432 \pm 0.0013$ Jy. It is listed as EMU ES J053527.8−691611 in the BRONZE source list. Its proximity to 30 Doradus prevented its original inclusion in the catalogue due to the complexity of the surroundings even though SN 1987A itself stands out clearly; it was reinstated subsequently. The radio emission primarily arises from the $1\farcs6$-diameter torus, which was resolved at 92 GHz for the first time by Lakićević et al. (2011). The integrated flux density at 888 MHz have continued to increase by $\sim 27\%$ since 2013–2014 when it was ≈ 0.90 Jy (Callingham et al., 2016), but at the slower pace observed after an initial exponential growth until 2009 (day ~ 8000, Ng et al., 2013).

While radio detections are made at the locations of at least two dozen SNe in background galaxies, in none of the cases could we unequivocally ascribe the radio emission to the SN and in most cases the host galaxy dominates.

D.4 Supergiants

Various classes of massive supergiant stars could be detectable at low radio frequencies, including Luminous Blue Variables (LBVs), Wolf–Rayet (WR) stars – especially if in colliding-wind binaries – and Be or B[e] stars (massive B-type stars showing permitted emission lines from an excretion disk and/or forbidden emission lines from a more tenuous envelope, respectively).

Nine such sources from the far-IR study of van Loon et al. (2010) were pursued, resulting in the detection of four of these: (a) the dusty B[e] star IRAS 04530−6916 (van Loon et al., 2005) seen on the image as a point source with a peak flux density $\sim$ 0.9 mJy [3] though it had not made the cut for the catalogue, (b) the Of/Of?p–WR transition object Brey 3 (IRAS 04537−6922; Heydari-Malayeri & Melnick, 1992) at $F_{888} = 5.25 \pm 0.06$ mJy, (c) emission-line object IRAS 05047−6644 at $F_{888} = 12.27 \pm 0.06$ mJy and (d) IRAS 05216−6753 at $F_{888} = 3.77 \pm 0.11$ mJy – likely a massive early-type star illuminating a dusty envelope (Chen et al., 2009). It is likely the nature of the 888 MHz radiation mechanism in these sources is free–free emission from ionized circumstellar gas.

LBV S Dor may have been detected as a point source with a peak flux density $\sim$ 0.19 mJy, but it is blended with much brighter adjacent emission from the N 119 H II region. The radio emission near R 71, at $\sim$ 12 arcsec distance ($\sim$ 3 pc at the distance of the LMC) to the South–West in an otherwise sparse field, is unresolved and peaks at $\sim$ 0.21 mJy. R 71 had also not been detected within a 3σ level of 24 μJy at 5 GHz in 2015 (Mehner et al., 2017). While no circumstellar nebula appears to be present around R 71 (based on HST Hα imaging before the recent eruption of R 71), there is a hint for a very small (less than 1 arcsec or 0.25 pc) circumstellar nebula around S Dor due to the [N II] emission lines detected in a long-slit échelle spectrum (Weis, 2003). Unfortunately, the possible nebula is not detected in HST Hα imaging (Weis, 2003).

LBV nebulæ have been detected at higher radio frequencies before, for in-

[3] radio emission had been detected by Filipovic et al. (1995)

stance, R 127, R 143, S 61 and S 119 (Agliozzo et al., 2012, 2017a,b, 2019). Cross-matching these sources with the ASKAP LMC 888 MHz catalogue revealed integrated fluxes for R 127 (EMU ES J053643.4−692947) of $F_{888} = 2.69 \pm 0.18$ mJy, S 61 (EMU EC J054551.8−671426) of $F_{888} = 2.31 \pm 0.06$ mJy and S 119 (EMU ES J053125.5−690537) of $F_{888} = 1.09 \pm 0.11$ mJy, all point-like sources.

D.5 X-ray binaries

Simbad lists 66 X-ray binaries within the ASKAP footprint. With the only exception of LMC X-2 (a low-mass X-ray binary), these are high-mass X-ray binary (HMXB) systems formed by a compact object (in most cases a neutron star) accreting matter from a massive companion star (Antoniou & Zezas, 2016). From about 20 systems X-ray pulsations were discovered, which identify the spin period of the neutron star. In many of the others cases the nature of the X-ray source is less clear; low statistical data quality and/or uncertain X-ray position do not allow identification of the optical counterpart.

Four of the 66 (candidate) X-ray binaries were successfully recovered in the 888 MHz catalogue – three in the Gold and one in the Silver source list. They are all well detected, coincident and isolated in the radio.

The radio-brightest of the four at $F_{888} = 9.15 \pm 0.05$ mJy correlates with RX J0457.2−6612, a little-known X-ray source, listed in the ROSAT PSPC X-ray catalogue of the LMC region by Haberl & Pietsch (1999a) and proposed by Kahabka (2002) as X-ray binary. However, there is no star within $20''$ brighter than $V = 18.8$ mag, which most likely excludes an HMXB nature. The X-ray source is probably identified with a QSO listed in Bailer-Jones et al. (2019), which is further supported by the ASKAP radio detection.

The second brightest (EMU ES J050123.5−703329, $F_{888} = 3.25 \pm 0.05$ mJy) coincides with *Einstein* source Cal 9 (Cowley et al., 1984), which was identified with ROSAT source RX J0501.6−7034 by Schmidtke et al. (1999). The authors list as pos-

sible optical counterpart HV 2289 – an early-B type supergiant with an orbital period of $P = 6.94$ d (Bianchi & Pakull, 1984). X-ray positions from Einstein and ROSAT are not precise enough to allow a secure optical identification, but the source was also detected by XMM–Newton (4XMM J050123.3−703333 in the 4XMM-DR9 catalogue; Webb et al., 2020) with a position error of $1\farcs5$ (1σ). This position is $2\farcs4$ from the Gaia DR2 position (Gaia Collaboration et al., 2018) of HV 2289, consistent within the uncertainty. Also the radio position is consistent within errors with the 4XMM source. Negueruela & Coe (2002) give B0 Ve for the spectral type, which suggests a classical Be/X-ray binary instead of a supergiant system. On the other hand, the orbital period of 6.94 d would be too short for a Be/X-ray binary (e.g., Haberl & Sturm, 2016) and favour a supergiant system. The radio detection of this system makes it even more puzzling.

The two faintest radio detections are associated with RX J0535.8−6530 ($F_{888} = 0.57 \pm 0.05$ mJy) and [SG2005] LMC 15 (0.40 ± 0.05 mJy). The former was found to be a variable X-ray source by Haberl & Pietsch (1999b); an optical counterpart was suggested to be a star by Grazian et al. (2002), and it is also detected with Spitzer and WISE. The GLADE catalogue (Dálya et al., 2018) and Bailer-Jones et al. (2019) list a background galaxy with redshift 0.05. The radio detection supports this identification.

The latter, [SG2005] LMC 15 was proposed as an HMXB candidate by Shtykovskiy & Gilfanov (2005), but the optical counterpart is not coincident with the XMM–Newton and Spitzer source and the real counterpart of the X-ray and radio source is likely the QSO MQS J054433.81−682813.7 given in the Million Quasars catalogue (Milliquas V6.3, Flesch, 2015).

The most famous X-ray binaries in the LMC, LMC X-1 (a black hole + supergiant system) and LMC X-4 (a neutron star accreting from a Roche-lobe filling supergiant) were not detected. However, we do detect a radio arc with LMC X-1 in its focus (Figure D.2), peaking at $4''$ due South (1 pc in projection at the distance of the LMC). If this is a bow shock (Hyde et al., 2017) ahead of the supersonic motion of LMC X-1 then this means the binary must have originated from *outside* the prominent star-forming region LHA 120-N 159 into which it would seem to be heading. The system may have run

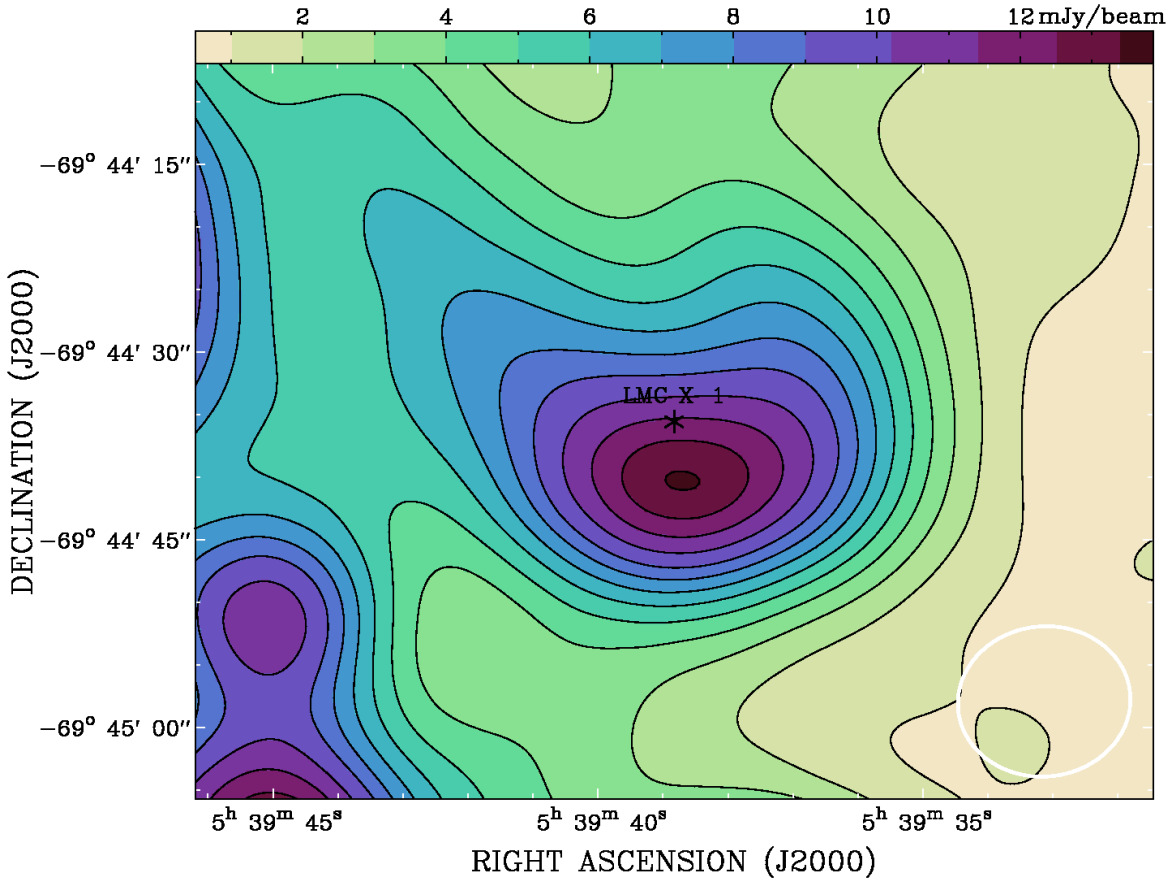

Figure D.2: X-ray binary LMC X-1 seen in the focus of a radio arc in the 888 MHz ASKAP LMC image. The beam size for ASKAP 888 MHz is $\sim 13\farcs9 \times 12\farcs1$ and is represented by the white ellipse in the bottom right.

away from N 159 and subsequently received a lateral kick when the supernova created the black hole.

D.6 Galactic active stars

Galactic stars seen in the direction of the Magellanic Clouds are at high Galactic latitude and radio-loud stars among them are most commonly nearby (∼100 pc) flare stars (M-dwarfs) or interacting tight binaries.

We detect the hierarchical young active system AB Dor (Pakull, 1981; Collier Cameron et al., 1988; Wolter et al., 2014; Azulay et al., 2015, 2017) and resolve the emission coming from the main (K-type) component and the B component (Figure D.3) – the latter of which itself is a spectroscopic M-dwarf binary. White et al. (1988) had ruled out 8-GHz emission coming from component B, but the system was later resolved at 5 GHz by Beasley & Cram (1993) and Lim (1993), with the latter work showing the highly polarized and variable nature of component B. The system has another component, C, which is a late-M dwarf with a sub-stellar companion (Climent et al., 2019) close to the main component but on the side of component B, so we cannot discard the possibility that some of the radio emission originates from it. The ASKAP emission is heavily blended, with a peak and integrated flux density of $F_{888,\mathrm{peak}} = 2.80 \pm 0.05$ mJy and $F_{888,\mathrm{total}} = 3.57 \pm 0.07$ mJy, respectively. Slee et al. (1984) detected it at 5 GHz (5.2 mJy); Vaughan & Large (1986) subsequently detected it at 843 MHz at a similar level (4 mJy) and our measurement confirms the flat spectrum (in F_ν). Collier (1982) characterised its activity as that of an FK Com system, which is associated with rapid rotation and evolution off of the main sequence (Oliveira & Foing, 1999). Rucinski (1983), however, argued for youth and a post-T Tauri status and their detection of lithium seemed to prove that. Innis et al. (1986) (see also Ortega et al., 2007) determined the association of the AB Dor system with the Pleiades moving group (later revisited as the AB Dor moving group – see, e.g., Barenfeld et al., 2013), confirming its relative youth. The activity of AB Dor is usually associated with coronal, magnetic activity (e.g., Donati & Collier Cameron, 1997; Donati et al., 1999; Jardine et al., 2002; Sanz-Forcada et al., 2003). AB Dor is at the same distance of 15 pc as the eruptive pair of M dwarfs CD $-38°11343$, which was detected with ASKAP by Riggi et al. (2021) at essentially the same integrated radio flux densities (at 912 MHz) as AB Dor.

The RS CVn system AE Men is detected, at $F_{888} = 1.51 \pm 0.06$ mJy. At a Gaia distance of $d = 288$ pc this corresponds to a radio luminosity of 2×10^{17} erg s^{-1} Hz^{-1}, placing it among the top 10% integrated radio flux densities of this kind of system (Morris & Mutel, 1988; Güdel, 2002). This system is of K2 III+F/G spectral type and

it is suggested to exhibit non-thermal (gyrosynchrotron) emission. What makes this object unusual compared to the $\sim 100+$ known population of RS CVn type systems is its X-ray (ROSAT), Hα and especially IR (AllWISE) detection. AE Men was observed on several occasions with MOST and ATCA (Slee et al., 1987; Vaughan & Large, 1987; Wendker, 1995) but this is its first radio continuum detection. We reprocessed the ATCA observations from 1994 and confirm that it was not detected.

The β Lyræ-type eclipsing binary BK Dor (= CD $-67°435$) is also detected for the first time at radio frequencies, with $F_{888} = 0.90 \pm 0.07$ mJy. With the Gaia distance of 279 pc this yields a radio luminosity of 8×10^{16} erg s^{-1} Hz^{-1}, not as high as AE Men but still luminous. It was discovered as a variable X-ray source by Fuhrmeister & Schmitt (2003) before being associated with the G7 III star by Torres et al. (2006) and found to be a short-period ($P = 2.2$ d) eclipsing binary by Szczygieł et al. (2008). In such a system the two stars are highly distorted and exchange mass.

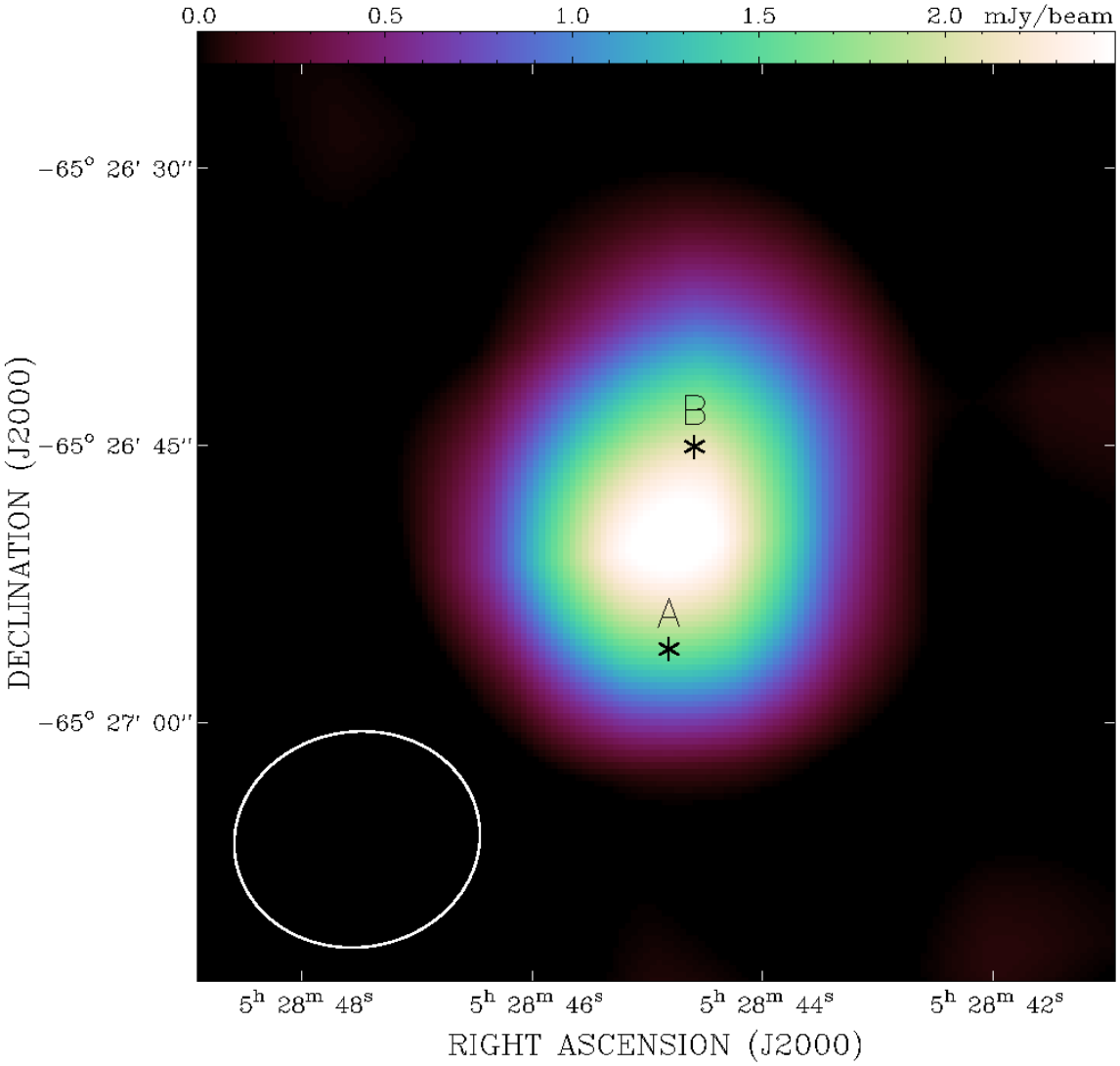

Figure D.3: LMC AB Dor, a hierarchical young active system detected in the 888 MHz ASKAP LMC image where emission coming from the main component and the B component contribute. ASKAP 888 MHz beam size is $\sim 13\farcs9 \times 12\farcs1$, represented by the white ellipse in the bottom left.

E Supplementary data for search for more AGN like SAGE0536AGN

E.1 Photometry

Here I show all the photometry used for the t-SNE selected sample shown in Figure E.1.

E.2 Additional spectra

Here I show the additional spectra of SAGE0534AGN that were not shown in the main text. SAGE0534AGN observation by SALT is seen in Figure E.1 and the observation by SAAO 1.9m telescope is seen in Figure E.2. All of the spectra are provided as online supplementary material. Note that those observed with SALT have not been flux calibrated and those that have been observed with SAAO 1.9m telescope have been flux calibrated.

E.3 CIGALE best fits

Here I show the CIGALE best fit models for the t-SNE selected sample. The sources with no X-ray observations are shown in Figure E.3 and E.4, and those with X-ray observations are shown in Figure E.5. Source 8 was fit without X-ray due to the best fit having $\chi^2 > 10$, but without including the X-ray the best fit has $\chi^2 = 2.9$. This fit was done before it was known Source 8 was a star.

Table E.1: Table format of the table of photometry of all sources in t-SNE selected sample. The table of photometry is available as online supplementary material. Note that if there is no value for the flux density error, this means that the corresponding flux density is an upper limit calculated from the survey image.

Column(s)	Format	Description
1–11	A11	ID. Source name.
13–20	E8.2	XMM-Newton 0.2 to 10 keV flux density (mJy).
22–29	E8.2	XMM-Newton 0.2 to 10 keV flux density error (mJy).
31–38	E8.2	XMM-Newton 0.2 to 12 keV flux density (mJy).
40–47	E8.2	XMM-Newton 0.2 to 12 keV flux density error (mJy).
49–54	F6.4	SMASH u band flux density (mJy).
56–61	F6.4	SMASH u band flux density error (mJy).
63–68	F6.4	SMASH g band flux density (mJy).
70–75	F6.4	SMASH g band flux density error (mJy).
77–82	F6.4	SMASH r band flux density (mJy).
84–89	F6.4	SMASH r band flux density error (mJy).
91–96	F6.4	SMASH i band flux density (mJy).
98–103	F6.4	SMASH i band flux density error (mJy).
105–110	F6.4	SMASH z band flux density (mJy).
112-117	F6.4	SMASH z band flux density error (mJy).
119–123	F5.3	Gaia bp band flux density (mJy). From Gaia EDR3.
125–129	F5.3	Gaia bp band flux density error (mJy). From Gaia EDR3.
131–135	F5.3	Gaia g band flux density (mJy). From Gaia EDR3.
137–141	F5.3	Gaia g band flux density error (mJy). From Gaia EDR3.
143–147	F5.3	Gaia rp band flux density (mJy). From Gaia EDR3.
149–153	F5.3	Gaia rp band flux density error (mJy). From Gaia EDR3.
155–159	F5.3	VISTA Y band flux density (mJy). From VMC survey.
161–165	F5.3	VISTA Y band flux density error (mJy). From VVMC survey.
167–171	F5.3	VISTA J band flux density (mJy). From VMC survey.
173–177	F5.3	VISTA J band flux density error (mJy). From VMC survey.
179–183	F5.3	VISTA Ks band flux density (mJy). From VMC survey.
185–189	F5.3	VISTA Ks band flux density error (mJy). From VMC survey.
191–196	F6.2	Spitzer IRAC 1 band flux density (mJy). From SAGE survey.
198–201	F4.2	Spitzer IRAC 1 band flux density error (mJy). From SAGE survey.
203–208	F6.2	Spitzer IRAC 2 band flux density (mJy). From SAGE survey.
210–213	F4.2	Spitzer IRAC 2 band flux density error (mJy). From SAGE survey.
215–220	F6.2	Spitzer IRAC 3 band flux density (mJy). From SAGE survey.
222–225	F4.2	Spitzer IRAC 3 band flux density error (mJy). From SAGE survey.
227–231	F5.2	Spitzer IRAC 4 band flux density (mJy). From SAGE survey.
233–236	F4.2	Spitzer IRAC 4 band flux density error (mJy). From SAGE survey.
238–243	F6.2	WISE 1 band flux density (mJy). From AllWISE survey.
245–248	F4.2	WISE 1 band flux density error (mJy). From AllWISE survey.
250–255	F6.2	WISE 2 band flux density (mJy). From AllWISE survey.
257–260	F4.2	WISE 2 band flux density error (mJy). From AllWISE survey.
262–267	F6.2	WISE 3 band flux density (mJy). From AllWISE survey.
269–272	F4.2	WISE 3 band flux density error (mJy). From AllWISE survey.

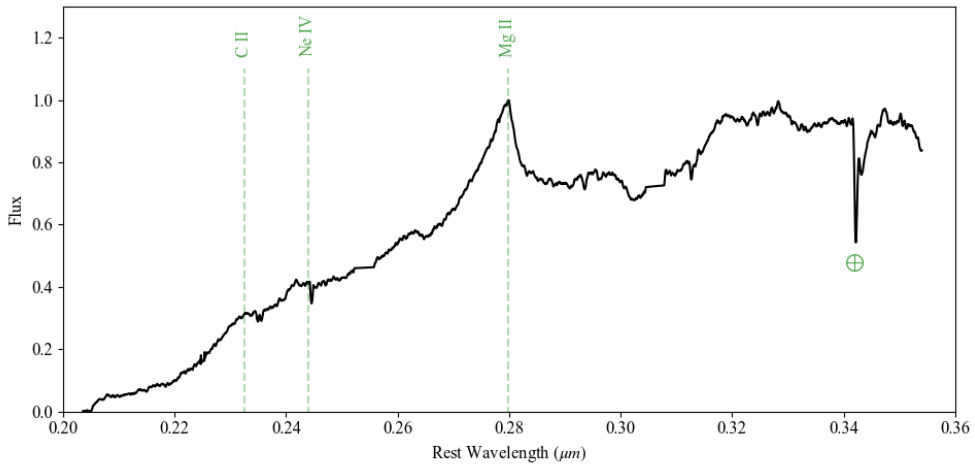

Figure E.1: Spectrum of SAGE0534AGN observed with SALT under program 2021-2-SCI-017 (PI: Joy Anih). We used the RSS, a combination of three CCD detectors with total 3172 $\times$ 2052 pixels and spatial resolution of 0.126700" per pixel. We used the long-slit with width 1".5, grating PG0900, filter PC03850 and an Argon arc lamp. The data was processed using the standard IRAF tools. This spectrum is not flux calibrated.

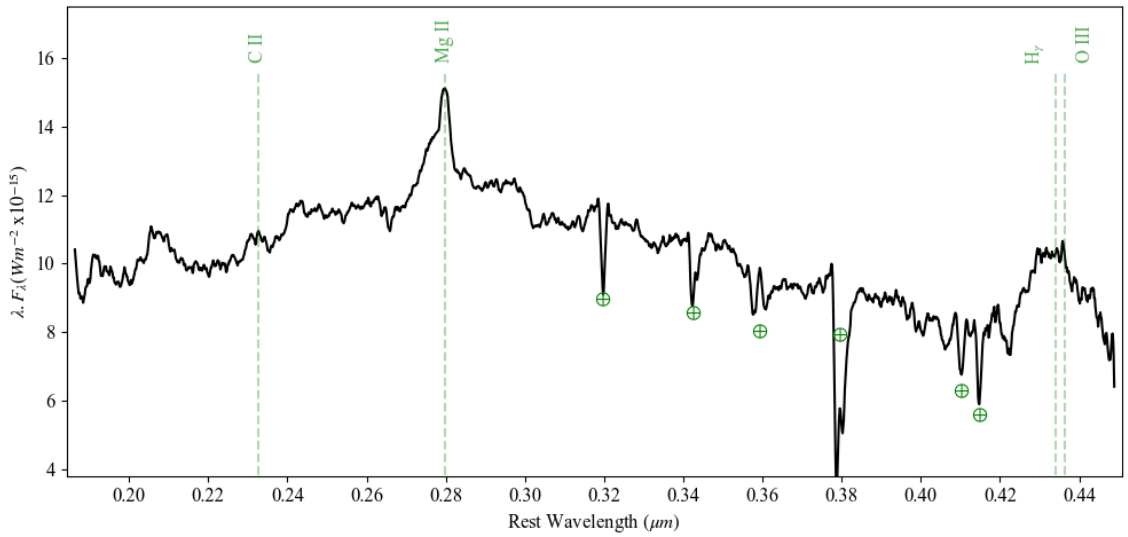

Figure E.2: Spectrum of SAGE0534AGN observed with SAAO 1.9m telescope with SpUpNIC. Grating 7 (grating angle of 16°) and the order blocking 'BG38' filter were used, delivering a resolving power R = $\frac{\lambda}{\Delta\lambda} \sim 500$ over a wavelength range of 3800Å — 9000 Å. The CuAr lamp was used for wavelength calibration. Three 600s exposures were obtained. The standard star EG 21 (Hamuy et al. 1994) was observed on the same night under the same conditions for 30s. The data was processed using the standard IRAF tools.

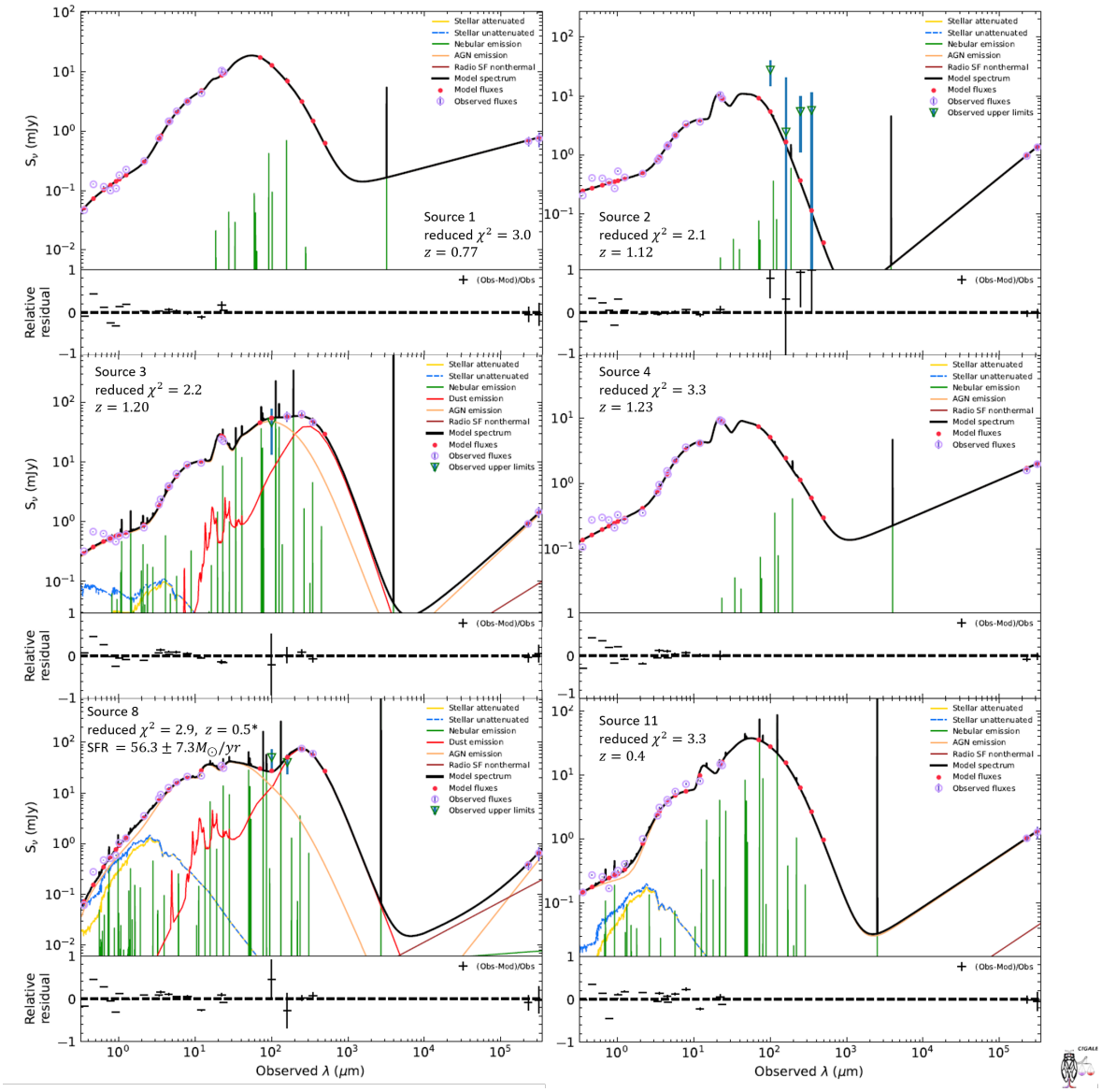

Figure E.3: CIGALE best fits of Sources without X-ray observations.

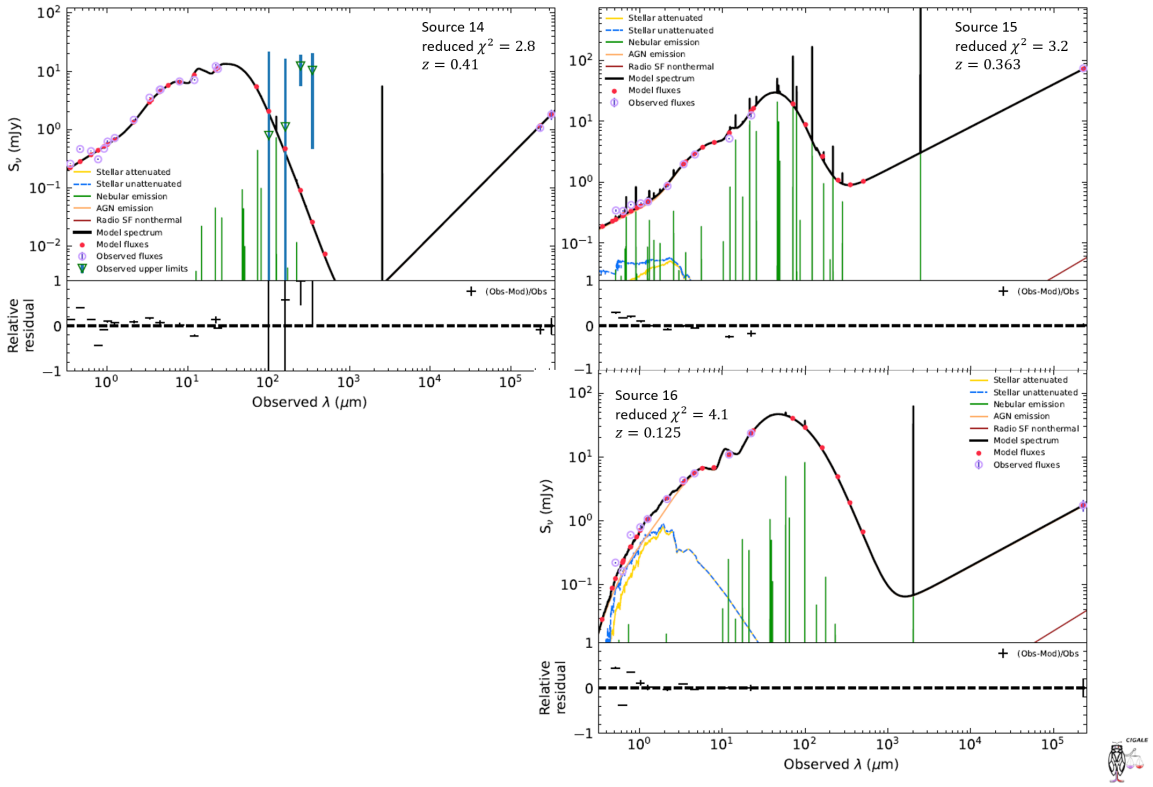

Figure E.4: (cont.) CIGALE best fits of Sources without X-ray observations.

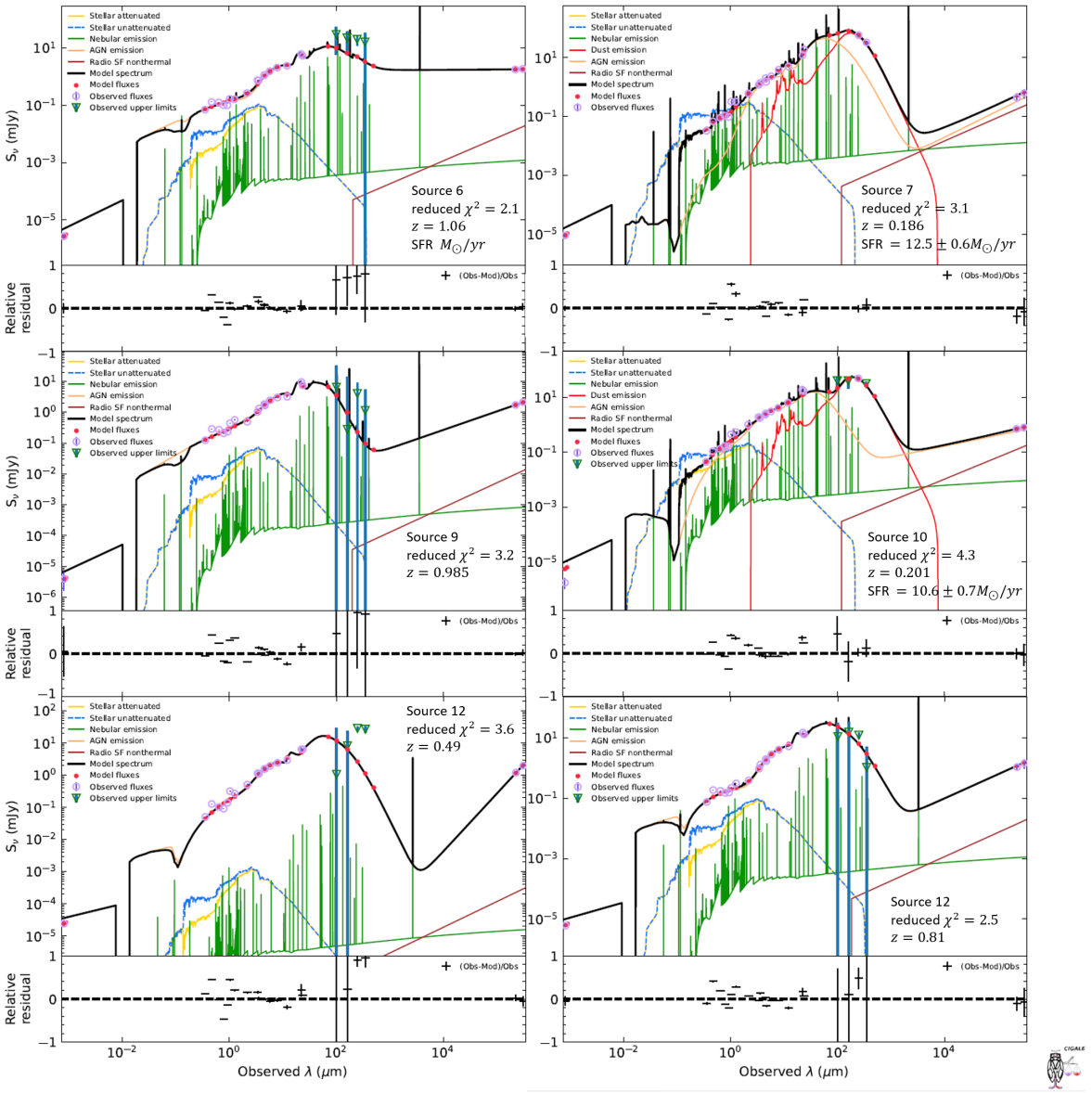

Figure E.5: CIGALE best fits for sources with X-ray observations.

Bibliography

Abbott T. M. C., et al., 2018, ApJS, 239, 18

Agliozzo C., Umana G., Trigilio C., Buemi C., Leto P., Ingallinera A., Franzen T., Noriega-Crespo A., 2012, MNRAS, 426, 181

Agliozzo C., et al., 2017a, MNRAS, 466, 213

Agliozzo C., et al., 2017b, ApJ, 841, 130

Agliozzo C., et al., 2019, A&A, 626, A126

Akras S., Guzman-Ramirez L., Leal-Ferreira M. L., Ramos-Larios G., 2019, ApJS, 240, 21

Anders F., Chiappini C., Santiago B. X., Matijevič G., Queiroz A. B., Steinmetz M., Guiglion G., 2018, A&A, 619, A125

Angles-Alcázar D., 2015, ApJ, 770

Antoniou V., Zezas A., 2016, MNRAS, 459, 528

Antonucci R. R. J., 1982, Nature, 299, 605

Antonucci R. R. J., 1984, ApJ, 278, 499

Antonucci R., 1993, ARA&A, 31, 473

Asmus D., Hönig S. F., Gandhi P., 2016, ApJ, 822, 109

Assef R., Kochanek C., Collaboration T. A., 2010, ApJ, 713

Assef R., Denney K., Kochanek C., Peterson B., Kozłowski S., 2011, ApJ, 742

Assef R. J., et al., 2013, ApJ, 772, 26

Astropy Collaboration et al., 2013, A&A, 558, A33

Astropy Collaboration et al., 2018, AJ, 156, 123

Azulay R., et al., 2015, A&A, 578, A16

Azulay R., et al., 2017, A&A, 607, A10

Baars J. W. M., Genzel R., Pauliny-Toth I. I. K., Witzel A., 1977, A&A, 500, 135

Bailer-Jones C. A. L., Fouesneau M., Andrae R., 2019, MNRAS, 490, 5615

Baldi R. D., et al., 2010, ApJ, 725, 2426

Barenfeld S. A., Bubar E. J., Mamajek E. E., Young P. A., 2013, ApJ, 766, 6

Barnbaum C., Stone R. P. S., Keenan P. C., 1996, ApJS, 105, 419

Baron D., Ménard B., 2019, MNRAS, 487, 3404

Baron D., Stern J., Poznanski D., Netzer H., 2016, ApJ, 832, 8

Bañados E., Venemans B. P., Decarli R., Farina E. P., Mazzucchelli C., 2016, ApJS, 227

Beasley A. J., Cram L. E., 1993, MNRAS, 264, 570

Belfiore F., et al., 2018, MNRAS, 477, 3014

Bell E. F., 2003, ApJ, 586, 794

Bentz M., Peterson B., Netzer H., 2009, ApJ, 697

Bianchi L., Pakull M., 1984, in Mead J. M., Chapman R. D., Kondo Y., eds, NASA Conference Publication Vol. 2349, NASA Conference Publication. pp 416–419

Bilicki M., Jarrett T. H., Peacock J. A., Cluver M. E., Steward L., 2014, ApJS, 210, 9

Bilicki M., et al., 2016, ApJS, 225, 5

Bilicki M., et al., 2018, A&A, 616, A69

Blandford R., Meier D., Readhead A., 2019, ARA&A, 57, 467

Boquien M., Burgarella D., Roehlly Y., Buat V., Ciesla L., Corre D., Inoue A. K., Salas H., 2019, A&A, 622, A103

Bornancini C. G., Oio G. A., Alonso M. V., García Lambas D., 2022, arXiv e-prints, p. arXiv:2204.05219

Boroson T., 2011, ApJL, 735

Boyer M. L., et al., 2011, AJ, 142, 103

Breiman L., 2001, Machine Learning, 45

Brescia M., Cavuoti S., D'Abrusco R., Longo G., Mercurio A., 2013, ApJ, 772, 140

Brescia M., Cavuoti S., Longo G., De Stefano V., 2014, A&A, 568, A126

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Buat V., et al., 2021, A&A, 654, A93

Buckley D. A. H., Swart G. P., Meiring J. G., 2006, in Stepp L. M., ed., Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 6267, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. p. 62670Z, doi:10.1117/12.673750

Burgess A. M., Hunstead R. W., 2006, AJ, 131, 114

Burgh E. B., Nordsieck K. H., Kobulnicky H. A., Williams T. B., O'Donoghue D., Smith M. P., Percival J. W., 2003, in Iye M., Moorwood A. F. M., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. pp 1463–1471, doi:10.1117/12.460312

Callingham J. R., et al., 2016, MNRAS, 462, 290

Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682

Carliles S., Budavári T., Heinis S., Priebe C., Szalay A. S., 2010, ApJ, 712, 511

Casey C. M., 2012, MNRAS, 425, 3094

Chabrier G., 2003, PASP, 115, 763

Chen C. H. R., Chu Y.-H., Gruendl R. A., Gordon K. D., Heitsch F., 2009, ApJ, 695, 511

Chen Y.-M., et al., 2016, Nature Communications, 7, 13269

Cheng T.-Y., et al., 2020, MNRAS, 493, 4209

Choudhury S., et al., 2021, MNRAS, 507, 4752

Cioni M. R. L., et al., 2011, A&A, 527, A116

Cioni M. R. L., et al., 2013, A&A, 549, A29

Cioni M., Girardi L., Moretti M., Piffl T., Ripepi V., Rubele S., 2014, å, 562

Clemens M. S., Scaife A., Vega O., Bressan A., 2010, MNRAS, 405, 887

Climent J. B., Berger J. P., Guirado J. C., Marcaide J. M., Martí-Vidal I., Mérand A., Tognelli E., Wittkowski M., 2019, AJ, 886, L9

Cole A. A., Tolstoy E., Gallagher John S. I., Smecker-Hane T. A., 2005, AJ, 129, 1465

Collier A. C., 1982, MNRAS, 200, 489

Collier Cameron A., Bedford D. K., Rucinski S. M., Vilhu O., White N. E., 1988, MNRAS, 231, 131

Cowley A. P., Crampton D., Hutchings J. B., Helfand D. J., Hamilton T. T., Thorstensen J. R., Charles P. A., 1984, ApJ, 286, 196

Cowley M. J., et al., 2016, MNRAS, 457, 629

Cowley M. J., et al., 2018, MNRAS, 473, 3710

Coziol R., Andernach H., Caretta C. A., Alamo-Martínez K. A., Tago E., 2009, AJ, 137, 4795

Crause L. A., et al., 2019, Journal of Astronomical Telescopes, Instruments, and Systems, 5, 024007

Crawford S. M., et al., 2010, in Silva D. R., Peck A. B., Soifer B. T., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7737, Observatory Operations: Strategies, Processes, and Systems III. p. 773725, doi:10.1117/12.857000

Croston J. H., Ineson J., Hardcastle M. J., 2018, MNRAS, 476, 1614

Cusano F., et al., 2021, MNRAS, 504, 1

Cutri R. M., et al., 2013, Explanatory Supplement to the AllWISE Data Release Products, Explanatory Supplement to the AllWISE Data Release Products

Cutri R. M., et al., 2021, VizieR Online Data Catalog, p. II/328

Czerny B., Li J., Loska Z., Szczerba R., 2004, MNRAS, 348, L54

Dale D. A., Helou G., Magdis G. E., Armus L., Díaz-Santos T., Shi Y., 2014, ApJ, 784, 83

Dalton G. B., et al., 2006, in McLean I. S., Iye M., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 6269, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. p. 62690X, doi:10.1117/12.670018

Dálya G., et al., 2018, MNRAS, 479, 2374

De Bortoli B. J., Parisi M. C., Bassino L. P., Geisler D., Dias B., Gimeno G., Angelo M. S., Mauro F., 2022, A&A, 664, A168

DeBoer D. R., et al., 2009, IEEE Proceedings, 97, 1507

Dickel J. R., McIntyre V. J., Gruendl R. A., Milne D. K., 2005, AJ, 129, 790

Dicken D., et al., 2014, ApJ, 788, 98

Dickey J. M., et al., 2013, PASA, 30, e003

Donati J. F., Collier Cameron A., 1997, MNRAS, 291, 1

Donati J. F., Collier Cameron A., Hussain G. A. J., Semel M., 1999, MNRAS, 302, 437

Donley J. L., et al., 2012, ApJ, 748, 142

Dorigo Jones J., Oey M. S., Paggeot K., Castro N., Moe M., 2020, ApJ, 903, 43

Edge D. O., Shakeshaft J. R., McAdam W. B., Baldwin J. E., Archer S., 1959, Mem. R. astr. Soc., 68

Ehman J. R., Dixon R. S., Kraus J. D., 1970, Astronomical Journal, 75

Ellison S. L., Teimoorinia H., Rosario D. J., Mendel J. T., 2016, MNRAS, 455, 370

Esquej P., et al., 2013, A&A, 557, A123

Evans C. J., et al., 2015a, A&A, 574, A13

Evans C. J., van Loon J. T., Hainich R., Bailey M., 2015b, A&A, 584, A5

Fabian A., 2011, ARA&A, 50

Fanaroff B. L., Riley J. M., 1974, MNRAS, 167, 31P

Fanti C., Fanti R., Dallacasa D., Schilizzi R. T., Spencer R. E., Stanghellini C., 1995, A&A, 302, 317

Feltre A., Charlot S., Gutkin J., 2016, MNRAS, 456

Ferrarese L., Merritt D., 2000, ApJL, 539, L9

Filipovic M. D., Haynes R. F., White G. L., Jones P. A., Klein U., Wielebinski R., 1995, A&AS, 111, 311

Filipovic M. D., White G. L., Haynes R. F., Jones P. A., Meinert D., Wielebinski R., Klein U., 1996, A&AS, 120, 77

Filipovic M. D., Jones P. A., White G. L., Haynes R. F., 1998a, PASA, 15, 128

Filipovic M. D., et al., 1998b, A&AS, 127, 119

Filipovic M. D., Haynes R. F., White G. L., Jones P. A., 1998c, A&AS, 130, 421

Filipovic M. D., Jones P. A., White G. L., Haynes R. F., 1998d, A&AS, 130, 441

Filipović M. D., et al., 2009, MNRAS, 399, 769

Filipović M. D., et al., 2021, MNRAS, 507, 2885

Finkbeiner D. P., et al., 2004, AJ, 128, 2577

Flesch E. W., 2015, PASA, 32, e010

Flesch E. W., 2019a, VizieR Online Data Catalog, p. VII/283

Flesch E. W., 2019b, arXiv e-prints, p. arXiv:1912.05614

Flesch E. W., 2021, VizieR Online Data Catalog, p. VII/290

For B. Q., et al., 2018, MNRAS, 480, 2743

Fritz J., Franceschini A., Hatziminaoglou E., 2006, MNRAS, 366, 767

Fuhrmeister B., Schmitt J. H. M. M., 2003, Astronomische Nachrichten Supplement, 324, 32

Gaia Collaboration et al., 2016, A&A, 595, A1

Gaia Collaboration et al., 2018, A&A, 616, A1

Gaia Collaboration et al., 2021, A&A, 649, A1

Gaia Collaboration et al., 2022, arXiv e-prints, p. arXiv:2208.00211

Gandhi P., Hoenig S. F., 2015, in TORUS2015: The AGN Unification Scheme After 30 Years.

Gaskell C. M., Benker A. J., 2007, arXiv e-prints, p. arXiv:0711.1013

Gaskell C. M., Goosmann R. W., Antonucci R. R. J., Whysong D. H., 2004, ApJ, 616, 147

Gebhardt K., et al., 2000, ApJL, 539, L13

Geha M., et al., 2003, AJ, 125, 1

Gordon K. D., Meixner M., Meade M., Whitney B. A., Engelbracht C. W., Bot C., 2011, AJ, 142

Graham A. W., 2008, PASA, 25, 167–175

Grazian A., Omizzolo A., Corbally C., Cristiani S., Haehnelt M. G., Vanzella E., 2002, AJ, 124, 2955

Griffin M. J., et al., 2010, A&A, 518, L3

Griffith M. R., Wright A. E., 1993, AJ, 105, 1666

Grin N. J., et al., 2017, A&A, 600, A82

Groenewegen M. A. T., Blommaert J. A. D. L., 1998, A&A, 332, 25

Groenewegen M. A. T., et al., 2019, A&A, 622, A63

Groenewegen M. A. T., et al., 2020, A&A, 636, A48

Güdel M., 2002, ARA&A, 40, 217

Gullieuszik M., et al., 2012, A&A, 537, A105

Haberl F., Pietsch W., 1999a, A&AS, 139, 277

Haberl F., Pietsch W., 1999b, A&A, 344, 521

Haberl F., Sturm R., 2016, A&A, 586, A81

Hale C. L., Robotham A. S. G., Davies L. J. M., Jarvis M. J., Driver S. P., Heywood I., 2019, MNRAS, 487, 3971

Hamuy M., Suntzeff N. B., Heathcote S. R., Walker A. R., Gigoux P., Phillips M. M., 1994, PASP, 106, 566

Hancock P. J., Murphy T., Gaensler B. M., Hopkins A., Curran J. R., 2012, MNRAS, 422, 1812

Hancock P. J., Trott C. M., Hurley-Walker N., 2018, PASA, 35, e011

Hao L., Weedman D. W., Spoon H. W. W., Marshall J. A., Levenson N. A., Elitzur M., Houck J. R., 2007, ApJL, 655, L77

Häring N., Rix H.-W., 2004, ApJL, 604, L89

Hatziminaoglou E., Omont A., Stevens J. A., Amblard A., Arumugam V., Auld R., Aussel H., Babbedge T., 2010, å, 518

Heydari-Malayeri M., Melnick J., 1992, A&A, 258, L13

Hine R. G., Longair M. S., 1979, MNRAS, 188

Ho L. C., 2008, ARA&A, 46

Hönig S. F., 2019, ApJ, 884, 171

Hönig S. F., Kishimoto M., Antonucci R., Marconi A., Prieto M. A., Tristram K., Weigelt G., 2012, ApJ, 755, 149

Hönig S. F., et al., 2013, ApJ, 771, 87

Hony S., et al., 2011, A&A, 531, A137

Hora J. L., et al., 2008, PASP, 120, 1233

Hotan A. W., et al., 2014, PASA, 31, e041

Hotan A. W., et al., 2021, PASA, 38, e009

Houck J. R., et al., 2004, ApJS, 154, 18

Huchra J. P., et al., 2012, ApJS, 199, 26

Huertas-Company M., Rouan D., Tasca L., Soucail G., Le Fèvre O., 2008, A&A, 478, 971

Hughes A., Staveley-Smith L., Kim S., Wolleben M., Filipović M., 2007, MNRAS, 382, 543

Hyde E. A., Russell D. M., Ritter A., Filipović M. D., Kaper L., Grieve K., O'Brien A. N., 2017, PASP, 129, 094201

Innis J. L., Thompson K., Coates D. W., 1986, MNRAS, 223, 183

Ivanov V. D., et al., 2016, A&A, 588, A93

James G., Witten D., Hastie T., Tibshirani R., 2017, An Introduction to Statistical Learning: with Applications in R. New York: Springer

Jansen F., et al., 2001, A&A, 365, L1

Jardine M., Collier Cameron A., Donati J. F., 2002, MNRAS, 333, 339

Jarrett T. H., et al., 2011, ApJ, 735, 112

Jarvis M. J., et al., 2013, MNRAS, 428, 1281

Johnston S., et al., 2008, Experimental Astronomy, 22, 151

Jonas J. L., 2009, IEEE Proceedings, 97, 1522

Jones H., Saunders W., Colless M., Read M., Parker Q., Watson F., Campbell L., 2005, in Fairall A. P., Woudt P. A., eds, Astronomical Society of the Pacific Conference Series Vol. 329, Nearby Large-Scale Structures and the Zone of Avoidance. p. 11

Jones D. H., et al., 2009, MNRAS, 399, 683

Jones O. C., et al., 2017, MNRAS, 470, 3250

Joseph T. D., et al., 2019, MNRAS, 490, 1202

Joye W. A., Mandel E., 2003, in Payne H. E., Jedrzejewski R. I., Hook R. N., eds, Astronomical Society of the Pacific Conference Series Vol. 295, Astronomical Data Analysis Software and Systems XII. p. 489

Kahabka P., 2002, A&A, 388, 100

Kamath D., Wood P. R., Van Winckel H., 2014, MNRAS, 439, 2211

Kemper F., et al., 2010, PASP, 122, 683

Kim S., Staveley-Smith L., Dopita M. A., Freeman K. C., Sault R. J., Kesteven M. J., McConnell D., 1998, ApJ, 503, 674

Kim S., Staveley-Smith L., Dopita M. A., Sault R. J., Freeman K. C., Lee Y., Chu Y.-H., 2003, ApJS, 148, 473

Kim D.-W., Protopapas P., Byun Y.-I., Alcock C., Khardon R., Trichas M., 2011, ApJ, 735, 68

King A., Pounds K., 2015, ARA&A, 53

Kishimoto M., Antonucci R., Blaes O., Lawrence A., Boisson C., Albrecht M., Leipski C., 2008, Nature, 454, 492

Kobulnicky H. A., Nordsieck K. H., Burgh E. B., Smith M. P., Percival J. W., Williams T. B., O'Donoghue D., 2003, in Iye M., Moorwood A. F. M., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. pp 1634–1644, doi:10.1117/12.460315

Koenig X., Leisawitz D., Benford D., Rebull L., Padgett D., Assef R., 2012, ApJ, 744

Kormendy J., Ho L. C., 2013, ARA&A, 51, 511

Kovačević M., Chiaro G., Cutini S., Tosti G., 2020, MNRAS, 493, 1926

Kozłowski S., Kochanek C. S., Udalski A., 2011, ApJS, 194

Kozłowski S., et al., 2012, ApJ, 746, 27

Kozłowski S., et al., 2013, ApJ, 775, 92

Krogager J. K., et al., 2018, ApJS, 235, 10

Lacy M., et al., 2004, ApJS, 154, 166

Lakićević M., van Loon J. T., Patat F., Staveley-Smith L., Zanardo G., 2011, A&A, 532, L8

Lamb J. B., Oey M. S., Segura-Cox D. M., Graus A. S., Kiminki D. C., Golden-Marx J. B., Parker J. W., 2016, ApJ, 817, 113

Large M. I., Mills B. Y., Little A. G., Crawford D. F., Sutton J. M., 1981, MNRAS, 194, 693

Lawrence A., et al., 2007, MNRAS, 379, 1599

Leftley J. H., Tristram K. R. W., Hönig S. F., Kishimoto M., Asmus D., Gandhi P., 2018, ApJ, 862, 17

Leighly K. M., Cooper E., Grupe D., Terndrup D. M., Komossa S., 2015, ApJL, 809, L13

Leverenz H., Filipović M. D., Vukotić B., Urošević D., Grieve K., 2017, MNRAS, 468, 1794

Lim J., 1993, ApJL, 405, L33

Lochner M., McEwen J. D., Peiris H. V., Lahav O., Winter M. K., 2016, ApJS, 225, 31

López-Gonzaga N., Burtscher L., Tristram K. R. W., Meisenheimer K., Schartmann M., 2016, A&A, 591, A47

Lynden-Bell D., 1969, Nature, 223

Magorrian J., et al., 1998, AJ, 115, 2285

Maitra C., Haberl F., Ivanov V. D., 2018, in 42nd COSPAR Scientific Assembly. pp E1.12–27–18

Małek K., et al., 2013, A&A, 557, A16

Markowitz A. G., Krumpe M., Nikutta R., 2014, MNRAS, 439, 1403

Marocco F., et al., 2021, ApJS, 253, 8

Martin D. C., et al., 2005, ApJL, 619, L1

Martínez-Paredes M., et al., 2020, ApJ, 890, 152

Martocchia A., Karas V., Matt G., 2000, MNRAS, 312, 817

Masci F. J., Hoffman D. I., Grillmair C. J., Cutri R. M., 2014, AJ, 148

Mateos S., et al., 2012, MNRAS, 426, 3271

Mauch T., Murphy T., Buttery H. J., Curran J., Hunstead R. W., Piestrzynski B., Robertson J. G., Sadler E. M., 2003a, VizieR Online Data Catalog, p. VIII/70A

Mauch T., Murphy T., Buttery H. J., Curran J., Hunstead R. W., Piestrzynski B., Robertson J. G., Sadler E. M., 2003b, MNRAS, 342, 1117

McConnell N., Ma C.-P., 2013, ApJ, 764

McConnell D., Sadler E. M., Murphy T., Ekers R. D., 2012, MNRAS, 422, 1527

McConnell D., et al., 2016, PASA, 33, e042

McConnell D., et al., 2020, PASA, 37, e048

McMahon R. G., Banerji M., Gonzalez E., Koposov S. E., Bejar V. J., Lodieu N., Rebolo R., VHS Collaboration 2013, The Messenger, 154, 35

McNamara B., Nulsen P., 2007, ARA&A, 45

Mehner A., et al., 2017, A&A, 608, A124

Meixner M., et al., 2006, AJ, 132, 2268

Meixner M., et al., 2010, A&A, 518, L71

Meixner M., et al., 2013, AJ, 146, 62

Meyssonnier N., Azzopardi M., 1993, A&AS, 102, 451

Mills B. Y., 1985, Proceedings of the Astronomical Society of Australia, 6, 72

Morgan D. H., Watson F. G., Parker Q. A., 1992, A&AS, 93, 495

Morris D. H., Mutel R. L., 1988, AJ, 95, 204

Mullaney J. R., Alexander D. M., Goulding A. D., Hickox R. C., 2011, MNRAS, 414, 1082

Murphy T., et al., 2010, MNRAS, 402, 2403

Möller A., et al., 2016, Journal of Cosmology and Astroparticle Physics, 2016

Nakoneczny S., Bilicki M., Solarz A., Pollo A., Maddox N., Spiniello C., Brescia M., Napolitano N. R., 2019, å, 624

Negueruela I., Coe M. J., 2002, A&A, 385, 517

Neugent K. F., Levesque E. M., Massey P., Morrell N. I., Drout M. R., 2020, ApJ, 900, 118

Ng C. Y., Zanardo G., Potter T. M., Staveley-Smith L., Gaensler B. M., Manchester R. N., Tzioumis A. K., 2013, ApJ, 777, 131

Nidever D. L., et al., 2017, AJ, 154, 199

Nikutta R., Elitzur M., Lacy M., 2009, ApJ, 707, 1550

Nikutta R., Hunt-Walker N., Nenkova M., Ivezić v., Elitzur M., 2014, MNRAS, 442, 3361

Noll S., Burgarella D., Giovannoli E., Buat V., Marcillac D., Muñoz-Mateos J. C., 2009, A&A, 507, 1793

Norris R. P., Hopkins A. M., Afonso J., Brown S., Condon J. J., 2011, PASA, 28

Oliveira J. M., Foing B. H., 1999, A&A, 343, 213

Oliveira J. M., et al., 2011, MNRAS, 411, L36

Oliveira J. M., et al., 2013, MNRAS, 428, 3001

Oliveira J. M., et al., 2019, MNRAS, 490, 3909

Ortega V. G., Jilinski E., de La Reza R., Bazzanella B., 2007, MNRAS, 377, 441

Padovani P., Alexander D., Assef R., et al. 2017, ARA&A, 25

Pakull M. W., 1981, A&A, 104, 33

Parisi M. C., Grocholski A. J., Geisler D., Sarajedini A., Clariá J. J., 2009, AJ, 138, 517

Parisi M. C., Geisler D., Grocholski A. J., Clariá J. J., Sarajedini A., 2010, AJ, 139, 1168

Parisi M. C., Gramajo L. V., Geisler D., Dias B., Clariá J. J., Da Costa G., Grebel E. K., 2022, A&A, 662, A75

Peacock J. A., Gull S. F., 1981, MNRAS, 196

Peacock J. A., Miller L., Longair M. S., 1986, MNRAS, 218, 265

Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825

Pei Y. C., 1992, ApJ, 395, 130

Peng C., 2007, ApJ, 671

Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2002, AJ, 124, 266

Peng C. Y., Ho L. C., Impey C. D., Rix H.-W., 2010, AJ, 139, 2097

Pennock C. M., et al., 2021, MNRAS, 506, 3540

Pennock C. M., et al., 2022, MNRAS, 515, 6046

Peterson B., 1993, PASP, 105

Peterson B. M., 1997, An Introduction to Active Galactic Nuclei. Cambridge, New York Cambridge University Press

Pichara K., Protopapas P., Kim D. W., Marquette J. B., Tisserand P., 2012, MNRAS, 427, 1284

Pilbratt G. L., et al., 2010, A&A, 518, L1

Pingel N. M., et al., 2022, PASA, 39, e005

Poglitsch A., et al., 2010, A&A, 518, L2

Ponti G., Morris M. R., Terrier R., Goldwurm A., 2013, in Torres D. F., Reimer O., eds, Astrophysics and Space Science Proceedings Vol. 34, Cosmic Rays in Star-Forming Environments. p. 331 (arXiv:1210.3034), doi:10.1007/978-3-642-35410-6_26

Pounds K., Reeves J., Kings A., Page K., 2003, MNRAS, 345

Pović M., et al., 2012, A&A, 541, A118

Pović M., Sánchez-Portal M., Pérez García A. M., Otelo Group 2013, in Sun W. H., Xu C. K., Scoville N. Z., Sanders D. B., eds, Astronomical Society of the Pacific Conference Series Vol. 477, Galaxy Mergers in an Evolving Universe. p. 177

Prevot M. L., Lequeux J., Maurice E., Prevot L., Rocca-Volmerange B., 1984, A&A, 132, 389

Qu M., Shih F., Jing J., Wang H. M., 2003, in AAS/Solar Physics Division Meeting #34. p. 02.06

Raban D., Jaffe W., Röttgering H., Meisenheimer K., Tristram K., 2009, MNRAS, 394

Rees M., 1966, Nature, 211

Reid W. A., Parker Q. A., 2012, MNRAS, 425, 355

Reines A. E., Volonteri M., 2015, ApJ, 813, 82

Reis I., Baron D., Shahaf S., 2018a, AJ, 157

Reis I., Poznanski D., Baron D., Zasowski G., Shahaf S., 2018b, MNRAS, 476, 2117

Reynolds C. S., 2021, ARA&A, 59, 117

Rieke G. H., et al., 2004, ApJS, 154, 25

Riggi S., et al., 2021, MNRAS, 502, 60

Ripepi V., et al., 2015, MNRAS, 446, 3034

Risalti G., Harrison F., Madsen K., Walton D., 2013, Nature, 494

Roman-Duval J., et al., 2019, ApJ, 871, 151

Rosario D. J., et al., 2015, A&A, 573, A85

Roseboom I. G., Lawrence A., Elvis M., Petty S., Shen Y., Hao H., 2013, MNRAS, 429, 1494

Rubele S., et al., 2015, MNRAS, 449, 639

Rucinski S. M., 1983, A&AS, 52, 281

Ruffle P. M. E., et al., 2015, MNRAS, 451, 3504

Salim S., 2014, Serbian Astronomical Journal, 189, 1

Salpeter E., 1964, ApJ, 140

Sanders D. B., Soifer B. T., Elias J. H., Madore B. F., Matthews K., Neugebauer G., Scoville N. Z., 1988, ApJ, 325, 74

Sanz-Forcada J., Maggio A., Micela G., 2003, A&A, 408, 1087

Saulder C., van Kampen E., Chilingarian I. V., Mieske S., Zeilinger W. W., 2016, A&A, 596, A14

Schartmann M., Wada K., Prieto M. A., Burkert A., Tristram K. R. W., 2014, MNRAS, 445, 3878

Schlafly E. F., Meisner A. M., Green G. M., 2019, ApJS, 240, 30

Schmidt M., 1963, Nature, 197

Schmidtke P. C., Cowley A. P., Crane J. D., Taylor V. A., McGrath T. K., Hutchings J. B., Crampton D., 1999, AJ, 117, 927

Schumacher G., Mondaca E., Warner M., Martinez M., Estay O., Abbott T. M. C., 2010, in Radziwill N. M., Bridger A., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7740, Software and Cyberinfrastructure for Astronomy. p. 77402H, doi:10.1117/12.857986

Seale J. P., Looney L. W., Chu Y.-H., Gruendl R. A., Brandl B., Chen C. H. R., Brandner W., Blake G. A., 2009, ApJ, 699, 150

Sekiguchi K., Kilkenny D., Winkler H., Doyle J. G., 1989, MNRAS, 241, 827

Sewiło M., et al., 2018, AJ, 853, L19

Seyfert C. K., 1943, ApJ, 97

Shafter A. W., 2013, AJ, 145, 117

Shaw R. A., Stanghellini L., Mutchler M., Balick B., Blades J. C., 2001, ApJ, 548, 727

Shtykovskiy P., Gilfanov M., 2005, A&A, 431, 597

Skowron D. M., et al., 2020, arXiv e-prints, p. arXiv:2006.02448

Skrutskie M. F., et al., 2006, AJ, 131, 1163

Slee O. B., Haynes R. F., Wright A. E., 1984, MNRAS, 208, 865

Slee O. B., Nelson G. J., Stewart R. T., Wright A. E., Innis J. L., Ryan S. G., Vaughan A. E., 1987, MNRAS, 229, 659

Smith H. J., Hoffleit D., 1963, Nature, 198

Smolčić V., et al., 2009, ApJ, 696

Srinivasan S., Boyer M. L., Kemper F., Meixner M., Sargent B. A., Riebel D., 2016, MNRAS, 457, 2814

Stalevski M., Fritz J., Baes M., Nakos T., Popović L. Č., 2012, MNRAS, 420, 2756

Stalevski M., Ricci C., Ueda Y., Lira P., Fritz J., Baes M., 2016, MNRAS, 458, 2288

Steinhardt C. L., Weaver J. R., Maxfield J., Davidzon I., Faisst A. L., Masters D., Schemel M., Toft S., 2020, ApJ, 891, 136

Stern D., Eisenhardt P., Gorjian V., Kochanek C., 2005a, ApJ, 631

Stern D., et al., 2005b, ApJ, 631, 163

Stern D., Kirkpatrick J., Allen L., Bian C., Blain A., 2007, ApJ, 663

Stern D., et al., 2012, ApJ, 753, 30

Sturm R., et al., 2013, A&A, 558, A3

Subrahmanyan R., Saripalli L., Hunstead R. W., 1996, MNRAS, 279, 257

Szczygieł D. M., Socrates A., Paczyński B., Pojmański G., Pilecki B., 2008, AcA, 58, 405

Taylor M. B., 2005, in Shopbell P., Britton M., Ebert R., eds, Astronomical Society of the Pacific Conference Series Vol. 347, Astronomical Data Analysis Software and Systems XIV. p. 29

Tazaki R., Ichikawa K., 2020, ApJ, 892, 149

Tchekhovskoy A., 2015, The Formation and Disruption of Black Hole Jets, Astrophysics and Space Science Library, 414

Terrell J., 1967, ApJ, 147

Thilker D. A., Bianchi L., Simons R., 2014, in American Astronomical Society Meeting Abstracts #223. p. 355.11

Tody D., 1986, in Crawford D. L., ed., Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 627, Instrumentation in astronomy VI. p. 733, doi:10.1117/12.968154

Tody D., 1993, in Hanisch R. J., Brissenden R. J. V., Barnes J., eds, Astronomical Society of the Pacific Conference Series Vol. 52, Astronomical Data Analysis Software and Systems II. p. 173

Tombesi F., Cappi M., Reeves J., Palumbo G., Braito V., Dadina M., 2011, ApJ, 742

Torres C. A. O., Quast G. R., da Silva L., de La Reza R., Melo C. H. F., Sterzik M., 2006, A&A, 460, 695

Trakhtenbrot B., Netzer H., 2012, MNRAS, 427, 3081

Treister E., et al., 2009, ApJ, 693, 1713

Tremaine S., et al., 2002, ApJ, 574, 740

Tristram K. R. W., Burtscher L., Jaffe W., Meisenheimer K., Hönig S. F., Kishimoto M., Schartmann M., Weigelt G., 2014, A&A, 563, A82

Urry C. M., Padovani P., 1995, PASP, 107, 803

Valdes F., Gupta R., Rose J. A., Singh H. P., Bell D. J., 2004, ApJS, 152, 251

Vaughan A. E., Large M. I., 1986, Proceedings of the Astronomical Society of Australia, 6, 319

Vaughan A. E., Large M. I., 1987, Proceedings of the Astronomical Society of Australia, 7, 42

Vavilova I. B., Dobrycheva D. V., Vasylenko M. Y., Elyiv A. A., Melnyk O. V., Khramtsov V., 2021, A&A, 648, A122

Vestergaard M., Peterson B. M., 2006, ApJ, 641

Viaene S., et al., 2020, A&A, 638, A150

Walborn N. R., et al., 2014, A&A, 564, A40

Wang Y., et al., 2022, ApJ, 930, 38

Webb N. A., et al., 2020, A&A, 641, A136

Wegner G., et al., 2003, AJ, 126, 2268

Weis K., 2003, A&A, 408, 205

Wendker H. J., 1995, A&AS, 109, 177

Wenger M., et al., 2000, A&AS, 143, 9

Werner M., Roellig T., Low F., Rieke G., Rieke M., Hoffmann W., 2004a, ApJS, 154

Werner M. W., et al., 2004b, ApJS, 154, 1

White G. L., Jauncey D. L., Batty M. J., Peters W. L., Gulkis S., 1988, PASP, 100, 825

White S. V., et al., 2020a, PASA, 37, e017

White S. V., et al., 2020b, PASA, 37, e018

Whiting M. T., 2020, in Ballester P., Ibsen J., Solar M., Shortridge K., eds, Astronomical Society of the Pacific Conference Series Vol. 522, Astronomical Data Analysis Software and Systems XXVII. p. 469

Wolf C., et al., 2018, PASA, 35, e010

Wolter U., Czesla S., Fuhrmeister B., Robrade J., Engels D., Wieringa M., Schmitt J. H. M. M., 2014, A&A, 570, A95

Woods P. M., et al., 2011, MNRAS, 411, 1597

Wright E. L., et al., 2010, AJ, 140, 1868

Xue Y. Q., et al., 2010, ApJ, 720, 368

Yang G., et al., 2020, MNRAS, 491, 740

Yang G., et al., 2022, arXiv e-prints, p. arXiv:2201.03718

Zhang X., Feng Y., Chen H., Yuan Q., 2020, ApJ, 905, 97

Zhuang M.-Y., Ho L. C., Shangguan J., 2018, ApJ, 862, 118

Zivkov V., et al., 2018, A&A, 620, A143

Zivkov V., et al., 2020, MNRAS, 494, 458

van Aarle E., van Winckel H., Lloyd Evans T., Ueta T., Wood P. R., Ginsburg A. G., 2011, A&A, 530, A90

van Gelder M. L., et al., 2020, A&A, 636, A54

van Loon J. T., Sansom A. E., 2015, MNRAS, 453, 2341

van Loon J. T., et al., 1998, A&A, 329, 169

van Loon J. T., Zijlstra A. A., Groenewegen M. A. T., 1999a, A&A, 346, 805

van Loon J. T., Groenewegen M. A. T., de Koter A., Trams N. R., Waters L. B. F. M., Zijlstra A. A., Whitelock P. A., Loup C., 1999b, A&A, 351, 559

van Loon J. T., Zijlstra A. A., Bujarrabal V., Nyman L. Å., 2001, A&A, 368, 950

van Loon J. T., Cioni M. R. L., Zijlstra A. A., Loup C., 2005, A&A, 438, 273

van Loon J. T., Marshall J. R., Cohen M., Matsuura M., Wood P. R., Yamamura I., Zijlstra A. A., 2006, A&A, 447, 971

van Loon J. T., Cohen M., Oliveira J. M., Matsuura M., McDonald I., Sloan G. C., Wood P. R., Zijlstra A. A., 2008, A&A, 487, 1055

van Loon J. T., et al., 2010, AJ, 139, 68

van der Maaten L., Hinton G., 2008, Journal of Machine Learning Research, 9, 2579